



Implementation Framework: Toward a National Genomic Surveillance Network

Supported by



Executive Summary

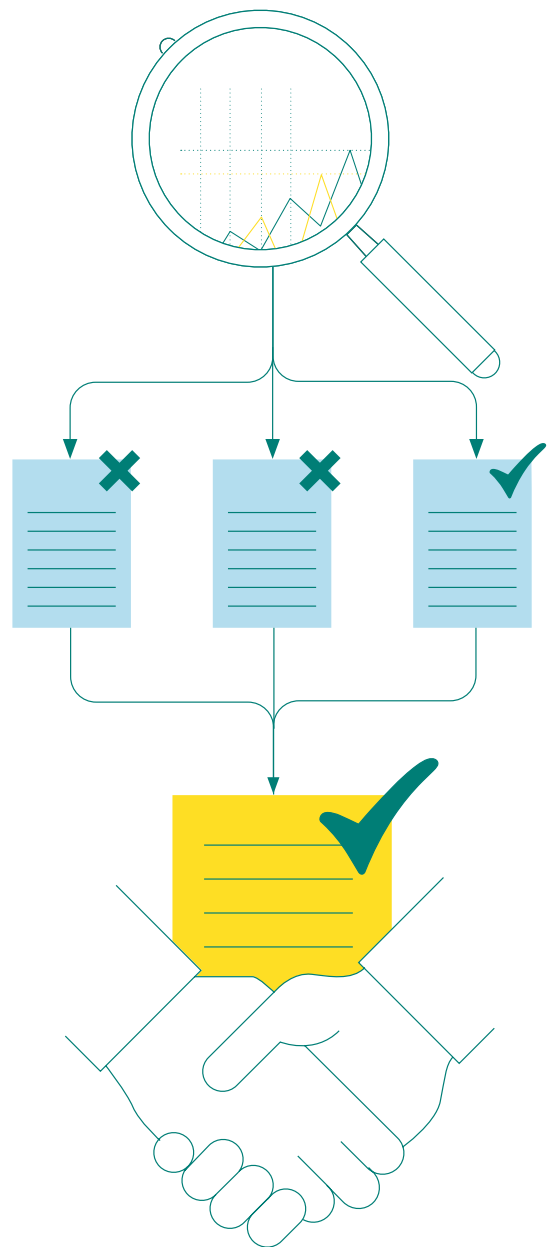
The rapid emergence and spread of SARS-CoV-2 make clear the need for robust systems to detect and track viral threats. A robust surveillance system could have identified SARS-CoV-2 community prevalence and spread sooner, and better prepared the United States for the 2019 outbreak.

When SARS-CoV-2 was finally identified, the virus had spread for several weeks undetected.

It appeared in several states simultaneously, and in people with limited or no travel history to regions with high numbers of SARS-CoV-2. If a more robust surveillance system were in place, it could have tracked SARS-CoV-2 from the start of the U.S. outbreak, giving policymakers a critical tool to limit its spread. The emergence of new variants that keep catching the U.S. flat-footed suggests a more robust genomic surveillance system is urgently needed.

The Rockefeller Foundation is releasing an accompanying document titled [Accelerating National Genomic Surveillance](#).

This Action Plan outlines broad priorities that should be considered when establishing a national genomic surveillance network, including: (1) Build a Viral Defense System, (2) Be Diverse and Deeply Human, (3) Connect the Dots, (4) Provide More to Do More, (5) Build New Analytic Tools and Make Them Visual, and (6) Rapidly Assess the Threat Posed by New Variants.



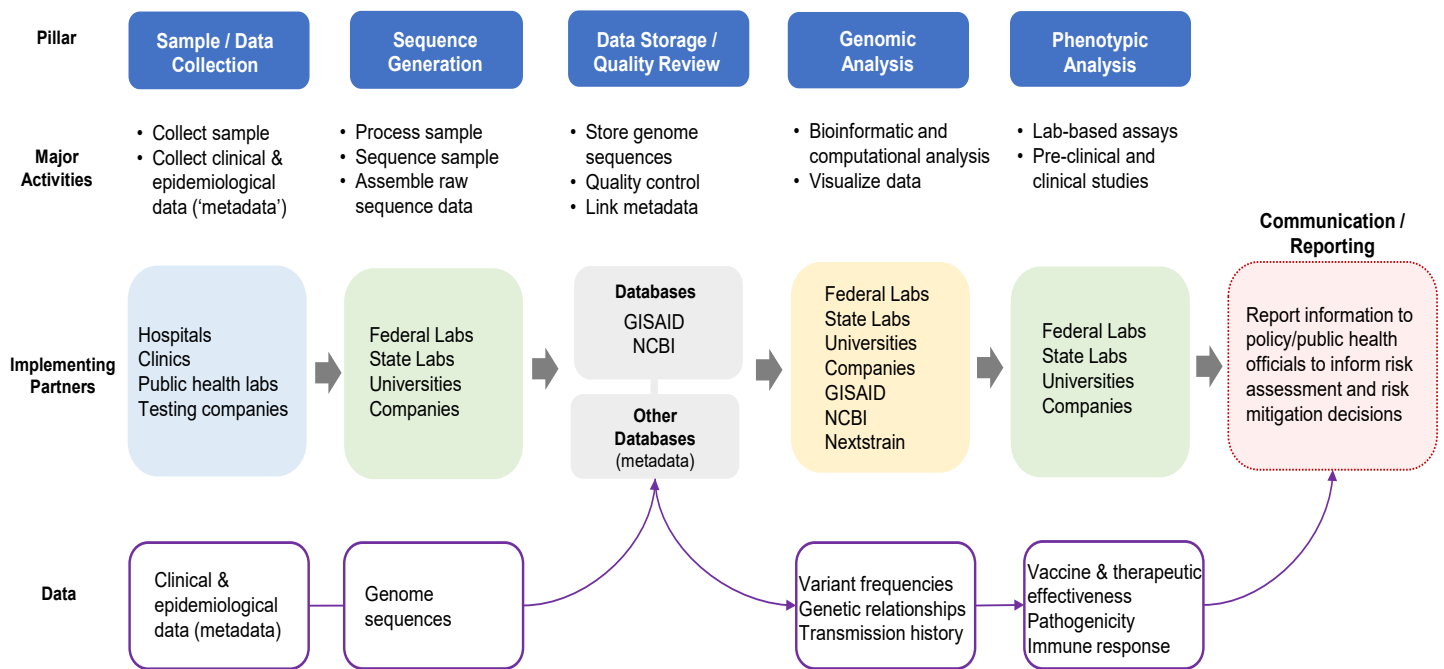
Genomic surveillance is an essential tool for tracking viruses as they emerge, spread, and continuously evolve.

Genomic surveillance integrates clinical, epidemiological, genomic, and phenotypic data to track changes in virus transmission, virulence, and effectiveness of medical countermeasures. Recent advances in next-generation sequencing make it possible to quickly and cost-effectively sequence large numbers of SARS-CoV-2-positive cases. Parallel advances in bioinformatics, computational biology, and molecular virology make it possible to analyze the virus in context to assess risk in close to real-time.

The integration of these data sets and analyses produces actionable information for public health officials and policymakers. Many countries have invested in genomic surveillance, and their investments have paid significant dividends. For example, the UK established a national genomics surveillance network that recently identified a critical variant, B.1.1.7, that enhances viral transmission and increases the risk of death compared to other variants.

In March of 2020, the U.S. government launched a nationwide SARS-CoV-2 genomic surveillance system called SPHERES. This initiative was an effort to coordinate academic and private sector sequencing efforts to better align them with public health, and to use these efforts as a bridge to more formal national genomic surveillance programs.

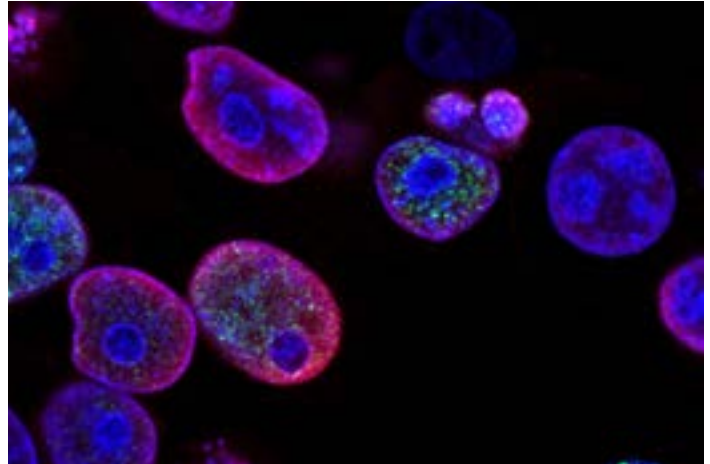
Genomic Surveillance



Overview of Genomic Surveillance

The Rockefeller Foundation is exploring how to support accelerating genomic surveillance in the United States. One area of need is greater coordination and collaboration among a diverse group of stakeholders to address barriers that impede the multifaceted functioning of robust genomic surveillance. More specifically, a consortium of stakeholders from public, private and non-governmental organizations is needed to:

- Support national leadership that establishes priorities and standards for data collection, sequencing, and analysis.
- Reduce barriers that prevent the collection and rapid sharing of clinical and viral samples as well as information (e.g., metadata) compiled from various data sources.
- Improve communication and coordination among all partners involved in genomics surveillance through regional centers and national working groups.
- Transmit information and analyses so policymakers and public health officials can act quickly to mitigate the impact of emerging pathogens and their variants.
- Monitor genomic surveillance in the US to ensure adequate numbers and representation of SARS-CoV-2 positive cases are sequenced and analyzed to produce actionable information for public health officials.
- Lead U.S. efforts to align and coordinate with a global genomics surveillance network to identify, track and mitigate emerging pathogens before they spread globally.



OPPORTUNITIES TO REDUCE BARRIERS:

On February 16, 2020, The Rockefeller Foundation convened a diverse group of experts engaged in genomic surveillance. Although genomic surveillance stakeholders are often siloed, and each faces unique challenges, several opportunities to enhance genomic surveillance emerged from the convening. These include:

- Improving linkage to metadata
- Increasing sequencing volume and representation
- Bolstering information sharing and integration
- Advancing genomic analysis
- Accelerating phenotypic analysis
- Enhancing communication and data visualization

A national consortium might focus on several areas where it sees opportunities to improve coordination and address barriers that hamper genomic surveillance. Opportunities to enhance genomic surveillance will pay immediate dividends by providing actionable information to public health officials to increase awareness and reduce the spread of SARS-CoV-2 variants of interest and concern; and lay the foundation for national and global surveillance systems that prevent future pandemics.

Improving Linkage to Metadata



Metadata includes clinical and epidemiological information that can add greater context to better understand the impact of factors driving evolution of the virus and impact of these properties of the virus on special populations and communities. Examples of metadata include age, location, vaccine status, and clinical symptoms. Metadata combined with genotypic and phenotypic data is a powerful predictor of how the virus is evolving and the medical interventions needed to keep pace. Unlike viral genomic sequences (which in the U.S. are not protected under HIPAA privacy laws), metadata is sensitive and must be protected to ensure patient privacy.

Several barriers impede the access and use metadata as part of viral genomic surveillance. These include the nature of the data itself, which can include sensitive patient data protected by federal, state, and local regulations. Academic labs may need Institutional Review Boards or waivers to access certain types of data. Limited communication between data users (researchers analyzing data) and data collectors (hospitals, commercial labs, and public labs) also impedes information flow. Guidance on metadata collection—such as what data should be collected—is sometimes vague, making it difficult for hospitals, clinics, and public and private labs that perform SARS-CoV-2 testing to know what they should or should not collect. When metadata is collected, the lack of standards (e.g., descriptions of clinical symptoms) can make it difficult to analyze across populations. Finally, secure databases that enable the integration of metadata with viral genomic sequences are lacking.

OPPORTUNITIES TO IMPROVE LINKAGE TO METADATA:

1. Create metadata priorities and standards. National alignment is needed to clarify metadata priorities (i.e., establish a minimal and specific set of data required to address critical public health questions). National templates for metadata collection and reporting are also needed to compare data across geographic regions and populations. In the United Kingdom (UK) the best and strongest analyses are those that combine contact tracing with genomics (e.g., the most robust evidence for increased transmissibility and severity of the B.1.1.7 variant was based on such data). Critically, as part of this effort to generate priorities and standards, a communications ‘mechanism’ needs to be established between data generators, such as hospitals, and data users, such as scientists at universities. This ‘mechanism’ will facilitate communication between two groups to align research and public health priorities.

2. Strengthen existing databases and explore additional ones to improve access to metadata, while ensuring compliance with applicable regulations. A minimal set of metadata that significantly enhances genomic surveillance and does not affect patient privacy (such as geographic location of a viral sequence) should be uploaded to publicly accessible databases. These are metadata that are allowed to be collected and shared under federal, state and local regulations. Separately, certain types of metadata may need to be stored in non-publicly accessible databases separate from the databases holding genomic and phenotypic data. Ultimately, these metadata will need to be accessible and paired with epidemiological data to generate more useful information for public health officials.

Increasing Sequencing Volume and Representation

Genomic sequencing involves several steps, such as accessing samples stored under proper conditions, sample preparation, sequencing, and processing the raw sequence data (i.e., assembly of raw sequences). Sequencing is performed using a high-throughput sequencer. The processing of raw sequencing data often requires access to computing clusters or cloud computing resources able to compute and store terabytes of data. Bioinformatics specialists are needed with the ability to process raw sequencing data (i.e., perform quality control), assemble consensus genomes, and analyze the assembled sequences.

Several barriers impede the scaling of sequencing. One barrier has been limited and patchwork funding for sequencing. This barrier, however, could be partially addressed with the recent announcement by the CDC to invest \$200 million in new funding for genomic sequencing. Besides funding, limited workforce capacity, especially in state public health labs, also limits genomic surveillance. The limited numbers sequencing experts and bioinformaticians constrains sequencing, sharing and analysis of genomic data. Finally, viral genomic sequencing for public health surveillance purposes is mandated to be done in a CLIA facility. The need to meet CLIA regulations prevents some groups from engaging in genomic sequencing.

OPPORTUNITIES TO SCALE SEQUENCING:

1. Strengthen workforce capacity. To rapidly train sequencing experts and bioinformaticians in public health labs, offering technical assistance and fellowship programs through the Centers for Disease Control (CDC) and Association of Public Health Laboratories (APHL), should be expanded. Programs focused on placing professionals qualified in sample processing, sequencing and raw data

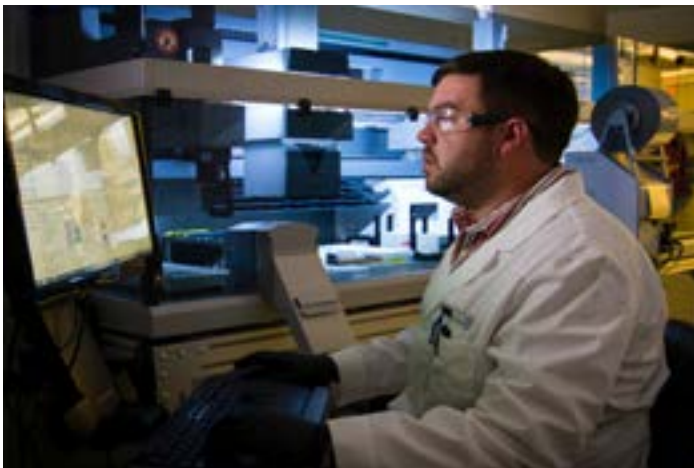
EXAMPLE IN ACTION

An example of a particularly successful program is the CDC/APHL Bioinformatics Fellowship Program (<https://www.aphl.org/bioinformatics>). This program has transformed U.S. federal and state public health preparedness. After completing the program, more than 70% of its graduates continue in public health careers.

Public health laboratories also need advocacy to address state limits on the number of full-time employees (FTEs) that can be hired. Hiring additional, permanent staff with genomic sequencing and analysis skills in all state public health labs is critical to sustaining genomic capacity in the public health laboratory system.

processing into public health laboratories should also be expanded and further developed.

2. Ensure sequencing is representative of all regions and all groups, particularly minority and underserved populations. Future increases in funding will increase sequencing volume. However, it is critical that genomic surveillance represents geographical and ethnic diversity of the U.S., particularly underserved populations that infrequently interact with the U.S. healthcare system. In addition to underserved groups, sequencing and analysis must be carried out on high-priority clinical cases, such as individuals infected long-term (i.e., ‘long-haulers’), immunocompromised individuals, and vaccine failures.



Bolstering Information Sharing and Integration

Although sharing of genomic sequencing information during the COVID-19 pandemic has been excellent, the lack of data integration is a significant impediment to scaling genomic surveillance. Clinical and epidemiological data (i.e., metadata) is often not linked to genome sequences, genomic analysis, or phenotypic characterization. Several barriers prevent data integration. In the U.S., certain metadata is more sensitive

than viral genomic and phenotypic data, and often includes components protected by state and federal laws. These data must be handled differently from genomic and phenotypic data, which do not reveal a patient's information. Because of the limitations on metadata, it can be difficult to collect, transmit and store. In particular, sensitive metadata generally cannot be stored in publicly accessible databases. The physical separation of data makes it difficult to integrate.

OPPORTUNITIES TO IMPROVE DATA INTEGRATION:

1. Strengthen existing databases, while also exploring databases for genomic surveillance metadata. Both of the major genomic sequence databases, the National Center for Biotechnology Information (NCBI) and The GISAID Initiative, store and have the ability to store metadata. NCBI has an existing capability to connect open access sequence records with controlled access metadata, which can be aligned with principles of open and rapid data exchange for the public good.

2. Establish a network that links publicly accessible databases containing genomic and phenotypic data with databases containing metadata. A network is needed that facilitates public health researchers' access to sensitive metadata and permits them to easily link such sensitive metadata to genomic and associated non-sensitive metadata (including epidemiological and clinical data) in publicly accessible databases, while ensuring patient privacy. Federal regulation may be necessary. This system must be able to integrate all available data to predict the impact of new variants rapidly. It should have a precise mandate in terms of the reports it produces and provides actionable information to public health officials and policymakers.

NCBI is building such a network as part of the U.S. government's efforts to understand whether emerging variants affect vaccine effectiveness. GISAID already integrates genomic data with certain types of metadata. Thus, it may be possible to leverage and expand these efforts, rather than build another platform.

An alternative to a network that links metadata with genomic analysis, is to develop analytic tools that can be used by those who routinely collect and access metadata, such as hospitals, clinics, and public health labs. In other words, bring the tools to the data rather than the data to the tools (i.e., make software designed for use by bioinformaticians in public health departments).

Advancing Genomic Analysis

A national genomics surveillance effort will generate thousands of terabytes of raw sequencing data. These data must be standardized, uploaded, curated, and analyzed. Data analysis generally involves processing raw genomic data to ensure specific quality standards and then running a series of bioinformatic and computational algorithms to achieve a specific analytical goal. Data analysis is complex and computationally demanding. It often requires access to large computing clusters capable of processing data quickly.

Analysis of genomic data faces several challenges. Improvements are needed to existing tools to make them more user-friendly and better able to scale. Additional tools are also needed for new angles of analysis. These include tools for finer-scale "genomic epidemiology", where the utility derives from addressing fine-scale epidemiological questions, such as "is this cluster of cases in a workplace part of the same epidemiologically-linked outbreak?" These types of questions require access to non-public detailed epidemiological metadata and should generally be occurring within public health departments, so they can be co-located with the sensitive epidemiological data. Realizing this goal requires the development of user-friendly analytic tools specifically developed for

public health workers with access to sensitive metadata. Tools are also needed for broader scale genomic surveillance, where the questions surround more general epidemiological and evolutionary questions regarding the circulation patterns and spread of genetic variants. These analyses require larger amounts of data and more sophisticated methods, but do not require the same level of access to detailed metadata. For this broader-scale genomic surveillance, at a minimum, a system is needed that can estimate the prevalence of different variants through time and across space.

OPPORTUNITIES TO ADVANCE GENOMIC DATA ANALYSIS:

1. Establish analytic priorities and standards for genomic data. Priorities for minimum analysis could include sample-level reporting of SARS-CoV-2 lineages and clades along with mutations relative to reference strains. Other priorities could be population-level analysis of phylogenetic trees linking samples from different geographies and analysis of spatio-temporal frequencies of mutations, lineages and clades. Templates should describe optimal parameters for quality control and analysis.

2. Create and share analytical tools that are easy-to-use, scalable, and capable of addressing unanswered questions. A next generation set of analytical tools should be developed that are easy-to-use, even by analysts with limited viral genomics experience. These tools must also be capable of handling high volumes of complex and variable data. They should be freely available and updated continuously by a dedicated group of software engineers. Analytical tools need to serve national surveillance priorities; thus, they should focus on generating reports that are understandable by a diverse audience with varying levels of expertise and use case scenarios.

Accelerating Phenotypic Analysis

For a holistic approach to pandemic surveillance, functional characterization is essential in determining how genetic changes in the viral genome impact viral transmission, virulence, and vaccine or treatment response. This type of characterization typically involves lab-based assays and clinical or non-clinical models to study the virus's functional impact on a host or population.

Several barriers slow phenotypic analysis. First, it can be difficult to obtain viral isolates from domestic or international sources because of complicated or restrictive material transfer agreements and shipping regulations. Second, patient samples and serum from SARS-CoV-2 cases and vaccinated individuals are generally not stored for long periods. The lack of access to patient samples makes it difficult to evaluate the immune response to novel SARS-CoV-2 variants quickly.

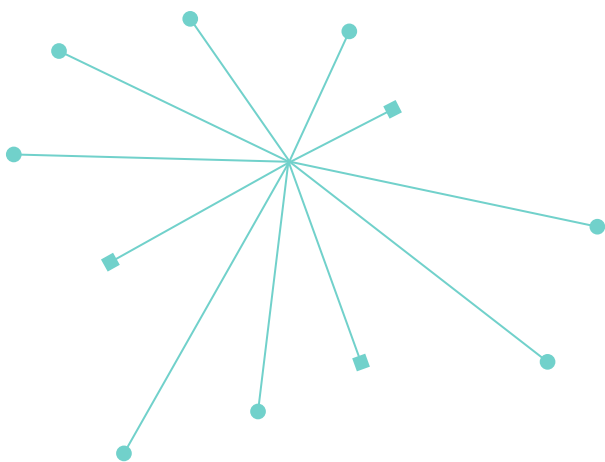
Improving phenotypic analysis of SARS-CoV-2 variants is a priority for the U.S. government. An interagency group has been established to coordinate phenotypic analysis across federal agencies and NIH has separately engaged in efforts to coordinate phenotypic analysis with industry through the NIH ACTIV consortium. The goal of these programs is to facilitate cooperation across the U.S. government and with the private sector to improve risk assessment of newly detected variants.

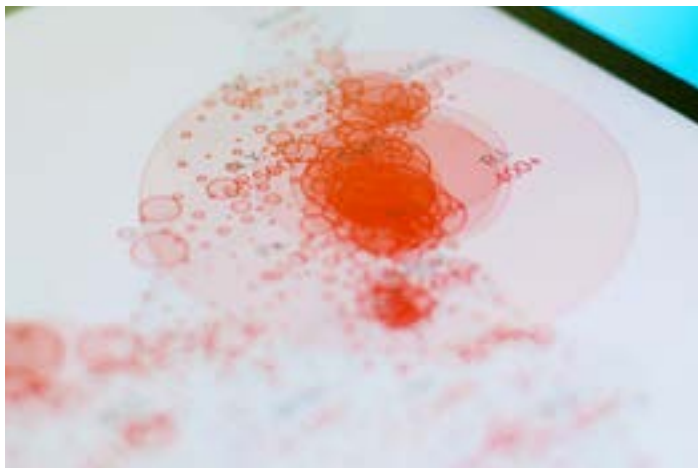


OPPORTUNITIES TO ACCELERATE PHENOTYPIC ANALYSIS:

1. Improve access to clinical samples and viral isolates. Several initiatives may allow greater access to clinical samples and viral isolates for immune and phenotypic analysis, including alignment of material transfer agreements among partners, improved coordination between testing sites where samples are collected and laboratories doing phenotypic studies, and support for the storage of high-priority clinical samples (such as long-term infected patients or vaccine and treatment failures). Standardized language in institutional review board (IRB) protocols and patient consent forms should be encouraged that allows for sharing of virus isolates and serum samples.

2. Invest in international laboratory capacity so functional analysis can be performed in countries where the viruses emerge, or variants are detected. There must also be greater coordination among international and domestic labs involved in phenotypic analysis so overlapping studies are reduced and data generated are shared more rapidly.





OPPORTUNITIES TO ENHANCE COMMUNICATION AND DATA VISUALIZATION:

1. Increase communication and coordination between public health officials, researchers and industry. A national consortium can bring these disparate groups, which infrequently interact, together. In particular, genomic epidemiologists who specialize in integrating clinical, epidemiological, genomic, and phenotypic data should be at the forefront of this engagement with public health officials.

2. Improve tools to visualize data. A set of tools that visualizes genomic epidemiology information for non-experts is needed to improve the framing, flow, and accessibility of information transmitted to public health officials.

Enhancing Communication and Data Visualization

Information must be shared quickly and accurately with other scientists, public health officials, and policy makers. Many officials, however, are not experts in viral genomics. Therefore, information must be presented in clear and concise formats. Interpretation of the information is also needed to help non-experts assess the potential public health impact of a particular virus or variant.

Measuring Success

Monitoring and evaluating the success of a national genomics surveillance system should be guided by several metrics. Examples of metrics include the ones listed in the following table:

Metric	Minimum Standard	Target
Cases sequenced	SARS-CoV-2 positive cases representative of regional, racial and ethnic diversity	5% of cases* Cases sequenced
Time from sample received by sequencing facility to uploading of consensus genome	Consensus genome (i.e., the most common nucleotide at any specific position in the genome) that passes quality control standards	10 days Time from sample received by sequencing facility to uploading of consensus genome
Time required for genomic analysis	Identify novel variants, determine viral lineage, and place into appropriate city strain phylogeny.	24-48 hours Time required for genomic analysis
Time required for threat assessment (phenotypic analysis)	Neutralizing assays using vaccinated serum	14-21 days* Time required for threat assessment (phenotypic analysis)

*This number is illustrative. It is based on a model by Vavrek et al. (MedRxiv, January 15, 2021) that shows 5% sampling of all positive tests allows the detection of emerging strains when they are at a prevalence of 0.1% to 1.0%. *Depends on rapid access to viral isolate and vaccinated serum.

Moving Forward Together

There is immense opportunity, but advancements toward a more robust genomic surveillance network require equally robust coordination.

As a first step, The Rockefeller Foundation intends to support working groups that will seek to address major barriers to scaling of national genomics surveillance. These four working groups could include:

- 1. Improving Access to Metadata and Data Integration**
 - 2. Building Workforce Capacity to Increase Genomic Sequencing**
 - 3. Advancing Genomic Analysis**
 - 4. Accelerating Phenotypic Analysis**
-

Through these working groups, The Rockefeller Foundation will convene stakeholders from public, private, and non-governmental organizations across each of these focus areas. By bringing together a diverse set of actors in the space, working groups will help quickly surface issues, and collectively develop solutions. Working groups will also facilitate communication and coordination between different stakeholders. An entity will then also be responsible for coordinating across different working groups to document cross-cutting issues, and lay the groundwork for a highly effective and efficient national network.

In addition, consideration should be given to building a network of regional hubs of public and private partners working together to share information and leverage genomic sequencing and analytical capacity within the region. Coordination within each region could be facilitated by a large genome sequencing center with expertise in pathogen surveillance.

This center could coordinate genomic surveillance among several of the region's hospitals and clinics, diagnostic labs, public health labs, universities, and companies. In each region these entities will vary. A national level entity could coordinate genomic surveillance among the regional hubs. This arrangement would allow regional hubs to leverage resources across the network and promote coordination on key issues affecting all regions.

Conclusion

A robust genomics surveillance system is needed in the U.S. to detect and track SARS-CoV-2 variants and mitigate their impact on transmission, virulence, vaccines, therapeutics, and natural immunity. The ideal system will be able to transmit information in near real-time to public health officials, policymakers, the public, and health industry executives.

To achieve this vision, existing U.S. genomic surveillance capacity and expertise must be leveraged and expanded. This is achievable with greater coordination and collaboration to across the full range of concerned stakeholders. By working to remove barriers that impede scale-up and to develop even more effective ways of working together, the U.S. has the opportunity to build a genomic surveillance with the scope and scale the country so urgently needs.

A robust U.S. genomic surveillance network will help to reduce the impact of the current pandemic, while also laying the foundation for a future system that protects people from a range of threats, including future pandemic viruses, antibiotic resistant bacteria, and life-threatening fungal infections. It also will become a critical hub in a global genomics surveillance system that detects, tracks, and mitigates the spread of pathogens anywhere in the world.

Participants

The Rockefeller Foundation is grateful to the following people who have contributed to this report. Some may differ with aspects of it or have stressed other matters of primary focus. All have contributed with the greatest sense of shared purpose at this time of national need.

Kristian G. Andersen
Professor, Scripps Research

Gregory Armstrong
Director, Advanced Molecular Detection Program
Centers for Disease Control and Prevention

Mara G. Aspinall
Professor of Practice
Arizona State University

Scott J. Becker
Chief Executive Officer
Association of Public Health Laboratories

Trevor Bedford
Associate Professor
Fred Hutchinson Cancer Research Center

Jason Blanton
Molecular Administrator
Bureau of Public Health Laboratories, Florida
Department of Health

David Blazes
Deputy Director, Surveillance
Bill and Melinda Gates Foundation

Rick Bright
Senior Vice President,
Pandemic Prevention &
Response
The Rockefeller Foundation

J. Rodney Brister
Chief, Viral Resources,
Information and Engineering
Branch
National Center for
Biotechnology
National Library of Medicine
National Institutes of Health

Liliana L. Brown
Director, Office of Genomics
and Advanced Technologies
Division of Microbiology and
Infectious Diseases
National Institute of Allergy and
Infectious Diseases
National Institutes of Health

Cristina Casetti
Deputy Director
Division of Microbiology and
Infectious Diseases
National Institute of Allergy and
Infectious Diseases
National Institutes of Health

Jessica Dymond
Assistant Program Area
Manager, Health Protection and
Assurance, National Health
Johns Hopkins University
Applied Physics Laboratory

David M. Engelthaler
Director, Pathogen Genomics
Division, Translational Genomics
Research Institute

Phillip G. Febbo
Chief Medical Officer & Senior
Vice President, Illumina

Marta Fernandez-Suarez
Senior Director of R&D
FIND (Foundation for Innovative
New Diagnostics)

Rachael Fleurence
Special Assistant to the
NIH Director for COVID-19
Diagnostics
National Institutes of Health

Matthew Frieman
Associate Professor
Department of Microbiology
and Immunology
University of Maryland, School
of Medicine

Dylan George
Vice President, BNext
In-Q-Tel

Gardiner Harris
Independent Consultant
The Rockefeller Foundation

Zia Khan
Senior Vice President,
Innovation
The Rockefeller Foundation

Moritz Kraemer
Research Fellow
University of Oxford, United
Kingdom

Florian Krammer
Professor
Icahn School of Medicine at
Mount Sinai

Greg Kuzmak
Manager, Health
The Rockefeller Foundation

Bruce Lowry
Senior Advisor
Skoll Foundation

James Lu
Co-Founder, President
Helix

Duncan MacCannell
Office of Advanced Molecular
Detection
Centers for Disease Control and
Prevention

Bronwyn MacInnis
Director of Pathogen Genomic
Surveillance
Broad Institute of MIT and
Harvard

Sebastian Maurer-Stroh
GISAID Initiative

Jon Meltzer
Senior Vice President
Specialty and Esoteric Testing
Labcorp

Martha Nelson
Scientist
National Institute of Allergy and
Infectious Diseases
National Institutes of Health

Kelly F. Oakeson
Chief Scientist: Bioinformatics &
Next-Generation sequencing
Utah Public Health Laboratory

Kim D. Pruitt
Chief, Information Engineering
Branch
National Center for
Biotechnology Information
National Library of Medicine
National Institutes of Health

Jonathan Quick
Managing Director
The Rockefeller Foundation

Naveen Rao
Senior Vice President, Health
The Rockefeller Foundation

Caitlin Rivers
Senior Scholar
Johns Hopkins Center for
Health Security

Pardis Sabeti
Professor
Harvard University & The Broad
Institute

Rajiv Shah
President
The Rockefeller Foundation

Stephen Sherry
Acting Director
National Center for
Biotechnology Information
National Library of Medicine
National Institutes of Health

Viviana Simon
Professor
Department of Microbiology
Icahn School of Medicine at
Mount Sinai

Mark Stevenson
Executive Vice President and
Chief Operating Officer
Thermo Fisher Scientific

Matthew Stremmlau
Founder and Chief Executive
Officer, Equator Labs &
Independent Consultant
The Rockefeller Foundation

Marc A. Suchard
Professor, University of
California, Los Angeles

Anita Suresh
Head of Sequencing
Programme
FIND (Foundation for Innovative
New Diagnostics)

Eric Topol
Gary & Mary West Endowed
Chair of Innovative Medicine
Executive Vice President
Scripps Research

Debra A. Wadford
Chief, Viral and Rickettsial
Disease Laboratory
Infectious Diseases Laboratory
Branch
California Department of Public
Health

Lane Warmbrod
Senior Analyst
Johns Hopkins Center for
Health Security

Rachel West
Post-doctoral associate
Johns Hopkins Center for
Health Security

Michael Worobey
Professor and Head
Department of Ecology and
Evolutionary Biology
University of Arizona

Kelly Wroblewski
Director, Infectious Disease
Association of Public Health
Laboratories



The
ROCKEFELLER
FOUNDATION

rockefellerfoundation.org