2010

Enabling Responsible Public Genomics

John M. Conley
University of North Carolina School of Law, jmconley@email.unc.edu

Adam K. Doerr

Daniel B. Vorhaus

Publication: *Health Matrix*

ENABLING RESPONSIBLE PUBLIC GENOMICS

John M. Conley, Adam K. Doerr, and Daniel B. Vorhaus[†]

ABSTRACT

As scientific understandings of genetics advance, researchers require increasingly rich datasets that combine genomic data from large numbers of individuals with medical and other personal information. Linking individuals' genetic data and personal information precludes anonymity and produces medically significant information – a result not contemplated by the established legal and ethical conventions governing human genomic research. To pursue the next generation of human genomic research and commerce in a responsible fashion, scientists, lawyers, and regulators must address substantial new issues, including researchers' duties with respect to clinically significant data, the challenges to privacy presented by genomic data, the boundary between genomic research and commerce, and the practice of medicine.

This Article presents a new model for understanding and addressing these new challenges – a "public genomics" premised on the idea that ethically, legally, and socially responsible genomics research requires openness, not privacy, as its organizing principle. Responsible public genomics combines the data contributed by informed and fully consenting information altruists and the research potential of rich datasets in a genomic commons that is freely and globally available. This Article examines the risks and benefits of this public genomics

---

[†] John M. Conley is a William Rand Kenan, Jr. Professor of Law at the University of North Carolina, Chapel Hill and of counsel to the law firm of Robinson, Bradshaw & Hinson, P.A. Adam K. Doerr and Daniel B. Vorhaus are attorneys at Robinson, Bradshaw, publisher of the Genomics Law Report, http://www.genomicslawreport.com/. Robinson, Bradshaw has represented the Personal Genome Project and PersonalGenomes.org, and Daniel B. Vorhaus currently serves as an ELSI advisor to the Personal Genome Project, including with respect to its informed consent agreements.

The authors wish to thank participants in a seminar at Duke University's Center for Genome Ethics, Law & Policy, at the Genetics & Ethics in the 21st Century conference on Genomics & Personalized Medicine, and at the Consumer Genetics Show in Boston for their helpful comments.

model in the context of an ambitious genetic research project currently under way – the Personal Genome Project. This Article also (i) demonstrates that large-scale genomic projects are desirable, (ii) evaluates the risks and challenges presented by public genomics research, and (iii) determines that the current legal and regulatory regimes restrict beneficial and responsible scientific inquiry while failing to adequately protect participants. The Article concludes by proposing a modified normative and legal framework that embraces and enables a future of responsible public genomics.

TABLE OF CONTENTS

INTRODUCTION

In 1909, Danish biologist Wilhelm Johanssen suggested the term "gene" to describe the (then-conceptual) mechanism by which inherited traits passed from parent to offspring. A century's worth of research has revealed that the gene – a stretch of both coding and non-coding DNA sequences – is indeed the primary biochemical mechanism of inheritance. This research has also revealed that the traditional understanding of genetic inheritance as a direct linkage between a single gene and a single trait is, for the vast majority of traits, simplistic and incomplete.[1] Instead, most gene-trait relationships appear to be complex and frequently inexplicable by today's knowledge,[2] requiring geneticists to continue to attempt to unravel the tangled pathways between genes and traits by developing new research, theories, and analytical tools.[3] Many geneticists look forward to a second century of genomic research that they hope will enable personalized medical guidance to be offered to patients and consumers, based on genetic factors, long before the manifestation of disease.

Genomics thrives on data. Data sets that combine genomic and trait data, including medical histories and environmental measurements, are the lifeblood of genomic research. To continue exploring the pathways by which a genome[4] is ultimately manifest as a unique human being, researchers require increasingly rich and large-scale datasets. Until quite recently, amassing such datasets was impossible because of inadequate technology and prohibitive costs. The original

---

[1] *See, e.g.*, Jose L. Badano & Nicholas Katsanis, *Beyond Mendel: An Evolving View of Human Genetic Disease Transmission*, 3 NATURE REVIEWS GENETICS 779, 779 (2002).

[2] *See, e.g.*, David Botstein & Neil Risch, *Discovering Genotypes Underlying Human Phenotypes: Past Successes for Mendelian Disease, Future Approaches for Complex Disease*, 33 NATURE GENETICS 228 (2003).

[3] In addition to traditional Mendelian inheritance, research is investigating the role of common and rare genetic variants, coding and non-coding regions and structural variation within the genome, and a host of other factors.

[4] The term "genome" refers to the entire human genome, a copy of which is contained within the nucleus of every cell in the body. The human genome is comprised of roughly 3 billion base pairs. The term "gene," which refers to an identified portion of the genome, consists of a stretch of protein-coding and non-coding regions. Genes vary greatly in size, and there are thought to be 20,000 to 25,000 genes in the genome, which represents approximately one and a half percent of the entire human genome. For purposes of this Article, we use the term "genome" to refer to research and commerce that investigates the structure and function of human genomic data including – but not limited to – the role that genes and genetic inheritance plays in determining human traits.

effort to sequence a complete human genome took more than a decade and cost several billion dollars. In recent years, however, the cost of genomic sequencing has declined precipitously, spurred by scientific and technological advances that seem likely to render genomic sequencing a commodity within the next ten years, if not much sooner.[5] This decline in genomic sequencing costs means that researchers will soon be able to generate the massive amounts of information that large-scale genomic research requires. Developing the technological capacity to generate these rich datasets at affordable costs overcomes the scientific obstacle, but it inevitably results – as have other disruptive advances, from railroads to the Internet – in the creation of new legal and social challenges. This Article focuses on three such challenges: the difficulty of guaranteeing privacy, the clinical significance of genomic information, and the difficulty of obtaining truly "informed" consent.

By combining individuals' genomic data with their personal information – medical history, physical traits, diet, lifestyle, and environmental exposures – researchers will be able to create integrated datasets to analyze the complete range of factors that contribute to complex human traits and conditions. As the American Society of Human Genetics has recognized, however, such datasets implicate substantial privacy issues.[6] Correlating genotypic and phenotypic information in the same database creates "an accurate and unique identifier"[7] that makes it difficult to guarantee participants that their data will remain private. Further, longitudinal studies, which enable researchers to track participant traits, exposures, and outcomes over time, require that researchers must be able to identify and interact with participants to update personal and health information. Thus, unlike many human research projects, where the nature of the research permits or even requires[8] participant anonymity, the structural features

---

[5] *See* MARK GUYER & ADAM FELSENFELD, NAT'L HUMAN GENOME RESEARCH INST., NAT'L INSTS. OF HEALTH, THE FUTURE OF GENOME SEQUENCING, http://www.genome.gov/27529702 (last visited May 3, 2010).

[6] Joann Boughman, Am. Soc'y of Human Genetics, ASHG Response to NIH on Genome-Wide Association Studies, Nov. 30, 2006, http://www.ashg.org/pages/statement_nov3006.shtml.

[7] *Id.*

[8] For example, clinical trials for the development of many pharmaceutical products utilize double-blind studies, "[a] clinical trial design in which neither the participating individuals nor the study staff knows which participants are receiving the experimental drug and which are receiving a placebo (or another therapy). Double-blind trials are thought to produce objective results, since the expectations of the doctor and the participant about the experimental drug do not affect the outcome . . . ." ClinicTrials.gov, Glossary of Clinical Trials Terms, http://clinicaltrials.gov/ct2/

and the success of large-scale and long-term genomic research dictate a different, less anonymous model.

Complicating matters further, an individual's genomic information often contains clinically actionable information – information that could lead subjects to seek specific testing or treatment based on their genetic makeup. Not only is the significance of this genomic data, and the appropriate clinical response, often not well-understood by patients and providers alike, it is constantly evolving as new scientific discoveries are made and new gene-trait associations are advanced. As a result, participants may feel pressed to act on genomic data in the face of incomplete information as to the advisability or efficacy of the treatment they pursue (or forego).

Similarly, researchers must make difficult decisions regarding how much and what type of information to share with participants. Should researchers be obliged to share everything? Should there be some sort of cutoff to distinguish information with established medical value from preliminary findings? If so, who should determine where the line falls? Regardless of how the balance is struck, some participants will be dissatisfied – to say nothing of the challenge presented in interpreting such data, including where the burden of interpretation should fall.

Finally, ensuring that subjects provide truly informed consent to participate in large-scale genomic research projects presents serious challenges. A deep understanding of the underlying science of genomics is critical to informed consent – a high bar. Finding sufficient numbers of knowledgeable participants,[9] although challenging, is not even the most significant issue related to informed consent. The greater difficulty lies in the genomic unknowns, as it is impossible to inform participants about benefits, risks, and implications that the researchers themselves do not fully understand. What does it mean to require an individual to provide informed consent when the significance of genomic data changes with each new discovery?

These challenges of genomic research – the inability to guarantee privacy, the clinical significance of genomic information, and the difficulty of obtaining truly informed consent – must be resolved to pursue the next generation of human genomic research and commerce in an ethically and legally responsible fashion. This Article offers a

---

info/glossary (last visited May 3, 2010).

[9] The challenge is to locate not only a sufficient number of participants that have an adequate understanding of genomics in order to provide informed consent, but to do so in a way that produces a representative sample population. A research cohort consisting entirely of English-speaking genetics professors will be of limited value for obvious reasons.

novel approach to the host of emerging legal and ethical questions surrounding the next generation of genomic research: a revised normative and legal framework for the pursuit of "public genomics" that is premised on the idea that ethically, legally, and socially responsible genomic research requires openness, not privacy, as its organizing principle. To be both responsible and effective, large-scale genomic research should take place in full view of the public and the law, adopting openness in risk disclosure and informed consent, extensive dialogue and data-sharing among researchers and participants, and thoughtful self-regulation.

This Article details how public genomics projects may be conducted in an ethically, legally, and socially responsible fashion and what features of the current legal, regulatory, and normative landscape may need to be adjusted in order to enable such research. This Article also examines the risks and benefits of the public genomics model, using as an example the Personal Genome Project (PGP), a project sponsored by Harvard Medical School which aims to establish a publicly available, fully consented dataset containing comprehensive genotypic, phenotypic, and personal information of 100,000 participants.

Part I reviews the present state of genomic science and research, concludes that large-scale genomic research projects are achievable and desirable, and delineates the core characteristics of the "public genomics" model. Part II delves deeper into the risks and challenges presented by public genomics research, including: (i) the feasibility and desirability of conducting anonymized or confidential genomics research; (ii) the implications of informed consent given a rapidly evolving knowledge base; and (iii) the difficulty of distinguishing between research and clinical activities. Part III assesses the current legal and regulatory scheme applicable to public genomics and determines that it is presently inadequate to meet the objective of protecting individual research participants while simultaneously enabling responsible and promising scientific inquiry. Finally, the Article articulates the key principles and characteristics of a new legal and normative regime that would advance these objectives far more effectively than the current morass of state and federal statutes, regulations, and common law decisions.

I. THE PATH TO PUBLIC GENOMICS

The path to public genomics began even before the official launch of the Human Genome Project in 1990, an extraordinarily ambitious and publicly funded attempt to develop and publish a complete roadmap of the human genome. The widely publicized – and, at the time, widely criticized – venture saw an international consortium of gov-

ernment agencies and research centers[10] produce a "rough draft" of the consensus human genome in 2000, followed by a "final draft" in 2003. This historic achievement – the genomics moonshot – is well chronicled,[11] and the sheer magnitude of the undertaking is often conveyed by reference to the tremendous time and budget of the project: more than a decade and nearly three billion dollars.[12]

The pace of technological and scientific innovation has significantly accelerated since the Human Genome Project's completion, enabling the construction of several large-scale human genomic databases designed to build upon the Human Genome Project's work.[13]

---

[10] *See* Nat'l Human Genome Research Inst., Nat'l Insts. of Health, The Human Genome Project Completion: Frequently Asked Questions, http://www.genome.gov/11006943 (last visited May 3, 2010).

[11] *See generally* KEVIN DAVIES, CRACKING THE HUMAN GENOME: INSIDE THE RACE TO UNLOCK HUMAN DNA (2001) (providing an historical explanation of the human genome project); JAMES SHREEVE, THE GENOME WAR: HOW CRAIG VENTER TRIED TO CAPTURE THE CODE OF LIFE AND SAVE THE WORLD (2004).

[12] *See* Nat'l Human Genome Research Inst., *supra* note 10. For a more detailed budget and timeline, see Human Genome Project Info., Human Genome Project Budget, http://www.ornl.gov/sci/techresources/Human_Genome/project/budget.shtml (last visited May 3, 2010). Less attention-grabbing is the fact that, despite significant opposition and a reliance on technologies that were far from fully developed when the project commenced, the Human Genome Project was completed ahead of schedule and under budget, the result of an unexpectedly rapid development in the underlying genomic sequencing science and technology. *See* Nat'l Insts. of Health, Dep't of Health & Human Servs., Human Genome Project: Fact Sheet, http://www.nih.gov/about/researchresultsforthepublic/HumanGenomeProject.pdf (last visited May 3, 2010).

[13] The rate of change has not slowed in the months following the completion of this Article. Since this article was written, sequencing costs have further declined and commercial and clinical offerings have continued to expand. For instance, between the summer of 2009 and the spring of 2010, (i) the number of genetic tests for diseases reported by GeneTests rose from 1,790 to 1,975 and (ii) the cost of commercial whole-genome sequencing continued to decline, with Illumina and Life Technologies introducing new sequencing machines that offer whole-genomes at or below the $10,000 threshold (this is a "materials only" cost). *See infra* note 29; *see also* GeneTests, Growth of Laboratory Directory, http://www.ncbi.nlm.nih.gov/projects/GeneTests/static/whatsnew/labdirgrowth.shtml (last visited May 3, 2010); *see* Dan Vorhaus, *Life Technologies Fires Latest Sequencing Salvo*, GENOMICS L. REP., http://www.genomicslawreport.com/index.php/2010/01/28/life-technologies-fires-latest-sequencing-salvo/ (last visited May 3, 2010). Nor have genomic research projects remained static. As discussed below, the PGP has submitted a revised protocol and informed consent agreement reflecting comments from IRB members, participants and other advisors, including this Article's authors. *See infra* note 97. Despite the rate and breadth of changes in the scientific and commercial areas of genomics, the social, legal, and regulatory challenges addressed by this Article – particularly the expanding role of public genomic research and the need to conduct such research in a responsible fashion – remain substantively unchanged.

International efforts, including the International HapMap Project[14] and the 1000 Genomes Project,[15] continue to dramatically increase the level of detail and sensitivity of available human genome reference data, expanding upon the limited consensus genome produced by the Human Genome Project. Other projects, such as the UK Biobank,[16] are attempting to categorize genomic variation within geographically delimited populations.[17] Still others aggregate the results from targeted genome-wide association studies that seek the genetic bases of specific diseases and other complex human traits.[18] A host of new

---

[14] *See* Nat'l Ctr. Biotech. Info., U.S. Dep't Health & Human Servs., International HapMap Project, http://www.hapmap.org (last visited May 3, 2010). Begun in 2002, the International HapMap Project produced a haplotype map of the human genome. Haplotypes refer to sections of DNA on particular chromosomes that are typically inherited in blocks. This map helped produce a rough catalogue of common human genetic variation (those occurring in at least five percent of the population). *Id.*

[15] The 1000 Genomes Project, launched in 2008, builds on the International HapMap Project and aims to establish a much more detailed catalogue (variations occurring in at least one percent of the population) of human genetic variation by sequencing complete genomes for approximately twelve hundred individuals, with the resulting genome map to be "made swiftly available to the worldwide scientific community through freely accessible public databases." 1000 Genomes Project: A Deep Catalog of Human Genetic Variation, Project Overview, http://www.1000genomes.org/page.php?page=about (last visited May 3, 2010).

[16] Biobank, Welcome to UK Biobank, http://www.ukbiobank.ac.uk/ (last visited May 3, 2010).

[17] A number of other regional or national genomic biobanks have been proposed, and several have even proceeded to the point of producing data. *See, e.g.*, Henry T. Greely, *The Uneasy Ethical and Legal Underpinnings of Large-Scale Genomic Biobanks*, 8 ANN. REV. GENOMICS & HUM. GENETICS 343, 348 (2007).

[18] Both the Human Genome Epidemiology Network (HuGENet), part of the Centers for Disease Control and Prevention's (CDC) Office of Public Health Genomics and the National Human Genome Research Institute's (NHGRI) Office of Population Genomics maintain databases tracking published GWA studies, with HuGENet reporting 335 GWA studies since 2001 and NHGRI identifying 306 as of May 6, 2009. Nat'l Human Genome Research Inst., Nat'l Insts. of Health, Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Disease and Traits, http://www.genome.gov/gwastudies/ (last visited May 3, 2010). These numbers are rising almost daily. *See* Ctrs. for Disease Control & Prevention, Human Genome Epidemiology Network, http://www.cdc.gov/genomics/hugenet/default.htm (last visited May 3, 2010); Nat'l Human Genome Research Inst., Nat'l Insts. of Health, A Catalog of Published Genome-Wide Association Studies, http://www.genome.gov/26525384 (last visited May 3, 2010); HuGE Navigator, HuGE Literature Finder, http://www.hugenavigator.net/HuGENavigator/searchSummary.do?firstQuery=gwa&publitSearchType=now&whichContinue=firststart&check=n&dbType=publit&Mysubmit=go (last visited May 3, 2010) (HuGENet reporting 335 GWA studies since 2001).

large-scale genomic research projects recently established or under consideration[19] indicates that the rapid pace of development will continue.

The proliferation of large-scale genomic databases exhibits the characteristics of a simple supply and demand relationship: researchers demand ever-richer datasets to solve genomic puzzles while a burgeoning industry of genomic sequencing hardware and service providers compete to supply the tools and technologies needed to generate those datasets.

For genomic researchers, the identification of the genetic bases for certain traits was underway prior to the Human Genome Project's

---

In addition, the National Center for Biotechnology Information (NCBI) hosts the Database of Genotypes and Phenotypes (dbGaP), which serves as a central point of research access for certain research that investigates genotype and phenotype interaction. NCBI, dpGaP: Genotype and Phenotypes, http://www.ncbi.nlm.nih.gov/gap (last visited May 3, 2010). Although previously dbGaP data was made available in two tiers, with aggregate study data provided as "open-access" to the general public and individual, de-identified study data available only as "controlled-access data" to authorized users, the NCBI removed its "open access" data in September of 2008 due to re-identification concerns, restricting all dbGaP study data to authorized researchers. *See* discussion *infra* Part II. *See* Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays*, PLOS GENETICS, Aug. 2008, at 1, 4; Catherine Clabby, *DNA Research Commons Scaled Back*, 97 AM. SCIENTIST 113, 113 (2009), *available at* http://www.americanscientist.org/issues/num2/dna-research-commons-scaled-back/1.

In response to these events other major institutions, including the Wellcome Trust, and the Broad Institute, also removed pooled or aggregate human DNA data from their websites. *See* Jennifer Couzin, *Genetic Privacy. Whole-Genome Data Not Anonymous, Challenging Assumptions*, 321 SCIENCE 1278, 1278 (2008).

[19] In addition to the Personal Genome Project, large-scale genomic research projects are being conducted with support of research institutions, health care providers, commercial testing companies and federal, state, and local government organizations. For background information on the Personal Genome Project, see Pers. Genome Project, Project Overview, http://www.personalgenomes.org/project.html (last visited May 3, 2010). For examples of large-scale genomic research, see, e.g., Nat'l Human Genome Research Inst., Nat'l Insts. of Health, ClinSeq: A Large-Scale Medical Sequencing Clinical Research Pilot Study, http://www.genome.gov/25521305 (last visited May 3, 2010); Coriell Personalized Med. Collaborative, About the Study, http://cpmc.coriell.org/Sections/About/?SId=9 (last visited May 3, 2010); Navigenics, Overview, http://www.navigenics.com/visitor/about_us/ (last visited May 3, 2010); 23andMe, The 23andMe Research Revolution, https://www.23andme.com/researchrevolution/overview/ (last visited May 3, 2010); The Genome Ctr. at Wash. Univ., The Cancer Genome Atlas, http://genome.wustl.edu/highlights/article/the_cancer_genome_atlas (last visited May 3, 2010); M.U.R.D.O.C.K. Study, The MURDOCK Study: Measurement to Understand the Reclassification of Disease of Cabarrus/Kannapolis, https://www.murdock-study.com/ (last visited May 3, 2010).

completion, particularly for those traits, like Huntington's disease, where family histories demonstrated a high degree of heritability and strongly suggested a single gene-trait relationship.[20] These monogenic traits – traits defined by a single gene – were the first to have their genetic underpinnings clearly identified. It is now widely believed that the majority of this "low-hanging fruit" has already been plucked.

Although public resources including the Human Genome Project, the International HapMap Project, and the 1000 Genomes Project continue to produce increasingly refined maps of human genomic variation, there is a growing recognition that a host of factors beyond common genetic variation, including rare genetic variants, copy number variation (CNV) and other rearrangements of the genome and epigenetic or environmental factors, are key contributors to many complex traits, demonstrating the need for richer datasets that include whole-genome sequence data.[21] The challenge now is to characterize and understand the genomic underpinnings of more complex phenotypes, such as height or heart disease. Although some progress has been made in this area,[22] the reality is that the genotype-phenotype associations identified to date account for only a very small portion of the exhibited population variance for nearly every complex trait.[23]

For example, although genomic research has identified more than fifty regions of the genome that contribute to height, a recent study found that those genetic loci were able to predict only four to six per-

---

[20] In the case of Huntington's for instance, by using DNA marking methods researchers had already determined the general location of the Huntington's gene by 1983; their work served as a departure point for the broader work conducted by the Human Genome Project. *See* James F. Gusella et al., *A Polymorphic DNA Marker Genetically Linked to Huntington's Disease*, 306 NATURE 234, 234-35 (1983); *see also* John Hardy & Andrew Singleton, *Genomewide Association Studies and Human Disease*, 360 NEW ENG. J. MED. 1759, 1764 (2009) ("The vast majority of success in defining genetic risk in disease has been a result of traditional gene-hunting efforts to find mutations that underlie monogenic diseases. . . . In considering the genetics of complex disease and particularly the role of common variants that affect expression, a more nuanced perspective is useful.").

[21] *See* David B. Goldstein, *Common Genetic Variation and Human Traits*, 360 NEW ENG. J. MED. 1696, 1698 (2009); Mildred K. Cho, *Translating Genomics Into the Clinic: Moving to the Post-Mendelian World*, 1 GENOME MED. 7.1 (2009); Kelly A. Frazer et al., *Human Genetic Variation and Its Contribution to Complex Traits*, 10 NATURE REV. GENETICS 241 (2009).

[22] *See, e.g.*, Joel N. Hirschhorn, *Genomewide Association Studies – Illuminating Biologic Pathways*, 360 NEW ENG. J. MED. 1699 (2009) (discussing recent developments that lead to a potential path to gene discovery for polygenic diseases and traits).

[23] *See, e.g.*, Goldstein, *supra* note 21, at 1696; Nicholas Katsanis, *From Association to Causality: The New Frontier for Complex Traits*, 1 GENOME MED. 23, 23-23.2 (2009); Cho, *supra* note 21, at 7.2.

cent of the variance in individual height, while a much less sophisticated (and substantially less expensive) predictive model developed by Sir Francis Galton in 1886, which relies solely on knowing the heights of an individual's parents, was able to predict forty percent of that same variance.[24] Accordingly, to identify the genomic factors that meaningfully contribute to complex phenotypes such as height, researchers require datasets that allow them to peer deeply into individual genomes in search of rare genetic modifications and other sources of variation and test their hypotheses against large cohorts of both similar and dissimilar populations.[25]

Not to be overlooked is the increasing demand from non-research users of genomic sequence data, including medical practitioners and individuals seeking sequence data for clinical and recreational purposes. Although these groups have not traditionally demanded sequence data on a large scale, the number of genomic data providers continues to increase.[26] Direct-to-consumer (DTC) sequencing companies are already providing whole-genome sequencing[27] and facilitating large-

---

[24] *See* Yurii S. Aulchenko et al., *Predicting Human Height by Victorian and Genomic Methods*, 17 EUR. J. HUM. GENETICS 1070 (2009). Depending on your perspective, the news is either better or worse for other traits. Eye color, for example, is a complex trait insofar as it is influenced by a number of genes. Unlike height, however, eye color is not a continuum trait (i.e., there a finite number of phenotypes that will be expressed). A recent paper published a method for predicting eye color from genotype, with between seventy percent and ninety percent accuracy depending on the color to be predicted. That prediction rate is certainly better than in the height example, but it is limited to (a) "Dutch people of European ancestry" and (b) a phenotypically-simple trait (i.e., eye color is not only a discrete trait but it is also, unlike cancer or diabetes, easy to diagnose/observe and is fully penetrant because everyone's eyes have color). *See* Cheryl Jones, *Genetic Test Predicts Eye Color*, NATURE.COM, Mar. 9, 2009, http://www.nature.com/news/2009/090309/full/news.2009.145.html.

[25] *See, e.g.*, Goldstein, *supra* note 21, at 1697-98 (discussing how genome-wide association studies may benefit from the study of rare variants); Hirschhorn, *supra* note 22, at 1699 (explaining that genomewide association studies are effective at "[d]iscovery of biologic pathways underlying polygenic diseases and traits").

[26] As of July 2, 2009, GeneTests, a publicly funded medical genetic information resource, reported a total of 1,171 clinics and 605 laboratories conducting tests for a total of 1,790 diseases. *See* NCBI, Welcome to GeneTests, http://www.ncbi.nlm.nih.gov/sites/GeneTests/?db=GeneTests (last visited May 3, 2010). In addition, an expanding group of genomic service providers, including, most prominently, 23andMe, Navigenics, deCODEme, DNADirect, Pathway Genomics, Knome, and Illumina, are offering their services – frequently single nucleotide polymorphism (SNPs) genotyping, with the exception of Knome and Illumina, which provide whole-genome sequencing – directly to consumers and/or to healthcare providers. These DTC providers market their services for a variety of uses ranging from the exploration of an individual's genealogical history to a personalized disease risk assessment to a form of clinical diagnostic or preventative care.

[27] At the end of 2008, the least expensive commercially available whole-

scale genomic research,[28] work that will both meet and create demand for the next-generation sequencing that public genomics researchers require.

On the supply side, an expanding number of companies[29] are competing to meet genomic researchers' demands for increasingly rich datasets at grant-friendly prices. Persistent and rapid technologi-

---

genome sequence was provided by Knome, and came with a $350,000 price tag. Since then, Knome lowered its price to $99,000, and held an eBay charity auction in which the starting price for a whole-genome sequence was $68,000. Genomic sequencing platform provider Illumina also entered the consumer market in June, 2009, offering a $48,000 whole-genome sequence. The service imposes several requirements, however, as it must be ordered in consultation with a physician and individuals must turn to a third party – including any of several other DTC genomic service providers - for interpretive guidance. *See* Knome, Inc., X PRIZE Foundation and Knome Spotlight DNA Day with Historic Charity Auction of Whole Genome Sequence (2008), http://www.knome.com/about/news/20090423.html; Rachel Lehmann-Haupt, *$99,000 to See Your Future?*, MSN MONEY, Apr. 10, 2009, http://articles.moneycentral.msn.com/Investing/HomeMortgageSavings/99000-to-see-your-future.aspx#pageTopAchor; Kevin Davies, *Illumina to Offer $48,000 Personal Genome Sequencing Service*, BIO-IT WORLD.COM, June 11, 2009, http://www.bio-itworld.com/news/06/11/09/illumina-personal-genome-sequencing-service.html. Other commercial sequencers include CoFactor Genomics. *See* CoFactor Genomics, New to Next-Gen?: Next Generation Sequencing in the Context of Today's Science, http://cofactorgenomics.com/next-gen/ (last visited May 3, 2010) (listing the cost of sequencing at $23,000). Not commercially available as of this writing but expected to be $5,000 and on sale in 2009. Julia Karow, *Complete Genomics to Offer $5,000 Human Genome as a Service Business in Q2 2009*, GENOMEWEB, Nov. 4, 2008, http://www.genomeweb.com/sequencing/complete-genomics-offer-5000-human-genome-service-business-q2-2009.

[28] Examples include the Scripps Genomic Health Initiative and 23andMe Parkinson's Community: Strength in Numbers, which aims to recruit 10,000 people to participate in a GWA study of Parkinson's disease by making 23andMe's genetic testing product available to eligible participants at a discounted price. *See supra* note 19; 23andMe, 23andMe Parkinson's Community: Strength in Numbers, https://www.23andme.com/pd/join/ (last visited May 3, 2010); *see also* Andrew Pollack, *Google Co-Founder Backs Vast Parkinson's Study*, N.Y. TIMES, Mar. 12, 2009, at B1.

[29] In the past several years, the traditional Sanger sequencing method has been replaced by "next-generation" commercial DNA sequencing platforms offered by Roche, Illumina, Applied Biosystems, and Helicos BioSciences. Even as these new sequencing techniques provide tremendous sequencing output at previously unimaginable resolution, speed and, most importantly, price, they are already being challenged by newer entrants to the market, including sequencing-as-a-service providers Knome and Complete Genomics and so-called "third-generation sequencers" Oxford Nanopore and Pacific Biosciences. The next wave of commercial sequencers – companies like NABsys and BioNanomatrix – are making still more astounding promises, like NABsys's goal of sequencing a genome "for less than $100 in under an hour." Ryan McBride, *NABsys Secures $4M First Round to Develop Electronic DNA Sequencing*, XCONOMY, May 5, 2009, http://www.xconomy.com/boston/2009/05/05/nabsys-secures-4m-first-round-to-develop-electronic-dna-sequencing/.

cal advances in DNA sequencing have driven the price of genomic sequencing down sharply over the past decade, with cost falling at a pace that outstrips standard models for rapid technological innovation.[30] These advances promise to transform genomic sequencing into a commodity, accessible and affordable for researchers, and even individuals.[31] As prices continue to fall, there seems little doubt that the era of the $1,000 genome,[32] and perhaps even the $100[33] genome, is nearly upon us.

---

[30] *See* GUYER & FELSENFELD, *supra* note 5 ("During this time, and especially in the past few years, sequencing costs have decreased faster than Moore's Law and sequencing capacity has increased at an ever-greater pace."); *see also* David Ewing Duncan, *Discount DNA*, PORTFOLIO.COM, Dec. 17, 2008, http://www.portfolio.com/views/columns/natural-selection/2008/12/17/Discount-DNA ("Right now, gene-testing companies will sell you a near-complete perusal of your own DNA for somewhere between $100,000 and $350,000. A year ago it cost $1 million. Five years ago, the cost of the first complete human genome was $2.7 billion." Duncan's pricing, of course, is already far out of date.).

[31] DTC genomic service companies continue to reduce the price of their services while expanding their offerings. DTC genotyping is available for as little as $99 and whole-genome sequencing is available for $48,000. *See supra* note 28. While some critics question the value of the genetic association information provided by even the most reputable DTC companies. *See, e.g.*, Grace Wong, *Why Screening Your Genes Is Big Business*, CNNHEALTH.COM, May 6, 2009, http://www.cnn.com/2009/HEALTH/05/05/consumer.genomics ("David Goldstein, head of the Center for Human Genome Variation at Duke University, believes that given the current knowledge of how genes influence disease, there is only a 'recreational' use for retail tests."). The field has nonetheless attracted several companies rushing eager to take advantage of consumer enthusiasm for genetic testing, including Atlas Sports Genetics, which offers a $149 test that "[a]ims to predict a child's natural athletic strengths" on the basis of a single gene. Juliet Macur, *Born to Run? Little Ones Get Test for Sports Gene,* N.Y. TIMES, Nov. 30, 2008, at A1.

Whole-genome sequencing is offered to consumers by Knome for $99,000 and Illumina for $48,000 and commercially by companies like CoFactor and Complete Genomics for around $20,000, with the promise of complete sequencing for a mere $5,000 in 2009. *See supra* note 27; *see also* Duncan, *supra* note 30; Erika Check Hayden, *Genome Sequencing: The Third Generation*, 457 NATURE 768 (2009); Daniel MacArthur, *Complete Genomics Back in Action After 6 Month Funding Delay*, GENETIC FUTURE, http://scienceblogs.com/geneticfuture/2009/08/complete_genomics_back_in_acti_1.php (last visited May 3, 2010).

[32] While the $1,000 genome presents a reasoned guess at the level where individual genomic sequencing will be cost-effective, it is far from a rigid threshold value above which truly personalized medicine will remain unattainable. *See, e.g.*, G. M. Church, *The Personal Genome Project,* MOLECULAR SYS. BIOLOGY 1, 1 (2005). Instead, "[t]he '$1,000 genome' has become shorthand for the promise of DNA-sequencing capability made so affordable that individuals might think the once-in-a-lifetime expenditure to have a full personal genome sequence . . . is worthwhile." G.M. Church, *Genomes for ALL*, SCI. AM. 47, 48 (2006); *see also* Nature Genetics, Question of the Year, http://www.nature.com/ng/qoty/index.html (last visited May 3,

Not surprisingly, this combination of prevalent but poorly understood complex traits and plummeting genomic sequencing costs has produced a new generation of research studies. This work seeks to untangle the connection between genotype and phenotype by collecting and categorizing comprehensive genomic and trait data of large numbers of individuals. In contrast to the reference data produced by the Human Genome Project, HapMap Project, and 1000 Genomes Project or genome-wide association studies that narrowly focus on specific traits, this next generation of genomic research aims to develop large and detailed databases of individual human genotypic and phenotypic variation. These complex databases will match individualized genomic data – likely whole-genome sequence data in the near future – with equally detailed phenotypic and environmental data, providing researchers more insight into the genomic and environmental bases of complex traits. These projects, of which the Personal Genome Project is the clearest example, represent in size, scope, and structure a new model of human genomic research – what we shall define and refer to throughout this Article as "public genomics."

The notion that genomic research should be "public" is hardly novel. The Human Genome Project was publicly funded, and its results are publicly available.[34] Subsequent milestone genomic research

---

2010) (compiling responses of prominent geneticists to the question, "What would you do if it became possible to sequence the equivalent of a full human genome for only $1,000?"); Julia Karow, *PGP to Publish Initial Data Sets Next Month as Church Predicts $1,000 Genome in 2009*, IN SEQUENCE, Sept. 23, 2008, http://www.genomeweb.com/sequencing/pgp-publish-initial-data-sets-next-month-church-predicts-1000-genome-2009.

[33] *See, e.g.*, ScienceDaily, The 15-Minute Genome: Faster, Cheaper Genome Sequencing on the Way, July 29, 2009, http://www.sciencedaily.com/releases/2009/07/090727191912.htm (describing Pacific Biosciences and its attempt to produce a $100 genome using single molecule real-time sequencing); McBride, *supra* note 29 (discussing NABsys's goal of a $100 genome in under an hour); Lauren Gravitz, *TR10: $100 Genome*, TECH. REV., Mar.-Apr. 2009, *available at* http://www.technologyreview.com/biomedicine/22112/ (describing BioNanomatrix's belief that it can reach the $100 genome target in five years).

[34] The HGP information was published in a database known as GenBank and has been annotated and packaged for public consumption by both public (e.g., University of California Santa Cruz) and public-private (e.g., Ensembl) entities. The UCSC browser incorporates data from the Encyclopedia of DNA Elements (ENCODE) Consortium, an international research collaborative funded by NHGRI. Ensembl is a joint collaboration of EMBL-EBI and the Wellcome Trust Sanger Institute (with additional funding from NHGRI, among other bodies) designed to automatically annotate the human genome sequence. *See* UCSC Genome Bioinformatics, About the UCSC Genome Bioinformatics Site, http://genome.ucsc.edu/ (last visited May 3, 2010); e!Ensembl, About the Ensembl Project, http://uswest.ensembl.org/info/about/intro.html (last visited May 3, 2010).

efforts have featured similar public elements, and many other projects receive public funding and deposit their data in widely accessible databases.[35] As used in this Article, however, "public" is given a more expansive meaning, embodying what we believe to be the crucial characteristics of the next generation of human genomic research projects.

Public genomics, as that term is used throughout, refers to projects that satisfy each of the following criteria:

Publicly Available Data. Raw participant data is made available to the public without charge and with minimal or no restrictions.[36] Research datasets that are subject to access controls (e.g., where only a limited number of authorized researchers are permitted to analyze the data)[37] fail to satisfy this criterion, as do privately operated commercial datasets.[38] Publishing data directly into the public domain permits unfettered access by Nobel Laureates and self-taught patients alike, allowing an unrestricted universe of researchers to use the data in seeking answers to an unlimited number of questions. Developing open-access datasets democratizes genomics research[39] and harnesses

---

[35] *See supra* note 19 (discussing HuGENet and dbGaP). Other examples include the National Institute of General Medical Sciences (NIGMS), which is providing grant money to create a Pharmacogenomics Knowledge Base (PharmGKB) that will "present complete, comprehensive, and current knowledge in pharmacogenomics, backed by critical datasets . . . and the most compelling literature." *See* U.S. DEP'T HEALTH & HUMAN SERVS., PHARMACOGENOMICS KNOWLEDGE BASE, PHARMGKB (R24), http://grants.nih.gov/grants/guide/rfa-files/RFA-GM-10-002.html (last visited May 3, 2010).

[36] The PGP, for example, publishes submitted participant data – both phenotypic and genotypic data – directly to a publicly accessible database. *See* Pers. Genome Project, http://www.personalgenomes.org/ (last visited May 3, 2010).

[37] The dbGaP data is a typical example of this. Similarly restricted genomic databases include the Genetic Association Information Network (GAIN) and the Wellcome Trust Case Control Consortium (WTCCC). *See* Jane Kaye et al., *Data Sharing in Genomics – Re-Shaping Scientific Practice,* 10 NATURE REVIEWS GENETICS 331, 333 (2009).

[38] A recent trend in the DTC genomics space has seen commercial service providers, including 23andMe, utilize crowd-sourcing techniques to recruit research participants. *See* 23andMe, The 23andMe Research Revolution, https://www.23andme.com/researchrevolution/ (last visited May 3, 2010). However, public participation in research, while important, is not sufficient to satisfy this criterion of public genomics. In order to do so, the data generated by the research must be provided in an open-source fashion. *See* 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/ (discussing distinction between crowd-sourcing and open-source).

[39] To make public genomics datasets truly accessible they will need to be coupled with appropriate interpretive tools, as with the genome browsers already

the wisdom of crowds to and encourages innovation and collaboration in building more useful datasets and conducting more collaborative and productive research.[40]

Linked Participant Data. Genotype and phenotype data supplied by participants are directly linked within the dataset. There are no promises, explicit or implicit, of anonymization, confidentiality, or any other form of de-identification. This explicit association of data enriches the core dataset and enables researchers to more directly involve participants in their research, including permitting the return of research results directly to participants.[41] This feature is also critically

---

offered by UCSC and Ensembl. *See* Kaye et al., *supra* note 37. With the expected proliferation of genomic sequence data driven by next-generation sequencing technology, there is a clear need for next-generation interpretive tools. *See* Daniel MacArthur, *Next-Generation Genome Browsers*, GENETIC FUTURE, http://scienceblogs.com/geneticfuture/2009/03/next-generation_genome_browser.php (May 3, 2010) (discussing visual and organizational methods in which genomic information can be conveyed). One early example is the GENESTAT information portal, which is a Wikipedia-style resource that focuses on epidemiological genetic association studies and collates key papers, websites and software. *See* Samuli Ripatti et al., *GENESTAT: An Information Portal for Design and Analysis of Genetic Association Studies*, 17 EUR. J. HUM. GENETICS 533 (2009).

[40] The concept of the wisdom of crowds originated in 1906 with the ubiquitous Francis Galton. He observed that 800 visitors to an agricultural fair had little individual luck estimating the weight of a butchered ox, but that the average of their individual estimates was almost exactly correct. JAMES SUROWIECKI, THE WISDOM OF CROWDS xi-xiv (2004). For contemporary examples in science and research, see, e.g., Pascal Hingamp et al., *Metagenome Annotation Using a Distributed Grid of Undergraduate Students*, 6 PLOS BIOLOGY 2362, *available at* http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0060296 (last visited May 3, 2010); Richard Van Noorden, Chemistry Paper Pwned by Live-Blog Experiments, NATURE.COM, July 27, 2009, http://blogs.nature.com/news/thegreatbeyond/2009/07/chemistry_paper_pwned_by_liveb.html. Other examples of the successful crowd-sourcing of scientific research include Open Wetware, in which a broad community of researchers makes their raw data freely accessible. OpenWetWare, Overview, http://www.openwetware.org/wiki/OpenWetWare:About (last visited May 3, 2010); SNPedia, SNPedia:FAQ, http://www.snpedia.com/index.php/SNPedia:FAQ (last visited May 3, 2010), a decentralized and open-access resource that collates research and resources surrounding SNPs; open-access resource that collates research and resources surrounding SNPs, and the open-access journals published by the Public Library of Science in which every article is freely accessible and searchable. *See* Pub. Library of Sci., Home, http://www.plos.org/ (last visitedMay 3, 2010). In addition, the NIH recently renewed its policy that all published results of NIH-funded research be deposited into the publicly searchable, full-text database PubMed Central no later than 12 months after publication. To view NIH's Public Access Policy, see Nat'l Insts. of Health, U.S. Dept. of Human Servs., Overview, http://publicaccess.nih.gov/ (last visited May 3, 2010).

[41] The decision of whether or not to return research results directly to patients is a controversial one that demands a case-by-case analysis of the participant population and the results in question. One recent survey interviewed potential study

important to longitudinal studies, which are likely to become increasingly common in large-scale genomic research.[42] Although some genomic research projects provide such data confidentially to participants[43] or only to certain authorized researchers, satisfying both this criterion and the criterion of publicly available data necessitates that public genomics participants consent to share their genomic data openly with the world. Although the linkage of phenotypic and genotypic research data does not necessarily entail that the individuals supplying such data be individually *identified*,[44] the content of the data renders it inherently *identifiable* when placed in the public domain, as described in more detail in Part II. Due to the impracticability of preserving participant privacy in the context of such linked datasets, public genomics depends on "information altruists"[45] who are willing to forego traditional expectations of informational privacy and confiden-

---

participants across the country and concluded that, on the whole, "[f]ocus groups participants voiced a strong desire to be able to access individual research results." Juli Murphy et al., *Public Expectations for Return of Results from Large-Cohort Genetic Research*, 11 AM. J. BIOETHICS 36, 41 (2008). Murphy, her coauthors, and other commentators also discuss the risks of such disclosure, including such concerns as blurring the line between research and the practice of medicine and subjecting individuals to risks or uncertainty based on genomic research results. *See also* Michelle N. Meyer, *The Kindness of Strangers: The Donative Contract between Subjects and Researchers and the Non-Obligation to Return Individual Results of Genetic Research*, 8 AM. J. BIOETHICS 44, 44-45 (2008) (arguing that the results of Murphy and his coauthors, indicating that participants express a preference for research results, does not create a binding obligation or duty on researchers to return such results). These concerns are taken up by this Article in detail in Parts II and III, *infra*.

[42] *See, e.g.*, Genetics & Pub. Policy Ctr., Genetics and Public Policy Center Receives Funding for Public Consultation Project, Aug. 26, 2009, *available at* http://www.dnapolicy.org/news.release.php?action=detail&pressrelease_id=134 (describing a $1.6 million follow-on funding award by NHGRI in 2009 in support of assessing public attitudes about a large-scale U.S. population-based study that "likely would involve the participation of hundreds of thousands of U.S. volunteers, who would be followed for a period of many years to ascertain and quantify the major environmental and genetic contributors to common illnesses.").

[43] See, for example, the CPMC and Scripps research projects, both of which intend to provide certain data directly and confidentiality to participants as part of their research protocols. *See supra* note 20.

[44] Although the PGP does not publicly link participant names to individual genotypic and phenotypic information – and while we are aware of no other study that does so – the PGP did publicly identify its first ten participants. *See* Pers. Genome Project, PGP-10, http://www.personalgenomes.org/pgp10.html (last visited May 3, 2010); Daniel MacArthur, Look into the Eyes of the PGP-10, http://scienceblogs.com/geneticfuture/2008/12/look_into_the_eyes_of_the_pgp1.php (last visited May 3, 2010) (discussing the online identification of the Personal Genome Project participants).

[45] *See* Isaac S. Kohane & Russ B. Altman, *Health Information Altruists – A Potentially Critical Resource*, 353 NEW ENG. J. MED. 2074, 2074-77 (2005).

tiality in order to further scientific research and, frequently, who are also willing to participate directly in that research.[46] While the open publication of individually identifiable genomic data carries known and unknown risks, it can also produce unanticipated benefits.[47]

Portability of Data. Public genomic datasets are enriched by combining the detailed information profiles of large numbers of identifiable individuals, providing researchers with the ability to compare individual genotype-phenotype data against a wide variety of sufficiently-sized populations. Although projects with large cohorts are becoming more common – for instance, the PGP and the Coriell Personalized Medicine Collaborative intend to enroll 100,000 individuals, while the UK BioBank aspires to 500,000[48] – the principles of public genomics promote data-sharing and interoperability. Public genomics data, therefore, is not tied to any particular project and may move freely to other genomics projects, with each such project – whether conducted by a government, a university, a business, or an individual and whether consisting of one or one million participants – able to harvest and, in some cases, seed the public genomics commons.[49] The PGP database, for example, is governed by a Creative

---

[46] Examples of public genomics pioneers include Craig Venter, James Watson and George Church; each was among the very first to release identifiable genomic information into the public domain.

[47] George Church, the first participant in the PGP, experienced one such benefit of public genomics. Church increased the dose of his cholesterol medicine after a complete stranger – who happened to be a hematologist – reviewed the medical records Church had previously made available online. Matthew Herper, *Going to Church*, FORBES.COM, Apr. 4, 2009, http://www.forbes.com/forbes/2009/0427/020-biofuels-biotech-genetics-ideas-opinions.html. In addition, one of the members of the "PGP-10" learned of a potential genetic predisposition to hypertrophic cardiomyopathy which prompted a follow-up visit to his physician. *See* Luke Timmerman, *Google, Microsoft May Help Usher in Personalized Medicine Wave, Says George Church,* XCONOMY, May 12, 2009, http://www.xconomy.com/boston/2009/05/12/google-microsoft-may-help-usher-in-personalized-medicine-wave-says-george-church/.

[48] In addition, the Genomics and Personalized Medicine Act, which was first introduced in 2006 by Senator Barack Obama and introduced in Congress again in 2007 and 2008, would, if passed, instruct the Secretary of HHS to "advance the field of genomics and personalized medicine" through establishment of "a national biobanking distributed database . . . a system for the integration of data, including genomic data and associated environmental and clinical health information, which shall facilitate the pooled analysis and synthesis of such data." S. 3822, 109th Cong. § 5 (2006). Although not specifically discussed in the proposed legislation, such a national biobank would be likely to rival the UK BioBank in size and scope.

[49] The notion of a "genome commons" traces back at least to Steven Brenner in 2007, when he proposed the idea of a public resource to store, verify and annotate information on human genetic variation. Steven E. Brenner, Commentary, *Common Sense For Our Genomes*, 449 NATURE 783, 783-84 (2007); Jing Tang et al., *Integrat-*

Commons CC0 "no rights reserved" tool[50] which permits transfer to and from, and commingling with, other public genomics datasets without restriction.

Voluntary and Informed Participation. This last characteristic is admittedly a normative judgment rather than a descriptive attribute. Voluntary, informed participation is not unique to public genomics, but the field's other features add an additional dimension to informed consent. Given the sometimes sensitive nature of genomic data and the still-novel structure of public genomics research, we argue that it would be irresponsible to publish an individual's genomic and phenotypic data in a publicly available, portable, and linked format, to be used by all comers for any purpose whatsoever, in the absence of anything less than truly voluntary and informed participant consent. For this reason robust informed consent protocols[51] and other human subjects research protections[52] are integral features of public genomics.

"Public genomics" is characterized by openness: information altruists publishing identifiable genotypic and phenotypic information to an open-access database that facilitates the collaboration and dialogue that genomic researchers desire and that will be necessary to understand the genomic underpinnings of our most complex traits and

---

ing Post-Genomic Approaches as a Strategy to Advance Our Understanding of Health and Disease, GENOME MED., Mar. 4, 2009, at 35.1, 35.4 (discussing the need to combine multiple datasets, including multiple types of datasets, and concluding that "[w]hile there are clearly obstacles that need to be overcome, biological models based upon multiple datasets are likely to become the basis that drives future research"); Kaye et al., *supra* note 37 (discussing the changing landscape of genomic data sharing and the challenges for both participants and researchers in facilitating such sharing).

[50] *See* Creative Commons, About CC0 – "No Rights Reserved," http://creativecommons.org/about/cc0 (last visited May 3, 2010) (describing the CC0 tool, which operates as a waiver of copyright and associated rights and, in the event the waiver is ineffective, is backed by a more traditional unconditional, irrevocable, non exclusive, royalty free license for any use). For more on the implementation of the CC0 license in the context of the Personal Genome Project, see Creative Commons, Expanding the Public Domain: Part Zero, http://creativecommons.org/weblog/entry/13304 (last visited May 3, 2010).

[51] One immediate consequence of this commitment to truly informed consent is that it restricts another desirable but not immediately attainable feature of public genomics: representative public enrollment and participation. Although some public genomics projects are generally open to the public (given certain minimum eligibility criteria), the rigorous informed consent criterion effectively limits the potential participant population, at least at this point in time. One of the ultimate goals of conducting responsible public genomics is to improve public understanding and acceptance of genomic research such that, in time, public genomics participant populations will become increasingly representative of the population as a whole.

[52] *See infra* Part IV.A.

intractable diseases. But such a high degree of openness comes with risks and challenges that cannot be ignored or minimized. Part II examines the ethical and social dimensions of public genomics and illustrates why voluntary participation and an aggressive commitment to openness and truly informed consent are of fundamental importance to responsible public genomics.

II. ETHICAL AND SOCIAL CHALLENGES IN PUBLIC GENOMICS

Public genomic research has engendered contentious debate in almost every conceivable forum: the scientific and legal literatures, the research and regulatory communities, Congress and state legislatures, and the popular media.[53] The specific points of ethical and policy disputes are many, but this Article highlights several areas where public genomics raises difficult policy questions that challenge traditional principles of research ethics.

First, can research capable of generating the rich datasets that define public genomics, and that are necessary to unravel the genomic bases of complex traits, satisfy traditional privacy expectations of research subjects, or is the promise of such protection illusory?

Second, what can and should be done to ensure that informed consent is truly "informed" in light of these privacy concerns and the shifting landscape of genomic science which, among other complexities, demands that genomic datasets be amenable both to future uses and to research questions that cannot be specifically identified or consented to at the outset?

Third, given that certain genomic information inherently possesses both research and clinical value, what duties does a genomic researcher have to identify and inform a participant of the clinical significance of genomic information? If researchers have either an ethical or a legal duty of disclosure, is it a continuing duty or only an initial duty? To what extent should researchers be subject to legal regulations that are intended to regulate the practice of medicine that may simultaneously restrict researchers' ability to disclose certain findings or to discuss the implications of those findings with research participants?

Each of the foregoing questions implicates a tension between the societal benefits of public genomics research and the interests of indi-

---

[53] *See* Lori Andrews, *Harnessing the Benefits of Biobanks*, 33 J.L. MED. & ETHICS 22, 23-24; Mark A. Rothstein, *Expanding the Ethical Analysis of Biobanks*, 33 J.L. MED. & ETHICS 89 (2005) (recognizing that biobanks complicate traditional issues of informed consent and privacy).

vidual participants in that research. As discussed in Part III, the current legal and regulatory regime does not adequately address these tensions. Some of this inadequacy is due to such generic factors as divided authority, regulatory inertia, jurisdictional confusion, and the reality that scientific and technological innovation inevitably outpaces the corresponding and necessary legal and regulatory response. But there are also factors inherent in public genomics that challenge three of the traditional pillars of human subjects research: privacy, informed consent, and the distinction between scientific research and the practice of medicine.

A. Problems of Privacy

The traditional expectation that genomic data supplied by research participants should be subject to robust privacy protections arises out of a complex mixture of basic ethical principles and a general concern that the release of identifiable genomic data into the wrong hands would pose significant risk for individual participants. At the ethical core of this concern lies the principle of beneficence. Beneficence is one of the three pillars of the Belmont Report, the foundational source of modern human subjects research protections in this country, which requires researchers to minimize possible harms.[54] Leaving aside the question of whether an individual's genomic data warrants or is amenable to robust privacy protections, it is indisputable that the current norm in genomic research requires investigators and funders to devote significant resources to the protection of participants' genomic data.

Large and influential genomic research projects, including the HapMap Project and the UK BioBank, have strongly emphasized their commitment to privacy.[55] More recently, influential genomic research undertakings such as the 1000 Genomes Project and the Coriell Personalized Medicine Collaborative (CPMC) have taken a similar ap-

---

[54] "In this document, beneficence is understood in a stronger sense, as an obligation. Two general rules have been formulated as complementary expressions of beneficent actions in this sense: (1) do not harm and (2) maximize possible benefits and minimize possible harms." NAT'L COMM'N FOR THE PROT. OF HUMAN SUBJECTS OF BIOMEDICAL & BEHAVIORAL RESEARCH, U.S. DEP'T. OF HEALTH, EDUC., & WELFARE, THE BELMONT REPORT: ETHICAL PRINCIPLES AND GUIDELINES FOR THE PROTECTION OF HUMAN SUBJECTS OF RESEARCH (1979) [hereinafter BELMONT REPORT], *available at* http://ohsr.od.nih.gov/guidelines/belmont.html#gob.

[55] Jeantine E. Lunshof et al., *From Genetic Privacy to Open Consent*, 9 NATURE REV. GENETICS 406, 408 (2008) ("The HapMap informed-consent protocol does not unambiguously guarantee anonymity or confidentiality of participants' genetic information. On the contrary, it even mentions the risk of tracing identity through publicly available HapMap data. Nevertheless, the consent protocol clearly suggests that the risk of re-identification is vanishingly small.").

proach, emphasizing their strong commitment to privacy in both their public pronouncements and their informed consent materials.[56] The CPMC, for instance, makes "every effort to ensure [participant] privacy," including the use of "the most sophisticated computer security tools currently available to store and protect [participants'] information," sample barcoding and secure storage procedures, as well as the provision of a Certificate of Confidentiality (discussed below).[57] The resulting norm in genomic research, as well as the likely expectation of most participants in such projects, is a structure in which genomic data is guarded by the most sophisticated and robust data protection and privacy protocols available.

Unfortunately, growing evidence suggests that both explicit and implicit promises to maintain participant privacy may be misleading, if not entirely illusory. Under the broad heading of "privacy," which is concerned with limiting access to certain information – in this case the genomic data of research participants – fall the related concepts of "confidentiality" and "anonymity." "Confidentiality" typically involves using coded identifiers, such as barcodes, to link subjects to their samples and information. "Anonymity," by contrast, usually refers to the stripping of all or most of the potential identifiers from samples and information.[58]

In recent years, commentators and researchers alike have raised challenges to the privacy norm – whether manifest as promises of anonymity or of confidentiality – and have attempted to show that privacy guarantees are likely not as ironclad as they appear to potential participants at the time of consent.[59] Perhaps the strongest warn-

---

[56] *See* 1000 GENOMES PROJECT, 1000 GENOMES PROJECT: DEVELOPING A RESEARCH RESOURCE FOR STUDIES OF HUMAN GENETIC VARIATION – CONSENT TO PARTICIPATE, http://www.1000genomes.org/bcms/1000_genomes/Documents/Informed%20Consent%20Form%20Template.pdf (last visited May 3, 2010) [hereinafter CONSENT TO PARTICIPATE]; Coriell Personalized Med. Collaborative, Consent to Participate in a Research Study, http://cpmc.coriell.org/Docs/PDF/Informed_Consent.pdf (last visited May, 3, 2010); Coriell Personalized Med. Collaborative, Data and Sample Security, http://cpmc.coriell.org/Sections/About/Security.aspx?PgId=16 (last visited May 3, 2010); Coriell Personalized Med. Collaborative, Privacy and Confidentiality, http://cpmc.coriell.org/Sections/About/Confidentiality.aspx?PgId=141 (last visited May 3, 2010).

[57] Coriell Personalized Med. Collaborative, Privacy and Confidentiality, *supra* note 56.

[58] Greely, *supra* note 17, at 349, 351.

[59] *See, e.g.*, Lunshof et al., *supra* note 55; Kaye et al., *supra* note 37; Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays*, 4 PLoS GENETICS 1 (2008), *available at* http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1000167; Bradley Malin & Latanya Sweeney, *How (Not) to*

ings come from the lawyer and ethicist Henry Greely, who goes so far as to accuse some genomics projects of "twisting the meanings of words like 'confidential,' 'anonymous,' 'human subjects,' 'informed consent,' and 'justice' in ways that jeopardize, for the sake of administrative convenience and short-term research gains, the interests, the wishes, and the rights of those who contributed to these resources."[60] Similarly, Lunshof *et al.* suggest that these illusory promises of privacy violate the key ethical principle of veracity, undermining the validity of informed consent and thereby failing to adequately protect human subjects research participants.[61,62]

Focusing specifically on confidentiality, Greely argues that "[t]he problem with coded identifiers is that they cannot be made definitely secure."[63] He suggests a number of plausible security breach scenarios, including misconduct by the person who retains the key to the code (often called the "trusted intermediary"),[64] theft of the key by hackers,[65] and the loss of laptops or other storage devices that contain the keys or the sensitive information itself.[66] These kinds of possibilities have been rendered even more plausible by recent breaches, which include the compromise of military personnel records by the loss of a laptop,[67] Kaiser Permanente's serious breach of confidentiali-

---

Protect Genomic Data Privacy in a Distributed Network: Using Trail Re-identification to Evaluate and Design Anonymity Protection Systems, 37 J. BIOMEDICAL INFORMATICS 179 (2004). *See also supra* note 19 (discussing dbGaP data pullback).

[60] Greely, *supra* note 17, at 344.

[61] Lunshof et al., *supra* note 55, at 408.

[62] In addition to the technical, ethical and legal concerns discussed above, there is another more pragmatic concern associated with the potential failure to deliver on privacy promises: the potential for a loss of public trust. Public genomics is capable of flourishing only with the trust and participation of both the information altruists that supply the projects with data and of society at large, which sanctions and largely funds such public genomics research. While there are numerous ways to violate the trust of individuals and of the public, an unexpected and high-profile genomic privacy breach could pose a serious threat to the future of public genomics. *See, e.g.*, Kaye et al., *supra* note 37, at 335 ("Establishing and maintaining global public goodwill and trust is an ongoing task to ensure the future of sound, and hence ethical, scientific research."); David Gurwitz et al., *Children and Population Biobanks*, 325 SCIENCE 818, 819 (2009) (discussing the importance of informed consent in population biobanks involving children and concluding that "[t]he long-term benefits of maintaining public trust in biomedical research . . . justify extra governance efforts and added costs").

[63] Greely, *supra* note 17, at 350.

[64] *See* Rothstein, *supra* note 53, at 95.

[65] *See* Greely, *supra* note 17, at 351.

[66] *See id.*

[67] *See* Elect. Privacy Info. Ctr., Veterans Affairs Data Theft, *available at* http://epic.org/privacy/vatheft/ (last visited May 3, 2010).

ty as a result of a complex information technology accident,[68] the use of file-sharing applications by researchers to hack into medical records,[69] the theft of encrypted data from hard disks,[70] the prediction of social security numbers using statistical techniques,[71] and the inappropriate access and sale of patient medical data by employees of healthcare insurers and providers, who were eventually convicted for these crimes.[72]

Anonymization, according to Greely, takes all of the problems associated with confidential data and introduces a few new problems. He provides a detailed explanation of how it would be possible to identify a hypothetical subject – him – in an ostensibly "anonymized" (information stripped of all identifiers) DNA biobank. Others have tested the proposition by attempting to breach the privacy protections of real-world and purportedly anonymized biobanks. For example, researchers at Carnegie Mellon used two pieces of publicly available information – date of birth and zip code of residence – to identify the health records of the governor of Massachusetts.[73] In another example, Homer *et al.* used "high-density single nucleotide polymorphism (SNP) genotyping microarrays" – a widely available technology – to determine whether a particular "individual's genomic DNA is present at trace amounts within a complex mixture containing DNA from numerous individuals." They claim to have done so "accurately and robustly."[74] Although Homer *et al.* cite the example of identifying an individual's DNA in a single degraded forensic sample, the technique should also apply to pooled collections of samples, even when "sharing only summary statistics."[75]

In specific response to Homer *et al.*'s research, the National Human Genome Research Institute (NHRGI) has restricted public access to previously unrestricted portions of pooled, de-identified data in its Genotype and Phenotype (dbGaP) dataset.[76] The open-access

---

[68] *See* Lunshof et al., *supra* note 55.

[69] *See, e.g.*, Bob Brewin, *File-sharing Networks Used to Uncover Thousands of Medical Records*, NEXTGOV.COM, Feb. 27, 2009, http://www.nextgov.com/nextgov/ng_20090227_9147.php.

[70] John Markoff, *A Method for Stealing Critical Data*, N.Y. TIMES, Feb. 22, 2008, at C1.

[71] John Markoff, *A Weakness in Social Security Numbers*, N.Y. TIMES, July 7, 2009, at A14.

[72] Charles Ornstein, *Breaches in Privacy Cost Kaiser*, L.A. TIMES, May 15, 2009, at A3.

[73] Malin & Sweeney, *supra* note 59, at 181-86.

[74] Homer et al., *supra* note 59, at 1.

[75] *Id.* at 9.

[76] *See* Jennifer Couzin, *Genetic Privacy: Whole-Genome Data Not Anonym-*

portion of dbGaP data included information on genetic variation observed in people with such serious diseases as asthma, prostate cancer, cardiovascular diseases, diabetes, Parkinson's disease, and amyotrophic lateral sclerosis. The new protocol will require NIH approval before access is granted, a change that closes off a section of the genomic commons to all but a handful of researchers and a development that some fear "could delay or maybe even prevent some studies . . . ."[77]

The cautionary lesson is that critics of the privacy promise are correct: at least as a purely technical matter, anonymity and confidentiality simply cannot be guaranteed in the context of most (if not all) large-scale genomic research projects. Moreover, such guarantees further test the ethical principle of veracity when they rest on Certificates of Confidentiality and purport, in the case of the CPMC, to render participant records "strictly private."[78] Certificates of Confidentiality are issued by the U.S. Department of Health and Human Services and intended "to protect the privacy of research subjects by protecting investigators and institutions from being compelled to release information that could be used to identify subjects within a research project."[79] Although such certificates are an important indicator of a research project's commitment to the privacy norm, these certificates hardly provide an absolute guarantee of data privacy. As a 2006 North Carolina case involving Duke University researchers reveals,[80] Certificates of Confidentiality are of uncertain legal effect in the face of a criminal defendant's rights to discover exculpatory evidence and to confront and cross-examine witnesses.[81] Accordingly, as with other privacy protection measures, they are insufficient to underwrite absolute privacy promises.

---

ous, Challenging Assumptions, 321 SCIENCE 1278, 1278 (2008) (describing the dbGaP data restrictions, as well as similar restrictions implemented by the Wellcome Trust and the Broad Institute, among other genomic research institutions).

[77] Catherine Clabby, *DNA Research Commons Scaled Back*, 97 AM. SCIENTIST 113, 113 (2009).

[78] Coriell Personalized Med. Collaborative, *supra* note 56.

[79] OFFICE OF EXTRAMURAL RESEARCH, U.S. DEP'T HEALTH & HUMAN SERVS., CERTIFICATES OF CONFIDENTIALITY: BACKGROUND INFORMATION, http:/grants.nih.gov/grants/policy/coc/background.htm (last visited May 3, 2010).

[80] State v. Bradley, 634 S.E.2d 258 (N.C. Ct. App. 2006).

[81] *See* Laura M. Beskow et al., *Certificates of Confidentiality and Compelled Disclosure of Data*, 322 SCIENCE 1054, 1054 (2008); Clabby, *supra* note 77, at 114 ("[L]aw enforcement is expanding the way it uses DNA to track suspects, even innocent relatives of suspects in crime cases. In one case, investigators secretly obtained from a medical clinic a DNA sample that belonged to the daughter of a suspected serial killer to hasten their detective work.").

Some projects are heeding these warnings and designing their research and informed consent protocols accordingly. At one end of the spectrum, the PGP provides no privacy promises whatsoever, stating simply that participants' "genetic and trait information will not be maintained or made available in a confidential or anonymous fashion."[82] Taking an intermediate approach, the 1000 Genomes Project balances the traditional norm of robust privacy protection measures, including over-enrollment and over-sampling, to ensure that "nobody – not even you or us – will know for sure whether your sample was used" in the project," with an open acknowledgement of the limitations of such protections.[83] The project's Informed Consent Form Template addresses both known risks of re-identification ("although only experts will know how to interpret [your genomic information], there is a small chance that somebody could figure out how to connect you with [that] information") and unknown risks ("as technology advances, there may be new ways of linking information back to you that we cannot foresee now").[84] Finally, at the far end of the privacy spectrum lie research projects such as the CPMC. By assuring participants that "no personal identifying information about you will be released to anyone outside of the CPMC study without your authorization,"[85] such projects downplay or even ignore what are categorized as the the "remote" risks of re-identification of participant data and other privacy breaches.[86]

---

[82] PGP CONSENT FORM, *infra* note 90, at § X(1.1). The PGP informed consent form further elaborates: "Your genetic and trait information will be made available via a publicly accessible website and database, according to the procedures described above. Public disclosure of your information due to unintended data breaches, including hacking or other activities outside of the procedures described above, is also possible." *Id.* It is important to note that, as in the case of the PGP, a lack of privacy promises does not mean that privacy is lacking entirely. Although identifiable – particularly in the case of the first ten participants, the so-called "PGP-10" – the PGP does not explicitly identify its participants. This distinction – between identifiable and identified – is a meaningful one: public genomics projects may well take steps to prevent participants from becoming *identified*, including through the application of robust privacy measures, while still acknowledging that participation nevertheless carries a risk of identification based on the *identifiable* nature of the contributed data.

[83] CONSENT TO PARTICIPATE, *supra* note 56, at 4-5.

[84] *Id.*

[85] Coriell Personalized Med. Collaborative, *supra* note 56.

[86] *Id.* The Coriell Personalized Medicine Collaborative revised Informed Consent form continues on to say that "however, since there is a remote chance of identification, all qualified investigators who have access to your data will sign legal documents requiring them to protect the privacy of your information." *Id.* Similarly, subsection 3, entitled "Confidentiality Protections Through the Federal Government," states that "if you choose to release your genetic variant information to others . . .

The practical difficulties of effecting perfect privacy – whether complete confidentiality or anonymity – in the genomics research context need not result in public genomics researchers' abandonment of all privacy protection measures, nor should it forbid them from accurately describing such measures to potential participants. Rather, these challenges strongly suggest that genomic researchers must diligently examine the tradition of implicit and explicit *promises* of privacy to participants – a suggestion that public genomics projects like the 1000 Genomes Project and the PGP are heeding and one that carries broad implications for traditional assumptions of both privacy and informed consent more generally. While beneficence may require that privacy be protected to the extent possible in the public genomics context, veracity demands that it not be guaranteed unqualifiedly.

B. The Difficulty of "Informed" Consent

Informed consent is indisputably one of the foundational ethical principles of human subjects research. From the trials of Nazi war criminals[87] to the Belmont Report[88] to current federal regulations,[89] the ethical balance between advancing important scientific research and protecting human subjects has been struck only with the willing and fully informed consent of participants. While the risk of physical harm to participants in public genomics research seems likely to be low – preparing a genomic sequence in most cases requires little more than a saliva sample or cheek swab from the participant – the intangible risks present a more difficult challenge. Accepting the model of public genomics as set forth in Part I – including publicly available and individually identifiable genotypic and phenotypic data – requires that researchers and participants alike carefully think through the associated risks. Because the researchers themselves cannot precisely predict where genomics research will lead or what challenges it will face, however, both parties must understand that researchers are in-

---

Coriell cannot guarantee the confidentiality of this information." *Id.* at § H(3). All told, the risk of breach or re-identification is minimized – especially in comparison to the efforts at disclosure undertaken by the 1000 Genome Project or the PGP – and the implication is that a breach, should it occur, would be most likely due to a participant's own carelessness with her information.

[87] Informed Consent is the first article of The Nuremberg Code. *See* 2 NUERNBERG MILITARY TRIBUNALS, TRIALS OF WAR CRIMINALS 181-82 (1949).

[88] Informed Consent is the first application of the three ethical principles (respect for persons, beneficence, and justice) set forth in the Belmont Report. *See* BELMONT REPORT, *supra* note 54.

[89] Informed Consent is one of the core criteria for IRB approval of human subjects research under the Common Rule. *See* 45 C.F.R. § 46.111(a)(4) (2008).

capable of fully "informing" participants of all such risks. For example, although the informed consent form for the Personal Genome Project provides a lengthy but "non-comprehensive list of hypothetical scenarios that could pose risks," it advises participants to check back frequently "to obtain the most current information regarding potential risks and discomforts as they become apparent."[90] Even determining what information should be given to prospective sample donors in order to make sure that their "consent" is "informed" at a meaningful level is no easy task. The NHGRI spokesperson who announced that agency's decision to make the dbGaP dataset private framed the problem succinctly: "How much do you tell people without scaring them? How do you communicate the level of risk? What level of risks are people willing to tolerate? How frequently do they want to be asked?"[91]

Here, again, the various public genomics projects now in development illustrate a range of approaches to this problem. Returning to the Coriell project, its informed consent form generally restricts its discussion of participants' risks to those contained in the specific genetic data returned to participants, including the risks of being provided with (i) inaccurate data, (ii) inaccurate or incomplete disease risk assessment, (iii) data related to non-medically actionable diseases, or (iv) misinterpretation of or apprehension about that data on the part of the participant.[92] As previously discussed, the risk of a breach of participant privacy is largely discounted in Coriell's informed consent process.[93]

The PGP's informed consent agreement takes the opposite approach and includes a lengthy but "non-comprehensive list of hypothetical scenarios that could pose risks" for participants and their families.[94] Although the disclosed risks range from discrimination – by employers or insurers, regardless of whether it is lawful to do so, or even friends or family members – to the planting of synthetic DNA

---

[90] HARVARD MED. SCH., PERSONAL GENOME PROJECT CONSENT FORM § VII (2009), http://www.personalgenomes.org/consent/PGP_Consent_Approved03242009.pdf [hereinafter PGP CONSENT FORM]. By the time of publication, the PGP Consent Form will have been updated again to reflect changes in the project design and participation, as well as clarifications of the risks involved in participation. While these are important improvements to the PGP Consent Form, these changes do not materially alter the PGP or the PGP Consent Form for purposes of this Article.

[91] Clabby, *supra* note 77, at 114.

[92] *See* Coriell Personalized Med. Collaborative, *supra* note 56, at §§ E (2), (3), (5), (6), (7).

[93] *See supra* notes 81-84 and accompanying text.

[94] PGP CONSENT FORM, *supra* note 90, at § VII.

at a crime scene,[95] participants are repeatedly warned that "the complete set and magnitude of the risks that the public availability of [your genomic data] poses to you and your relatives is not known at this time."[96] Participants are also advised to "check the [PGP's] website regularly to obtain the most current information regarding potential risks and discomforts as they become apparent."[97] The 1000 Genomes Project is similarly sensitive to the possibility of unknown or uncertain future risks, warning participants that "we cannot always foresee the results of research, so new risks may come up in the future that we cannot predict now."[98]

"Informed," in the public genomics context, means both that researchers use their best efforts to diligently identify risks to participants – even remote risks or risks that the *researchers* judge to be of insignificant value[99] – and that participants are capable both of understanding the nature and the magnitude of the risks presented. More importantly, participants must also understand and internalize the fact that participation entails as yet unknown (and potentially serious) risks. From an ethical perspective, truly informed consent helps to satisfy the dual goals of autonomy and beneficence. From a more pragmatic perspective, informed consent will prepare participants prior to enrollment so that any of the known or unknown risks that later materialize will not cause participants to retrospectively view their consent as uninformed.

Given the current state of genomic knowledge and the tremendous anticipated size of many public genomics research projects, it is statistically probable that risks – both predicted and unanticipated –

---

[95] *Id.* Although discussed as a hypothetic risk in the PGP's informed consent agreement, a recent paper in the journal Forensic Science International: Genetics demonstrated that it is indeed possible to fabricate synthetic DNA to match the profile of an individual without obtaining any actual tissue from that person, prompting the lead author of the paper to comment, "You can just engineer a crime scene." Andrew Pollack, *DNA Evidence Can Be Fabricated, Scientists Show*, N.Y. TIMES, Aug. 18, 2009, at D3.

[96] PGP CONSENT FORM, *supra* note 90, at § VII.

[97] *Id.* The lengthy and aggressive disclosure of risks, including even those that would appear to be remote, has not seemed to hamper, at least initially, the PGP's ability to attract interested participants. As of early 2009 the PGP was oversubscribed, with well over 13,000 indications of interest for 100 available openings in the second phase of participant enrollment, despite minimal recruitment, enrollment or publicity efforts up to that point.

[98] *Id.*

[99] It may be that a *de minimis* threshold for risk disclosure could and should be developed, but such a threshold would need to be established over time and with participation from researchers, participants, IRBs, and community members and other stakeholders.

will materialize over the life of a project and that some participants will feel regret at having participated. There is likely no way to completely eliminate bad outcomes and participant regret other than ceasing to perform socially valuable genomic research. Informed consent allows participants – including the information altruists of public genomics – to accept certain risks. The dichotomy for these participants is not between a choice made with either full knowledge or with complete ignorance of the risks of public genomics. All choices will be made in ignorance of risks that are unknown but presumed significant. In such circumstances, an informed consent that fully discloses all known risks while duly acknowledging and accepting an ignorance of unknown risks ensures that participants are as informed as possible and less likely to experience episodes of post-enrollment regret.

C. The Clinical Significance of Genomic Data

An individual's genomic sequence data inescapably holds clinical significance. As understanding of the relationship between genotype and phenotype expands, so too will the clinical significance of this data.[100] From the researcher's perspective – as well as from a legal perspective, as described in Part III – providing clinically significant information or interpretation blurs the line between genomic research and the practice of medicine. From the participant's standpoint, receiving a gigabyte-long[101] string of Gs, Ts, As, and Cs without any discussion of the relationship between that string of letters and the individual's predicted future trait or disease status would surely amount to a disappointment, regardless of the content of the informed consent.[102] Public genomics participants may also expect some clinical benefit (or at least some degree of understanding) as a result of the

---

[100] Although it may not increase linearly or proportionally with scientific understanding. It is possible, if not likely, that a deepening understanding of genotype-phenotype relationships will disclose relationships of such complexity or with such attenuated genomic or genetic bases that the clinical significance of genomic data in such cases will be minimal in comparison to the value of other environmental, medical history or statistical data. Nevertheless, it is likely that the clinical utility of an individual's genomic information – in purely absolute terms – is likely to increase over time as it becomes more fully understood and additional genomic-drive diagnostics and treatments are developed.

[101] *See* Daniel MacArthur, *How Much Data Is a Human Genome? It Depends How You Store It.*, GENETIC FUTURE, June 28, 2008, http://www.genetic-future.com/2008/06/how-much-data-is-human-genome-it.html (discussing different ways of storing human genome data).

[102] *See, e.g.*, Murphy et al., *supra* note 41, at 40-41 (discussing research participants' desire for return and explanation of research data).

receipt of their genomic data, including an expectation of access to higher levels of counseling or other clinically-relevant forms of follow-up services, regardless of the researchers' ability to provide such services, especially in large-scale public genomics projects. Although the simpler solution might be to avoid returning any data of potential clinical significance, strong arguments exist for the proposition that researchers have a duty to inform participants of incidental findings if their failure to do so could result in serious and avoidable harm.[103] This creates a particular difficulty for genomic researchers, insofar as findings of clinical significance are likely to be far more common in this area than in other areas of scientific research. This is especially true in the case of studies that generate complete exomic[104] or genomic data, where findings with clinical significance cannot properly be termed "incidental" at all; these findings are simply an expected part of the data generated for each participant.

The difficulty of how to respond to such significant findings is compounded by the fact that researchers and clinicians may differ in their understandings of whether data is "significant" or actionable. Just as researchers may be unqualified to understand the clinical significance of or proper clinical response to a participant's genomic data, so too are clinicians likely to find themselves behind the curve in their knowledge and implementation of the latest significant genomic research. Even researchers who focus most diligently on understanding the clinical implications of their data will be unfamiliar with the latest

---

[103] *See* Susan M. Wolf et al., *The Law of Incidental Findings in Human Subjects Research: Establishing Researchers' Duties*, 36 J.L. MED. & ETHICS 361, 362 (2008).

> We argue that researchers owe research participants duties that are both ethical and legal obligations: to disclose in the informed consent process the possibility of discovering IFs and the plan for management; to recognize an IF that arises during the course of research; to verify the presence of an IF and evaluate its probable importance, obtaining expert consultation if necessary; and to offer to disclose an IF of likely clinical or reproductive importance to the research participant.

Id.

[104] The exome is the part of the genome formed by exons, the portions of genes that are expressed in the form of synthesized proteins and other functional gene products. It is often described as the functional or coding portion of the genome, despite increasing recognition that regions of the genome that do not directly code for proteins are functionally significant. For the human genome, the exome is estimated to comprise approximately 1.5% of the total genome. *See* Pauline C. Ng et al., Genetic *Variation in an Individual Human Exome*, PLOS GENETICS, August 2008, at 1, *available at* http://www.plosgenetics.org/article/fetchObjectAttachment.action?uri=info%3Adoi%2F10.1371%2Fjournal.pgen.1000160&representation=PDF.

findings to some degree or may fail to reach a consensus, given the rapid expansion of knowledge in the field.[105]

Moreover, unless a researcher is a licensed physician, providing a subject with interpretation of the subject's genomic sequence data could easily exceed that researcher's expertise. As discussed in Part III, researchers who share interpretative guidance with their subjects also risk running afoul of state statutes regulating genetic counseling and the unauthorized practice of medicine. Researchers who share clinically significant data with their research subjects, even when motivated only by a perceived moral or ethical obligation, could even risk medical malpractice liability in state court.[106]

The longitudinal nature of many public genomics research projects further complicates the issue. If permitted by the study protocol, researchers may obtain express consent to recontact study participants in the future. In such cases, what obligation – and at what expense – does a researcher have to notify participants of subsequent research or knowledge developments that imbue previously returned genomic sequence data with new or altered clinical significance? Expecting researchers to continuously monitor the clinical significance of new knowledge for all study participants would dramatically increase the costs and legal risk of genomic research – likely prohibitively.[107]

Yet the problem may not be a completely intractable one. The PGP, for instance, has attempted to address this problem by providing participants with a "Preliminary Research Report," a "non-comprehensive list of genetic variants present in [the participant's] DNA sequence data that are currently believed by the PGP to be of significance in clinical practice."[108] The PGP provides the Preliminary Research Report alongside prominent explanations of its limitations, including that it should not to be used for any clinical or medical purpose and will not be updated on an ongoing basis. In this way, the PGP attempts to balance the desire of researchers to aid participants in the basic interpretation and understanding of their genomic data with the legal necessity of complying with professional, legal, and financial realities that collectively prevent PGP researchers from

---

[105] Genomic research projects take varying approaches to this issue. *See* Coriell Personalized Med. Collaborative, Informed Cohort Oversight Board (ICOB), http://cpmc.coriell.org/Sections/About/SAB.aspx?PgId=51 (last visited May 3, 2010).

[106] This concern is hypothetical at the moment. *See generally* Larry I. Palmer, *Should Liability Play a Role in Social Control of Biobanks?*, 33 J.L. MED. & ETHICS 89 (2005) (discussing possibility of such liability).

[107] Murphy et al., *supra* note 41, at 37.

[108] PGP CONSENT FORM, *supra* note 90, at § V, 4.2.

playing a more active role in interpreting individual participant data. If researchers are able to provide participants with some indication of the clinical utility of the data – along with the raw data – without subjecting themselves to legal liability or additional regulatory obstacles, both participant and ethical concerns may be satisfied.

Of course, this entire issue may be moot before long. If technological advances continue to drive down the cost of genomic sequencing, genomic sequencing could become an affordable commodity by the end of this decade.[109] If that is the case, individuals – of their own accord or at the encouragement of their healthcare providers or insurers – may routinely have their full genome sequenced as part of routine clinical care and in advance of their participation in any particular genomic research project.[110] In that case, the burden of preparing and interpreting genomic sequence data may largely shift from the research to the clinical setting, significantly lessening participants' reliance on researchers to interpret such data.

D. Legal Dimensions of Ethical and Social Issues

Public genomics has the potential to vastly expand our understanding of the interaction between genotype, phenotype and environment and, in the process, to provide immense benefits to human health, but it must deliver on that promise within the framework of the law and in a manner that adheres to core ethical principles of human subjects research. While most genomic researchers implicitly accept this principle, few understand how the norms and guidelines of human subjects research protection and the developing field of public genomics intersect with the current legal regime. For example, if study participants provide informed consent to public genomics research only by understanding the magnitude of both known and *unknown* risks of participation, then negligence or medical malpractice actions tried before a state court judge accustomed to acting with the benefit of hindsight (including knowledge of risks unknown at the time of consent that have since become manifest) present public genomics researchers with a potentially serious litigation risk. To take another example, state direct-to-consumer genetic testing statutes,[111] intended

---

[109] *See supra* notes 29-31 and accompanying text.

[110] Then again, even with a substantial additional drop in the cost of genomic sequencing, it is not yet a given that a whole-genome sequence will become a standardized part of routine clinical activity. *See, e.g.*, Dan Vorhaus, *The Genome* In Silico *and the Future of Whole-Genome Sequencing*, GENOMICS L. REP., July 27, 2009, http://www.genomicslawreport.com/index.php/2009/07/27/the-genome-in-silico-and-the-future-of-whole-genome-sequencing/.

[111] For examples, see *infra* notes 136-41 and accompanying text.

to protect consumers from harm due to unregulated or unprofessional medical services, may cast their nets wide enough to sweep in public genomic research projects, even those not intended to target consumers or provide clinical analysis to participants or healthcare providers.

But no matter how robust the self-reflection and self-regulation, risk-free public genomic research simply does not exist. With that in mind, the object of researchers and regulators, consistent with the fundamental ethical norm of beneficence to subjects, should be to develop research projects and legal frameworks that acknowledge and responsibly address the risks and rewards of public genomics. A variety of legal structures have evolved to balance the costs and benefits of general scientific research, but the novel social and policy issues of public genomics threaten to disrupt that legal-scientific equilibrium. In the final Part of this Article, we sketch the characteristics of a new normative, legal, and regulatory framework designed to enable public genomics research while simultaneously protecting the rights of its participants. To modify a legal regime that currently does neither, however, requires an understanding of that regime.

III. THE CURRENT LEGAL REGIME

Because issues of privacy, informed consent, and the clinical significance of genomic data implicate a wide range of statutory and common law issues, the legal regime presently governing public genomics resists easy categorization. In some cases, as with state statutes governing research data privacy, public genomics research falls within the intended scope of coverage. In other cases, as with medical malpractice or regulation of direct-to-consumer genomic service providers, public genomics falls only incidentally within the application of a particular legal or regulatory regime. In all cases, however, the indeterminacy and inconsistency of the existing legal structure fails to adequately protect research subjects while posing severe risks to the progress of responsible public genomics.

This Part examines the shortcomings of the prevailing legal structure governing public genomics from the standpoint of previously identified policy issues. The goal is not to provide a comprehensive overview of the legal landscape but rather to illustrate the ways in which the current legal regime is incompatible with public genomics. Beginning with federal law, this Part considers some of the various statutes and regulations relevant to public genomics research. Next, the common law doctrines that courts may find useful in tort cases involving public genomics researchers and research projects are evaluated. Finally, this Part concludes that the scattered, ad-hoc regulatory approach adopted by both state and federal agencies is inconsistent

with the legitimate goals of responsible public genomics research and, in many cases, the policy interests underlying the current legal regime itself.

A. Statutes and Regulations

Both state and federal legislatures have addressed issues affecting public genomics research, most notably in the areas of participant privacy, informed consent, and the regulation of the practice of medicine. International law has also played an important role in shaping ethical and normative understandings of the permissible bounds of human subjects research.

1. Federal Law

At the federal level, the Genetic Information Nondiscrimination Act of 2008 (GINA)[112] represents the most comprehensive effort thus far to regulate the use of genetic information. GINA bars discrimination on the basis of genetic information by health insurance companies and employers. Because GINA will not become effective until implementing regulations are enacted,[113] it is difficult to ascertain exactly how, if at all, its passage will directly impact public genomics. More important than any direct impact, however, is the public message conveyed by GINA's passage, including the reassurance that information altruists may need to willingly participate in public genomics research.

Given public concern with discrimination on the basis of genetic information,[114] GINA may provide certain study participants with some limited comfort that their genetic information, should it become

---

[112] Pub. L. No. 110-233, 122 Stat. 881 (codified as amended in scattered sections of 42 U.S.C.).

[113] Title I of GINA (Genetic Nondiscrimination in Health Insurance) took effect one year after the effective date of enactment, May 21, 2008. *See* Genetic Information Nondiscrimination Act (GINA) of 2008, Pub. L. No. 110-233, 122 Stat. 881 §§ 101(f), 102(d), 103(f), 104(c), 105(b) (codified in scattered sections of 42 U.S.C.). Title II of GINA (Prohibiting Employment Discrimination on the Basis of Genetic Information) will take effect eighteen months after the effect date of enactment. § 213.

[114] *See, e.g.*, GENETICS & PUB. POLICY CTR., U.S. PUBLIC OPINION ON USES OF GENETIC INFORMATION AND GENETIC DISCRIMINATION (2007), *available at* http://www.dnapolicy.org/images/reportpdfs/GINAPublic_Opinion_Genetic_Information_Discrimination.pdf (concluding that "[n]early all Americans believe that health insurers and employers should not be able to deny or limit insurance coverage or to make decisions about hiring and promotion based on genetic tests results about their risk of future disease" and that "[t]hree of every four Americans support a law forbidding genetic discrimination by health insurers and employers.").

identified, cannot be used to discriminate against them in the context of health insurance or employment. Unfortunately, GINA is non-comprehensive in several key respects, failing to prohibit the use of genetic information in setting rates or determining coverage for life, disability, or long-term care insurance.[115] Moreover, while GINA's passage makes certain forms of genetic discrimination illegal, it cannot ensure that genetic discrimination will not occur in the future. For these reasons, some commentators view GINA's limited protections as doing more harm than good, especially insofar as these protections encourage unwarranted individual reliance on the law.[116] Nonetheless, GINA represents both an important step towards the protection of participants in public genomics research and a potential model, both good and bad, for future, more comprehensive legislation against genetic discrimination.

Proposed and pending federal legislation, although increasingly recognizing the challenges and opportunities presented by the field of genomics, brings little clarity. One of the most high-profile pieces of proposed legislation is the Genomics and Personalized Medicine Act (the "GPMA"), which was introduced into the U.S. Senate in 2006 and 2007. The GPMA was most recently introduced into the House of Representatives in 2008,[117] where it stalled in the Ways and Means Committee. The GPMA contemplates the creation of a Personalized Medicine Interagency Working Group (IWG) charged with the crea-

---

[115] *See, e.g.*, Mark A. Rothstein, *Is GINA Worth the Wait?*, 36 J.L. MED. & ETHICS 174, 174 (2008) (discussing the "essential issues of discrimination and disclosure of genetic information in health insurance and employment that GINA does *not* address"); Amy L. McGuire & Mary Anderlik Majumder, *Two Cheers for GINA?*, 1 GENOME MED. 6.1, 6.2 (2009); OFFICE FOR HUMAN RESEARCH PROTS., U.S. DEP'T HEALTH & HUMAN SERVS., GUIDANCE ON THE GENETIC INFORMATION NONDISCRIMINATION ACT: IMPLICATIONS FOR INVESTIGATORS AND INSTITUTIONAL REVIEW BOARDS (2009), http://www.hhs.gov/ohrp/humansubjects/guidance/gina/html.

[116] *See, e.g.*, Mark A. Rothstein, *Genetic Exceptionalism and Legislative Pragmatism*, 35 J.L. MED. & ETHICS 59, 63 (2007) (arguing that GINA may "lead the public to rely on the law to their detriment. It may also encourage people to believe in genetic exceptionalism, thereby making it a self-fulfilling prophesy, and it may erroneously convince legislators that they have resolved the underlying issues, thereby delaying enactment of meaningful legislation."). For this reason, the PGP strengthened its precautionary language regarding the risk of genetic discrimination following the passage of GINA. For this very reason the PGP *strengthened* the precautionary language contained in its informed consent agreements regarding the risk of genetic discrimination following the passage of GINA.

[117] *See, e.g.*, S. 976, 110th Cong. (2007); S.3822, 109th Cong. (2006); H.R. 6498, 110th Cong. (2008).

tion of a national biobank, or the provision of national funding for non-governmental biobanks such as the one begun by the PGP.[118]

Under the GPMA, the IWG must "develop[] and promulgat[e] guidelines regarding procedures, protocols, and policies for the *safeguarding of the privacy of biobank subjects*. . . ."[119] The GPMA also contemplates the award of grants to biobanking initiatives, provided they comply with the IWG's regulations and, as a separate requirement, "establish mechanisms to ensure patient privacy and protection of information . . . ."[120] The GPMA presupposes that privacy is not only possible, but of paramount importance.[121] As discussed in Part II, this assumption is frequently incompatible with the public genomics model. Equally important, although the GPMA purports to "encourage development of genomic tests and therapies; and [to pay] greater attention to the quality of genetic tests, direct-to-consumer advertising and the use of personal genomic information,"[122] it fails to distinguish between research, clinical and commercial genetic testing.

More broadly, the increasing availability of genomic data is driving the development of a range of genomic activities. These include genomic research, use of genomic data in clinical care and commercial genetic testing and other genomic services targeted at consumers (as opposed to patients or healthcare providers). Increasingly, the distinctions between research, clinical care and consumer products are becoming blurred. Thus, enabling public genomic research will require appropriate distinctions between the generation and return of genomic data in the research, clinical and commercial contexts, including distinguishing public genomics research from other forms of non-public genomic research.[123] Federal regulation to date misses this

---

[118] *Id.*

[119] H.R. 6948, § 4(b)(5) (emphasis added).

[120] *Id.* § 5(c)(3)(G).

[121] This is consistent with other national legislative efforts, including the UK Biobank project. *See* UK Biobank, Confidentiality, http://www.ukbiobank.ac.uk/faqs/confidentiality.php (last visited May 3, 2010); Untersuchungen bei Menschen [Human Genetic Examination Act] 374/09 (Apr. 24, 2009), http://www.eurogentest.org/uploads/1247230263295/GenDG_German_English.pdf (providing a translation of the act).

[122] The GPMA "focuses on the safety, efficacy and availability of information about genetic tests, including pharmacogenetic and pharmacogenomics tests." These and other provisions are meant to aid in paying "greater attention to the quality of genetic tests, direct-to-consumer advertising and use of personal genomic information." 153 Cong. Rec. S3709 (daily ed. Mar. 23, 2007) (statement of Sen. Obama).

[123] *See, e.g.*, Dan Vorhaus, *Genomic Research Goes DTC*, GENOMICS L. REP., July 9, 2009, http://www.genomicslawreport.com/index.php/2009/07/09/genomic-research-goes-dtc/.

mark, and as described in the next section, the scene at the state level is not much better.

Finally, although not specific to genomic research, protections applicable to human subjects research must also be included in any list of federal laws affecting public genomics. In the United States, the "Common Rule" is the definitive standard for human subjects research.[124] The Common Rule grew out of a series of regulations from the Department of Health and Human Services (HHS) that was later adopted by a number of Federal departments and agencies. While not perfectly comprehensive or preemptive, because it does not cover every Federal agency and does not cover certain privately sponsored research,[125] the Common Rule does apply to "all research involving human subjects conducted, supported or otherwise subject to regulation by a federal department or agency" that has adopted it.[126] Although the cost of sequencing an individual genome is falling quickly, the large cohorts sought by many public genomics projects means that research costs remain high, with researchers generally seeking federal support. Moreover, genomic researchers are frequently associated with research universities and other institutions that have signed Federal-wide Assurances, which are institution-wide agreements to comply with Common Rule protections on all research projects, regardless of whether any individual project receives federal funds. Accordingly, most public genomics projects will be subject to the Common Rule.[127]

---

[124] The Federal Policy for the Protection of Human Subjects is often referred to as the "Common Rule" and is codified at 45 C.F.R. §§ 46.101-46.124 (2008). The Common Rule sets forth fundamental human subjects research protections (HSRP) for almost all government-funded research in the United States, and most academic and other research institutions require compliance with the Common Rule regardless of the source of founding. While the Common Rule supplies a floor and not a ceiling for HSRP, there appears to be little evidence of state or local lawmakers exercising their authority to "provide additional protections for human subjects." *See* § 46.101(f) (2008).

[125] Vorhaus, *supra* note 123. While most DTC genomic research is not, at this point in time, covered by the Common Rule or similar regulations, research conducted by pharmaceutical, medical device and other companies subject to FDA regulation, although not covered by the Common Rule, is subject to FDA human subject protection regulations that closely track the Common Rule. *See generally* U.S. Food & Drug Administration, Comparison of FDA and HHS Human Subject Protection Regulations, http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/EducationalMaterials/ucm112910.htm (last visited May 3, 2010).

[126] 45 C.F.R. § 46.101(a) (2008).

[127] To the extent that a new DTC research model develops and proves to be sustainable, certain public genomics projects may not be covered, at least initially, by existing human subjects research protections, including the Common Rule. Any such gap in coverage for research participants would be addressed, eventually, by legisla-

Compliance with the Common Rule is the responsibility of "each institution engaged in research which is covered by [the] policy."[128] At most institutions, responsibility for Common Rule and other human subjects research compliance falls to an Institutional Review Board (IRB), and the institution is required to "certify that each application or proposal for research covered by the assurance . . . has been reviewed and approved by the IRB,"[129] the composition and operations of which is also spelled out in the Common Rule.[130] The Office for Human Research Protection (OHRP), a division of HHS, reviews institutional compliance with the Common Rule. While the OHRP has the authority to investigate reported violations of the Common Rule and, in limited cases, to undertake "not-for-cause compliance oversight evaluations,"[131] the federal government does not conduct any initial review or approval of human subjects research projects. This task has been left exclusively to the individual IRBs.

Under the Common Rule, the level of IRB supervision corresponds to the level of risk that a project presents to its human subjects,[132] with a conception of "risk" that generally involves the likelihood of physical injury to study participants.[133] Because public genomics is non-therapeutic, physical risks to participants are low. Under a broader conception of risk, however, participants in public genomics research face the possibility of significant adverse consequences from their participation. These adverse consequences include social harms and other risks that are difficult to forecast, and thus to proscribe. For example, participants of childbearing age risk disclosing to potential spouses that they carry genes associated with an increased risk of developing diseases such as Huntington's disease,

---

tive and regulatory action, likely by extending the Common Rule (or a close analogue) to DTC or other private or commercial research activities.

[128] 45 C.F.R. § 46.103(a) (2008).

[129] § 46.103(f) (2008).

[130] *See, e.g.*, IRB Membership, § 46.107 (2008); IRB Functions and Operations, § 46.108 (2008); IRB Review of Research, § 46.109 (2008); Criteria for IRB Approval of Research, § 46.111 (2008).

[131] 42 U.S.C. § 289(b) (2006).

[132] *See, e.g.*, 45 C.F.R. § 46.109(e) (2008) ("An IRB shall conduct continuing review of research covered by this policy at intervals appropriate to the degree of risk . . . ."); § 46.116(a)(6) (placing more stringent informed consent requirements upon projects exposing subjects to more than minimal risk).

[133] *See, e.g.*, § 46.102(i) ("Minimal risk means that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests."); *but see* § 46.101(b)(2)(ii) (discussing research that "could be damaging to the subjects' financial standing, employability, or reputation").

Alzheimer's disease, or breast cancer that could have been inherited by their children or be passed along to future children. Such risks can be mitigated – in the case of the PGP, participants are given the option to withdraw from the study after receipt of their results but prior to publication[134] – but only as to genes that are identified and whose function is understood at that particular point in time. Indeed, as previously discussed, the most significant risks to participants in public genomics research may be presently unknown. Accordingly, while the Common Rule plays an important role in protecting human research subjects, its limited conception of risk leaves it ill-equipped to protect human subjects and guide public genomics investigators in conducting responsible research.

2. State Law

As the clinical and commercial applications of genomic research become more widespread, state legislatures have also jumped into the fray, enacting a morass of widely varying regulations. While state legislation presents a challenge to genomic research in many forms, the problem is especially acute with regard to public genomics in areas ranging from privacy to the unauthorized practice of medicine.

a. Regulation of DTC Genomic Service Providers

The sentiment behind much of the legislation directly applicable to public genomics involves licensing and disclosure requirements motivated by suspicion of direct-to-consumer (DTC) genomic services, including genetic testing. DTC service providers usually reach consumers through the Internet, television, or in some cases, over-the-counter, with the consumer providing a tissue or fluid sample and returning it to a laboratory for testing.[135] The company then provides the customer with individualized information about that person's genetic heritage, paternity, or propensity for certain traits or diseases. DTC genomic companies vary in the degree to which they interact directly with consumers and patients or rely on genetic counselors, physicians, or other healthcare professionals to mediate that relationship.[136]

---

[134] PGP CONSENT FORM, *supra* note 90, at § V, 4.3.

[135] Fed. Trade Comm'n, At-Home Genetic Tests: A Healthy Dose of Skepticism May Be the Best Prescription, *available at* http://www.ftc.gov/bcp/edu/pubs/consumer/health/hea02.pdf (last visited May 3, 2010).

[136] Examples of the former include 23andMe and Pathway Genomics, which return results directly to individuals via a website interface. Examples of the latter include Navigenics, Knome, and Illumina which, while returning information directly

Public genomics projects in which a participant provides a tissue or fluid sample to be sequenced in a laboratory, frequently without consultation with a physician, would likely qualify as a DTC service or test under many legislative definitions.[137] Because DTC testing often removes one-on-one professional counseling and associated safeguards, a great deal of public concern has been voiced about the use of DTC testing in complex medical practice areas such as genetics.[138]

Public genomics projects and other biobanking organizations, unlike DTC service providers, do not collect or sell – or conduct research using – genomic data for commercial benefit. And yet, despite their non-commercial focus, and despite robust informed consent procedures and existing human subjects research protections that comprehensively address the concerns underlying DTC statutes, legislation governing DTC clinical tests fails to acknowledge those distinc-

---

to individuals, rely on healthcare professions to mediate at least one key stage (either ordering or receiving and interpreting) of the process.

[137] Both California and New York, for instance, took regulatory action against DTC testing and service providers in 2008 for their alleged failure to appropriately involve a physician or meet other clinical or regulatory requirements. *See, e.g.*, Robert Langreth & Matthew Herper, *States Crack Down on Online Gene Tests*, FORBES.COM, Apr. 4, 2008, http://www.forbes.com/2008/04/17/genes-regulation-testing-biz-cx_mh_bl_0418genes.html.

[138] *See* Leslie Pray, *DTC Genetic Testing: 23andMe, DNA Direct and Genelex*, 1 NATURE EDUC. (2008), *available at* http://www.nature.com/scitable/topicpage/DTC-Genetic-Testing-23andme-DNA-Direct-and-674; *see also* Howard Wolinsky, *Do-It-Yourself Diagnosis*, 6 EMBO REP. 805, 806-07 (2005) (discussing the benefits and risks of DTC testing); Meredith Wadman, *Gene-Testing Firms Face Legal Battle*, 453 NATURE 1148 (2008) (discussing the legal battle over DTC testing in New York and California). Critics of direct-to-consumer genetic testing are concerned that the results of genetic tests are not as accurate as the public perceives them to be, and fear that customers might "make unwarranted, and even irrevocable, decisions on the basis of test results and associated information, such as the decision to terminate a pregnancy, to forgo needed treatment, or to pursue unproven therapies." Kathy Hudson et al., Am. Soc'y of Human Genetics, *ASHG Statement on Direct-to-Consumer Genetic Testing in the United States*, 81 AM. J. HUM. GENETICS 635, 635-36 (2007); Barbara Ameer & Norberto Krivoy, *Direct-to-Consumer/Patient Advertising of Genetic Testing: A Position Statement of the American College of Clinical Pharmacology*, 49 J. CLINICAL PHARMACOLOGY 886, 888 (2009). Accordingly, much of the existing and proposed state legislation requires licensing and disclosure and/or the involvement of physicians or other healthcare professionals to ensure paying consumers of genetic testing are not duped or led to believe the results of such tests are more predictive than they actually are. Unfortunately, depending on how such legislation is drafted, it may be a fairly simple task to circumvent certain requirements, for example a physician involvement requirement, by simply structuring the service model to accommodate the legislation which serves to satisfy the letter but not the spirit of the requirement, while simultaneously creating additional administrative expenses for companies, individuals and states.

tions. Most state statutes appear to assume that any direct interaction between individuals and a genomic activity involves clinical genetic testing. As of May 2008, thirteen states prohibited all DTC testing.[139] Of the twelve states that allowed limited DTC testing, many of them permitted only extremely specific, non-genetic tests, thereby implicitly banning DTC genetic tests altogether.[140]

This failure to recognize the distinction between clinical and research applications has the potential to subject public genomics research to incidental regulation under clinically focused laws. For example, Idaho's outright prohibition on DTC testing precludes the ordering of tests by anyone other than "practitioners legally authorized to *diagnose, treat and prescribe*."[141] This regulation, while intended to prevent unauthorized access to clinical testing, lacks the nuanced understanding of commercial, clinical, and research genomics necessary to produce efficient and targeted regulation of DTC testing. Indeed, despite the fact that the Idaho regulations appear to be directed at the clinical concerns caused by DTC testing, state officials have confirmed that all DTC testing is prohibited in Idaho.[142]

Ultimately, to the extent that DTC laws or regulations sweep up true public genomics activity, they likely do so either unintentionally or without a complete understanding of the distinctions between the variety of existing genomic services and research. To avoid unintentionally restricting the legitimate activities of responsible public genomics, more finely-tuned drafting and enforcement of existing and future regulations is needed.

b. Licensing of Genetic Counselors and Physicians

Public genomics researchers who desire to share research results with study participants face special difficulties in complying with state laws, whether those laws are specific to DTC genomic services or part of broader efforts to regulate genetic counseling and the prac-

---

[139] Genetics & Pub. Policy Ctr., Direct-to-Consumer Genetic Testing: Empowering or Endangering the Public?, http://www.dnapolicy.org/policy.issue.php?action=detail&issuebrief_id=32 (last visited May 3, 2010) (discussing the dangers of false or misleading advertising in the context of DTC testing).

[140] *See, e.g.*, MD. CODE REGS. 10.10.01.03B(28) (2009) (allowing DTC "Health Awareness Tests," which are limited only to cholesterol screening).

[141] IDAHO ADMIN. CODE r.16.03.14.350.07 (emphasis added). Indeed, the regulation assumes that tests will be ordered in the context of a doctor-patient relationship and that the test results will be incorporated into the patient's medical record.

[142] Genetics & Pub. Policy Ctr., Survey of Direct-to-Consumer Testing Statutes and Regulations, June 2007, *available at* http://www.dnapolicy.org/resources/DTCStateLawChart.pdf.

tice of medicine. At least twelve states have enacted or considered laws establishing licensure requirements for genetic counselors.[143] Massachusetts, for example, defines the "practice of genetic counseling" to include "helping the individual . . . understand genetic or prenatal tests . . . and interpret complex genetic test results."[144] Would investigators in a public genomics research project need to acquire state genetic counseling licenses to share results with study participants in Massachusetts or other states where similar legislation has been enacted? What level of information-sharing would trigger the need to obtain such a license? What if the test results were based only on a computerized analysis of the sequencing data and, other than the variations in the data itself, were not otherwise tailored to the research participant? At present, none of these questions have clear answers, and public genomics researchers are left to balance the legal risks against uncertain professional, ethical, and legal obligations to return data to participants.

Sharing research data with participants also implicates state laws that forbid the practice of medicine by unlicensed doctors, known as the unauthorized practice of medicine.[145] Violating the prohibition against the unauthorized practice of medicine constitutes a misdemeanor or felony under most state medical practice acts. State medical practice acts generally provide that a violation of the prohibition against the unauthorized practice of medicine is a misdemeanor or felony,[146] and licensed physicians who aid the unauthorized practice of medicine may be subject to disciplinary action.[147] Depending on the phrasing of the law and the specifics of the research protocol, a public genomics project that shares research data with participants may arguably fall within the scope of the law. For example, in New York, the practice of medicine is defined as "diagnosing, treating, operating or prescribing for any human disease, pain, injury, deformity or physical condition."[148] Would the identification of a monogenic, dominant condition – for example, Huntington's disease – constitute the diagnosis of a disease or medical condition within the meaning of

---

[143] ALISSA JOHNSON, NAT'L CONFERENCE OF STATE LEGISLATURES, NCSL GENETICS BRIEF: GENETIC COUNSELOR LICENSING, http://www.ncsl.org/default.aspx?tabid=14276 (last visited May 3, 2010).

[144] MASS. GEN. LAWS ch. 112, § 227 (2009).

[145] *See, e.g.*, N.C. GEN. STAT. § 90-18 (2009) ("No person shall perform any act constituting the practice of medicine . . . without being duly licensed and registered.").

[146] *See, e.g., id.*; MINN. STAT. § 147.091(1)(i) (2009); GA. CODE § 43-34-37(a)(9) (2008).

[147] *See, e.g.*, 225 Ill. Comp. Stat. 60/22 (2008).

[148] N.Y. EDUC. LAW § 6521 (Gould 2001).

such statutes? Because activities that involve diagnosis may be considered the unauthorized practice of medicine, a public genomics project could be found to be "practicing medicine" if it shared its findings with participants.

c. Genetic Privacy

Finally, a number of states directly regulate the privacy of genetic and genomic information. In most cases, state legislation simply requires informed consent before an individual's genetic information may be shared with a third party. For example, Minnesota law provides that genetic data "may be disseminated only with the individual's written informed consent."[149] Accordingly, a research protocol that complies with the Common Rule would also likely comply with this law. Other state laws governing disclosure are not so clear. California, for instance, requires written authorization "for each separate disclosure of the test results, and shall include to whom the disclosure would be made."[150] This language clearly contemplates private (i.e., consumer or clinical) genetic testing, but how would it apply to public genomics research? When an individual's genomic data is published to a public Web site, as with the PGP, what constitutes a "disclosure"? Is a disclosure the act of placing the information on a public server? Does a disclosure occur each time a visitor to the Web site access the information? Would a statement that the disclosure is being made to the world at large satisfy the requirement that the written authorization "include to whom" the information is being disclosed, or is something more specific required?[151]

---

[149] MINN. STAT. § 13.386(3)(4) (2009).

[150] CAL. INS. CODE § 10149.1(g). Violations of this statute are punishable by a civil penalty of between one and five thousand dollars, plus court costs. The California authorization is similar to that imposed by the Health Insurance Portability and Accountability Act (HIPAA) at 45 C.F.R. § 164.508(c) (2008). With respect to HIPAA, the so-called "Privacy Rule," which took effect in 2003, regulates the use and disclosure of certain medical data by "covered entities." § 164.501. Although the "protected health information" covered by the Privacy Rule certainly encompasses the information received and published by most DTC genomic service providers and genomic research projects, the term "covered entity" includes health care providers, health care clearinghouses and health plans and a health care provider "furnishes, bills, or is paid for health care in the normal course of business." § 160.103; *see also* CTRS. FOR MEDICARE & MEDICAID SERVS., U.S. DEP'T HEALTH & HUMAN SERVS., COVERED ENTITY CHARTS, http://www.cms.hhs.gov/HIPAAGenInfo/Downloads/CoveredEntitycharts.pdf (last visited May 3, 2010). It is doubtful that, at least under current interpretations, either commercial or research genomic entities of the type generally described in this Article would be covered by HIPAA's Privacy Rule or by HIPAA, generally.

[151] Statutes that require authorizations to take a particular form or meet spe-

Even more troubling for public genomics is a complete bar against any disclosure. Under Georgia law, "genetic testing may only be conducted to obtain information for therapeutic or diagnostic purposes."[152] The legislature included an exception for scientific research, but with the limitation that "any research facility may conduct genetic testing and may use the information derived from genetic testing for scientific research purposes *so long as the identity of any individual tested is not disclosed to any third party . . . .*"[153] The law provides no apparent exception for disclosure with informed consent.

* * *

The nuances, ambiguities, and outright prohibitions created by state legislative activity create significant hurdles for the progress of public genomics, rendering compliance extremely difficult in the large, multi-state research protocols typical of such research. While many of these laws apparently target clinical testing, research-oriented initiatives such as the PGP are loathe to run afoul of the plain language of state law for obvious reasons, including their reliance on public funding, their desire to carry out research responsibly, and their inability to shoulder ancillary expenses such as compliance penalties or litigation expenses. Yet few if any research budgets account for the costs of ensuring legal and regulatory compliance. In any case, public genomics projects should be able to conduct research with something more concrete than a mere hope that, if and when an issue arises, the state enforcement agency will adopt a lenient interpretation of the relevant statute, regulation, or rule. Until a broader regulatory scheme clarifies the parameters of permissible genomics research in the various state jurisdictions, the absence of predictable legal outcomes may discourage or even preclude research-oriented projects from operating in certain states. Significant genetic populations and their data will be excluded from participation in public genomics research, biasing the results and depressing enrollment numbers[154] while simultaneously failing to fully protect research subjects.

---

cialized prescriptions present a similar problem. For example, Alaska law provides for a state-authored consent form. *See* ALASKA STAT. § 18.13.010(c) (2008) ("The Department of Health and Social Services may by regulation adopt a uniform informed and written consent form to assist persons in meeting the requirements of this section.").

[152] GA. CODE ANN. §33-54-3(a) (2005).

[153] § 33-54-6 (emphasis added).

[154] Although the PGP has not had this problem to date, at least one other large-scale genomic research project, a collaborative project between the Scripps Translational Science Institute and the DTC genomics company Navigenics, has had difficulty meeting its target enrollment numbers. *See* Daniel MacArthur, *Enrollment in Navigenics-Scripps Study Disappointing*, GENETIC FUTURE, http://scienceblogs

3. International Law

Although international law generally does not bind either researchers or regulators to follow particular standards in conducting genetic research, it has played a critical role in the development of ethical understandings of human subjects research and is a key component of responsible public genomics. The Nuremberg Code, a set of ethical principles intended to guide research involving human subjects, has played an especially important role.[155] Drafted by the American judges who tried twenty-three Nazi doctors and researchers for crimes against humanity, including grotesque experiments that exposed concentration camp prisoners to mustard gas and freezing temperatures, the Nuremberg Code sets forth a series of ethical principles governing research on human subjects.[156] Voluntary consent is "absolutely essential," and obtaining consent is the "personal duty and responsibility" of the researcher.[157] As with the Common Rule, proportionality principles that balance potential harms and gains also play an important role in the Nuremberg Code.[158] Other important international law statements on human subjects research include the World Medical Association's Declaration of Helsinki,[159] a nonbinding but highly influential set of ethical principles for human subjects research, and the International Convention on Civil and Political Rights.[160] Although these principles lack binding force in the United States, courts faced with cases involving human research subjects have looked to the Nuremberg Code and the Declaration of Helsinki to guide their analysis.[161] In addition to guiding courts faced with litigation involv-

---

.com/geneticfuture/2009/02/enrollment_in_navigenics-scrip.php (last visited May 3, 2010).

[155] *See* NUERNBERG MILITARY TRIBUNALS, *supra* note 87, at 181.

[156] *See, e.g.*, NAT'L INSTS. OF HEALTH, OFFICE OF HUMAN SUBJECTS RESEARCH, DIRECTIVES FOR HUMAN EXPERIMENTATION: NUREMBERG CODE, http://ohsr.od.nih.gov/guidelines/nuremberg.html (last visited May 3, 2010).

[157] *See* NUERNBERG MILITARY TRIBUNALS, *supra* note 87, at 181-82.

[158] *See id.* at 182 ("The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment."). Other key principles involve protecting research subjects from harm and justifying experiments by their expected result.

[159] World Med. Ass'n, *World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*, ¶¶ 20, 22, Oct. 2008, *available at* http://www.wma.net/en/30publications/10policies/b3/index.html.

[160] International Covenant on Civil and Political Rights art. 7, Dec. 16, 1966, 999 U.N.T.S. 171, 6 I.L.M. 368 (stating "[N]o one shall be subjected without his free consent to medical or scientific experimentation.").

[161] *See* Grimes v. Kennedy Krieger Inst., Inc., 782 A.2d 807, 835 (Md. 2001)

ing human research subjects, the ethical norms embodied in these international documents will likely play a role in structuring future international agreements or domestic regulations. Because these international norms also emphasize research subject privacy, they also raise concerns for public genomics similar to those discussed with regard to federal and state legislation.

In the specific area of genomic research, the United Nations Educational, Scientific and Cultural Organization (UNESCO) published the Universal Declaration on the Human Genome and Human Rights in 1997[162] and the International Declaration on Human Genetic Data in 2003.[163] Both of these nonbinding statements set forth principles intended to govern genomic research and recognize the ways that the particular attributes of genomic data impact issues of research ethics. The Declaration on Human Genetic Data, for example, accords human genomic data a special status because it "may contain information the significance of which is not necessarily known at the time of the collection of the biological samples" and "may have a significant impact on the family, including offspring, extending over generations, and in some instances on the whole group to which the person concerned belongs."[164] The 2003 Declaration also contains useful protections for research subjects, including the principle that "no one should be denied access to his or her own genetic data . . . ."[165] While the Declaration does not bind the United States, other countries may look to it for guidance, and state courts may consider the principles it elucidates in deciding cases involving public genomics research, much as they have done with the Nuremberg Code and the Declaration of Helsinki.

(holding that the Nuremberg Code "should be the preferred standard for assessing the legality of scientific research on human subjects"); *cf.* Abdullahi v. Pfizer, 562 F. 3d 163 (2d Cir. 2009) (determining that the Nuremberg Code and Declaration of Helsinki are evidence of an international norm against nonconsensual human experimentation).

[162] Universal Declaration on the Human Genome and Human Rights, UNESCO, 29th sess., 26th plen. mtg. (Nov. 11, 1997), *available at* http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html.

[163] International Declaration on Human Genetic Data, UNESCO, 32d sess., 20th plen. mtg. (Oct. 16, 2003), *available at* http://portal.unesco.org/en/ev.php-URL_ID=17720&URL_DO=DO_TOPIC&URL_SECTION=201.html

[164] *Id.* art. 4.

[165] *Id.* art. 13. This requirement is in tension with the potential legal liabilities, discussed earlier in this Part, associated with the return of individualized genomic data.

B. Common Law

Researchers in the field of public genomics must also consider the role that courts may play in answering the difficult policy questions posed by this developing field. While some state laws governing genomic information specifically provide a private right of action for their violation,[166] most litigation related to public genomics will likely be based on tort law claims such as researcher negligence or medical malpractice. As discussed in Part II, participants in public genomics research face a variety of risks, many of which cannot be articulated at the time participants consent to join the study. When these risks are realized and a participant is harmed – an outcome that is statistically likely in public genomic projects with large numbers of participants – the form of any resulting litigation will likely depend upon the form of the harm, neither of which can be known in advance.

When a participant in a public genomics project becomes sick or dies as a result of a condition with an identified genetic component, medical malpractice[167] litigation could result. The theory would be that the physician-researchers owed a duty of due care to the participant as a "patient," which they breached by failing to discover the condition or disclose it to the participant. The primary barrier to a plaintiff's claim for medical malpractice liability would be the requirement that the plaintiff show the existence of a physician-patient relationship that imposes a duty of care. While researchers have generally been considered outside the scope of medical malpractice claims,[168] the nature of genomic data may change a court's analysis. From an ethical standpoint, providing participants with a preliminary interpretation of their raw genomic data, as the PGP does, provides usable information to research subjects, encourages participation, and aids in satisfying the researchers' ethical and legal duties to return

---

[166] *E.g.*, ALASKA STAT. § 18.13.020 (2008) (providing civil right of action against a person who collects, analyzes, or discloses the results of a DNA sample in violation of Alaska's genetic testing laws).

[167] Different states permit different claims, and "medical malpractice" claims might also come in the form of any of the following: wrongful death action, wrongful pregnancy action, wrongful birth action, wrongful life action, infliction of emotional distress, and even negligence. Note that some courts have opined that certain medical malpractice actions, such as the action for wrongful birth following negligent genetic testing or for the failure to diagnose fetal defect, do not fall under traditional tort analysis and rather involve matters of public policy better left to legislature. *See, e.g.*, Coleman v. Dogra, 844 N.E.2d 1190 (Ohio 2006).

[168] *See* Ande v. Rock, 647 N.W.2d 265 (Wisc. Ct. App. 2002) (holding that parents suing physician-researchers for failing to disclose that their newborn tested positive for cystic fibrosis failed to allege any physician-patient relationship between themselves and physician-researchers).

and/or interpret data in certain contexts. Yet by returning or interpreting genomic data, a researcher may mitigate one form of potential liability while exacerbating another, as the researcher's decision to provide analysis of raw genomic data blurs the conceptual divide between non-clinical research and the practice of medicine.[169]

A participant-plaintiff could claim a physician-patient relationship with an investigator in a public genomics project, or that a genetic counselor associated with the project – perhaps put in place precisely to satisfy a duty to provide guidance to participants receiving research results – should be liable for the participant-plaintiff's harms, depending on the exact interactions between the genetic counselor and the participant-plaintiff. Possible claims here include the failure to be properly informed about available follow-up genetic tests, especially in the cases where further diagnostic testing would be required to determine the nature and extent of a fetal defect or other significant disease;[170] failure to diagnose and/or discover an inheritable genetic impairment;[171] failure to give sufficient or adequate genetic counseling, including ascertainment of carrier status; failure to timely refer a participant to a genetic specialist; and loss of chance.[172] Although

---

[169] Refusing to return data to participants does not resolve the issue, especially in states where the legislature has enacted legislation making an individual's genetic data his or her personal property. *See, e.g.*, COLO. REV. STAT. § 10-3-1104.7(1)(a) (2006) ("Genetic information is the unique property of the individual to whom the information pertains."). And returning data without any interpretation whatsoever presents its own difficulties, as described in Part II.

[170] For example, Maryland recognizes a cause of action in tort:

> for wrongful birth when the doctor does not inform the patient about available diagnostic tests which might reveal the possibility of neural tube defects of the fetus, when these defects are genetically caused, when further diagnostic testing would be required to determine the nature and extent of any fetal defects, and when the plaintiff asserts she would have aborted child had she been made aware of the fetal deformities.

Reed v. Campagnolo, 630 A.2d 1145, 1146-52 (Md. 1993) (quoting Reed v. Campagnolo, 810 F.Supp. 167, 172-73 (D. Md. 1993)); *see also* Ferrell v. Rosenbaum, 691 A.2d 641 (D.C. 1997) (finding that a physician-geneticist breached applicable standard of care by failing to review and follow up on information contained in child's medical records indicating abnormal blood test results where the failure to timely diagnose lead to harm. By the time diagnosis was made, mother-plaintiff and father were estranged, and therefore, geneticist's negligence robbed mother-plaintiff of the opportunity to conceive additional children who might serve as bone marrow donors for the sick child.).

[171] *See* Chen v. Connors, 896 N.E.2d 656 (Mass. App. Ct. 2008) (holding that a physician's alleged failure to evaluate a patient as a potential thalassemia carrier was not a clear deviation from good medical practice where the physician was not a genetic counselor and there was no record showing that the patient informed the physician that her twin sister was found to be a carrier for thalassemia).

[172] The "loss of chance" doctrine involves issues of causation in medical

some of these claims can be avoided with preemptive structuring on the part of the public genomics project,[173] not all of these potential claims are amenable to structural solutions. Furthermore, such structural changes restrict the universe of potential participants, introducing bias into the population and weakening the quality of the research data generated by the project.

Depending on the specific facts of its particular research protocol, a public genomics project has a variety of arguments in defense of its structure, including the fact that there is no "patient" in the relationship at all. Researchers could also argue that the plaintiff-participant did not consult the public genomics project or the genetic counselor as a health care provider; that the public genomics project and the genetic counselor are not engaging in the practice of medicine; and that the participant-plaintiff should not expect to receive medical services for a medical condition. All of these defenses may be bolstered by the informed consent protocols implemented by the project.[174] Nonetheless, because the existence of a physician-patient relationship is ultimately a question of fact for the jury,[175] public genomics projects face substantial litigation risk related to medical malpractice claims, even if those claims are ultimately rejected by the courts. Finally, even frivolous or meritless litigation is extremely worrisome given the shoestring budgets of many scientific research projects.

If the participant is unable to prove the existence of a physician-patient relationship, researchers may still owe other common law duties toward the project participants. State courts have only recently begun to grapple with this issue, but at least one case suggests that state courts may closely review the merits of research protocols in tort suits brought by participants. In *Grimes v. Kennedy Krieger Institute, Inc.*, the Maryland Court of Appeals decided a case involving claims

---

malpractice claims. When a doctor fails to diagnose a condition and the patient is injured, some courts hold that the doctor can be held liable in tort. *See, e.g.*, Matsuyama v. Birnbaum, 890 N.E.2d 819, 823 (Mass. 2008) (explaining "Where a physician's negligence reduces or eliminates the patient's prospects for achieving a more favorable medical outcome, the physician has harmed the patient and is liable for damages. Permitting recovery for loss of chance is particularly appropriate in the area of medical negligence.").

[173] The PGP, for instance, initially declined to enroll individuals that had not finished having children, in part to address such potential risks.

[174] *See, e.g.*, PGP CONSENT FORM, *supra* note 90, at § VII. (containing several provisions designed to inform prospective participants that (i) data supplied by the project is not intended or suitable for clinical or medical use, and (ii) the project will not supply any clinical, health or medical care to project participants).

[175] RICHARD E. SHANDELL & PATRICIA SMITH, THE PREPARATION AND TRIAL OF MEDICAL MALPRACTICE CASES 1-2 (rev. ed. 1994).

against researchers studying lead exposure in substandard housing in Baltimore.[176] The cost of mandatory lead abatement in Baltimore had become more expensive than the rental value of some properties, causing landlords to leave the properties vacant rather than to repair them.[177] Researchers from Johns Hopkins decided to investigate whether less costly abatement methods could allow the vacant units to be returned to the housing stock.

The researchers devised a protocol under which they would provide subsidies to landlords to perform varying levels of lead abatement, and then test the lead level in the residents' blood over a period of five years.[178] The researchers were especially interested in learning the effects on young children, who are the most likely to ingest ambient lead. Although the researchers obtained informed consent forms from the participants, authorizing the researchers to test the children's blood and to measure levels of lead dust within the house,[179] the informed consent failed to "clearly disclose[] . . . that the researchers contemplated that, as a result of the experiment, the child might accumulate lead in her blood, and that in order for the experiment to succeed it was necessary that the child remain in the house as the lead in the child's blood increased or decreased."[180] The lead levels in some of the children's blood increased, and several of the parents sued the researchers, arguing that the researchers breached a duty of care owed to the human subjects to warn them about and fully abate the lead in their homes, with consequent harm to the children.[181]

After an extended discussion of standards for human subjects research, including the Nuremberg Code and the Declaration of Helsinki, the Maryland Court of Appeals held that "the very nature of nontherapeutic scientific research on human subjects can, and normally will, create special relationships out of which duties arise."[182] This duty was justified in part because "the investigators are in a better position to anticipate, discover, and understand the potential risks to the health of their subjects" and should therefore be legally required to inform the participants of dangers to their health that might arise from time to time.[183] Furthermore, in addition to the duties in tort, the researchers owed contract duties to the participants, because the in-

---

[176] Grimes v. Kennedy Krieger Inst., 782 A.2d 807, 811-12 (Md. 2001).
[177] *Id.* at 821.
[178] *Id.* at 823.
[179] *Id.* at 824.
[180] *Id.*
[181] *Id.* at 832-33.
[182] *Id.* at 834-35.
[183] *Id.* at 851.

formed consent form constituted a binding contract supported by consideration.[184] The court concluded that the informed consent in this case had been insufficient because it did not inform the parents that the children would ingest lead, that testing the degree of ingestion was the entire point of the experiment, and that the child would have to remain in the house for the entire study period for the experiment to work. All the risks involved were clearly foreseeable to (and even intended by) the researchers from the inception of the experiment.[185] The court also expressed doubt regarding the impartiality of the IRB, which the court viewed as having helped the researchers avoid the application of Common Rule provisions applicable to research involving children.[186]

Viewed narrowly, *Grimes* is simply a case about informed consent, holding that researchers, just like physicians with their patients, must not expose their subjects to potential harm without their consent. And it seems that the researchers in the case were deficient in this regard, if they actually, as alleged, did not properly warn the participants and otherwise keep promises made to their subjects. Under this reading, *Grimes* is similar to other existing human subject research cases that examine whether the researchers obtained the participants' informed consent.[187]

The *Grimes* court, however, also seemed to assert a more expansive authority – the authority to go beyond the issue of consent and examine the validity of the research protocols themselves. The opinion, for example, notes that "there are conflicting views in respect to nontherapeutic research, as to whether consent, even of a person capable of consenting, can justify a research protocol that is otherwise

---

[184] *Id.* at 818, 859. The contract holding was based on the fact that the parents agreed to the testing in return for the test results and a small cash payment.

[185] *Id.* at 844.

[186] *Id.* at 814 ("[T]he IRB was willing to aid researchers in getting around federal regulations designed to protect children used as subjects in nontherapeutic research. An IRB's primary role is to assure the safety of human research subjects – not help researchers avoid safety or health-related requirements."). Similar concerns likely motivated the Harvard Medical School IRB's requirement that the PGP utilize a separate Data Safety Monitory Board (DSMB) to "monitor the progress of the PGP, including the risks to study participants," on an ongoing basis. PGP CONSENT FORM, *supra* note 90, at §VII(1.6). The PGP's informed consent agreement also provides for ongoing direct participant monitoring through the collection of regular (every three months) "safety questionnaires" that attempt to assess, among other items, whether participation has resulted in any adverse experiences. *Id.* at § VI(5).

[187] *See* Greenberg v. Children's Hosp. Research Inst., Inc., 264 F. Supp. 2d 1064, 1070 (S.D. Fla. 2003); Whitlock v. Duke Univ., 637 F. Supp. 1463, 1467 (M.D.N.C. 1986).

unjustifiable."[188] The court also quoted at length from a law review article that said:

> This "justifying" side of consent raises some timeless and thorny questions. What if people consent to activities and results which are repugnant, or even evil? . . . We have three ways out: We can say, first, "Yes, consent justifies whatever is consented to – you consented, so case closed;" second, "This particular consent is deficient – you did not really consent and so the result or action is not justified;" or third, "You consented, but your consent cannot justify this action or result."[189]

Although the *Grimes* court focuses mostly on the second approach, the opinion leaves little doubt that the court believes that it is empowered to examine the third, at least in research where children are involved.[190]

This case raises the question of who should decide whether human subject research is justified, even in the presence of truly informed consent to that research. Should the researchers themselves decide, under the guidance of expert institutional review boards? Should state courts play a role? The *Grimes* court left no doubt of its position on this question, stating that "the assessment of the legal effect of research on human subjects must always be subject to judicial evaluation."[191] Indeed, throughout the opinion, the court illustrates the tension between judicial and scientific evaluation of issues surrounding human subjects research, noting that "to turn over human and legal ethical concerns solely to the scientific community[] is to risk embarking on slippery slopes, that all too often in the past, here and elsewhere, have resulted in practices we, or any community, should be ever unwilling to accept."[192]

---

[188] Grimes v. Kennedy Krieger Inst., 782 A.2d 807, 856 (Md. 2001); *see also id.* at 848 (relying on a federal regulation that provided "[t]he research will be conducted in accordance with sound ethical principles").

[189] *Id.* at 856-57 (quoting Richard W. Garnett, *Why Informed Consent? Human Experimentation and the Ethics of Autonomy,* 36 CATH. LAW. 455, 458-60 (1996)); *see also* Richard W. Garnett, *Why Informed Consent? Human Experimentation and the Ethics of Autonomy*, 34 CATH. LAW. 455, 490 (1996).

[190] *See Grimes*, 782 A.2d at 857 ("[N]o degree of parental consent, and no degree of furnished information to the parents could make the experiment at issue here, ethically or legally permissible. It was wrong in the first instance.").

[191] *Id.* at 817.

[192] *Id.* at 853.

The institutional competence argument underlying *Grimes* – that courts must have a role in determining ethical issues involving human subjects research – has important implications for public genomics. A participant in a public genomics project who suffered a negative consequence, even one discussed in the informed consent materials and explicitly acknowledge by the participant to be a risk, might be able to sue the project if he or she can convince the court that the project was "wrong in the first instance."[193] Researchers should also consider the fact that a court's power is not limited to after-the-fact awards of money damages – issues of validity could just as easily arise in a proceeding where an opponent of the research project sought to stop it by obtaining a temporary restraining order or preliminary injunction. The likelihood of a court finding a particular research project unjustifiable is unfortunately heightened by the novelty and uncertainty surrounding public genomics and the challenge it brings to traditional notions of research ethics. This problem is exacerbated by the fact that the benefits of long-term scientific research (including public genomics) may take years to become apparent, while the harms participants are warned about (such as loss of a job or insurance) could take place in the near future. Consequently, a state court confronted with a near-term harm, and long-term but as-yet unrealized benefits, could be inclined to find the research unjustifiable. As with federal and state statutes and regulations therefore, common law adjudication of issues raised by public genomics imposes serious and uncertain risks.

C. Problems with the Current Legal Regime

Most of the deficiencies of the current legal regime should now be apparent. Much of the regulation is incidental, with genomic research falling only accidentally within the coverage of laws governing medical practice or direct-to-consumer genomic services. Even when regulation is intentionally directed towards genomic research, the assumptions underlying that regulatory policy – for example, the assumptions that anonymity or privacy are inviolate components of any genomics research protocol – are often inconsistent with the public genomics model.

Beyond the scope of any particular law or regulation, the fact that each state has largely taken its own approach to regulating genomics services and research is itself a serious problem. As with GINA, the federal government generally endeavors to avoid preemption of state law when it enacts federal laws, preferring instead to set a minimum

---

[193] *Id.* at 858.

standard that states may expand upon as they see fit. While federalism may support this approach in principle, the resulting patchwork of regulation means that the compliance costs will be very high for any public genomics projects with national or international ambitions. Moreover, as demonstrated by *Grimes*, compliance with state and federal law will not necessarily mitigate the legal risk faced by public genomics projects.

Finally, among the most interesting questions raised by public genomics is how the current legal structure will handle problems that researchers, regulators, or participants cannot anticipate. The risk that an insurer will discriminate against an individual with a genetic predisposition for a life-threatening condition, while difficult to quantify precisely, represents a risk that everyone can grasp. But what of the risks that no one has even considered? The existing legal structure, which strains to accommodate even the known, present-day issues presented by genomic research, lacks any mechanism to respond promptly and effectively to new developments. Without both a consensus understanding of the goals of public genomics – including the permissible methods or practices for pursuing such goals – and a legal regime flexible enough to adjust to unanticipated developments and risks as they arise, public genomics research will be stymied by the confusing, inconsistent morass that is the current legal regime.

IV. THE WAY FORWARD

While public genomics promises immense human benefit, it differs from traditional research models in ways that create a tension among competing principles. How does one protect the legitimate privacy expectations of individuals while engaging in research whose benefits derive in part from broadly sharing individually identifiable data? Can consent to participation be truly informed when the risks of participation cannot be fully known? And at what point do efforts to protect participants from risk violate the principles of individual autonomy and respect?

Finding a way to enable a future of responsible public genomics requires lawmakers and researchers to cooperatively develop a legal and regulatory structure that protects participants while enabling important scientific research to proceed in a responsible fashion. Researchers must learn to cope with the legal duties imposed by regulators, and regulators must adapt the present legal structure to the new challenges presented by the advance of genomic research. The goal of this Article is to explore these tensions, and to propose ways to harmonize scientific and legal views on research ethics by providing principles to guide the development of a new regulatory regime.

This final Part articulates several key principles underlying a new legal regime designed to enable responsible public genomics. These principles are drawn from traditional notions of research ethics, viewed from both the legal and scientific perspective and updated to meet the challenges presented by genomics. They also represent an effort to outline a broad conceptual framework and to encourage a wider public discourse. This Article does not purport to offer positive law solutions or to set forth a fully developed legal regime that responds to each of the questions; rather, it sketches a framework of principles that present a way forward for public genomics.

A. Respect for Individual Participants

The principle that stands at the heart of responsible public genomics is that no public genomics research project will ever succeed without evincing the utmost respect for its participants. In the case of competent adult[194] subjects, researchers must have respect for the safety and autonomy of subjects, and must diligently strive to protect and to give life to that autonomy.[195] A core principle of such respect is honesty. As noted by the Belmont Report, "[i]n most cases of research involving human subjects, respect for persons demands that subjects enter into the research voluntarily and with adequate information. In some situations, however, application of the principle is not obvious."[196]

Public genomics is just such a non-obvious situation. Our ignorance dwarfs our knowledge in the field of genomics. Thus, in the case of public genomics, "adequate information" requires that researchers comprehensively and regularly consider the risks associated with their work and pursue an honest and full disclosure of both known risks and the fact that unknown risks exist. If the structure of a study's protocol and research goals means that promises of privacy cannot be honored, then such promises must not be made, although privacy should still be protected to the extent feasible and appropriate given the particular project. Finally, individual respect and autonomy require that subjects be permitted to consent, within limits, to both known and unknown risks, provided such consent reflects the sub-

---

[194] This Article does not purport to address research involving children as subjects. *See* John A. Robertson, *The $1000 Genome: Ethical and Legal Issues in Whole Genome Sequencing of Individuals*, 3 AM. J. BIOETHICS 1 (2003), http://www.utexas.edu/law/faculty/jrobertson/1000genome.pdf.

[195] *See, e.g.*, BELMONT REPORT, *supra* note 54 (researchers must have respect for the autonomy of persons and protect persons with diminished autonomy).

[196] *Id.*

ject's considered judgment, in the face of all reasonably available information, and does not pose an unacceptable risk to others.[197]

In the case of public genomics research, respect for autonomy is evidenced through the development of robust informed consent procedures and through the diligent efforts of researchers and participants alike to ensure that the agreements that embody such informed procedures are living documents, constantly under scrutiny to ensure that they produce the most informed consent possible. To reach the ideals set forth in the Belmont Report, researchers must devote considerable time and energy to self-regulation, including the consent process, to the point that such ethical, legal, and social considerations command a meaningful part of the resources of any truly responsible public genomics research project.[198]

No matter how robust, however, self-regulation and self-reflection can only take public genomics so far. To truly enable responsible public genomics, legal and regulatory regimes must be modified so that the risks of public genomics to participants, as well as to researchers, are reduced even as they are fully disclosed.

B. Respect for Researchers

We start from the principle that the vast majority of public genomics researchers – both those that conduct the research and those that approve it – do so with the sincere desire to protect the subjects of their research, and that they act diligently and in good faith to do so. This principle reflects the practical recognition that, if the structures regulating public genomics research were to be crafted on any assumption other than good faith, the strictures would need to be so tight and pervasive that much beneficial research would never see the light of day.

If we assume the general good faith of researchers – including the IRBs and other institutional structures[199] that approve and oversee

---

[197] *See id.* ("To respect autonomy is to give weight to autonomous persons' considered opinions and choices while refraining from obstructing their actions unless they are clearly detrimental to others. To show lack of respect for an autonomous agent is to repudiate that person's considered judgments, to deny an individual the freedom to act on those considered judgments . . . .").

[198] The Human Genome Project recognized the importance of devoting resources to the consideration of the ethical, legal and social issues associated with genomic research. The ELSI Research Program, a legacy of the original Human Genome Project, continues to devote substantial resources to this topic today. *See* Nat'l Human Genome Research Inst., Nat'l Insts. of Health, ELSI Planning and Evaluation History, http://www.genome.gov/10001754 (last visited May 3, 2010).

[199] For instance, the independent DSMB in the case of the PGP. *See* PGP CONSENT FORM, *supra* note 90, at § VII(1.6).

their work – two of the best ways to balance competing principles are (i) to ask how researchers in practice have been balancing and proposing to balance those principles, and (ii) within broad limits, to permit researchers the flexibility to determine the best balance in the context of their own research. This does not mean that researchers should operate without formal limits but that those limits should be drawn to provide specific researchers, and public genomics as a field, with the flexibility they need to adapt participant protections in the manner that is best suited to the changing landscape of public genomics. The Helsinki Principles and the Common Rule each present a set of broad guidelines within which practitioners can apply their informed, good faith judgment in support of broad, agreed-upon principles rather than in compliance with undifferentiated restrictions.

C. Respect for Common Law

This article has identified the risks of lawsuits for negligence or malpractice that might be brought against public genomics researchers, primarily in the fifty state courts. In one sense, this possibility seems a nightmare scenario for researchers in new fields of knowledge, particularly one as intimately connected to human health as genomics. As recognized herein, a research institution and individual researchers could find themselves defending actions that, with the greater knowledge of hindsight (possibly derived from the very research being challenged), presented greater than anticipated risks to participants. As a practical matter, however, absent comprehensively preemptive federal regulation – which is unlikely to meet with a favorable response in the current political and legal climate – such risk is unavoidable. Not only that, it is also beneficial.

The common law of torts embodies a rich history of both reflecting and leading the development and articulation of the duties that members of society owe each other, and of adapting – albeit gradually – to changes in technology and moral sensibilities. A research community that takes its ethical obligations seriously, diligently pursues self-regulation, and ultimately formalizes and instantiates those ethical obligations in regulatory form will shape over time, for appropriate reasons and in appropriate ways, the common law's understanding of research duties in the context of public genomics. Being tested, sometimes to failure, by the application of evolving common law duties will provide a powerful corrective when individuals, or the research community as a whole, strays too far from its recognized ethical, legal, and social obligations. In other words, judicial oversight is both inevitable and necessary.

D. Ensure that Regulation is Focused and Informed

This article has reviewed the existing framework of federal and state statutes and regulations and found it lacking along several dimensions. This should come as no surprise. Positive law struggles whenever knowledge expands because yesterday offers less understanding of tomorrow, and laws that attempt to address future technologies – as in the case with genomic sequencing and interpretation – suffer even more acutely the burden of unintended consequences.

Because regulation is an inevitable response to technological advancement, however, the genomics community should work diligently and cooperatively to ensure that an emerging regulatory framework incorporates certain fundamental distinctions where applicable. For instance, to the extent feasible, the clinical practice of medicine should be clearly and unmistakably distinguished from basic scientific research. This distinction becomes even more important in a field such as genomics, where the raw data of research has unavoidable clinical significance. Researchers need to know which methods or models are acceptable – especially how they can and cannot interact with their participants – without being subject to regulation for practicing medicine, an activity in which they neither qualify nor intend to engage.

Especially within the realm of public genomics, formal regulation must also recognize challenges to traditional notions of participant privacy protection and permit – where consistent with underlying research goals and coupled with sufficiently informed and freely given consent – the public disclosure and use of personal genotypic, phenotypic, and other information. Thorough and responsible self-regulation on the part of the public genomics community in advance of the inevitable formal regulation will provide a blueprint for future lawmakers and regulators and ensure that such regulation is both cognizant of and informed by existing public genomics research.

E. Establish a Forum for Consolidating Information and Guidance on Best Practices

The federal regulatory scheme should include an expanded regulatory forum or oversight body that consolidates information and expertise on public genomics best practices. This Article takes no position regarding where such consolidation should occur, and does not consider consolidation a replacement for the core missions of existing federal agencies, such as the mission of the FDA to ensure the safety and efficacy of pharmaceuticals, biologics, and medical devices. The writ of such organization should include all public genomics

research conducted in the United States (with attention paid to international efforts), and it should have as its objective providing both information and guidelines on the development of responsible public genomics.

This Article does not propose that such agency promulgate a comprehensive set of research regulations, which would risk producing the sort of rigid "one size fits all" approach that characterizes existing laws and regulations. Rather, this agency would provide information and guidance to: (i) permit research institutions, researchers and other stakeholders, acting in good faith, to collaborate in the design and implementation of responsible, effective research protocols; (ii) allow research participants access to data regarding the risks and benefits of research projects; and (iii) provide courts with a standard to follow when formulating common duties and standards of care. Although one purpose of such an organization would be to try to reduce the incentive of states to enact overlapping regulatory schemes, to the extent states do put such schemes in place, the existence of a comprehensive source of information and guidance may at least inform such schemes for the better and encourage consistency in regulatory aims and implementation.

F. Ongoing Review and Re-Appraisal

Finally, any public genomics regulatory framework should provide the means for all of the applicable stakeholders in this field – researchers and research participants, educators, policymakers and regulators, public and private clinical care providers, and commercial genomic service providers – with the opportunity to measure, publicize, and assess on a continuing basis the effectiveness of existing ethical, legal, and social practices in the field of public genomics research.

The dialogue should permit public as well as industry and regulatory voices, and the institutional structures should be biased toward a flexible, guidelines-based regime (rather than a more rigid rules-based approach) so that researchers can continue to achieve the best balance between conducting desirable, responsible research and protecting the well-being and autonomy of research participants. Many of these goals may be achieved by expanding upon the existing government genomics advisory and oversight framework including, for example, the Secretary's Advisory Committee for Genetics, Health, and Society (SACGHS).[200]

---

[200] The SACGHS "advises the Secretary of Health and Human Services on the broad range of human health and societal issues raised by the development and use and potential misuse of genetic technologies" and is already relatively interdiscip-

CONCLUSION

The public genomics era is arriving rapidly and carries with it great promise for the progress of scientific knowledge and human health. Nevertheless, the era is developing gradually enough that policymakers, regulators, and researchers have no excuse for failing to develop a structural framework adequate to the demands of public genomics and designed to promote its continued development. All who would participate in or benefit from a future of public genomics research have a shared responsibility to contribute to that process, for to do nothing would jeopardize the future of responsible public genomics.

---

linary in nature. Office of Sci. Policy, Nat'l Insts. of Health, Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS), http://oba.od.nih.gov/sacghs/sacghs_home.html (last visited May 3, 2010); Office of Sci. Policy, Nat'l Insts. of Health, SACGHS Member Roster, http://oba.od.nih.gov/SACGHS/sacghs_members.html (last visited May 3, 2010). However, SACGHS is limited to serving the needs of HHS, and even in that capacity its advisory authority is limited. Expanding its focus and/or authority, perhaps through partnerships with other agencies and public genomics stakeholders, would be highly desirable.