THE HINDU

NATIONAL

Published: May 21, 2016 02:39 IST | Updated: May 21, 2016 02:39 IST NEW DELHI, May 21, 2016

Largest gene database of Indians soon

- Jacob Koshy

Medgenome teams up with consortium that has committed to sequence 1 lakh Asian genomes

In a step to create one of the largest repositories of Indian genomes, Bangalore-based Medgenome has teamed up with a southeast Asian consortium that has committed to sequence 100,000 Asian genomes. Were it to work to plan this could mean a consolidated storehouse of at least 30,000 Indian genomes and could help understand the wide genetic variety in India's various ethnic groups and midwife customized medications for cancer and heart disease as well as identify possible new genetic aberrations that cause untreatable diseases.

Ever since the human genome was first sequenced in 2003, meaning, that the entire DNA pattern in the cell that lends people their unique identity was deciphered, several countries have announced initiatives to map genomes of their resident populations. The so-called 1000 Genomes project is a collection of gene samples from across the world to capture the variety of genes that are typical to different population groups. The United Kingdom announced a plan in 2014 to create a bank of 100,000 genomes in the nation and 100,000 Asia genomes project —called GenomeAsia 100K-- echoes similar ambitions. "Indian populations are greatly neglected in such databases," said Mahesh Pratapneni, Executive Director, Medgenome and a top official with the GenomeAsia 100K. "and glaring considering we're one-sixth of humanity."

The project will develop in phases with an initial 1000 genomes, consisting of India and East Asian populations, sequenced within this year and the entire database to be ready by 2020. Medgenome already has a bank of 200 Indian genomes.

$120 million project

The project will cost $120 million (approx. Rs 800 crore) though only about half of that has been firmed up. Other key collaborators in the project are Singapore's Nanyang Technological Institute, Singapore and Macrogen, a genetics diagnostic company in Seoul. Nearly 60 petabytes of data—equivalent to 30 trillion pages of text—are expected to be churned out in this study. Though all this data would be publicly available to researchers, access to it would be staggered. "We will release it all over 3-4 years but the main contributors to the project would access this earlier," Pratapneni told The Hindu.

Though human genome sequencing is a frontier area of biotechnology, it was prohibitively expensive. Technology advancement has made prices dramatically drop, enabling several companies to offer genome sequencing services. Experts however say that while the cost of sequencing has fallen it's the analysis of genes that adds value and that would mean being able to access and compare huge datasets.

While many diseases are linked to genes going awry, afflictions such as diabetes, cardiovascular diseases, cancer etc are usually the result of several genes malfunctioning, and often in a domino-like effect. Identifying such culprits are impossible without comparing genes, across individuals and population groups, in large numbers. Thus BRCA 1 and BRCA 2—genes associated with breast cancer—are found in as many as one-third of women. Several of them go on to live without ever contracting the cancer. These genes come in several varieties that can very on the level of families as well as ethnicities. Genome sequence studies are effective in studying such variations.

'Good initiative'

"It's a good initiative and could throw up valuable data provided there is a good study design in place," said Samir Brahmachari, former head of the Council of Scientific and Industrial Researcj and who led the country's first initiative to sequence an Indian genome.
