

What is the Human Genome Project?

Nowadays, the science of genetics is indispensable. The use of highly evolved technology to read and analyze DNA has revolutionized our world. Much of the population, however, lacks an understanding of one of the world's most crucial discoveries and sciences. This project will discuss the Human Genome Project, as well as various important factors in genetics, such as DNA synthesis, and sequencing. It will also cover diseases caused by the deformation of DNA, as well as discuss important individuals and Canadian connections in the world of genetics. The goal of this project is to expand and amplify the reader's understanding of the world of genetics, and the key factors and individuals involved in its development.

The Human Genome Project (HGP) began in 1990 (Genomics, 1999). It has been developing and expanding for 16 years, and has cost millions of dollars (National Geographic News, 2006). This project has mapped the basics of the human genome, which proved to be an extremely difficult goal (Suzuki S. & Knudtson P, 1988; National Geographic News, 2006). This consists of reading human DNA in order to gain some insight as to the defects in genes leading to diseases, since all living organisms are defined by their DNA (Genomics, 1999; National Geographic News, 2006). These chemical molecules determine our 'blueprint for biological development', determining our hair colour, blood type, eye colour, etc (National Geographic News, 2006). Scientists worldwide have stored over 500 million sequences of DNA bases (Genomics, 1999).

Geneticists at the National Human Genome Research Institute at the U.S. National Institutes of Health are trying to cut the cost of DNA mapping. People could therefore have their DNA tested, as well as learn about their origins in an efficient and

economic way (National Geographic News, 2006). By knowing a patient's genetic background, doctors could diagnose illnesses years before the symptoms appear in an individual. The Human Genome Project has advanced with the growth of the internet and computing. This endeavour is currently digitally mapping the final gene sequences of DNA, allowing the genome to be read and replicated much faster and more efficiently (National Geographic News, 2006). This publication states that one fifth of all human genes, numbering roughly 4000 thus far, have been patented. This is done by private firms and universities. Gene patents allow for the company to have the rights to those genes: if the gene is found to help in curing diseases, they get the credit and profit (National Geographic News, 2006).

DNA Synthesis

Contrary to popular belief, it is possible to create DNA (Genome Canada, 2006; M. Belouchi, personal communication, March 29, 2007). These extremely complex molecules are composed of different chemical compounds called nucleotides. They are: Guanine, Adenine, Thymine, and Cytosine. These component molecules fit together. Adenine links with Thymine, and Cytosine links with Guanine. A strand of DNA, made up with a given number of nucleotides, follows a certain pattern. All DNA strands, however, link with their exact opposites, forming one DNA molecule, the famed double helix. On occasion, however, DNA molecules may deform. This phenomenon may lead to a mismatch in the pairing of DNA links. For example; if Guanine linked with Adenine, the strand of deoxyribonucleic acid would be defective. This mutation is called SNP: Single Nucleotide Polymorphism.

One mismatch in the bonding of DNA bases could lead to cystic fibrosis, or even cancer. It has not yet been discovered, however, how to alter nucleotide sequences, making a healthier being. There are 70 to 100 trillion cells in the human body. Each cell is composed of cytoplasm, and a nucleus. The 46 chromosomes are located in the nucleus. Although each cell is identical in its housing of 30000 genes on the chromosomes, cells in each different tissue function in a unique way. There are 280 different tissues and organs in the body. These differences are due to proteins which turn on a series of genes in each cell. If the genetic material in a nucleus were to be stretched out, it would reach three meters in length. Dr. Belouchi, of Genizon Biosciences states that only 5 percent of each chromosome is genetic material, while the rest is structural backbone (M. Belouchi, personal communication, March 29, 2007). Since our body registers on average 10000 mutations a day, proteins replace and repair faulty DNA on a consistent basis. Scientists have been hypothesizing about artificially controlling proteins to replace and repair hazardous SNPs.

DNA Sequencing

The Human Genome Project involves mapping human DNA, our biological coding (National Geographic News, 2006). To do this, geneticists separate DNA molecules, duplicate them, and read the molecules. The reading is rigorous work, and can be done many different ways. First, scientists heat the DNA double helix at about 90* Celsius, in water, making two single strands. Dr. Belouchi states that each nucleotide is composed of one phosphate molecule, one ribose molecule, and a base molecule: either Adenine (A), Cytosine (C) Thymine (T), and Guanine (G) (M. Belouchi, personal communication, March 29, 2007). Next, they must get rid of the terminal phosphate at

one end of the DNA strand. They do this by using an enzyme, called phosphatase, to cleave the phosphate.

They must then place the DNA in a test tube, add a radioactive phosphate, called P32, and add an enzyme kinase. Kinase, works like a nanomachine, attaching P32 to the end of the DNA. Next, copies of the DNA are made, and placed into 4 different tubes, each containing different chemicals. Each type of chemical will cut the DNA in different areas. Chemical 1 will sever the DNA after some of the A nucleotides. Chemical B will cut after some of the A or G nucleotides. Chemical 3 will cut after some of the T or C nucleotides, and the final chemical will sever the DNA following some of the T nucleotides. This chemical digestion only partially cuts the DNA, skipping some bases, leaving some DNA intact. This allows for DNA with a base sequence of AACGT to be cut at various nucleotides of the same family.

DNA is eventually cut at all possible places, giving many different sized strands. Sometimes, chemical 1 skips an A, cleaving after the next one, intentionally. It is at this stage that the biochemists may start sequencing. The strands of DNA are placed in electrophoresis gel. This gel has a Jell-O-like texture. A negative charge is put at one end of the well of the machine, while a positive charge is put into the other end of the well. The strands slowly move from the negative to positive charge. As they enter the gel, the pieces move at different speeds, depending on their size (the smaller, the faster). For smaller DNA strands, or for better resolution, one may use a polyacrylamide gel.

When the strands are at different positions in the gel, the charges are removed, halting the DNA's movement. The gel is then removed, and placed on clear plastic. It is then wrapped in saran wrap, and covered by an unused X-ray film. The ensemble is

placed in an X-ray reading cassette, developed and then read in a dark room. The radioactive elements can be seen as lines on the X-ray, in each of the four columns. By deduction, one can identify in what order the bases appear. If two lines appear four, and eight spaces up in the first well, this means that the original strand tested had A bases at positions four and eight in the molecule. By going through this slow process, one may determine the sequence of nucleotides in a gene.

There are, however, 30000 genes in the human genome. After 1977, many fast and efficient techniques were used to extract and read DNA (Genomics, 1999). These became automated, and began to show genes causing certain diseases in humans and other organisms. In 1995, the first complete DNA base sequence of an entire organism was produced by Crig Venter from the Institute for Genomic Research. Many complete genomes of micro organisms have been sequenced (Suzuki S. & Knudtson P, 1988). Dr. Belouchi states that newer technology results in each DNA base being coloured in a specific tone, allowing for rapid and massive gene sequencing (M. Belouchi, personal communication, March 29, 2007). Instead of taking 9 months to sequence a gene, a DNA analyzing machine, costing 500,000 dollars, can sequence an entire gene in merely 2 and a half hours.

This machine uses nanotechnology and capillary electrophoresis to determine the base sequences. A light source creates a different reaction in each of the four nucleotides, allowing each one to be identified as it reaches the end of the thin electrophoretic wire. This wire replaces the agarose gel used in traditional techniques. Scientists can compare an individual's DNA to the healthy human genome, detecting any mutations. SNPs are not always dangerous, and can be caused by the inheritance of traits from both parents.

Three nucleotides are called a codon. Each codon represents one amino acid. In cystic fibrosis, one mutated amino acid can result in a whole host of complications and increase mortality rate. Some simple diseases are caused by one genetic mutation, while others, such as asthma and cancer, are caused by multiple mutations. This process allows scientists to recognize diseases at early or non-existent stages. The information retrieved is then stored in a public library. A company called 454 life came out with new reading techniques (National Geographic News, 2006). Now, DNA strands pass through a very small machine called a nanopore. This machine uses nanotechnology to read the bases of the DNA strands and could allow geneticists to read an entire human genome in just a few hours. This method, however, is still being perfected (H. Sleiman, personal communication, March 29, 2007).

The application of the Genome Project to the curing of disease

The Human Genome project proposes to map one's genetic coding. An individual's mapped genome is compared to the already read and mapped parts of the Human Genome (Genome Canada, 2006; M. Belouchi, personal communication, March 29, 2007). Thus, doctors could theoretically treat patients before symptoms of diseases appear (National Geographic News, 2006). This is done either by identifying any deficiencies in the individual's genome or by examining the family history of the individual (National Geographic News, 2006). Any problems in the family history could potentially lead to diseases such as cancer, since the offspring inherits genes from each parent. In fact, geneticists are able to now find genetic mutations in DNA that lead to certain diseases. Thus, scientists can discover any mismatches (DNA molecules that

don't bond together perfectly) in the specific individual's DNA (Genome Canada, 2006; M. Belouchi, personal communication, March 29, 2007).

It is, however, at this point in time, impossible for geneticists to cure these mismatches, or mutations in DNA. Ideally, they could send in artificially created strands of DNA to pair up with other strands, getting rid of mutated molecules. However, scientists have not been able to cure mismatches, for when the DNA is sent into the body, white blood cells and anti-bodies attack and destroy the DNA molecule. On one occasion, these molecules were sent in the body, disguised as viruses. They successfully reached the mutation. However, the virus still contained some of its dangerous contents, killing the infected cells (H. Sleiman, personal communication, March 29, 2007). Geneticists know what the problem is, but cannot cure mutations in the human genome.

The Search for a Cure

Geneticists worldwide are battling head to head in a competition (H. Sleiman, personal communication, March 29, 2007). They are trying to discover how to analyze the human genome very quickly and efficiently. So far, geneticists are trying to see how few nucleotides (component molecules of DNA) can be read to determine the pattern of the genome. Since single DNA strands can be made up of over 600 nucleotides, mapping the genome takes a long time, and lots of money. Scientists have discovered that by analyzing 18 base pairs, they can detect any similarities in that one strand of DNA. With this number, reading is done efficiently. They must therefore repeat this for all strands of DNA collected.

Huntington's disease

The Human Genome Project has the potential to cure many diseases. It has already greatly helped reverse diabetes in mice (Collins, F. & Barker, A.D, 2007; National Geographic News, 2006). As Collins and Barker (2007) stated, the internal body structures of mice are much like ours, thus this is the first step to curing diabetes. The Human Genome Project has made monumental advances in studying ovarian cancer as well (Collins, F. & Barker, A.D, 2007). Geneticists are able to identify diseases only by comparison. By reading and analyzing the DNA of patients with the same disease, they will likely find an SNP (gene mutation) in common. On average, there is one genetic mutation per person in every 1000 genes analyzed. There are, however, in one's body, 10000 genetic mutations a day. As professor Majid Belouchi (2007), from Genizon Canada stated: proteins, microscopic "nanomachines" repair these SNPs by removing and replacing faulty DNA bases with perfectly accurate ones.

In some cases, however, cells have to commit suicide, rather than contaminate neighboring cells (M. Belouchi, personal communication, March 29, 2007; Harmon, 2007). Genetic mutations during one's lifetime, however, are not the only cause for disease. Genetically inherited diseases, such as Huntington's disease, are passed on from parent to child. Huntington's disease, a horrible condition passed on genetically, is a progressive degenerative neurological disease. In this new age of technology, one can now choose to know whether they possess the disease or not. This is done by performing a series of tests in which geneticists retrieve DNA, comparing it with other affected patients.

Scientists can simply examine the genetic mutations to see if the tested patients have genetic problems in common. Once the SNPs is found in common, the patient

knows they have Huntington's disease. In this case, genetic mutations cannot be reversed. The disease is passed on with the inheritance of genes. In this disease, the sickness' gene is dominant. If each parent possesses a dominant and recessive gene, the child has a fifty percent chance of receiving the dominant gene and thus developing the disease. The dominant gene over-rules the recessive one. If either parent possesses two dominant genes, the child will be diseased. In most families, however, neither parents possess the diseased gene (M. Belouchi, personal communication, March 29, 2007; New York Times, 2007). These basic principles are known as Gregor Mendel's, the father of modern genetics, Laws of Inheritance (Chromosome theory of heredity, 1999).

Important Individuals

Monk Gregor Mendel is considered to be the father of genetics. While working as a monk in the mid 1800s, he performed experiments on over 29,000 pea plants, and put forth the concept of genetic transmission of traits (Wikipedia, 2007). These ideas are known as "Mendel's Laws of Inheritance." Gregor Mendel's ideas were not accepted or believed until the 20th century. In around 1900, Hugo de Vries and Carl Correns realized the importance of his discoveries. Mendel discovered the roles of dominant and recessive genes, and their importance in the inheritance of characteristics. In 1865, Mendel said that genes were inherited through generations. These genes were particles that maintained their identity across generations (Chromosome theory of heredity, 1999). Mendel stated that each child receives traits from both parents (Chromosome theory of heredity, 1999). He believed that genes exist as pairs in body cells.

In experiments, he represented dominant alleles as uppercase letters, and recessive ones as lower case. If one child inherits genes A and B from one parent, an a and b from another, the possible genetic combinations are: Ab, AB, aB, ab (Chromosome theory of heredity, 1999). Mendel's laws can be classified into three subgroups (Snedden, R. 1995). He stated that all organisms had particles of inheritance, known today as genes. He proposed that a certain characteristic opposed the other. Mendel strongly believed that genes must exist for each characteristic, and that when fertilization takes place, the sperm and egg combine, thus giving one particle, or gene, to the child, per characteristic.

The next great advance in genetics was the discovery of the structure of the DNA molecule by Francis Crick and James D. Watson in 1953 (Wikipedia, 2007). They were jointly awarded the 1962 Nobel Prize for Physiology or Medicine "for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material." (Wikipedia, 2007) Their discovery became the foundation for the Human Genome Project. This allowed scientists to think about the possibility of testing the components of individual DNA. Watson was the head of the Cold Spring Harbor Laboratory, and was named the head of the Human Genome Project.

Francis Collins was another major contributor to the world of modern genetics. He was born on April 14, 1950, in Virginia, and home-schooled till the sixth grade (Wikipedia, 2007). Early on, he had no interest for biology. Educated at Yale University, Collins received a Ph.D. in physical chemistry in 1974. After earning his M.D. at the University of North Carolina at Chapel Hill in 1977, he served as a resident of internal medicine at the North Carolina Memorial Hospital, from 1978 to 1981. As Wikipedia (2007) states, after joining the University of Michigan in 1984, Collins continued

developing methods of reading large stretches of DNA to discover disease genes. His gene-hunting approach to genetics led him to the position of Professor of Internal Medicine and Human Genetics at the University of Michigan, in 1981. This method, called “positional cloning” has become an integral part of modern day genetics. This approach allows scientists to discover mutated genes, without prior knowledge of the abnormalities of the gene.

In 1989, Collins applied this method to hopefully find a cure for cystic fibrosis. Some other successes include the discovery of the genes coding for: Huntington's disease, neurofibromatosis, multiple endocrine neoplasia type 1, as well as adult Leukemia. In 1993, Collins accepted the position of director of the National Center for Human Genome Research (NHGRI), succeeding James Watson. As director, Collins supervises mass genetic sequencing. With new technology, Francis Collins hopes to discover genes leading to cancer, heart disease, and mental illness. Collins’ efforts have allowed the Human Genome Project (HGP) to work under budget, and quickly.

In June 2000, Collins and Bill Clinton announced that all 3 billion base pairs had been mapped, completing the human genome. Collins’ most ambitious plan was to create the Haplotype map. This is a huge database consisting of all the SNPs, helping in the discovery of the treatment for the diseases. This attempt was successful. Furthermore, Collins is known for incorporating religion into his works. He recently published “The Language of God”, a smash hit (Wikipedia, 2007). Collins’ progress has been recognized by numerous awards, such as his election into the Institute of Medicine and the National Academy of Sciences.

Many Canadians, such as Thomas J. Hudson, have made monumental advances in genetics. Tom Hudson was born June 12, 1961, in Arvida, Quebec (Wikipedia, 2007). Hudson received his Medical degree (M.D.) at the Université de Montréal, in 1985, later working as a resident in Internal Medicine, Clinical Immunology and Allergy at McGill University. In 1990, Hudson decided to attend the Massachusetts Institute of Technology, working with David Housman. In Massachusetts he joined in an attempt to construct a physical map of the human genome. As Wikipedia (2007) states, by the year 2006, he and his devoted team mapped over 10,000 genes. From 1995 to 2001, Hudson became assistant director of the Whitehead Institute/MIT Center for Genome Research. Thomas Hudson was attracted back to Montreal, at the McGill University Health Center, in 1996, where he started the Montreal Genome Center. He and his team eventually moved into the McGill University and Genome Quebec Innovation Centre.

Until 2006, Tom Hudson worked as a doctor and geneticist in McGill, with his main scope on the detection of diseases. His most amazing discoveries include discovering genes for Type II diabetes, susceptibility to leprosy, multiple sclerosis, asthma and inflammatory bowel disease. One of Hudson's publications about inflammatory bowel disease in 2001 led to the start of the Haplotype map project. This map was completed in 2005. From 2001 to 2005, Hudson was associate director of the Canadian Genetics Disease Network, and started Genome Quebec and Genome Canada. In 2006, Hudson left Montreal for Toronto, where he would lead the new Ontario Institute for Cancer Research in Toronto, Ontario. Hudson, due to his progress in science, has received innumerable awards such as: a winner of Canada's Top 40 under 40, in 1998, the Radio-Canada scientist of the year award in 2000. In 2001, Hudson won the

Robert H. Haynes Young Scientist Award by the Genetics Society of Canada. He received a Burroughs-Wellcome Clinician-Scientist Award in 2002, and was voted as the person who made the most significant contribution to healthcare by the readers of MacLean's magazine, in 2005. Finally, in 2006, Hudson was elected to the Royal Society of Canada.

Another extremely important contributor to Canadian genetics is Professor Majid Belouchi. Professor Belouchi was born in Morocco (M. Belouchi, personal communication, March 29, 2007). Dr. Belouchi received his Ph.D. in Molecular Biology from the Université de Montréal. He was the former director of Scientific Affairs at a gene discovery company called Algene Biotechnologies (Genizon, 2007). Dr. Belouchi has over 20 years of experience in his fields of genetics and molecular biology, and played a tremendously important role in mapping genes involved in common diseases. These include such diseases as Schizophrenia. Dr. Belouchi was the first person to sequence DNA in Canada, and worked as the scientific director of affairs at Algène Biotechnology Inc. He has also mapped many diseases in mice, with relevance to diseases in human beings. He was recently implicated in the treatment of diabetes in mice. Genizon (2007) also states that Dr. Belouchi is a professor at the Université de Montréal, as well as the original founder of Genizon Canada, and sells genetic information to Pfizer and Genetech.