Accelerating the Defensive Deployment of Pathogen Sequencing

Simon Grimm

SUMMARY

America's inability to quickly detect new biological threats endangers national security. US biosurveillance could be improved through the use of metagenomic sequencing, a technology that would allow the detection of unknown pathogens. But although American companies dominate the sequencing market, no government program uses metagenomic sequencing at scale, and the use of sequencing in clinics remains low.

To remedy this, the United States should accelerate the deployment of sequencing through i) increased adoption of metagenomic biosurveillance by the Centers for Disease Control and Prevention (CDC), ii) investment by the Advanced Research Projects Agency for Health (ARPA-H), and iii) improved regulation of clinical metagenomics by the Food and Drug Administration (FDA). Specifically:

1. The CDC should establish public-private partnerships to run metagenomic sequencing on 10-20 percent of samples collected through its federal wastewater and international traveler surveillance programs.
2. To increase the adoption of clinical metagenomic sequencing, the FDA should release a public update of its 2016 draft guidance on the regulatory approval of sequencing-based pathogen diagnostics.
3. ARPA-H should stand up a program for the development of faster and more robust sample processing methods for metagenomic sequencing.

PROBLEM

America is vulnerable to biological threats. No major government program has the goal of continuously monitoring the emergence of new pathogens, whether they are natural, accidentally leaked, or intentionally released. Common-sense improvements to disease surveillance have yet to be implemented: monitoring of well-known pathogens like H5N1 is often delayed by months, and data sharing remains slow or incomplete. All the while, biological risks are increasing, as malicious actors can use new technologies to engineer biology for nefarious ends.

If we want to reliably mitigate new pathogen outbreaks, aggressive steps to improve biosurveillance are required. The key technology enabling better pathogen detection is metagenomic sequencing: a method for reading out all genetic material in a sample, allowing the identification of both known and unknown biological threats (Appendix 1).

The United States has the technological capacity to use metagenomic sequencing for nation-scale pathogen monitoring. American companies provide more than half of global sequencing capacity and comprise the majority of the sequencing market. But metagenomic sequencing is not routinely deployed in America's biosecurity architecture. No FDA-cleared sequencing-based diagnostics exist, leaving both hospitals and military bases with few tools to reliably detect unknown pathogens. Similarly, though CDC monitors incoming international travelers and wastewater for known viruses, the agency does not use metagenomic sequencing to get ahead of unforeseen biological threats.

The administration's interest in government reform and its embrace of private sector innovation both provide an opportunity to strengthen American biosurveillance. Through improved FDA guidance, companies could develop and sell more affordable sequencing-based pathogen diagnostics. By working with the private sector, the CDC could pilot metagenomic sequencing within federal biosurveillance programs. And US government research and development (R&D) bodies like ARPA-H could further the development of sequencing-based pathogen detection. These changes would both strengthen America's security against future biological threats, and further US companies' lead in sequencing technology.

SOLUTION

The Centers for Disease Control and Prevention

The CDC has entered into multiple public-private partnerships to build biosurveillance systems: the National Wastewater Surveillance System (NWSS), launched in 2020, collects wastewater across the US, covering more than 100 million citizens. Meanwhile, the Traveler-based Genomic Surveillance (TGS) Program collects nasal swabs and airplane waste from thousands of international travelers each month.

- Working with private companies, both of these programs should pilot metagenomic sequencing of collected samples.
- Under the PREVENT Pandemics Act, CDC is allowed to use more flexible Other Transactional Authority awards (OTA) for biosurveillance purposes. OTA awards should be used for pilot public-private partnerships that trial sequencing of 10 to 20 percent of samples from NWSS and TGS.
- For NWSS, samples should be taken from a set of five to ten major metropolitan areas.
- For TGS, nasal swab samples and airplane wastewater should be collected from three or more major international airports and pooled before sequencing to maintain anonymity.

After removing human data, the resulting sequencing data should be shared in real-time (less than a day after data generation) to allow analysis by actors beyond CDC.

Currently, programs like TGS squander the value of uploaded data by omitting crucial information like flight or airport origin. This should be changed; following guidance on metagenomic sequencing data sharing released by HHS (see below), metagenomic sequencing data should always be linked to precise contextual metadata.

Following actions by the Department of Government Efficiency (DOGE), CDC will likely undergo structural reform. Instead of cutting NWSS or TGS, agency reform should create space to make these pathogen-agnostic monitoring platforms a centerpiece of infectious disease surveillance.

The Department of Health and Human Services (HHS) should provide data sharing guidance for metagenomic sequencing data. This guidance would establish how human genomic material should be removed prior to upload and clarify that pooled sequencing samples that had human genomic material removed do not fall under Health Insurance Portability and Accountability Act (HIPAA) privacy rules. Additionally, the provision of precise contextual metadata should be mandatory (Appendix 2).

The Food and Drug Administration

The FDA originally released draft guidance on the approval of sequencing-based diagnostics in 2016. Given rapid technological advances, the agency decided to not further develop this guidance. However, sequencing technology has now become cheaper and more precise, allowing the development of sample-to-answer sequencing devices in the next 5-10 years. Based on these developments, FDA should publish updated draft guidance:

- This guidance should clearly specify that metagenomic pathogen diagnostics must demonstrate very high specificity, while allowing moderate sensitivity. Additionally, FDA should clarify under which conditions metagenomic detection methods can be used to detect newly emerging pathogens without additional regulatory review.
- Finally, the guidance should outline conditions under which sequencing-based diagnostics would qualify for reimbursement by the Centers for Medicare & Medicaid Services (CMS).

ARPA-H

ARPA-H should set up a new program to develop faster and more robust sample processing methods for clinical metagenomic sequencing.

- At present, getting clinical metagenomic results takes anywhere from twelve hours to several days, with sample preparation as the primary bottleneck.
- Building on the Biomedical Advanced Research and Development Authority's (BARDA) DRIVe's Agnostic Diagnostics work, a new ARPA-H program should focus on technologies that prepare heterogeneous clinical samples for sequencing more quickly.
- The ultimate aim should be to get different sample types onto a sequencer in under two hours, at less than $50 per sample, while maintaining detection capability for pathogens at clinically relevant concentrations.

JUSTIFICATION

The status quo will leave us exposed to new biological threats. Routine metagenomic sequencing has only become viable in recent years, as the cost of sequencing has plummeted. Despite these technological advances, it remains uncertain whether the government will adopt metagenomic sequencing anytime soon.

The CDC's failure to look beyond known threats is explained by its fragmented structure. Founded in 1946 with a specific mandate for disease surveillance and epidemiology, the agency has splintered into subunits dedicated solely to specific known pathogens. This structure makes it challenging for the CDC to incorporate technologies like metagenomic sequencing, which detect many pathogens at once, including unknown ones. Similarly, it remains unclear if the CDC will fully integrate new biosurveillance programs into its disease surveillance apparatus, despite their affordability and effectiveness. Together, TGS and NWSS public-private components cost just $37 million a year—less than half a percent of the CDC's 2024 budget—and reduce reliance on state public health labs, many of which struggle to share timely data or adopt advanced detection methods.

The FDA's past approach to regulating diagnostics is a similarly bad fit for modern detection technologies. To approve diagnostics, companies need to show reliable performance for detecting specific pathogens. This approach makes little sense when

evaluating diagnostics that can detect as-of-yet unknown pathogens. With preliminary guidance released nearly a decade ago, the FDA has left regulatory uncertainty unresolved. As a result, only companies with close relationships to the FDA understand the agency's expectations for approving metagenomics-based diagnostics, slowing innovation and adoption.

Adopting metagenomic sequencing

Both domestically and abroad, more actors are exploring metagenomic sequencing for pathogen detection. In the United Kingdom, the National Health Service has launched an ambitious sequencing-based pathogen detection system, partnering with Oxford Nanopore Technologies to diagnose severe respiratory illness. Next door, the European Union has committed €24 million to develop a rapid point-of-care metagenomic sequencing diagnostic.

The US has taken some small steps to embrace sequencing-based pathogen detection. The Department of Defense's Defense Innovation Unit launched the ANTI-DOTE project, a program for detecting engineered pathogens in military base wastewater, and BARDA DRIVe spent $3–4 million dollars on developing metagenomic sequencing tools through its "Agnostic Diagnostics" program. However, these initiatives remain too limited in scope and scale to enable the United States to reliably detect unknown or engineered pathogens in the near future.

Through the actions outlined in this policy brief, HHS agencies can make large improvements to American pathogen early detection. The combination of cheaper sequencing, established sampling infrastructure, and increasing biological risks make the current moment well-suited for accelerating the deployment of metagenomic sequencing, both in clinics and within the federal government's biosecurity infrastructure. ■

FURTHER RESOURCES

- CDC, "SARS-CoV-2 Sample Positivity in Travellers Can Predict Community Prevalence Rates: Data from the Traveller-Based Genomic Surveillance Programme," 2024
- National Wastewater Surveillance System, "About CDC's National Wastewater Surveillance System (NWSS)"
- FDA, "Infectious Disease Next Generation Sequencing Based Diagnostic Devices: Microbial Identification and Detection of Antimicrobial Resistance and Virulence Markers," *Federal Register*, 2016
- Defense Innovation Unit, "DIU Demonstrates Capability To Find Novel Threats in High Complexity Wastewater Data," 2025
- Simon L. Grimm et al. "Inferring the Sensitivity of Wastewater Metagenomic Sequencing for Pathogen Early Detection," medRxiv, 2023
- Charles Y. Chiu and Steven A. Miller, "Clinical Metagenomics," *Nature Reviews Genetics*, 2019

- Nava Whiteford et al., "Towards Ubiquitous Metagenomic Sequencing: A Technology Roadmap"
- Arielle D'Souza and Janika Schmitt, "Mapping America's Biosurveillance," Institute for Progress, 2024

. . . .

Simon Grimm is a medical doctor working on pathogen early detection at SecureBio and the Massachusetts Institute of Technology.

APPENDIX

1. A primer on metagenomics

Unlike targeted approaches such as antigen tests or quantitative polymerase chain reaction (qPCR), metagenomic sequencing works by breaking up all genetic material (DNA and RNA) in a sample into short fragments, reading the sequence of DNA/RNA bases in each fragment, and then matching these reads against reference databases to determine which organisms they came from.

The cost of sequencing has dropped precipitously in recent years, making it likely that metagenomic sequencing will become cost-competitive with targeted pathogen detection approaches in three to ten years. In the mean-time, there is already research showing that metagenomic sequencing is viable for both detecting a large number of pathogens in wastewater and in a large swath of clinical sample types.

2. Barriers to metagenomic data sharing

There are three main concerns around the generation and sharing of metagenomic data.

First, metagenomic datasets can be noisy, increasing the risk of false-positive findings. Even for more simple types of data, a concern about false positives has traditionally made CDC averse to data-sharing. However, quickly resolved false positives are much less harmful then not identifying a new biological threat as fast as possible. Public health agencies should thus move toward more data sharing, rather than less. This would allow a larger number of actors to review public health data, increasing transparency and accelerating threat detection.

A second concern is that metagenomic sequencing data can contain human genomic material. Uploading large amounts of human genomic data could both pose privacy risks, and could enable exploitation by US adversaries. However, researchers have developed robust methods to remove human reads from sequencing material. For instance, data that gets uploaded to the National Center for Biotechnology Information's Sequence Read Archive (SRA) can be run through its Human Read Removal Tool (HRRT), or researchers can use many publicly available tools to remove human data before any further analysis takes place.

Finally, CDC's wastewater and international traveler programs routinely omit large amounts of raw data and crucial metadata due to fears that associating pathogens with

specific counties or origin countries encourages discrimination. However, these omissions harm transparency and significantly reduce the value of publicly-funded biosurveillance data. To not have this happen with metagenomic sequencing data, fast data sharing and provision of appropriate metadata should become mandatory within federal biosurveillance programs.