HUMAN GENOME PROJECT (HGP)

The genetic blueprint of an organism or individual is encoded within their DNA sequences. Variations between two individuals naturally manifest as differences in their nucleotide sequences. To fully comprehend these discrepancies, it became imperative to map the entire human genome. The advent of genetic-engineering techniques, facilitating the isolation and cloning of DNA fragments, alongside the development of rapid and efficient DNA sequencing methods, paved the way for an ambitious endeavor: the Human Genome Project (HGP).

The HGP was an extensive international research initiative aimed at comprehensively mapping and understanding all human genes. Collectively, the genes residing on the haploid set of chromosomes constitute the genome. Initiated in 1990, the HGP was a 13-year endeavor coordinated by the US Department of Energy and the National Institutes of Health. The project expanded with the Welcome Trust (UK) joining as a major partner, alongside contributions from Japan, France, Germany, China, and other nations.

The completion of the project in 2003 marked a significant milestone in scientific history. The HGP earned the moniker of a megaproject due to several factors:

a. The exorbitant cost, estimated at 9 billion US dollars, with each sequenced base pair costing approximately US$3.
b. The vast number of base pairs (around 3 billion) requiring identification and sequencing.
c. The need for a substantial workforce comprising scientists, technicians, and support staff.
d. The immense volume of data generated, equivalent to approximately 3300 books, each containing 1000 pages filled with 1000 typed letters. However, advancements in high-speed computational devices facilitated the storage, retrieval, and analysis of this data.
e. The emergence of the science of Bioinformatics during this period played a pivotal role in aiding the HGP's progress.

Goals of HGP

The Human Genome Project (HGP) aimed to achieve several significant objectives:

(i) Identify all the approximately 20,000-25,000 genes present in human DNA.
(ii) Determine the sequences of the 3 billion chemical base pairs constituting human DNA.
(iii) Store this vast amount of information in databases.
(iv) Enhance tools for the analysis of genetic data.
(v) Facilitate the transfer of related technologies to other sectors, including industries.
(vi) Address ethical, legal, and social issues (ELSI) arising from genetic research and genomic data.
(vii) Foster the development of Bioinformatics, a burgeoning field closely intertwined with the HGP's advancements.
(viii) Sequence the DNA of model organisms: Studying the DNA sequences of non-human organisms offers insights into their inherent capabilities, with potential applications in healthcare, agriculture, energy production, and environmental remediation. Various non-human model organisms, such as bacteria, yeast, Caenorhabditis elegans (a non-pathogenic nematode), Drosophila, and plants like rice and Arabidopsis, have been sequenced as part of this initiative.

Methodology

Two methodologies have been acknowledged for scrutinizing the human genome:

(i) Expressed Sequence Tags (ESTs): This approach focuses on identifying all genes expressed as RNA.
(ii) Sequence Annotation: This method involves sequencing both coding and noncoding regions of the entire genome and attributing various functions to different regions. The sequencing aspect of the project was concluded in 2003, with the sequencing of chromosome I being the final task, completed in May 2006.

Salient Features of Human Genome

(1) The human genome encompasses approximately 30,000 genes, a figure significantly lower than the previous estimate ranging from 80,000 to 100,000.
(2) The human genome comprises 3.1647 billion nucleotide bases.
(3) Less than 2% of the genome consists of structural genes responsible for encoding proteins.
(4) The average size of a gene is around 3000 base pairs, although sizes vary considerably, with the largest known human gene, dystrophin, spanning 2.4 million bases.
(5) Chromosome I boast the highest gene count, with 2968 genes, while the Y-chromosome harbors the lowest number of genes, totaling 231, within the human chromosome set.
(6) Approximately 99.9% of nucleotide bases are identical across all individuals.
(7) Roughly 1.4 million locations exhibit single nucleotide differences known as single nucleotide polymorphisms (SNPs). These variations hold promise for identifying disease-associated sequences and tracing human ancestry.
(8) Repeated sequences constitute a substantial portion of the human genome.
(9) Repeated sequences are stretches of DNA sequences replicated hundreds to thousands of times. While they may lack direct coding functions, they offer insights into chromosome structure, dynamics, and evolution. Several non-human organisms, including bacteria (E. coli), yeast, Caenorhabditis elegans, Drosophila, and plants (such as rice and Arabidopsis), have also undergone sequencing efforts.

Applications and Future Challenges

1) The completion of the initial phase of the Human Genome Project has been likened to the discovery of antibiotics due to its role in unveiling a vast database of knowledge regarding numerous facets of the human genome.
2) In the near future, we anticipate mapping all human genes, sequences, transposons, and non-coding DNA.
3) There exist over 1200 genes associated with common cardiovascular diseases, endocrine disorders like diabetes, Alzheimer's disease, various cancers, and other neurological conditions. By obtaining comprehensive insights into these genes, it may become feasible to understand how to modify them and potentially eliminate the risk of developing such disorders.
4) Single-gene defects are responsible for a multitude of hereditary diseases that have the potential for correction.
5) The ability to study the interactions among different genes, proteins, and the mechanisms underlying tissue and organ formation, tumor development, or transitions between developmental stages will soon become achievable.
6) This advancement holds promise for promoting healthier and longer lives, developing designer drugs, and tailoring genetically modified diets to suit the individual needs of human beings.