Human Genome project

- In uncoiled state the molecule of DNA from a human cell approximately measure about 2 meters
- This DNA is accommodated in compressed form inside the microspace of the nucleus
- The human genome compriss about 3 billion nucleotide pairs
- DNA is packaged in chromosomes with the help of Histone proteins forms nucleosome (fundamental packing unit is nucleosome)
- These nucleosomes are subsequently coiled in to solenoids
- Chromatin is differentiated in to Euchromatin (light stained) Heterochromatin (Deeply stained)
- A nucleosome is a sub unit of chromatin composed of short DNA wrapped twice arounda central core of 8 histone molecules. A second Hstone (Hj) fastens the DNA to the nucleosome core.
- The genome is an organism's complete set of DNA instructions *(*whole hereditary information encoded in DNA). Genomes vary widely in size.
- The smallest known genome for a free living organism (a bacterium) contains about 6,00,000 DNA base pairs. Bacterial genome consists of a single 'chromosome' (circular naked DNA with out proteins)
- All human cells contain a complete genome except cells such as mature RBC.
- All stretches of DNA do not act as'*coding'* genes.
- Only about 3% of genome is *functional* (consists of DNA encoding proteins and regulative sequences).The remaining 97% of DNA is referred to as "Junk DNA" (non-functional)

Goals of HGP

- To identify all 20,000 to 25,000 protein coding genes in human DNA.
- To determine the sequences of the 3 billion base pairs of human DNA.
- To store the information in Human DNA Data Base and make it available to all.
- To improve tools for data analysis.
- To transfer related technologies to industries.
- To look in to ELSI (Ethical, Legal and Social Issues) that may crop up from HGP.
- Initial goal is to locate at least 3000 Genetic
- Genetic Markers(genes or other identifiable loci).

Launching of **Human Genome** Pro**ject** (HGP)

- The genetic make up of an organism lies in the "sequences of bases in its DNA.
- The difference between two individuals is due to the differences in their DNA sequences (at least at some places).
- The availability of simple and fast techniques (DNA tech) for determining DNA sequences paved the way for the launching of an ambitious & megaproject (HGP) of sequencing human genome in 1990.
- Human genome is supposed to have about $3x10^9$BP
- Human genome project (HGP) was associated with the rapid growth of bioinformatics.
- HGP was a 13 year project co-ordinated by U.S Department of Energy (DOE) and National Institute of Health (NIH).
- HGP was completed in 2003 (earlier than planned).

Approaches of HGP

- The four complementary approaches that are used in mapping the complete human genome are as follows.

A. Genetic mapping of the human genome:

- Initial goal is to locate at least 3000 genetic markers spaced evenly throughout the chromosomes.
- It is made feasible by the abundant occurrence of certain repeated sequences of DNA which can be separated / 'cut'by using 'Restriction Endonucleases' These fragments are called "restriction fragment length polymorphs" - RFLPS
- The resulting map will make it easy to find the loci of the other markers.

B. Physical mapping of human genome:

- It is done by cutting each chromosome into a number of identifiable fragments and then determining their actual order in the chromosome.

C. Sequencing the human genome is the process of determining the exact order of the nucleotide base pairs of each chromosome. This is the most time consuming part of the project.

D. Analyzing the genomes of other species such **as** *E.coli,* yeast**,** *Arabidopsis* and mouse**.**

- The above four approaches are yielding data which together provide a complete map of the human genome.

Salient features of HGP

- Each human chromosome contains hundreds of genes separated by intergenic regions which may contain regulatory sequences and ‘jun k DNA’ .
- Genes and gene related sequences account for 30% of DNA.10% of it has genes encoding proteins or RNA. 90% of this DNA is non-coding like regulatory sequences, gene families and pseudogenes.

- 70 % of DNA occurs as extragenic DNA. 20 % of it is repetitive DNA which includes dispersed and clustered repeats.80 % of it is low copy number DNA
- About 3164.7 million nucleotide base pairs are present in human genome.The average gene contains about 3000 base pairs.The largest known gene contains 2.4 million bps.

The total numbers of protein encoding genes are about 30,000. They are less than 1.5% of the human genome.

- Chromosome 1(number 1 of the 23 chromosomal set) has the highest number of genes (2968).
- Y chromosome has only 231 genes. SNPS ('snips') single nucleotide polymorphs. They are 1.4 million single base DNA differences which are called single nucleotide polymorphs. They are useful in 'DNA Finger printing'.

Note: A set of similar genes that occur in different species is called a Gene family.
Genes that have lost their protein coding ability are called pseudo genes.

Advantages of human genome project:

- Application of genome research in the area of human health care and molecular medicine: The identification and mapping of genes responsible for genetic diseases will help in diagnosis, treatment and prevention of such diseases like Alzheimer's and familial breast cancer (cancer running in families - genetically inherited/transmitted cancers).
- Application of microbial genomics research for safe, efficient and environmental protection.

- Knowledge of microbial genomics can also be used in the production of "biofuels".
  Understanding genomics will help us understand human evolution.
- Application of DNA forensics help to identify potential suspects whose DNA may match evidence left at crime scenes.
- Understanding plant and animal genomes will allow us to create several disease resistant varieties lowering the production costs of agriculture and live stock.
- Incorporation of edible vaccines into food products.
- Use of **bioengineered seeds** to grow insect and drought resistant crops.
- To produce bacterial enzymes that breakdown explosives and wastes (domestic and industrial).

Gene mapping:

- Assigning of fragments of DNA to chromosomes / mapping of genes to specific location with in chromosomes is known as genetic mapping.

There are two types of gene mapping, **i) Genetic mapping ii) Physical mapping**.

Genetic mapping:

- It is the process of measuring the distance between two or more loci on a chromosome. using "linkage analysis" or "cross over analysis" first studied in the fruit fly by Morgan.
- A single copy of genome includes protein coding genes, non-protein coding genes and junk DNA.
- Protein coding genes are of two types namely solitary coding genes and duplicated genes.

Solitary coding genes

- If, the sequence of protein coding genes occurs only once in the haploid genome, it is known, as **solitary coding genes**. About50% of the protein coding genes are solitary coding genes.

Duplicated genes

- The protein coding genes that are formed by the duplication of the ancestral gene and subsequent independent mutations are called duplicated genes.
- They occur in clusters and encode closely related proteins which have homologous but non-identical amino acids sequences. Such proteins differ slightly in their properties.

- Moderately repeated DNA includes the tandemly repeated genes encoding rRNA, tRNA, histones and mobile genes called transposons. (Sequence of DNA that can move around to different positions within the chromosome / genome).
- TVanposons are involved in the transposition of DNA.
- Simple sequence of DNA which includes very short sequences repeated in long tandem arrays is located in centromeres, telomeres and specific spots in the arms of chromosomes.

Physical mapping

- It is mapping by using available techniques like Somatic cell hybridization (SCH), Fluorescent in situ hybridization (FISH) etc., to determine the absolute position of a gene on a chromosome.
- To identify a gene responsible for a disease is also referred to as "gene mapping".
- Genetic maps provide powerful tools for gene identification, study, utilization and isolation.

Fine mapping

It involves the identification of **markers** that are very tightly linked to a targeted gene.

- The advent of DNA markers has made "fine mapping" possible in any plant that can be crossed.
- A 'genetic fine map' of a specific locus usually aims at the identification and location of markers that are closely linked to the targeted gene and are within one or fewer centi Morgans (units of relative distances betwen two genes calculated on the basis of cross over percentages).

DNA Finger printing / DNA typing / DNA profiling

- The patterns of ridges of skin of fingers, toes, palms and soles in human beings constitute "finger prints". They are unique to each individual and can be altered only through surgery.
- Just as human beings are identified from one another by their unique finger prints, they can also be identified with unique base sequences in their DNA too. Hence the name DNA fineer printing.
- DNA finger printing technique was developed by Alec Jeffreys at Leicester University (UK-1984). While he was studying the gene for myoglobin, he found that the myoglobin gene contains many short segments which vary in size and composition from individual to individual.

Basis of DNA Finger printing

- **Over** 99% **of the** 3 billion nucleotide pairs in human DNA are identical among all individuals. A large portion of DNA **is** junk DNA.
- Changes in the junk DNA (which occur often) are not expressed. Restriction Fragment Length Polymorphs **(RFLPs pronounced** 'riflips' **)** are characteristic to every person **.**
- **They are called** Variable Number Tandem Repeats **(**VNTRs) and are useful as 'GENETIC MARKERS'**.**
- They are bits of chromosomes cut by Restriction Endonucleases. The 'Fundamental Technique ' involved in DNA Finger Printing **was** pioneered and perfected **by** Jeffrys **of** Great Britain.

- The VNTRs of two persons differ in the number of tandem repeats or the sequence of *bases*. Such changes are caused due to mutations and gene recombinations.

- They serve as basis of a technique called DNA FINGER PRINTING

- Note the numbers of 'Tandem Repeats' in the chromosomes of the mother and father
- Forensic experts have developed Short Tandem Repeats (STRs) a set of genetic markers analyzed by PCR. Even trace samples can be typed by STR.
- In STR the repeated nucleotides are 2 to 9 base pairs.

Protocol of a genetic fingerprint

A. Obtaining DNA **:** The first step is to o btain a sample of DNA from blood**,**saliva ,hair roots ,semen etc.
B. Fragmenting DNA*:* Treating DNA with Restrition Enzymes which cut the DNA into smaller fragments by cutting it at specific sites .
C. the DNA can be amplified using PCR or rDNA methods **(**DNA amplification).
D. Denaturing DNA*:* The DNA on the gel is 'denatured' using alkaline chemicals or by heating' (denaturation means separation or splitting of the double helix **into**'single strands by breaking hydrogen bonds' between the two strands *).*

E. Separation of DNA fragments**:** DNA fragments are applied at one end of agarose gel plate When an electric current is applied to the gel **,** the DNA fragments (which are slightly negatively charged travel across the gel (smaller and more mobile pieces travel farther**).** This technique of separation of DNA fragments into individual bands is called 'Gel Electrophoresis'.

F. Blotting**:** A thin nylon membrane is placed over the 'size fractionated DNA strands' and covered by paper towels .As the towels draw moisture the DNA strands are transferred **on** to the nylon membrane through capillary action. This process is called 'Blotting' more precisely SOUTHERN BLOTTING **,** after the name of its inventor E.M. Southern.

G. Using probes to identify specific DNA: **A** radio active probe' (DNA labelled with a radio active substance)is added to the DNA bands. The 'PROBE' is a single stranded DNA molecule that is 'complementary' to the gene of interest (the probe attaches by base pairing to those restriction fragments that are complementary to it).

H. Hybridisation with probe **:**After the probe 'hybridises' and the excess probe washed off, a sheet of photographic film is placed on the membrane bearing 'DNA hybrids' **.**

I. Exposure on film to make a Genetic /DNA Finger Print*:* The radioactive label exposes the film to form an image (image of bands) corresponding to specific DNA bands. The thick and thin dark bands form a pattern of BARS which constitute **a'**Genetic Finger Print'

Applications of DNA finger printing

- Useful in the preservation of endangered species.
- Useful in the study of phylogentic relationship of different groups.
- Useful in the study of pedigree analysis ,
- Useful in Anthropological studies, of origin and migration of human populations through ages .
- It is useful in forensic medicine to establish the identity of the criminals among the suspected persons.
- In establishing paternity / maternity of a 'child under dispute'.

ELEMENTARY IDEA OF GENE THERAPY

- When genes are altered so that the encoded proteins are unable to carry out their normal functions, genetic disorders can result.
- Gene therapy aims at supplementing a defective mutant gene/allele with a normal functional sene**,** to treat genetic disorders. A normal gene may be inserted into a nonspecific location within the genome to replace a nonfunctional gene.

- **A** normal gene is introduced into a patient's target cells ,with the help of a VECTOR. Scientists took the logical step of trying to introduce genes straight into human cells, focusing on diseases caused by single-gene defects, such as cystic fibrosis, hemophilia, muscular dystrophy **and** sickle cell anemia**.**

The most common vectors used to carry therapeutic gene' **into target cells are the** viruses **(genetically altered viruses which carry normal human DNA).**

- Gene therapy may be i. Somatic gene therapy ii. Germ Line Gene therapy (sperms , ova).

Germ Line Therapy is controversial because it involves bringing changes in the 'genetic makeup' of the future generations. ETHICAL reasons are involved in it.

- Somatic gene therapy is done in two ways i. In *vivo* (genes are changed while cells are in the body) using techniques such as "MICRO INJECTION" ii. *Ex vivo* (defective cells are modified out side the body and then 'transplanted' back into the body).

- The First Gene Therapy was started with the treatment of 'Severe Combined Immuno Deficiency (SCID) which is caused due to a 'mutation' in the gene that codes for the enzyme 'Adenosine Deaminase' *(ADA)*. 'T' cells of SCID patient are isolated from the body of the patient and are mixed with genetically modified retroviruses carrying the 'normal copy' of ADA gene.
- These viruses insert the gene into human cell's genome'. These cells cultured for a while to make sure that they are active and then introduced into the blood stream of the patient.

- Various methods include l. Insertion of a normal gene into the genome of the person suffering from the disorder concerned ii. Exchanging an 'abnormal gene' with a 'normal gene' through 'Homologous Recombination' iii. Repair of 'abnormal gene' through' SELECTIVE REVERSE MUTATION'(site directed mutagenesis) iv. Altering a gene by controlling the 'Gene Regulator Mechanism'.
- Viral Vectors include i. RNA containing 'Retroviruses' ii. dsDNA (double stranded DNA) containing 'Adenoviruses' iii. ssDNA containing Adeno-associated viruses.
- To improve /ensure safe delivery of 'new DNA' into defective cells , DNA is protected from damage by using 'LIPOPLEXFS' and 'POLYPLEXES' during 'Transfection'.
  (LIPOPLEX= DNA covered with lipids ; POLYPLEX= DNA covered with 'complex polymers').
- Naked DNA can also be introduced into target cells by electroporation using Gene Gun. **Gene guns shoot** DNA coated sold particles **into cells using high pressure gases. Defective Genes** are **inactivated** by using 'synthetic' **oligo deoxv nucleotides.**

NOTE: **Introduction of foreign material into eukaryotic cells is called** Transfection. **It is done by using chemicals such as calcium chloride or by using 'Gene guns' which shoot the DN A directly into the nucleus of the target cell. Transfer through viral media is called** 'Transduction'