The life-saving sequence

Posted by By ink at 8 December , at 18 : 17 PM

A start-up in Bengaluru is part of the Asian genome project, the genetic study that could open up a world of personalised medicine.

BY GOVIND KRISHNAN V.

As the microprocessor revolution was sweeping Silicon Valley in the Seventies, and tech giants Apple and Microsoft came into being, a silent revolution was gathering steam in the world of molecular biology. In 1976 a team of researchers in Harvard, led by future Nobel winner Martin Gilbert, was inaugurating the genomic era with a project that sought to develop a dependable method to sequence DNA. A continent away, at the University of Cambridge, another team headed by Frederick Sanger was trying to achieve the same thing. Sanger had won the Nobel Prize in 1953 for sequencing the protein structure of insulin. Computers, which were to play such a significant part in the history of genetic sequencing, were just making their appearance in Sanger's lab, storing the data generated by the experiments. Both teams achieved their objective in 1977, finding a reliable and easy method to sequence DNA molecules into its component bases. This won Sanger his second Nobel Prize for chemistry, which he shared with Gilbert.

The Sanger method for sequencing DNA was the key that a generation of scientists would use to unlock the mysterious world of genetic inheritance. It became the driving force behind scientific efforts to map the genomes of various organisms—bacteria, microbes, and even human beings. In 1990, an international collaboration of scientists started sequencing the human

genome using Sanger's method. The human genome project sequenced the combination of all the 25,500 genes that make up the human chromosome. It took them 13 years and cost $2.7 billion.

Today, in Bengaluru, just beyond Electronic City, genome sequencing costs a little over Rs. 1 lakh ($1461), and takes three days. The computer revolution has made it possible to sequence entire genomes with less investment of time and money than Sanger or Gilbert could have fathomed.

***

MedGenome, an Indian biotechnological startup which does gene sequencing and clinical diagnostics, recently became part of the Genome Asia 100 K project, a not-for-profit international collaboration to sequence 1 lakh human genomes from Asia. The project will sequence the genomes of Asian populations to map the genetic variations among various ethnic sub-groups. It will take genome samples from more than 40 Asian countries and sequence them to create a standard reference genome. The database will be hosted by the Nanyang Technological University, Singapore. While members of the consortium will enjoy early access to the data, it will be released to the public and medical researchers within 36 months of completion.

Following in the steps of the 100 K Genome project in the UK, the Asian Genome Project builds on research which suggests that disease mutations present themselves differently in various ethnic groups, and aims to identify rare versions of the same gene associated with different populations. The immediate aim of the project is to help tailor medicine and healthcare to individual patients.

"Each individual's genetic makeup is unique," says Venkataswamy, a scientist with MedGenome, "While there is a standard human reference genome, each gene can have various versions called alleles. Four base pairs (AGCT) that code a particular gene can have slight variations in their combinations. This is why individuals differ. But so far drugs and medical care targets only the standard population." In a paper published in the *Annual Review of Genomics and Human Genetics*, Isaac Chan and Geoffrey Ginsburg wrote: "A broad aspiration of the Human Genome Project is the concept of personalised medicine—a rapidly advancing field of healthcare that is informed by the uniqueness of the individual. Although this concept is not entirely new, many clinicians have held great expectations for the development of medical diagnosis, prognosis, and treatment that could be based on an individual's genetic information. This variability of human health has long been recognized, but only with the advent of the genome sciences have the tools to understand this diversity become available."

DNA molecules are double helix shaped, like two rope ladders that twine around each other. In most living cells the DNA molecules which carry genetic information are in the shape of thread-like chromosomes. Human cells have 23 pairs of chromosomes. A pair comprises parts inherited from both parents, carrying different traits which determine the physical characteristics of the offspring. A gene is a small geographical section of the chromosome, like the sub-sections of a long highway, each marking a different town or city. A gene is the basic unit of heredity and controls certain traits. Just as computers are coded by zeroes and ones, genes are coded by 4 nucleobases, adenine, guanine, thymine and cytosine. The combinations of A, G, T and C make up the genetic code of a particular gene. When scientists sequence a human genome, they decipher the nucleobases that make up each gene.

***

When the paper was published in 2011, sequencing a single human genome cost $10,000. Today it costs about $1,000 in the United States, and between Rs. 1 and Rs. 1.25 lakh in India. The demand for low-cost sequencing led to the development of high-throughput Next Generation Sequencing (NGS) techniques. The rate decrease in sequencing cost per human genome outpaced Moore's law (a law in computing that predicts processing power doubles every year) in 2007-08, falling from $10 million to $1million in a year. Over the next two years, the cost fell further by ninety per cent.

The drastic fall in cost has had two effects: making population-wide sequencing projects possible, and allowing companies like MedGenome to offer physicians clinical diagnostics of their patients by analysis of genetic samples.

The Asia Genome consortium says the first short-term benefits of the data they gather would be in cancer research, a field directly linked to genomics.

Arathi Khanna Gupta joined MedGenome after researching the molecular basis of leukaemia at Brigham and Women's Hospital, Harvard Medical School. It is her interest in cancer therapy that led her to Bengaluru.

She says: "Medical care for cancer has so far followed a one-size-fits-all-approach. This is not necessarily true. There are variations at the genetic level among various populations, which medical care has to recognise better. Right now, chemotherapy for lung cancer is like hitting a nail with a sledge hammer. Chemotherapy kills all fast growing cells. So it also kills hair, nails, skin cells, etc. Another huge factor is that all drug trials are conducted on Western populations. We haven't studied their effectiveness in non-white and Asian populations. Asian populations also display higher genetic diversity than the West, because of the higher prevalence of endogamy. India has particularly high genetic variation because of caste endogamy. Endogamous communities often develop genetic disorders because weak traits do not get weeded out of the closed gene pool and stronger genes do not enter it. Thus a genomic population study is extremely important."

As we enter the laboratory wing of MedGenome, Venkataswamy gestures to a rack filled with plastic floaters of the brightest colours—red, purple, blue. Scientists in white lab coats slipping in and out, exchange their outdoor footwear for the funky lab-only floaters. The lab is divided into three areas behind glass wall partitions—sampling where the DNA is extracted, preparation, and sequencing. Venkataswamy leads me into the sampling section, where vials of blood tissue are stored using anti-clotting chemicals. Some are blood samples from various clinics intended for the genome projects. Others are samples sent for analysis by doctors for clinical genetic tests.

"Many come from oncologists or neurologists. When cancer is suspected before there is a tumour, the presence of mutations in genes associated with a cancer type can be used to detect it. Many neurological diseases share the same symptoms, so sequencing genes that cause different diseases is used as a diagnostic tool," he says. In some cases, doctors ask for a test on specific genes. But usually, they ask to screen for a particular disease, and MedGenome's scientists create a gene panel associated with that disease to sequence it.

While DNA can be extracted from all kinds of tissues including saliva and skin, almost all the 2,000 samples collected so far for the Asia Genome project have been blood tissues, Venkataswamy says. He goes on to explain how DNA is extracted from the blood samples. Cells, which are protected by a cell

membrane, consist of a cytoplasm in which the nucleus is immersed, along with other organelles. The DNA is contained as chromosomes in the nucleus.

To isolate the DNA from a sample, a scientist lyses the blood tissue with a protein and a reagent grade detergent. The protein and detergent are harmless to the DNA but soon break down everything else in the cell; dissolving the proteins, fats and lipids that make up the cell membrane, organelles and cytoplasm. The DNA is freed. The solution is now transferred into a tube with a column in the middle. By adding a chemical buffer the DNA gets attracted and stuck to the column. Dozens of such tubes are now put into a centrifugal machine that rotates at blinding speed, draining away the solution of cell debris. In minutes, he picks out the column tubes containing almost pure DNA. This is now transferred to a sonication machine, where it will be prepared for sequencing. The sonication device produces waves of ultrasonic sound which fragments the entangled DNA strands into smaller bits that can be analysed. The scientist adjusts the strength of the waves on the device, breaking the DNA into fragments of pre-determined size.

The final stage is the sequencing of the DNA. MedGenome uses NGS sequencing machines from Illumina, a company that partners in the Asia Genome project, and which pioneered current NGS methods. The method allows the parallel sequencing of several DNA samples, with the entire process requiring around three days. A chemical process attaches oligoneucleotides to the end of the each DNA fragments. Each set of the oligonucleotides has copies which are loaded on to the flow-cell of the sequencing machines. The flow-cell has millions of small perorations where they attach themselves. When the fragmented DNA is introduced into the flow-plate, the oligonucleotides bond to their copies and the whole DNA sequence starts multiplying into large clusters, which can now be identified optically by the machine. Knowing which oligonucleotide has attached to which kind of DNA, the sequence can be understood and recorded.

***

In 1977, an unusual pair of virologists in California was cracking open one of medicine’s deepest mysteries—how is cancer caused? Michael Bishop had turned from the study of history to medicine and then virology. An even more colourful character, Harold Varmus, studied mediaeval English literature before becoming a doctor and practising medicine across the world, which included a year’s stint in Bareilly in Uttar Pradesh. His academic wanderlust finally led him to Bishop and the investigation of the Rous Sarcoma Virus, which caused cancer in chickens. Even after decades of research, scientists did not know what caused mutations that turned normal cells cancerous. The two major competing theories pitched virus infections that introduced a cancer gene in the cells of the host organism, against environmental factors like smoking in humans, which caused genetic mutations that led to cancer. The young virologists discovered something entirely different—the Rous Sarcoma Virus did not introduce a new gene into an animal’s body, it shared a gene with the organism, which it activated. In other words, oncogenes, or cancer producing genes, were the internal enemy, part of the genetic map of all living organisms.

Biologists already knew that cancer cells were different from normal cells in only one respect: they never stop reproducing. The discovery of oncogenes allowed scientists to analyse the mechanism behind the genetic mutations that turn healthy cells into cancer cells. Oncogenes regulate cell growth and reproduction. Mutations in oncogenes can lead to over-expression of that gene; or mutations can lead to new oncogenes forming. Another mechanism by which cancer is caused involves mutations in tumour-suppressing genes.

These genes prevent cells from growing indefinitely. Mutations can switch off the action of the tumour-suppressing genes or underplay their expression. More than one gene is associated with a particular kind of cancer and usually mutations in multiple genes happen before a normal cell becomes a cancer cell.

From 1990 to 2015, the mortality rate of almost all cancers fell by one per cent a year. By the mid-Nineties, several oncogenes and tumour-suppressing genes had been identified. Drugs that target these genes were developed. Unlike chemotherapy, which uses toxic drugs to destroy cancer cells, targeted drugs focus on cancer genes and deactivate them. Twenty-four targeted drugs are now available for cancers, including leukaemia, prostrate, breast, sarcoma and cervical cancers. More are in the pipeline.

***

Arathi Gupta believes the next step in the fight against cancer is evolving from targeted therapy to personalised or precision medicine, using population genomics. “Your genetic predisposition can determine therapy in many ways. It can help tailor the drugs used. In lung cancer, the most common is non-small cell carcinoma. Thirty per cent of patients with this type of cancer have mutations in the EGFR gene. If the patient has this mutation, they are receptive to certain drugs, which does not work in the rest of the patients whose cancer is not caused by EGFR mutations. It can prolong life by nine months. You might respond to a particular medicine because of your predisposition. Cancer medicine must move towards treatment regimes tailored for a particular individual in deciding drug type, drug dosage and preventive care. For other diseases, too, most of the drug trials have been carried out on western subjects. Genome sequencing of Asian populations would help identify different alleles, disease carrying mutations, associations of genetic sub-groups with various diseases and differing disease rates.”

Apart from benefits to medical research, the project expects the data to feed into studies in the history of population migration as well as future applications that might be developed. Mahesh Pratapneni, the CEO of the 100 K Asia Genome project says that from the samples collected so far, they have identified a set of sub-groups that will be the basis of further sampling and population analysis. “We are only interested in genetically unique sub-groups. None of the metadata will be in the public database. We are stratifying the sub-groups through further analysis. From an initial set of about 2,000 genomes we have identified 50 plus sub-groups. We think at least 50 samples are needed from each group for population genomics analysis. We have a sampling strategy that is being detailed right now. It is a combination of ethnic diversity discovered through initial 2,000 sample analysis, clinical conditions of interest and availability. The (initial) population genomics data is being analysed. We are working towards a reference resource that can enable drug discovery and precision medicine initiatives at the least. Just as with any platform project we believe that many innovative use cases will emerge as we move forward just as we could not have predicted an Uber 15 years ago when the US government put GPS out of military use.”

There are scientists who remain sceptical of the viability of personalised medicine as well as whether genome sequencing is the way to achieve it. They see it as a case of the availability of technology overriding research goals. Aswin Sai Narain Seshasayee, a biologist with the National Centre for Biological Sciences (NCBS) in Bengaluru, says: “Mutations on a genome do not act alone, and the manner in which one mutation might affect the effect of another on a cell or organism’s characteristics is not well understood. This is further complicated by the fact that any individual genome of a complex

organism may not carry just a few variations with respect to a ‘healthy reference’, but is likely to carry even thousands or hundreds of thousands of such variations, depending on the size of the genome. I am not sure that we will be able to achieve a deep understanding of such genetic interactions in the near future. One could argue that statistical analysis of big data could overcome the lack of understanding of genetic interactions; however, I am not convinced that we will have enough power to get the statistics to the point that it is predictive enough to drive ‘precision’ ‘personalised’ medicine. Biochemical readouts, derived from metabolic processes, may be a better way to approach this, because these might mask the complexity arising from such genetic interactions. However, at this point, genome sequencing works very well as a technology, because it is a lot easier to do than any biochemical experiments on a large scale.”

Oncologists, however, are more hopeful. Dr. Santosh Gowda, who heads the oncology department at the Mazumdar Shaw Cancer Centre in Bengaluru, thinks personalised medicine in cancer is the way forward. “Though cancer is generated by environmental factors also, it is mainly germline mutations that cause it. But each individual’s DNA is unique, like a molecular fingerprint. Right now, cancer drugs are randomised for the entire population and do not take into account the 10 per cent or so variation that exists. Metabolic cancer drugs for prostate cancer are very aggressive (with side effects). Individual variations are not taken into account, treatment modifications are not in place. Can we tailor them to respond to genetic reasons? Among Asian women lung cancer patients, the percentage of non-smokers is 50 per cent, while it is 20 to 25 per cent among Caucasian women. This is why the argument for genetic sequencing of populations becomes important.”

In the 2016 US Federal Budget, President Barack Obama allotted $215 million for the Precision Medicine Initiative (PMI) a programme to promote biomedical research that would help doctors and healthcare organisations tailor treatment to individual patient needs. Among many of the projects PMI has launched is one that plans to enlist a million volunteers to share their medical and genetic details to create a massive databank. Similar genetic databanks are under construction in other countries including the UK and China.

PMI allotted $71 million to the National Cancer Institute (NCI) for research into precision oncology. The NCI has done a series of clinical trials including an ongoing one where patients are treated on the basis of the genes that are driving their cancer, rather than the type of cancer. Precision medicine includes not only genetics and genomics, but also big data and analytics. In the case of cancer, NCI notes that there are disparities between population groups, with minority groups showing higher incidence and mortality.

In 2012, a genome sequencing study of 14,000 Asian women from China, South Korea, Taiwan, Singapore and Hong Kong found variations in three places in the genome which predisposed non-smokers to lung cancer. The study found that the molecular basis of lung cancer can be different for smokers and non-smokers.

A survey released this year by the SAP and Oxford Economic Survey (not connected with Oxford University) found healthcare organisations in Europe, Canada and the US prioritising precision medicine around diabetes, cancer, neurological diseases, ageing and autoimmune diseases. For example, some cardiologists reported that they have moved away from standard doses of medicine for heart attacks since the metabolism rates for individuals vary.

Pharmacogenetic testing on how different people’s genetic makeup influence their response to drugs, allow them to calibrate drug dosage to individual

patients. The study, which surveyed 126 physicians, said that 49 per cent of the doctors surveyed reported significant improvement in patient care because of personalised medicine practices, while 53 per cent expected this to happen in the next two years. The survey also found that some pharmaceutical companies have started using biomarkers in their drug testing, identifying patient sub-populations and utilising genomic information to understand their responses.

***

“Precision medicine would individualise diagnosis, treatment and prevention,” Dr. Gowda says. In his own practice, he uses genetic analysis for preventive treatment of patients at high risk of breast cancer. Current studies estimate that 55 to 65 per cent of women who inherit the BRCA gene will develop breast cancer, while 45 per cent of the women who inherit BRCA2 will develop the disease in their lifetimes. With the lowering cost of genome sequencing, more Indians are able to afford the testing. Gowda also advises preventive diagnosis for patients who have a family history of cancer. “In the US, they have now developed a 21-gene panel to test for inherited germline mutations. The test is not available in India, so in some cases we sent the samples to clinics there,” he says.

Precision medicine is in its early stages, and both optimism and scepticism has been expressed about its potential. Critics of the PMI programme point at the high cost, uncertainty and concerns over data privacy. The SAP and Oxford Economic Survey said only 32 per cent of doctors it interviewed expressed confidence in their ability to protect the genomic information of their patients. In an article in *The New York Times* Dr. Michel J. Joyner, a physician at Mayo Clinic, wrote: “Given the general omertà about researchers’ criticising funding initiatives, you probably won’t hear too many objections from the research community about President Obama’s plan for precision medicine. But I am deeply sceptical. Like most moonshot medical research initiatives, precision medicine is likely to fall short of expectations. Medical problems and their underlying biology are not linear engineering exercises, and solving them is more than a matter of vision, money and will.”

On the other side of the fence, a prominent scientific voice who has put his weight behind precision medicine is Harold Varmus, the virologist who discovered that cancer is a genomic disease. In an article co-authored with Francis Collins (head of the human genome project) in the *New England Journal of Medicine* (January 30, 2015) he says about the PMI project: “But the prospect of applying this concept (precision medicine) broadly has been dramatically improved by the recent development of large-scale biologic databases (such as the human genome sequence), powerful methods for characterising patients (such as proteomics, metabolomics, genomics, diverse cellular assays, and even mobile health technology), and computational tools for analysing large sets of data… With sufficient resources and a strong, sustained commitment of time, energy, and ingenuity from the scientific, medical, and patient communities, the full potential of precision medicine can ultimately be realised to give everyone the best chance at good health.”

Towards the end of his book on the history of cancer, *The Emperor of Maladies*, Dr. Siddartha Mukherjee constructs an imaginary journey through time. The time traveller is Atossa, the Persian Queen who lived around 500 BCE. She is the first known personality in history to suffer from cancer. Finding a tumour in her breast, Atossa had her Greek slave perform a primitive mastectomy. It is not known whether Atossa was able to excise her cancer and for how many years she survived the surgery. Mukherjee imagines

Atossa travelling the temporal landscape of cancer, arriving at various points in history to receive treatment. Her chances start to improve as she enters the 20th century and peaks as she enters the 1990s. She is diagnosed at an early stage and screened to see if a tumour will appear in her unaffected breast. Her genome is sequenced and a mutation is detected in BRCA1. Her two daughters are also tested and given preventive treatment. When one of them develops a tumour, early surgery saves her life.

But the scenario is completely different when Mukherjee fast-forwards Atossa to 2050. Mukherjee, so far the grim chronicler of humanity's trench warfare against an ever-mutating foe, allows himself the audacity of hope.

"In 2050, Attosa will arrive at her breast oncologist's clinic with a thumb-size flash drive containing the entire sequence of her cancer's genome, identifying every mutation in every gene. The mutations will be organised into key pathways. An algorithm might identify the pathways that are contributing to the growth and survival of her cancer. Therapies will be targeted against these pathways to prevent a relapse of the tumour after surgery. She will begin with one combination of targeted drugs, expect to switch to a second cocktail when her cancer mutates, and switch again when the cancer mutates again. She will likely take some form of medicine whether to prevent, cure, or palliate her illness, for the rest of her life." Atossa survives.