The challenges of genetic research in India

S K Mahajan

One day in early 1993 a doctor from BARC hospital brought to my office a friend from Kota in Rajasthan. The latter, a middle-aged medical practioner, had a son afflicted with DMD (Duchenne Muscular Dystrophy). He also had an unaffected daughter but did not know whether she also had the defective dmd gene on the X chromosome inherited from her mother. He had read that molecular biologists in the US could now carry out such carrier detection assays in a majority of DMD families and wanted to know whether we in India do the same. I told him that we had the basic infrastructure and capability for performing the PCR-based assay for detecting deletion mutation which are present in nearly two-thirds of all dmd cases. We also had a programme for detecting microsatellite polymrphisms in humans which could be used as linked markers for dmd carrier detection. However, genetic disease detection was not a part of our molecular biology progamme at that time. We therefore did not have the specific primers and a hands-on experience in dmd carrier detection and could not be of direct help to our visitor. I also did not know any one else who was conducting such a test at that time.

A Sunday afternoon a few weeks later, my friend Dr. Hemu Adhikari, who was then a senior scientist in our Food Technology Division but is better known to outsiders as an actor on Marathi stage and screen, dropped in at my home with two of his friends. The latter had a BMD (Becker's Muscular Dystrophy) patient in their family and were extremely well-informed about DMD, BMD and many other genetic diseases including the latest molecular diagnosis techniques that were becoming available for more and more diseases, thanks to the excitement and interest generated by the Human Genome Project (HGP). They naturally wanted to know what Indian molecular biologists were doing in the matter.

Genetic diseases and the HGP did not then invoke much interest within the Indian medical community. Even the more concerned and informed medical friends felt that our concern in India should be largely confined to infectious diseases which affect many more of our people and could be managed inexpensively by us. But the family members of those afflicted with a genetic disease thought very differently. They could see that the cost of keeping a single such patient alive and reasonably well looked after far exceeded the expenses in managing hundreds of infection episodes. Though the parents of a DMD boy loved their son they were not keen to have another affected son or grandson. They had seen the suffering of the child, experienced the trauma of the family members and were willing to go any extent to know more about the disease, cling to any hope of remedy of symptoms or the cause, and avoid births of more such children. To them the HGP was not a luxury relevant only to the rich nations who had largely conquered the infectious diseases. They saw in it a ray of hope. They almost demand work on genetic diseases, which, we now know, nearly all diseases are. I had no option but to convince my colleagues to initiate work on DMD diagnosis and they agreed even though a routine disease diagnosis is not considered an exciting prospect by most research scientists.

Why is it that the Indian medical community has been so little interested in human genetic diseases? Though the attitude of the younger doctors and medical students is changing, the Human Genetics Task Force of Department of Biotechnology had a major difficulty in establishing clinics for diagnosis and counseling even for beta Thalassemia which affects a large number of Indians in several communities. The money was available, and it was usually enough to interest researchers struggling to fund their laboratories, but the patients go to doctors and not many of them found it worth their while to get interested.

The situation is not unique to India. The reasons lie in the history of human genetics. Genetics is itself a young science, having just completed a hundred years of its formal existence. Before that very little was understood about the mechanisms of inheritance, though the fact that humans, like other living organisms, inherited their characteristics, including some diseases, from their ancestors was known to even primitive people. But not knowing how this happened, there was no question of intervention, and we reconciled to our helplessness by passing the entire blame to the divine who created us.

The best we could do was to choose better partners for ourselves and for the plants and animals that we bred. The real progress started only with Mendel and the pace accelerated after the rediscovery of Mendel's insights by more influential quarters in 1900. However, that was not enough for the health of human genetics. Humans had to wait for the birth of genomics around 1980. The story is briefly worth telling.

Though the A, B, O blood groups were discovered in 1900 and the Mendelian nature of the inheritance of the determinants of human genetic diseases was recognised soon afterwards, progress in human genetics was relatively slow during the first seven decades of the 20th century. The main reason for this was the non-availability, in humans, of the two main tools of the classical genetic analysis — namely, the induced mutation and the experimental genetic cross involving large purebred populations. This made mapping of the genes associated with specific diseases and other phenotypes, normally the first step in further analysis of the related character, extremely difficult. Only in some cases, the genes could be traced to the X chromosome due the peculiar inheritance pattern of the disease phenotype in these cases.

In the 1970s, the development of the techniques of somatic cell hybridisation on one hand, and gene cloning

***S.K.Mahajan**, Head,Molecular Biology and Agriculture Division, Bhabha Atomic Research Centre, Mumbai 400 085. Email: sk_mahajan@yahoo.com*

on the other, accelerated the pace of genetic analysis in humans. But the real breakthrough came with the development of DNA sequencing techniques in 1977. This very soon revealed the presence of widespread sequence polymorphisms which could be used as high density genetic markers on the human genome and permit gene mapping by studying linked segregation of heritable phenotypes (such as genetic diseases) in naturally occurring families with one of the easily scorable and tightly linked marker with a known map location. It may be mentioned here that even before the whole genome sequencing was started, other high density markers, such as RFLP (restriction length polymorphism) and microsatellites became available in the 1980s. These, particularly the latter, played a crucial role in preparing detailed genetic maps of the human chromosomes and in positional mapping of disease genes.

However, as we see below, with the completion of the HGP the markers of choice have become the SNPs (Single Nucleotide Polymorphisms) whose number on the human genome runs into millions, making several of them tightly linked to any gene under investigation.

It was also recognised that the availability of a large number of genomic markers likely to result from whole genome sequencing would permit detection of small increases in mutation frequencies caused by environmental mutagens like ionising radiations even when the exposed populations were relatively small. The whole genome sequence was further expected to provide the repertoire of all the genes present on the genome, thereby accelerating the discovery and characterisation of genes associated with heritable genetic diseases, permitting molecular dissection of non-heritable genetic diseases like cancer and, in general, helping in understanding human biology.

It is no wonder that, from the early 1980s onwards, the idea of sequencing the entire human genome, though a daunting task for the then available technologies, started gaining steady momentum and eventually led to the launch of the human genome project in October 1990.

Initially two US government agencies, the Department of Environment and the National Institutes of Health, were the main players though very soon many others, including several European governments, and private companies in the US and elsewhere, joined hands. The initially projected budget was three billion dollars and the target date for sequencing the entire human genome was 2005. The original plan was to first prepare genetic and physical maps of all 24 chromosomes, generate ordered overlapping clones of sequenceable length for the entire genome, develop improved technologies for rapid, automated sequencing and sequence analysis, and then launch a large scale sequencing effort in many laboratories on a division of labor basis.

The first five years of the project saw the generation of high density genetic and physical maps, which permitted the identification of a large number of new disease genes, on schedule. However, rapid developments in technology and sequencing strategies, and a large influx of private money looking for patents, soon led to major departures from the original plan and the project implementation was accelerated. The big turning point came in 1995 when Craig Venter proposed the so-called shotgun sequencing strategy which was to totally alter the rules of the game and eventually lead to the generation of a draft sequence of the entire genome five years ahead of the original target. The existence of two draft sequences, one generated by the DOE/NIH /Sanger Centre consortium and the other by CELERA, was jointly announced on June 26, 2000, and detailed versions were published in March 2001.

The public sector consortium, which involves 16 sequencing centres, makes its data available to the public periodically, while the CELERA database is accessible only to subscribers. In the public sector consortium, more than 90% of the 3.2 billion bases of the genome have now been sequenced at better than 99% accuracy. Of this about 55% is in its final form with less than 1 error in 10,000. The entire sequence is expected to be available in finished form by the summer of 2003, or earlier, with an accuracy of 99.99%. The next step of annotation, i.e., identification of the coding sequences (genes) and the functions of their protein products, may take much longer.

It is now expected that about 2% of the human genome may consist of exons (coding sequences) and the total number of genes may be 70,000 or so**,** though this number appeared much smaller (approx 34,000) when the sequence was published early this year.

Nearly 7,000 annotated protein sequences are available in public data banks. The 3-D structure of some 800 of these is known. The number of protein variants with a known disease mutation or polymorphism is nearly 9,900.

The identification of the functions of the protein products of these genes follows several strategies which have given birth to the new fields of functional genomics, proteomics and structural genomics. These are changing the very nature of biological research which has necessitated training of a new kind of biological scientists with a broad interdisciplinary background in computer sciences, physics, chemistry. And even engineering.

While 99.9% of the sequence in different humans is identical, point mutations or single nucleotide polymorphisms (SNPs) occur every few hundred bases apart. Specific SNPs can affect individual response to diseases and environmental agents like infections, toxins, and chemicals including therapeutic drugs. Even when present outside the coding regions, due to their high density SNPs can be extremely useful in monitoring segregation of specific alleles in small populations and will help in identifying the multiple genes associated with complex diseases like cancer, diabetes and vascular diseases. They may also be helpful in designing customised, individual- or ethnic group-specific drugs. A great deal of work is therefore going on in identifying SNPs in various population groups, both in public- and privately-funded sequencing groups. More than two million SNPs are now available in the public database.

The goals of functional genomics include the creation

of full length cDNA and clones for all human genes, and the study of the expression and control of these genes in different tissues, in genetic as well as infectious diseases, and in response to drugs and other environment changes. Several new technologies such as DNA and protein chips, MALDITOF mass spectrometry, large-scale protein structure determination using Synchrotrons and Bioinformatics are proving useful in this effort. Another major tool is the generation and study of knockouts in non-human organisms, especially the mouse. These should provide a better understanding of human biology and also make disease management much more sophisticated and effective.

Sequencing of several model organisms, namely, *E. coli, yeast, D. melanogaster, C. elegans*, and mouse, was conceived as a part of the HGP from the very beginning. Of these the first four genomes have been completely sequenced and the work on the mouse genome is progressing parallel to the human genome. It is expected to be completed by 2005 or earlier. Several other genomes have also been sequenced, or are being sequenced, by various groups. Fifty-one bacteria have been completely sequenced and another 76 are being sequenced. A large number of animal and plant genome sequencing projects are either in progress or on the anvil. They would provide a major resource for comparative genomics and proteomics and help in ascertaining the functions of specific genes and their protein products and give new insights into evolutionary, biochemical, genetic, metabolic and physiological pathways.

The results of the HGP and the allied genomic sciences (functional genomics, proteomics, structural genomics, and bioinformatics) are likely to revolutionise medicine and health care. The most important impact will be in the area of predictive diagnosis of genetic diseases and susceptibility to infectious, environmental and multi-factorial genetic diseases. Neonatal diagnosis coupled with counseling may prevent more serious genetic diseases while suggested changes in life style indicated by an individual's SNP profile may reduce development of such diseases as cancer, asthma, diabetes or CAD.

Many new drug targets are likely to be suggested by the structures of genes and proteins involved in specific pathways, and the repertoire of new drugs is likely to increase from the present 2,000 to nearly 12,000 during the next 20 years. Since individual genetic profiles would make adverse reactions and side effects of specific drugs predictable, medical intervention may become more specific, precise and successful. Clinical trials may require matching of genetic profiles of the control and test individuals and their results may show much smaller statistical fluctuations. Finally, several of the common single gene diseases, and perhaps even multiple gene diseases, may become amenable to gene therapy.

The availability of the tools of predictive diagnosis raises several ethical, legal, and social issues, and medical practitioners will have a crucial role to play in discussing these questions. The most important of these will be the question of genetic privacy with bearing on the question of discrimination in insurance, employment, classroom, and courts of law, and even in family set-ups. Another, and perhaps even more serious, issue arises from the fact that predictive diagnosis, especially when carried out on early embryos, automatically provides a basis for genetic selection, with eugenic potential. Deciding about which genome has a right to life and propagation can be very tricky.

These issues need serious discussion within the medical community to guide the public at large. Since both the repertoire of genetic mutations and the socio-cultural parameters are different in India it will not be wise to depend on the decisions made by other nations and cultures. The Indian medical community has a responsibility here. Is it ready to accept the challenge?