Towards Responsible Genomic Surveillance:
A Review of Biosecurity and Dual-use Regulation

Kenza Samlali
Concordia University, Montreal, QC, Canada
Bricobio Community Bio Lab, Montreal, QC, Canada
Home country: Belgium

Julie Stern
George Mason University, Arlington, VA, USA
Georgia Institute of Technology, Atlanta, GA, USA
Home country: USA

Elicana Nduhuura
Faculty of Medicine, Mbarara University of Science and Technology, Mbarara, Uganda
Home country: Uganda

2021 Next Generation for Biosecurity Competition
August 3, 2021, Revised October 27, 2021, November 8, 2021

Abstract

Genomic surveillance has become a widely used tool to monitor disease dynamics and track the emergence of new variants with an accurate and timely response. This is supported and encouraged by international health regulations and the effect of the COVID-19 pandemic on national health security. In this report, we analyze how genomic surveillance pipelines used for disease monitoring and genomic population research studies pose risks for dual-use malfeasance, potentially resulting in the exploitation of marginalized groups of people. By exploring responsible genomic surveillance along with frameworks and norms currently in place, we identified potential risks in the genomic surveillance research pipelines themselves, in the use of similar techniques for population research beyond disease monitoring, and in the current health security frameworks that present genomic surveillance as a biorisk management technique. We conclude that samples collected as part of a genomic surveillance pipeline for disease surveillance should not be used for research if it involves more than just the pathogenic genetic information. When genomic data are sampled for disease monitoring, for COVID-19 diagnostic testing for example, only the viral portion of the genomic information collected is needed, while the human genetic information is not. That means, no human genomic information[1] should be analyzed through these pipelines. Human genetic data collected incidentally in the course of public health research and surveillance should be subject to stringent safeguards to ensure that these data are not exploited and, with the consent of the donor, stored and used safely, securely, and responsibly.

[1] Human genomic information may include Y-chromosome, whole genome, mitochondrial sequence information, other human Omic information (transcriptomic, proteomic, metabolomic).

Introduction

Genomic data hold the key to our identities; characteristics at the individual, family, and population levels but also valuable information about diseases. The rapid pace of genomic-related technology development as well as the burgeoning growth of data-driven genomic surveillance pipelines exposes new potential biosecurity and public health risks and vulnerabilities. Technological advancement is outpacing our ability to develop and implement an updated regulatory infrastructure fast enough. The intention of this paper is to describe these risks and regulatory gaps, as well as express the urgency of strengthening safety and security protocols for human genomic data both nationally and internationally.

The main argument of this paper is that genomic surveillance for the purpose of disease monitoring, especially in times of crisis such as a pandemic, poses the risk of capturing human genomic information, which goes beyond the purpose of disease monitoring. Greater access to human genomic data creates new risks, and vulnerabilities. Pandemic disease monitoring creates the opportunity for greater access to data. For instances where human genomic data are desired for population studies, which may have medical research benefits, regulation is needed to ensure that such data are stored and used safely, securely, and responsibly. Furthermore, the inability to truly anonymize human genome data poses further risks to privacy (Shabani and Marelli, 2019; Yousefi et al., 2018; Zaaijer et al., 2017).

Real world historical examples give an idea of the existing risk landscape, a landscape that will only grow as new vulnerabilities due to genomic surveillance are added to it. Governments collecting deoxyribonucleic acid (DNA) samples directly from populations are posing various health security risks around the world, including human rights violations, civil liberties violations in terms of invasion of privacy, and national security risks with respect to the export of information to other countries. Biospecimens can originate from collection of DNA by law enforcement and forensics leading to another category of health security vulnerabilities. DNA collection for medical research opens up another set of vulnerabilities. Last, there are aspects of the current COVID-19 pandemic that are paving the way for an infrastructure of malfeasance if it were not addressed.

In this report, we first describe some historical examples that shine a light on the possible misuse of genetic data. Next, we provide a detailed background describing genomic disease surveillance frameworks, governance of research and use of genomic information, and the current landscape of genomic disease surveillance. Finally, we perform a risk analysis and identify technological and data security risks and harms, and the lack of governance in the domain of genomic surveillance. This is followed by some brief policy recommendations and a conclusion.

Vulnerabilities include harms to human rights. Privacy violations can also occur affecting civil liberties. Discrimination is inevitably a result of these harms and violations. Genomic information becomes part of the tools empowering government and law enforcement agencies. Health security vulnerabilities lead to public mistrust in genomic methods.

Human Rights and Civil Liberties Violations

China's human rights record has long been the subject of international criticism. Among their alleged abuses, China has been collecting DNA samples from ethnic minority populations and individuals detained in China, which imposes an immediate and severe health security risk. Minority populations targeted for DNA collection by China include Christians, the Falun Gong, Muslims, Tibetans, and Uighurs. A recent report from the UN (Martin, 2019; UNHROC, 2021) alleges that these DNA samples are being used in conjunction with an "organ harvesting on demand" program.

China uses genomic information for longer term studies that investigate ways to exploit a population's characteristics. China has been harvesting DNA from prenatal women using leftover blood samples and genetic data from the Non-Invasive Fetal Trisomy (NIFTY) test operated by the Beijing Institute of Genomics (BGI). For these data, prevalence of viruses in Chinese women as well as mental health indicators were investigated in an exploratory fashion to data mine and discover useful patterns that can exploit a population. China has been using military supercomputers along with artificial intelligence algorithms for these studies (Needham and Baldwin, 2021).

The misuse of population genetics data is also reported in life sciences research. One historical example involves Native Americans. In 2004, the Navajo Nation sued Arizona State University for taking DNA samples specifically meant for studies of diabetes and using those samples in studies of "schizophrenia, migration and inbreeding without their consent" (Reardon, 2017). The Navajo Nation took control of their own data by banning DNA analysis on tribal lands for 15 years.

Kuwait sought to collect DNA samples from its entire population. In 2015, the government of Kuwait passed a law requiring all Kuwaiti citizens, foreign residents, and visitors to provide DNA samples. In 2017, a Kuwait court overturned this law, protecting the right to privacy (Human Rights Watch, 2017).

Forensics and Law Enforcement Risks

Another vulnerability includes using genomic data to inform government and law enforcement. China has also been obtaining the DNA of the male population including children for future use in law enforcement databases, where the connections between "multigenerational family trees created by the police" make it easier to link an unknown male back to a specific family or even individual (Dirks and Leibold, 2020). The U.S. Department of Homeland Security (DHS) was also collecting DNA in a pilot program at the U.S.-Mexico and U.S.-Canadian borders from detainees and only those entering illegally which may also have included children and asylum seekers, for law enforcement purposes (Merchant, 2020; Moreau, 2019). Furthermore, Hong Kong authorities collected DNA samples from protestors in the aftermath of a new security law imposed by Beijing (Wee and May, 2020). But the government of Hong Kong said the samples would not be transported to the mainland. Additionally, scientific research on population genetics related to ethnicity was co-authored by law enforcement personnel. Recently, more than 50 papers have been flagged in high impact journals such as *Human Genetics*. One of these papers described the analysis of 38,000 Male Chinese Y chromosome samples, many of which were of Uygur descent. This triggered the resignation of eight out of 25 editorial board members from *Molecular Genetics & Genomic Medicine* (Normile, 2021).

Health Security Vulnerabilities

A theme emerges in several of these examples, that of collecting DNA without the knowledge of the person, or with any transparency about where the data could end up being stored and what could be done with them. This is referred to as a lack of informed consent, which in turn can lead to a large-scale mistrust in the healthcare system.

The BGI NIFTY studies collected data for eight million women globally. Women outside of China had no idea their DNA data could end up stored in China. Furthermore, Native Americans did not consent to their data being used for medical studies with outcomes that were not only publicly disseminated, but likely to lead to discrimination. Although governments such as Kuwait may have passed a law that everyone's DNA be collected, they didn't provide information about what they would do with that DNA. Furthermore, sharing DNA due to law enforcement or national security would override privacy expectations, and provide legal gray areas.

The concern over the dual-use of genomic surveillance during disease outbreaks can be justified through historical examples of discrimination against marginalized[2] groups during epidemics (Sekalala et al., 2021). To our knowledge, no clear precedents exist regarding the misuse of genomic surveillance pipelines for gathering human genomic information. During the pandemic, Beijing sent a COVID testing team to Hong Kong without being asked, to help increase the number of COVID tests performed per day. However, due to the DNA collection of protestors, (Moreau, 2019; Wee and May, 2020), past mainland Communist government intrusions, the lack of transparency, China collecting DNA for minority populations, the people of Hong Kong did not trust that the COVID tests weren't collecting full population DNA.

Vulnerabilities due to the COVID-19 pandemic may also involve COVID-19 test kits and sampling as the entry point of data exploitation. Keeping track of which companies provide the kits and where the data are going, may help ascertain these risks and vulnerabilities. For example, ThermoFisher sampling kits are still being sold to Chinese police forces (Wee and May, 2020). BGI, the company that was collecting women's DNA data for the NIFTY studies, is also providing test kits for COVID-19. Because BGI's privacy policy and user agreements made it possible for them to perform the women's NIFTY DNA studies, worries exist that potentially these COVID-19 test kits could end up engaged in a possible dual-use scenario. In addition, if the data were deemed to be a matter of national security or defense, BGI would be allowed to legally share this information.

Although we have yet to observe proof of the misuse of genomic surveillance pipelines for political purposes, with the Hong Kong BGI example coming close, the above determined vulnerabilities, added to the gaps in the regulatory landscape, indicate the need for developing urgent mitigation strategies. To mitigate the risks of the dual-use of genomic surveillance strategies for disease monitoring, we need to first focus on recognizing these risks in current frameworks that guide genomic surveillance pipelines. These risks include using private genomic information without consent, using genomic information to target minorities at risk, losing beneficial disease genomic surveillance, and losing trust in the healthcare system on a large-scale. Next, we can mitigate these risks by improving the transparency in operation of national genomic surveillance programs. This is elaborated on in the Risk Analysis section below.

Background

This section provides background information describing the existing state-of-the art in the genomic surveillance technology, the current genomic surveillance frameworks, and genomic, data, privacy, and security regulation. This background helps to understand the technological aspects that make these new risks and vulnerabilities possible as well as our baseline starting points in frameworks and regulation pertaining to this type of and related genomic data.

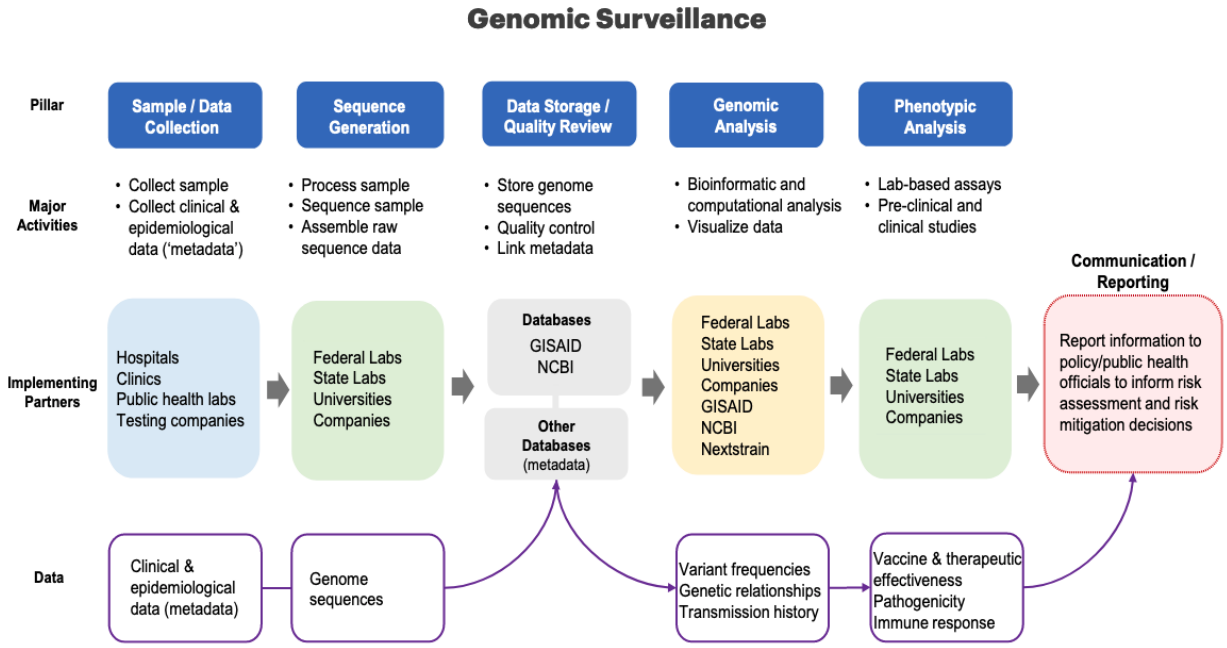
What Is Genomic Surveillance?

Genomic surveillance is a disease monitoring method that integrates clinical, epidemiological, genomic, and phenotypic data to track changes in the transmission and virulence of a pathogen and the effectiveness of medical countermeasures nationally or globally (Rockefeller Foundation, 2021, Sanger Institute, 2021). The pipeline of a genomic surveillance program consists of sampling, sequence generation, data storage and quality review, genomic data analysis, and phenotypic analysis of pathogens (Figure 1). Multiple life sciences research methods are used in this pipeline. Sampling requires appropriate tools and experience,

[2] Marginalized groups may include minorities, indigenous populations, people with medical conditions, or any other at-risk population.

and sample preparation requires the extraction of all the DNA or ribonucleic acid (RNA) from the sample, after which often a specific region is amplified or the whole genome of the pathogen is fragmented and barcoded. The infrastructure of this pipeline includes trained medical personnel for sampling, trained research personnel for sample preparation and data analysis, trained technical personnel and field services for sequencing equipment, implementing partners such as hospitals, test centers, research institutes, government labs, sequencing centers, data centers and computing servers, and all the equipment. The sequencing is often outsourced to specialized centers with sequencing machines. After sequencing, the resulting data are analyzed and combined with phenotypic information, resulting in conclusions about strain variants, infectivity, and other epidemiological measures. Genomic surveillance differs from disease surveillance programs by adding a genome sequencing component to the picture. During the COVID-19 pandemic, genomic surveillance became an established method for disease surveillance. Although the cost of sequencing was a bottleneck in the past, genomic sequencing of the pathogen is now more common thanks to the rapid development of sequencing technologies[3] with increasingly lower price points over the last 20 years.[4] Sharing this sequence information data with international partners through databases such as the Global Initiative on Sharing Avian Influenza Data (GISAID) (GISAID, 2021, Bogner et al., 2006) or the National Center for Biotechnology Information (NCBI) (NCBI, 2021) helps the data become globally accessible and is a critical part of genomic surveillance.

Figure 1. Genomic Surveillance Pipeline



The genomic surveillance pipeline consists of five major activities, executed by different implementing partners. Source: Rockefeller Foundation, 2021.

[3] Today, you can buy a sequencer the size of a USB drive (Oxford Nanopore MinION ~ US$1,500 (Nanopore Tech, 2021)), or you can get your whole genome sequenced commercially for around US$99 (Nebula Genomics, 2021).

[4] On February 15, 2001, 20 years ago, the first draft of a full human genome was published (Venter et al., 2001). The international and multidisciplinary Human Genome Project cost millions. Sequencing surpassed Moore's law in 2007. Now, the cost of high coverage whole genome sequencing dropped below US$1000.

Genomic Surveillance Frameworks

Genomic surveillance in a global healthcare context is strongly promoted through international frameworks to increase our preparedness against possible disease outbreaks.[5] Guidelines on surveillance methods for specific disease outbreaks have typically been issued by the WHO, such as for the SARS outbreak in 2004 (WHO, 2004). The International Health Regulations (IHR) (2005) (WHO/CDS/EPR/IHR/2007.1) outline a country's general duties in case of disease outbreaks. The IHR is agreed to by 196 countries, all of which are required to have the ability to detect, assess, report, and respond to public health events. Genomics did not have a central role in past disease surveillance methods.[6] In response to the complexity of the H5N1 avian flu outbreak in 2006, the WHO Pandemic Influenza Preparedness (PIP) Framework was unanimously adopted on May 24, 2011 (WHO, 2021a). This was one of the first frameworks to support responsible surveillance, with a focus on sharing pathogen genomic sequence information, which has now become the basis of the response against the SARS-CoV-2 outbreak.

Beyond the IHR and PIP, several additional non-binding guidelines are available that can guide responsible genomic surveillance. Regarding human genomic information resulting from health research, the most inclusive responsible frameworks and policy guidelines can be sourced from the Global Alliance for Genomics and Health (GA4GH, n.d.), an organization that aims to set global standards for responsible genomic data sharing within a human rights framework (Knoppers, 2014). GA4GH shares several toolkits to facilitate the implementation of consent clauses in genomic research (Regulatory & Ethics Toolkit), responsible genomic data analysis and sharing (Genomic Data Toolkit), and genomic data security (Data Security Toolkit). In response to the COVID-19 genomic surveillance program, they released the *Responsible Data Sharing to Respond to the COVID-19 Pandemic* document outlining responsible practices on public health ethics and data sharing (Beauvais, 2021).

The IHR and the PIP are primary frameworks for disease surveillance and management and are legally binding. IHR compliance is encouraged through joint statements with other international organizations. For example, at various points during the H1N1 pandemic, the Food and Agricultural Organization (FAO), the World Organisation for Animal Health (OIE), and the WHO issued joint statements (FAO/OIE/WHO, 2009). And when one country's active violation of the IHR causes specific damage to either the population or the economy of a second country, the International Court of Justice (ICJ) could be used (Gostin and Katz, 2016). Furthermore, to implement the IHR, countries can volunteer to have a Joint External Evaluation (WHO JEE) performed to assess their emergency preparedness, which includes an evaluation of their surveillance capacities. The Global Health Security Index of 2019[7] (GHS Index, 2019) is an additional assessment performed for all 195 IHR signatories. Its goal was to create transparency, expose gaps in health security including disease surveillance capacities under which genomic surveillance resides.

Genomic Data Privacy and Security Norms and Regulations

Genomic surveillance for purposes of disease monitoring poses a risk of intentional or unintentional exposure of human genomic information. Therefore, it is useful to perform an analysis of the current norms and regulations in place that protect individuals' genomic information. A great resource summarizing all

[5] Disease surveillance is part of a methodology to maintain responsible science for global healthcare, within the WHO biorisk management framework (World Health Organization, 2010).

[6] Past diseases such as SARS (2004), infections from antibiotic resistant strains (e.g., *Staphylococcus aureus*) or Zika did not use genomic surveillance.

[7] The GHS Index was a first in that it created an assessment to benchmark health security using open-source information for 195 IHR signatory countries for how they meet various indicators. The overarching categories are prevention, detection and reporting, rapid response, health system, compliance with international norms, and risk environment (GHS Index, 2019).

the current (inter)national regulations and norms in place concerning the use of genomic data, specifically related to consent, privacy, data access, and benefit-sharing is the Genomic Data Policy Resource List compiled by the World Economic Forum (WEF, 2019). The norms and regulations in place relating to human genomic information can be divided into four categories: genomic data rules that apply to researchers and research institutes, genomic data rules that apply to private entities like employers or health insurance, civil and human rights concerning genomic data, and genomic surveillance specific guidelines and regulation (see Table 1).

For research institutes, there are rules like the Common Rule (1991) in the United States, which governs the ethical conduct of research and provides a framework for consent. The Common Rule was revised in 2018 to include new rules[8] on human subject research, biospecimen handling, health data, and consent (National Human Genome Research Institute, 2017). It now allows "broad consent"[9] on the use of secondary data (which previously needed informed consent), facilitating for example secondary genomic data analysis[10] or the sharing of genomic data without having to inform the patient each time (Redhead, 2017). There are attempts to balance individual privacy with the power that population-level health data can provide by adding for example "identifiable information and biospecimens" under the definition of "human subject" within the Common Rule. Beyond the Common Rule, Institutional Review Boards (IRB) can place additional restrictions on research involving genomic data.

Further protection of an individual's genomic data is provided by the protection of genomic information as personal health information.[11] In the United States, the Health Insurance Portability and Accountability Act (HIPAA) (U.S. 104th Congress. 1996), requires covered entities (health plans and healthcare providers) to de-identify personal health information, which includes genomic information, according to their guidelines (Office of Civil Rights, 2020). The Genetic Information Non-discrimination Act (GINA) (EEOC, 2008) protects individuals from discrimination by health insurance providers and employers based on genetic information or asking for genetic information. In Canada, the equivalent is the Genetic Non-Discrimination Act (GNDA) (Consolidated Federal Laws of Canada, 2017). In Europe, individual genetic information is largely protected under the General Data Protection Regulations (GDPR) ((EU) 2016/679) (Radley-Gardner, Beale, and Zimmermann, 2016). Under Art. 9§1, processing of genetic data is prohibited if it can identify a natural person. The H3Africa Consortium outlines genomic data protection rules for the African continent (Ramsay, 2015).

It is worth noting the United Nations (UN) Convention Biological Diversity (CBD) Nagoya Protocol (NP)[12] provides a framework to aid countries in maintaining oversight over local genetic resources, including digital sequence information (DSI).[13] The NP offers interesting arguments on the ethical considerations around DSI, such as the need for financial models of benefit sharing in case the DSI is used for profit (e.g., pharmaceutical use of indigenous plants), and questions around the monetary value of DSI (CBD/DSI/AHTEG/2018/1/3) (Laird and Wynberg, 2018, GISAID, 2019). Although currently the NP only protects biodiversity genetic resources, conversations are ongoing about including human genetic

[8] Common Rule, §46.116(d) ("broad consent for the storage, maintenance and secondary research use of identifiable private information or identifiable biospecimens (collected for either research studies other than the proposed research or non-research purposes) is permitted as an alternative to the informed consent requirements.").

[9] Permission provided for future research.

[10] Future analysis or secondary analysis that goes beyond the initial purpose for generating the sequencing data.

[11] Genomic data are only protected if it is considered personal information or health data. So-called unidentified genomic information and aggregated data, stripped from identifiers such as name, age, health history, etc. fall out of their scope.

[12] Nagoya Protocol (NP) on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity (CBD) (NP).

[13] "Digital sequence information" as used by the Nagoya Protocol currently has no agreed definition. Differences in terminology in scientific circles reflect differences in the material referred to, which makes it difficult to harmonize terminology. In this paper, DSI is used to refer to digital information resulting from sequencing (DNA or RNA) of biospecimens.

information. Whether giving countries ownership over population genomics is a good method to improve privacy and regulate data sharing, is still under consideration.

Lastly, using genetic-data-derived information is also protected under the civil rights and human rights acts, based on whether someone is discriminated against due to their genetics. For example, in the United States, the Office of Civil Rights (OCR) has additional laws against discrimination based on race, color, national origin, disability, age, sex, and religion by certain healthcare and human services. International human rights declarations concerning human genomics include the Convention on Human Rights and Biomedicine (Council of Europe, 1997), and the Universal Declaration on the Human Genome and Human Rights (UNESCO, 1997). These acts set out the basic principles of human rights protection, bearing on research in genetics and biology and the application of its results. The Universal Declaration has been agreed upon to stimulate Member States to establish national legislation to enforce outlined principles.

Table 1. Select List of Current Genomic Surveillance Norms and Regulations (sampling, sequence information, pipeline)

Name	Publication Date	Latest update	Entity	Jurisdiction	Type of regulation	Topic
Genetic Information Nondiscrimination Act (GINA)	2008	/	U.S. Equal Employment Opportunity Commission (EEOC)	United States	Federal law (Act)	Protections against access to genetic information and genetic discrimination in both the health insurance and employment settings
Health Insurance Portability and Accountability Act (HIPAA)	1996	2013	U.S. Department of Health & Human Services	United States	Federal law (Act)	Protecting sensitive patient health information from being disclosed without the patient's consent or knowledge
Common Rule	1991	2018	Code of Federal Regulations	United States	Legislation	Rules and regulations surrounding research conducted on human subjects
Framework for Responsible Sharing of Genomic and Health-Related Data	2014	/	Global Alliance for Genomics and Health (GA4GH)	International	Framework	Principled and practical framework for the responsible health-related data
GA4GH Consent Policy	2015	/	Global Alliance for Genomics and Health (GA4GH)	International	Policy Guideline	Patient consent for genomic research studies
Model Framework for Governance of Genomic Research and Biobanking in Africa	2018	/	Human Heredity and Health in Africa (H3Africa)	African continent	Framework	Governance elements for genomics research studies or biobanking sharing of genomic
International Ethical Guidelines for Health-Related Research Involving Humans	2016	/	World Health Organization (WHO)	International	Guideline	Informed consent. Guideline 9, 10, 16, and 17
General Data Protection Regulations (GDPR)	2018	/	European Parliament and Council of the European Union	Europe	European Private Law	Prohibited use of identifiable genetic data
Genetic Non-Discrimination Act (GNDA)	2017	/	Canadian Government	Canada	Consolidated act in Canadian Constitution (Bill)	Act to prohibit and prevent genetic discrimination
Universal Declaration on the Human Genome and Human Rights	1997	2003	United Nations UNESCO	International	Declaration	Rights of the persons concerned by human genome research; legal framework for stimulating ethical debate and harmonizing of the law worldwide
Statement on Benefit-sharing	2000	/	HUGO Committee on Ethics, Law and Society (CELS)	International	Statement	Recommendations on benefit sharing in human genome research
Nagoya Protocol	2010	/	United Nations	International	International Treaty	Guidelines on using genetic resources
International Health Regulation (IHR)	2005	/	World Health Organization (WHO)	International	International Treaty	Health security regulations, including disease surveillance framework
Export Administration Regulations (EAR)	1997	continuous	U.S. Department of Commerce's Bureau of Industry and Security (BIS)	United States	Regulation	Export controls and sanctions
Pandemic Influenza Preparedness (PIP) Framework	2011	/	WHO	International	Framework	Disease surveillance framework suggesting genomic surveillance methods with a focus on influenza

Current Evolving Landscape of Genomic Surveillance

Whole-genomic sequencing of pathogens in a genomic surveillance context became widely adopted with the COVID-19 pandemic (Brady, 2020, Park, 2021). Genomic surveillance programs have expanded to multiple nations. In the African continent, multiple efforts are ongoing to build capacity for genomic surveillance (see Table 2) (National Institutes of Health, 2021). A bibliometric analysis reveals that the number of genomic surveillance related publications increased fourfold between 2019 to 2021 (Figure 2), which is linked to the increasing amount of sequence data available and shared on platforms such as GISAID. The GISAID database now contains more than one million SARS-CoV-2 sequences (Maxmen, 2021). Before the COVID-19 pandemic, influenza virus genomic data such as avian H5 influenza strains (World Health Organization, 2021) or seasonal human influenza such as A(H3N2) (GISAID EpiFlu Database, 2021), was typically shared (Global consortium on H5N8 and Related Influenza Viruses, 2016). Several publications have highlighted the importance of genomic surveillance as a tool to monitor pandemics, for example the mapping variant emergence through time throughout the African continent (Wilkinson et al., 2021), California (Deng et al., 2020) and Australia (Rockett, 2020). However, these publications do not portray the efficiency of rapid pandemic response of the genomic surveillance systems in question. Even though frameworks such as the IHR have now included genomics and pathogenic sequence data sharing as part of genomic surveillance strategies, national genomic surveillance programs are not established equally everywhere. If we measure the percentage of positive COVID-19 cases that went through a *genomic* surveillance pipeline that performed SARS-CoV-2 viral RNA sequencing of the samples and shared (i.e., submitted the data to the GISAID database), we can see that national genomic surveillance programs look like they are on their way to becoming well established in Australia , Denmark, Iceland, and New Zealand (Table 3) with pathogen sequence information of more than 40 percent of positive disease cases in the population being shared on the platform.[14] However, the larger countries have sequenced and shared orders of magnitude more positive disease cases in number, although less of a percentage of the total positive cases population (Wadman, 2021, Stevens, 2021). Hence, questions of scalability and what constitutes program successes may be brought up. It illustrates the importance of the issues of metrics and their interpretability. Another metric that could give an indication of genomic surveillance efficiency, is the lag between sample sequencing and submission of sequence data to GISAID. This lag is larger than a couple of days in many low- and middle-income countries (LMIC). These metrics could be good indicators to include in the GHS Index, but further research into why certain nations score high or low on either metric is needed. This could be due to lack of reagents owing to disruptions in global supply chains, lack of equipment and infrastructure, scarcity of technical skills, or hesitancy to share data (Wilkinson et al., 2021). This in turn can reveal novel biosecurity risks within the genomic surveillance pipeline.

As a result of the global COVID-19 pandemic, multiple institutes are publishing frameworks to reduce barriers to genomic surveillance, supporting infrastructure building in LMIC and accelerating the availability of additional clinical metadata. For example, in the U.S. a Genomic Surveillance Working Group established in April 2021, co-led by WHO and The Rockefeller Foundation and coordinated by FIND,[15] with a focus on increasing Next-Generation Sequencing (NGS)[16] infrastructure and expertise in LMIC to expedite the end of the COVID-19 pandemic. The Rockefeller Foundation report called for a GDPR (Rockefeller Foundation, 2021). The fast and urgent nature of pandemic responses have recently led

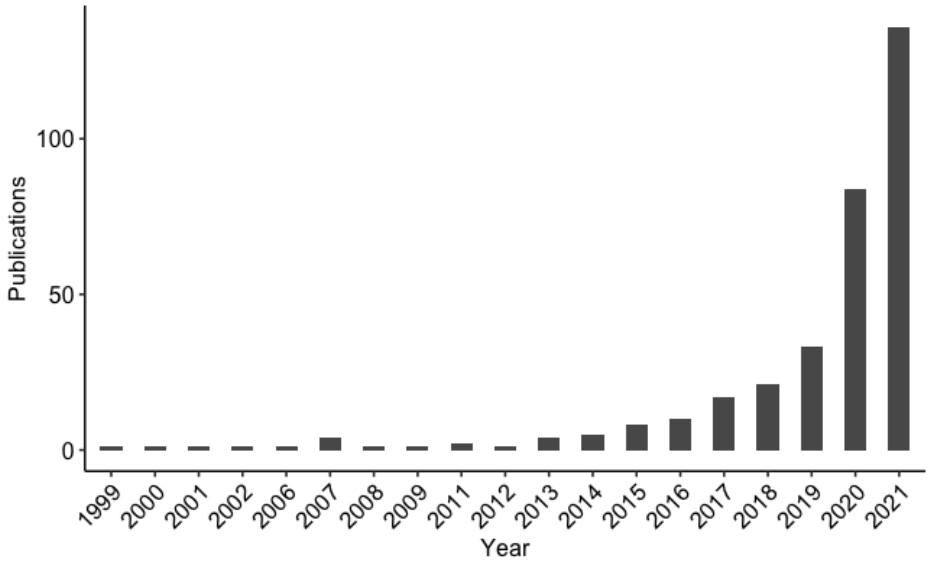
[14] SARS-CoV-19 is a positive single-stranded RNA virus. GISAID contains a database with specifically raw data of the whole-genome, not just spike protein information, for example.

[15] FIND, the global alliance for diagnostics. https://www.finddx.org/about/

[16] A sequencing technique first introduced in 2004, with which independent sequence data from millions of individual DNA molecules can be read, allowing each fragment to be classified independently. This has enabled easier and faster sequencing of large genomes.

to HIPAA waivers[17] for private companies that are contracted by the U.S. Government.[18] In February 2021, a new bill introduced by Congress requested new funds for infrastructure building for sequencing, drastically improving data-linkage between health data such as sequence data and social indicators such as sexual orientation, race, and ethnicity (Tracking COVID-19 Variants Act, 2021). This recent data-linkage and infrastructure support has been proposed without addressing potential biosecurity risks.

Figure 2. Bibliometric Analysis of Genomic Surveillance Literature



Number of articles published yearly with topic "Genomic surveillance." The year 2021 contains articles published through October 27, 2021. Analysis performed with the Web of Science database and R.

Table 2. Selected List of National Disease Genomic Surveillance Programs

Country or Region	Program Name	Pathogen Interest	Reference
India	Indian SARS-CoV-2 Genomics Consortium (Insacog)	SARS-CoV-2	https://www.mohfw.gov.in/
United States	CDC Spheres	SARS-CoV-2	https://www.cdc.gov/coronavirus/2019-ncov/variants/spheres.html
United States	CDC Serology Survey	SARS-CoV-2	https://www.cdc.gov/coronavirus/2019-ncov/covid-data/serology.html
Australia	Communicable diseases genomics network	SARS-CoV-2, Salmonella, Listeria, Tuberculosis	https://www.cdgn.org.au/
United Kingdom	COVID-19 Genomics U.K. Consortium (COG-UK)	SARS-CoV-2	https://www.cogconsortium.uk/

[17] The Secretary of the U.S. Department of Health and Human Services may waive certain provisions of the Rule under the Project Bioshield Act of 2004 (PL 108-276) and section 1135(b)(7) of the Social Security Act (United States).

[18] For example, during the COVID-19 pandemic, Helix Ltd. and Illumina Inc. were contracted by the CDC to provide sampling kits, perform sampling, the sample preparation, and sequencing. Helix mentions the following on their program website: "We have received a waiver of consent for a limited data set under HIPAA regulations for purposes of public health (Privacy Rule (45 CFR 165.512b)). The Helix research team receives no identifiable information beyond the limited dataset that includes sample collection date and zip code..."

Country or Region	Program Name	Pathogen Interest	Reference
U.K., Colombia, India, Philippines, Nigeria	NIHR Global Health Research Unit Genomic Surveillance of Antimicrobial Resistance	Antimicrobial Resistance bacteria	https://Ghru.Pathogensurveillance.Net/
Kenya	Kemri Wellcome Trust Kilifi	SARS-CoV-2	https://Shrts.Net/A6ity
West Africa	Genomics of Infectious Diseases	SARS-CoV-2, Malaria	https://Ace.Aau.Org/Ace-1-Centers/Acegid/
African continent	Institute of Pathogen Genomics - Africa CDC	SARS-CoV-2	https://Africacdc.Org/Institutes/Africa-Pathogen-Genomics-Initiative
Nigeria	Centre for Human Virology and Genomics	SARS-CoV-2, viral hemorrhagic diseases	https://Nimr.Gov.Ng/Molecular-Biology-Biotechnology
South Africa	Network for Genomic Surveillance	SARS-CoV-2	https://www.Krisp.Org.za
Uganda	Uganda Virus Research Institute	SARS-CoV-2, Ebola, HIV	https://www.Uvri.Go.U
African continent	H3Africa Consortium	Sickle Cell Disease, Cardiovascular Diseases	https://H3africa.Org/Index.Php/Consortiumfcgv
South Africa, Senegal	United World Antiviral Research Network (UWARN)	SARS-CoV-2	https://cerid.uw.edu/uwarn/our-partners

Table 3. Select Data from GISAID SARS-CoV-2 Submission Tracker

	Viral genomes shared	Reported COVID-19 cases	Percentage of positive cases sequenced and shared
New Zealand	1,120	2,499	44.8
Denmark	124,810	309,420	40.3
Australia	19,248	32,427	59.4
Iceland	5,070	6,967	72.8
United States of America	650,002	33,875,385	1.92
United Kingdom	586,066	5,602,325	10.5
India	38,961	31,293,062	0.125
Germany	142,828	3,752,592	3.81

GISAID is a database of internationally shared genetic sequences, particularly for COVID viral strains. "Viral genomes shared" indicates the number of viral genomes (the genetic sequence information of the SARS-CoV-2 virus) shared with GISAID. "COVID cases" indicates the number of positive cases in that specific country. "Percentage of positive cases sequenced and shared" provides an indicator for the "efficiency" of national SARS-CoV-2 genomic surveillance, with Australia, Iceland, and New Zealand leading worldwide. Source: GISAIDRisk analysis

Risk Analysis

Gaps In Dual-Use Research Governance for Genomic Surveillance Data

Technology-dependent norms for responsible life science research have been established in the past. The resulting risk assessments of dual-use research typically focus on engineered organisms or pathogens, and experiments that could demonstrate enhancing capabilities of these pathogens (Burnette, 2021). For example, the Fink report (National Research Council, 2004) presents seven experimental categories of high-risk pathogenic use with dual-use potential. A similar dual-use assessment strategy is used in the International Federation of Biosafety Associations framework, and a similar application focused bioweapons and terror dual-use assessment is proposed in the Malice Analysis Framework (EBRC, n.d., National Research Council, 2011). The WHO Responsible Science framework (WHO, 2010) mentions disease surveillance as a biorisk management tool yet does not consider possible bio risks generated by disease surveillance methods. Risks related to sequence information, patient information or other digital information are typically not included in the definition of dual-use research of concern (DURC) (WHO, 2021b). With the continuing convergence of Big Data and the life sciences, the definition of dual-use research needs to expand. Political and security concerns with the help of data can identify novel risks (Mahfoud et al., 2018). A recently suggested risk assessment framework designed around data security issues in biology has been outlined by the American Association for the Advancement of Science (AAAS) (AAAS, FBI, UNICRI, 2014). In this report, the AAAS examines the risks and benefits associated with Big Data analytics in the life sciences and develops frameworks for risk and benefit assessments of emerging or enabling technologies. Beyond this example, frameworks built around a dual-use data and health security perspective have not been fully established.

In order to develop risk mitigation strategies, we first analyze the hazards, risks and resulting harms related to the use of genomic surveillance strategies for disease monitoring methodologically. Here, we take the liberty to expand the definition of dual-use risk assessment by analyzing risks at every step in the genomic surveillance pipeline for disease monitoring. These include security risks, data risks, technological risks, and risks posed by a lack of governance by current norms and regulations.

Security Risks

Wrong implementation of disease genomic surveillance infrastructure and logistics, and the use of these capacities for acquiring human genomic information, can entail security risks. The economic benefit of sequencing large sets of samples, the ease of access to population-scale human biospecimens, the lack of genomic disease surveillance frameworks (see following section), and the current infrastructure support opens doors to potential new harms. If security of genomic surveillance data and data infrastructure can result in the manipulation of the data infrastructure (increasing lag time between sequencing and data sharing) or the data itself to provide misinformation on epidemiological status of a country, this could result in a misportrayal of the potential public health risk. Misuse of genomic surveillance pipelines can include their use for population-wide genomic studies to gain intentional insights into disease and population dynamics for drawing discriminating correlations between them. Population genetics knowledge can be used to gain information on subpopulations. This information can be health related (e.g., hereditary diseases which are protected by genetic non-discrimination acts GNDA and GINA) or identifying (biomarkers that can aid in the identification of population subgroups). This can lead to discrimination based on sex at birth, ancestry, disease (mental and physical), or disability. It could also potentially lead to the design of ethnic targeted bioweapons. In turn, human rights can be violated through repression, genocide, and systematic discrimination (Koblentz, 2013; Mahfoud, 2018). The requirements of actors to weaponize genomic surveillance pipelines include mainly the control over all pipeline implementing partners and sufficient

financial means. Because genomic surveillance infrastructure is deployed heavily during epidemics, it is up to national policy and budget allocation to follow IHR. If this infrastructure has less or no use, for example by lack of funds or inter-epidemic periods, there is an opportunity for repurposing this infrastructure to other large-scale population genomic studies that are potentially nefarious. Guidelines for disease surveillance during inter-epidemic periods are available, but don't detail potential security issues resulting from available logistics and infrastructure.

Data Risks

Vulnerabilities in the genomic surveillance data and cyber infrastructure can exist due to a lack of access to computational infrastructure. This can lead to potential harms (AAAS, FBI, UNICRI, 2014). For example, inappropriate access to genomic information by cyberattacks include data harvesting or disruptive malware. Furthermore, cyberattacks with data harvesting malware can retrieve intellectual property data relating to COVID-19 vaccine development. This has happened throughout 2020 and 2021 in the U.K., U.S., and Canada on multiple occasions (Muthuppalaniappan and Stevenson, 2021). A second potential harm is misdirection of public health information, by making changes to specific data entries within the database or by flooding the database with false information (AAAS, FBI, UNICRI, 2014). This is a high-risk scenario for disease genomic surveillance programs, as they directly inform on viral variant origin and epidemiology. This could hide the origin of certain variants or present a wrong picture of epidemiological information of states. These data can also be exploited beyond health security and can support other activities with national security concerns such as discrimination against population minorities or weaponization (including biological agents).

Technological Risks and Harms

The usability of the technology can reveal points of potential risk. Being able to sequence, store, and analyze genomic data on a population level entails multiple levels of complexity. First, to obtain purified biological material such as DNA or RNA, sampling using commercially available kits needs to be performed.[19] Human samples that contain disease vector information also contain human genome and other Omic information. At the sample preparation step, only part of the information, such as viral RNA, is maintained by extraction and selective amplification steps when this is desired. The sample preparation kits are particularly tuned to the end application. For example, SARS-CoV-2 whole genome sequencing requires RNA extraction from viral particles, conversion into DNA, and preparation for sequencing (CDC Advanced Molecular Detection, 2021). This kit differs from a genealogy sample preparation kit where human DNA is extracted from human cells and relevant areas are amplified and prepared for sequencing. Sample preparation (purification, amplification, barcoding) might need trained laboratory personnel, and is frequently outsourced to sequencing centers (private or government). The logistics of sampling and sample preparation can become complex on a large scale, often relying on many third parties, but also becomes more affordable. For example, next generation sequencing technology hardware is often inherently designed to sequence multiple samples at the same time.[20] Additionally, logistics do not differ whether performing a surveillance project for disease vectors, forensic purposes, or human genomic studies. This can make it attractive for large-scale population studies. The sequencing itself is also fairly accessible. Although traditional sequencing providers or sequencing machine providers screen their customers based on affiliation, use case etc., from historical examples we've seen this is no longer a barrier. Sequencing technologies are also

[19] For example, popular SARS-CoV-2 genomic surveillance kits are ThermoFisher's TaqPath™ COVID-19 Combo Kit. These are SARS-CoV-2 sample preparation kits for RNA extraction and amplification of PCR target sites.

[20] For example, up to 3,072 samples per run on Illumina NovaSeq 6000 S4 and SP flow cells, up to 384 samples per run on Illumina NextSeq 2000, Illumina NextSeq 500/550/550Dx sequencing machines.

becoming more portable and accessible.[21] The pipeline to go from sample to data with such devices is not complex, with kits and devices being offered commercially. An unregulated market of sequencing technologies can be of concern with dropping price points and a competitive market. Next, analyzing genomic data requires strong computing power, which could limit the accessibility (Oakeson et al., 2017). Some sequencing machines such as those from Illumina, have solved this problem by performing the alignment and initial data handling in the cloud on their private servers. The raw sequence info is sent back to the user, after Illumina's analysis. Shared servers and in-the-cloud computation are currently widely used for further epidemiological analysis and simplify the hardware needs. Because these are based on commercial platforms such as Amazon Web Services, security and data risks exist. Interception of epidemiological data, and manipulation are possibilities. Interestingly, the genomic revolution has now led us to a point where it is cheaper to store biospecimens and re-sequence the samples over time, rather than storing the genomic data on servers.[22] How this will affect genomic surveillance is unclear yet. To conclude, genomic surveillance pipelines are technologically built to accommodate handling of hundreds of samples at a time. It is much more accessible financially and practically compared to a handful of samples. The technology facilitates large scale work and does not restrict the use case.

Lack of Governance: Norms and Regulations

We established that disease genomic surveillance pipelines provide opportunity for dual-use risks. Do frameworks that guide disease genomic surveillance pipelines recognize these risks? The GHS Index assessment also discovered that international health and security norms were lacking in compliance (e.g., only 5 percent had a genetic data sharing policy or sharing biological materials beyond influenza, public health emergency cross-border agreements lacking (in 69 percent), and less than half have conducted a WHO JEE (Bell et al., 2017)). There is a lack of compliance with IHR (2005) (WHO/CDS/EPR/IHR/2007.1) (WHO, 2008) and the PIP, which provide the main guiding framework for disease surveillance and management. This is not only due to a lack in surveillance capacity but could also be declared by WHO IHR recommendation counteracting national regulations during disease outbreaks. Furthermore, the IHR does not address certain modern biosecurity issues such as concerns with sampling, or sample and genetic information sharing (Gostin and Katz, 2016).[23]

One framework that does identify biosecurity risks within disease genomic surveillance pipelines is the European Centre for Disease Prevention and Control (ECDC) strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations. The ECDC "Expert opinion on whole-genome sequencing for public health surveillance" (ECDC, 2016) and "strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations" (ECDC, 2016), are to our knowledge one of the few surveillance guidance documents we found to thoroughly outline possible risks in implementing genomic surveillance. Although this strategic framework again focuses on the benefits of connecting patient health data to epidemiological data, it does address the potential security hazards in sharing genomic information.

[21] For example, Oxford Nanopore Technologies sequencing currently leading in this domain.

[22] "A typical Oxford Nanopore PromethION run will generate a terabyte of raw data. This is usually what you want to store as the current algorithms don't extract all biological information yet. However, storing a TB on DNA nexus, a popular genomics cloud computing service, will cost you USD 360 a year, excluding any operations on the data or downloading the data again. A PromethION flowcell will set you back USD, …. , it becomes cost-effective to re-sequence instead of storing the data after a few years …." (Private communication, Prof. Martin Smith, August 1, 2021).

[23] Article 6(2) of the IHR offers some guidance for information sharing, stating that States Parties shall continue to provide relevant public health information to the WHO following a potential PHEIC notification, however there is no explicit mention of sharing biological materials or genetic sequence data.

We established that disease genomic surveillance pipelines have technical infrastructures that provoke dual-use risks, and frameworks to guide disease genomic surveillance pipelines do not recognize these risks. Can existing norms and regulations regarding genetic data protect individuals in the event of possible misuse of human genomic information within a disease genomic surveillance framework? Structures to guide and govern the genomic revolution are lagging behind the science and, in their absence, existing country-level laws on health data are often applied to the ways genomic information will be collected and used (WEF, 2019). National legislation such as HIPAA and GINA in the U.S., GNDA in Canada, and GDPR in Europe are put in place to protect patient genomic information yet they do not protect patients de-identified genomic information, do not apply to law enforcement, can be waived during a public health crisis, or can be violated with subpoenas. Examples include the Chinese government collecting DNA from several marginalized groups, some of which are in vulnerable or at-risk status, others for purposes of research in order to build associations between genes and characteristics, and a third category of building DNA databases to help with future law enforcement.

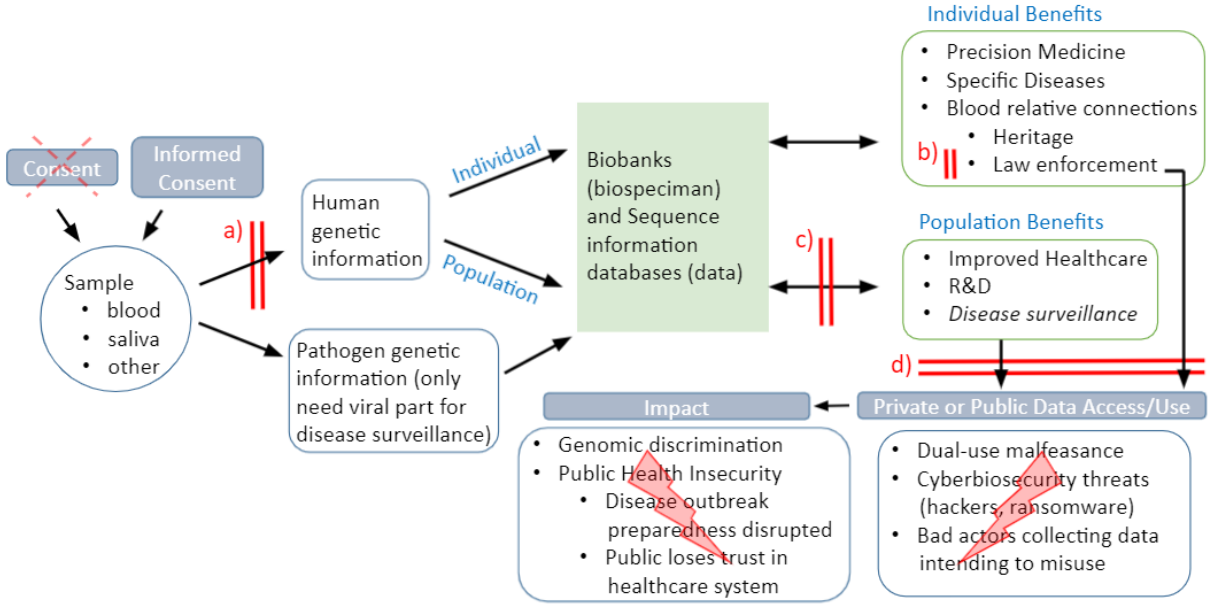
It is also worth noting that recently, the benefits of data sharing and phenotypic-genotypic data linkage have been presented to policy makers as a priority (e.g., Rockefeller Foundation, 2021). The security risks of genomic data are not well considered, which has led to discussion of changing regulations to facilitate data sharing or ownership. For example, while a Notice of Proposed Rulemaking (NPRM) in September 2015 suggested the Common Rule biospecimens definition include non-identifiable information, this was not adopted, meaning that any non-identifiable information does not require informed consent. This is more practical when performing health research that involves bioinformatics. The Common Rule only protects identifiable information. However, this carries questions about identifiability of biospecimens. Is whole-genomic sequence information, stripped from identifiers, truly not identifiable? Research suggests otherwise (Yousefi, 2018). Similarly, GDPR is receiving strong backlash from the research community. Furthermore, while DSI for the NP currently does not include human genetic information, suggestions are on the table to change this definition. Countries have suggested that aboriginal communities are a biodiversity asset as well, and therefore the country should be able to protect this DSI under the NP. But, this method of DSI protection can bring about a risk, as it grants ownership of population DSI to a country, which can be powerful for nefarious regimes (WHO, 2018).

The United States had enacted another type of regulation to counter the misuse of specifically genetic profiling enabling tools and equipment. Foreign policy measures can have quite significant effects on the use of certain technologies such as sampling kits or sequencing tools in distinct areas, by export control measures (Export Administration Regulations, n.d.). These measures have failed to control the export of U.S. materials as reports came out from private investigations into public tenders indicating that Promega and ThermoFisher were selling sequencers and sampling kits to Chinese law enforcement (Wee, 2021).

Balancing Risks and Benefits of Genomic Disease Surveillance

Disease genomic surveillance pipelines have technical infrastructures that provoke a dual-use risk. Frameworks to guide disease genomic surveillance programs do not recognize these risks, and existing norms and regulations regarding genetic data do not protect individuals' human genomic information adequately. This leaves us with potential harmful scenarios for public health and safety, individuals' privacy, and opportunities for political or cultural discrimination. We conclude that the disease genomic surveillance pipelines have possible dual-use scenarios that have not been explored thoroughly. As a summary, Figure 3 illustrates a diagram for the genomic surveillance process and depicts where we can draw red lines.

Figure 3. Dual-Use Risks of the Disease Genomic Surveillance Pipeline



Informed consent should be given to individuals. (a) Samples enter the disease genomic surveillance pipeline, after which potentially human genome data can be collected. Genetic information is stored in databases that could pose security risks. (b) Law enforcement and forensic data use can be of malicious intent when performing large population studies. (c) Regulation is needed for potential misuse in population genome studies. (d) Dual-use risks lead to downstream impacts.

One major factor affecting the establishment of new norms is the rapid pace of technology development. The advancement in sequencing technology, data storage and bioinformatics analysis has been hugely beneficial for global health security. In particular, the population-wide large-scale sequencing and monitoring of potential public health threatening pathogens, have taught us about the emergence of pathogen variants, viral dynamics, and the effectiveness of our response mechanisms. Balancing the societal benefit of public health surveillance against the state infringement of individual rights, the revised IHR for the first time incorporated human rights into infectious disease control, focusing on human rights considerations in disease surveillance. Ultimately, global health security remains a priority and we need to accelerate, facilitate, and support the sampling and sharing of pathogen sequences. At the same time, we need to be able to provide basic human privacy rights and foresee possible malicious use of genomic data during surveillance programs. This is the biggest challenge when drafting new norms and regulations.

Policy Recommendations

We see a first opportunity for control at the level of infrastructure, materials, and methods. The particular sample preparation kit used can reveal the end-user intentions.[24] Through company distribution channels and technical field service agents we can find out who the end-users of these specific kits are or whether these kits end up at resellers. Similarly, transparency on where the sequencing is outsourced to and what data analysis frameworks are being used can help with prevention. Misuse of equipment can also be

[24] Pathogen genotyping sample preparation kits can easily be distinguished from paternity testing, human whole-genome sequencing, HLA genotyping, or mitochondrial DNA kits.

prevented by enforcing duty of due diligence for large businesses providing equipment/products used in the genomic surveillance pipeline. However, we recognize the challenge in this. A more feasible method of control on the technological level is investing in technological methods to secure genomic data and improve secure storage and data sharing.

Another opportunity for control is at the level of consent. Consent methods for sampling in surveillance programs need to be standardized due to the huge difference in consent norms globally (broad consent, informed consent, or other norms)[25] (Ali et al., 2021). Indiscriminate and compulsory collection of samples should be sanctioned.

To guide the development of new genomic data policies and adjust existing approaches, we recommend the using the WEF Genomic Data Policy Framework for an overview of the current state of genomic privacy governance, and the GA4GH disease surveillance guidelines on consent, data privacy, data access, and benefit-sharing. In these documents, elements that are important to maintain and lead a responsible genomic surveillance program such as good leadership, transparency, and secure infrastructure, are central. Current frameworks such as the IHR can be improved by providing further detail on responsible practices related to data sharing, data infrastructure security, sampling and sample preparation, or outsourcing of sequencing. Furthermore, the IHR can detail concerns or solutions around inter-pandemic use of genomic surveillance infrastructure. Finally, it has the potential to enforce and define government genomic surveillance process transparency. The WHO JEE and the GHS Index tools could also include more metrics that evaluate data sharing capacities (Razavi, Erondu, and Okereke. 2020). Other metrics that can detect, identify, and monitor misuse beyond public health and pandemic preparedness can be included in the WHO JEE. Furthermore, we suggest frequent assessments that review the technological advances in the space of sequencing and data security. Good examples are the United States Common Rule agencies and departments, which are tasked with reexamining the definition of the terms "identifiable private information" and "identifiable biospecimen" at least every four years, based on novel analytical technologies and techniques such as sequencing (National Human Genome Institute, 2017).[26] An assessment of the IHR for example can include creating a list of sequencing, bioinformatics, and big data technologies that could influence the risk landscape of genomic disease surveillance, along with recommendations that go through a public comment process.

Finally, although working with human genetic information in genomic surveillance for disease monitoring pipelines is imminent, we need more awareness of the complexity of trying to protect the privacy of human genomic data while making the most of the public health benefit from large population studies and data sharing. This awareness will help us mitigate the potential risk of use of genomic surveillance programs for malicious purposes such as epidemiological data manipulation, or infrastructure repurposing for population control. Acknowledging these risks, scientists and policy makers alike can be facilitated by expanding our current definition of DURC in national biosafety and biosecurity guidelines to include a data security perspective. Based on that, we can develop guidelines to identify misuse of genomic surveillance for disease monitoring pipelines. We suggest that this identification methodology could include (1) auditing genomic surveillance programs, including sampling units, and sequencing centers; (2) screening publication records for suspicious collaborations between law enforcement and genomic researchers; (3) monitoring genomic surveillance programs in areas with existing political tension or human rights violations; (4) screening public tenders to investigate government contracts with sequencing providers. These can possibly then be used in WHO JEE, or even be implemented in the WHO surveillance system for attacks (SSA) on healthcare to identify countries where genomic surveillance infrastructure is misused (WHO, 2019).

[25] Ali et al. (2021) reviewed genetics and genomic research in Africa and offered an insight into global health issues such as broad consent, data sharing, and governance of biobanks that are still not addressed and challenged stakeholders into precise discussions to establish approaches via reputable institutions to protect and benefit researchers and populations.

[26] Under the Common Rule, information is identifiable if the subject's identity "may readily be ascertained" by the researcher.

Conclusion

Genomic surveillance pipelines for disease monitoring have possible dual-use vulnerabilities. A dual-use risk analysis indicates that a genomic surveillance pipeline can be repurposed to collect and analyze human genomic information using the same exact genomic surveillance pipeline. The rapid pace of biotechnological advances and the growing ability to sequence high throughput data creates a lag where new security risks are emerging more quickly than policy makers can address them. Historical and current examples of targeting and data misuse illustrate human rights violations. National legislation such as HIPAA and GINA in the United States and GDPR in Europe are put in place to protect patient genomic information, yet they do not protect patients' de-identified genomic information, do not apply to law enforcement, can be waived during a public health crisis, and be subject to subpoenas. Additionally, research and clinical settings navigate a completely different set of legislation and norms when using commercial partners or contractors. Some open questions we're left with include:

- With inherent lack of transparency, if genomic surveillance information was already stolen/misused from sampling kits, could human genomic information have already been retrieved?
- If a blueprint of a population is available, what power does this give to countries?
- What does it mean if governments obtain property rights over their population genomic data?
- If accessibility to human genomic data will be normalized due to the effect of sequencing technology development on the market, what guidelines and legislation should be in place to protect an individual's privacy?

To conclude, the risks of genomic surveillance include the using private genomic information without consent, using genomic information to target minorities at risk, losing beneficial disease genomic surveillance, and losing trust in the healthcare system on a large scale. This will affect our pandemic response, drug development and human health in general, restricting the future use of DSI. Strengthening our genomic disease surveillance frameworks by acknowledging these risks, calling for transparency, and improving security can help in maintaining the benefits of genomic disease surveillance.

Acknowledgments

We would love to thank the following experts for their advice, time, and valuable contributions. This report does not reflect their personal opinions.

- Prof. Gregory Koblentz, George Mason University
  Private Communication with J.S., initial and overarching advice

- Prof. Martin Smith, CHU Saint-Justine, University of Montreal
  Private spoken communication with K.S. Email conversation with K.S., J.S., E.N. Topics include Nanopore sequencing, sequencing workflow, genomic data handling, and future perspectives

- Bev-Heim Meyers, Canadian Coalition on Genetic Fairness
  Private spoken communication with K.S. on the Genetic Non-Discrimination Act (GNDA)

- Prof. Yves Moreau, KU Leuven, GeneWatch and Human Rights Watch
  Private spoken communication with K.S, J.S, E.N. on identifying misuse in genomic surveillance research

- Prof. Mahsa Shabani, KU Leuven
  Private spoken communication with K.S., on genomic data protection by GDPR

- Dr. Misaki Wayengera, Makerere University
  Private communication with E.N. on genomic surveillance for SARS-CoV-2 in Uganda

- Prof. Andrew Kilianski, GMU and IAVI
  Private spoken communication with K.S., J.S. on dual use and risk in genomic surveillance research

We thank NTI, the Next Generation Global Health Security Network, and the competition judges, including Gabby Essix, Hayley Severance, Taylor Winkelman, Mohammed Jibriel, Doreen Nyakato, Kate Kerr, Ryan Ritterson, Jessica Smrekar, and Prof. Gregory Koblentz, for their support.

We thank the paper reviewers Prof. Gregory Koblentz, Dr. Mark Kazmierczak, and Hayley Severance for their feedback.

Acronyms

BGI	Beijing Institute of Genomics
CBD	Convention on Biological Diversity
CDC	U.S. Centers for Disease Control and Prevention
DSI	digital sequence information
DURC	dual-use research of concern
EAR	Export Administration Regulations
ECDC	European Centre for Disease Prevention and Control
FAO	Food and Agriculture Organization of the United Nations
GA4GH	Global Alliance for Genomics and Health
GDPR	General Data Protection Regulations
GHS Index	Global Health Security Index
GINA	Genetic Information Nondiscrimination Act
GISAID	Global Initiative on Sharing Avian Influenza Data
HIPAA	Health Insurance Portability and Accountability Act of 1996
ICJ	International Court of Justice
IHR	International Health Regulations
IRB	Institutional Review Board
JEE	World Health Organization Joint External Evaluation
LMIC	Low- and middle-income countries
NPRM	Notice of Proposed Rulemaking
OIE	World Organisation for Animal Health
OCR	Office of Civil Rights
PIP	Pandemic Influenza Preparedness
NCBI	National Center for Biotechnology Information
NIFTY	Non-Invasive Fetal Trisomy
NGS	next-generation sequencing
NP	Nagoya Protocol
UN	United Nations
WEF	World Economic Forum
WHO	World Health Organization

References

AAAS, FBI, and UNICRI. 2014. *National and Transnational Security Implications of Big Data in the Life Sciences*. Washington, DC: American Association for the Advancement of Science. http://www.aaas.org/sites/default/files/AAAS-FBI-UNICRI_Big_Data_Report_111014.pdf.

Ali, Joseph, Betty Cohn, Erisa Mwaka, Juli M. Bollinger, Betty Kwagala, John Barugahare, Nelson K. Sewankambo, and Joseph Ochieng. 2021. "A Scoping Review of Genetics and Genomics Research Ethics Policies and Guidelines for Africa." *BMC Medical Ethics* 22 (1): 39. https://doi.org/10.1186/s12910-021-00611-9.

Beauvais, Michael, et al. n.d. Responsible Data Sharing to Respond to the COVID-19 Pandemic_ Ethical and Legal Considerations v. 3.0. Accessed October 27, 2021. https://www.ga4gh.org/covid-19/.

Bell, Elizabeth, Jordan W. Tappero, Kashef Ijaz, Maureen Bartee, Jose Fernandez, Hannah

Burris, Karen Sliter, et al. 2017. "Joint External Evaluation—Development and Scale-Up of Global Multisectoral Health Capacity Evaluation Process." *Emerging Infectious Diseases* 23 (13). https://doi.org/10.3201/eid2313.170949.

Bogner, Peter, Ilaria Capua, David J. Lipman, and Nancy J. Cox. 2006. "A Global Initiative on Sharing Avian Flu Data." *Nature* 442 (7106): 981–981. https://doi.org/10.1038/442981a.

Burnette, Ryan N., ed. 2021. *Applied Biosecurity: Global Health, Biodefense, and Developing Technologies*. Advanced Sciences and Technologies for Security Applications. Springer International Publishing. https://doi.org/10.1007/978-3-030-69464-7.

Centers for Disease Control and Prevention. 2021. "How It Works | Advanced Molecular Detection (AMD) | CDC." August 30, 2021. https://www.cdc.gov/amd/how-it-works/index.html.

Collins, Francis. "Genome Data Help Track Community Spread of COVID-19." 2020. NIH Director's Blog. July 21, 2020. https://directorsblog.nih.gov/2020/07/21/genome-data-helps-track-community-spread-of-covid-19/.

Consolidated Federal Laws of Canada. 2017. *Genetic Non-Discrimination Act*. https://laws-lois.justice.gc.ca/eng/acts/G-2.5/page-1.html#h-1.

Council of Europe. 1997. "Convention for Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Biomedicine: Convention of Human Rights and Biomedicine." *Kennedy Institute of Ethics Journal* 7 (3): 277–90. https://doi.org/10.1353/ken.1997.0021.

Deng, Xianding, Wei Gu, Scot Federman, Louis du Plessis, Oliver G. Pybus, Nuno R. Faria, Candace Wang, et al. 2020. "Genomic Surveillance Reveals Multiple Introductions of SARS-CoV-2 into Northern California." *Science* 369 (6503): 582–87. https://doi.org/10.1126/science.abb9263.

Dennis, Brady, Chris Mooney, Sarah Kaplan, and Harry Stevens. 2020. "Scientists Have a Powerful New Tool for Controlling the Coronavirus: Its Own Genetic Code." *Washington Post.* Accessed July 25, 2021. https://www.washingtonpost.com/graphics/2020/health/coronavirus-genetic-code/.

Dirks, Emile, and James Leibold. 2020. "Genomic Surveillance: Inside China's DNA Dragnet." 34. ASPI International Cyber Policy institute.

Engineering Biology Research Consortium (EBRC). n.d. "Malice Analysis: Assessing Biotechnology Research for Security Concerns." Accessed August 3, 2021. https://ebrc.org/malice-analysis/.

European Centre for Disease Prevention and Control (ECDC). 2016. *Expert Opinion on Whole Genome Sequencing for Public Health Surveillance: Strategy to Harness Whole Genome Sequencing to*

Strengthen EU Outbreak Investigations and Public Health Surveillance. LU: Publications Office. https://data.europa.eu/doi/10.2900/12442.

Export Administration Regulations (EAR). n.d. Accessed August 3, 2021. https://www.bis.doc.gov/index.php/regulations/export-administration-regulations-ear.

Food and Agriculture Organization of the United Nations (FAO), World Organisation for Animal Health (OIE), World Health Organization (WHO). *FAO/OIE/WHO Tripartite Statement on the Pandemic Risk of Swine Influenza.* 2009. https://www.who.int/influenza/gip/FAO-OIE-WHO_RiskSwineFlu_082020.pdf?ua=1.

GISAID. "GISAID—Submission Tracker Global." 2021. Accessed July 25, 2021. https://www.gisaid.org/index.php?id=208.

GISAID EpiFlu Database. 2021. "GISAID - Influenza Genomic Epidemiology." 2021. https://www.gisaid.org/epiflu-applications/influenza-genomic-epidemiology/.

GISAID. 2019. *Notice and Request for Comments on the Implications of Access and Benefit Sharing (ABS) Regimes on Global Health and Biome.* Accessed July 25, 2021. https://www.gisaid.org/fileadmin/gisaid/files/pdfs/GISAID_Comments_DOS-10789.pdf.

Global Alliance for Genomics and Health (GA4GH). n.d. *Global Alliance for Genomics and Health: Consent Policy.* Accessed August 2, 2021. https://www.ga4gh.org/wp-content/uploads/Consent-Policy-Final-27-May-2015.pdf.

Global Consortium for H5N8 and Related Influenza Viruses. 2016. "Role for Migratory Wild Birds in the Global Spread of Avian Influenza H5N8." *Science* 354 (6309): 213–17. https://doi.org/10.1126/science.aaf8852.

Global Health Security Index 2019. Accessed August 3, 2021. https://www.ghsindex.org/.

Gostin, Lawrence O. and Rebecca Katz. 2016. "The International Health Regulations: The Governing Framework for Global Health Security." *The Milbank Quarterly* 94 (2): 264–313. https://doi.org/10.1111/1468-0009.12186

Human Rights Watch. 2017. "Kuwait: Court Strikes Down Draconian DNA Law." *Human Rights Watch*, October 17, 2017. https://www.hrw.org/news/2017/10/17/kuwait-court-strikes-down-draconian-dna-law.

Khoury, Muin J. July 19, 2017. "Integrating Genomics into Public Health Surveillance: Ushering in a New Era of Precision Public Health." n.d. Centers for Disease Control and Prevention. Accessed July 25, 2021. https://blogs.cdc.gov/genomics/2017/07/19/integrating-genomics/.

Knoppers, Bartha Maria. 2014. "Framework for Responsible Sharing of Genomic and Health-Related Data." *The HUGO Journal* 8 (1): 3. https://doi.org/10.1186/s11568-014-0003-1.

Koblentz, Gregory D. 2013. "Regime Security: A New Theory for Understanding the Proliferation of Chemical and Biological Weapons." *Contemporary Security Policy* 34 (3): 501–25. https://doi.org/10.1080/13523260.2013.842298.

Laird, Sarah A. and Rachel P. Wynberg. 2018. *Fact-Finding and Scoping Study on Digital Sequence Information on Genetic Resources in the Context of the Convention on Biological Diversity and the Nagoya Protocol*. CBD/DSI/AHTEG/2018/1/3.

Maxmen, Amy. 2021. "One Million Coronavirus Sequences: Popular Genome Site Hits Mega Milestone." *Nature* 593 (7857): 21–21. https://doi.org/10.1038/d41586-021-01069-w.

Mahfoud, Tara, Christine Aicardi, Saheli Datta, and Nikolas Rose. 2018. "The Limits of Dual Use." *Issues in Science and Technology* (blog). July 31, 2018. https://issues.org/the-limits-of-dual-use/.

Martin, Will. September 25, 2019. "China Is Harvesting Thousands of Human Organs from Its Uighur Muslim Minority, UN Human-Rights Body Hears." *Business Insider*. https://www.businessinsider.com/china-harvesting-organs-of-uighur-muslims-china-tribunal-tells-un-2019-9.

Merchant, Nomaan. 2020. "U.S. to Start Collecting DNA from People Detained at Border." *AP NEWS*, January 6, 2020, sec. Immigration. https://apnews.com/article/immigration-tx-state-wire-donald-trump-us-news-8e7d4e3d6e2ef24dcc4fd9e79552a3d0.

Moreau, Yves. 2019. "Crack Down on Genomic Surveillance." *Nature* 576 (7785): 36–38. https://doi.org/10.1038/d41586-019-03687-x.

Muthuppalaniappan, Menaka and Kerrie Stevenson. 2021. "Healthcare Cyber-Attacks and the COVID-19 Pandemic: An Urgent Threat to Global Health." *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care* 33 (1): mzaa117. https://doi.org/10.1093/intqhc/mzaa117.

National Center for Biotechnology Information (NCBI). 2021. "National Center for Biotechnology Information." 2021. https://www.ncbi.nlm.nih.gov/.

National Human Genome Research Institute. n.d. "The Human Genome Project." Genome.Gov. Accessed July 29, 2021. https://www.genome.gov/human-genome-project.

National Human Genome Research Institute. March 2017. "Highlights of Revisions to the Common Rule." 2017. National Human Genome Research Institute.Gov. https://www.genome.gov/about-genomics/policy-issues/Human-Subjects-Research-in-Genomics/Highlights-of-Revisions-to-the-Common-Rule.

National Institutes of Health (NIH). "South Africa Study Shows Power of Genomic Surveillance Amid COVID-19 Pandemic." 2021. February 18, 2021. https://www.nih.gov/south-africa-study-shows-power-genomic-surveillance-amid-covid-19-pandemic.

National Research Council. 2004. *Biotechnology Research in an Age of Terrorism*. 2004. Washington, DC: The National Academies Press. https://doi.org/10.17226/10827.

National Research Council. 2011. *Research in the Life Sciences with Dual Use Potential: An International Faculty Development Project on Education About the Responsible Conduct of Science*. https://doi.org/10.17226/13270.

Needham, Kirsty, and Clare Baldwin. 2021. "Special Report: China's Gene Giant Harvests Data from Millions of Women." *Reuters*, July 7, 2021, sec. APAC. https://www.reuters.com/article/us-health-china-bgi-dna-idUSKCN2ED1A6.

Normile, Denis. 2021. "Genetic Papers Containing Data from China's Ethnic Minorities Draw Fire." *Science* 373 (6556). https://www.science.org/content/article/genetic-papers-containing-data-china-s-ethnic-minorities-draw-fire.

Oakeson, Kelly F., Jennifer Marie Wagner, Michelle Mendenhall, Andreas Rohrwasser, and Robyn Atkinson-Dunn. September 2017. "Bioinformatic Analyses of Whole-Genome Sequence Data in a Public Health Laboratory." *Emerging Infectious Diseases Journal* 23, (9). — Accessed August 3, 2021. https://doi.org/10.3201/eid2309.170416.

Office for Civil Rights (OCR). 2020. "Important Notice Regarding Individuals' Right of Access to Health Records." Text. HHS.Gov. January 28, 2020. https://www.hhs.gov/hipaa/court-order-right-of-access/index.html.

Park, Alice. "Genetic Sequencing Could Revolutionize Public Health." June 11, 2021. *Time*. Accessed July 25, 2021. https://time.com/6071577/genetic-sequencing-covid-19/.

Radley-Gardner, Oliver, Hugh Beale, and Reinhard Zimmermann, eds. 2016. "Regulation (EU) 2016/679 of the European Parliament and of the Council." In *Fundamental Texts on European Private Law*. Hart Publishing. https://doi.org/10.5040/9781782258674.

Ramsay, Michèle. 2015. "Growing Genomic Research on the African Continent: The H3Africa Consortium." *South African Medical Journal* 105 (12): 1016. https://doi.org/10.7196/SAMJ.2015.v105i12.10281.

Razavi, Ahmed, Ngozi Erondu, and Ebere Okereke. 2020. "The Global Health Security Index: What Value Does It Add?" *BMJ Global Health* 5 (4): e002477. https://doi.org/10.1136/bmjgh-2020-002477.

Reardon, Sara. 2017. "Navajo Nation Reconsiders Ban on Genetic Research." *Nature* 550 (7675): 165–66. https://doi.org/10.1038/nature.2017.22780.

Redhead, C. Stephen. 2017. "Updated Common Rule: Research Using Stored Biospecimens." IF10653. Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF10653/2.

Rockefeller Foundation. 2021. "Implementation Framework: Toward a National Genomic Surveillance Network."

Rockett, Rebecca J., Alicia Arnott, Connie Lam, Rosemarie Sadsad, Verlaine Timms, Karen-Ann Gray, John-Sebastian Eden, et al. 2020. "Revealing COVID-19 Transmission in Australia by SARS-CoV-2 Genome Sequencing and Agent-Based Modeling." *Nature Medicine* 26 (9): 1398–1404. https://doi.org/10.1038/s41591-020-1000-7.

Sanger Institute. March 11, 2021. "Genomic Surveillance—The World's Binoculars Focused on Infectious Diseases." Wellcome Sanger Institute Blog. https://sangerinstitute.blog/2021/03/11/genomic-surveillance-the-worlds-binoculars-focused-on-infectious-diseases/.

Secretariat of the Convention on Biological Diversity. 2011. *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from Their Utilization to the Convention on Biological Diversity*. Montreal: United Nations.

Sekalala, Sharifah, Katrina Perehudoff, Michael Parker, Lisa Forman, Belinda Rawson, and Maxwell Smith. 2021. "An Intersectional Human Rights Approach to Prioritising Access to COVID-19 Vaccines." *BMJ Global Health* 6 (2): e004462. https://doi.org/10.1136/bmjgh-2020-004462.

Shabani, Mahsa, and Luca Marelli. 2019. "Re-Identifiability of Genomic Data and the GDPR." *EMBO Reports* 20 (6): e48316. https://doi.org/10.15252/embr.201948316.

United Nations Educational, Scientific and Cultural Organization (UNESCO). 1997. *Universal Declaration on the Human Genome and Human Rights*. *Res. 29 C/Res.16*. https://www.ohchr.org/EN/ProfessionalInterest/Pages/HumanGenomeAndHumanRights.aspx.

United Nations Human Rights Office of the High Commissioner (UNHROC). 2021. "China: UN Human Rights Experts Alarmed by 'Organ Harvesting' Allegations." June 2021. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=27167&LangID=E.

U.S. 104th Congress. 1996. Health Insurance Portability and Accountability Act of 1996.

U.S. Equal Employment Opportunity Commission (EEOC). 2008. *The Genetic Information Nondiscrimination Act of 2008*. https://www.eeoc.gov/statutes/genetic-information-nondiscrimination-act-2008.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, et al. 2001. "The Sequence of the Human Genome." *Science* 291 (5507): 1304–51. https://doi.org/10.1126/science.1058040.

Wadman, Meredith. 2021. "U.S. Rushes to Fill Void in Viral Sequencing as Worrisome Coronavirus Variants Spread." *Science*. February 9, 2021. https://www.sciencemag.org/news/2021/02/us-rushes-fill-void-viral-sequencing-worrisome-coronavirus-variants-spread.

Stevens, Harry and Miriam Berger. December 23, 2020. "U.S. Ranks 43rd Worldwide in Sequencing to Check for Coronavirus Variants like the One Found in the U.K." *Washington Post*. Accessed July 25, 2021. https://www.washingtonpost.com/world/2020/12/23/us-leads-world-coronavirus-cases-ranks-43rd-sequencing-check-variants/.

Tracking COVID–19 Variants Act. April 2, 2021. S.236, 117th Congress (2021–2022). https://www.congress.gov/bill/117th-congress/senate-bill/236/text.

Wee, Sui-Lee. 2021. "China Still Buys American DNA Equipment for Xinjiang Despite Blocks." *The New York Times*, June 11, 2021, sec. Business. https://www.nytimes.com/2021/06/11/business/china-dna-xinjiang-american.html.

Wee, Sui-Lee, and Tiffany May. 2020. "China's Offer to Help with Virus Testing Spooks Hong Kong." *The New York Times*, August 6, 2020, sec. Business. https://www.nytimes.com/2020/08/06/business/hong-kong-china-coronavirus-testing.html.

Wilkinson, Eduan, Marta Giovanetti, Houriiyah Tegally, James E. San, Richard Lessells, Diego Cuadros, Darren, P. Martin, et al. 2021. "A Year of Genomic Surveillance Reveals How the SARS-CoV-2 Pandemic Unfolded in Africa." *Science* 374 (6566): 423–31. https://doi.org/10.1126/science.abj4336.

World Economic Forum (WEF). 2019. "Genomic Data Policy: Resource List." Geneva.

World Health Organization (WHO). 2004. "Introduction." *WHO Guidelines for the Global Surveillance of Severe Acute Respiratory Syndrome (SARS): Updated Recommendations*. http://www.jstor.org/stable/resrep28003.3.

World Health Organization. 2008. *International Health Regulations (2005)*. 2nd ed. Geneva: WHO Press.

World Health Organization (WHO). 2010. "Responsible Life Sciences Research for Global Health Security: A Guidance Document." WHO/HSE/GAR/BDP/2010.2. http://www.ncbi.nlm.nih.gov/books/NBK305040/.

World Health Organization (WHO). 2018. "Implementation of the Nagoya Protocol in the Context of Human and Animal Health, and Food Safety: Access to Pathogens and Fair and Equitable Sharing of Benefits." Questions and answers. Convention on Biological Diversity and WHO. https://www.who.int/influenza/pip/QA_NP_Public_Health.pdf.

World Health Organization (WHO). 2019. *Surveillance System for Attacks on Health Care (SSA): Methodology*. Version 1.0. WHO/WHE/EMO/2019.2/BRO. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/312330.

World Health Organization (WHO). 2021a. *Pandemic Influenza Preparedness Framework*. 2nd ed. Geneva.

World Health Organization (WHO). 2021b. *Emerging Technologies and Dual-Use Concerns: A Horizon Scan for Global Public Health*. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/346862.

World Health Organization. 2021c. "Global Influenza Surveillance and Response System (GISRS)." 2021. https://www.who.int/initiatives/global-influenza-surveillance-and-response-system.

Yousefi, Soheil, Tooba Abbassi-Daloii, Thirsa Kraaijenbrink, Martijn Vermaat, Hailiang Mei, Peter van 't Hof, Maarten van Iterson, et al. 2018. "A SNP Panel for Identification of DNA and RNA Specimens." *BMC Genomics* 19 (1): 90. https://doi.org/10.1186/s12864-018-4482-7.

Zaaijer, Sophie, Assaf Gordon, Daniel Speyer, Robert Piccone, Simon Cornelis Groen, and Yaniv Erlich. 2017. “Rapid Re-Identification of Human Samples Using Portable DNA Sequencing.” Edited by Andrew P Morris. *ELife* 6 (November): e27798. https://doi.org/10.7554/eLife.27798.