



---

# Strengthening the genomic surveillance ecosystem

Guiding principles to harness  
India's genomic potential



The  
**ROCKEFELLER**  
**FOUNDATION**

October, 2021



## Context

**This report draws insights from consultations with India's leading scientific and public health experts, as well as highlights fundamental principles in order to harness genomic surveillance and build resilient system that augment response time to future outbreaks.** The purpose of these consultations was to deliberate on pathogen genome surveillance in India, the country's long-term and short-term goals in this area, and its capabilities.

In December 2020, the Ministry of Health and Family Welfare (MoHFW), under the Government of India, set up the Indian SARS-CoV-2 Genomics Consortium (INSACOG), a national, multi-agency consortium of Genome Sequencing Laboratories (RGSs) established to expand genome sequencing of the SARS-CoV-2 virus across the nation. The National Centre for Disease Control (NCDC) is the coordinating agency for the consortium. One of its objectives is to sequence at least five percent of the total SARS-CoV-2 positive clinical specimens. Since its constitution, INSACOG has rapidly expanded its capacity from a network of ten labs to 28

labs and quickly established standard operating procedures for sample collection and sequencing. However, in March 2021, it was reported<sup>1</sup> that less than one percent of the total positive samples were sequenced from January 2021 to March 18, 2021. INSACOG managed to identify virus variants with concerning mutations as early as March 2021 but could not leverage genomic insights into forecasting the sudden surge in infections that the country experienced in April and May 2021. Therefore, higher sequencing levels, greater data-sharing, and better integration with clinical and epidemiologic data are needed for high-confidence



forecasting in order to translate the data into proposed action to contain a rapidly spreading pathogen such as SARS-CoV-2.

As the third wave of Covid-19 looms over India, there are increased efforts to boost genomic surveillance capabilities across India. One such effort includes INSACOG's increased capacity to leverage sequencing samples from the Northeast region.<sup>2</sup>

The Rockefeller Foundation is keen to use these insights to inform their existing investments and help complement the Government of India's efforts to strengthen India's public health systems. In the first quarter of 2021, The Rockefeller Foundation announced

\$20 million to help strengthen global capabilities to detect and respond to pandemic threats within the first 100 days of an outbreak. In line with this vision, The Rockefeller Foundation is supporting a coalition of organizations led by the Centre for Cellular and Molecular Biology (CCMB) to strengthen institutional capabilities to harness genomic data for actionable insights against pathogen outbreaks in India.

This report summarizes the challenges and opportunities shared by the experts. It serves as a vital resource for policymakers, public health experts, and the global scientific community to strengthen the genomic surveillance ecosystem in India.



<sup>1</sup> <https://www.downtoearth.org.in/news/health/dismal-figures-india-sequenced-less-than-1-of-total-Covid-19-samples-in-nearly-3-months--76114>; <https://www.indiatoday.in/coronavirus-outbreak/story/ncdc-director-genome-sequencing-india-positive-corona-covid-samples-uk-double-triple-variants-1794357-2021-04-23>

<sup>2</sup> <https://www.thehindu.com/news/cities/Hyderabad/genome-surveillance-for-north-east-region-begins/article35430755.ece>

#### Disclaimer

The information provided in this report is for informational purposes only. The Rockefeller Foundation expressly disclaims and assumes no responsibility for any losses, damages, claims, or other liabilities arising out of or relating to the use of this information.

It is expressly understood that The Rockefeller Foundation, by providing this information, has no obligation to update the information or provide additional support or information to the recipient.



## Overview of Genomic Surveillance

**Genomic surveillance helps track novel pathogens and variants of known pathogens.** Pathogens, including the SARS-CoV-2 virus, continuously mutate to adapt to their hosts, resulting in variations in their genetic code.

Typically, mutations occur over time due to a host of factors like the genetic profile of the population, disease burden, migration, use of antibiotics, immunity derived by natural infection or vaccination, and nutrition status, among others. While most of these mutations and resulting pathogen variants do not translate into a more virulent pathogen, certain mutations alter the pathogen's virulence and transmissibility. Such variants are classified by the World Health Organization (WHO) as variants of concern. The devastating second wave of the pandemic that hit India is believed<sup>3</sup> to have been caused by one such variant of the SAR-CoV-2 virus, B.1.617.2, now commonly known as the Delta variant.<sup>4</sup> It is paramount that genomic surveillance systems are strengthened to anticipate and rapidly respond to prevent a future

outbreak in India. In the past, genomic surveillance systems have proven to be instrumental in controlling polio in India and around the globe. In the case of polio, the gradual scale-up of vaccination and public health measures have helped to reduce the circulation of polio genotypes from 33 to 2. It also aided the enterovirus laboratories in India in tracking the genotypes globally.

As the country emerges from the devastating second wave of Covid-19, it faces the looming threat of subsequent viral surges. Thus, it is imperative to analyze existing genomic surveillance capacities and identify avenues to complement the government's efforts in this space.



## Objectives of genomic surveillance



### To enhance pandemic preparedness

Smart genomic surveillance helps governments to strengthen predictive interventions and serves as an early warning system. The data from genomic surveillance can be integrated with other epidemiological and metadata to provide real-time, actionable insights that enable governments to take immediate measures like issuing advisories, changing testing strategies, and stockpiling essential supplies.



### To support vaccine development for emerging infectious diseases

Genomic surveillance helps to track and analyze circulating strains and informs vaccine development and redesign for emerging and existing infectious diseases. Since pathogens mutate over time due to a host of factors, inputs from the genomic analysis can help determine the vaccine efficacy vis-a-vis the variants. This helps identify vaccine breakthroughs (when a vaccinated individual becomes sick from the same pathogen that the vaccine is meant to prevent), which can adversely impact pandemic containment efforts and present vaccinologists with an opportunity to adapt vaccines to accommodate the genetic changes in pathogens.



### To support diagnostics and therapeutics

Smart surveillance also aids in improving diagnostics, therapeutics, and other medical countermeasures like safety and treatment protocols. Early in the Covid-19 pandemic, tests were developed based on the information available from the genomic sequencing of SARS-CoV-2. Tests could become unreliable due to the emergence of new variants over time, and smart surveillance can help provide insights in real-time to inform the development and calibration of diagnostics.

Similarly, the emergence of variants has threatened the efficacy of therapeutics. Data from genomic sequencing can help inform pharmaceutical companies and help shape their research and design efforts.



<sup>3</sup> <https://dbtindia.gov.in/sites/default/files/Note%20on%20INSACOG%20Data%20and%20Mutants%2021-04-2021.pdf>

<sup>4</sup> <https://economictimes.indiatimes.com/industry/healthcare/biotech/healthcare/who-renames-virus-strain-first-found-in-india-as-delta-variant/articleshow/83135554.cms?from=mdr>



## Guide to a Smart Surveillance System

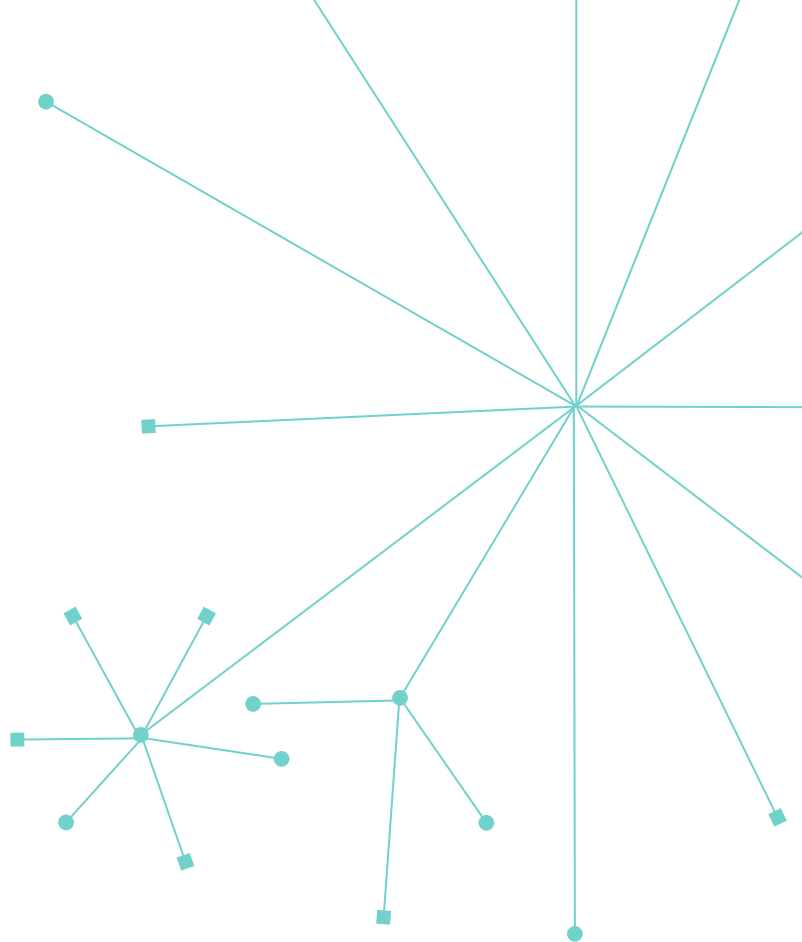
**A smart surveillance system is founded on the premise of providing clinical information in real-time to immediately mobilize all relevant stakeholders to sequence pathogens and derive actionable insights.**

This report aims to set out the following broad guiding principles to serve as a reference framework for governments, health care professionals, and scientists to build a smart genomic surveillance system.

- a) **Concerted efforts to scale sequencing** by volume and across geographies to better understand the changing characteristics of the pathogen and its evolution. Deploying decentralized technologies for scale-up is crucial for wider surveillance measures.
- b) **Robust genomic analysis to address critical public health questions** and design population-level interventions against pathogens.
- c) **Develop a conducive data ecosystem that strategically leverages existing systems and integrates** them for optimal use amongst key stakeholders. Elements to support the strengthening of this includes the following:
  - Focusing on efforts to decentralize genomic sequencing and developing central-level protocols for sequencing to synergize efforts among various data sets.
  - Strengthening the link between genomic sequencing with epidemiological and clinical surveillance networks and streamlining various streams of complex and variable data through a federated database system.
  - Nurturing sustained communication channels between public institutions through a consortium model that enables timely and transparent data sharing. Correspondingly, embedding responsible data sharing to ensure the protection of individual identity.



- d) **Develop protocols for responsible data sharing** between the scientific community and others, including public health professionals, the medical community, policymakers, and the public.
- e) **Recognize and build discourse around enhancing data visualization** to enable data systems' effective delivery of insights relevant to drive real-time decision-making efforts amongst key stakeholders.





# 1

## Scaling Genomic Sequencing

Sequencing a pathogen's genome aids in classifying strains, understanding patterns of transmission and spread, and providing information to inform public health response. In fast-spreading pathogens, the frequency of genetic evolution is higher, and sequencing assumes greater significance to monitor variants and distinguish their characteristics from the original strain.

For instance, mutations in the SARS-CoV-2 strain became a concern when the Alpha variant (B.1.1.7 lineage) revealed genetic changes in the spike protein. The intricate exercise of detecting and identifying new strains thus involves designing a strategy to select suitable samples, retrieving specimens, and using high-throughput sequencing technologies<sup>5</sup> to assemble consensus genome sequences. High-quality sequencing data on different variants underpin a robust genomic surveillance system to investigate how they originate, circulate, and reproduce.

In India, at the beginning of the Covid-19 pandemic, the sequencing of SARS-CoV-2 was undertaken to understand the virus but largely remained an esoteric

activity. The emergence of the Alpha variant made clear the need for greater genomic sequencing, and the INSACOG was established. The Government of India awarded additional funding of INR 70 crores (approximately \$9 million) to augment sequencing capacities, bioinformatic analyses, scientific training, and the supply of required reagents and consumables.<sup>6</sup>

However, some barriers continue to impede the genome sequencing programs in India. For instance, the limited diversity in the clinical specimens collected for viral sequencing has been a critical barrier. At the same time, the non-availability of domestic reagent production and shortage of a trained workforce continues to impact the volume of sequencing. Finally, pathogen genomic sequencing for public health surveillance is concentrated in public laboratory facilities, preventing private labs with the requisite capacity to engage in genomic sequencing. Due to India's geographic and socio-economic heterogeneity, there is a need for a strategy to overcome these barriers to expedite the pace and improve the comprehensiveness of sequencing.

<sup>5</sup> High throughput sequencing technology is capable of sequencing billions of bases (unit components of a genome) in a single sequencing reaction and allows researchers to sequence entire genomes on a population-wide scale at reasonable cost. Source: <https://pubmed.ncbi.nlm.nih.gov/22931062/>

<sup>6</sup> <https://dbtindia.gov.in/sites/default/files/Note%20on%20INSACOG%20Data%20and%20Mutants%2021-04-2021.pdf>





## OPPORTUNITIES TO SCALE SEQUENCING

---

- **Develop differential sequencing strategies.**

Differential strategies must be adopted to carry out targeted sequencing of known pathogen variants alongside detecting emerging ones. Tracking the prevalence of SARS-CoV-2 and other pathogen variants imported into India must involve targeted sample collection from transit hubs and specific transient populations (such as travelers). At the same time, an extensive collection representing the expanse of both space and time (Spatio-temporal diversity) must be undertaken to identify new variants of pathogens. It should include collection from laboratories and clinical establishments to waste-water sources such as natural pathogen reservoirs, sewage systems, wet markets, and outbreak sites.

- **Deploy decentralized technologies.**

To rapidly sequence genomic data and determine sources of clusters, portable and decentralized technology must be widely adopted.<sup>7</sup> In addition to genomic sequencing, such next-generation sequencing approaches investigate the interactions between the causative pathogen and the infected host. Additionally, technology that can both identify unknown infectious agents and establish relationships with other microorganisms present in the collected sample, broadening the impact of single sample sequencing evaluation. Such technologies should be scaled up throughout the country, considering its advantages in portability, turnaround time, and cost.<sup>8</sup> Even beyond the Covid-19 pandemic, these technologies can become critical tools for undertaking broader surveillance of outbreaks in a population.

- **Expand local production of sequencing reagents.**

To ensure the availability of reagents, domestic manufacturing capacities of reagent kits should be developed in a similar capacity to how RT-PCR reagent production was globally scaled. Distributed manufacturing models can decrease the cost of sequencing and improve availability of supplies during a surge when more reagents are needed.

Availability of affordable reagent kits that meet rigorous quality standards will positively impact the scaling of sequencing across the country. For instance, the C-CAMP-InDx Program<sup>9</sup> was conceptualized to build the country's self-reliance in molecular diagnostics. It brought together key stakeholders such as reagent manufacturers and component-makers of the kit, on one platform. This played a crucial part in scaling up India's test kit-making capacity.

- **Leverage existing sentinel surveillance networks.**

Several countries have instituted National Influenza Centers to expand influenza monitoring under the WHO's Global Influenza Surveillance and Response System (GISRS). The network<sup>10</sup> has played an effective role in serving as a platform to undertake surveillance, monitor influenza epidemiology, issue alerts on virus evolution and aid in vaccine development. A similar mechanism should be considered and further expanded to institutionalize surveillance for all respiratory infections, beyond severe acute respiratory infections (SARIs) and influenza-like infections (ILIs). It must be noted, the latest guideline on "Genome Surveillance for SARS-CoV-2 in India" by INSACOG released on 15 July 2021, already makes substantial mention of the role of sentinel sites in the implementation of genomic surveillance strategy, including defining a standard operating procedure (SOPs) for sentinel sites.<sup>11</sup> Ensuring compliance will be critical to effective scaling of genomic surveillance efforts in the country.

- **Sustain active and passive efforts of pathogen surveillance.**

During periods of non-emergency, the optimum level of genomic surveillance will depend on the epidemiology, infectiousness, and evolutionary rate of pathogens. Genomic surveillance activities must be used in steady state, non-emergency times to ensure surge capabilities in an emergency response. For instance, in the case of polio,

surveillance efforts have continued despite polio being eradicated in India, which is helping to identify and track emerging polio viruses. As proposed under the Global Immunological Observatory model, serological baselines must be defined using representative samples via routine collection of blood and plasma specimens. Ongoing surveillance will enable quick detection of anomalies reflecting immune responses to emerging threats. In parallel, monitoring of environmental factors must be maintained and integrated with the larger disease surveillance efforts. It is also critical to evaluate wastewater for emergence of newly emerging or re-emerging pathogens in a community setting.

When the infection rate of a specific pathogen is low in the population as detected in clinical specimens, sequencing wastewater samples becomes more important. Thus, it is important that public health officials and policymakers consider a mix of surveillance methods to optimize efforts and forecast the emergence of infectious disease outbreaks.



8 <https://www.nature.com/articles/s41467-020-20075-6>

9 <https://www.ccamp.res.in>

10 <https://idsp.nic.in/index1.php?lang=1&level=1&sublinkid=5789&lid=3722>

11 file:///C:/Users/ijjad/Downloads/INSACOGGuidanceDocumentdated15July2021final.pdf

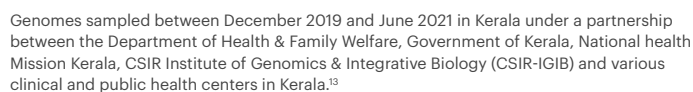




# Advancing Genomic Analysis

Maintaining regular public reports from such dashboards enables stakeholders to determine if a pathogen interacts differently with human hosts. This helps in addressing some critical public health questions about the pathogen: Is it more infectious? Is it more virulent? Is it likely to evade countermeasures or escape immunity from vaccine or prior infection? This also offers a more refined

Built with banjolily/Phylvis/. Maintained by Vinod Scaria Lab.  
Showing 4164 of 4164 genomes sampled between Dec 2019 and Jun 2021



understanding for improving molecular diagnostics, supporting the design and sensitivity monitoring of serological assays, informing vaccine development, and tailoring therapies, including monoclonal antibodies.<sup>14</sup>

It is key to note that an outbreak's spatiotemporal dynamics provide important insights to help direct public health resources intended to control transmission. Genomic surveillance gives us temporal signals, and a substantial database of clinical specimens collected across time and geographical locations. In a phylogenetic and phylogeographic context, this would help in identifying patterns within the datasets that correspond to geographic distributions, migratory histories, and mutations of the virus, informing subsequent public health responses. Such a nationwide effort of genomic epidemiology reflecting spatio-temporal frequencies of variants is currently absent. A key impediment is the limited collection of granular data across four key indicators (cases, tests, deaths, and hospitalizations) at the local level. This limits analysis on the correlation of variants with diverse factors, such as distinct population groups and geo-climatic patterns. Moreover, the existing bioinformatics capacity in public laboratories is in need of expansion to standardize, process, and characterize pathogen properties. The recent funding announced by the Cabinet Ministers towards NCDC for strengthening Epidemic Intelligence Services (EIS) under Phase II of India's Covid-19 Emergency Response and Health Systems Preparedness Package could partially address these concerns.<sup>15</sup>



## OPPORTUNITIES TO ADVANCE GENOMIC ANALYSIS

---

- **Reinforce the value of data collection, sharing and analysis at the local-level.** The collection of data must be prioritized at the district level. It will enable the tracking of genomic epidemiology of circulating variants reducing the need to normalize massive datasets. The seroprevalence surveys undertaken in the city of Pune in Maharashtra by IISER illustrated the merit of local data in informing effective public policy decisions and non-pharmaceutical interventions. Similarly, for genomic analysis, key inputs with a suitable sampling strategy and clinical metadata can address fine-scale epidemiological questions, such as linking a local outbreak to a specific variant. In the long run, these capabilities will enable districts to localize infectious disease outbreaks within their territory and address the outbreaks with improved precision.
- **Establish a hazard index to analyze the correlation between geographical patterns and disease outcomes.** Analyzing the geographic clustering tendency of a virus,<sup>16</sup> there is a need to draw correlations between geo-climate patterns and the epidemiology of different strains. This will help explain the reasons behind the enhanced spread of select variants in a few geographies and their relation to regional and climatic factors. This analysis can also offer insights into whether specific populations are vulnerable and whether socio-economic and nutritional status, profession, disease propensity, immunity levels, or genetic variations play a role in affecting disease severity and outcome. The existing Infectious Diseases Hazard Map project developed by IISER Pune,<sup>17</sup> underpinned on migration patterns, can be leveraged for this purpose to assign risk index and hazard rankings to different cities and towns. This will serve as a guide for the phylodynamic analysis of the virus and aid in detailing the local control strategies against outbreaks.

<sup>12</sup> <https://www.who.int/publications/i/item/9789240018440>

<sup>13</sup> <https://sites.google.com/view/genepikerala/home?authuser=3>

<sup>14</sup> <https://www.who.int/publications/i/item/9789240018440>

<sup>15</sup> <https://pib.gov.in/PressReleaseDetail.aspx?PRID=1733841>

<sup>16</sup> <https://doi.org/10.1371/journal.pone.0238344>

<sup>17</sup> <http://www.iiserpune.ac.in/-hazardmap/>



● **Bolster bioinformatics tools to characterize pathogens properties.** In tandem with increased sequencing, efforts should be made to develop scalable analytical tools and computational algorithms to assemble, analyze, and interpret genomic data. An institutional platform that successfully evaluates pathogen genomes and examines the spread of diseases in real-time can serve national surveillance priorities for broad-based disease outbreak management. To expand analytical capabilities, bioinformatics and genomic analysis must become more accessible to all audiences. Traditionally, resources are directed towards monitoring the transmission of the known pathogens.

Still, investments are needed to develop capabilities to respond to emerging pathogen variants and locate unusual outbreaks and cluster infections in real-time. In parallel, a dedicated cadre of engineers and biostatisticians must be appointed and trained within the country's public health systems. Authorities must also consider harnessing the existing bioinformatics capabilities of the private sector. Like the European Bioinformatics Institute, India should encourage the development of ifs, tools, and software to align, verify and visualize the diverse data produced in publicly funded research, and make that information freely available to all.<sup>18</sup>



<sup>18</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/920378/Genome\\_UK\\_-\\_the\\_future\\_of\\_healthcare.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/920378/Genome_UK_-_the_future_of_healthcare.pdf)



# 3

## Building a Conducive Data Ecosystem

A health-data ecosystem is founded on prototypes, data analytics, and visualizations to fuel data-driven decisions. It is a collection of different systems that capture, analyze, and, where appropriate, distribute information across the ecosystem. Network interoperability ensures that data and information are accessible to a variety of decision-makers and stakeholders at all levels.

In the Covid-19 context, India currently has four data systems: the Indian Council of Medical Research (ICMR)'s testing database, the MoHFW's Covid-19 India website, the CoWIN vaccination dashboard, and the INSACOG database on genomic surveillance. While bilateral database linkages exist between ICMR, MoHFW and INSACOG, the need for stronger, multilateral integration of all four databases (particularly INSACOG) still persists. Considering the non-contiguous nature and specific purposes of these databases, data is collected and disseminated by multiple agencies, leading to increased fragmentation of the data ecosystem. Ideally, genomic sequencing data needs to be accompanied by appropriate metadata, and for this, the data systems must be made interoperable. For example, a positive test sample that is sequenced should be linked to the patient's CoWIN data and other testing data to analyze the immune escape of any variant and the efficacy of testing against such a variant.

A significant barrier to data integration is the deficiency of communication channels between public institutions. Public health decision-making suffers from inefficiency and shortcomings without a composite database to catalog different interventions and their results. Further, it is critical to note that in addition to making the genomic data more meaningful, there is also a need to make it more accessible. Since the onset of the Covid-19 pandemic, scientists in India have highlighted the inability to access genomic data.<sup>19</sup> This can be attributed to the absence of a structured and open data ecosystem that is easy to access. Additionally, in the lack of a uniform central genomic surveillance and analysis protocol, state governments may maneuver independent genomic programs, leading to multiple data banks and increased fragmentation.

Against this backdrop, it is important to acknowledge that a robust data ecosystem is not activated only during a crisis. There is immense merit in ensuring it remains operational and undertakes continuous surveillance, enabling the country to respond proactively to any potential future outbreaks.

<sup>19</sup> <https://www.thehindu.com/news/national/at-least-100-scientists-petition-pm-modi-for-better-access-to-icmrs-data-bank/article34442648.ece> ; <https://scroll.in/latest/993670/covid-19-over-160-scientists-urge-modi-to-publicise-genomic-and-testing-data-of-virus>



## OPPORTUNITIES TO BUILD A CONDUCTIVE DATA ECOSYSTEM

---

- **Decentralize efforts of genomic sequencing.**

State governments need to spearhead genomic sequencing and its related policy interventions instead of a centrally controlled sequencing program. This is largely because India is widely heterogeneous, and testing and sequencing need to account for local needs and preferences for the data to be actionable for public health officials and patients. There is a great need to simultaneously encourage states to utilize their capabilities and institutions to play a more significant role in sequencing efforts, while the federal government works towards establishing common standards and platforms. The synergies between these diverse datasets can offer insight into the migration route of variants across the country and guide interventions at the local and national levels. At the same time, a uniform protocol can assist states that cannot design their independent genomic sequencing programs. Thus, a common protocol can incentivize states to engage in sequencing while ensuring that the data meets the national program's goal.

Many state governments have shown interest in collaborating with scientific and research organizations to enhance genomic surveillance. For example, a part of the work undertaken by the Bangalore Bioscience Cluster looks at existing information about the SARS-CoV-2 variants and their trajectories in populated cities in India. They work closely with state governments on sequencing retrospective samples to unpack the available information and identify future variants of concern. Leveraging these interests is vital to open up opportunities at a state level, which helps in making data more readily available to the state governments for discussions and policy actions.

- **Strengthen the collection of clinical metadata.**

Genome surveillance undertaken by INSACOG is currently linked with epidemiological surveillance and clinical correlation. While INSACOG's current network predominately has linkages with Government labs, strengthening their linkage with private and academic labs will help comprehensively assess the representatives of the data gathered. This system can be further bolstered by linking genomic sample collection and clinical metadata to a universal health identifier for aggregating all necessary data. This will ensure that the patient's clinical outcome (a mild, moderate, or severe infection) is recorded and analyzed vis-à-vis the viral variants. These factors are critical to establishing a causal relationship between the different variants and the disease severity and outcome. Even beyond the Covid-19 crisis, there is a need to link genomic sequencing exercises with epidemiological and clinical surveillance networks to rapidly identify, characterize and label variants of concern (VOCs) and determine their functional differences from other related pathogen strains.

- **Introduce a federated data system.**

Currently, the country has separate data systems for testing, seroprevalence, vaccination, and genomic surveillance, with limited communication channels. A federated data system is needed to streamline the various streams of complex and variable data and ensure interoperability for meaningful analysis. This will involve some aspects of data to be shared at the sensitive metadata while others may be non-personal and publicly accessible. A federated data architecture provides secure yet open access to disparate data systems and data formats.<sup>20</sup> This will require collaboration between data controllers to craft rules and core principles to preserve security, interoperability, and performance.



- **Nurture sustained communication channels between public institutions.**

To integrate different data channels, necessary relationships must exist between various organizations to foster trust and transparency. The principal institutions, including the MoHFW, Ministry of Science and Technology, the state health departments, ICMR, INSACOG, NCDC, must agree on a consortium model to share data in a timely, sustained, and predictable manner. Creating such collaborations early on, instead of operating in silos, is in the interest of the public and the agencies involved to drive the data-to-action continuum.

- **Embed responsible data sharing to protect individual privacy.**

Data that is generated from public funds are for the global public good. Moreover, as far as data access

and harmonization is concerned, people who have access to data must make it available for others, aligned to stipulated guidelines. However, one of the critical elements is to ensure identities are anonymized. Steps must be taken to ensure individual identities are not being traded off.

Confidentiality, if it is compromised, could lead to severe challenges and expose an individual's vulnerabilities. Therefore, it is pertinent to ensure that socio-demographic data does not include personal identifiers and is only collected through consent, including consent to share after adequate anonymization/de-identification. A categorization of the sensitivity of the data to define commensurate restrictions around its access (like suggested by the Department of Biotechnology's 2021 Biotech-Promotion of Research and Innovation through Data Exchange Guidelines<sup>21</sup>) could also be considered.



<sup>20</sup> [http://www3.weforum.org/docs/WEF\\_Sharing\\_Sensitive\\_Health\\_Data\\_2020.pdf](http://www3.weforum.org/docs/WEF_Sharing_Sensitive_Health_Data_2020.pdf)

<sup>21</sup> <https://dbtindia.gov.in/sites/default/files/Biotech%20Pride%20Guidelines%20July%202021.pdf>



# 4

## Enhancing External Communication and Data Visualization

Data sharing within the global scientific community and those outside of it, including public health professionals, the medical community, policymakers, and the public, is the mainstay of reproducible science. It is a crucial resource for overall human development, benefits the scientific community, and boosts transparency with the public.

However, a challenge within the Indian landscape is inadequate sharing of available data and limited accessibility through open space platforms, making data sharing less transparent. Therefore, it is pertinent to recognize that accelerating data sharing is important and has ethical implications, and must be executed through responsible means.

Correspondingly, data visualization is instrumental for exploring and communicating insights from genomic datasets. When genomic data is combined with epidemiological data generated by public health laboratories, its analyses are vital to various public health professionals, researchers, epidemiologists, and policymakers.<sup>22</sup> However, this often results in interpretability challenges as many data users may not necessarily have experience routinely working with genomic data that these analyses rely upon. As research highlights, such challenges may arise even in the initial phases of data exploration and model building, not just in later phases of

investigation or when attempting to communicate the results of an analysis.<sup>23</sup> Therefore, as the discourse around building robust genomic surveillance in India develops, it is critical to recognize and account for measures that would help enhance communication and data visualization in the system.



<sup>22</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6513170/>

<sup>23</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6513170/>



## OPPORTUNITIES TO ENHANCE COMMUNICATION AND DATA VISUALIZATION

- **Build and sustain strong relationships with global surveillance systems.**

A crucial step towards generating responsible data sharing is engaging with institutions at national and international levels. In addition to strong communication between data systems, it is essential to develop efficient platforms for responsible data sharing. Creating an independent value and ethics-based agency is necessary to maintain relationships with multiple organizations, including international organizations. For instance, The GISAID Initiative is the global database mechanism, comprised of a community of scientists that share genomic data. GISAID makes available bioinformatic tools to enable scientists to analyze their data in local, regional and global context and enables these data to be shared within the community of users. Scientists from over 182 countries and territories share data through the GISAID mechanism. Understanding its value system could be essential in developing and strengthening a national-level genomic surveillance system contextualized to India. More importantly, access to such data and building strong relationships with such institutions will play a prominent role in building India's genomic data sharing processes and practices in a responsible, ethical, and practical manner.

- **Develop an efficient interface with the public**

There is a need to build systems that enable appropriate appraisal of the quality and relevance of data. Following this, relevant and authorized bodies should develop processes to share data into the global health system as a way of gaining global context to strengthen national and local data relevance. Further, variants and mutations evoke many misconceptions and fear in the community. It is also important to allay fears regarding reinfections or breakthrough infections. Therefore, the public interface should be designed to communicate critical information effectively. Additionally, recognizing that many consumers of genomic data may not be experts in viral genomics is essential to driving transparent, easily understandable, and accessible communication and information sharing.







## Conclusion

Since the beginning of this pandemic, India witnessed a host of seesaw policy measures, where positions and discourse changed quickly with the evolving situation. While it is still premature for a conclusive discussion around genomic surveillance, as highlighted in this report, there is widespread recognition of its scope and potential as a critical element to strengthen over-all public health response efforts.

It is equally pertinent to reaffirm that surveillance of any kind will not meet its objectives if done siloed or in a one-off manner and needs to be longitudinal. Therefore, whether genomic, serological, or pathogen surveillance needs to be done more routinely with adequate representative sampling strategies for a comprehensive surveillance system to function. Efforts to scale sequencing, develop a comprehensive data ecosystem, and build stronger partnerships between data stakeholders must also be undertaken. At the same time, it is crucial to focus on decentralized systems that adequately build local capacities and enable granular data generation and analysis at state levels while still linking systems to a central command system.

While India has made significant strides in its pandemic response since the outbreak, including greater genomic analysis capabilities, there is consensus amongst key stakeholders that the nation currently stands at a critical juncture. It finds itself in an opportune moment to reflect on the current system's limitations while building a more resilient one.

Moreover, there is a broad consensus that creating this genomic surveillance system transcends the need for merely strengthening systemic impediments and logistics around it. It is equally important to focus on linking relevant stakeholders from the policy and governance space with the public health and scientific communities and building trust, transparency, and cooperation across systems.

As the road ahead remains uncertain and the potential for additional pandemic waves looms, building the discourse around optimal genomic surveillance in India and exploring promising models to harness genomic sequencing remains essential. Insights highlighted in this report can stand as a directional guide to build a robust genomic surveillance system. Not only that, but it can also serve as a relevant resource for policymakers, public health experts, and the Indian and global scientific community to enhance overall public health systems as the world continues to fight the deadliest health emergency in over a century.

---

# Acknowledgements

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.

## KEY CONTRIBUTORS

---

### Anurag Agrawal

Director, CSIR Institute of Genomics & Integrative Biology (IGIB)

### Bhramar Mukherjee

Professor and Chair, Department of Biostatistics; Professor, Department of Epidemiology, Professor, Global Public Health, University of Michigan (UM)

### Gagandeep Kang

Professor of Microbiology, Wellcome Trust Research Laboratory, Division of Gastrointestinal Sciences, the Christian Medical College (CMC) Vellore

### Harpreet Singh

Head, Division of Biomedical Informatics, Indian Council of Medical Research (ICMR)

### K. Srinath Reddy

President, Public Health Foundation of India

### L S Shashidhara

Professor of Genetics, Molecular Biology and Evolutionary Biology at Ashoka University

### Neeraj Jain

Country Director India, PATH

### Nirmal Kumar Ganguly

Former Director General, Indian Council of Medical Research (ICMR)

### Rakesh Mishra

Director, Centre for Cellular and Molecular Biology (CCMB)

### Satyajit Mayor

Director, National Centre for Biological Sciences (NCBS)

### Shahid Jameel

Director, Trivedi School of Biosciences at Ashoka University

### Manisha Bhinge

Managing Director, Programs, Health Initiative, The Rockefeller Foundation

### Abhinav Verma

Consultant, Policy and Strategy Lead, Asia Regional Office, The Rockefeller Foundation

## REPORT DEVELOPMENT TEAM

---

### Suryaprabha Sadasivan

Vice-President, Chase India

### Jija Dutt

Sr. Associate, Chase India

### Aashna Kothiyal

Associate, Chase India

### Torsha Dasgupta

Associate, Chase India

Design by [ahoystudios.com](https://ahoystudios.com)





The  
**ROCKEFELLER**  
**FOUNDATION**

[rockefellerfoundation.org](http://rockefellerfoundation.org)