http://www.inosr.net/inosr-experimental-sciences/
Alexander and Aiden
INOSR Experimental Sciences 2(1): 1-9, 2016.

International Network Organization for Scientific Research ISSN: 2705-1692

Overview of Human Genome Project (HGP)

Alexandar Abigail and Aiden Samuel

Department of Health Science California State University, Monterey Bay USA

ABSTRACT

This study is an overview of human genome project (HGP). The Human Genome Project originally aimed to map the nucleotides contained in a human haploid reference genome (more than three billion). The "genome" of any given individual is unique; mapping the "human genome" involved sequencing a small number of individuals and then assembling these together to get a complete sequence for each chromosome. Therefore, the finished human genome is a mosaic, not representing any one individual. A model of the human genome could reveal where and why such breakdowns occur, providing an unprecedented asset for the diagnosis and prevention of their associated diseases. It should also prove enormously valuable in the diagnosis and treatment of genetic damage caused by external factors, such as radiation, toxic chemicals and other environmental pollutants. The Human Genome Project was an international research effort to determine the sequence of the human genome and identify the genes that it contains. The Project was coordinated by the National Institutes of Health and the U.S. Department of Energy. Additional contributors included universities across the United States and international partners in the United Kingdom, France, Germany, Japan, and China. The Human Genome Project formally began in 1990 and was completed in 2003, 2 years ahead of its original schedule.

Keywords: Overview, Human, Genome, Project. HGP

INTRODUCTION

The Human Genome Project (HGP) was an international scientific research project with the goal of determining the base pairs that make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and a functional standpoint [1]. The Human Genome Project is expected to produce a sequence of DNA representing the functional blueprint and evolutionary history of the human species. However, only about 3% of this sequence is thought to specify the portions of our 50,000 to 100,000 genes that encode proteins. Thus an important part of basic and applied genomics is to identify and localize these genes in a process known as transcript mapping. When genes are expressed, their sequences are first converted into messenger RNA transcripts, which can be isolated in the form of complementary DNAs (cDNAs). Approximately half of all human genes had been sampled as of 15 June, 1996

The Human Genome Project is a national effort to decipher the human genetic blueprint that is being spearheaded by the U.S. Department of Energy (DOE) and the National Institutes of Health (NIH). It will require "mapping" the location of some 100,000 genes along the 23 pairs of human chromosomes, then "sequencing," or determining the order of, the three billion base pairs of nucleotides that make up these chromosomes.

It remains the world's largest collaborative biological project [2]. After the idea was picked up in 1984 by the US government when the planning started, the project formally launched in 1990 and was declared complete on April 14, 2003. [3] Funding came from the US government through the National Institutes of Health (NIH) as well as numerous other groups from around the world. A parallel project

was conducted outside the government by the Celera Corporation, or Celera Genomics, which was formally launched in 1998. Most of the government-sponsored sequencing was performed in twenty universities and research centers in the United States, the United Kingdom, Japan, France, Germany and China [4].

A small portion of each cDNA sequence is all that is needed to develop unique gene markers, known as sequence tagged sites or STSs, which can be detected in chromosomal DNA by assays based on the polymerase chain reaction (PCR). To construct a transcript map, cDNA sequences from a master catalog of human genes were distributed to mapping laboratories in North America, Europe, and Japan. These cDNAs were converted to STSs and their physical locations on chromosomes determined on one of two radiation hybrid (RH) panels or a yeast artificial chromosome (YAC) library containing human genomic DNA. This mapping data was integrated relative to the human genetic map and then cross-referenced to cytogenetic band maps of the chromosomes.

The Human Genome Project originally aimed to map the nucleotides contained in a human haploid reference genome (more than three billion). The "genome" of any given individual is unique; mapping the "human genome" involved sequencing a small number of individuals and then assembling these together to get a complete sequence for each chromosome. Therefore, the finished human genome is a mosaic, not representing any one individual.

The histograms reflect the distributions and densities of genes along the chromosomes. Because the individual genes are too numerous to represent, images have been chosen to illustrate the myriad aspects of human biology, pathology, and relationships with other organisms that can be revealed by analysis of genes and their protein products

REVIEW OF HUMAN GENOME

Are you male or female? What color is your skin? Your eyes? Your hair? Are you tall? Short? Fat? Thin? Somewhere in between? In appearances, temperment, talents, and skills, do you favor your mother or your father, a grandmother or grandfather, or are you an original, a one-of-a-kind? In the nucleus of each of the hundred trillion cells in your body is a very special heirloom, a "recipe book" of sorts, containing about a hundred thousand individual recipes, organized into 24 chapters that, taken as a whole, provide a complete set of instructions for making the being that is you. This recipe book, made from your full complement of genetic material, is called a genome, and it is written in an arcane language that is universal for all life on this planet. Were we able to read this genetic language well enough to decipher the human genome, we would have at our disposal what would almost certainly be the most powerful resource for biological and medical research ever developed. Although the genome of each of us is uniquely ours, the variation from person-to-person averages but two-tenths of one-percent (less for "identical" twins). This means that a model of one human genome could serve as a standard reference for us all. More than 4,000 diseases, including cancer and heart disease, have been identified as occurring because of a breakdown in the genetic process [5]. A model of the human genome could reveal where and why such breakdowns occur, providing an unprecedented asset for the diagnosis and prevention of their associated diseases. It should also prove enormously valuable in the diagnosis and treatment of genetic damage caused by external factors, such as radiation, toxic chemicals and other environmental pollutants.

Just as archaeologists needed the Rosetta stone to decipher Egyptian hieroglyphics, biologists will need a "key" to decipher the human genome. This "key" is expected to be provided by the mapping and sequencing of the genome, which is the main objective of the Human Genome Project. To understand "mapping" and sequencing" it helps to review the process

of how the information in the genome is translated into a human being, a living, four- dimensional space-time entity [6].

Your body is a community of different types of tissue, bone tissue, skin tissue, muscle tissue, nerve tissue, etc., which in turn are communities of specific types of cells. What determines the structure of a given cell, and also regulates much of the chemical activity that drives and defends the body, is protein. The ingredients for making protein are chemical compounds called amino acids, of which there are 20 varieties. Amino acids from one or more of these different varieties link together to form a "polypeptide chain". A typical protein is a polypeptide chain of some 300 amino acids. The varieties of amino acids linked and the order in which the individual acids are joined determines the nature of the protein made [7] [8]. This assemblage of amino acids into polypeptide chains, crucial to determining whether the final product is a fish, a flower, a microorganism, or a human being, is directed by the instructions contained in the genetic code.

Mapping the Human Genome

Among the main goals of the Human Genome Project (HGP) was to develop new, better and cheaper tools to identify new genes and to understand their function. One of these tools is genetic mapping. Genetic mapping - also called linkage mapping - can offer firm evidence that a disease transmitted from parent to child is linked to one or more genes. Mapping also provides clues about which chromosome contains the gene and precisely where the gene lies on that chromosome.

Genetic maps have been used successfully to find the gene responsible for relatively rare, single-gene inherited disorders such as cystic fibrosis and Duchenne muscular dystrophy. Genetic maps are also useful in guiding scientists to the many genes that are believed to play a role in the development of more common disorders such as asthma, heart disease, diabetes, cancer, and psychiatric conditions.

Human genes are found in the rungs of a DNA double helix. DNA makes up the 23 pairs of chromosomes in the human body [9]. If you think of the human body as big, complicated, encrypted code, then the scient-ists mapping the human genome are attempting to break that code. Once the code is broken, it will reveal many secrets of how the human body works, and it could lead to greater disease prevention. In June 2000, scientists from the Human Genome Project and from Celera Genomics both announced that they had assembled a working draft sequence of the human genome, a major step in cracking the code.

What researchers are trying to do is construct a detailed genetic map of the human genome and determine the entire nucleotide sequence of human deoxyribonucleic acid (DNA). A nucleotide is the basic unit of nucleic acid, which is found in the 23 pairs of chromosomes in the human body. According to the Human Genome Project, there are between 26,000 and 40,000 genes in the human body. Each of these genes is composed of a unique sequence of pairs, each with four bases, called base pairs.

In a DNA molecule, which is shaped like a twisted ladder, the bases are the chemicals that interlock to form the rungs of the ladder. The sides of the ladder are made of sugar and phosphate molecules [10]. The human body has about 3 billion base pairs, but only about 4 percent of those pairs constitute DNA that affects gene function. We don't have any idea about the purpose of the other 96 percent of base pairs, consequently termed junk-DNA.

The finished sequence was completed, with 99.99-percent accuracy, in 2003. Discovering the secrets of the sequence is still in its infancy.

Better understanding the human genome will tell us a lot about how life works. It could lead to preventing or curing diseases, because genetics is what getting sick is all about -- our genes are trying to fight off the genes of a virus or bacteria. The next step will be to determine how this battle is played out [11]. Today,

researchers know the positions of some genes that control our medical traits. Other genes have been located but their functions are unknown, and still others remain entirely elusive. The point of genome research is to locate the genes and determine just how the four bases are sequenced, and then to learn what the genes actually do.

Importance of Human Genome Project

The Human Genome Project was an international research effort to determine the sequence of the human genome and identify the genes that it contains. The Project was coordinated by the National Institutes of Health and the U.S. Department of Energy. Additional contributors included universities across the United States and international partners in the United Kingdom, France, Germany, Japan, and China. The Human Genome Project formally began in 1990 and was completed in 2003, 2 years ahead of its original schedule.

The work of the Human Genome Project has allowed researchers to begin to understand the blueprint for building a person [12]. As researchers learn more about the functions of genes and proteins, this knowledge will have a major impact in the fields of medicine, biotechnology, and the life sciences.

It is important because it uses information from DNA to develop new ways to treat, cure, or even prevent the thousands of diseases that afflict humankind. But the road from gene identification to effective treatments is long and fraught with challenges. The Department of Energy's (DOE) Human Genome Program, directed by Ari Patinus and the National Institute of Health's National Human Genome Research Institute (NHGRI) directed by Francis Collins, together make up the United States Human Genome Project. At least eighteen countries have established genome research programs. Some of the larger programs are in Australia, Canada, Italy and Mexico [13]. The Human Genome Organization (HUGO) helps to organize international collaboration in the genome project. In June 2000, international leaders of the HGP confirmed that the rough drafts of the human genome had been completed a year ahead of schedule. The draft sequence will provide a gibbet of sequence across ninety percent of the human genome. Efforts are still underway to complete the high quality DNA reference sequence by 2003.

The goal of the Human Genome Project is to identify all the approximate 30,000 genes in human DNA. It determines the sequence of the three billion chemical base pairs that make up human DNA and store this information in data bases. It improve tools for data analysis, transfer related technologies to the private sector and addresses the ethical, legal and social issues that may arise form the project [14].

Technology and resources created by the HGP already have a major input on research across the life sciences. There are some potential benefits of HGP research to they are molecular medicine, microbial genome, risk assessment, evolution, DNA forensic and agriculture. The HGP is "starting to have profound impact on the biomedical research and promise to revolutionize the wider spectrum of biological research and medical medicine"

With the HGP molecular medicine can improve diagnosis of disease, detect genetic predispositions to disease earlier, ration drug design, control gene therapy and control systems of drugs, and create pharamacogenomics customs drugs. The potential for using genes themselves to treat disease, known as gene therapy, has captured the imaginations of the public and the biomedical community for good reason. This rapidly developing field holds great potential for treating or even curing genetic and acquired diseases [15].

In 1994 DOE initiated the Microbial Genome project to sequence the genomes of bacteria that is useful in energy production, toxic waste reduction, environmental remediation, and industrial processing. Microbial genomic will also help pharmaceutical researchers gain a better understanding of how pathogenic microbes cause disease.

Sequencing their microbes will help reveal vulnerabilities and identify new drug tests." "Gaining a deeper understanding of the microbial world also will provide insights into the strategies and limits of life on this planet". Within the next decade, researchers will find most human genes. A major challenge for the 21st century is to show how faulty genes play a role in disease causation. Drug design will be revolutionized as researchers create new classes of medicines based on a reasoned approach using gene sequence and protein structure function information. The drugs, targeted to specific sites in the body, promise to have fewer side effects than many of today's medicines [16].

Genomic study will have a huge impact on the ability to assess risks posed to individuals by exposure to toxic agents. This knowledge will address DOE's mission to understand the effects of low-level exposures to radiation and other energy-related causes especially in terms of risk of cancer.

Genetics will help us understand human evolution and the common life that we all share in biology. In the Human Genome Project human evolution research is study evolution through germline mutations in lineages, study migration of different population groups based on female genetic inheritance, study mutations of the Y chromosome to trace lineage and migration of males, and compare breakpoints in the evolution of mutations with ages of populations and historical events.

Any type of organism can be identified by examination of DNA sequences to identify individuals, forensic scientist perform DNA fingerprints which is when scientist scan about ten DNA regions that vary from person to person and use the data to create a DNA profile of the individual. Plant and animal genomes allow us to create more disease resistant plants. This reduces the costs of agriculture and providing consumers with more nutritious, pesticides-free foods. Farmers have been able to increase outputs and reduce waste because their corps and herds are healthier [17].

Some of the legal, ethical and social issues concerns arising from the new genetics are the fairness in the use of genetic information by usurers; privacy and confidentiality of genetic information; psychological impact and stigmatization due to individuals' genetic differences. Most people think that science is remote from the work they do, the lives they lead, and the decisions that they make day by day. The progress of science can potentially invade your life in the most direct ways, affecting the choices you make at the grocery store, your own health care and that of your family, and even your reproductive decisions.

The Human Genome Project (HGP) has created the field of genomic understanding genetic material on a large scale [18]. The medical industry is building upon the knowledge, resources, and technologies emanating from the HGP to further understanding of genetic contributions to human health. As a result of this expansion of genomic into human health treatment, the field of genomic medicine has been born. Genetics is playing an increasingly important role in the diagnosis, monitoring, and treatment of diseases.

All diseases have a genetic component, whether inherited or resulting from the body's response to environmental stresses like viruses or toxins. The successes of the Human Genome Project (HGP) have even enabled researchers to pinpoint errors in genes, the smallest units of heredity that cause or contribute to disease.

An increasing number of gene tests are becoming available commercially although the scientific community continues to debate the best way to deliver them to the public and medical communities that are often unaware of their scientific and social implications. While some of these tests have greatly improved and even saved lives, scientists remain unsure of how to interpret many of them. Also, patients taking the tests

face significant risks of jeopardizing their employment or insurance status. And because genetic information is shared, these risks can extend beyond them to their family members as well.

The connection between science and health is a direct one, and your ability to understand the science behind health affects your ability to understand the issues and the stakes. Science may seem difficult, because scientists often use technical language to talk about abstract ideas. Most people are curious about the way their bodies work and this curiosity goes beyond immediate concerns about any specific health condition.

Genetic Markers

Markers themselves usually consist of DNA that does not contain a gene. But because markers can help a researcher locate a disease-causing gene, they are extremely valuable for tracking inheritance of traits through generations of a family.

The development of easy-to-use genetic maps, coupled with the HGP's successful sequencing of the entire human genome, has greatly advanced genetics research. The improved quality of genetic data has reduced the time required to identify a gene from a period of years to, in many cases, a matter of months or even weeks.

Genetic mapping data generated by the HGP's laboratories is freely accessible to scientists through databases maintained by the National Institutes of Health and the National Library of Medicine's National Center for Biotechnology Information (NCBI) [ncbi.nlm.nih.gov], as well as the Genome Browser of University of California, Santa Cruz.

Genome donors

In the IHGSC international public-sector HGP, researchers collected blood (female) or sperm (male) samples from a large number of donors. Only a few of many collected samples were processed as DNA resources. Thus the donor identities were protected so neither donors nor scientists could know whose DNA was sequenced. DNA clones taken from many different libraries were used in the overall project, with most of those libraries being created by Pieter J. de Jong's. Much of the sequence of the reference genome produced by the public HGP came from a single anonymous male donor from Buffalo, New York

HGP scientists used white blood cells from the blood of two male and two female donors (randomly selected from 20 of each) – each donor yielding a separate DNA library. One of these libraries was used considerably more than others, due to quality considerations. One minor technical issue is that male samples contain just over half as much DNA from the sex chromosomes (one X chromosome and one Y chromosome) compared to female samples (which contain two X chromosomes). The other 22 chromosomes (the autosomes) are the same for both sexes.

Although the main sequencing phase of the HGP has been completed, studies of DNA variation continued in the International HapMap Project, whose goal was to identify patterns of single-nucleotide polymorphism (SNP) groups (called haplotypes, or "haps"). The DNA samples for the HapMap came from a total of 270 individuals; Yoruba people in Ibadan, Nigeria; Japanese people in Tokyo; Han Chinese in Beijing; and the French Centre d'Etude du Polymorphisme Humain (CEPH) resource, which consisted of residents of the United States having ancestry from Western and Northern Europe.

Genetic Prescreening

When doctors first performed in vitro fertilization (IVF) in 1978, it gave many otherwise infertile couples a way to have a child of their own. IVF works by removing the eggs from the woman's uterus, fertilizing them in a laboratory and then, a few days later, transferring the fertilized egg, called a zygote, back into the uterus. IVF has also led to a procedure that allows parents to weed out genetically defective embryos. This procedure is called preimplantation genetic diagnosis (PGD).

PGD is often used during IVF to test an embryo for genetic disorders before

inserting it into the woman's uterus. Once the egg is fertilized, a cell from each embryo is taken and examined under a microscope for signs of genetic disorders. Many couples use this procedure if there are any inherited disorders in their genes to decrease the possibility that the disorder will passed to their child. Currently, PGD can be used to detect many disorders, including cystic fibrosis, Down syndrome, Tay-Sachs disease and hemophilia A.

Some genetic disorders are specific to one gender or another, such as hemophilia, which usually affects boys. Doctors may examine the cells to determine the gender of the embryo. In a case where a family has a history of hemophilia, only female embryos are selected for placement in the uterus. This practice is at the center of a larger debate about whether parents should be able to choose embryos purely on the basis of gender. Some people worry that it could lead to an imbalance between genders in the general population, especially in societies that favor boys over girls, such as China.

While PGD enables us to pick out embryos that don't have genetic disorders, and even chose the gender we want, it is only the beginning of what genetic engineering can do. Parents could someday custom-order babies with certain traits.

Analytical review of Human Genome

The process of identifying the boundaries between genes and other features in a raw DNA sequence is called genome annotation and is in the domain of bioinformatics. While expert biologists make the best annotators, their work proceeds slowly, and computer programs are increasingly used to meet the high-throughput demands of genome sequencing projects. Beginning in 2008, a new technology is known as RNA-seq was introduced that allowed scientists to directly sequence the messenger RNA in cells. This replaced previous methods of annotation, which relied on the inherent properties of the DNA sequence, with direct measurement, which was much more accurate. Today, annotation of the human genome and other genomes relies primarily on deep sequencing of the transcripts in every human tissue using RNA-seq. These experiments have revealed that over 90% of genes contain at least one and usually several alternative splice variants, in which the exons are combined in different ways to produce 2 or more gene products from the same locus.

The genome published by the HGP does not represent the sequence of every individual's genome. It is the combined mosaic of a small number of anonymous donors, all of the European origin. The HGP genome is a scaffold for future work in identifying differences among individuals. Subsequent projects sequenced the genomes of multiple distinct ethnic groups, though as of today there is still only one "reference genome.

FINDINGS

Key findings of the complete genome sequences include:

1. There are approximately 22,300 protein-coding genes in human beings, the same range as in other mammals.
2. The human genome has significantly more segmental duplications (nearly identical, repeated sections of DNA) than had been previously suspected.
3. At the time when the draft sequence was published fewer than 7% of protein families appeared to be vertebrate specific.

CONCLUSION

In conclusion Human genome, all of the approximately three billion base pairs of deoxyribonucleic acid (DNA) that make up the entire set of chromosomes of the human organism. The human genome includes the coding regions of DNA, which encode all the genes (between 20,000 and 25,000) of the human organism, as well as the noncoding regions of DNA, which do not encode any genes. By 2003 the DNA sequence of the entire human genome was known. The

human genome, like the genomes of all other living animals, is a collection of long polymers of DNA. These polymers are maintained in duplicate copy in the form of chromosomes in every human cell and encode in their sequence of constituent bases (guanine [G], adenine [A], thymine [T], and cytosine [C]) the details of the molecular and physical characteristics that form the corresponding organism. The sequence of these polymers, their organization and structure, and the chemical modifications they contain not only provide the machinery needed to express the information held within the genome but also provide the genome with the capability to replicate, repair, package, and otherwise maintain itself. In addition, the genome is essential for the survival of the human organism; without it no cell or tissue could live beyond a short period of time. For example, red blood cells (erythrocytes), which live for only about 120 days, and skin cells, which on average live for only about 17 days, must be renewed to maintain the viability of the human body, and it is within the genome that the fundamental information for the renewal of these cells, and many other types of cells, is found.

REFERENCES

1. Bryant, J. A (2007). Design and information in biology: From molecules to systems. p. 108. ISBN 9781853128530.
2. Wellcome Sanger Institute. "The Human Genome Project: a new reality". Wellcome Trust Sanger Institute, Genome Research Limited. Archived from the original on 2013-08-01.
3. Naidoo N; Pawitan Y; Soong R; Cooper DN; Ku CS (2011). "Human genetics and genomics a decade after the release of the draft sequence of the human genome". Hum Genomics. 5 (6): 577–622.
4. "Celera: A Unique Approach to Genome Sequencing". ocf.berkeley.edu. Biocomputing. 2006.
5. Davidson College (2002). "Sequencing Whole Genomes: Hierarchical Shotgun Sequencing v. Shotgun Sequencing". bio.davidson.edu. Department of Biology, Davidson College.
6. Human Genome Project Information Archive (2013). "U.S. & International HGP Research Sites". U.S. Department of Energy & Human Genome Project.
7. Vizzini, Casimiro (March 19, 2015). "The Human Variome Project: Global Coordination in Data Sharing". Science & Diplomacy. 4 (1).
8. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A. (2001-02-16). "The sequence of the human genome". Science. 291 (5507): 1304–1351. ISSN 0036-8075.
9. Roach JC; Boysen C; Wang K; Hood L (1995). "Pairwise end sequencing: a unified approach to genomic mapping and sequencing". Genomics. 26 (2): 345–353.
10. Center for Biomolecular Science & Engineering. "The Human Genome Project Race". Center for Biomolecular Science and Engineering.
11. International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome" (PDF). Nature. 409 (6822): 860–921.
12. Venter, JC; et al. (2001). "The sequence of the human genome" (PDF). Science. 291 (5507): 1304–1351.
13. Osoegawa, Kazutoyo; Mammoser, AG; Wu, C; Frengen, E; Zeng, C; Catanese, JJ; De Jong, PJ (2001). "A Bacterial Artificial Chromosome Library for Sequencing the Complete Human Genome". Genome Research. 11 (3): 483–96.

14. Tuzun, E; et al. (2005). "Fine-scale structural variation of the human genome". Nature Genetics. 37 (7): 727–737.
15. Kennedy D (2002). "Not wicked, perhaps, but tacky". Science. 297 (5585): 1237.
16. Venter D (2003). "A Part of the Human Genome Sequence". Science. 299 (5610): 1183–4.
17. Wadman, Meredith (2008-04-16). "James Watson's genome sequenced at high speed". Nature News. 452 (7189): 788.
18. Gonzaga-Jauregui C; Lupski JR; Gibbs RA (2012). "Human genome sequencing in health and disease". Annu Rev Med. 63 (1): 35–61.