McKinsey & Company

Healthcare Systems and Services

Genomic sequencing: A vital tool in combatting the pandemic

As variants of the COVID-19 virus emerge, genomic sequencing could help boost the public health response and provide the impetus to create value for future crises.

This article is a collaborative effort between Matt Craven, Ashka Davè, Laura Furstenthal, Rohit Kumar, and Hrishika Vuppala representing the views of McKinsey's public & social sector, healthcare, and pharmaceuticals & medical products practices.

March 2021

Since the beginning of the COVID-19 pandemic, genomic sequencing efforts, particularly in the United States, have fallen behind. While the federal government is taking measures at the national level to increase sequencing for COVID-19, it is equally important that states build on recent momentum to establish their own infrastructure to ensure a timely and precise local response to sequencing data.

The United States has sequenced 200,000 cases, whereas the United Kingdom has sequenced 350,000.[1] US states have been consistently undersequencing, making it harder to take targeted public health action, react to variants, and efficiently contact trace. With the influx of federal funding, states have an opportunity to build a sequencing utility that not only helps address COVID-19, but also builds the sequencing infrastructure for a more effective future public health system. Additionally, states may consider developing a sequencing network across existing diagnostic and sequencing labs to ensure adequate surveillance, building a local database to aggregate data for timely public health action, and convening a corps of experts to interpret sequencing data for local response.

The critical role of sequencing data

Without sequencing data, public health authorities are blind to viral mutations that may reduce the efficacy of current response tools (for example, safety protocols, tests, vaccines, and clinical care) and are unable to take precise public health action. Sequencing is a powerful tool that enables public health authorities to improve precision, efficacy, and efficiency of public health response. This response includes the following initiatives:

Take informed public health actions. Genomic sequencing of SARS-CoV-2 in a given population can serve as an early warning system to guide the public-health response. The Netherlands sequenced its first positive samples in late February. Based on the analysis, the government introduced strict mitigation measures within weeks (for example, canceling large events, encouraging working from home, and closing schools).[2]

Conduct more efficient case investigations. Sequencing can help trace viral lineage, transmission, and rate of evolution. These insights, provided by data that can range from cases in a given facility to intercontinental transmission, are an important aid to outbreak investigations. In the case of New York City's initial outbreak in March 2020, once epidemiologists identified the probable date of introduction and the strain, public health leaders were able to estimate how many people in the area were infected. Such information is vital to support effective response.[3]

Gauge the real-time effectiveness of the public-health response and adjust accordingly. New variants, including B.1.1.7 (the "UK variant") and B.1.351 (the "South Africa variant") have highlighted the risk that new strains will increase transmissibility or reduce the efficacy of vaccines. This information has enabled a rapid pivot to the development of vaccine boosters effective against variants.[4]

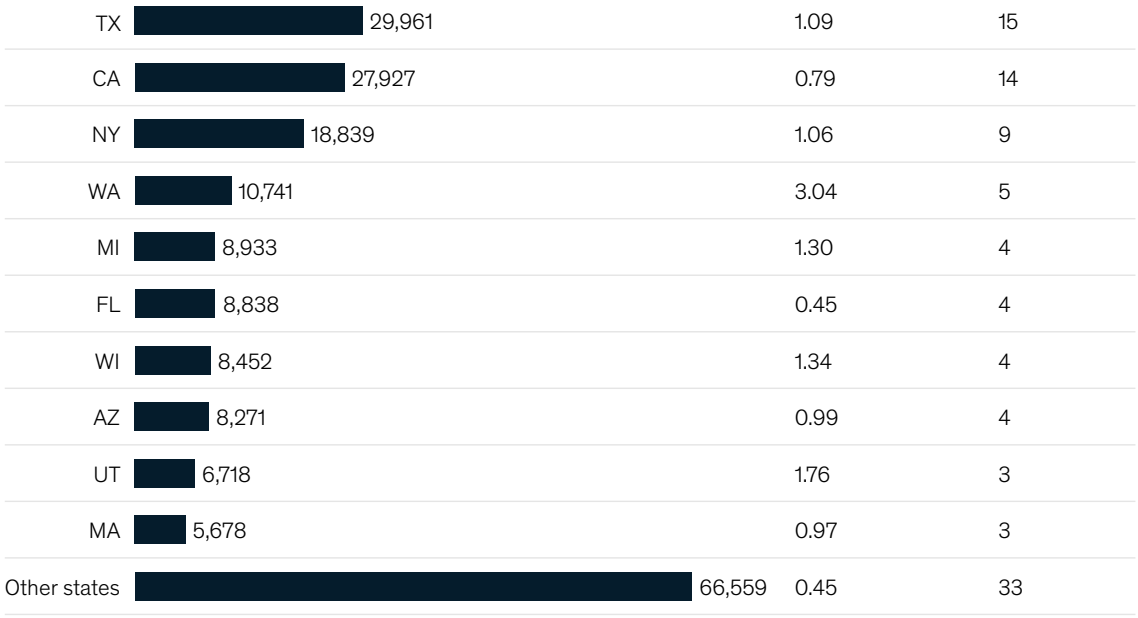
The United States, despite being home to a quarter of the world's cases, has sequenced less than half a percent of its COVID-19 samples (Exhibit). In contrast, the United Kingdom has sequenced 8 percent of its samples,[5] accounting for 45 percent of the world's sequences. As a result of its large volume of genome sequencing, the United Kingdom was quick to detect and act on a new, more transmissible variant of the virus in December. By contrast, the United States had little to no sequencing data to inform its response.

A state-wide genomic sequencing infrastructure not only helps address COVID-19 today but is also a tool that enhances

Exhibit

The United States has sequenced 0.71% of all positive cases.

State breakdown, March 21, 2021



	Cumulative cases sequenced, %	Portion of US sequences, %
TX	1.09	15
CA	0.79	14
NY	1.06	9
WA	3.04	5
MI	1.30	4
FL	0.45	4
WI	1.34	4
AZ	0.99	4
UT	1.76	3
MA	0.97	3
Other states	0.45	33

Source: CDC National Genomic Surveillance Database (3/21/2021)

public health effectiveness in the long-term (see Sidebar 1, "The four components of sequencing programs"). Sequencing can be used for more efficient source identification of foodborne disease (for example, norovirus, *E. coli*, *Salmonella*, and Hepatitis A), ascertaining the origins of 'outlier' cases—those with no known connection to other cases—to contain further outbreaks (as was done in the latter phases of the 2014–16 Ebola outbreak in Guinea),[6] and developing diagnostics and vaccines for novel viruses (as was the case with Zika in May of 2015).[7]

The role of US states in expanding genomic sequencing

In an effort to catch-up, the American Rescue Plan of 2021 directs $1.75 billion to the Centers for Disease Control and Prevention (CDC) to support genomic sequencing and surveillance initiatives. While bolstering national surveillance is vital, in order for states to take targeted and timely action, the turnaround time from sample collection to sequence must be short and the sequencing data must be easily paired with patient metadata. To do this effectively, local sequencing is imperative.

Substantive funds have been made available to states to establish these sequencing programs. The Coronavirus Response and Relief Supplemental Appropriations Act of 2021 provides an additional $19.1 billion to states to strengthen laboratory testing by expanding the use of SARS-CoV-2 genomic sequencing and molecular epidemiology for state and local surveillance and response.[8]

Public health officials have options for how to direct funds. Efforts could include rapid sequencing and analysis of SARS-CoV-2 genomes by contractors and staff within public health laboratories, enhanced work-

force or bioinformatics capabilities (including improved access to cloud computing resources), or the establishment or expansion of partnerships with academia and the private sector (see Sidebar 2, "California's genomic sequencing partnership").[9]

States have several options for establishing a sequencing capability. The right archetype for each state depends on its existing infrastructure and level of coordination.

Matchmaker. State officials can use existing surveillance systems and data to identify the appropriate diagnostic lab partners and connect them with existing sequencing capacity. Since September, the CDC has been providing funding to labs for sequencing,[10] but a missing link remains access to the right samples. As a matchmaker, the state would leverage state surveillance data to ensure appropriate geographic representativeness of sequenced samples and would need to partner with multiple labs to source samples. The states that are best positioned to play this role already have sufficient sequencing capacity

Sidebar 1

The four components of sequencing programs

Effective sequencing programs comprise four components: samples, sequencers, software, and staff. Many jurisdictions already have access to many or all of these components.

1. **Samples.** A sequencing program must have a systematic way to source the positive samples needed to maintain a timely and sufficiently comprehensive view of variants in a given area. Samples might be collected based on three use cases: surveillance samples that geographically represent a population, public health areas of concern such as outbreaks in congregate living facilities, and critically ill patients in hospital settings.
2. **Sequencers.** Positive samples must be routed to entities with sequencing platforms and bioinformatics capabilities. These platforms are typically available at universities, academic medical centers, and public health laboratories. Recent next-generation sequencing methods have substantially reduced the cost of sequencing a single genome, resulting in a marked increase in the number of sequencing platforms. Furthermore, the CDC has financed partnerships with several labs (for example, Labcorp and Quest) as well as sequencing-equipment manufacturers (such as Illumina) to boost the country's sequencing capacity.
3. **Software.** Free open-source software, such as NextStrain, enables the effective visualization of sequenced data to inform public-health action. However, organizations will need appropriate data infrastructure to receive, store, and transmit information about sequences and the associated patient metadata. A core part of the sequencing-data infrastructure is a database that stores sample metadata (including protected health information), manages sample inventory, streamlines the selection of samples to be sequenced, and generates reports.
4. **Staff.** Experts are needed to determine the necessary sample size for outbreak response and general surveillance and to train staff to interpret sequencing data for public health actions. These professionals will need to have computational and bioinformatics expertise to interpret sequencing data. Jurisdictions must also have a sufficient number of regional- and state-level laboratory personnel who can carry out genomic sequencing.

and a surveillance system that receives electronic reports from diagnostic laboratories; they also have the administrative capabilities to manage multiple diagnostic lab partners.

Sequencer. Some states have the ability to serve as the end-to-end sequencing provider, from receiving, testing, and storing diagnostic samples to sequencing the relevant positive samples. In these situations, the state serves as both the diagnostic and the sequencing lab. This functionality reduces operational burden and dependence on multiple lab partners for both supply of samples and sequencing capacity. States that already operate diagnostic-testing laboratories with sufficient coverage across communities and sequencing capabilities are a good fit for this role.

Convener of experts. In situations where diagnostic testing and sequencing are highly federated, states can function as a convener of data and experts to use genomic information for public health action. In this situation, a state would need to build a central database or repository that allows for the merging of patient metadata and sequencing data, and recruit experts to analyze genomic data.

How states can get started

States can follow four steps to assess their existing capabilities and select the appropriate archetype. The level of necessary investment in sequencing programs increases with each step.

1. **Start building data infrastructure.** Regardless of the archetype, a state must have a data infrastructure to connect sequencing data to patient metadata. States can consider establishing this capability immediately and begin integrating it with other public health reporting streams (for example, surveillance systems and contact tracing).
2. **Identify sequencing gaps.** By overlaying genomic sequences to corresponding geographies, states can determine whether they have sufficient sequencing coverage. This lens may also uncover local hotspots where additional sequencing may be needed.
3. **Begin establishing a network.** If coverage gaps exist, states should assess whether the bottleneck is caused by a failure to source samples or insufficient sequencing capacity. This evaluation can be conducted by reaching out to laboratories that contribute to GISAID

Sidebar 2

California's genomic sequencing partnership

COVIDNet is a collaboration between the state of California, county departments of public health, the Chan-Zuckerberg Biohub, the Invitae Corporation, and local academic and commercial laboratories. Through its early sequencing efforts, the state was able to identify a new variant, which seems to have originated in the state, spread quickly toward the end of last year, and now constitutes more than half the viral genome samples in Los Angeles. This same variant has also been detected in a few other states across the country.[1] The partnership estimates that the state will sequence 40,000 genomes over the next year.[2]

[1] Catherine Ho, "California and Zuckerberg lab team up to map coronavirus genome, one of the largest efforts to discover how it spreads," *San Francisco Chronicle*, July 30, 2020, sfchronicle.com.

[2] "Current genomic epidemiology efforts related to SARS-CoV-2," in *Genomic Epidemiology Data Infrastructure Needs for SARS-CoV-2: Modernizing Pandemic Response Strategies* (Washington, DC: The National Academies Press, 2020), 33–46.

(Global Initiative on Sharing All Influenza Data)[11] or are a part of the CDC's Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance (SPHERES) program and confirming their sequencing capacity. If sufficient capacity is in place, states will want to play a matchmaker role to connect this capacity to sources of positive samples.

4. **Expand sequencing capabilities.** To augment sequencing capabilities, states may need to partner with local, high-throughput diagnostic labs or augment public-health laboratory capabilities with sequencing.

All organizations—public, private, and nonprofit—have a vested interest in surveillance and genomic sequencing. The more samples that are analyzed, the more accurate and valuable the insights from this data. Further, the investments that entities make to build sequencing capabilities can help to address public-health events in the future. Stakeholders may consider the elements needed to enable effective genomic sequencing and then choose a path that makes the most of existing infrastructure and partnerships. Time is of the essence.

Matt Craven (Matt_Craven@mckinsey.com) is a partner in McKinsey's Silicon Valley office. **Ashka Davè** (Ashka_Dave@mckinsey.com) is a consultant in the Detroit office. **Laura Furstenthal** (Laura_Furstenthal@mckinsey.com) is a senior partner in the San Francisco office. **Rohit Kumar** (Rohit_Kumar@mckinsey.com) is a partner in the Chicago office. **Hrishika Vuppala** (Hrishika_Vuppala@mckinsey.com) is a partner in the San Francisco office.

The authors would like to thank Avani Venkateswaran and Gaurav Agrawal for their contributions to this article.

This article was edited by Elizabeth Newman, an executive editor in the Chicago office.

---

1 "Global sequencing coverage," GISAID, last updated March 26, 2021, covidcg.org.
2 Bas B. Oude Munnink et al., "Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands," *Nature Medicine*, July 16, 2020, Volume 26, pp. 1405–10.
3 Ana S. Gonzalez-Reiche et al., "Introductions and early spread of SARS-CoV-2 in the New York City area," *Science*, July 17, 2020, Volume 369, Number 6501, pp. 297–301.
4 "Moderna COVID-19 vaccine retains neutralizing activity against emerging variants first identified in the U.K. and the Republic of South Africa," Moderna, January 25, 2021, moderna.com.
5 Jaimy Lee, "Fear of new virus variant pushes U.S. toward more genomic sequencing," MarketWatch, last updated January 29, 2021, marketwatch.com.
6 Marta Gwinn, Duncan MacCannell, and Gregory L. Armstrong, "Next generation sequencing of infectious pathogens," *Journal of the American Medical Association*, March 5, 2019, Volume 321, Number 9, pp. 893–4, jamanetwork.com.
7 R.N. Faria et al., "Establishment and cryptic transmission of Zika virus in Brazil and the Americas," *Nature*, June 15, 2017, Volume 546, Number 7658, pp. 406–10, nature.com.
8 Department of Health and Human Services, "HHS announces $22 billion in funding to support expanded testing, vaccination distribution," Centers for Disease Control and Prevention, last reviewed January 7, 2021, cdc.gov.
9 "ELC enhancing detection through coronavirus response and relief (CRR) supplemental funds," CDC, January 12, 2021, cdc.gov.
10 National Center for Immunization and Respiratory Diseases, "Genomic surveillance for SARS-CoV-2 variants," CDC, last updated February 2, 2021, cdc.gov.
11 GISAID is a global science initiative and primary source, established in 2008, which provides open access to genomic data of influenza viruses and the coronavirus responsible for the COVID-19 pandemic.
