Accelerating National Genomic Surveillance

The ROCKEFELLER FOUNDATION

Foreword

One secret weapon has helped beat every disease outbreak over the last century; but it is not masks or social distancing, lockdowns, or even vaccines. Instead it is data. Data tells us which masks work, how far is far enough to socially distance, whether lockdowns are necessary or even working, and who is immune. In all, data is what moves us from a panic-driven response to a science-driven one, telling us how to fight back and which tools are best.

Too often, however, the world learns to value data too late and at too high of a cost. When I oversaw the U.S. response to the West African Ebola crisis, we began with incomplete data and did not gain ground until we could better understand the basics: who was positive and where? As Covid-19 swept the world one year ago, many countries, including the United States, under-prioritized the need for data and the tests that produce it. It is why The Rockefeller Foundation worked with partners since the outbreak's earliest days to develop, produce, and field the tests needed in the United States and abroad.

The data-deficient response to Covid-19 is why this pandemic's been so costly. The human toll has been staggering: more than 2.5 million killed and at least 115 million infected at the time of this report's publication. The pandemic, and resulting lockdowns meant to bend its curve, have cost the global economy trillions of dollars in growth. More than just economics, the long-term impacts are likely to disproportionately impact people of color, minorities, and vulnerable populations. In addition, millions of school children in America have been forced to learn virtually since the pandemic's start, diminishing educational outcomes while increasing absences, hunger, abuse, and more.

Even as we use data to finally beat this pandemic, we must think about the data needed to prevent the next pandemic, whether it is the result of a SARS-CoV-2 variant or another virus altogether.

Unfortunately, right now, the United States and most of the rest of the world are in no better position to stop a variant from going global today than they were before the pandemic started. Currently, only a handful of countries have analyzed more than 5 percent of their Covid-19 cases. In the United States, most localities are sequencing less than 1 percent of cumulative cases. What little genomic data we are collecting, is not being analyzed or shared fast enough to help public health authorities and scientists make informed decisions about relaxing precautions or adapting vaccines and treatments.

This report—based on a recent convening of scientists, lab administrators, public health officials, and entrepreneurs—provides a blueprint for dramatically expanding genomic surveillance in the United States. By amplifying warning signals and sharing information and best practices, the system presented here could save countless lives and billions of dollars by helping to forestall new variant-driven surges. It could be deployed against more routine—but no less devastating—illnesses as well.

This report is also a first step. The Rockefeller Foundation will invest part of a recent $1 billion commitment to incubate a broad, data-driven platform so the world can better anticipate, visualize, and respond to future outbreaks with pinpoint precision. We look forward to working with partners in government, the academia, and the private sector, in the United States and abroad to develop, analyze, and share the data needed today as well as tomorrow.

Onward,

Dr. Rajiv J. Shah,
President of
The Rockefeller Foundation

The United States has been flying virally blind. By the time the first cases of Covid-19 were confirmed last year, the pandemic was already spreading uncontrollably. Outbreaks tore undetected through New York City, Boston, San Francisco, Chicago, and Seattle.

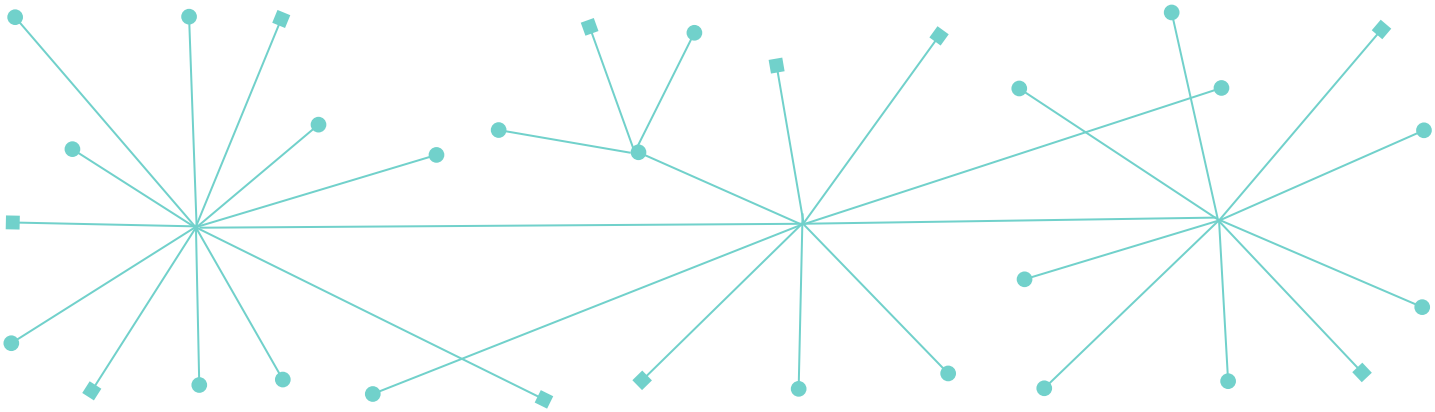
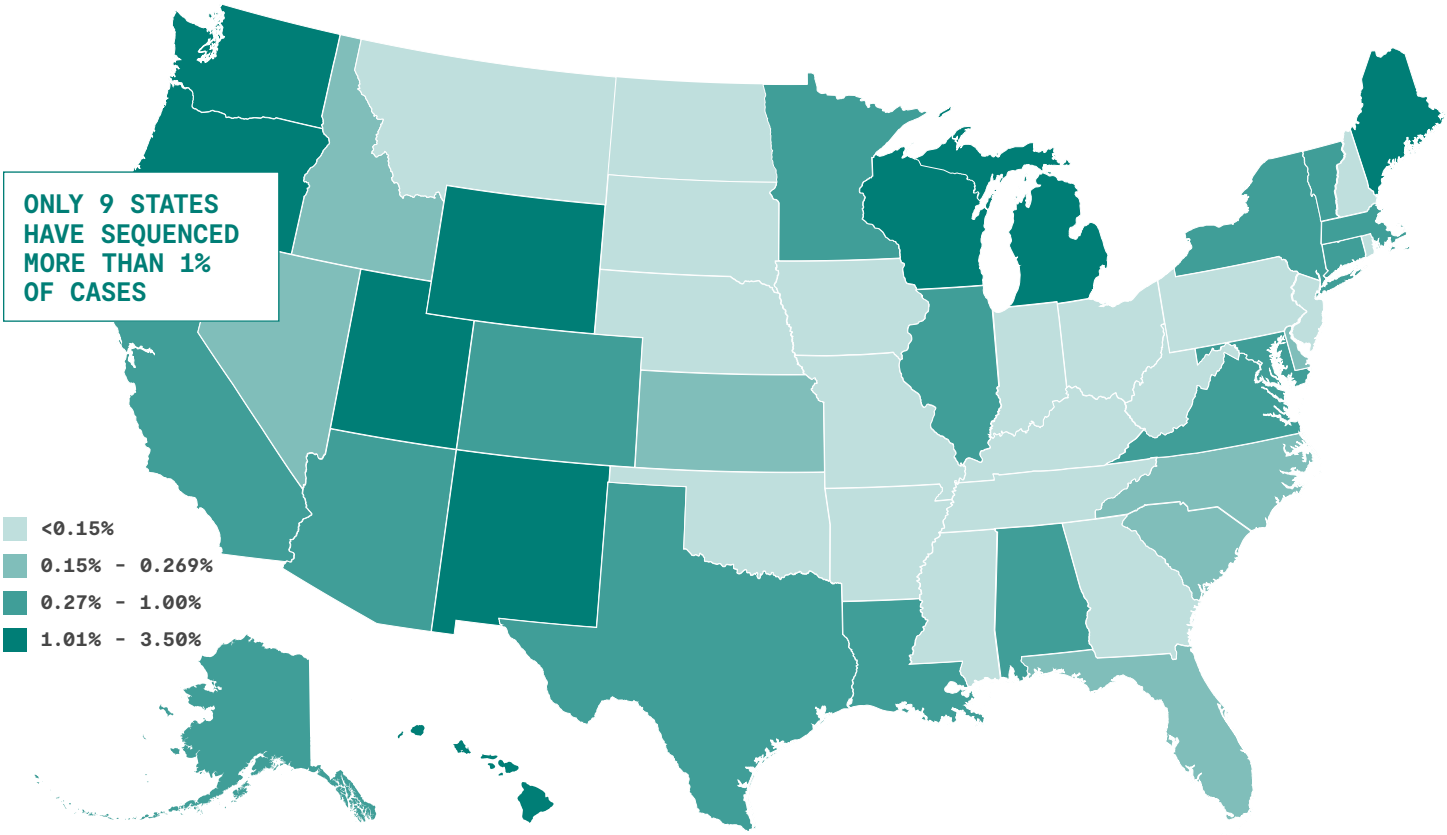


FIGURE 1 **Proportion of SARS-CoV-2 Cases Sequenced in the United States**



Source: CDC (cdc.gov) as of February 28, 2021

A more contagious variant of the coronavirus is now spreading rapidly and mostly invisibly throughout the country. Reports are emerging of additional new and very worrisome variants in the U.S. If authorities could track variants, sophisticated mitigation firebreaks could be deployed, and manufacturers could quickly create targeted multi-variant vaccines and alternate therapeutics.

While the U.S. has made significant strides in recent weeks, the country must accelerate efforts to increase SARS-CoV-2 genomic surveillance. Since the start of the pandemic, the United Kingdom has sequenced and submitted more than 300,000 cases, while the U.S. has sequenced and submitted less than half that number for a much larger population.

Until the U.S. is able to boost genomic surveillance, economically devastating shutdowns remain among the few effective tools to check rapid variant spread—an expensive and disruptive sledgehammer when a scalpel is needed.

As a country, the U.S. must take crucial steps to support a national viral genomic surveillance network to defeat Covid-19 and prepare for the next pandemic.

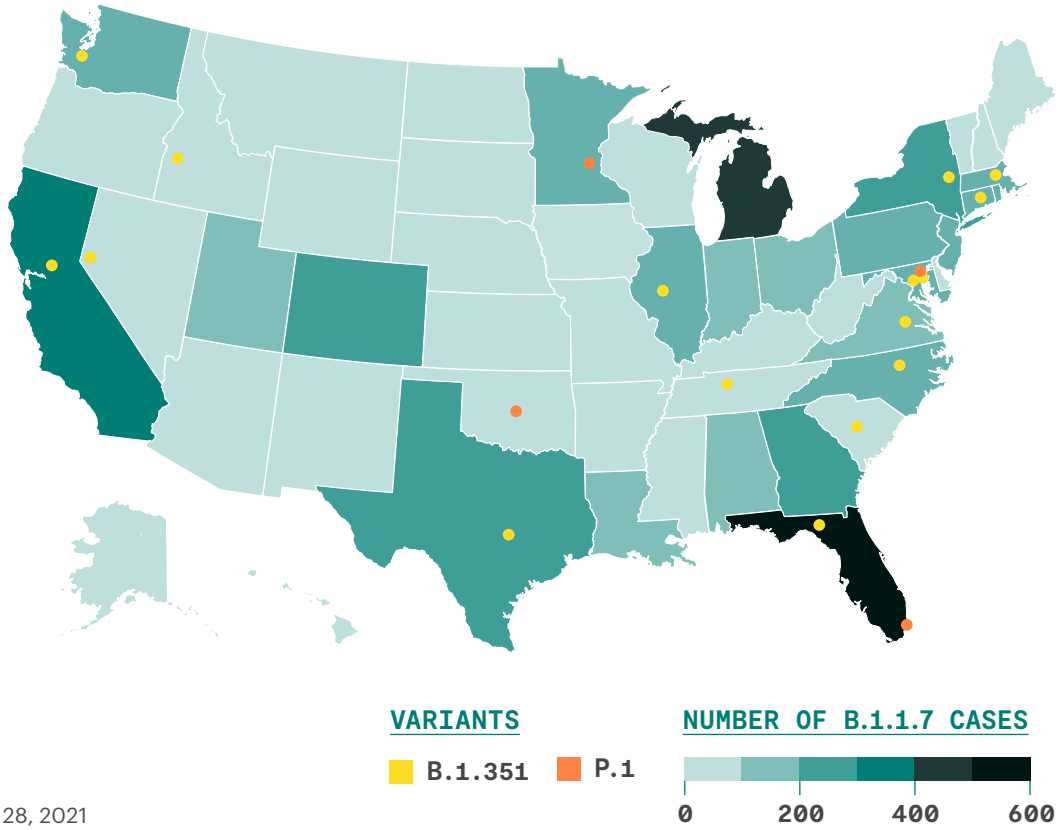
The following plan is part of a larger commitment by The Rockefeller Foundation to incubate a global data-driven, pandemic prevention institute. The goal is to have an early alert system to increase awareness of a newly emerging or spreading pathogen before it has an opportunity to spread undetected. Such a system could save millions of lives and trillions of dollars, by leveraging rapid detection for effective mitigation planning and response activities. And it could be deployed not only to track the rare, yet potentially catastrophic, pandemic pathogen, but also to enhance current influenza monitoring and contain the growing scourge of antimicrobial and antifungal resistance.

FIGURE 2 Cases By SARS-CoV-2 Variant in the U.S.

B.1.1.7
Associated with increased transmissibility and preliminary evidence suggests it is 35% more deadly than other variants. First identified in the United Kingdom in September, 2020.

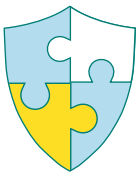
B.1.351
Associated with reduced immune response to some vaccines and treatments. First identified in South Africa in October, 2020.

P.1
Contains a set of mutations that may affect its ability to be recognized by antibodies. First identified in Japan, among travelers from Brazil, in early January 2021.



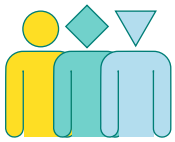
Source: CDC (cdc.gov) as of February 28, 2021

Action Plan to Accelerate Genomic Surveillance in the US



1. Build a viral defense system

Create sentinel sites around the country that detect a new virus or new variant as soon as it appears. Set basic protocols for specimen and data collection, sequencing, and structural genotypic and functional phenotypic analysis. Address barriers that prevent the diverse set of players from effectively cooperating. Connect this system to a global system that shares data and analytic outcomes so threats can be detected anywhere in the world.

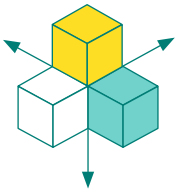


2. Be diverse and deeply human

Black, Hispanic, and Native American. The old, those with chronic conditions, and the immunocompromised. Covid-19 has hit such groups particularly hard, and tracing its uneven effects is vital to an equitable recovery. Ensuring genomic surveillance benefits all Americans by analyzing samples representative of all Americans. Also, ensuring sequencing of highly atypical clinical cases.

3. Connect the dots

The Tower of Babel that characterizes the nation's electronic medical records cannot be allowed to curse this effort. Critical epidemiological and clinical data is often impossible to access and usually not connected to viral genomic data. A national standard for the collection, aggregation, and sharing of risk-related medical information that safeguards privacy and adheres to the law must be created. A network that links these disparate datasets is needed.

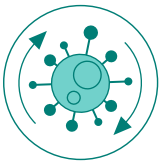


4. Provide more to do more

Expanding genomic sequencing will require more people, sequencing equipment, computers, reagents, sample storage, and a host of other supplies. This cannot be done on the usual public health shoestring. Although the Biden administration's $200 million down payment is a good start, much more support is required to establish, expand, and sustain a system designed to stay in front of a rapidly mutating and spreading virus.

5. Build new analytic tools and make them visual

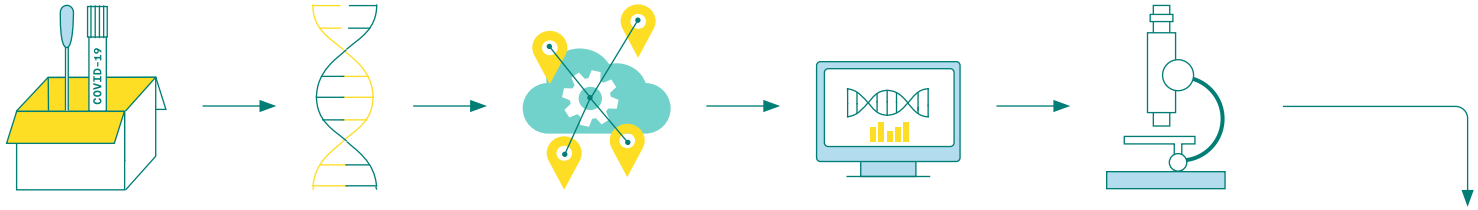
New software tools for analyzing genomic data that are easy-to-use and scalable are urgently needed. Visualization tools must be built and distributed to help state and local officials and eventually national officials around the world to understand and explain viral spread and threats.



6. Rapidly assess the threat posed by new variants

When significant mutations occur, there is a need to understand their biology at the molecular level and determine if these changes in the virus make it more transmissible, more virulent, or impact the effectiveness of vaccines and therapeutics. This research requires storage of more samples and improved access to viral isolates.

FIGURE 3 Genomic Surveillance System



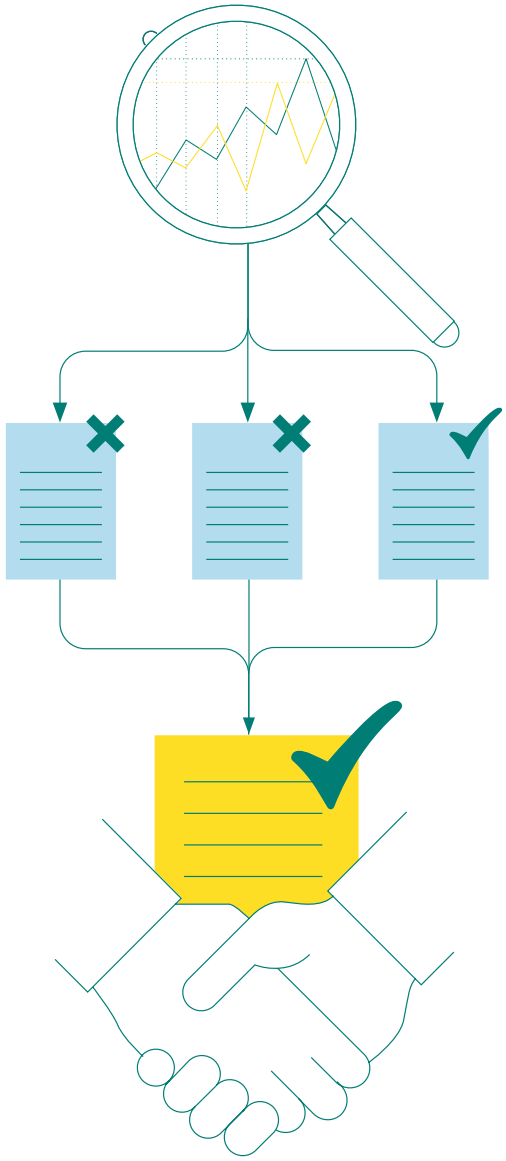
PILLAR	SAMPLE INTAKE	SEQUENCING	DEPOSIT DATA	GENOMIC ANALYSIS	BIOLOGICAL ANALYSIS
Implementing Partners	Hospitals, Clinics, Public Health Labs, Testing Companies	Federal Labs, State Labs, Universities, Companies	Public Databases Restricted Databases	Federal Labs, State Labs, Universities, Companies, Databases	Federal Labs, State Labs, Universities, Companies
Major Activities	Collect samples Collect clinical and epidemiological data	Process sample Sequence sample Assemble raw sequence data	Store genome sequences Quality control Store clinical and epidemiological data	Bioinformatic and computational analysis Visualize data	Lab-based experiments Clinical and non-clinical studies
Data	Clinical and epi-demiological data ('metadata')	Genome sequences		Variants Genetic relationships Transmission	Vaccine and therapeutic effectiveness Virulence Transmission

Communication
Share information

Evidence informs clinical and public health decisions

Adapt diagnostics, vaccines and therapeutics

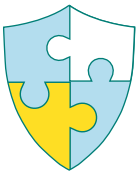
Developing the Action Plan



In February 2021, The Rockefeller Foundation held a virtual conclave of scientists, lab administrators, public health officials, and executives to discuss expanding genomic surveillance in the U.S.. Several common barriers emerged, including the need to access clinical and epidemiological data; standardize data and information to encourage sharing and integration; share viral and clinical specimens for functional analysis; and solve communication problems between data generators and users. Importantly, the sense of improved urgency for each of these steps to occur was repeatedly emphasized to optimize the flow of information to effect a rapid response.

Creating a national consortium for genomic surveillance will require leadership, collaboration, funding, and sustained commitment. Such projects normally take years. This one is needed within weeks. Urgency mandates that existing capacities be promptly mobilized and conclave members had a host of suggestions for rapid expansion. The immediate payoff for these actions will be improved efforts to reduce the spread of dangerous variants. The long-term reward will be a foundation for a global surveillance system that prevents or ameliorates future pandemics and might also trace multi-drug resistant bacterial and fungal infections that are a growing menace.

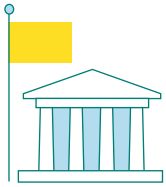
Building on insights derived from the convening, a consortium of cross-sector actors and institutions could help to resolve key issues in this U.S. genomic surveillance space. Potential action items are outlined in the following pages.



1. Building A Viral Defense System

A national surveillance network that coordinates public, academic, and industry partners to collect, share, and analyze genomic and phenotypic data and information is urgently needed. This network's goal should be to rapidly scale genomic surveillance activities so emerging variants can be identified and assessed and their impact blunted. This network must be capable of transmitting information and guidance to public health officials so they can act quickly.

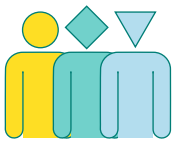
This national surveillance network should:



I. Establish national leadership that sets priorities and standards for data collection, sequencing, analysis, sharing, and assessing potential threats posed by new variants.

II. Address barriers that prevent the collection and sharing of genome sequences and viral isolates and ensure national-scale genomics, paired with high-resolution rapid sequencing, at state and local levels.

- **Improve communication** and coordination among all partners involved in genomic surveillance through regional centers and national working groups so policymakers and public health officials can act quickly to mitigate the impact of emerging variants.
- **Fold U.S. efforts into a global genomic surveillance network** to identify, track, and contain emerging viruses before they spread globally.

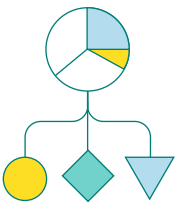


2. Be Diverse and Deeply Human

Covid-19 has had a profoundly unequal impact. Minority populations, including Black, Hispanic, and Native American communities, have been especially hard hit, as have the elderly, those with chronic conditions, and the immunocompromised. Much of the available data—whether genomic, epidemiological, or clinical—is not representative of these groups. In certain cases, these data are biased; in others, they are simply missing. Consequently, important questions about how the virus affects these populations go unanswered.

Opportunities to ensure equal representation:

I. Viral samples should be representative of all groups, including underserved and vulnerable populations. In particular, the country needs to ensure samples are collected from high-risk groups that rarely interact with the healthcare system, such as undocumented immigrants. Additional resources are needed to ensure that every region of the U.S., including regions with less genomic surveillance capacity, are adequately surveyed, analyzed, and represented in relevant databases.

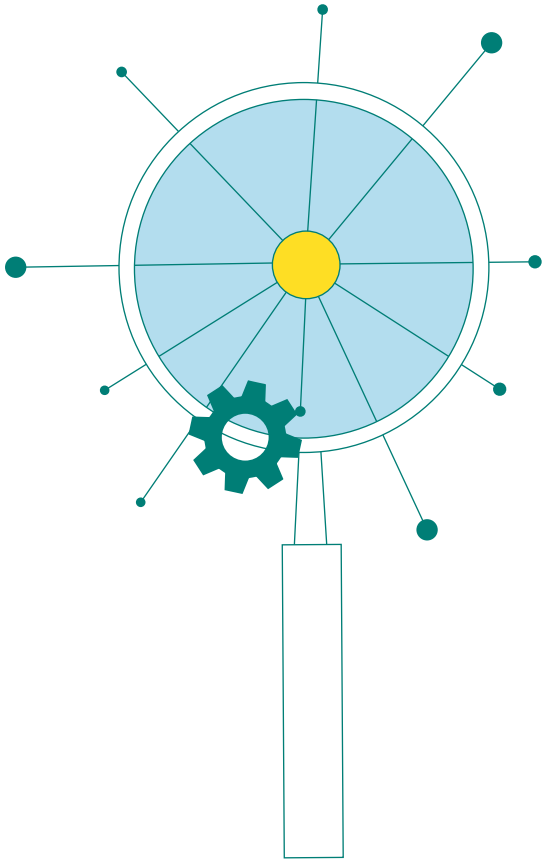


II. Clinical and epidemiological data must also be representative of America's diversity. These data must be collected from individuals from every race, ethnicity, gender, age, and geography. They must contain socioeconomic factors such as housing type. The data must be protected and safeguarded to ensure privacy and compliance with all federal, state, and local laws.

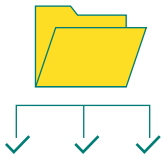
3. Connect the Dots

The frustrating complexity and diversity of electronic medical records is an ongoing challenge in the clinical setting, and this Tower of Babel has established a beachhead in genomic surveillance. Basic information—such as age, vaccine status, and clinical symptoms—is often not linked to genome sequences, genomic analysis, or phenotypic characterization.

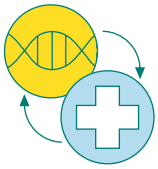
Several barriers prevent integration. Personal and clinical information is sensitive and often includes components protected by state and federal laws. These data must be handled differently from impersonal genomic and phenotypic information. The resulting restrictions make the collection, storage, and transmittal of such information fraught. So, each research effort often involves untying a unique Gordian knot.



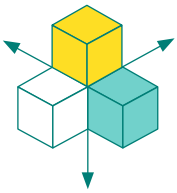
Opportunities to improve access and integration of data:



I. Increase access to clinical and epidemiological data. A minimal set of standard questions must be adopted along with simple guidance about collection reporting and use of information. Existing data stores, such as the National Center for Biotechnology Information (NCBI) and the GISAID Initiative should be strengthened. Additionally, new data stores should be created that are specifically designed to protect clinical and epidemiological data. When data cannot be transferred to researchers for analysis, new tools should be created that go to the data, while respecting the terms of access and sharing of data for a designated use. Analyzing clinical and epidemiological data shouldn't just be the province of research universities, hospitals, and clinics should be empowered to explore it as well.



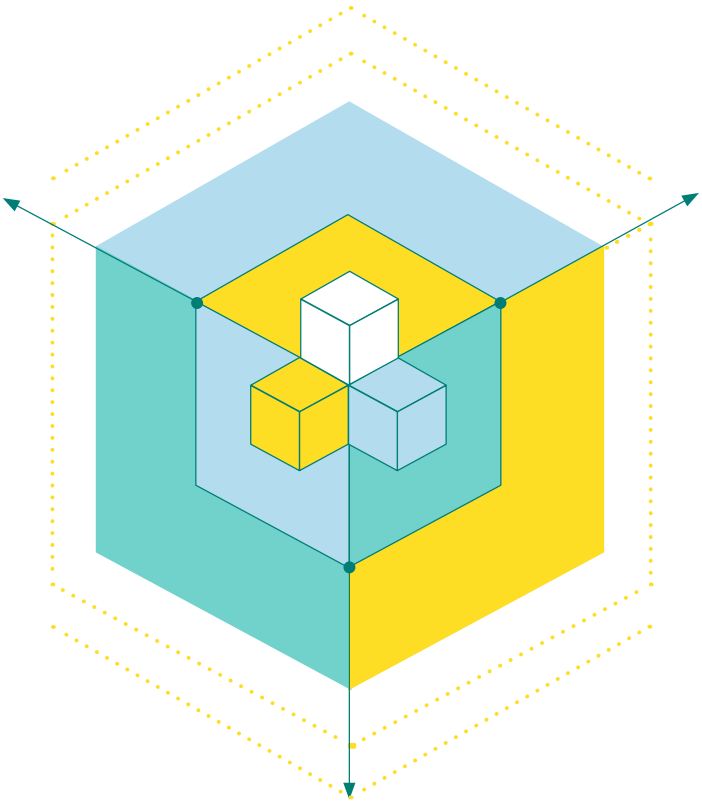
II. Establish a network that is able to link clinical and epidemiological data to genomic data. Combining diverse data sets, some of which contain risk-related medical data, is nearly impossible. A network capable of linking a variety of independently governed data systems would enable a more comprehensive analysis of an emerging pathogen and its impact and would add great value to pandemic risk assessment and mitigation planning. This network needs to work closely with national, state, and local authorities to ensure critical information about the virus is gathered while patients' privacy rights are protected. This network should have a precise mandate regarding the reports it produces and provide actionable information to public health officials and policymakers.



4. Provide More to Do More

Several barriers impede the scaling of surveillance. The first has been historically limited and patchwork funding for surveillance and analytical capacity. Although billions of dollars have been directed to clinical diagnostic testing, relatively little money has been allocated to genomic surveillance systems. This is short-sighted since the costs of genomic surveillance are a fraction of the economic costs of finding and tracking emerging variants that can accelerate the spread, increase the severity, or blunt the effectiveness of vaccines and treatments.

Limited workforce capacity, especially in state public health labs, also impedes the scaling and responsiveness of pathogen sequencing and reporting. The limitation of trained technicians constrains the sequencing and sharing of genomic data. Many technicians in public health labs have shifted to performing clinical diagnostic tests. Also, some state labs may not have sufficient expertise in bioinformatics to process and analyze data in a timely manner.



Opportunities to scale surveillance:

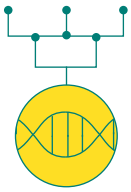
I. Increase funding. The Biden administration has recognized the value of and need to enhance national genomic surveillance. A recent $200 million down payment to jumpstart sequencing is a good start. Much more will be needed to sufficiently expand and sustain an effective surveillance system. One option to consider is to offer pre-paid contracts for a fixed number of sequenced samples from state public health labs, universities, and private companies. These contracts can mandate specific turnaround times, ensure appropriate representation of all populations, and require information be shared by depositing sequences in designated databases. The U.S. must ensure that increases in funding don't just produce more data—but instead produce high-quality, actionable data that serves public health.

II. Strengthen workforce capacity. To rapidly train public health laboratory experts, fellowship programs offered through the Centers for Disease Control (CDC) and the Association of Public Health Laboratories (APHL) should be expanded. Programs that place professionals qualified in sample processing, sequencing, and raw data processing into public health laboratories should also be expanded and further developed.

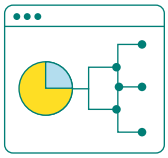
5. Build New Analytic Tools and Make Them Visual

A national genomic surveillance effort will generate thousands of terabytes of raw data. These data must be analyzed and then communicated to public health officials. Unfortunately, many analytic tools are not user-friendly and cannot handle mountains of data. After data are analyzed, they are often not presented in a form that is intuitive to non-experts.

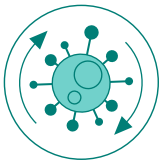
Opportunities to advance genomic data analysis:



I. Establish analytic priorities and standards for genomic data. Priorities for minimum analysis could include reporting of viral lineages and mutations. They could also include relationships linking samples from different geographies or clinical outcomes. Standards should describe optimal parameters for quality control so datasets from a diverse array of sources can be compared.



II. Create analytic tools that are easy-to-use, scalable, and capable of revealing new insights. Software tools to analyze genomic data must be easy-to-use, even by analysts with limited viral genomics experience, and capable of handling a high volume of genomic data. They should be freely available and overseen by a dedicated group of software engineers that updates the tools. These tools need to serve national surveillance priorities; thus, they should focus on generating information and reports that are understandable by a diverse audience with varying levels of expertise.



6. Rapidly Assess the Threat Posed by New Variants

The evolution of a virus will continually create new variants of concern that may impact its infectiousness, deadliness, and immune response to vaccines, treatments, or natural infections. Assessing the threat posed by such new variants is vital to determine when vaccines, therapeutics, or lab tests need to be modified or targeted to specific regions. It is also essential to inform clinical and public health decisions.

A complete understanding of the threat posed by new variants requires lab-based experiments and evaluation in non-clinical and sometimes clinical models. But viral and clinical samples are often difficult to share and ship. Challenges of getting samples from those infected to measure immune responses can be immense.

Opportunities to accelerate threat assessment:

I. Improve access to clinical samples and viral isolates. An intense effort of scientific diplomacy is needed to improve cooperation and access both domestically and internationally. Standard material transfer agreements must be created and accepted. Common storage protocols should be created for high-priority clinical samples, including serial specimens from patients with extended infections, re-infections, or those for whom vaccines failed.

II. Increase investment in international laboratory capacity and collaboration. Create opportunities so functional analysis can be performed in countries where viruses emerge or variants are detected.

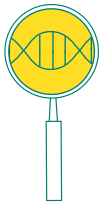
Measuring success

A national genomic surveillance system's goal should be to detect, track, and stop viral threats rapidly. Continuously monitoring and evaluating this system ensures it is meeting this goal. Several metrics can be used to gauge success, including:

- the number of SARS-CoV-2 positive cases sequenced; ensuring they are representative of regional, racial, and ethnic diversity and reflective of the dynamic prevalence of the virus in the community;
- the time required from obtaining a sample to depositing a high-quality genome into an appropriately designated database;
- the time required to perform baseline genomic analysis, such as identifying novel variants and determining viral lineage;
- the time required to assess the pathogen's threat, such as functional phenotypic assessment, to understand pathogenicity and potential impact on existing vaccines, diagnostics, and treatments.

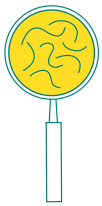
Building a global surveillance system

Although the country and global community must move quickly to increase genomic surveillance to end the SARS-CoV-2 pandemic, there is significant benefit to expanded genomic surveillance of other pathogens. The coordinating structures strengthened as part of enhancing SARS-CoV-2 genomic surveillance will also be useful for future threats. Ideally, the partnerships formed to support SARS-CoV-2 genomic surveillance will be sustained and leveraged to:



USE GENOMIC SURVEILLANCE TO PREVENT FUTURE PANDEMICS BY RAPIDLY IDENTIFYING EMERGING PATHOGENS THAT MAY POSE SERIOUS THREATS;

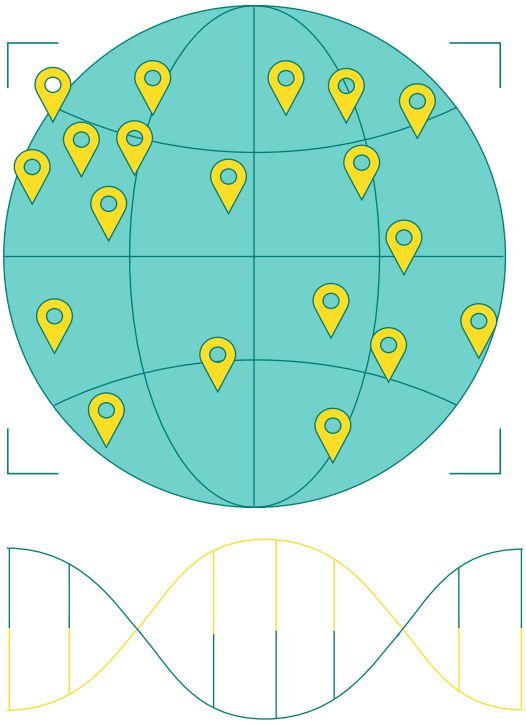
CONTINUALLY ASSESS VACCINES, THERAPEUTICS, AND OTHER MEDICAL COUNTERMEASURES FOR ENDEMIC AND SEASONAL PATHOGENS TO DETERMINE WHEN UPDATES ARE NEEDED TO MAINTAIN THEIR EFFECTIVENESS;



ESTABLISH GENOMIC SURVEILLANCE FOR A BROAD RANGE OF PATHOGENIC THREATS, SUCH AS ANTIBIOTIC-RESISTANT BACTERIA, LIFE-THREATENING FUNGI, AND AGRICULTURAL PESTS.

Toward global surveillance

A robust national surveillance system is essential to detect and respond to emerging pathogens within the U.S. However, this alone is not sufficient to provide security against emerging pandemic threats. Global participation and coordination are vital to protect people in every nation from emerging threats. Outbreaks can happen anywhere in the world. Rapid globalization, urbanization, and climate change mean that outbreaks of pathogens will occur more frequently and spread more quickly. Rapidly detecting and responding to these outbreaks requires a coordinated global partnership. A robust U.S. surveillance system can serve as a key partner in a global system and ensure emerging viral threats don't become the next pandemic.



The ROCKEFELLER FOUNDATION

rockefellerfoundation.org