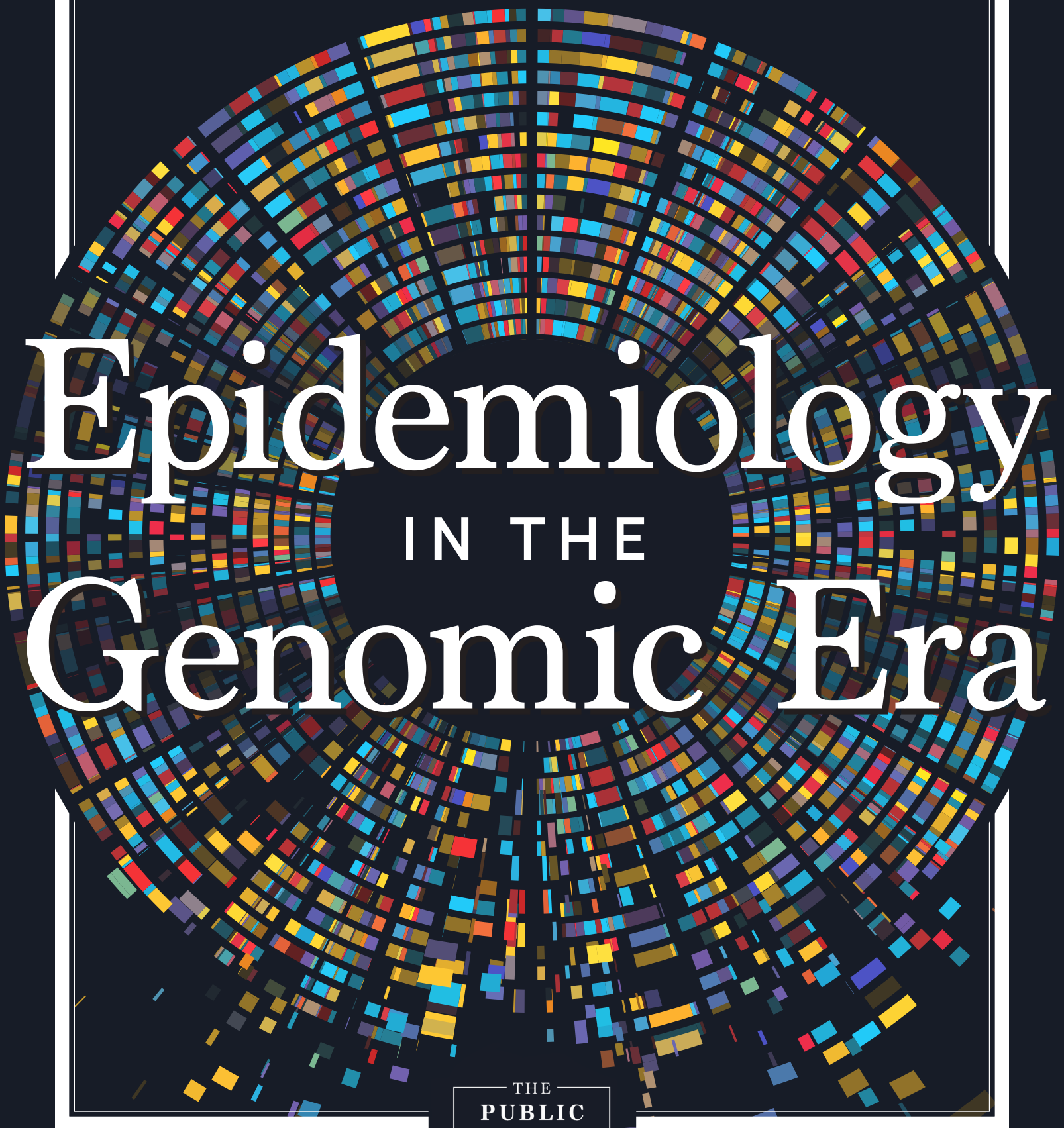
THE PUBLIC HEALTH COMPANY - WHITE PAPER



Epidemiology IN THE Genomic Era

THE PUBLIC HEALTH COMPANY

THE PUBLIC HEALTH COMPANY

The Public Health Impact of Genomic Data

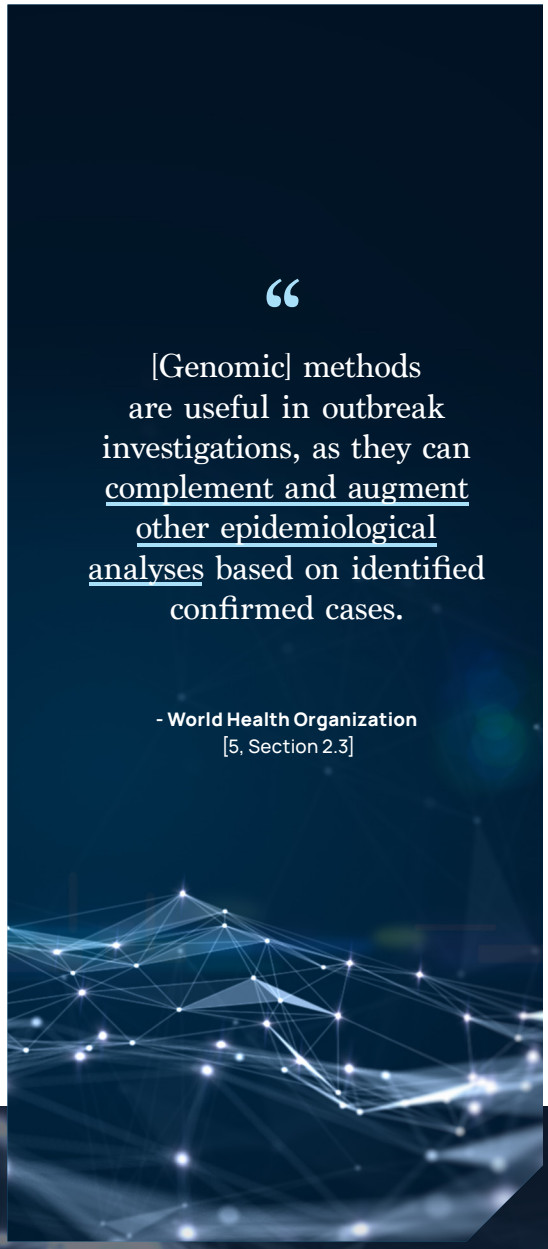
Abstract

In the COVID era, public health jurisdictions around the world have used SARS-CoV-2 genomic data in their pandemic response at unprecedented scale. As of April 23, 2021, over 1.2 million viral genomes have been sequenced for public health purposes. These data have enabled policy decisions (e.g., Denmark increasing interventions despite falling cases in response to detecting the UK variant), cluster investigations (e.g., uncovering transmission clusters across healthcare facilities that share staff), diagnostic test and vaccine development (e.g., establish diagnostic assays to detect the UK variant), and more.[1-5]

> “
>
> Recent advances have allowed the genomes of SARS-CoV-2—the causative agent of COVID-19—to be sequenced within hours or days of a case being identified. As a result, for the first time, genomic sequencing in real time has been able to inform the public health response to a pandemic.
>
> – **World Health Organization** [5, Section 1]

In this whitepaper we detail some of the ways that pathogen genomic data has had significant public health impact around the world during the COVID-19 pandemic. Pathogen genomic data is one part of a constellation of public health tools available. Basic interventions like masks, social distancing, contact tracing, and maintaining societal trust and compliance are vital to controlling any pandemic. However, genomic data provides an inexpensive high-quality data source that has been combined with traditional epidemiological data sources to turbocharge the whole spectrum of public health responses around the world.

Although the examples we discuss have demonstrated the significant impact genomic data has had on public health, this success came about largely through ad-hoc partnerships, last-minute data sharing agreements, and quickly repurposed software infrastructure, all of which limited the scale and reproducibility of these efforts. Being fully prepared for future public health emergencies will require building new technology tools which span the chasm of our institutional silos, to create an enduring network which harnesses the lessons learned from COVID-19 genomic efforts so this capability becomes routine and widely available.



Public Health Use of Genomic Epidemiology in the COVID Era

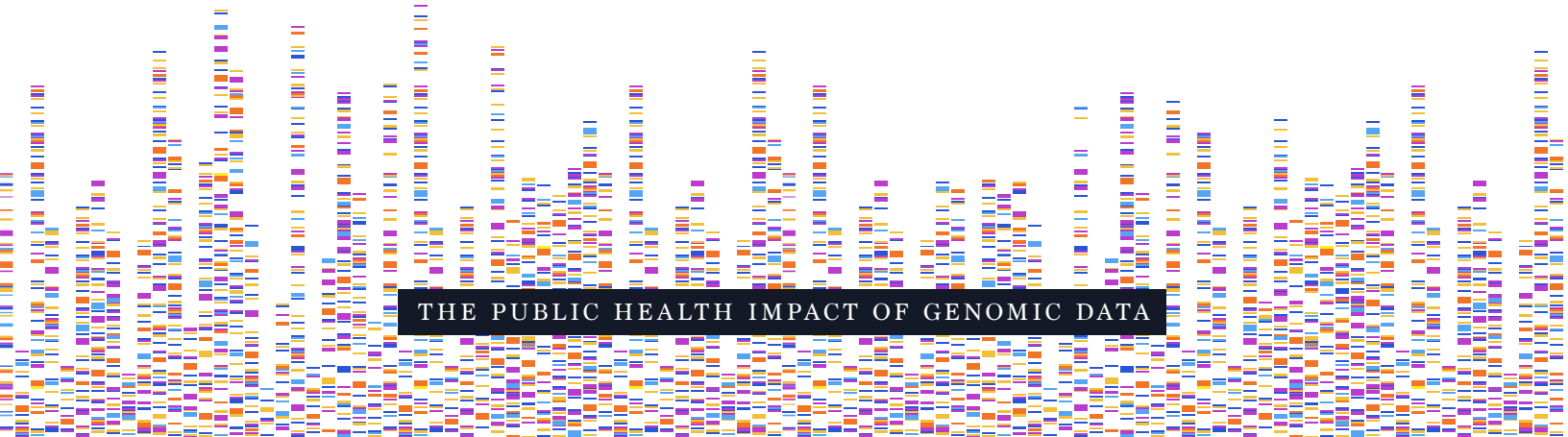
Over the past 15 years pathogen genomic sequencing has become dramatically less expensive and more widely available, prompting the increased use of genomic epidemiology in public health worldwide.[5] Public health use of genomic data has increased in sophistication and impact through application in the 2009 H1N1, 2012 MERS, and 2013-2016 Ebola public health emergencies.[5] The COVID pandemic has seen genomic data used on a massive scale—over 1.2 million SARS-CoV-2 sequences were produced during the first 14 months of the COVID-19 pandemic, compared with only 35,000 H1N1 genomes during the 18-month 2009 H1N1 pandemic.[1] Moreover, rapid turnaround times are now allowing genomics to inform public health response nearly in real time.[5] We detail various ways that SARS-CoV-2 genomic data has been used in public health interventions around the world.

Variant Detection

As the pandemic spreads through our communities the SARS-CoV-2 virus is constantly evolving. The adaptation of the virus to humans through natural selection as well as selective pressure due to pharmaceutical and non-pharmaceutical interventions causes new variants to emerge, which can exhibit increased transmissibility, virulence, and potentially worse clinical outcomes.

While an individual viral genome by itself is not sufficient to predict the increased public health risk of a variant, public health viral genomic surveillance is a necessary component to:

1. **Detect and highlight variants of interest (VOI) circulating in a region**
2. **Allow public and private decision-makers to correlate circulating VOIs with contact tracing and other epidemiological and clinical metadata to evaluate if a VOI poses increased risk and is in fact a public health variant of concern (VOC)**
3. **Plan and implement public health policies and interventions across businesses and communities to address regional VOCs, including vaccine escape mutants**
4. **Focus the collective efforts of public, private, and academic laboratories on:**
    - Conducting rapid clinical studies to establish increased transmissibility, virulence, or clinical significance of VOCs impacting a region
    - Ensuring diagnostic assays are keeping pace with VOCs that defeat existing tests[4]
    - Rapidly investigating and containing local VOCs that escape vaccines



These types of public health actions based on genomic data have been repeatedly demonstrated during the COVID-19 pandemic. For example, in South Africa, unusually high case counts in Nelson Mandela Bay prompted public health officials to consult genomic data in this region, which revealed a cluster consisting of a new variant (B.1.351) containing potentially worrisome mutations in the viral spike protein mediating cell entry.[6] At the same time, UK public health officials discovered a variant (B.1.1.7) in COG-UK, the UK's national public health genomic surveillance system, that shared some of these same worrisome spike protein mutations. A cascade of subsequent analyses by the public health community correlated genomic surveillance data with epidemiological metadata and provided mounting evidence for increased transmissibility of B.1.1.7 and sounded the public health alarm.[7,8] The discovery of these variants led directly to large-scale action, including targeted travel restrictions[9,10] and detailed recommendations on how public guidance, outreach, and the UK's transmission mitigation strategies need to be modified in response to B.1.1.7.[2] Denmark is maintaining stringent measures in spite of falling overall case counts because of the exponential rise of B.1.1.7 as monitored by a new nationwide sequencing effort.

As the pandemic rages, we will continue to see the emergence of new variants that may pose increased threats to lives and livelihood. As present restrictions are relaxed on the road back to normalcy, pathogen genetic surveillance can provide advanced warning of threats posed by a new variant, allowing for a proactive, targeted, tightening of restrictions rather than a return to broadly disruptive lockdowns. Establishing genomic surveillance–with robust participation by both private and public sectors–is the necessary first step in detecting and responding to the variants circulating in a county, state, or nation.

Genomic Data Protects Underserved Communities

The COVID pandemic has disproportionately impacted historically marginalized communities that suffer from lower access to healthcare and have fewer options to avoid essential frontline work when ill. The insights provided by genomic epidemiology data allow public health officials to understand the pandemic's progression through these communities and offer interventions that help protect livelihoods and prevent spread.

For example, a large genomic epidemiology study undertaken in the Mission District of San Francisco, a predominantly Latinx community, established that frontline workers were bringing viral strains into their households that originated from a wide geographic region.[11] This combination of genomic data with epidemiological data highlighted the elevated exposure risk of frontline work and dispelled the myth that this neighborhood's residents were simply being infected due to social interactions within their community. These data subsequently led to the introduction of a "Right to Recover" program that provides replacement income for frontline workers to quarantine after testing positive for COVID.[12]

In the UK, where the circulating B.1.1.7 variant prompted increased shutdowns across all sectors of society to mitigate its spread, genomic data coupled with epidemiological metadata demonstrated that the variant had lower attack rates in children than adults.[7] This genomic data led experts to recommend targeted interventions, instead of broad school closures which disproportionately impact marginalized communities.

Genomic Data Improves Our Understanding of Pandemic Dynamics

Advances in modeling allow us to use genomic data to estimate epidemiological dynamics not possible with traditional epidemiological approaches alone. Moreover, genomic data makes it possible to understand the dynamics of SARS-CoV-2 (or a variant) in a community even before initial cases are reported.[5]

It is now possible to use genomic data for robust estimates of epidemiological parameters such as reproductive numbers,[5] attack rates, and numbers of undetected cases.[15] Combined with traditional public health metadata, this enables the identification of cryptic community transmission[16] and even allows pinpoint identification of when and where the virus enters into specific populations.[11] Crucially, this powerful intelligence requires representative viral genomic sampling across sectors and communities. Without broad, representative surveillance the ability to estimate pandemic dynamics is limited and community-specific introductions will be missed.[5]

Geographic models integrate genomic data with movement data, population density, and geographic proximity to understand the dispersion of virus between regions and assess public health policy decisions. For example, analysis of SARS-CoV-2 genomic data in Brazil established that spread was mostly occurring within states and that inter-state spread decreased after travel restrictions were put in place, suggesting that public health efforts should be focused on local travel around population centers.[14] Interlinked communities cannot be safe unless all are safe, so it is crucial to understand the balance between new introductions (which can be limited by travel restrictions and quarantine) and community circulation (which requires interventions to reduce person-to-person contact in daily life).

Genomic Genomic Data Supercharges Outbreak Investigations

The SARS-CoV-2 genome is constantly mutating, on average once every 2-3 times transmission occurs between individuals. These mutations can be used to fingerprint specific viral infections and build viral family trees that precisely quantify how outbreak cases are related, establish duration of transmission in large outbreaks, and distinguish whether an outbreak is due to multiple introductions or local circulation.[5]

In areas where overall case counts are low, rapid genomic analysis of new transmission chains is vital to prevent things from worsening, and genomic data substantially enhances containment efforts. **For example, in New Zealand, a country whose rapid response and geographic advantages allowed community transmission to be repeatedly extinguished in 2020, public health officials used genomic data[3] to:**

1. Track which cases lead to ongoing transmission chains
2. Conclude that the virus was not circulating for long before the first cases were identified
3. Track a large cluster linking previously unlinked cases and establishing that it was a superspreader event seeded by a strain from the U.S. and that the reproductive number for this cluster dropped from 7 to 0.2 due to rapid containment efforts

In particular the use of genomic data to quantify the drop in reproductive number served as robust evidence that strict early lockdown was highly effective.[3]

In California, two dozen county public health departments began using viral genomic data in 2020 for COVID-19 outbreak response. Driven by an informal network of academic sequencing laboratories who engaged directly with local health jurisdictions (LHJs), the initiative provided genomic epidemiology expertise to support local public health efforts.[17-19] These collaborations involved painstakingly arranging approval and specimen transport from each participating LHJ, as well as returning data to each LHJ's epidemiologists for analysis and action. **The partnership has produced thousands of SARS-CoV-2 sequences from California specimens, and these data have been used in dozens of local public health investigations statewide throughout 2020. For example:**

1. Genomic epidemiology definitively linked outbreaks at three skilled nursing facilities (SNFs) in Southern California that were sharing staff. Public health officials used genomic data as part of their outreach effort with SNF management to change staffing policy.

2. An initial genomic epidemiology investigation at a Northern California county jail established that inmates had closely related SARS-CoV-2 viruses, indicating transmission within the jail. In response, officials changed infection control practices. Ongoing genomic surveillance showed that subsequent SARS-CoV-2 positive inmates had unrelated strains, proving that the modified infection control practices were effective, and confirming the importance of screening during intake.

3. Intensive contact tracing combined with genomic epidemiology in another Northern California county, a low-burden region, allowed public health officials to definitively link the county's cases to strains from outside the region. This established that local community transmission was not a major driver of the pandemic in this county and allowed public health resources to be prioritized accordingly. Later in the pandemic, genomics provided a link between cases in a population of agricultural workers and a major public university, demonstrating how protecting the most vulnerable protects the entire community.

4. Genomic epidemiology combined with contact tracing and a workplace investigation at a Northern California county food processing facility definitively established that employees at adjacent workstations were part of a single infection cluster. Public health officials took genomic data back to facility management to demonstrate the need for improved infection control infrastructure in this workplace.

The impact of genomic data in outbreak investigations is greatest when targeted sequencing of outbreak samples is combined with routine large-scale genomic surveillance. Statewide genomic surveillance facilitates an understanding of the baseline variants circulating in communities, which then allows facility-specific outbreaks to be distinguished from community transmission.

Linking Metadata to Relevant Sequence Data is Vital

In order for genomic data to have an impact, decision-makers must be able to act rapidly on it. This, in turn, requires that sequencing efforts are targeted to relevant public health situations and that the resulting sequence data is returned quickly and linked to epidemiological metadata. In other words, decision-makers can do much more with 100 representatively sampled sequences from their community linked to patient age, demographics, domicile, and comorbidities than they can with 100 anonymized sequences divorced from location, patient information, or any other metadata. Table 1

highlights the varying metadata, turnaround times, and sampling strategies required for the effective use of genomic data.

Previous California genomic epidemiology efforts[17-19] linked metadata to genomic data manually, with LHJs collating line lists that were overlaid onto genomic information. Automating these types of time-intensive processes will be crucial for increasing the scale and impact of genomic data in the U.S.

To make the kinds of analyses described above routine and universally available, genomic epidemiology efforts will need to link relevant metadata, improve sequencing turnaround times, and target sequencing to critical scenarios where that insight will empower containment. The operational barriers to accomplishing this are particularly severe in the decentralized U.S. public health system where thousands of local health jurisdictions have developed their own disparate sequencing logistics, IT systems, and data standards. Overcoming these challenges will require significant investment and consistent effort by new partnerships that include the public sector, private sector industries, and academia.

Table 1: Metadata, turnaround and sampling requirements for various public health uses of genomic data.

Use of genomic data	Metadata required	Turnaround time required	Sampling approach
Detect and contain variants	Case information	ASAP	Comprehensive
Enhance outbreak investigations	Case and investigation information	ASAP	Focused
Characterize variants	Vax status, clinical outcomes	Moderate (~1-2 weeks)	Focused or representative
Identify importations and circulation patterns	Geographic, travel history, demographic	Moderate (~1-2 weeks)	Representative
Forecast case curve	Geographic	Moderate (~1-2 weeks)	Representative
Refine diagnostic tests (to avoid gene dropout)	N/A	Slow (>2 weeks)	Representative
Estimate variant prevalence (for updating vaccines)	N/A	Slow (>2 weeks)	Representative
Estimate epidemiological parameters (R0, etc.)	Geographic	Slow (>2 weeks)	Representative

Conclusion

Genomic epidemiology presents a tremendous opportunity to harness the power of biotechnology and data science to rapidly detect and contain outbreaks while enabling centralized intelligence on pathogen movements through populations. Realizing the full potential of this opportunity will require translating the individual successes of genomic epidemiology during the COVID-19 pandemic into robust operational tools widely available to both public and private sectors.

Progress will depend on new types of collaborations between government, the private sector, and academia focused on building the sequencing logistics, data infrastructure, and software tools needed to empower decision-makers with insights previously unavailable within disease control. The rapid worldwide application of viral genomic analysis by multiple sectors during this pandemic lends hope that a streamlined, innovative approach can be developed to detect and contain future pathogen threats.

References

1. https://www.nature.com/articles/d41586-021-01069-w
2. Mitigations to reduce transmission of the new variant SARS-CoV-2 virus, 22 December 2020. GOV.UK. Available from: https://www.gov.uk/government/publications/emgspi-btweg-mitigations-to-reduce-transmission-of-the-new-variant-sars-cov-2-virus-22-december-2020
3. Geoghegan JL, Ren X, Storey M, et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. Nat Commun. Nature Publishing Group; 2020; 11(1):6351.
4. Vanaerschot M, Mann SA, Webber JT, et al. Identification of a Polymorphism in the N Gene of SARS-CoV-2 That Adversely Impacts Detection by Reverse Transcription-PCR. J Clin Microbiol. American Society for Microbiology Journals; 2020; 59(1). Available from: https://jcm.asm.org/content/59/1/e02369-20
5. Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. World Health Organization; 2021.
6. Tegally H, Wilkinson E, Giovanetti M, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv. Cold Spring Harbor Laboratory Press; 2020; :2020.12.21.20248640.
7. Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, Cevik M. Public health actions to control new SARS-CoV-2 variants. Cell. 2021. Available from: http://www.sciencedirect.com/science/article/pii/S0092867421000878
8. COG-UK report on SARS-CoV-2 Spike mutations of interest in the UK. COG-UK Consortium; 2021.
9. Covid-19: UK bans travel from South Africa over variant. BBC News. 2020. Available from: https://www.bbc.com/news/uk-55434086
10. Shepardson D. Exclusive: Biden to impose South Africa travel ban to combat new COVID-19 variant - CDC. Reuters. 2021. Available from: https://www.reuters.com/article/uk-health-coronavirus-usa-travel-exclusi-idUSKBN29T0P0
11. Chamie G, Marquez C, Crawford E, et al. Community Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 Disproportionately Affects the Latinx Population During Shelter-in-Place in San Francisco. Clin Infect Dis. 2020; (ciaa1234). Available from: https://doi.org/10.1093/cid/ciaa1234
12. Chávez L. SF will distribute $2 million to COVID-positive residents and workers without a safety net. Mission Local. 2020. Available from: https://missionlocal.org/2020/06/sf-will-distribute-2-million-to-covid-positive-residents-and-workers-without-a-safety-net/
13. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel-human interface. Ferguson NM, editor. eLife. eLife Sciences Publications, Ltd; 2018; 7:e31257.
14. Candido DS, Claro IM, Jesus JG de, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. medRxiv. Cold Spring Harbor Laboratory Press; 2020; :2020.06.11.20128249.
15. Li LM, Ayscue P. Using viral genomics to estimate undetected infections and extent of superspreading events for COVID-19. medRxiv. Cold Spring Harbor Laboratory Press; 2020; :2020.05.05.20092098.
16. Bedford T, Greninger AL, Roychoudhury P, et al. Cryptic transmission of SARS-CoV-2 in Washington state. Science. American Association for the Advancement of Science; 2020; 370(6516):571–575.
17. COVID-19 Variant First Found in Other Countries and States Now Seen More Frequently in California. COVID-19 Var. First Found Ctries. States Seen More Freq. Calif. UC San Franc. Available from: https://www.ucsf.edu/news/2021/01/419656/covid-19-variant-first-found-other-countries-and-states-now-seen-more
18. CZI, CZ Biohub & State of California Partner to Track COVID-19 Spread [Internet]. Chan Zuckerberg Initiat. 2020. Available from: https://chanzuckerberg.com/newsroom/chan-zuckerberg-initiative-chan-zuckerberg-biohub-the-state-of-california-partner-to-track-covid-19-spread-statewide/
19. Scientists and public health officials use genomics to understand how COVID-19 spreads in San Diego, neighboring regions. Available from: https://www.scripps.edu/news-and-events/press-room/2020/20201120-andersen-cdc.html