THE HUMAN GENOME PROJECT

ADEBAMOWO C. A.

Dr. C. A. Adebamowo MBBCh, FWACS, FICS is a Consultant Surgical Oncologist to the University College Hospital Ibadan.

"all the best answers only generate more questions"

The human genome project is one of the most ambitious human endeavors of recent history and whatever its outcome, the world is not likely to be the same again after its completion. The output of the program and others like it in the biological sciences that are unraveling the genetic code of human and other organisms are likely to lead to a new way of understanding biology and modulating it either in the form of prevention of diseases, understanding the processes underlying the development, progress, resolution, and treatment of disease processes. The impact is no less profound in the other areas of biology including environment management, development of new chemical and lately the use of deoxyribonucleic acid (DNA) as the building blocks of the CPU of future computers.

"first, a remembrance of things past"

The set of instructions for the making of an organism are contained in the nucleus of living organisms. They consist of tightly wound coils of DNA molecules that are called the **genome**, these, along with about equal amount of associated proteins are together referred to as **chromosomes**. In the human there are 24 distinct separate chromosomes. The nucleus of the human cell contains 23 pairs of these chromosomes, one set from each parent, made up of 22 autosomes and an X or Y sex chromosome. These chromosomes can be seen under light microscopy and their peculiar staining pattern coupled with difference in sizes allow each of the 24 chromosomes to be distinguishable from one another **(karyotype)**. The genome unwound and tied together will measure 5 feet and be 50 trillionth of an inch in diameter. These coils of DNA contain all the instructions necessary for building and maintaining of life.

In humans, DNA consists of two strands wrapped around each other in the shape of a "double helix". Each nucleotide is made up of a sugar, phosphate molecules, and nitrogen containing molecules called bases. Each strand of DNA is made of repeating units of these compounds. There are four different bases – adenine (A), thymine (T), cytosine (C), and guanine (G). The particular order of arrangement of these bases specifies the exact genetic instruction that is required to create a unique organism with its own traits.

The two strands of DNA are held together by weak bonds between the bases on each strand thus forming base pairs (bp). There are approximately 3 million bps in the human genome. The genome obeys strict base-pairing rules such that A pairs only with T while C pairs with G. Adherence to this rule ensures that during transcription (copying of the DNA), there is minimization of the risk that errors will occur that may somehow affect the organism or its offspring (**mutations**). Genes are the functional units of heredity and they consist of a specific sequence of nucleotide bases that carry the instruction for the construction of structural proteins and enzymes. The human genome consist of approximately 80,000 to 100,000 genes. These genes vary markedly in length and protein coding sequences (exons) account for only 10%. The rest include introns that have no coding function, control sequences and intergenic regions whose function are currently obscure. Within genes, 3 specific sequence of bases (codons) code for each of the 20 amino acids that are found in the body.

Changes in DNA, if gross can be seen microscopically as missing, extras, breaks or rejoinings (translocation) of chromosomes, for example in diseases such as Down's syndrome. However most of the changes in DNA that are responsible for diseases and differences between individuals are too subtle to be visualized microscopically. These require molecular analysis hence the need for detailed genetic maps.

"the journey of a thousand miles starts with the first step"

In 1990, the Department of Energy and the National Institute of Health of the United States of America officially began a $3 billion, 15 years program to find the estimated 80, 000 to 100, 000 human genes and determine the sequence of the 3 billion DNA building blocks that underlie all of human biology and diversity. This initiative, called the "**Human Genome Program**" (HGP) was initially characterized by the development of the biological instrumentation, bio-informatic and computational techniques that are required for efficient production scale DNA sequencing. The effort was later joined by genome centers in the UK, France, Germany, Canada and Japan. Progress on the project has been better than anticipated such that there has been a revision of the time scale and a rough draft of the human genome is expected to be released by the spring of the year 2000 instead of 2001 as previously anticipated.

The method for the achievement of this objective is to make a series of maps of the human genome by:

1. Dividing the chromosomes into small bits that can be propagated and characterized
2. Ordering them so that they correspond to their respective locations on the chromosomes
3. Determine the sequence of the base pairs
4. Develop tools for making use of this information in human biology and medicine

Gene maps are constructed at different resolutions ranging from genetic linkage maps (coarsest) that depict chromosomal location of DNA markers based on their pattern of inheritance to physical maps that describe the chemical characteristics of the DNA molecules themselves. The HGP is committed to producing a complete and accurate reference (error estimate of 1 in 10,000 base pairs) physical map of the human genome that is expected to jump-start this field of human biology in the new millennium. When completed, it is estimated that the data generated will fill 200 volumes of 1000 pages each and it will take an individual reading round the clock 26 years to go through. It is obvious

therefore that this work also requires improvement in the creation and management of large databases, information management and computational technology.

"there is a chase afoot"

For a while attention has been on the mapping of the genome, however it is becoming obvious that the utilization of the information requires a new goal. Therefore as part of a review of the program, the identification of individual sequence variations is now a major focus of research. Although 99% of human DNA sequences are the same across populations, variations in individual DNA sequences may have a major impact not only on how individuals differ from each other but also in how they differ in their response to diseases, disease-causing endogenous and exogenous (environmental) factors such as bacteria, viruses, chemicals, food, drugs and other therapies. The most common type of variations seen is the so-called SNPs (pronounced "snips", meaning single nucleotide polymorphisms), which occur about once in 100 to 300 bases. SNP maps may help to identify the multiple genes that are operative in "complex" diseases such as cancer, diabetes mellitus, vascular diseases and some mental illnesses.

Beyond the identification of the sequences is the efficient interpretation of the function of the genes. This area of research, which is referred to as '**functional genomics**" is a technically challenging endeavor that requires the generation of complete sets of DNA clones and sequences of human and model organism genes. Other goals of functional genomics include studies of gene expression and control, creation of mutations that cause loss or alteration of function in nonhuman organisms and the development of experimental and computational methods for protein analyses - **proteomics**.

The functions of human genes are often elucidated by studying their parallel in non-human organisms in the field of "**comparative genomics**". Hence the HGP is also involved in the study of *E. coli, Saccharomyces cerevisiae, Drosophila melanogaster, Caenorhabditis elegans* and the laboratory mouse. The result of these studies is leading to new breakthroughs in fundamental biology as scientists compare entire genomes in order to gain insights into the evolutionary, biochemical, genetic, metabolic, and physiological pathways.

The field of epigenetics – the study of how the activity of DNA can be altered semi-permanently, not by mutations but by other chemical processes like methylation of the bases – has also been influenced by the HGP. It is suggested that epigenetic traits may be inheritable and that inheritability is not permanent, hence it may be lost, a proposal that is currently the subject of hot debates and exciting research.

The amount of data and the analysis required to make sense of the tremendous information that is being generated requires new techniques in bioinformatics, computational technology, data management, statistical analysis, software and hardware development and engineering. The degree of variability that is obvious in the penetrance and expressivity of relatively simple monogenic diseases foretell the complexity that is to be expected in the application of the new information to the understanding and manipulation of complex diseases.

"all that is necessary for evil to triumph is for good men to do nothing"

The development of any new technology especially one as profound and as fundamental as the unraveling of the genome is bound to be accompanied by questions about the ethical, legal and social implications of the "new science". Right from its inception, the HGP has had as an integral part, programs designed to study these particular issues. Concerns about privacy, disclosure, how the findings may affect concepts of race and ethnicity, the potential uses and misuses of genetic data at work, at school, and the courts are quite real. The mechanism for the commercial exploitation and the possible impact of advances in genetic concepts on our ideas of humanity and personal responsibility are also areas that have excited controversy especially with the competition to the publicly funded HGP by privately funded programs like that of Celera Genomics.

The reality of these concerns argues for a vigorous program of research, education and training in order to ensure that good science does not fall victim to irrational fears.

"chance favors the mind that is prepared "

For the medical profession, the HGP represents both opportunity and challenge. There are whole new fields of research and practice that are developing and new ones will arise in the nearest future. There are computational, biological, physical, chemical, mathematical, engineering and management challenges that will require skill, dedication and foresight to surmount. The revolution has already started and there are new opportunities in molecular medicine, waste control, environmental management, agriculture, animal husbandry, biotechnology, energy resources, risk assessment, law, bioethics, philosophy, etc, so the sky is the limit. The new developments in medicine will happen, the onus is on members of the medical profession all over the world, but especially in the developing countries to be part of the effort since ...

"genes help only those who help themselves"

Table 1: Potential benefits of the Human Genome Project

1. Elucidation of the basis of monogenic diseases.
2. Development of DNA based screening tests for monogenic diseases to identify at risk individuals and intervention to reduce their risk.
3. Disease prediction and prevention on an individual and community basis.
4. Gene therapy based on methods of repairing, modulating, replacing or ameliorating the effect of a defective gene.
5. Pharmacogenomics. The use of a persons genetic makeup to design the best therapy from the point of view of side-effects and therapeutic benefits
6. Unraveling the genetic basis of complex diseases.
7. Understanding gene-gene and gene-environment interactions.
8. Genetic counseling.
9. Ethical, legal and social issues relating to the new science

REFERENCE

http://www.ornl.gov.hgmis/resource/medicine/html

All correspondence to
Dr. Clement A. Adebamowo
Division of Oncology,
Department of Surgery,
University College Hospital,
Ibadan, Oyo State, Nigeria
E-mail: adebamowo@alpha.linkserve.com
Office Phone/Fax: (234) 02 2410995 (DL)