# Managing the Transition to Widespread Metagenomic Monitoring: Policy Considerations for Future Biosurveillance

Chelsea Liang, James Wagstaff, Noga Aharony, Virginia Schmit, and David Manheim

The technological possibilities and future public health importance of metagenomic sequencing have received extensive attention, but there has been little discussion about the policy and regulatory issues that need to be addressed if metagenomic sequencing is adopted as a key technology for biosurveillance. In this article, we introduce metagenomic monitoring as a possible path to eventually replacing current infectious disease monitoring models. Many key enablers are technological, whereas others are not. We therefore highlight key policy challenges and implementation questions that need to be addressed for "widespread metagenomic monitoring" to be possible. Policymakers must address pitfalls like fragmentation of the technological base, private capture of benefits, privacy concerns, the usefulness of the system during nonpandemic times, and how the future systems will enable better response. If these challenges are addressed, the technological and public health promise of metagenomic sequencing can be realized.

**Keywords:** Metagenomic sequencing, Health policy, Biosurveillance, Disease X, Pandemic preparedness, Technology planning

## Introduction

THE COVID-19 pandemic has created tremendous political will to bolster pandemic preparedness. As such, this is an opportune time to ensure that investments and technology adoption policy are geared to prevent not just the next pandemic, but all future pandemics. Metagenomic shotgun sequencing is a set of methods that extracts genetic sequence data directly from environmental samples. While metagenomic sequencing is limited in

Chelsea Liang is an Independent Researcher, University of New South Wales, School of Biotechnology and Biomolecular Sciences, Sydney, Australia. James Wagstaff, PhD, is a Research Fellow, Future of Humanity Institute, University of Oxford, Oxford, UK. Noga Aharony, MS, is a PhD Student, Department of Systems Biology, Columbia University, New York, NY. Virginia Schmit, PhD, is Director of Research, 1DatSooner, DE, and a Policy Specialist, National Institute of Allergy and Infectious Diseases, Bethesda, MD. David Manheim, PhD, is Head of Policy and Research, ALTER, Rehovot, Israel; Lead Researcher, 1DaySooner, Claymont, DE; and Visiting Researcher, Humanities and Arts Department, Technion – Israel Institute of Technology, Haifa, Israel.

various ways, and other technologies could plausibly be adopted instead, it offers a disease-agnostic approach to monitoring, detecting, and characterizing pathogens and variants, and many disparate groups are working toward or promoting this future pathway.[1-3]

In this article we discuss policy obstacles to the establishment of a universal, scalable, One Health-upholding,[4] and pathogen-agnostic monitoring system, along with relevant technical and operational issues. While policy planning under uncertainty is always difficult, and concrete plans are premature, strategic thinking can convert technological and policy uncertainties into specific questions, which can then be addressed by the relevant academic, policy, and professional communities.

As with any transitional planning, we identify the starting point, the destination, and then consider transitional challenges. Accordingly, in this article we start with an overview of the current state of biosurveillance, then consider certain representative future systems, and finally describe some common challenges. We use the term "widespread metagenomic monitoring" (WMGM) to refer to future sequencing-based systems to detect infectious diseases and potential pandemic risks, WMGM is distinct from the current and often disjointed biosurveillance efforts and from other visions that are either not pathogen agnostic or are more narrowly focused on specific geographies or sources. We identify the following issues as needing policy solutions:

- Suboptimal use and high prices
- Privacy and data abuse
- Peacetime usefulness
- Enabling crisis response

If not adequately addressed, we expect these issues to massively delay or even prevent the implementation of a system to detect pandemic risks.

## Present State of Biosurveillance

In the United States, a complex set of programs exists where state-level control over some biosurveillance activities competes with multiple national programs. Meanwhile in many low-income countries, regional and global cooperation, often funded by international partners, is more common. Globally, current approaches track a limited number of patients using tests specific to a single disease and, for the most part, known diseases only.

### Geographic Heterogeneity

The practice of surveillance varies greatly around the world. In the United States, not only does the Centers for Disease Control and Prevention (CDC) run several disease-specific and syndromic biosurveillance programs, but the Department of Homeland Security's Countering Weapons of Mass Destruction Office runs both the BioWatch Program and the National Biosurveillance Integration Center.[5] Separate systems like the US Department of Agriculture's Animal and Plant Health Inspection Service are in place for agricultural and livestock disease monitoring.[6] These government systems tend to have limited data sharing between each other, or with other systems internationally. But even when open and widely used systems such as the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) are used,[7] public health officials more often flag outbreaks of notifiable diseases via doctor diagnoses, rather than via syndromic or other monitoring methods.

Meanwhile, many low-income countries have, at best, partial coverage of the population for basic health services. If the governments or health departments in the areas affected have the capacity to gather data on prevalence, they do so, but they often do not even aggregate extant data. Outbreaks are reported to the World Health Organization (WHO) when they are identified, and limited real-time analysis capacity exists, although the Africa Centres for Disease Control and Prevention (Africa CDC) and others are starting to address this gap.[8,9] At the same time, these countries are collaborating with the use of open-source tools like IDseq for analysis of metagenomic sequencing data.[10] The open nature of these systems enables faster analysis and increased operational resilience. We note a trend toward the increasing modularity of various nodes of biosurveillance.

### Genomic Data Gathering Paradigms

In the current paradigm, a given node that gathers disease data points is the same as, or is highly coupled to, the node that performs the analysis of such data. This vertical approach is often tied to a specific jurisdiction or data-gathering method, as seen among US agencies mentioned earlier. The vertical integration of analysis into gathering has meant that it is at best awkward—and at worst impossible—to aggregate data between different systems for a more comprehensive disease landscape. This is a key issue in current global infectious disease monitoring (Table 1).[33,34]

### Funding Infrastructure and Payment Systems

Funding for biosurveillance has always been uneven, with costs borne largely by high-income countries but with inconsistencies even there.[33,35] In some high-income countries, costs for tests even during a pandemic are often borne by consumers. The resulting implicit discrimination against lower-income and neglected communities has important direct impacts—which also leads to insufficient data and biases—that undermine surveillance efforts. The parts of surveillance that tend to maintain funding are for lower-risk issues like foodborne pathogens and rare reportable diseases, rather than robust infrastructure for detecting future outbreaks. In low-income countries, there is also a constant

Table 1. Current Popular Genomic Data-Gathering Paradigms

| | |
|---|---|
| Wastewater epidemiology | Prior to the COVID-19 pandemic, wastewater epidemiology was already in use, albeit primarily for monitoring of illicit drug use,[11-15] exposure to pesticides on a population level,[13] and some limited investigations of detection of poliovirus in South Africa.[16] |
| | • In the early use of monitoring for illicit drug use there was at least 1 indication of a nation monitoring migrant workers and targeting that specific area for mandatory drug testing and prosecution. Clearly, even for the "anonymized" sample that is wastewater, protections must be in place for vulnerable populations.[13] |
| | Even before COVID-19, it was recognized that wastewater epidemiology could be used more broadly for disease monitoring.[17] |
| | • During the COVID-19 pandemic it was used to monitor levels of this infectious organism both at the regional level in Australia,[18] Germany,[19] and Japan,[20] and at the more granular level of dormitory buildings or neighborhoods served by specific trunk lines.[21,22] |
| Clinical diagnostics | Point-of-care and point-of-use diagnostics using metagenomic sequencing are already available, albeit not ubiquitous in high-income countries.[23,24] Without comprehensive data on individuals testing positive for "disease X" in an emergent epidemic or pandemic setting, the possible availability of real-time epidemiological data is severely hampered. |
| | During the COVID-19 pandemic, in the United States, public health quarantine and isolation instructions were given based on testing data. |
| | • It was determined that weekly testing of all residents and staff at long-term care facilities was much more effective at identifying positive cases, and subsequently quarantining infected individuals, than waiting for observable symptoms to trigger testing.[25,26] |
| | • This further demonstrates the importance of testing in protecting communities, particularly for infectious diseases that demonstrate infectivity prior to symptom onset. |
| Reservoir biosurveillance | The US Department of Agriculture Animal and Plant Health Inspection Service currently monitors agricultural endeavors domestically in the United States and collaborates internationally on monitoring of agricultural models.[6] |
| | Known zoonoses have been monitored in large-scale agricultural settings.[27-30] |
| | • Animal producing farms located at the intersection of the urban–rural divide could also be settings where diseases previously unknown can spill from the wild population to the domesticated herd or flock, and from there infect hundreds or thousands of animals.[31,32] |

battle to maintain funding for surveillance systems, which can seem superfluous until they are vital (Table 2).

## POTENTIAL METAGENOMIC MONITORING FUTURES

Accurate long-term planning is challenging, and even more so when a plan is predicated on major technological progress. Thus any vision for WMGM must remain tentative and flexible. However, to get to WMGM responsibly and with maximized biosecurity benefits, there needs to be a common understanding of qualities we expect to see in a high-investment scenario.

## *Gather*

An expansive future WMGM system collects data from many nucleic acid sequencing data sources in a coherent set of formats. Other data-types are still available, but given the extent and rapidity of sequencing data, they are largely ancillary. The geographic coverage of nucleic acid data sources feeding into the system is extensive and global, and the cost per sample is minimal. Clinical use of metagenomic sequencing is routine and nearly universal for any suspected respiratory, urinary, and other infections, displacing disease-specific polymerase chain reaction (PCR)-, antigen-, or CRISPR (clustered regularly interspaced short palindromic repeats)-based tests. Similar to the use of syndromic surveillance today, subsets of this medical data are used for biosurveillance. Beyond clinical use for diagnosis, sampling and sequencing capacity is deployed directly for biosurveillance. This encompasses high-risk "sentinel" populations and civic-minded volunteers,[41] as well as agricultural and wilderness ecosystems,[42] built environments,[43] and urban wastewater, often at a neighborhood level. One potential example is a Nucleic Acid Observatory that monitors wastewater and waterways.[44] The rapid gathering of data by this decentralized network is routine and automated, wherever possible, enabling temporal trends to be quickly identified.

## *Analyze*

Analysis is possible in both a centralized and decentralized fashion. Local analysis includes diagnostics in clinical settings that replace and supersede (current) vertically integrated monitoring systems; the details of such systems are important but not our focus.

Table 2. Summary of Policy Shortcomings in Current Biosurveillance Efforts

| Aspect | Description | References |
|---|---|---|
| Data gathering | Gaps in coverage exist—not only geographic but temporal—with data being collected relatively infrequently and without sufficient metadata. | 33 |
| Epidemiological analysis | Analysis is often cumbersome and manual, especially for more detailed data, such as genomic data, and even aggregate and symptomatic data itself is not always widely shared or available. | 36,37 |
| Reporting | Reporting from individual sources is often slow and nonstandardized, especially internationally, and often does not feed into any unified analysis. | 38 |
| Analysis | The analysis that does occur is reported in haphazard ways, with no real standards. | 39 |
| Connection to policy response | Critically, there is no coherent link between the analysis and policy response. As a result, decisions made on the basis of the data are often made ad hoc or without understanding limitations of the analysis. | 40 |
| Privacy preservation | Even for the anonymized sample that is wastewater, protections must be in place for vulnerable populations. | 13 |
| Funding for biosurveillance | Funding is unstable and has always been uneven, with costs borne largely by high-income countries but is inconsistent even there. | 33,35 |

Note: Shortcomings exist in almost every area of the system, including gathering, analyzing, storing, and reporting data.

The data collection points tended by farmers and animal biosurveillance researchers, wastewater investigators, healthcare providers, and others can obtain genomic health insights more rapidly, cheaply, and of higher quality than they can generate themselves. The routinely collected data is also analyzed by academic, local government, and international biosurveillance experts, in near real time. These centers have strong links to political entities responsible for public health and biosecurity. Separating analysis from data gathering also encourages data standards and systems that enable subsequent analysis and broader sharing.

Other data sources for WMGM, such as indicators from other forms of data analysis (eg, internet search data) are integrated. Analysis of data streams in both clinical and public health applications may use a variety of publicly available and/or open-source software, which allows for continually improving and diverse ecosystems of analysis and prediction for clinical applications, national and international public health early warning, and research.

## Scale

All else being equal, a WMGM program seeking to maximize public health benefits (1) maximizes sampling density in space, time, and in terms of sequencing depth, and (2) minimizes the time between nucleic acid sampling and data analysis (ideally everywhere, often, and instant). Achieving increased sampling density in space and time at a sufficiently low cost, and with results on sufficiently fast timelines, implies a future with substantial increases in automation at all steps of data acquisition and the extent of decentralization of nucleic acid sequencing. A key component of a future system is therefore field-deployable nucleic acid sequencing machines that perform sample collection, sample preparation, and sequencing autonomously at an extremely low cost. To enable this scale of

ubiquity, data formats, information protocols, and downstream analysis are carefully designed to derive maximum insight from the integration of these diverse data streams while protecting against abuse of data collection at such an intensive scale.

## Store

Data and metadata collected by the distributed sequencing network are at least partly public, but systems also account for societal preferences regarding privacy. Maintaining privacy is necessary to earn social licensing and trust in storing potentially identifying information. Data from these systems are housed in publicly funded repositories and are used for both real-time monitoring and research. These systems provide sufficiently granular monitoring to afford transformational biosecurity benefits, and sufficiently strong privacy protections via a combination of (1) high levels of information secured by operational security and/or (2) statistically or cryptographically deidentified public representations of data streams for monitoring activities that guarantee individual privacy.[45-47]

## Reporting and Usefulness

The distribution and prevalence of diseases is routinely reported in a standardized fashion. Worrying clusters, mutations, and novel crossovers are flagged to both local public health officials and international infectious disease monitoring organizations. The timeliness, sensitivity, and specificity of monitoring approaches are well characterized and a range of candidate threat profiles have been identified, enabling well-calibrated, predetermined but flexible response plans to be activated quickly and with accountability. Funding for the system is supported politically as cost-effective medical infrastructure and as a crucial global warning and response system.

## Feasibility

The expansive, yet anticipated, future system will be impossible to realize with current technologies and systems. Current policy and legal structures are also insufficient. In the coming years, any advances in this direction will involve restrictive tradeoffs between coverage, depth, cost, usefulness, and privacy. Fortunately in the longer term, it seems possible to find solutions that make minimal compromises on each front.

With almost certain decreases in sequencing costs and increases in compute capabilities, both real and perceived threats to genomic privacy are potential limiting factors. To address this concern, statistical or cryptographic privacy in data collection is important, as gathering will not be carried out by a single entity. Additionally, privacy-preserving representations can lower barriers to data sharing between entities. Compressed, privacy-preserving representations of sequence data may also limit "information hazards" associated with gaining a deeper understanding of natural variation in the genomes of environmental organisms.[48]

WMGM depends on numerous technological advances, changes to policy, and new systems. None of these are simple, but some of them, or at least their direct antecedents, are already being built.

## WAY POINTS AND OBSTACLES IN A TRANSITION

Near-term applications of metagenomic sequencing in biological monitoring foreshadow longer-term futures for wider deployment. Crucially, near-term proposals differ significantly in how sampling is accomplished, but also in the read length and speed of the underlying sequencing technology.

For example, Shean and Greninger[49] propose a near-term future resting on widespread deployment of clinical sampling. In their vision, metagenomic sequencing has increased analytic sensitivity (achieved through deeper sequencing—sequencing a higher proportion of the nucleic acid molecules in the sample, more slowly) such that data can be used reliably and cost-effectively for diagnosis of infectious disease and determination of antimicrobial sensitivity. They also suggest using these methods for outbreak clustering and transmission tracking. In this vision, it seems that some preliminary analysis at least, will occur "on machine" (ie, automatically and locally). Ideally, this would be expanded to collecting more than just the immediately clinically relevant data to increase the usefulness of each data point. This increase in metadata would require trustworthy privacy mechanisms that allow for use of the data for WMGM and advantages for reservoir and other monitoring.

Another near-term possibility is the Nucleic Acid Observatory,[44] which proposed ongoing wastewater and watershed sampling across the United States to find sequences that recently emerged or are increasing in frequency, indicating a potential new pathogen or other notable events.

## Critical Technological Advances

For a technical review of current metagenomic techniques and their use in biosurveillance, refer to Ko et al[1] and Simner et al.[3] In this article, we include both metatranscriptomics (analyzing collective RNA transcriptomes, specifically) and viral metagenomics in the definition of metagenomic sequencing. While current surveillance efforts focus on culture- or PCR-based methods, recent advances have been made in using metagenomics for the surveillance of both viruses and microbes, and sequencing both DNA and RNA. Metagenomics, but more so metatranscriptomics, are still limited, especially in terms of the extraction techniques. Such techniques are different depending on the organisms expected to be in the sample, but especially in terms of analysis because assembled metagenomes are still highly fragmented and especially difficult to compare using current algorithms. Table 3 outlines possible critical advances, including extraction protocol, genome assembly and characterization, and privacy and storage considerations. We propose that efforts on the technologies outlined, in addition to identifying other technologies, will contribute to a WMGM future.

## Policy Planning Under Technological Uncertainty

It is possible that the most valuable policy in the long term is to do more technical research. However, the extent to which this might be true can be evaluated in discussions such as this article. Whether a metagenomics sequencing-based biosurveillance is technically viable will become apparent in the coming years, but much more neglected is the consideration of policy and systemic concerns. We identify 4 current drawbacks that, if unaddressed, will delay or prevent the implementation of a system to detect pandemic risks. Unfortunately, changes in policy move more slowly than advances in technology, which too often leads to both slow adoption and locking in subpar methods—for example, if PCR tests are adopted as a standard or if clinical guidelines indicate that sequencing is appropriate only after other testing is performed. Even after sequencing is demonstrated to be comparably inexpensive and rapid, it might remain reserved for unusual cases. For reasons such as these, it is crucial to flag the needs of future systems and current drawbacks now.

### Suboptimal Use and High Prices

The value of metagenomic sequencing will be limited if publicly beneficial uses of metagenomic monitoring are impossible due to patents, the collection of nonpublic data, or a lack of academic and clinical incentive to participate in broadly beneficial applications. A metagenomic monitoring

Table 3. Critical Technological Advances

| Technology Type | Desired Characteristics | References |
|---|---|---|
| Nucleic acid extraction protocols | Methods can be optimized for both RNA and DNA extracted from different organism and subcompartments, as well as viral particles, unlike current methods that require a different protocol for each scenario | 50-55 |
| Enzymes | Technologies with higher efficiency and lower cost for higher yield and purity, as well as a higher range of nucleic acids, reducing loss due to mechanical extraction techniques | 56-61 |
| Automation in sampling and nucleic acid extraction | Limits between-sampling bias, which is currently ubiquitous, and will help streamline the process and reduce time and costs | 62,63 |
| Short-read sequencing | Development of successive generations of sequencing machines | 64-67 |
| Short-read assembly algorithms | More accurate and informative assembly | 68-71 |
| Long-read sequencing | These technologies are experiencing more rapid improvements than short-read sequencing, so investing in their improvement will likely have higher payoff in the longer term | 72-80 |
| | While currently less suitable for metagenomics, portable real-time devices, like the Oxford Nanopore devices, can be deployed in real time for environmental samples | 75,81 |
| | High-fidelity long reads will reduce the need for sophisticated algorithms to reconstruct original sequences (for detection of novel and unknown pathogens) | 78,82 |
| Analysis of long- and short-read sequencing | Improved error correction | 83-86 |
| Filtering of metagenomic data and identifying potentially pathogenic sequences | Improved comparison between metagenomes | 70,87,88 |
| | Improved binning for metagenomics | 89-92 |
| | Improved prediction of gene function | 93,94 |
| | New methods will allow mobilization and multiplication of DNA segments | 95-98 |
| | Machine learning will allow improved error correction and analysis automation | 99,100 |
| Storage and data processing | The costs of computation and storage are already significantly lower than they have been historically, but costs should decline at least enough to make the promises of metagenomic sequencing realistic | 101,102 |
| | Development of interoperable data formats (consistent metadata is important for epidemiological analysis) | 103 |
| Secure computing | Privacy preserving methods of analysis (eg, federated learning, an approach that enables public analysis of private data and does not explore raw sequencing data will be used) | 104,105 |
| | Storing and sharing of data will be secured (eg, secure cloud, encrypted data) | 106,107 |

system could fail to be adopted if providing or accessing data is overly unattractive or difficult. To reduce the likelihood of private capture and fragmentation of data, one option is to accelerate data publication and provision of data at the earliest point, so that clinicians and scientists provide public data as immediately as possible. Because academic incentives push against prepublication data dissemination, and commercial incentives push for closed information systems, policymakers should promote immediate data availability.

Burdens of system change also have the potential to delay adoption, which can result in the paradigm of using first

paper tests, then PCR or culture-based tests, and only then using metagenomic sequencing. Varied sources will remain critical in the coming decade, but as metagenomic sequencing declines in price enough to be negligible, it should supplant PCR, lateral flow, or other testing, not just supplement them.

An issue related to suboptimal use is costs. The idiosyncratic nature of the US health system will pose additional challenges related to reimbursements and universal clinical access, but even internationally there will be challenges. Capital investment in biosurveillance may be difficult, especially in lower-income countries and less well-

populated areas, but international subsidies, tax credits, and other incentive schemes may be useful. Because metagenomic monitoring is particularly important in areas that otherwise may have less access to care, addressing disparities in access to both clinical and biosurveillance sequencing will be essential.

### Privacy and Data Abuse

Abuse of private genetic information is likely to occur as the use of sequencing proliferates, whether via metagenomic sequencing or otherwise. For this reason, near-term focus on addressing privacy concerns is crucial. These concerns will become increasingly salient as metagenomic monitoring becomes ubiquitous. Metagenomic samples may initially include human DNA, which is clearly identifiable, but even removed microbiotic signatures are potentially personally identifiable.[108] It is unclear if there are legal restrictions on analysis of sewage and similar sources, but the data are potentially predictive of otherwise personal information, so public discussion of privacy tradeoffs and preventing misuse is important.

Technological privacy solutions need to be adapted to the specific goals of metagenomic monitoring and need standardization to enable usage. Legal structures that allow for public use of healthcare data, as well as policy approaches for developing standards and encouraging or mandating compliance and data sharing, will be essential.

Closely related to the concern about abuse of personal data is the concern that wider availability of genomic data could affect biosecurity. However, it is unclear if widespread monitoring significantly increases availability of these data compared with other applications of already increasingly available metagenomic technologies.

Beyond these legitimate concerns, new technologies are often the subject of suspicion and misinformation. Clear public rules and enforcement can help address public mistrust. The messaging about the promise of such technology, and the rules to prevent misuse, should be emphasized earlier rather than later. An example is the Health Insurance Portability and Accountability Act of 1996,[109] which is seen as too restrictive, and likely as a result, few claims of misuse of medical data in the United States have emerged. At the same time, fully private data would not allow for infectious disease monitoring, so the specific approach is unlikely to be viable. For this reason, initial deployments in countries with higher institutional trust may be preferred. Alternatively, less privacy-invasive initial use cases, such as metagenomic analysis of wastewater, might circumvent many concerns.

### Peacetime Usefulness

The most important applications for health security are preventing and responding to crises; however, systems that are useful only during a crisis are likely to lack funding or be unavailable when a crisis occurs.[35,110] These challenges are compounded for new technologies. For this reason, it is vital to ensure that metagenomic technologies are used even when there is no crisis. Thankfully, studies show a wide variety of nascent uses, including early cancer screening and precision medicine.[111-114]

Clinical applications for diagnosis of infections are crucial for identifying pathogens based on symptoms, ruling out the possibility of infection as a cause, and identifying antibiotic resistances in a given bacterial infection, as well as contact tracing and identifying routes of transmission. At present, metagenomic sequencing is rarely used for any of these, which could change if adoption increases. Advancing new uses depends on the availability of sequencing, their speed, and their reliability. Changing clinical practice is also challenging, and in addition to demonstrable advantages, care should be taken to ensure the new systems will benefit clinicians—and that cost differences are minimized or compensated for. Similarly, building transmission tracing systems for routine outbreaks can provide continuing value to public health workers.

Metagenomic analysis of wastewater is also valuable for routine public health monitoring. Benefits may include the identification of variants of seasonal influenza or SARS-CoV-2 or the locations of foodborne pathogen outbreaks, potentially even before clinical detection. Emphasizing these routine benefits will help ensure that the systems are maintained.

Despite the seeming convergence of interests between public health and metagenomic sequencing, identifying places where the routine uses of sequencing are misaligned with crisis uses is also critical to ensure systems are not built myopically. For example, short-term uses of pathogen metagenomics focused on viral metagenomic sequencing might compete with more valuable technology, such as parallel host transcriptomics, which can maximize clinical information per sample.

### Enabling Crisis Response

The value of metagenomic sequencing for biosurveillance lies in its ability to function as a warning system for imminent outbreaks. However, a warning is useful only if it enables a response. Government willingness to respond requires a trustworthy warning signal across the possible scenarios where response is needed. The value of such a response depends on the speed of warning signals and subsequent interventions. Building a system that provides alerts within hours instead of days is superfluous if a response take weeks. Similarly, the value of a system is limited by the accuracy of the warning signal. False alarms both reduce willingness to respond and make the system more expensive to use.

Tradeoffs may exist in such a system. For example, distributed analysis allows for faster independent confirmation of an incipient outbreak but may lead to more false positives. Similarly, nonpublic government analysis could be more inclusive of otherwise private data but may be

less trusted. In either case, governments need to plan responses when a warning signal is detected, regardless of the source.

Key questions about the use of metagenomic sequencing for biosurveillance extend far beyond the scope of this article. One of the most vital questions is how analysis will be translated into policy, and by whom. Governments, nongovernmental organizations, multilateral institutions, and academics are all capable of analysis, but their analyses will enable different types of response. Different response options will require different types of interactions between the systems and government and international planning. And perhaps most important, different funding models for the biosurveillance systems may be needed. An excellent system that is then defunded is far worse than a modest system that can be maintained.

### Solutions Seeking Implementations

In this decade, we expect that at a minimum the challenges described earlier in the sections on suboptimal use and high prices, as well as enabling crisis response, need to be substantively addressed. This will enable a system as integrated and encompassing as we describe to have an adequate foundation.

If buy-in from clinicians, hospitals, government bodies, academics, and other data gatherers is insufficient and if data from metagenomic sequencing are not readily available or easily stored and analyzed, we will never move past the current paradigm of disconnected biosurveillance systems. A WMGM future must also have enough financial, political, and operational support to make the transition.

Historically, supporting a biosurveillance system across many groups has often required that an institution or protocol manages the sharing and storing of data. Entities of comparable scale are the US National Center for Biotechnology Information or European Bioinformatics Institute, which require primary research authors to upload their genomic data. However, before a government-run data-sharing system is commonplace and regulated, a third party may need to be developed.

### CONCLUSION

In the coming years, next-generation sequencing and more widespread use of clinical and environmental sequencing for biosurveillance are likely. Drawbacks of such biosurveillance systems include data privacy and long-term viability of funding, which are unlikely to be fully remedied before deployment and require further attention. The question we address in this article is what policy issues are likely to arise in the coming years.

Among the most important policy issues are market failures, in several forms—for example, private capture of the market in ways that make widespread use expensive or that fragment the data and analysis ecosystem, as well as potential abuse of data and privacy concerns. Relatedly, it is possible that the system could become economically nonviable during nonpandemic times and funding is lost. Important in a different way is how planning and response activities are able to capitalize on these systems and data.

To address these questions and concerns, a variety of projects seem useful, and are best led by different groups. Until various price levels and market penetration are reached, research or expert forecasting of timelines is useful for planning. Building public or interoperable data systems and standards will be important for industry groups and government or nongovernment agencies. Policy planning to ensure that payment systems or regulations do not lock in current or near-term technologies is also needed. Of course, none of these will supplant the technological advances that are needed, but each will help unlock their potential. The most important tasks, however, must be started now, because if problems are addressed post-hoc instead of preemptively, much of the biosecurity potential of next-generation sequencing will be unnecessarily delayed, or lost.

### ACKNOWLEDGMENTS

### REFERENCES

1. Ko KKK, Chng KR, Nagarajan N. Metagenomics-enabled microbial surveillance. *Nat Microbiol*. 2022;7:486-496.
2. Carleton HA, Besser J, Williams-Newkirk J, Huang A, Trees E, Gerner-Smidt P. Metagenomic approaches for public health surveillance of foodborne infections: opportunities and challenges. *Foodborne Pathog Dis*. 2019;16(7): 474-479.
3. Simner PJ, Miller S, Carroll KC. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin Infect Dis*. 2018;66(5):778-788.
4. Gibbs EPJ. The evolution of One Health: a decade of progress and challenges for the future. *Vet Rec*. 2014;174(4): 85-91.
5. Homeland Security. Detecting bioterrorism attacks. Updated July 14, 2020. Accessed June 15, 2022. https://www.dhs.gov/biowatch-program
6. US Department of Agriculture Animal and Plant Health Inspection Service. National Animal Health Monitoring System. Updated August 18, 2022. Accessed September 16, 2022. https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/monitoring-and-surveillance/nahms
7. Burkom H, Loschen W, Wojcik R, et al. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE): overview, components, and public health applications. *JMIR Public Health Surveill*. 2021;7(6):e26303.

8. African Union and Africa Centers for Disease Control and Prevention (CDC). Establishing the Africa Centres for Disease Control and Prevention: the upside of a crisis. *Africa Health*. May 2017. Accessed September 16, 2022. https://africacdc.org/download/establishing-the-africa-centres-for-disease-control-and-prevention-the-upside-of-a-crisis/

9. World Health Organization. WHO Hub for Pandemic and Epidemic Intelligence. Accessed September 16, 2022. https://pandemichub.who.int/

10. GitHub. chanzuckerberg/czid-web. Accessed September 16, 2022. https://github.com/chanzuckerberg/czid-web

11. Feng L, Zhang W, Li X. Monitoring of regional drug abuse through wastewater-based epidemiology—a critical review. *Sci China Earth Sci*. 2018;61:239-255.

12. van Nuijs ALN, Castiglioni S, Tarcomnicu I, et al. Illicit drug consumption estimations derived from wastewater analysis: a critical review. *Sci Total Environ*. 2011;409(19):3564-3577.

13. Choi PM, Tscharke BJ, Donner E, et al. Wastewater-based epidemiology biomarkers: past, present and future. *Trends Analyt Chem*. 2018;105:453-469.

14. McCall A-K, Bade R, Kinyua J, et al. Critical review on the stability of illicit drugs in sewers and wastewater samples. *Water Res*. 2016;88:933-947.

15. Zarei S, Salimi Y, Repo E, et al. A global systematic review and meta-analysis on illicit drug consumption rate through wastewater-based epidemiology. *Environ Sci Pollut Res Int*. 2020;27(29):36037-36051.

16. Grabow WO, Botma KL, Villiers JC de, Clay CG, Erasmus B. Assessment of cell culture and polymerase chain reaction procedures for the detection of polioviruses in wastewater. *Bull World Health Organ*. 1999;77(12):973-980.

17. Sinclair RG, Choi CY, Riley MR, Gerba CP. Pathogen surveillance through monitoring of sewer systems. *Adv Appl Microbiol*. 2008;65:249-269.

18. Ahmed W, Angel N, Edson J, et al. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. *Sci Total Environ*. 2020;728:138764.

19. Ho J, Stange C, Suhrborg R, Wurzbacher C, Drewes JE, Tiehm A. SARS-CoV-2 wastewater surveillance in Germany: long-term PCR monitoring, suitability of primer/probe combinations and biomarker stability. *Water Res*. 2022;210:117977.

20. Haramoto E, Malla B, Thakali O, Kitajima M. First environmental surveillance for the presence of SARS-CoV-2 RNA in wastewater and river water in Japan. *Sci Total Environ*. 2020;737:140405.

21. McClary-Gutierrez JS, Aanderud ZT, Al-Faliti M, et al. Standardizing data reporting in the research community to enhance the utility of open data for SARS-CoV-2 wastewater surveillance. *Environ Sci (Camb)*. 2021;7:1545-1551.

22. Reeves K, Leibig J, Feula A, et al. High-resolution within-sewer SARS-CoV-2 surveillance facilitates informed intervention. *Water Res*. 2021;204:117613.

23. Derda R, Gitaka J, Klapperich CM, et al. Enabling the development and deployment of next generation point-of-care diagnostics. *PLoS Negl Trop Dis*. 2015;9(5):e0003676.

24. Toskin I, Blondeel K, Peeling RW, Deal C, Kiarie J. Advancing point of care diagnostics for the control and prevention of STIs: the way forward. *Sex Transm Infect*. 2017;93(suppl 4):S81-S88.

25. Quicke K, Gallichotte E, Sexton N, et al. Longitudinal surveillance for SARS-CoV-2 RNA among asymptomatic staff in five Colorado skilled nursing facilities: epidemiologic, virologic and sequence analysis. Preprint. *medRxiv*. Posted November 5, 2020. Accessed September 16, 2022. doi:10.1101/2020.06.08.20125989

26. Gallichotte EN, Nehring M, Young MC, et al. Durable antibody responses in staff at two long-term care facilities, during and post SARS-CoV-2 outbreaks. *Microbiol Spectr*. 2021;9(1):e00224-21.

27. Wood JLN, Leach M, Waldman L, et al. A framework for the study of zoonotic disease emergence and its drivers: spillover of bat pathogens as a case study. *Philos Trans R Soc Lond B Biol Sci*. 2012;367(1604):2881-2892.

28. Magouras I, Brookes VJ, Jori F, Martin A, Pfeiffer DU, Dürr S. Emerging zoonotic diseases: should we rethink the animal-human interface? *Front Vet Sci*. 2020;7:582743.

29. Graham JP, Leibler JH, Price LB, et al. The animal-human interface and infectious disease in industrial food animal production: rethinking biosecurity and biocontainment. *Public Health Rep*. 2008;123(3):282-299.

30. Jones BA, Grace D, Kock R, et al. Zoonosis emergence linked to agricultural intensification and environmental change. *Proc Natl Acad Sci U S A*. 2013;110(21):8399-8404.

31. Kracalik I, Malania L, Imnadze P, Blackburn JK. Human anthrax transmission at the urban–rural interface, Georgia. *Am J Trop Med Hyg*. 2015;93(6):1156-1159.

32. Makita K, Fèvre EM, Waiswa C, Kaboyo W, Eisler MC, Welburn SC. Evidence-based identification of the most important livestock related zoonotic diseases in Kampala, Uganda. *J Vet Med Sci*. 2011;73(8):991-1000.

33. Bajema NE, Beaver W, Parthemore C, et al. *Toward a Global Pathogen Early Warning System: Building on the Landscape of Biosurveillance Today*. Washington, DC: The Janne E. Nolan Center on Strategic Weapons; 2021. Accessed September 16, 2022. https://councilonstrategicrisks.org/wp-content/uploads/2021/07/Toward-A-Global-Pathogen-Early-Warning-System_2021_07_20-1.pdf

34. Bipartisan Commission on Biodefense. *Biodefense in Crisis: Immediate Action Needed to Address National Vulnerabilities*. Washington, DC: Bipartisan Commission on Biodefense; 2021. Accessed September 16, 2022. https://biodefense commission.org/reports/biodefense-in-crisis-immediate-action-needed-to-address-national-vulnerabilities/

35. Bipartisan Commission on Biodefense. *Diagnostics for Biodefense - Flying Blind with No Plan to Land*. Washington, DC: Bipartisan Commission on Biodefense; 2020. Accessed September 16, 2022. https://biodefensecommission.org/reports/diagnostics-for-biodefense-flying-blind-with-no-plan-to-land/

36. Morel B, Barbera P, Czech L, et al. Phylogenetic analysis of SARS-CoV-2 data is difficult. *Mol Biol Evol*. 2021;38(5):1777-1791.

37. Rao S, Cox T, Rao P. Computational challenges in genomic data analyses. Accessed December 12, 2022. https://tigerprints.clemson.edu/cgi/viewcontent.cgi?article=1008&context=hugedata

38. Chretien J-P, Rivers CM, Johansson MA. Make data sharing routine to prepare for public health emergencies. *PLoS Med.* 2016;13(8):e1002109.

39. Pollett S, Johansson M, Biggerstaff M, et al. Identification and evaluation of epidemic prediction and forecasting reporting guidelines: a systematic review and a call for action. *Epidemics.* 2020;33:100400.

40. Manheim D, Chamberlin M, Osoba OA, Vardavas R, Moore M. *Improving Decision Support for Infectious Disease Prevention and Control: Aligning Models and Other Tools with Policymakers' Needs.* Santa Monica, CA: Rand Corporation; 2016. https://www.rand.org/pubs/research_reports/RR1576.html

41. Molecular Reality Inc. Accessed September 16, 2022. https://www.molecularreality.com/

42. Piombo E, Abdelfattah A, Droby S, Wisniewski M, Spadaro D, Schena L. Metagenomics approaches for the detection and surveillance of emerging and recurrent plant pathogens. *Microorganisms.* 2021;9(1):188.

43. Shen J, McFarland AG, Young VB, et al. Toward accurate and robust environmental surveillance using metagenomics. *Front Genet.* 2021;12:600111.

44. Nucleic Acid Observatory Consortium. A global nucleic acid observatory for biodefense and planetary health. Preprint. arXiv:2108.02678 [q-bio.GN]. Submitted August 5, 2021. Accessed September 5, 2021. http://arxiv.org/abs/2108.02678

45. Baum C, Cui H, Damgård I, et al. Cryptographic aspects of DNA screening. SecureDNA. Published January 2020. Accessed September 16, 2022. https://www.securedna.org/download/Cryptographic_Aspects_of_DNA_Screening.pdf

46. Zhou B, Han Y, Pei J, Jiang B, Tao Y, Jia Y. Continuous privacy preserving publishing of data streams. Paper presented at: Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology; March 24-26, 2009; Saint Petersburg, Russia.

47. Pervaiz Z, Ghafoor A, Aref WG. Precision-bounded access control using sliding-window query views for privacy-preserving data streams. *IEEE.* 2015;27(7):1992-2004.

48. Lewis G, Millett P, Sandberg A, Snyder-Beattie A, Gronvall G. Information hazards in biotechnology. *Risk Anal.* 2019;39(5):975-981.

49. Shean RC, Greninger AL. One future of clinical metagenomic sequencing for infectious diseases. *Expert Rev Mol Diagn.* 2019;19(10):849-851.

50. Fremin BJ, Sberro H, Bhatt AS. MetaRibo-Seq measures translation in microbiomes. *Nat Commun.* 2020;11(1):3268.

51. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med.* 2018;50(8):1-14.

52. Blattman SB, Jiang W, Oikonomou P, Tavazoie S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat Microbiol.* 2020;5(10):1192-1201.

53. Dar D, Dar N, Cai L, Newman DK. In situ single-cell activities of microbial populations revealed by spatial transcriptomics. Preprint. *bioRxiv.* Posted February 25, 2021. Accessed September 16, 2022. https://doi.org/10.1101/2021.02.24.432792

54. Diebold PJ, New FF, Hovan M, Satlin MJ, Brito IL. Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR. *Elife.* 2021;10:e66834

55. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One.* 2015;10(5):e0128036.

56. Maghini DG, Moss EL, Vance SE, Bhatt AS. Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. *Nat Protoc.* 2021;16(1):458-471.

57. Coelho D, Lopes PA, Cardoso V, et al. Novel combination of feed enzymes to improve the degradation of *Chlorella vulgaris* recalcitrant cell wall. *Sci Rep.* 2019;9(1):5382.

58. Easparro BA, Garrett S, Atwood J. Are enzymes required for the extraction and isolation of DNA in conjunction with mechanical homogenization? *FASEB J.* 2016;30(1):1082.2.

59. Giovannoni M, Gramegna G, Benedetti M, Mattei B. Industrial use of cell wall degrading enzymes: the fine line between production strategy and economic feasibility. *Front Bioeng Biotechnol.* 2020;8:356.

60. Zou Y, Mason MG, Wang Y, et al. Nucleic acid purification from plants, animals and microbes in under 30 seconds. *PLoS Biol.* 2017;15(11):e2003916.

61. Luong T, Salabarria A-C, Edwards RA, Roach DR. Standardized bacteriophage purification for personalized phage therapy. *Nat Protoc.* 2020;15(9):2867-2890.

62. McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. *Elife.* 2019;8:e46923.

63. Manen J-F, Sinitsyna O, Aeschbach L, Markov AV, Sinitsyn A. A fully automatable enzymatic method for DNA extraction from plant tissues. *BMC Plant Biol.* 2005;5:23.

64. Na HS, Yu Y, Kim SY, Lee J-H, Chung J. Comparison of the performance of MiSeq and HiSeq 2500 in a microbiome study. *Microbiol Biotechnol Lett.* 2020;48(4):574-581.

65. Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 2012;13:341.

66. Ravi RK, Walton K, Khosroheidari M. MiSeq: A next generation sequencing platform for genomic analysis. In: DiStefano JK, editor. *Disease Gene Identification: Methods and Protocols.* New York: Springer; 2018:223-232.

67. Modi A, Vai S, Caramelli D, Lari M. The Illumina sequencing protocol and the NovaSeq 6000 system. In: Mengoni A, Bacci G, Fondi M, eds. *Bacterial Pangenomics: Methods and Protocols.* New York: Springer; 2021:15-42.

68. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics.* 2015;31(10):1674-1676.

69. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 2012;40(20):e155.

70. Coleman I, Korem T. Embracing metagenomic complexity with a genome-free approach. *mSystems.* 2021;6(4):e0081621.

71. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 2017;27(5):824-834.

72. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics*. 2016;14(5):265-279.

73. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*. 2018;36(4):338-345.

74. Liu Q, Wu H, Wu L, et al. Voltage-driven translocation of DNA through a high throughput conical solid-state nanopore. *PLoS One*. 2012;7(9):e46014.

75. Loman NJ, Watson M. Successful test launch for nanopore sequencing. *Nat Methods*. 2015;12(4):303-304.

76. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol*. 2020;38(6):701-707.

77. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*. 2015;13(5):278-289.

78. Hon T, Mars K, Young G, et al. Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci Data*. 2020;7(1):399.

79. Wenger AM, Peluso P, Rowell WJ, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol*. 2019; 37(10):1155-1162.

80. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*. 2020;21(1):30.

81. Castro-Wallace SL, Chiu CY, John KK, et al. Nanopore DNA sequencing and genome assembly on the International Space Station. *Sci Rep*. 2017;7(1):18022.

82. Lang D, Zhang S, Ren P, et al. Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore. *Gigascience*. 2020;9(12):giaa123.

83. Kolmogorov M, Bickhart DM, Behsaz B, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods*. 2020;17(11):1103-1110.

84. Salmela L, Rivals E. LoRDEC: accurate and efficient long read error correction. *Bioinformatics*. 2014;30(24):3506-3514.

85. Zhang H, Jain C, Aluru S. A comprehensive evaluation of long read error correction methods. *BMC Genomics*. 2020; 21(suppl 6):889.

86. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540-546.

87. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through dereplication. *ISME J*. 2017;11(12):2864-2868.

88. Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G. De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nat Genet*. 2012;44(2):226-232.

89. Kang DD, Li F, Kirton E, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. 2019;7:e7359.

90. Lin H-H, Liao Y-C. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. *Sci Rep*. 2016; 6:24175.

91. Maguire F, Jia B, Gray KL, Lau WYV, Beiko RG, Brinkman FSL. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic islands. *Microbial Genomics*. 2020;6(10):mgen 000436.

92. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 2016;32(4): 605-607.

93. Zeevi D, Korem T, Godneva A, et al. Structural variation in the gut microbiome associates with host health. *Nature*. 2019;568(7750):43-48.

94. Olekhnovich EI, Vasilyev AT, Ulyantsev VI, Kostryukova ES, Tyakht AV. MetaCherchant: analyzing genomic context of antibiotic resistance genes in gut microbiota. *Bioinformatics*. 2018;34(3):434-444.

95. Kent AG, Vill AC, Shi Q, Satlin MJ, Brito IL. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nat Commun*. 2020;11(1):4379.

96. Bertrand D, Shaw J, Kalathiyappan M, et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat Biotechnol*. 2019;37(8):937-944.

97. Durrant MG, Li MM, Siranosian BA, Montgomery SB, Bhatt AS. A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe*. 2020;27(1):140-153.e9. Published correction appears in *Cell Host Microbe*. 2020; 28(5):767.

98. Suzuki Y, Nishijima S, Furuta Y, et al. Long-read metagenomic exploration of extrachromosomal mobile genetic elements in the human gut. *Microbiome*. 2019;7(1):119.

99. LaPierre N, Ju CJ-T, Zhou G, Wang W. MetaPheno: a critical evaluation of deep learning and machine learning in metagenome-based disease prediction. *Methods*. 2019;166: 74-82.

100. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput Biol*. 2016; 12(7):e1004977.

101. Yang C, Chowdhury D, Zhang Z, et al. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. *Comput Struct Biotechnol J*. 2021;19:6301-6314.

102. Krumm N, Hoffman N. Practical estimation of cloud storage costs for clinical genomic data. *Pract Lab Med*. 2020;21:e00168.

103. Ten Hoopen P, Finn RD, Bongo LA, et al. The metagenomic data life-cycle: standards and best practices. *Gigascience*. 2017;6(8):1-11.

104. Gardiyawasam Pussewalage HS, Oleshchuk VA. Privacy preserving mechanisms for enforcing security and privacy requirements in e-health solutions. *Int J Inf Manage*. 2016; 36(6):1161-1173.

105. Yin X, Zhu Y, Hu J. A comprehensive survey of privacy-preserving federated learning: a taxonomy, review, and future directions. *ACM Comput Surv*. 2021;54(6):131.

106. Mohammed Yakubu A, Chen Y-PP. Ensuring privacy and security of genomic data and functionalities. *Brief Bioinform*. 2020;21(2):511-526.

107. Carter AB. Considerations for genomic data privacy and security when working in the cloud. *J Mol Diagn*. 2019; 21(4):542-552.

108. Hampton-Marcell JT, Lopez JV, Gilbert JA. The human microbiome: an emerging tool in forensics. *Microb Biotechnol.* 2017;10(2):228-230.

109. Health Insurance Portability and Accountability Act. Pub. L. No. 104-191 (1996). https://www.govinfo.gov/app/details/PLAW-104publ191

110. *Biosecurity for the Future: Strengthening Deterrence and Detection, Hearing Before the Subcommittee on Asia, the Pacific, Central Asia, and Nonproliferation of the House Committee on Foreign Affairs*, 117th Cong, 1st Sess (2021) (testimony of Amesh Adalja, MD, senior scholar, Johns Hopkins Center for Health Security). Accessed January 31, 2022. https://docs.house.gov/meetings/FA/FA05/20211208/114290/HHRG-117-FA05-Wstate-AdaljaA-20211208.pdf

111. Cheng M, Cao L, Ning K. Microbiome big-data mining and applications using single-cell technologies and metagenomics approaches toward precision medicine. *Front Genet.* 2019;10:972.

112. Dixon M, Stefil M, McDonald M, et al. Metagenomics in diagnosis and improved targeted treatment of UTI. *World J Urol.* 2020;38(1):35-43.

113. Yu J, Feng Q, Wong SH, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut.* 2017;66(1):70-78.

114. Jiang Y, Guo X, Liu L, et al. Metagenomic characterization of lysine acetyltransferases in human cancer and their association with clinicopathologic features. *Cancer Sci.* 2020; 111(5):1829-1839.

Address correspondence to:
*David B. Manheim*
*8734 First Avenue*
*Silver Spring, MD 20910*

*Email:* davidmanheim@gmail.com