Review

HUMAN GENOME PROJECT AND ITS IMPACT ON DIABETES MELLITUS

G.R. Sridhar

ABSTRACT

The human genome project (HGP) has made available the draft sequence of 3.1 billion nucleotide bases of human DNA, through a confluence of diverse fields. It marks the historic beginning of a new era when health and disease would be understood on the basis of genetic susceptibility. Post HGP, advances in diabetes can be expected in understanding its etiology and management. Putative susceptibility genes were localized to chromosomes 2q37, 12q, 3q and 20. A hitherto unsuspected gene for calpain 10 protein, may be a new marker. Ultimately, it would be possible to identify genes as the most 'proximal phenotype.' Armed with this information, interventions can be planned to favorably alter the gene-protein expression.

KEY WORDS: Gene-expression; Genome; Proteinomics; Bioinformatics; Calpain; Ethics.

INTRODUCTION

Health and conversely disease, result from an interaction of genes and the environment. Medical practice, has until now, relied on modifying the environment to seek health and to control disease by depending on anti-infective agents, chemicals, temperature, diet etc. That was because we did not know, nor could we do much, about the genes.

Post Human Genome Project (HGP), we are poised at begining of a historic era. We would be able to understand health and disease in terms of genetic background.

There has been exceptional media coverage of the event. HGP was called, 'biology's holy grail'. Others said it was the greatest intellectual moment in history. Yet another exclaimed! 'history books will mark this as the ceremonial start of the genomic era'. In this explosion of superlatives, we should consider what the HGP means to clinical practice.

HISTORICAL BACKGROUND OF HGP

The first proposal for sequencing the entire human genome was made in 1984, with resultant formation of The Human Genome Organization, four years later (1). The body of international cientists, who collaborated with the National Center for Human Genome Research, formally began work on HGP in 1990. Initially, sequencing the human genome was planned to be completed in 15 years. However, with the independent establishment of "The Institute for Genetic Research", headed by Craig Venter, a hectic pace was set, and the joint announcement that the human genome was sequenced, was made in June 2000, much ahead of the scheduled time frame.

26th June 2000

On 26th June 2000, the Human Genome Consortium and Celera Genomics, jointly announced the successful completion of the human genome project. What this meant was the availability of a draft sequence of the 3.1 billion nucleotide bases of the human DNA. The number is large. To put it into perspective, if one could read one nucleotide base every second, it would take 100 years to complete reading the entire sequence.

GOALS OF HGP

The goal of the HGP was to first ‘construct a comprehensive genetic map of the human genome, complemented by the sequencing of model organisms' (1). The other goals were to (a) analyze sequence variations in the human genome (b) develop technology for functional genomics, (c) consider ethical, legal and social implications and (d) develop bioinformatics and computational biology (2). It was for the first time that advances in biological sciences depended on advances in information technology. Similarly, evaluating ethical and social aspects that may occur as a result, were a noteworthy aim, and not an afterthought.

SOURCE OF DNA

More than 99.8% of the 3.2 billion base pairs, between any two humans, are the same. DNA samples from multiple anonymous donors were collected in accordance with the international review board protocol. Blood was taken from female donors and sperm from male donors (3). 20 to 50 samples were combined and used for making a genomic library and sequencing followed this. Total anonymity was observed for all the procedures.

Endocrine and Diabetes Centre, 15-12-16 Krishnanagar, Vishakhapatnam 530 002

TOP DOWN APPROACH OF HGP

The HGP consortium used the top down approach, in which the genome was segregated stepwise, into small segments. When the pieces were small enough, they were inserted into a vector and sequenced. These were then pieced together by back tracing, to the point of origin on the chromosomes.

BOTTOM UP APPROACH OF CELERA

In contrast, Celera Genomics applied the 'shotgun approach', where DNA was broken down into small fragments and then sequenced in an unbiased, brute fashion. Overlaps among fragments were identified and through computing-intense procedures, the sequence pairs were identified. They used high-speed sequencing with fluorescence markers.

CURRENT STATUS, PUBLISHED RESULTS

The draft sequence of the human genome was published in early February 2001 by both groups: by the International Consortium in *Nature* (Nature, 2001; 409:860-921) and by the Celera group in Science (Venter JC et al, Science 2001; 291:1304-51).

Essentially the number of bases range from 2.69 billion (International Consortium) to 2.91 billion (Celera group). The numbers of genes appear to be around 30,000-40,000, much less than the earlier estimates. The human genome is mostly a 'barren stretch of DNA bases, that do not code for anything' (Curr Sci 2001;80:483). The number of proteins, i.e. the total complement of proteins or the 'human proteome' seems to be higher, around 60,000.

GENOME RELATED WEB SITES

Data of the human genome may be accessed from publicly available web sites, that include

(a) http://www.nhgri.nih.gov/
(b) http://gdbwww.gdb.org/
(c) http://hgmp.mrc.ac.uk/
(d) http://www.celera.com

GENOMIC SCANS FOR TYPE 2 DIABETES MELLITUS GENES

Attempts are on to complete genome scans for type 2 diabetes mellitus. Published data suggests significance on chromosome 2q37, 12q, 3q and 20 (4-7). A more recent report uncovered an unexpected gene, which could make individuals susceptible to type 2 diabetes mellitus (Nature Genetics 2000; 26:163-75). The calpain 10 gene was associated with the development of type 2 diabetes mellitus, and could be a new marker, indeed possibly a new therapeutic agent.

Using positional cloning, which assumes no prior knowledge of a gene's function, calpain 10 was picked up as a candidate gene. However, positional cloning utilizes the principle that genes that are physically close would segregate ('guilt by association'). The calpain 10 with 15 exons, was shown to have three single nucleotide polymorphisms (SNPs), which were consistently associated with type 2 diabetes mellitus. It is possible that calpain 10 could function as a regulatory protein that modulates glucose response, insulin secretion or receptor interaction. Identification of such 'proximal phenotypes' could offer treatments of the underlying molecular defects, rather than merely the symptoms.

APPLICATION OF INFORMATION FROM HUMAN GENOME PROJECT

These are still early times to directly apply information obtained from the HGP to clinical use. Much work still needs to be done. One can foresee progress in the following areas:

Identification of Genes

More than 98% of the DNA is described as 'junk DNA', i.e. it does not directly code for any known genes, and we do not still know what the exact role of these intervening sequences is. The first aim would be to identify which part of the DNA codes for known proteins, which are the intervening sequences, and what, if any, the function of the different sequences is.

Function of Genes

Once the genes are identified, the next logical step is to understand the function of the identified gene and what role it plays in the entire physiological process. With less than 5% of the genomic DNA coding for proteins, identifying the functional genes is currently akin to picking a pin from the proverbial haystack.

DNA Chips

It is possible to construct DNA chips, in which a sample of DNA can be applied to a chip containing known gene fragments. A marker molecule can then identify the presence of DNA in the sample. In this way it is possible to rapidly isolate the genes of interest, in a given specimen.

Proteinomics

Attempts are on to identify the exact three - dimensional structures of nearly 10,000, biologically important, proteins. These are the logical amplifiers of information which is coded by the DNA. Functional proteinomics is the next step following sequencing the human genome, because proteins are, after all, the ultimate distal mediators of physiological action. Attention is now being focussed on this exciting new field of research, made possible by genomics.

Designer Drugs

It may be possible to develop drugs that are useful in an individual with a particular genotype. These drugs can be tailored to be effective and to have no side effects in that particular person. The downside of this is that drugs would be so specifically designed, that they cannot be used generically for persons with different genotypes. We would thereby lose pharmaceutical social security, when drugs may perhaps be available for only those who have access to such designer molecules.

INDIA AND THE HUMAN GENOME PROJECT

India did not actively participate in sequencing the human genome. Recently a 'Functional genomics project' has been initiated to use 'genome sequence data to identify genes arid the association to specific genes and their variation among populations.'

With a strong base in science, and a good manpower in information technology, India has the potential to develop bioinformatics (8), which fuses basic sciences including physics, chemistry, mathematics, molecular biology and information technology. It was proposed to set up a National Mission for analysis of genetic sequence data, including the human genome sequence. It would contain a database of all genome sequences, with appropriate annotation and generation of search engines. This would be supported by statistical and computational methodologies to identify disease causing genes and:mutations, identification of drug targets and enhance the study of evolution (Bhargava PM. The Hindu 9 Nov 2000; Sciencel Business section: pp7).

PROJECTED FUTURE SCENARIO

With the widespread access of the human genome sequences, identification of gene polymorphisms would permit testing the association of gene/s to type 2 diabetes mellitus (9). By cataloging the full complement of genes expressed in the beta cell and by understanding the function of proteins coded for by the genes, should help us to understand the pathophysiology of type 2 diabetes mellitus. In a way, identification of the genes would mean identifying the most 'proximal phenotype' (9). With this information, interventions can be planned to favorably alter the gene and protein expression.

Down the line, ultimately the aim is to comprehend the whole biological chain of events when disease strikes and then pick the best place in that chain to intervene. It is also possible to develop a network of databases so that clinical information as well as knowledge from molecular biology is available for preventive and curative service (10).

DRAWBACKS FROM THE AVAILABILITY OF GENETIC DATA

There are disadvantages that may result from knowledge of the human genome.

Privacy. Loss of privacy is a real threat. It could be possible to collect a biological specimen (e.g. a few cells from a thrown handkerchief or tissue paper) and then identify the genes that can determine information about oneself.

Genetic determinism. This is another important limitation. Genetic determinism presupposes that just because a gene for a particular quality exists, the phenotype automatically follows. Such a course is not true. Phenotypic expression is determined by a variety of factors other than the presence of genes (e.g. by environment, nutrition, psychological factors). There is a real danger of genetic determinism.

Patenting and ownership. Patenting and ownership is likely to be a serious problem. Scientists, commercial organizations or others may claim patency on genes, which raises a series of legal and ethical issues on who the 'owner' of the genes is.

Germ line modification. Modifying genes to produce individuals with 'desirable characteristics' is another possible conundrum. Technical advances should result in our ability to modify the germ line. *Where do we draw the line before playing God?*

SO WHAT IS THE FUTURE LIKELY TO HOLD?

As of now, the working draft of the human genome may provide merely a framework that is

still too crude for clinical use (11). With the availability of a complete picture, from the genes through to the proteins and their interactions with other molecules, one should be able to decipher the pathophysiology of diabetes mellitus. It would then be only a logical next step, to prevent and cure this recalcitrant disease.

However, it is humbling to remind ourselves of what Sir John Maddox said, "The most important discoveries of the next 50 years are likely to be the ones of which we cannot now even conceive" (12).

REFERENCES

1. Sinha S, Parthaprasad C. The Human Genome Project. Natl Med J India. 2000, 13:169-73

2. Murray RK. Biochemical and genetic bases of disease. *In* Murray RK, Granner DK, Mayes PA, Rodwell VS (*eds*). Harper's Biochemistry. Appleton Lange, Stamford. 2000; PP812-28

3. Kumar Singh K. The human genome: A first look at the blueprint of human inheritance. Curr Sci 2000;79:144-8

4. Hanis C, Boerwinkle E, Chakraborty R et al. A genome wide search for human non-insulin-dependent (type 2) diabetes genes reveals a susceptibility locus on chromosome 2. Nat Genet 1996;13:161

5. Mahtani MM, Widen E, Lehto M et al. Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. Nat Genet 1996;14:90

6. Pratley RE, Thompson DB, Prochazka M et al. An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians. J Clin Invest 1998;101:1757

7. Ji L, Malecki M, Warram JH et al. New susceptibility locus for NIDDM is localized to human chromosome 20q. Diabetes 1997: 46: 876

8. Sridhar GR, Rao YSV. Information technology and endocrine sciences in the 21st century. Indian J Endocrinol Metab 2000; 4: 70-80

9. Ghosh S. Positional cloning strategies to identify causal genes for type 2 diabetes mellitus. *In* LeRoith D, Taylor SI, Olefsky JM (*eds*) Diabetes mellitus. Lippincott, Philadelphia. 2000; pp664-74

10. Sridhar GR, Rao YSV. Clinical database network. Curr Sci 2000;79:549-50

11. Watkins LRCH. Waiting for the working draft from the human genome project. BMJ. 2000;320:1223-4

12. Sridhar GR. The last millennium and the next (Editorial). Indian J Endocrinol Metab 2000;4:1