Managing the Transition to Widespread Metagenomic Monitoring

Policy considerations for future biosurveillance

Chelsea Liang[1], James Wagstaff[2], Virginia Schmit[3], Noga Aharony[4,5], David Manheim[5,3,6,*]

1 University of New South Wales, Australia
2 University of Oxford, Future of Humanity Institute, United Kingdom
3 1Day Sooner, Delaware, United States
4 Columbia University, New York, United States
5 Technion, Israel Institute of Technology, Israel
6 Association for Long Term Existence and Resilience (ALTER), Israel

Abstract

Despite extensive discussion of the technological possibilities and public health importance of metagenomic sequencing, there has been scant discussion of what policy and regulatory issues need to be addressed to realize the potential of metagenomic sequencing . Here we review the current state of biosurveillance, and point to several ways future metagenomic monitoring may replace currently limited infectious disease monitoring models. We suggest that while many key enablers are technological, others are not. We therefore highlight key policy challenges and implementation questions which need to be addressed in order for the world to reach a state of 'Widespread Metagenomic Monitoring'. We find that policymakers must address pitfalls like fragmentation of the technological base, private capture of benefits, privacy concerns, the usefulness of the system during non-pandemic 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.

Contribution statement: DM conceptualized and initiated the project, supervised, and led investigation and writing of policy issues. CL led funding acquisition, administration, writing, reviewing and editing. JW led investigation and writing on applications of metagenomic sequencing. VS led investigation and writing on the present state of biosurveillance and technologies, and NA led investigation and writing about newer and promising technologies.

We are grateful to Joshua Monrad and Bill Beaver for helpful comments, and to participants in the Future of Humanity Institute Biorisk E-Seminar program for feedback, ideas, and suggestions. Thank you to the Long Term Futures Fund for financial support.

* Corresponding Author

Introduction

The COVID-19 pandemic has drawn an unparallelled amount of funding and political will to the space of pandemic preparedness. As such, this is an opportune time to ensure that the interventions considered by institutions - intentional investments and technology adoption policy - are geared to prevent not just the next pandemic, but all future pandemics. Metagenomic sequencing, referring to methods which extract genetic sequence data from many or all organisms in a sample, offers a disease agnostic approach to monitoring, detecting, and characterizing pathogens and variants.

This paper seeks to illuminate potential technical, operational and policy obstacles to the establishment of a universal, scalable and pathogen-agnostic monitoring system. With the plummeting costs of next generation sequencing, information dense metagenomic nucleic acid diagnostics and monitoring are a growing opportunity. Despite the challenge of policy planning under uncertainty, strategic thinking can convert technological and policy uncertainties into specific and scoped questions which can then be tackled by the relevant academic, policy, and professional communities.

As with any transitional planning, there needs to be an identification of the starting point, the destination, and challenges in between. Accordingly, we start with an overview of the current state of biosurveillance. Next, we note that there are various possible future systems, and describe some possibilities and common challenges. We refer to this system which aims to detect infectious diseases and potential pandemic risks as “Widespread Metagenomic Monitoring” to distinguish itself from the current and often disjoint “Biosurveillance” efforts. Finally, we identify the following critical issues as needing policy solutions (1) suboptimal use or high prices, (2) privacy and concerns about abuse of monitoring, (3) ensuring ‘peacetime’ usefulness, and (4) 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

At present, biosurveillance is a mix of techniques and systems which are both heterogeneous and overlapping, not only between countries, but even within them. In the United States (US), a complex set of programs exists where state-level control over some biosurveillance activities competes with multiple national programs. Meanwhile in many developing countries, regional and global cooperation, often funded by international partners, is more common. Current approaches track a limited number of patients using tests specific to one disease and, for the most part, known diseases only.

Geographic Heterogeneity:

There is significant variation around the world in surveillance. In the US, not only does the Center for Disease Control (CDC) run several disease-specific and syndromic biosurveillance

programs, but the Department of Homeland Security's Office of Health Affairs runs both Biowatch and an integration center for biosurveillance. [1] Separate systems like the US Department of Agriculture's Animal and Plant Health Inspection Service (APHIS) are in place for agricultural and livestock disease monitoring. [2] These systems used by governments tend to have limited data sharing between each other, much less internationally. But even when open and widely used systems like ESSENCE are used, [3] public health officials more often flag outbreaks of notifiable diseases via doctor diagnoses, rather than via syndromic or other monitoring methods.

Contrast this with the comparatively difficult situation in the developing world, where many countries have at best partial coverage of the population for even 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 WHO when they are identified, and limited real-time analysis capacity exists - though the Africa CDC and others are starting to address this gap. [4,5] At the same time, these countries are collaborating with the use of open-source tools like IDSeq for analysis of metagenomic sequencing data. [6] The open nature of these systems allows faster analysis and increased operational resilience. We note a trend towards the increasing modularity of various nodes of biosurveillance.

A closer look into current paradigms

In the current paradigm, a given node that gathers disease data points is the same, or is highly coupled to the node that performs the analysis of such data. This siloing is often tied to a specific jurisdiction or method of data gathering as seen in the various US agencies mentioned above. The vertical integration of analysis into gathering has meant that it is at best awkward - and at worst impossible - to aggregate data from different nodes which could ultimately benefit a nation or the international community. We see this as a key issue in the current instantiation of global infectious disease monitoring. [7,8]

Wastewater Epidemiology:

Prior to the SARS-CoV-2 pandemic, wastewater epidemiology (WWE) was already in use, albeit primarily for monitoring of illicit drug use, [9–13] exposure to pesticides on a population level, [11] and some limited investigations of detection of poliovirus in South Africa. [14] Further, in the early use of monitoring for illicit drug use there was at least one 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. [11]

Even before COVID-19, it was recognized that WWE could be used more broadly for disease monitoring, [15] and during the SARS-CoV-2 pandemic it was used to monitor levels of this infectious organism both at the regional level in Australia, [16] Germany, [17] and Japan, [18] and at the more granular level of dormitory buildings or neighborhoods served by specific trunk lines. [19,20]

Diagnostics:

Point of care (POC) and point of use (POU) diagnostics using metagenomic sequencing are already available, albeit not ubiquitous in High Income Countries (HIC) [21,22]. Without comprehensive data on individuals testing positive for ‘disease X’ in an emergent epidemic or pandemic setting, real time epidemiological data is severely hampered.

During the SARS-CoV-2 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 finding positive cases and quarantining infected individuals than waiting for observable symptoms to trigger testing. [23,24] This further demonstrates the importance of testing in protecting communities, particularly in infectious diseases that demonstrate infectivity prior to symptom onset.

Reservoir biosurveillance:

APHIS currently monitors agricultural endeavors domestically in the US and collaborates internationally on monitoring of agricultural models. [2] Known zoonoses have been monitored in large scale agricultural settings. [25–28] Animal producing farms located at the intersection of the urban / rural divide can also be a place 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. [29,30]

Funding infrastructure and payment systems:

Funding for biosurveillance has always been uneven, with costs borne largely by developed countries, but is inconsistent even there. [7,31] Even in some developed countries, costs for tests even during a pandemic are often borne by consumers. The resulting implicit discrimination against poorer and neglected communities has important direct impacts - but also leads to insufficient data, as well as biases, that undermine surveillance efforts. The parts of surveillance which tend to maintain funding are for the lower risk and less critical issues like foodborne pathogens and rare reportable diseases, rather than robust infrastructure for detecting future outbreaks. And in developing countries, there is also a constant battle to maintain funding for surveillance systems, which can seem superfluous until they are critical.

Summarizing the current situation

Obvious shortcomings exist in almost every area of the system, including gathering, analyzing, storing, and reporting data. While many of them can be feasibly addressed, critical shortcomings include gaps in coverage, not only geographic, but temporal, with data being collected relatively infrequently [7], and without sufficient metadata. Epidemiological analysis is often cumbersome and manual, especially for more detailed data such as genomic data [32,33], and even aggregate and symptomatic data itself is not always widely shared or available. Reporting from individual sources is often slow and non-standardized, [34] especially internationally, and often does not feed into any unified analysis. The analysis which does occur is reported in haphazard ways, with no real standards. [35] 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 ad-hoc or made without understanding limitations of the analysis. [36]

Potential Metagenomic Monitoring Futures

While the technology for metagenomic monitoring is still being developed and improved, there seems to be consensus that there are worthwhile goals to achieve. Accurate long-term planning is challenging, even more so when a plan is predicated on major technological progress. Thus any vision for Widespread Metagenomic Monitoring (WMGM) must remain tentative and flexible. It is, however, possible to outline some of the concrete features we would hope to see in order to consider how to get there.

Gather

The WMGM system we envision collects data from many nucleic acid data sources in a coherent set of formats. Other sources are still sometimes available, but given the extent and rapidity of sequencing data, they are ancillary at most. 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 PCR, antigen or CRISPR based tests. Beyond clinical use for diagnosis, sampling and sequencing capacity encompasses high-risk ‘sentinel’ populations in addition to civic-minded volunteers [37], as well as agricultural and wilderness ecosystems, [38] built environments, [39] and urban wastewater, often at a neighborhood level. One potential instantiation is “Nucleic Acid Observatory” which would monitor wastewater and waterways. [40] The gathering of data by this highly decentralized network is routine and automated wherever possible, at high frequency such that temporal trends could be sensitively identified.

Analyze

Analysis would be possible in both a centralized and decentralized fashion. Local analysis would start with diagnostics in clinical settings that replace and supersede the current vertically integrated monitoring systems. In exchange for their participation, (gathering) nodes such as farmers and animal biosurveillance researchers receive insights into animal and ecosystem health, and public health officials from wastewater and health practitioners would get clinical insights more rapidly, cheaply, and of higher quality than they could generate themselves. Analysis of data streams is near-real time as the distributed sequencing network is spread across multiple credible analysis centers, including both academic, local government, and international efforts. These centers have strong links to political entities holding responsibility for public health and biosecurity. Making analysis separate from data gathering also encourages data standards and systems that allow for subsequent analysis and broader sharing.

Scale

All else being equal, if we seek to maximize the public health benefits of a WMGM program we will want to (1) maximize sampling density in space, time, and also in terms of sequencing depth, and (2) minimize the time between nucleic acid (NA) sampling and data analysis (everywhere, often, and instant, described in Point C). Achieving increased sampling density in space and time at sufficiently low cost, and with results on sufficiently fast timelines, implies a significant increase in (1) the degree of automation at all steps of data acquisition and (2) the extent of decentralization of NA sequencing. A key component of a future system is therefore the existence of field-deployable NA sequencing machines that perform sample collection, sample preparation and sequencing autonomously at extremely low cost. To enable this scale of ubiquity, data formats, information protocols and downstream analysis should be carefully designed to derive maximum insight from the integration of these diverse data streams.

Store

Data and metadata collected by the distributed sequencing network would be at least partly public, but will also account for societal preferences regarding privacy. This is necessary to ensure that there is social licensing and trust in storing this potentially identifying information. Data from these systems is housed in publicly funded repositories, and is used for both real-time monitoring and research. These systems would 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 de-identified public representations of data streams for monitoring activities which guarantee individual privacy. [41–43] Other data sources for WMGM, such as various sources of symptomatic monitoring, as well as indicators from other forms of data-analysis, are integrated. Analysis of data streams in both clinical and public health applications use a variety of publicly available and/or open source software, which allows for continually improving and diverse ecosystems of analysis and prediction for both clinical applications, as well as national and international public health early warning, and research.

Feasibility

Of course, the stipulated future system is impossible with current technologies and systems, but it helps illuminate the specific shortfalls which exist, and what may be needed to address them. 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; and with planning, achievable to adequately address these concerns. . It seems useful to explore what will and will not be possible, and ensure that when building systems, near-term limitations do not hamper the potential for longer term improvements.

With almost certain decreases in sequencing costs and increasing compute capabilities, real or even imagined privacy is a potential limiting factor for WMGM. Statistically or cryptographically ensured privacy in data collection is potentially attractive, as gathering is unlikely to be carried out by a single entity, especially on a global scale. Additionally, privacy-preserving representations will 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. [44]

Reporting and usefulness

The distribution and prevalence of diseases would be routinely reported in a standardized fashion. Thereby, worrying clusters, mutations, and novel crossovers alike would be flagged both to local public health officials and to international infectious disease monitoring organizations. The timeliness, sensitivity and specificity of monitoring approaches would be well characterized, and a range of candidate threat profiles identified such that a range of well-calibrated, predetermined but flexible response plans can be activated quickly, and with accountability.

The public is aware of and supports both the monitoring system - which has privacy safeguards in place - and the response. Funding for the system is sufficient, and widely supported politically as both a cost-effective part of medical infrastructure, and a critical global warning and response system.

Challenges and Concerns

The system described 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. The critical path for these technologies is complex, and will not be explored, but there are many specific challenges which are already foreseeable.

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 propose a near-term future resting on widespread deployment of clinical sampling. [45] 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. Shean and Greaninger 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,” i.e. 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 entail a need for trustworthy privacy mechanisms which allow use of the data for WMGM - and advances will provide advantages for reservoir and other monitoring.

Another near-term proposal, mentioned above, is the Nucleic Acid Observatory (NAO), [40] which proposed ongoing wastewater and watershed sampling across the United States to find sequences which recently emerged or are increasing in frequency indicating a potential new pathogen or other notable events. In either case, there are both technical improvements expected which will allow the large-scale feasibility advances, as well as policy issues which must be addressed.

Critical Technological Advances

Across metagenomic monitoring approaches, it will be critical to develop nucleic acid extraction protocols that are optimized for many nucleic acid types, unlike current, more limited methods. [46–51] Such protocols would benefit from the development of enzymes with a higher efficiency and lower cost than those that currently exist. This would lead to a higher yield and purity, as well as a higher range of nucleic acids, reducing loss due to mechanical extraction techniques. [52–57] Finally, the further development of automation in sampling and nucleic acid extraction will limit between-sampling bias, which is currently ubiquitous, and help streamline the process and reduce costs. [58,59]

For short read sequencing, improvements to the technology [60–63] and assembly algorithms [64–67] will be useful. Even so, long-read sequencing will likely be more critical in the longer term to improve these systems. [68–76] While long-read sequencing is currently more costly, portable real-time devices like the Oxford Nanopore exist, and the technology may allow for environmental deployment in the near future. [71,77] In the long term, high-fidelity long-read sequencing will reduce the need for sophisticated algorithms to reconstruct the original sequences. [74,78] This will especially aid the detection of novel and unknown sequences in samples. However, even for long-read sequencing, advances in error correction [79–82] and binning [83–86] will be critical. Finally, better analysis techniques that allow the comparison of metagenomes, [66,87,88] the identification of functional regions [89,90] and the mobilization and multiplication of DNA segments [91–94] will be key in filtering metagenomic data and identifying potentially pathogenic sequences. In addition to all of this progress, advances in machine learning seem poised to provide significant additional power to error correction and assist in automation of analysis.

There are numerous challenges with storage and data processing. A set of reads for a single sample can be hundreds of gigabytes, and sample storage and processing have historically been significant. [95] Thankfully, it seems that costs of computation and storage are already low, [96] and are poised to continue to decline at least long enough to make most of the current promises of metagenomic sequencing realistic. Speed of analysis is also continuously improving based on newer computational methods, and faster hardware.

Computation in general is unlikely to be a key barrier, but secure computing and privacy preserving technologies for analysis may be critical. Federated learning, [97] a machine learning specific approach which allows public analysis of private data, and future work is promising. [98]

Other approaches using cryptographic obfuscation are potential avenues as well. However, no amount of technology can solve the policy and legal problems, which will be addressed below.

The visions outlined here are far from incompatible in theory, but could easily become so in practice. We propose that the priorities for those seeking to maximize health security benefits from widespread metagenomic monitoring should be to identify technology and policy initiatives which will retain strategic flexibility and technology interoperability, while baking in security and privacy considerations. For example, WMGM might necessitate the streaming of data collected in the course of metagenomic sampling to remote, global analysis platforms. The development of interoperable data streaming formats which do not expose raw sequencing data could address a variety of problems.

Policy Planning under Technological Uncertainty

A critical step which must be taken now is planning for the future technologies. It is possible that the policy that would generate the most value in the long term would purely be a prompt to do more fundamental technology research. However the extent to which this might be true can be evaluated in discussions such as this paper. Unfortunately, policy moves much more slowly than technology, and too often, the world of public health waits for technology to be available before planning how to utilize it. This leads to both slow adoption, and lock-in of sub-par methods. For example, if PCR tests are adopted as a standard, or clinical guidelines indicate that sequencing is only appropriate after other testing is performed, even once it is comparable inexpensive and rapid, sequencing might remain reserved for unusual cases. For that reason, it is critical to flag the needs of future systems and likely roadblocks now.

Critical issues which have been identified as needing policy solutions include (1) suboptimal use or high prices, (2) privacy and concerns about abuse of WMGM, (3) ensuring ‘peacetime’ usefulness, and (4) enabling crisis response. In each case, policy remedies are plausible, and should be further explored.

Suboptimal Use and High Prices

A loss of much of the value possible for metagenomic sequencing is possible if publicly beneficial uses of metagenomic monitoring are impossible due to: patents, the collection of non-public data, or a lack of academic and clinical incentive to participate in broadly beneficial applications. To reduce the likelihood of private capture and fragmentation of data, it is important to push down data publication and provision of data to the earliest point, so that clinicians and scientists provide public data as close to immediately as possible. To the extent that academic incentives push for delayed publications, and commercial incentives push for walled gardens, policymakers must instead push for immediate data availability. Practically, this may mean that grant funding for the development of new technology should require open standards and interoperability. Academic incentives which allow for non-publication of data and code should be addressed, and everyone should be as public as possible with data.

On the other hand, overly burdensome rules have the potential to delay adoption. The potential of widespread monitoring will be lost if we stay locked into 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 eventually supplant, not supplement, PCR, lateral flow, or other testing.

A related issue is costs. The idiosyncratic nature of the US health systems will pose additional challenges related to reimbursements and universal access, but even internationally there will be critical challenges. Capital investment may be difficult, especially in developing countries and less well populated areas, but international subsides, tax credits, and other incentive schemes may be useful. Because metagenomic monitoring is particularly critical in areas which otherwise may have less access to care, addressing disparities in access will be essential.

Privacy and Data Abuse

A focus on addressing privacy concerns is critical for allowing metagenomic monitoring These concerns could become increasingly salient as metagenomic monitoring becomes ubiquitous. Metagenomic data potentially includes human DNA, which is clearly identifiable, but even if that is removed, microbiotic signatures are potentially personally identifiable [99]. While it is unclear that there are any legal restrictions on analysis of sewage and similar sources, the data is potentially predictive of otherwise personal information like disease status or susceptibility. For that reason, public acceptance of privacy tradeoffs and/or preventing misuse may be important.

Technological privacy solutions need to be adapted to the specific goals of metagenomic monitoring, and need standardization to allow usage. Legal structures to 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.

Beyond these legitimate concerns, new technologies are often the subject of suspicion and misinformation. Clear public rules and enforcement can help combat public mistrust. The messaging about the promise of such technology, and the rules to prevent misuse should be emphasized earlier rather than later. The example of the Health Insurance Portability and Accountability Act of 1996 is perhaps useful, since it 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 severely limit 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 like wastewater might circumvent many of these concerns.

“Peacetime” usefulness

While the most important applications for health security are likely preventing and responding to crises, systems which are only useful during a crisis are likely to be ignored, under- or

un-funded, and less useful when a crisis occurs. [31,100] This challenge is particularly critical where we need uptake of a new technology. For this reason, a critical enabler for metagenomic sequencing is ensuring that it is used when there is no crisis. Thankfully, there seem to be a wide variety of nascent uses for early cancer screening and other precision medicine. [101–104]

Clinical applications for diagnosis of infections is critical for identifying pathogens given symptoms, ruling out the possibility of infection as a cause, and identifying antibiotic resistances in a given bacterial infection. It is also useful for contact tracing and identifying routes of transmission. At present, metagenomic sequencing is rarely used for any of these, but that could change. Such a change requires both technological advances and clinical changes. Pushing for new uses depends on the viability of the systems, their speed, and their reliability. But changing clinical practice is often challenging, and in addition to demonstrable advantages, care should be taken to understand and explain how the new systems will benefit clinicians - and cost differences should be minimized or compensated for. Similarly, building transmission tracing systems for routine uses can provide routine value to public health workers.

Wastewater sequencing is also valuable for routine public health monitoring. We can envision benefits for detecting which variants of seasonal influenza or SARS-CoV-2 are circulating, or identifying locations where there is an outbreak of a foodborne pathogen, potentially even in advance of clinical detection. Ensuring sufficient attention is paid to these routine benefits is critical for ensuring that the systems are used and maintained even when no novel threats emerge.

Despite the seeming convergence of interests, identifying places where these routine uses are misaligned with crisis uses is 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 biosurveillance value of metagenomic sequencing is as a warning of imminent events. But a warning is only useful if it enables a response. Government willingness to respond requires a trustworthy signal across the possible scenarios where response is needed, and the value of the response is directly proportional to speed of the signal, and the speed of response. Building systems that can provide alerts in hours instead of days is far less important if responses take weeks. Similarly, the value of the system is limited by the accuracy of the warning signal, as response activities have costs. False alarms both reduce willingness to respond, and make the use of the system more expensive.

This also suggests where trade-offs exist. For example, a distributed analysis system allows for faster independent confirmation of an incipient outbreak, but may lead to more false positives. Similarly, non-public government analysis may be able to be more inclusive of otherwise private

data, but may be less trusted. In either case, governments need clear plans to enable response if and when a signal is detected - but the reporting mechanisms will greatly differ.

For all of this, key questions extend far beyond the scope of the present discussion. One of the most critical is how analysis is translated into policy, and by whom. Governments, non-governmental or multilateral organizations, and academics are all capable of analysis, but these will enable different types of response. The different options will require different types of interaction between the systems and government and international planning. And finally but perhaps most critically, different funding models for the systems may be needed. An excellent system which is then defunded is far worse than a more modest system which can be maintained.

Conclusion

In the coming years, next-generation sequencing and more widespread use of clinical and environmental sequencing for biosurveillance is inevitable. The question we addressed is how this can happen, and what policy issues are likely to arise.

Among the most important issues are market failures, in several forms. Possible failure modes include private capture of the market in ways that make widespread use expensive, or fragment the data and analysis ecosystem, as well potential abuse of data and privacy concerns. Relatedly, there is a possibility that the system is economically non-viable when no pandemic occurs, and funding is lost. Critical in a different way is how planning and response activities are able to capitalize on these systems and data.

To address these questions, a variety of projects seem useful, and are best led by different groups. Research or expert forecasting of timelines until various price levels and market penetration are reached is useful for planning. Work on building public or interoperable data systems and standards is important for industry groups and government or non-government agencies. Policy planning to ensure that payment systems or regulations don’t lock-in current or near-term technologies is also needed. Of course, none of these will supplant the technological advances which are needed, but each will help unlock their potential. The most critical 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.

Citations

1. National Biosurveillance Integration Center. In: Department of Homeland Security [Internet]. 17 Oct 2011 [cited 17 Jan 2022]. Available: https://www.dhs.gov/national-biosurveillance-integration-center
2. USDA APHIS | National Animal Health Monitoring System. 25 Oct 2021 [cited 26 Oct 2021]. Available: https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/monitoring-and-surveillance/nahms
3. Burkom H, Loschen W, Wojcik R, Holtry R, Punjabi M, Siwek M, 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: e26303. doi:10.2196/26303
4. Nkengasong J. Establishing the Africa Centres for Disease Control and Prevention: the upside of a crisis. 2017; 3.
5. WHO Hub for Pandemic and Epidemic Intelligence. [cited 25 Jan 2022]. Available: https://www.who.int/initiatives/who-hub-for-pandemic-and-epidemic-intelligence
6. IDseq ·. Chan Zuckerberg Initiative; 2021. Available: https://github.com/chanzuckerberg/idseq-web
7. Bajema DNE, Beaver W, Parthemore C, Atanda A, Horn J, Manheim D, et al. Toward a Global Pathogen Early Warning System: Building on the Landscape of Biosurveillance Today. Janne E. Nolan Center on Strategic Weapons: Council on Strategic Risks; 2021 Jul p. 102. Available: https://councilonstrategicrisks.org/wp-content/uploads/2021/07/Toward-A-Global-Pathogen-Early-Warning-System_2021_07_20-1.pdf
8. Bipartisan Commission on Biodefense. Biodefense in Crisis: Immediate Action Needed to Address National Vulnerabilities. 2021. Available: https://biodefensecommission.org/reports/biodefense-in-crisis-immediate-action-needed-to-address-national-vulnerabilities/
9. 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. doi:10.1007/s11430-017-9129-x
10. van Nuijs ALN, Castiglioni S, Tarcomnicu I, Postigo C, de Alda ML, Neels H, et al. Illicit drug consumption estimations derived from wastewater analysis: A critical review. Science of The Total Environment. 2011;409: 3564–3577. doi:10.1016/j.scitotenv.2010.05.030
11. Choi PM, Tscharke BJ, Donner E, O'Brien JW, Grant SC, Kaserzon SL, et al. Wastewater-based epidemiology biomarkers: Past, present and future. TrAC Trends in Analytical Chemistry. 2018;105: 453–469. doi:10.1016/j.trac.2018.06.004
12. McCall A-K, Bade R, Kinyua J, Lai FY, Thai PK, Covaci A, et al. Critical review on the stability of illicit drugs in sewers and wastewater samples. Water Research. 2016;88: 933–947. doi:10.1016/j.watres.2015.10.040
13. Zarei S, Salimi Y, Repo E, Daglioglu N, Safaei Z, Güzel E, et al. A global systematic review and meta-analysis on illicit drug consumption rate through wastewater-based epidemiology. Environ Sci Pollut Res. 2020;27: 36037–36051. doi:10.1007/s11356-020-09818-6
14. 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. Bulletin of the World Health Organization. 1999;77: 973.
15. Sinclair RG, Choi CY, Riley MR, Gerba CP. Pathogen surveillance through monitoring of sewer systems. Adv Appl Microbiol. 2008;65: 249–269. doi:10.1016/S0065-2164(08)00609-6
16. Ahmed W, Angel N, Edson J, Bibby K, Bivins A, O'Brien JW, 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. doi:10.1016/j.scitotenv.2020.138764
17. 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. 2021 Sep p. 2021.09.16.21263575. doi:10.1101/2021.09.16.21263575
18. 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. The Science of the Total Environment. 2020;737: 140405. doi:10.1016/j.scitotenv.2020.140405
19. McClary-Gutierrez JS, Aanderud ZT, Al-Faliti M, Duvallet C, Gonzalez R, Guzman J, 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;9. doi:10.1039/d1ew00235j
20. Reeves K, Leibig J, Feula A, Saldi T, Lasda E, Johnson W, et al. High-resolution within-sewer SARS-CoV-2 surveillance facilitates informed intervention. 2021 May p. 2021.05.24.21257632. doi:10.1101/2021.05.24.21257632
21. Derda R, Gitaka J, Klapperich CM, Mace CR, Kumar AA, Lieberman M, et al. Enabling the Development and Deployment of Next Generation Point-of-Care Diagnostics. PLOS Neglected Tropical Diseases. 2015;9: e0003676. doi:10.1371/journal.pntd.0003676
22. 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: S81–S88. doi:10.1136/sextrans-2016-053073
23. Quicke K, Gallichotte E, Sexton N, Young M, Janich A, Gahm G, et al. Longitudinal Surveillance for SARS-CoV-2 RNA Among Asymptomatic Staff in Five Colorado Skilled Nursing Facilities: Epidemiologic, Virologic and Sequence Analysis. 2020 Jun p. 2020.06.08.20125989. doi:10.1101/2020.06.08.20125989
24. Gallichotte EN, Nehring M, Young MC, Pugh S, Sexton NR, Fitzmeyer E, et al. Durable Antibody Responses in Staff at Two Long-Term Care Facilities, during and Post SARS-CoV-2 Outbreaks. Microbiol Spectr. 2021;9: e0022421. doi:10.1128/Spectrum.00224-21
25. Wood JLN, Leach M, Waldman L, Macgregor H, Fooks AR, Jones KE, 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: 2881–2892. doi:10.1098/rstb.2012.0228
26. 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. doi:10.3389/fvets.2020.582743
27. Graham JP, Leibler JH, Price LB, Otte JM, Pfeiffer DU, Tiensin T, et al. The Animal-Human Interface and Infectious Disease in Industrial Food Animal Production: Rethinking Biosecurity and Biocontainment. Public Health Reports. 2008;123: 282. doi:10.1177/003335490812300309
28. Jones BA, Grace D, Kock R, Alonso S, Rushton J, Said MY, et al. Zoonosis emergence linked to agricultural intensification and environmental change. Proceedings of the National Academy of Sciences of the United States of America. 2013;110: 8399. doi:10.1073/pnas.1208059110
29. Kracalik I, Malania L, Imnadze P, Blackburn JK. Human Anthrax Transmission at the Urban–Rural Interface, Georgia. The American Journal of Tropical Medicine and Hygiene. 2015;93: 1156. doi:10.4269/ajtmh.15-0242
30. 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: 991–1000. doi:10.1292/jvms.11-0049
31. Bipartisan Commission on Biodefense. Diagnostics for Biodefense - Flying Blind with No Plan to Land. 2020. Available: https://biodefensecommission.org/reports/diagnostics-for-biodefense-flying-blind-with-no-plan-to-land/
32. Morel B, Barbera P, Czech L, Bettisworth B, Hübner L, Lutteropp S, et al. Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult. Molecular Biology and Evolution. 2021;38: 1777–1791. doi:10.1093/molbev/msaa314
33. Rao S, Cox T, Rao P. Computational Challenges in Genomic Data Analyses (Presentation). 2020.
34. Chretien J-P, Rivers CM, Johansson MA. Make data sharing routine to prepare for public health emergencies. PLoS medicine. 2016;13: e1002109.
35. Pollett S, Johansson M, Biggerstaff M, Morton LC, Bazaco SL, Major DMB, et al. Identification and evaluation of epidemic prediction and forecasting reporting guidelines: A systematic review and a call for action. Epidemics. 2020;33: 100400.
36. Manheim D, Chamberlin M, Osoba OA, Vardavas R, Moore M. Improving Decision Support for Infectious Disease Prevention and Control. Santa Monica, CA: Rand Corporation. 2016.
37. Molecular Reality Inc. Biology is Just Beginning. Molecular Reality Inc.; 2021. Available: https://www.molecularreality.com/
38. 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. doi:10.3390/microorganisms9010188
39. Shen J, McFarland AG, Young VB, Hayden MK, Hartmann EM. Toward Accurate and Robust Environmental Surveillance Using Metagenomics. Frontiers in Genetics. 2021;12: 151. doi:10.3389/fgene.2021.600111
40. Consortium TNAO. A Global Nucleic Acid Observatory for Biodefense and Planetary Health. arXiv:210802678 [q-bio]. 2021 [cited 5 Sep 2021]. Available: http://arxiv.org/abs/2108.02678
41. Baum C, Cui H, Damg I, Esvelt K, Gao M, Gretton D, et al. Cryptographic Aspects of DNA Screening. SecureDNA; p. 33. Available: https://www.securedna.org/download/Cryptographic_Aspects_of_DNA_Screening.pdf
42. Zhou B, Han Y, Pei J, Jiang B, Tao Y, Jia Y. Continuous privacy preserving publishing of data streams. Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology. New York, NY, USA: Association for Computing Machinery; 2009. pp. 648–659. doi:10.1145/1516360.1516435
43. Pervaiz Z, Ghafoor A, Aref WG. Precision-Bounded Access Control Using Sliding-Window Query Views for Privacy-Preserving Data Streams. IEEE Transactions on Knowledge and Data Engineering. 2015;27: 1992–2004. doi:10.1109/TKDE.2015.2391098
44. Lewis G, Millett P, Sandberg A, Snyder-Beattie A, Gronvall G. Information Hazards in Biotechnology. Risk Analysis. 2019;39: 975–981. doi:10.1111/risa.13235
45. Shean RC, Greninger AL. One future of clinical metagenomic sequencing for infectious diseases. Expert Review of Molecular Diagnostics. 19: 849–851. doi:10.1080/14737159.2019.1658524
46. Fremin BJ, Sberro H, Bhatt AS. MetaRibo-Seq measures translation in microbiomes. Nat Commun. 2020;11: 3268. doi:10.1038/s41467-020-17081-z
47. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med. 2018;50: 1–14. doi:10.1038/s12276-018-0071-8
48. Blattman SB, Jiang W, Oikonomou P, Tavazoie S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. Nat Microbiol. 2020;5: 1192–1201. doi:10.1038/s41564-020-0729-6
49. Dar D, Dar N, Cai L, Newman DK. In situ single-cell activities of microbial populations revealed by spatial transcriptomics. 2021; 40. doi:10.1101/2021.02.24.432792

50. Diebold PJ, Felicia N. New, Hovan M, Satlin MJ, Brito IL. Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR. Cold Spring Harbor Laboratory. 2021. doi:10.1101/2021.01.22.427834
51. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLOS ONE. 2015;10: e0128036. doi:10.1371/journal.pone.0128036
52. 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: 458–471. doi:10.1038/s41596-020-00424-x
53. Coelho D, Lopes PA, Cardoso V, Ponte P, Brás J, Madeira MS, et al. Novel combination of feed enzymes to improve the degradation of Chlorella vulgaris recalcitrant cell wall. Sci Rep. 2019;9: 5382. doi:10.1038/s41598-019-41775-0
54. Easparro BA, Garrett S, Atwood J. Are Enzymes Required for the Extraction and Isolation of DNA in Conjunction with Mechanical Homogenization? The FASEB Journal. 2016;30: 1082.2-1082.2. doi:10.1096/fasebj.30.1_supplement.1082.2
55. Giovannoni M, Gramegna G, Benedetti M, Mattei B. Industrial Use of Cell Wall Degrading Enzymes: The Fine Line Between Production Strategy and Economic Feasibility. Frontiers in Bioengineering and Biotechnology. 2020;8: 356. doi:10.3389/fbioe.2020.00356
56. Zou Y, Mason MG, Wang Y, Wee E, Turni C, Blackall PJ, et al. Nucleic acid purification from plants, animals and microbes in under 30 seconds. PLOS Biology. 2017;15: e2003916. doi:10.1371/journal.pbio.2003916
57. Luong T, Salabarria A-C, Edwards RA, Roach DR. Standardized bacteriophage purification for personalized phage therapy. Nat Protoc. 2020;15: 2867–2890. doi:10.1038/s41596-020-0346-0
58. McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. Turnbaugh P, Garrett WS, Turnbaugh P, Quince C, Gibbons S, editors. eLife. 2019;8: e46923. doi:10.7554/eLife.46923
59. 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. doi:10.1186/1471-2229-5-23
60. Na HS, Yu Y, Kim SY, Lee J-H, Chung* and J. Comparison of the Performance of MiSeq and HiSeq 2500 in a Microbiome Study. 2020;48: 574–581. doi:10.48022/mbl.2008.08003
61. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, 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. doi:10.1186/1471-2164-13-341
62. 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, NY: Springer; 2018. pp. 223–232. doi:10.1007/978-1-4939-7471-9_12
63. Modi A, Vai S, Caramelli D, Lari M. The Illumina Sequencing Protocol and the NovaSeq 6000 System. In: Mengoni A, Bacci G, Fondi M, editors. Bacterial Pangenomics: Methods and Protocols. New York, NY: Springer US; 2021. pp. 15–42. doi:10.1007/978-1-0716-1099-2_2
64. 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: 1674–1676. doi:10.1093/bioinformatics/btv033
65. 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 Research. 2012;40: e155. doi:10.1093/nar/gks678
66. Coleman I, Korem T. Embracing Metagenomic Complexity with a Genome-Free Approach. mSystems. 2021 [cited 2 Jan 2022]. doi:10.1128/mSystems.00816-21
67. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile

metagenomic assembler. Genome Res. 2017;27: 824–834. doi:10.1101/gr.213959.116
68. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics, Proteomics & Bioinformatics. 2016;14: 265–279. doi:10.1016/j.gpb.2016.05.004
69. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018;36: 338–345. doi:10.1038/nbt.4060
70. Liu Q, Wu H, Wu L, Xie X, Kong J, Ye X, et al. Voltage-driven translocation of DNA through a high throughput conical solid-state nanopore. PLoS One. 2012;7: e46014. doi:10.1371/journal.pone.0046014
71. Loman NJ, Watson M. Successful test launch for nanopore sequencing. Nat Methods. 2015;12: 303–304. doi:10.1038/nmeth.3327
72. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol. 2020;38: 701–707. doi:10.1038/s41587-020-0422-6
73. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics. 2015;13: 278–289. doi:10.1016/j.gpb.2015.08.002
74. Hon T, Mars K, Young G, Tsai Y-C, Karalius JW, Landolin JM, et al. Highly accurate long-read HiFi sequencing data for five complex genomes. Sci Data. 2020;7: 399. doi:10.1038/s41597-020-00743-4
75. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019;37: 1155–1162. doi:10.1038/s41587-019-0217-9
76. 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: 30. doi:10.1186/s13059-020-1935-5
77. Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins KH, McIntyre ABR, et al. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. Sci Rep. 2017;7: 18022. doi:10.1038/s41598-017-18364-0
78. Lang D, Zhang S, Ren P, Liang F, Sun Z, Meng G, 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: giaa123. doi:10.1093/gigascience/giaa123
79. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020;17: 1103–1110. doi:10.1038/s41592-020-00971-x
80. Salmela L, Rivals E. LoRDEC: accurate and efficient long read error correction. Bioinformatics. 2014;30: 3506–3514. doi:10.1093/bioinformatics/btu538
81. Zhang H, Jain C, Aluru S. A comprehensive evaluation of long read error correction methods. BMC Genomics. 2020;21: 889. doi:10.1186/s12864-020-07227-0
82. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37: 540–546. doi:10.1038/s41587-019-0072-8
83. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7: e7359. doi:10.7717/peerj.7359
84. 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. doi:10.1038/srep24175
85. Maguire F, Jia B, Gray KL, Lau WYV, Beiko RG, Brinkman FSLY 2020. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic Islands. Microbial Genomics. 6: e000436. doi:10.1099/mgen.0.000436

86. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32: 605–607. doi:10.1093/bioinformatics/btv638
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 de-replication. ISME J. 2017;11: 2864–2868. doi:10.1038/ismej.2017.126
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: 226–232. doi:10.1038/ng.1028
89. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, et al. Structural variation in the gut microbiome associates with host health. Nature. 2019;568: 43–48. doi:10.1038/s41586-019-1065-y
90. Olekhnovich EI, Vasilyev AT, Ulyantsev VI, Kostryukova ES, Tyakht AV. MetaCherchant: analyzing genomic context of antibiotic resistance genes in gut microbiota. Bioinformatics. 2018;34: 434–444. doi:10.1093/bioinformatics/btx681
91. 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: 4379. doi:10.1038/s41467-020-18164-7
92. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, Li C, et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat Biotechnol. 2019;37: 937–944. doi:10.1038/s41587-019-0191-2
93. 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: 140-153.e9. doi:10.1016/j.chom.2019.10.022
94. Suzuki Y, Nishijima S, Furuta Y, Yoshimura J, Suda W, Oshima K, et al. Long-read metagenomic exploration of extrachromosomal mobile genetic elements in the human gut. Microbiome. 2019;7: 119. doi:10.1186/s40168-019-0737-z
95. Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu A, Bian 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. doi:10.1016/j.csbj.2021.11.028
96. Krumm N, Hoffman N. Practical estimation of cloud storage costs for clinical genomic data. Practical Laboratory Medicine. 2020;21: e00168. doi:10.1016/j.plabm.2020.e00168
97. 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: 131:1-131:36. doi:10.1145/3460427
98. Gardiyawasam Pussewalage HS, Oleshchuk VA. Privacy preserving mechanisms for enforcing security and privacy requirements in E-health solutions. International Journal of Information Management. 2016;36: 1161–1173. doi:10.1016/j.ijinfomgt.2016.07.006
99. Hampton-Marcell JT, Lopez JV, Gilbert JA. The human microbiome: an emerging tool in forensics. Microbial Biotechnology. 2017;10: 228–230. doi:10.1111/1751-7915.12699
100. Biosecurity for the Future: Strengthening Deterrence and Detection. In: Johns Hopkins Center for Health Security [Internet]. [cited 31 Jan 2022]. Available: https://www.centerforhealthsecurity.org/our-work/publications/biosecurity-for-the-future-strengthening-deterrence-and-detection
101. Cheng M, Cao L, Ning K. Microbiome Big-Data Mining and Applications Using Single-Cell Technologies and Metagenomics Approaches Toward Precision Medicine. Frontiers in Genetics. 2019;10. Available: https://www.frontiersin.org/article/10.3389/fgene.2019.00972
102. Dixon M, Stefil M, McDonald M, Bjerklund-Johansen TE, Naber K, Wagenlehner F, et al. Metagenomics in diagnosis and improved targeted treatment of UTI. World J Urol. 2020;38: 35–43. doi:10.1007/s00345-019-02731-9
103. Yu J, Feng Q, Wong SH, Zhang D, Liang Q yi, Qin Y, et al. Metagenomic analysis of

faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. 2017;66: 70–78. doi:10.1136/gutjnl-2015-309800

104. Jiang Y, Guo X, Liu L, Rode S, Wang R, Liu H, et al. Metagenomic characterization of lysine acetyltransferases in human cancer and their association with clinicopathologic features. Cancer Sci. 2020;111: 1829–1839. doi:10.1111/cas.14385