THE HUMAN GENOME
AND
THE HUMAN GENOME PROJECT

BY – PRAMOD KU. MAHISH

INTRODUCTION

GENOME

Genome is a complet gene set of an organism.including coding and noncoding gene.

The human genome containt an estimated 100,000 genes,23 pair of chromosome(22 autosomal and one set of sex chromosome pair XX or XY) and the 3 billion base pair.

THE COMPLEXITY OF HUMAN GENOME

1. GENE ND GENE RELATED SEQUENCE:-

A. MULTIGENE FAMILIES:-

Prokariyotic genes are repeated only once in the genome,many Eukaryotic including human present in the multiple copies, called gene families. The human gene illustrated each o the 3 ways in which a multigene family can be organised.

1. With some families the individual genes are clustered together at a single position in the genome.

e.g.:- the growth hormone gene family, whose five members are clustered on chromosomes 17,and the 55 rRNA family, comprising 2000genes in a tandomarsy on the long arm of chromosome.

2. With other families the gene are dispersed around the genome.

e.g.:- The five members of the aldolase ggene families are located on chromosome 3,9,10,16,& 17.

3. Some large gene families are both clustered and dispersed at the same time.

e.g.:- rRNA gene about 280 copies clusters located on the short arms of chromosome 13,14,15,21,& 22.

The sequence of these member are difference that the gene product, they related on another have their own distinctive individual properties.

e.g.:- Alpha and Beta globin gene expressed at different stages in development from embryo to adult.

B. PSEUDOGENE:-

Genes they have lost their function containing non functional gene copies called Pseudogene.

Ussually these decayed genes containing one or more inactivating mutation, such as nonsense mutation, introduces a premature termination codon. A gene that has lost its function in this way is called Conventional Pseudogene.

A processed pseudogene is ont simply a mutated version of its parent gene, but instead in a DNA copy of the m-RNA.

=> A processed pseudogene therefore lacks introns, which would be removed from the m-RNA.

=> Also lack promotor region.

C. INTERONS AND EXONS:-

=> In molecular term, a gene can be define as a segment of DNA that is expressed to yield a functional product, which may be either RNA or a polypeptide chain.

However large a mounts of non coding DNA are also found between coding gene called Intron and the coding genes are called Exon.

=> The entire gene is transcribe to yield a long RNA molecule. The introne are removed by splicing machAnism after RNA synthesis.

=> Examples:- The human gene that incodes the blood clotting protein factor VIII, these genes spans approximatily 186kb of DNA and divided into 26 exons. After transcriptional processing we found that m-RNA is about 9kb long.

=>On averag introns are estimated to account for about 10 times more than exon.

=> Small nuclear RNAs (snRNAs) are critical component of spliceosome that splice introns out of pre-mRNA in the nucleus.

D. GENE CONTENT OF THE HUMAN GENOME:-

A. Transfer RNA genes :-

=> The classicl experiment estimate of the number of human t-RNA gene is 1.310 in the draft genome sequence we find only 497 human t-RNA genes.

=> More than 25% of t-RNA gene (140) are found in a region of only about 4mb on chromosome no.6. This small amount region only about 0.1 % of the genome.

=> More than half of the t-RNA gene (280 out of 497) reside on either chromosome 1 or 6.

=> Chromosome 3,4,8,9,10,12,18,20,21and x appear to have fewer than ~10 t-RNA genes each, and chromosome 22 and y have non at all .

B. RIBOSOMAL RNA GENES:-

Ribosome, the protein synthetic machine of the cell is made up of two sub unit.Both large sub unit found in chromosom13,14,15,21,22

2. EXTRAGENIC DNA:-

Approximately 30% of the human genomes is made up of genes and generelated sequence.

The remaining 70% of the genome over 200,0000 kb in total is extragenic DNA. About 80% of this fraction consist of unique or low copy number, sequence about which little can be said.

The rest about 400,000 kb of the genome as a whole is, made up of highly repetitive DNA.

A. DISPERSED REPETITIVE DNA:-

The current classification divides the dispersed repetitive DNA sequence in the human genom into two major categories :-

1.SINEs, 2.LINEs

1. SINEs (Short interspersed element):-

The major SINEs in mammals an genome are Alu sequence, so called because they usually contains a single site of or restiction endonuclease Alu.

The average length is about 280 bp and the are repeated some 700,000

to 1000,00 times in human genome.

2. LINEs:-

About 6000 bp long and repeated approcimately 60,000 – 100,000 times in the genome.

This is a type of transposon that can replicate and move around the genome by involving reverse transcription .

B. CLUSTERED REPETITIVE DNA :-

The human genome contains extensive tract up of repeat sequence arranged into long tandem arrays, this type of repetitive DNA is called satellite DNA.

1. Classical Satellite DNA:-

This was the first type of satellite DNA to be identified in the human genome and consist of clustered bettween 100 and 5000 kb in length.

2. Mini Satellite DNA :-

Mini Satellite DNA forms shorter cluster between 100 bp and 20 kb in length.

3. Micro Satellite DNA:-

5′ CACACACACACA 3′

3′GTGTGTGTGTGT 5′ 0.5% of the total genome

5′ AAAAAAA 3′

3′ TTTTTTTT 5′ 0.3% of the total genome

GENE MAPPING

1. GENETIC MAPPING :-

A. RESTICTION FRAGMENT LENGTH POLYMORTHISM:-

RFLP:- The difference in DNA sequence can result in variations of restiction sites and thus in the lingth of restiction fragment. An inherited difference in the pattern of restiction is known as RFLP.

RFLP is used as a marker in chromosomal mapping.

=> The presence and absece of restiction site a perticular chromosome provide a marker that can be used for gene mapping .

=> When a single base change occurs within a recognition site the sequence is no longer cut by restiction endonuclease, whereas an intract site on another chromosome can be cleaved.

=> DNA Fragment of different size are produced when the DNA is digested with these different site by restiction endonucleases for diagation.

METHOD:-

1. Large mol. Weight DNA is diagested with same restiction endonuclease.

2. Diagested DNA seperated in gel by electrophorsis.

3. Different lenght of DNA occurse

4. Seperated DNA are observed by hybridized by prob (rediotabelled).

5. Sequence is observed with the help of autorediogram.

B. SHORT TANDEM REPEAT POLYMORPHISM:-

STR:- These are about 100,000 block of the dinucleotide repeat CA/GT. These block contain 1- 40 repeating CA/GT units.

=> Amuch simple type of restiction unit exist in the form of dinucleotide (CA)n, trinucleotide (ATG)n and tetranucleotide(ATCG)n. These repeated unit are called short tndem repeats or microsatellites.

=> The CA repeat or microsatellites are used as genetic markes such variation ca n be easily scored by using DCR primer.

METHOD:-

1. The two DNA Fragment having CA repeated are used.
2. Both fragments repeat are different in lingth.
3. By using PCR (using primer) we amplify the both strand and allow to run in gel electrophoresis.
4. Most of variation can be scored by using PCR.
5. The band obtain in gel are differs.

PHYSICAL MAPPING

ASSEMBLING CONTIGE FROM BACs LIBRARIES :-

=>The method is based in the clonning, of DNA Fragment which can be going to mapping.

=>We used Bacterial artificial chromosome or yeast artificial chromosome for the clonning of DNA.

=> We take 100 – 200 kb DNA Fragment and inserting them in BACs.

=>The BACs were then positioned on individual chromosomes by looking for marker known as sequence tagged site (STS).

=> STS are short usually <500 bp.

=>Clone of the BACs were then broken into shortgunning.

=>Each fragment then sequence by PCR and amplified STS DNA then allow to electrophoresis.

1.WHATE IS HUMAN GENOME PROJECT ?

=>On octuber 1,1990 the human genome project (HGP) was officially launched in the under the department of energy and the national institutes of health.

=>Human genome project is a international effort whose principle goal were to sequence the entire human genome and the genome of several other model organism.Which provide platform lanching more detail reseach program into every genetic disorder.

⇒ The goals- To know most of the human gene,
to know location of the gene in chromosome,
to know sequence of the important genes,
to know function of the the genes.

=>the human genome refers to the international 13 year effort , formally begun in octuber 1990 and completed in aprile 2003,to discover all the estimated human gene and other model oranism genes and make them accessible for further biological study.

=> another project goal was to determine the complete sequence of 3 billion nitrogani bases.

CHROMOSOME WHICH ARE SEQUENCED

1. First chromosome sequenced (no. 22)- december 1999
2. Chromosome no. 5, 16,19 –aprile 2000
3. Chromosome no. 21-may 2000
4. Chromosome no. 20 december-2001
5. Chromosome no. 14 january - 2003

Post human genome process-

6 chromosome no. Y- july 2003
7 Chromosome no. 7- july 2003
8 Cromosome no. 6- octtuber 2004
9 Chromosome no. 13,19- march 2004
10 Chromosome no.9, 10- may 2004
11 Chromosome no. 5- septembre 2004
12 Chromosome no. 16- december 2004

13. Chromosome X –march 2005
14. Chromosome no. 2- aprile 2005
15. Chromosome no. 4- aprile 2005.

- Thankyou