**TITLE:** Claiming the public right of access to standardized genetic information

Stefanie Pina Escudero1, Laís Fajersztajn1, Clio Korn1, Maira Okada de Oliveira1, Stephanie Wankowicz1

*1Science Policy Group, University of California, San Francisco*

**Executive summary**

Currently, an individual’s access to her own genetic data is inconsistent. This is a violation of her rights and a detriment to the advancement of medicine for all. We propose policies that require clinical laboratories and companies to make genetic testing results available in a standardized, functional format both to individuals and to existing publicly accessible research databases.

**Current accessibility of genetic data**

Variations and mutations in the genetic code play a role in susceptibility to numerous diseases and conditions. Today, hundreds of thousands of U.S. citizens undergo clinical genetic testing for over 2000 conditions at over 500 clinical laboratories and companies (“NIH Fact Sheets - Genetic Testing: How It Is Used for Healthcare”). Genetic testing allows individuals to determine the cause of a disease and to make informed decisions about preventative measures or treatment plans that they may want to consider to mitigate their disease risk. In some cases, genetic testing can also give an individual a disease diagnosis or inform therapeutic decisions. For example, over 150,000 people per year receive genetic cancer risk results, which inform the treatment decisions they and their doctors make (Sándor 2018). Such test results may allow not only the person initially tested but also their family members to monitor disease risk and may thus have important implications for numerous individuals (Sándor 2018, Evans 2018). Importantly, the interpretation of genetic results, and therefore their use, is constantly evolving as research advances (Haeusermann et al. 2017, McGowan 2018).

Access to personal health information is an individual right protected by the Health Insurance Portability and Accountability Act (HIPAA) of 1996 in the United States and is regarded as an innate human right around the world (Hilary 2008). In light of the growing prevalence of genetic testing and its importance in health care, HIPAA was recently amended to explicitly state that patients have the right to their clinically derived genetic information (Office for Civil Rights 2015). Specifically, Health and Human Services stated that HIPAA-covered clinical laboratories and companies must, upon request, provide individuals whom they test with full gene variant information generated by such tests (Office for Civil Rights 2015).

However, this amendment did not include any specific information as to the type or standards of genetic testing results, often making the information provided by the clinical laboratory or company unusable or ineffective. This in turn makes it difficult for individuals to effectively use their testing results, either to make personal medical decisions or to help advance medical research by participating in studies. Indeed, under the current regulatory framework, many patients report difficulties in accessing their genetic data, and even when they do receive their test results, they are often in a format that cannot be read or used by either clinicians or researchers (Office for Civil Rights 2015, Kincaid 2019, Haeusermann et al. 2017, Wankowicz 2018).

Genetic disease risk is a numbers game. The more data on an individual genetic variant, the more confidence we can have about it potentially increasing risk for a disease. Access to genetic testing results in a standard, readable format therefore benefits not only the individual patient but medical research as a whole. Indeed, many patients are eager to make their anonymized genetic test data available for research (Goodman 2017). When clinical laboratories or genetic testing companies do not readily provide individuals with their genetic testing results, or do so in an unusable format, genetic research suffers. A prime example is Myriad Genetics, which for over 22 years has been sequencing the BRCA1 and BRCA2 genes for over 1 million individuals (Myriad Genetics Company Milestones). Despite the immense research potential of Myriad’s database on these genes, the company has not made this information available either to patients and physicians trying to make treatment decisions, to researchers, or to the greater public (Guerrini 2017). If this data were available to the greater research community and medical teams making treatment decisions, it would have the potential to change the way we look at risk for multiple types of cancer.

**Policy Recommendations**

*Standardization of data storage format*

Standardized formatting requirements for clinical genetic testing results would support patients’ access to their data. In particular, standard formatting would allow patients to more easily understand their data by facilitating interpretation and discussion of these data with a clinician or genetic counselor.

We propose mandating the use of the Variant Call Format (VCF), a standard text file format for storing variations in genetic sequences used across the world by large research and clinical databases (Tan 2015, 1000 Genomes Project Consortium, Danecek 2011). This format only requires storage of genome variants, rather than the entire genetic sequence, thus minimizing the storage capacity needed for such data. Additionally, VCF is a flexible file format specification that can represent many different variant types ranging from single nucleotide polymorphisms to indels to copy number variations (Tan 2015). As the majority of genomic tools and existing research databases use this format, it would allow easy transfer of data from clinical labs and companies to research studies (ClinVar - ClinGen 2018).

Additionally, we propose requiring individual variants to be annotated using the Human Genome Variation Society (HGVS) Nomenclature. HGVS nomenclature is an internationally accepted standard for descriptions of variations in genetic sequences that is easily readable by the human eye (den Dunnen 2016). HGVS names for genetic variants can be readily incorporated into VCF files, allowing integration of these two formatting standards.

To make the genetic data they provide fully usable, companies should additionally be required to release non-proprietary information on their variant calling pipeline, such as the human reference genome they are using. Several human reference genomes are currently in use, and this information has large impacts on the interpretation of genetic variants (Kaplun 2016, Guerrini 2017).

*Required deposit of genetic data into public databases*

In order to maximize the research potential of clinical genetic data, clinical laboratories and genetic testing companies should be required to deposit de-identified genetic variants into existing clinical interpretation databases.The more individual genetic information is available, the more informative the data becomes, and companies maintaining their own proprietary databases is therefore detrimental to patients (Cline 2018). Many companies already make their data publically available. However, without a legislative requirement that they do so, not all participate in this open data sharing, as noted above in the case of Myriad Genetics (McGowan 2018).

We recommend that genetic data be deposited in the ClinVar database, which is run by the National Center for Biotechnology Information at the National Institutes of Health (ClinGen 2018). This is the database used by companies that already make their data publically available. ClinVar uses HGVS Nomenclature, so required deposit of genetic testing results into this database would readily integrate with the standard formatting described above.

*Protecting privacy*

Companies that have resisted sharing the genetic data that they gather cite privacy concerns as justification (McGowan 2018). Although privacy of genetic data is paramount, there are ways to share clinical genetic data while still protecting patient privacy. The benefit of openly sharing clinical genetic data for individual patients, medical research, and society as a whole is too significant to justify allowing such data to remain siloed in private corporate databases. Companies can readily offer individuals an opt-out from having their genetic testing results added to public databases, so that individual consent for use of the data in research is assured. Additionally, data will be de-identified before being added to such databases, thus protecting patient privacy. Indeed, existing public databases such as ClinVar already have de-identification procedures in place, and numerous guidelines exist that can be used to further optimize the management of genetic data and assure individual privacy (Zook 2017, Global Alliance for Genomics & Human Health 2018).

**Impact**

Genetic testing interpretation is critical to being able to use genetics to understand risk and to treat diseases. Individuals who get genetic testing for any condition should be the owners of their genetic information. A policy that standardizes genetic results and requires them to be reported in an internationally accepted nomenclature is imperative. The policy changes we propose will allow individuals to share their information in a usable format with both health care providers and scientists. Information presented in this format will enable clinicians to provide accurate genetic consultation to individuals. Our proposed policy would prevent private genetic testing companies from maintaining proprietary ownership of the information of their clients and using if for their own commercial benefit. If consent is given, all the genetic information gathered by clinical laboratories and companies should be placed in publicly available databases where it can be requested by researchers. Wide sharing of genetic information leads to better risk identification, prevention, and treatment strategies for individuals with genetically determined diseases. Our policy addresses two key needs in clinical genetics: individual access to standardized genetic data as a health right for patients, and research access to anonymized genetic data to move towards medical advances for thousands of diseases.

**References**

Evans, Barbara J. 2018. “HIPAA’s Individual Right of Access to Genomic Data:

Reconciling Safety and Civil Rights.” American Journal of Human Genetics 102 (1): 5–10.

Haeusermann, Tobias, Bastian Greshake, Alessandro Blasimme, Darja Irdam, Martin Richards, and Effy Vayena. 2017. “Open Sharing of Genomic Data: Who Does It and Why?” PloS One 12 (5): e0177158.

Global Alliance for Genomics & Human Health: https://www.ga4gh.org/

Hilary, Charlesworth. 2008. “Universal Declaration of Human Rights (1948).” In Max Planck Encyclopedia of Public International Law.

Kincaid, Ellie. 2019. “Data Pirates: Patients And Scientist Battle To Liberate Genetic Testing Results.” Forbes. Forbes. January 17, 2019. https://www.forbes.com/sites/elliekincaid/2019/01/17/data-pirates-patients-and-scientist-battle-to-liberate-genetic-testing-results/.

McGowan, Kat. 2018. “One of America’s Biggest Genetic Testing Companies Refuses to Publicly Share Data That Could Save Countless Lives.” Mother Jones. June 6, 2018. https://www.motherjones.com/politics/2018/06/one-of-americas-biggest-genetic-testing-companies-refuses-to-publicly-share-data-that-could-save-countless-lives/.

“NIH Fact Sheets - Genetic Testing: How It Is Used for Healthcare.” n.d. Accessed February 24, 2019. https://report.nih.gov/nihfactsheets/ViewFactSheet.aspx?csid=43.

Office for Civil Rights (OCR). 2015. “Privacy.” HHS.gov. US Department of Health and Human Services. April 16, 2015. https://www.hhs.gov/hipaa/for-professionals/privacy/index.html.

“Our Genes, Our Data: Patients’ Right to Access Their Own Genetic Information.” n.d. American Civil Liberties Union. Accessed February 25, 2019. https://www.aclu.org/cases/our-genes-our-data-patients-right-access-their-own-genetic-information.

Sándor, Judit. 2018. “Genetic Testing between Private and Public Interests: Some Legal and Ethical Reflections.” Frontiers in Public Health 6 (January): 8.

Zook, Matthew, Solon Borocas, danah boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara A. Koenig, Jacob Metcalf, Arvind Narayanan, Alondra Nelson, Frak Pasquale. 2017. “Ten Simple Rules for Responsible Big Data Research.” PLOS Computational Biology 13(3): e1005399.

Tan, Adrian, Gonçalo R. Abecasis, and Hyun Min Kang. 2015. “Unified Representation of Genetic Variants.” Bioinformatics 31 (13): 2202–4.

Wankowicz, Stephanie, Brooke Maile, Ken Deutsch, Adam Hayden, Erika Brown, Tom Marsilje, Beth Caldwell, Tania Simoncelli, and Eliezer Van Allen. 2018. “Patient-Driven Efforts to Liberate Clinical Cancer Genomic Data.” <https://doi.org/10.31219/osf.io/gupvq>.