International Journal of Clinical and Biomedical Research

GENE MAPPING: BASICS, TECHNIQUES AND SIGNIFICANCE

*VIKRAM NIMBALKAR[1], DEEPALI CHIKTE[1], TEJAL RAMAN[1], SAHIL MANIYAR[1], PANDURANG GAIKWAD[1]

[1]Department of Pharmacology, P.D.V.V.P.F's College of Pharmacy, Vilad Ghat, Ahmednagar, Maharashtra .414001.

*Corresponding author: rajevikram@gmail.com

ABSTRACT

Watson says, "Like the system of interstate highways spanning our country, the map of the human genome will be completed stretch by stretch". It may be possible to use genetic information to diagnose the disease accurately and to predict a patient's likely response to a particular medicine or treatment. For whole genome mapping development and application of mapping, sequencing and computational tools are very essential and also linkage, physical and sequence maps are required to put the information together. For most genome mapping projects involve markers consisting of a unique site in the genome and should be independent of any particular experimental resource. For mapping purpose the DNA and RNA identification is essential. These genes are identified by hybridizing DNA clones against Northern blot, cDNA libraries, Zoo blot, Western blot and Southern blot of genomic DNA digested with rare cutter restriction endonuclease. The various experimental studies of gene mapping have extended our understanding of the genetics. This has allowed the investigators to detect a particular gene, which is responsible for the disease. Recent studies have shown the various effective and scientific gene mapping techniques and gene identification methods, which are helpful to diagnose a particular disease. It is easy for the doctor to give right medicine to the right patient to cure the disease when he can identify the defective gene responsible for disease. This article reviews the details of identification techniques of genes, gene mapping with broad applications.

KEY WORDS: Gene mapping, Restriction mapping, Fluorescent *in situ* hybridization (FISH), Sequenced tagged site (STS) mapping, Somatic cell hybridization.

INTRODUCTION

"All Human diseases are genetic in origin", 97% of genome is not functional. Gene map is essential for searching the genetic bases of complex disease, such as diabetes and cancer. A map has tremendous value for identifying disease genes.

These genes hold all our hereditary information and provide genetic code that allows our body to develop grow and function. These genes are made up of DNA (Deoxyribonucleic acid). DNA is a genetic alphabet. We inherit half of our genetic information from our mother and half from our father. Our genes are packed in structures called chromosomes. Human have 23 pairs of chromosomes in almost every cell of body. Mapping variations in human genome across population is essential if scientists are evergoing to understand its three billion letters and cure disease. There are many small seemingly innocuous variations in the human genome even though

humans are 99.9% identical to each other; the last 0.1% still contains millions of differences (according to Dr. Francis Collins, director of the U.S. National Human Genome Research Institute).

Mapping genes and the creation of genetic map is a fundamental part of science of genetics. A genetic map is essentially a set of locations or co-ordinates. These locations identify a specific place where each one of the gene or DNA sequence are specifically located in the genome. Gene mapping is the technology whereby the location of a gene or other DNA sequence is described. Gene is a small segment of polymorphic chain consisting of hundreds and thousands of nucleotides. These nucleotides are the subunits of the genes.

In gene mapping first step is to determine the approximate location of diseased gene on one of the 23rd chromosome. Each

section of chromosome has several hundreds of genes on it. So each must be located and tested individually to determine which one triggers the disease for e.g. Huntington's disease. Researchers took ten years to identify the gene after first marker was discovered. Recently much more effective technology has been introduced. To date, about 1700 of the estimated 50,000 to 1, 00,000 human genes (less than 2%) have been mapped. The Human Genome Project had been carried out in 1991-2003 using the various gene-mapping techniques.

In order to map the locus of a trait by genetic linkage, a panel of markers is tested for evidence of segregation with trait at meiosis. For mapping a new gene it is necessary to have a large number of different markers, ideally evenly spaced along each chromosome. Moving up the evolutionary tree, the generation of markers becomes increasingly difficult and in humans is ethically unaccepted. Comparison of the sequence of a diseased gene and its product with those of genes and proteins whose sequence and function are known can provide clues to the molecular and cellular cause of the particular disease. More than 10,000 human genes were catalogued online.

GENE MAPPING:

Gene mapping is a powerful technique used in a molecular genetics to identify an unknown gene in a family with a particular inherited disorder. This can be done by following types:

TYPES OF GENE MAPPING:

Gene mapping is of two types [5]

1. Genetic mapping
2. Physical mapping

1 GENETIC MAPPING:

Genetic mapping is based on the principles of inheritance as first described by Gregor Mendel in 1865 [13]. Using linkage analysis to determine the relative positions between two genes on a chromosome.

a) Genes were the first markers to be used: - The first genetic map constructed in early decades of 20th century for organisms for e.g. fruit fly, used genes as markers. It was understood that genes are the segments of DNA molecules.

b) DNA markers for genetic mapping: - Genes are very useful markers but they are not ideal. With the large genomes in vertebrates and flowering plant, one problem is that a map based entirely on genes is not detailed. Mapped features that are not genes are called DNA markers. As gene markers, DNA markers must have at least two allele. There are three types of DNA sequence feature that satisfy the requirement.

   I. Restriction Fragment Length Polymorphism (RFLPs).
  II. Simple Sequence Length polymorphism (SSLPs).
 III. Single Nucleotide Polymorphism (SNPs).

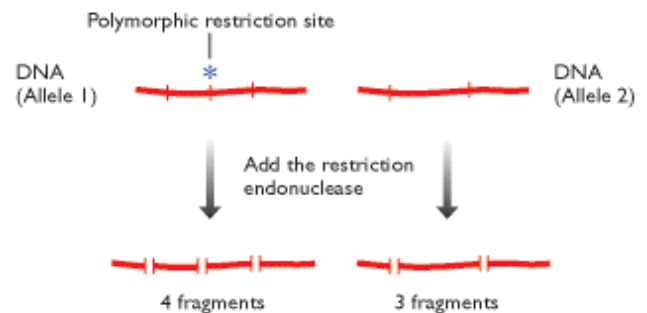## I. Restriction Fragment Length Polymorphism (RFLPs)[4] (Fig 1).

Fig 1: A restriction fragment length polymorphism (RFLPs)

The DNA molecule on the left has a polymorphic restriction site (marked with the asterisk) that is not present in the molecule on the right. The RFLP is revealed after treatment with the restriction enzyme because one of the molecules is cut into four fragments whereas the other is cut into three fragments

It is the first DNA marker to be studied. DNA molecule treatment with the restriction enzyme should always produce same set of fragments but it is not always because genomic DNA has some restriction sites are polymorphic existing two allele one displaying correct and other having a sequence alteration. The ultimate effect of this sequence alteration is that restriction fragments remain linked together after treatment with enzyme. There are 10^5 RFLPs human genome but for each RFLP there are only two alleles (with or without site).

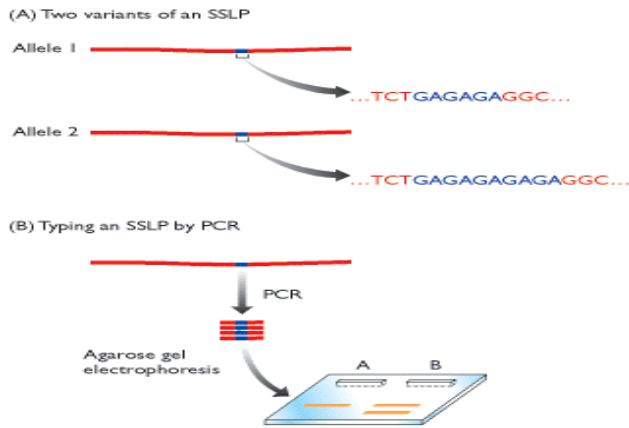## II. Simple Sequence Length polymorphism (SSLPs)

Fig 2: SSLPs and how they are typed

(A) Two alleles of a microsatellite SSLPs. In allele 1 the motif 'GA' is repeated three times, and in allele 2 it is repeated five times. (B) How the SSLPs could be typed by PCR. The region surrounding the SSLPs is amplified and the products loaded into lane A of the agarose gel. Lane B contains DNA markers that show the sizes of the bands given after PCR of the two alleles. The band in lane A is the same size as the larger of the two DNA markers, showing that the DNA that was tested contained allele 2.

These are the era of repeat sequence that displays the length variations. Different alleles containing different numbers of repeat units and these are multi allelic. There are two types of SSLPs
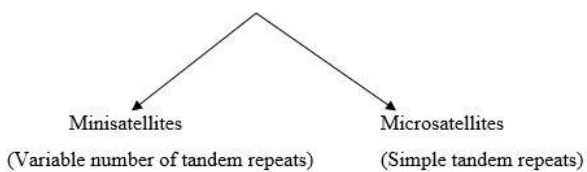
Minisatellites are more popular then microsatellites because minisatellites are not spread along the genomes but found in telomeric regions at the end of chromosome. On the other end microsatellites are more conveniently spaced throughout the genomes.

II. Single Nucleotide Polymorphism (SNPs).

These are the positions where some individuals have one nucleotide and others have different nucleotide. There are vast numbers of SNPs in every genome some also gives rise to RFLP

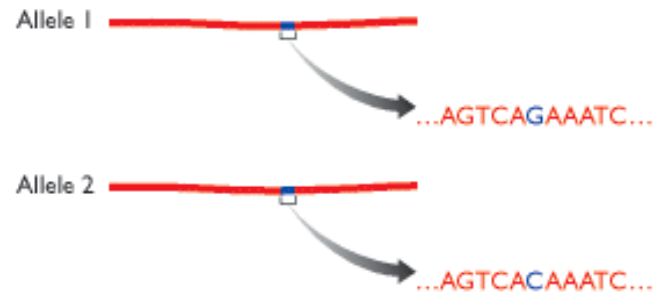but some does not because the sequence in which they lie is not recognized by restriction enzyme.

Fig 3: Single Nucleotide Polymorphism (SNPs)

In human genome atleast 1.42 million SNPs exists, only 1 lakh of which result in an RFLP[16]. These SNPs selection is more rapid because it is based on oligonucleotide hybridization analysis.

2   PHYSICAL MAPPING:

Somatic cell hybridization the map generated by genetic technique is sufficient for directing the sequencing phase of a genome project because of two reasons

a)   The resolution of genetic map depends on the number of crossovers that had been had scored.

b)   Genetic maps have limited accuracy[13].

Because of the limitations of genetic mapping a plethora of physical mapping technique has been developed, the most important being:

I.   Restriction mapping
II.   Fluorescent *in situ* hybridization (FISH).
III. Sequenced tagged site (STS) mapping.
IV. Somatic cell hybridization

I. Restriction mapping:

The results of single and double restrictions are shown at the top. The sizes of the fragments given after double restriction enable two alternative maps to be constructed, as explained in the central panel, the unresolved issue being the position of one of the three BamHI sites. The two maps are tested by a partial BamHI restriction (bottom), which shows that Map II is the correct one.
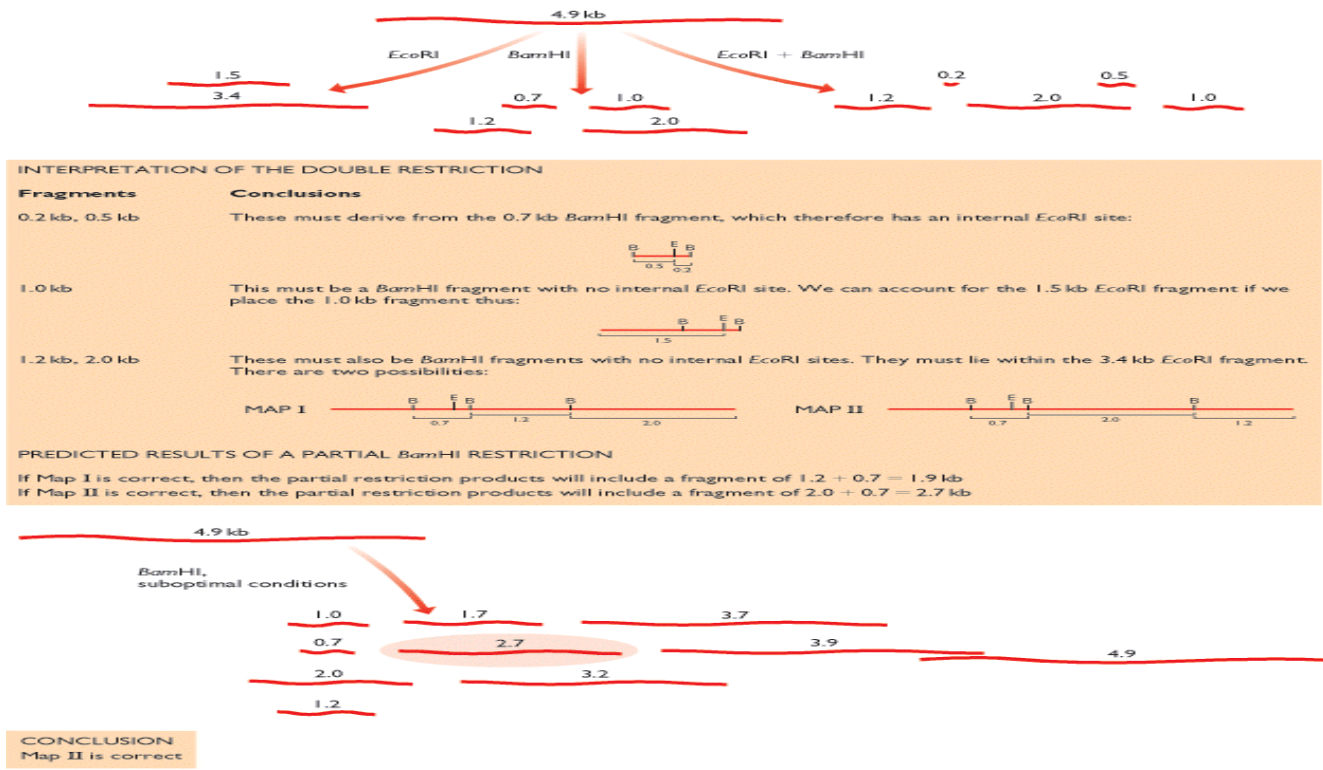
Fig 4: The objective is to map the EcoRI (E) and BamHI (B) sites in a linear DNA molecule of 4.9kb

This locates the relative positions on a DNA molecule of the recognition sequence for restriction endonucleases. In genetic mapping using RFLP, DNA markers can locate positions of polymorphic restriction sites within genome but very few restriction sites are polymorphic. So many sites are nit mapped by the technique so by increasing the marker density on a genome map by using alternative method to locate the positions of the nonpolymorphic restriction sites. This can be achieved by restriction mapping but it is applicable only for small DNA molecule.

It is also possible to use methods other than electrophoresis to map restriction sites in DNA molecules the technique called as Optical mapping [8, 15], restriction sites are directly located by looking at the cut DNA molecules with the microscope.

The image shows a 2.4-Mb segment of the Deinococcus radiodurans genomes after treatment with the restriction endonuclease nhel. The positions of the cut sites are visible as gaps in the white strand of DNA.

Fig 5: Optical mapping

The DNA must be first attached to a glass slide in such a way that the individual molecules become stretched out, rather than clumped together in a mass. There are two ways of doing this
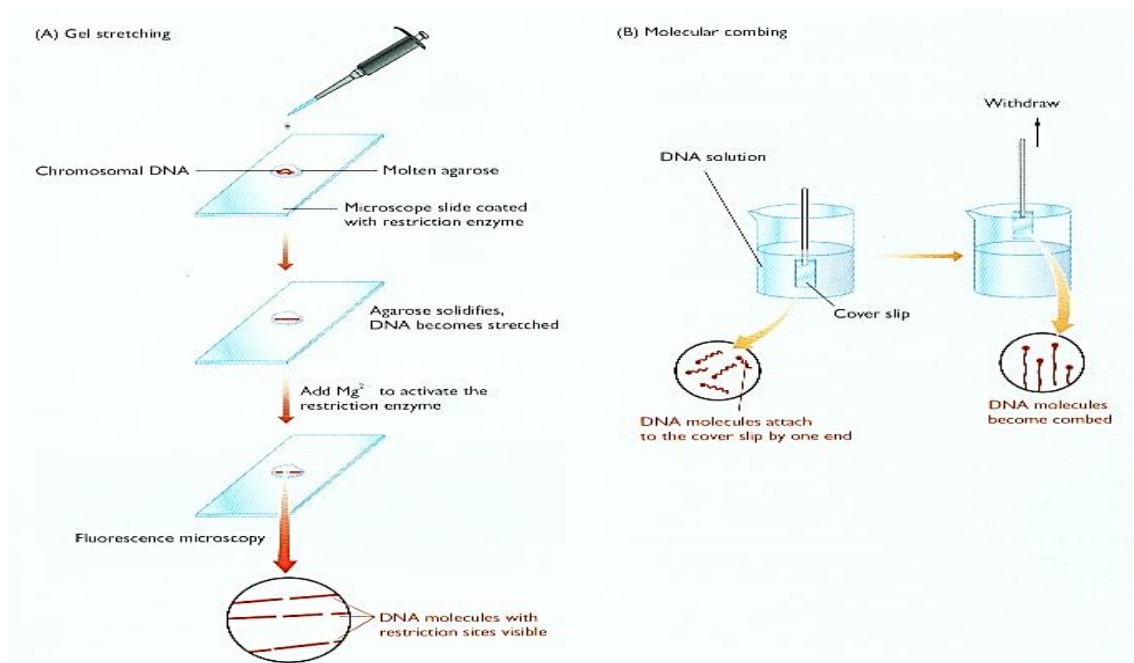
1.   Gel stretching

2.   Molecular combing.[10]

Fig 6: Gel stretching & Molecular combing

(A) To carry out Gel stretching, molten agarose containing chromosomal DNA molecules is pipetted on to a microscope slide coated with the restriction enzyme. As the gel solidifies, the DNA molecules become stretched. It is not understood why this happens but it is thought that fluid movement on the glass surface during gelatin might be responsible. Addition of magnesium chloride activates the restriction enzyme, which cuts the DNA molecules. As the molecules gradually coil up, the gaps representing the cut sites become visible.

(B) In Molecular combing, a cover slip is dipped in to a solution of DNA. The DNA molecules attach to the cover slip by their ends, and the slip is withdrawn from the solution at a rate of 0.3mm s^{-1}, which produces a 'comb' of parallel molecules.

II. Fluorescent *in situ* hybridization (FISH)[6]

A sample of dividing cell is dried onto a microscope slide and treated with formalmide so that the chromosomes becomes denatured but do not loose their characteristic metaphase morphologies. The position at which probe hybridizes to the chromosomal DNA is visualized by detecting the fluorescent signal emitted by the labeled DNA.

The optical mapping method (developed by Pinkel et al., 1986)[14] provides a link to a second type of physical mapping i.e. FISH[5]. In this technique the marker is a DNA sequence that is visualized by hybridization with a fluorescent probe
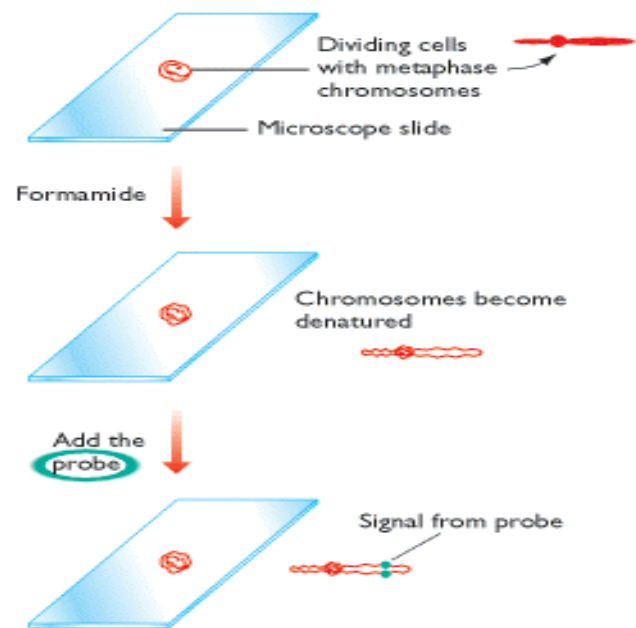

Fig 7: Fluorescent *in situ* hybridization

a)   In situ hybridization with radioactive or fluorescent probes-In this method the intact chromosome is examined by probing it with labeled DNA molecule. Denaturation of DNA is essential only then chromosomal DNA is able to hybridize with the probe. The standard method of denaturing chromosomal DNA without destroying the morphology of the chromosome is to dry the preparation on to a glass microscope slide and then treat with formamide. High resolution is possible only when the radiolabel probe with low emission energy such as 3H but these have low sensitivity. In late 1980 development of nonradioactive fluorescent probes. These having combine high sensitivity and high resolution for in situ hybridization

b)   FISH in action[20, 21] - FISH was originally used with metaphase chromosome so because of the highly condensed nature of metaphase chromosome only low-resolution mapping is possible. Since 1995 a range of higher resolution FISH technique has been developed. According to technique by changing the nature of the chromosomes preparation. This can be done by 2 ways

A. Mechanically stretched chromosome: - In this method centrifugation generates shear forces, which results the chromosomes being stretched upto 20 times their normal length[7].

B.   Non-Metaphase chromosome: - Metaphase chromosomes are highly condensed so others like prophase and interphase chromosomes are used. These interphase chromosomes contain the most unpacked of all cellular DNA molecules. To improve the resolution of FISH to better than 25 kb it is necessary to abandon intact chromosome and instead use purified DNA. This approach is called as fiber FISH[20, 21].

III. Sequenced tagged site (STS) mapping:

To generate a detailed physical map of a large genome we need ideally, a high resolution mapping procedure that is rapid and not technically demanding. At present most powerful physical mapping technique and one that has been responsible for generation of the most detailed map of large genomes is STS mapping. It is simply short sequence between 100 and 500 bp

in length that is easily recognizable and occurs only once in the chromosome or genome being studied.

To quantify as an STS a DNA sequence must satisfy two criteria
A. Its sequence must be known.
B. STS must have unique location in the chromosome being studied. STS can be obtained in many ways, the most common sources
1.   Expressed Sequence Tags (ESTs) [1,9]
2.   Simple Sequence Length Polymorphism (SSLPs).
3.   Random Genomic Sequence.
Having a map of entire human genome will make it theoretically possible to identify every gene that contributes to them, once a new gene identified, it is immediately sequenced to understand nature of protein it codes for and to identify mutations that are related to disease. The number of genes mapped grew from 579 in 1981 to 1879 in 1991.
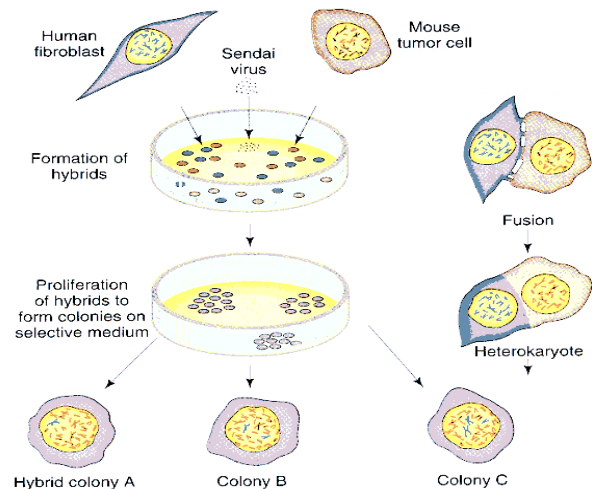
IV Somatic cell hybridization:

Fig 8: **Somatic Cell Hybridization**

Under the influence of Sendai virus, human fibroblast cells and mouse tumor cells grown together in culture will fuse to form a heterokaryote, in which the nuclei fuse. Because the cell machinery is primarily murine, over successive mitotic divisions, human chromosomes (in blue) are gradually lost. Eventually, a number of hybrid cell lines are formed that have only one (Colony A) or a few (Colonies B & C) human chromosomes, which are readily identifiable. These cell lines

can be screened for the presence of various human biochemical markers, which are then assignable to the retained chromosome. Sendai virus is used because it has several number of attachment, it can attach two different cells if closed together. Virus can mediate fusion of cell from different species and it is very small as compared to cell.

Chromosome walking[4] is the technique in which

> ➢ Overlapping clones to be used to build a contig (Contiguous Sequence).
> ➢ Clones are used to bridge the gap between different contigs.
> ➢ Ends of unconnected contigs are used as probes to 'walk' between chromosomes.

3. GENE IDENTIFICATION TECHNIQUES:

Gene identification is a very important tool in gene mapping. BLOTTING TECHNIQUES FOR DETECTION OF SPECIFIC DNA FRAGMENTS AND m-RNAs WITH DNA PROBES:

1. Southern Blotting
2. Northern Blotting
3. Western Blotting
4. Zoo blots
5. cDNA libraries
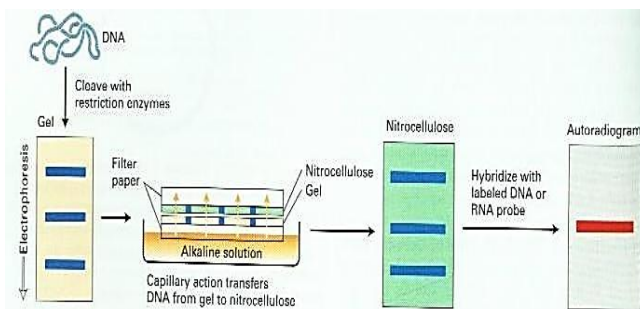
1. Southern Blotting:

Fig 9: Southern blot technique

It can detect a specific DNA fragment in a complex mixture of restriction fragments. The diagram depicts three different restriction fragments in the gel, but the procedure can be applied

to a mixture of millions of DNA fragments. Only fragments that hybridize to a labeled probe will give a signal on an autoradiogram. A similar technique called Northern blotting detects specific mRNAs within a mixture.

This method developed by a molecular biologist E.M Southern[17] in 1975 for analyzing the related genes in a DNA restriction fragment is called as southern blotting technique. This blot can easily provide a physical map of restriction sites within a gene located normally on a chromosome and reveal the number of copies of gene in the genome, and of gene when compared with the other complimentary genes. Thus the sequences of DNA are recognized following the sequence of nucleic acid probe.

2. Northern Blotting:[2,18]

Southern blotting is not directly blot transfer of m-RNA separated by gel electrophoresis because RNA does not bind to nitrocellulose filter.

☐ RNA bands are blot transfer from gel to chemically reactive paper.

☐ It is the reusable technique because covalent bonding of RNA with chemically reactive paper.

☐ m-RNA separated on gel transfer to nitrocellulose filter paper, which hybridized by single strand probe.

☐ Hybrid treats with SI nuclease and RNAase, which digest single stranded RNA/DNA probe.

☐ It does no affect double stranded nucleic acid structure of m-RNA. It protects nucleic acid probe.

3. Western Blotting:

Towbin et al.[19] in 1979 developed the western blotting technique to find out the newly encoded protein by a transformed cell. In this method radio labeled nucleic acid probes are not used. This technique follows the following steps.

1 Electrophoresis of protein in polyacrylamide gel.
2 Blotting of proteins onto nitrocellulose filter paper.
3 Hybridization of proteins by using radiollabeled antibodies of known structures.

4   Detection of hybridized sequences by autoradiography.

4.   Zoo blots:

A Zoo blot is a Southern Blot of genomic DNA samples from a wide variety of different species. A genomic DNA clone, which shows positive hybridization signals against the DNA of a variety of different species would be expected, therefore, to contain coding DNA sequence that have been strongly conserved during evolution.

5. cDNA Libraries[6]:

 M. Bento Soares (Columbia University) discussed strategies for constructing cDNA libraries for both gene discovery and characterization. To clone genes represented by low-abundance transcripts, subtractive hybridization strategies are being developed to eliminate pools of sequenced cDNAs. In addition, techniques are being optimized to produce libraries enriched for full-length cDNAs. These libraries will be very useful for increasing gene representation. Bernhard Korn (German Cancer Research Center) reported progress in constructing and gridding a full-length cDNA library from human fetal brain. His institution's current library has 120,000 clones with an average insert size of 1.8 kb. Some problems inherent in making such libraries were discussed. Several different approaches are available for gene identification and are reviewed by Monaco in 1994 (Table 1)[11, 12].

Table.1: Different   approaches   available   for   gene identification

Traditional approaches	Newer approaches
Conservation on Zoo lots	Exon amplification
CpG islands	cDNA selection
Northern blots	Genomic sequencing and computer analysis
cDNA library	Regionally mapped candidate genes

4. APPLICATIONS:

1. To find gene (quantitative trait loci) that are associated with traits of economic importance.

2. To develop comparative maps.

3. To discover gene causing major physiological defect as: -

☐ Locus BRCA2- Breast cancer

☐ E gene (ApoE)- Alzheimer's disease.

☐ Gene CRC- Melanoma

☐ Patched gene- Basal Cell Carcinoma

4.   Branches:

   a) Proteomics

   b) Pharmacogenomics

a) Proteomics

Proteomics is the term in the study of genetics, which refers to all the proteins expressed by a genome; proteomics involves the identification of proteins in the body and the determination of their role in physiological and pathophysiological functions. It is a new evolving field of science that seeks to specify all the proteins produced by the cell in all types of situations and environments and to understand how they function. Because proteins are the product of information coded for in DNA, proteomics is closely allied to the study of the genome.

Key Technologies for proteomics:

1. 1-D electrophoresis and 2-D electrophoresis are for the separation and visualization of the proteins

2. To identify and chararacterise proteins mass spectrometry, X=ray crystallography, and NMR are used.

3. To characterize protein-protein interactions affinity chromatography and protein expression systems like the yeast twos-hybrid and fluorescence resonance energy transfer (FRET) is used.

b) Pharmacogenomics:

Pharmacogenomics is the study of how an individual's genetic inheritance affects the body's response to drugs. This term comes from the word pharmacology and genomics and thus the intersection of pharmaceuticals and genetics. It is the branch of pharmaceutics, which deals with the influence of genetic variation of drug response in patients. It aims to develop

rational means to optimize pharmacotherapy, with respect to the patient's genotype, to ensure maximum efficacy with minimal adverse effects.

☐ Some severe but rear disorders are caused by the expression of a different form (allele) of a single gene is known as single gene or monogenic disorders. This disease includes –
I.   Cystic fibrosis
II.   Huntington's disease.
III. Hemophilia.

☐ Many common diseases that affect many billions of people arise through complex interaction between environment and number of genes is known as susceptibility genes, which alter risk of the disease developing or its severity.

This includes –
I.   Late onset Alzheimer's disease.
II.   Adult onset diabetes.
III. Cardiovascular disease.
IV.   Asthma
V.   Parkinson's disease.
VI.   Migraine.
VII. Schizophrenia.
VIII. Depression.
IX.   Chronic obstructive pulmonary disease.
X.   Osteoarthritis.

5.   Gene transposition in mammals:

The Minos transpons (jumping gene) can be used to randomly tag gene for   functional gene identification in mammalian cell lines and also mobilize, transpose genetic elements in somatic cells in mice.

Studies demonstrate germ lie transposition an insertion within genes in mice. This is helpful for identification of disease causing genes where the genetics is complex (disease of CNS, insulin resistance and Inflammation).

CONCLUSION

A DNA sequence, strung together, is the most precise type of map in that it contains both coding (gene-containing) and non-coding DNA. It is felt that obtaining the complete DNA sequence from the genome of many different organisms will provide scientists with vital information that will unlock many biological mysteries.

REFERENCES

1. Adams M. D. et al (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651-1656.

2. Alwine, J.C., Kemp, D.J., Parker, B.A., Reiser, J., Renart, J., Stark, G.R., et al, 1979. Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. Meths. Enzymol. 68, 220–242

3. Botstein,D., et al.1980.. Construction of a genetic linkage map in man using restriction fragmentlength polymorphism. Am. J. Genet. 32:314-331.

4. H.D.Keller,   P.Green,   C.Helms,   S.Cartinhour, B.Weiffenbach, K.Stephens, T.P. Keith et al. Cell 51: 319-337

5. Heiskanen M, Peltonen land Palotie A (1996) Visual mapping by high resolution FISH. Trends Genet., 12, 379-382

6. Laan M., Kallioniemi O-P.,Hellsten E., Alitalo K., Peltonen L. and Palotie A. (1995) Mechanically stretched chromosomes as targets for high resolution FISH mapping. Genome Research vol. 5, 13-20.

7.   Lin J, Qi R, Aston C,et al. (1999) Whole genome shotgun optical mapping of Deinococcus radiodurans. Science, 285, 1558-1562.

8. Marra MA, Hillier L and Warerstone RH (1998) Expressed sequence tags –ESTablishing bridges between genomes. Trends Genet. , 14, 4-7.

9. Michalet X, Ekong R, Fougerousse F, et al. (1997) Dynamic molecular combing: stretching the whole human genome for high-resolution studies. Science, 277, 1518-1523

10. Monaco A. F. (1994) Isolation of genes from cloned DNA. Current Opinion in Genetics and Development vol. 4, 360-365.

11. Monaco A. P., and Larin Z. (1994) YACs, BACs, and MACs: artificial chromosomes as research tools. Trends in Biotechnology 12, 280-286

12. Oliver S.G., Van der Aart QJM, Agostoni-Carbone ML, et al. (1992) The Complete DNA Sequence Of Yeast Chromosome III. Nature, 357, 38-46.

13. Pinkel D., Straume T. and Gray J. W. (1986) Cytogenetic analysis using quantitative, high sensitivity, fluorescence hybridization. Proc. Natl. Acad. Sci. USA 83, 2934-2938.

14. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ and Wang Y.k. (1993) ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science, 262,110-114.

15. SNP-SNP group (The International SNP Map Working Group 2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature, 409, 928-933.

16. Southern, E.M. (1975), Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

17. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA. 77:5201-5205.

18. Towbin, H.; Staehelin, T. and Gordon, J. (1979). Proc. Natl. Acad. Sci. USA. 76: 4350-4354.

19. Trask B.J., massa H, Kenwrick S and Gitschier J (1991) Mapping of human chromosome Xq28 by 2-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. Am. J. Hum. Genet.,48, 1-15

20. Trask B. J., Pinkel D and Van Den Engh G. J. (1989) The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids Spanning 250 kilobase pairs. Genomics 5, 710-717.

21. Wahl, G.M., J. L. Meinkoth, and A.R.Kimmel. 1987. Northern and Southern blots.Meth.Enzymol.152: 572-581.