FLOW OF GENETIC INFORMATION: GENOTYPE AND PHENOTYPE

Prof. Arjumand S. Warsy
Department of Biochemistry, College of Science
King Saud University, Riyadh

Organization of the Human Genome

- Diploid Human Genome: 6-7 billion base pairs (6-7 million kilo base pairs) of DNA organised into 23 pairs of chromosomes.
- No. of genes : 50,000 - 100,000.
- These genes are responsible for synthesis of proteins which are responsible for control of all aspects of embryogenesis, development, growth, reproduction and metabolism.
- The genes constitute only 10% of the total DNA (coding regions). 90% is involved in packing of DNA in chromosomes.

Nucleic Acids

Deoxyribonucleic acid (DNA) is present in the nucleus and mitochondria and is the carrier of genetic information.

Ribonucleic acid (RNA) is present in nucleus and cytoplasm and is involved in protein synthesis. There are 3 types of RNA.

mRNA	=	5% of total
rRNA	=	80% of total
tRNA	=	15% of total

DNA: The hereditary material

- Deoxyribonucleic acid (DNA) has been proven by various experiments to be the genetic material.
- It consists of two long strands of polynucleotides made up of a large number of nucleotides.
- Nucleotide: - Nitrogenous base:
  - Purine (A,G.)
  - Pyrimidine (C,T)
  - Sugar (deoxyribose)
  - Phosphate
- Structure of polynucleotide (Figure I)

- Structure of DNA Watson and Crick in 1953 proposed the following structure based on X-ray diffraction studies
  (Figure 2)

- DNA is made of 2 antiparallel, polynucleotide strands.
- The strands are twisted around each other forming the DNA helix (righthanded).
- Sugar-phosphate backbone is on the outside. Nitrogenous bases are on the inside.
- A is linked to T by 2 H-bands.
  G is linked to C by 3 H-bands.

<u>Replication</u>

- DNA multiplies prior to cell division.
- This process of multiplication is called replication.
- It is a semiconservative process i.e. the two DNA strands separate and each builds a new DNA. The two daughter DNA molecules each have the parent strand and a new strand.
- Replication starts at several points (replication units) on the DNA and finally the whole DNA is multiplied to two daughter DNA.

<u>DNA in Chromosomes</u>

- The DNA is tightly packed to form the chromatin which further condenses to form chromosomes during mitosis.
- Several order of DNA coiling occurs to form the chromosomes (Figure 3).

Types of DNA sequences in the Genome

- Unique or single copy DNA
    - Located throughout genome
    - ~ 75% of the total linear length of DNA
    - Includes 50,000 - 100,000 genes
    - < 10% of human genome are genes
- Repetitive DNA
    - ~ 15% of the proportion of genome
    - Repetitive sequences
    - Interspersed with genes and other single copy DNA
    - Two major families Alu and Li
    - 300,000 copies of a sequence of approx. 300 bp. these are Alu repeats as they contain Alu I restriction enzyme recognition site.
    - The Li family consists of approx. 100,000 copies of DNA sequences of upto 6000 bp each.
- Satellite DNA
    - 10% of the total DNA.
    - Highly repeated sequences - Short tendemly repeated sequences which code for rRNA and tRNA.
    - Highly localized in certain sites
        - Clustered around the centromere of certain chromosomes
        - In the telomeres
- Pseudogenes
    - DNA structures which resemble genes but are not expressed.
    - Due to mutations in the coding or regulatory sequences these have become silenced.
- Mitochondrial DNA (mt DNA)
    - Each mitochondria has its own DNA.

Gene Structure

- Each gene is a double stranded DNA portion dispersed throughout the

genome.

- Most human genes consist of coding sequence (exons) separated by non-coding sequences (introns) (Table 1).
- The number and size of introns in various genes in humans are extremely variable.
- Some introns are much longer than the coding sequences and some contain coding sequences for other genes.
- At 5' end of the gene are the promoter region consisting of `CAT' and `TATA' boxes. This also has the 'transcription initiation site'.
- At 3' end of the gene is the 'transcription termination site' proceeded by the polyadenylation site.

Regulation of Gene Expression

- Several control mechanisms exist to control the expression of genes.

- As shown by Jacob and Monad in 1961, structural genes have operator genes and repressor genes controlling their expression. Inducers increase the gene expression, while repressors decrease the gene expression.

- The promoter elements control gene expression. These are the GGGCGGG consensus sequences, TATA box (rich in AT) and CAT box.

- The Enhancers and Silencers increase and decrease gene expression, respectively.

Multigene Families

- Several copies of genes coding for the various rRNA are clustered as tandem arrays at the short arm of the 5 acrocentric chromosomes.

- tRNA gene subfamilies are located in numerous clusters interspersed throughout the human genouse.

Fundamentals of Gene Expression

- The central Dogma

	Transcription		Translation	
DNA	→	RNA	→	Protein

- Transcription: The process by which the genetic information is transmitted from DNA to mRNA.
  - Only one strand of DNA is transcribed. This is the `sense' or `coding' strand.
  - Synthesis of mRNA starts at the promoter site by RNA polymerase.
  - The introns and exons are transcribed and primary mRNA transcript (HnRNA) is formed.

- Post transcriptional process:
  - CAP is added at 5' end and poly A tail at 3' end.
  - Introns are spliced out.
  - The mature mRNA is transported to cytoplasm for translation.

- Translation: The process by which the information on mRNA is used to synthesise a specific protein is known as translation.

  - mRNA is translated to proteins on the ribosomes, by the action of a variety of tRNA's.
  - Each tRNA is specific for transport of an amino acid. Binding of amino acid to tRNA occurs by a process of activation which uses ATP.
  - The information for protein system is presented as the `genetic code'. The codons on mRNA are read by the anticodons on tRNA.
  - One amino acid is added at a time on the ribosomes move along the mRNA and joined together by peptide bond using by the enzume peptidyl transferase in the ribosomes.
  - At the stop codon (UAA, UAG or UGA), the protein synthesis stops and the polypeptide chain is released.
  - Polysomes: The structure in which several (up to 5) ribosomes are attached to RNA to make polypeptide's.
  - Post Translation Modifications.
  - Many proteins undergo modification after synthesis:

    - Folding of protein into a specific 3D structure.
    - Linking of two or more subunits to form the functional protein
    - Addition of carbohydrates
    - Formation of disulphide bond
    - Hydrolysis of peptide bond to make the active protein

All living cells have transcription and translation except RBC's which do not have a nucleus and ribosomes.

Table 1: Examples of genes

	Chromosomal location	Approx. gene size (b)	No. of exons
Alpha-globin	16	850	3
α1AT	14	1000	5
ß-globin	11	1600	3
Coagulation factor VIII	X	186000	26
CF transmembrane regulator protein	7	250000	24
Dystrophin	X	2400 000	80
G-6-PD	X	18000	13
HPGRT	X	44000	9
Inuslin	11	1430	3
Low density lipoprotein receptor	17	45000	18
Phenylalanine hydroxylase	12	90,000	13