

HOME / NEWS / INDIA

Explained | What is genome sequencing and why does the Genome India Project matter?

PREMIUM

Sequencing 10,000 Indian human genomes under the centre-backed project would mean that therapies could be customized for identified genetic variants among the Indian population

April 21, 2023 11:20 am | Updated 01:27 pm IST

DIKSHA MUNJAL

COMMENTS

SHARE

 READ LATER



The story so far: The Department of Biotechnology (DBT) recently said that the exercise to sequence 10,000 Indian human genomes and create a database under the Centre-backed Genome India Project is about two-thirds complete. About 7,000 Indian genomes have already been sequenced of which, 3,000 are available for public access by researchers.

The proponents of the project say it would enable researchers anywhere in the world to learn about genetic variants unique to the Indian population. Countries including the United Kingdom, China, and the United States have launched similar programmes to sequence at least 1,00,000 of their population's genomes.

What is genome sequencing?

The human genome is the entire set of deoxyribonucleic acid (DNA) residing in the nucleus of every cell of each human body. It carries the complete genetic information responsible for the development and functioning of the organism. The DNA consists of a double-stranded molecule built up by four bases – adenine (A), cytosine (C), guanine (G) and thymine (T). Every base on one strand pairs with a complementary base on the other strand (A with T and C with G) In all, the genome is made up of approximately 3.05 billion such base pairs. .

While the sequence or order of base pairs is identical in all humans, compared to that of a mouse or another species, there are differences in the genome of every human being that makes them unique. The process of deciphering the order of base pairs, to decode the genetic fingerprint of a human is called genome sequencing.

In 1990, a group of scientists began to work on determining the whole sequence of the human genome under the Human Genome Project. The first results of the complete human genome sequence were given in 2003. However, some percentage of repetitive parts were yet to be sequenced. The Human Genome Project released the latest version of the complete human genome in 2023, with a 0.3% error margin.

Costs of sequencing differ based on the methods employed or the accuracy expected. Since



Genomic sequencing has now evolved to a stage where large sequencers can process thousands of samples simultaneously. There are several approaches to genome sequencing — including whole genome sequencing or next generation sequencing — that have different advantages.

The process of whole-genome sequencing, made possible by the **Human Genome Project**, now facilitates the reading of a person's individual genome to identify differences from the average human genome. These differences or mutations can tell us about each human's susceptibility or future vulnerability to a disease, their reaction or sensitivity to a particular stimulus, and so on.

What are the applications of genome sequencing?

Genome sequencing has been used to evaluate rare disorders, preconditions for disorders, even cancer from the viewpoint of genetics, rather than as diseases of certain organs. Nearly 10,000 diseases — including cystic fibrosis and thalassemia — are known to be the result of a single gene malfunctioning.

In the past decade, it has also been used as a tool for prenatal screening, to investigate whether the foetus has genetic disorders or anomalies. The *New York Times* notes that the Nobel Prize-winning technology Crispr, which relies on sequencing, may potentially allow scientists to repair disease-causing mutations in human genomes. Liquid biopsies, where a small amount of blood is examined for DNA markers, could help diagnose cancer long before symptoms appear.

In public health, however, sequencing has been used to read the codes of viruses—one of its first practical usages was in 2014, when a group of scientists from M.I.T and Harvard sequenced samples of Ebola from infected African patients to show how genomic data of viruses could reveal hidden pathways of transmission, which might then be halted, thus slowing or even preventing the infection's spread. Experts say that as sequencing gets cheaper, every human's genome may feasibly be sequenced as part of routine health care in the future, to better understand personal molecular biology and health.



population, allowing to develop greater understanding of causative factors and potential treatments of diseases. This would be especially relevant for rare genetic diseases, which require large datasets to find statistically important correlations.

How did it help during the pandemic?

In January 2020, at the start of the pandemic, Chinese scientist Yong-Zhen Zhang, sequenced the genome of a novel pathogen causing infections in the city of Wuhan, a *New York Times* report states. Mr. Zhang then shared it with his virologist friend Edward Holmes in Australia, who published the genomic code online. It was after this that virologists, epidemiologists, and pharmaceutical firms began evaluating the sequence to try and understand how to combat the virus, track the mutating variants and their intensity and spread, and to come up with a vaccine. This information was also used to create diagnostic PCR machines.

To enable an effective COVID-19 pandemic response, researchers kept track of emerging variants and conducting further studies about their transmissibility, immune escape and potential to cause severe disease. Genomic sequencing became one of the first steps in this important process. Here, the purpose of genome sequencing was to understand the role of certain mutations in increasing the virus's infectivity. Some mutations have also been linked to immune escape, or the virus's ability to evade antibodies, and this has consequences for vaccines and vaccine makers.

Over the course of the pandemic, the United States and United Kingdom scaled up genomic sequencing, tracked emerging variants and used that evidence for timely actions.

India also put in place a sequencing framework, and the Indian SARS-COV-2 Genomics Consortia (INSACOG), a consortium of labs across the country, was tasked with scanning coronavirus samples from patients and flagging the presence of variants known to have spiked transmission internationally. The bulk of its effort was focussed on identifying international 'variants of concern' (VoC) marked out by the World Health Organization as being particularly infectious. Samples from international travellers who arrived in India



As of early December 2021, the INSACOG had sequenced about 1,00,000 samples. It was also tasked with checking whether certain combinations of mutations were becoming more widespread in India.

In the later stage of the pandemic, around December 2022, when over 90% of the adult population was already fully vaccinated and over one-fourth of adults boosted, sequencing helped in targeted efforts at ebbing infections. The Health Ministry urged States to ramp up sequencing (and not increase testing) to track new variants as the virus evolved by accumulating mutations.

What is the significance of the Genome India project?

India's 1.3 billion-strong population consists of over 4,600 population groups, many of which are endogamous. Thus, the Indian population harbours distinct variations, with disease-causing mutations often amplified within some of these groups. Findings from population-based or disease-based human genetics research from other populations of the world cannot be extrapolated to Indians, says a note from the Indian Institute of Science (IISc). But despite being a large population with diverse ethnic groups, India lacks a comprehensive catalogue of genetic variations.

Creating a database of Indian genomes allows researchers to learn about genetic variants unique to India's population groups and use that to customise drugs and therapies. About 20 institutions across India are involved in the project, with analysis and coordination done by the Centre for Brain Research at IISc, Bangalore. The Centre's Department of Biotechnology notes that the project will help "unravel the genetic underpinnings of chronic diseases currently on the rise in India, (for) example, diabetes, hypertension, cardiovascular diseases, neurodegenerative disorders, and cancer".

