Genetic Vision Team
فريـق الرؤيـة الجينية

Gene structure and function

فريق الرؤية الجينية
Genetic visions

Gene: sequence of DNA that give product, the product could be polypeptide or Non coding RNA.

If the polypeptide undergoes Specific modifications in endoplasmic reticulum (post translation modifications) and become functional Polypeptide we call it protein.

In 1970s a scientist Frederick sanger was created the first generation of Sequencing techniques " sanger sequencing"

Human genome Launch in 1990 and finish in 2003, It was token billions Dollars and 13 years.

Scientist successful to identifying number of Human genes and the locations of these genes.

Also at same time scientists worked with other model organism such as E. Coli, C. elegant, M. muscles and others .

Human contain approximately 20000 protein coding genes, also number of proteins non coding genes is 20000-25000 some organisms contain more than protein coding genes and these organisms are less complex than human.

Scientists try to explain this,

First scientists found that many genes have ability to give more than one product via Alternative splicing For example if gene X consist of 5 exons, via splicing exon number 1 and 2 could be interact to gather and give product, exon 3 and 5 could interact to gather and give another product, 1 and 4 and 5 could interact to gather and give another product and so...

Second ,proteins don't function by themselves, actually a complex Network of interactions between proteins and Non coding RNAs result a huge Diversity and highly organization and function.

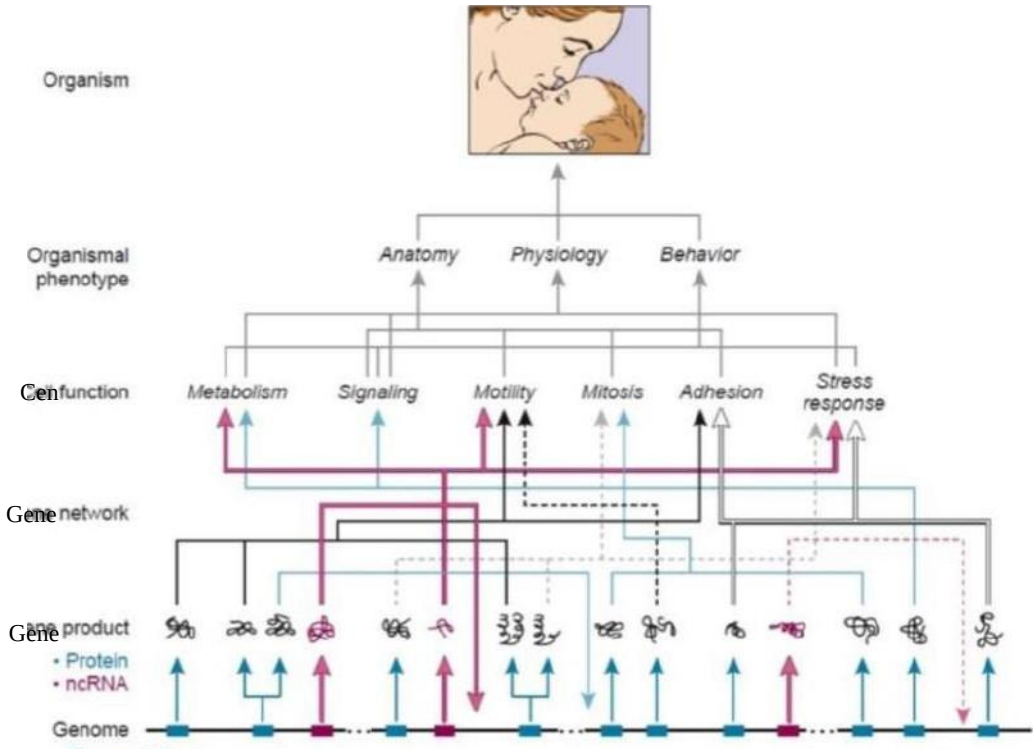
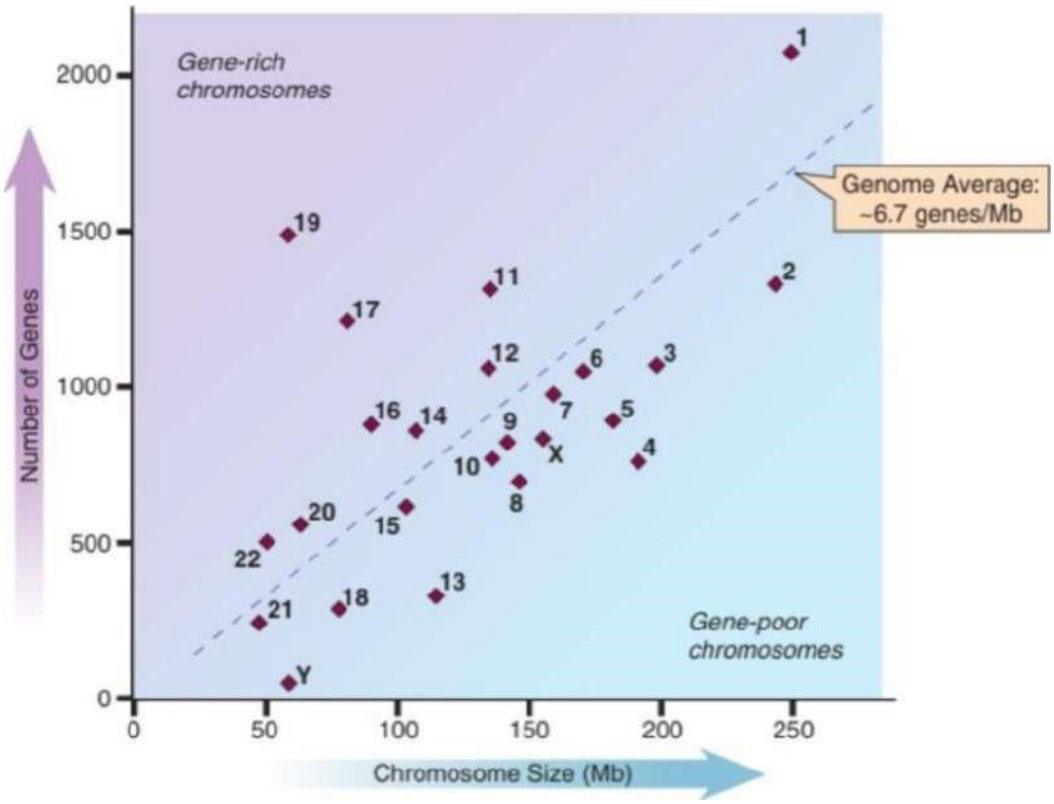


Figure 3-1 The amplification of genetic information from genome to gene products to gene net-

Figure The amplification of genetic information from genome to gene products to gene net. works and ultimately to cellular function and phenoty•e. The genome contains both protem-coding genes (blue) and noncoding RNA in.-RNA) genes (red). Many genes in the genome use alternative coding information to generate multiple different products. Both small and large ncRNAs participate In gene regulation. Many proteins participate in multigene networks that to cellular 5:22 AMsignals in a coordinated and combinatorial manner, thus further expanding the range of cellular functions that underlie organismal p}rnotypes.

As we said, scientists were identified the number of genes and the locations of all genes at different chromosomes

Scientists try to find a relationship between the number of genes and the size of chromosomes



From the figure , chromosomes are order according to size and number of genes

Chromosome number 1 is the largest chromosomes and the highest gene content (contains 2000 gene)

Chromosome number 19 is the second highest gene content and is not the second largest chromosome, which mean there is no Relationship between size of chromosome and gene content.

Some chromosomes are gene rich, or some regions on a chromosome are gene rich.

Some chromosomes are gene poor, even some region on a chromosome.

What is the classification criterion?

Scientist were calculated the Average of genes/ Mega

Base pair which approximately equal 6.7 gene /Mb

And draw the Dotted diagonal line, any chromosome above Diagonal line contain more than 6.7 gene /Mb considers gene rich chromosome, any chromosome below the diagonal line, considers gene poor chromosome.

In diagnosis, it is important where the mutation occur, at which Chromosome, is it a gene rich chromosome or not? even if the mutation in region of chromosome, is this region rich of genes or not?

Because the consequence dependent on region that mutated.

The only trisomy (extra chromosome- 47 rather than 46) Compatible of life are 13, 18, 21

Chromosomes 13, 18 and 21 are the lowest Autosomal gene content chromosomes.

Chromosome 21 is lowest size autosomal and more common between birth of other trisomy birth.

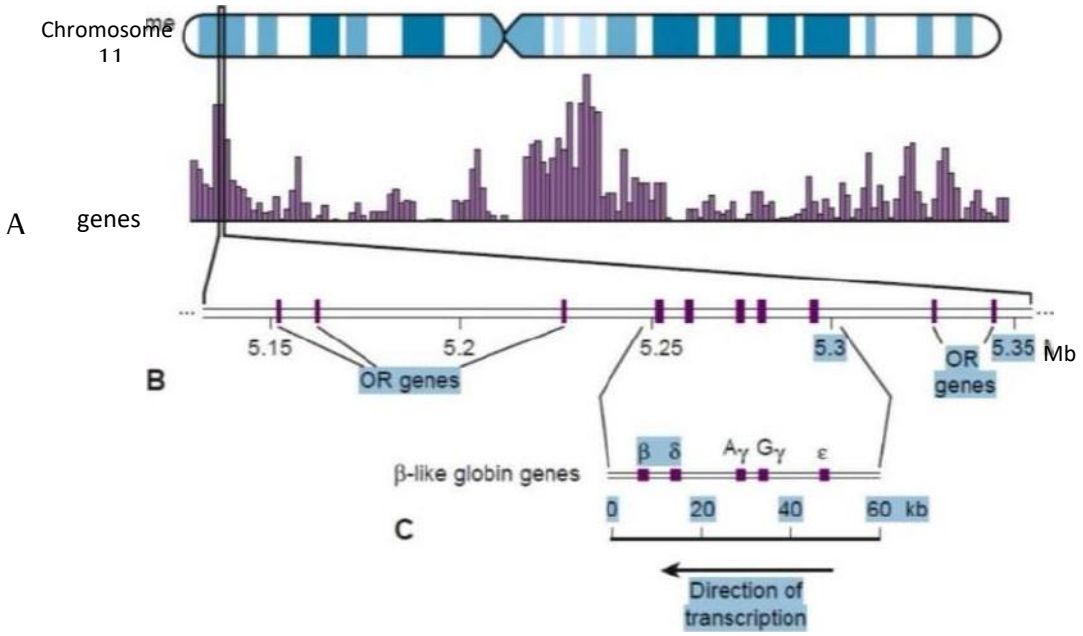
Gene density= Number of genes / Size of chromosome

The highest gene density chromosome is 19

gene density = slope

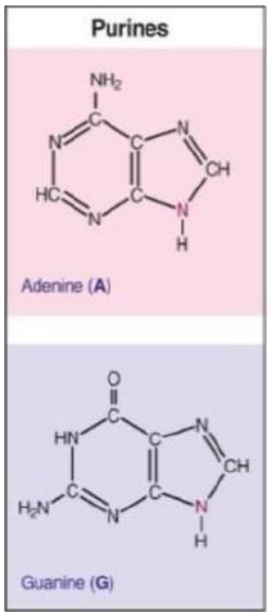
Slope = Y/X

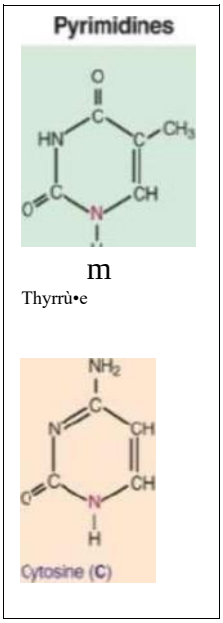
As we move from X axis to Y axis gene density increase



Look to this figure, we take a small region from short arm of Chromosomes 1 1, which contain 10 genes, 5 genes for olfactory Receptors and 5 genes for Beta like globin genes, this region is gene rich if you compare with other regions, we find that some regions are poor gene content,

We conclude the genes are distributed unequally along the chromosome.





m

Thyrrù•e

The Central Dogma of Molecular Biology

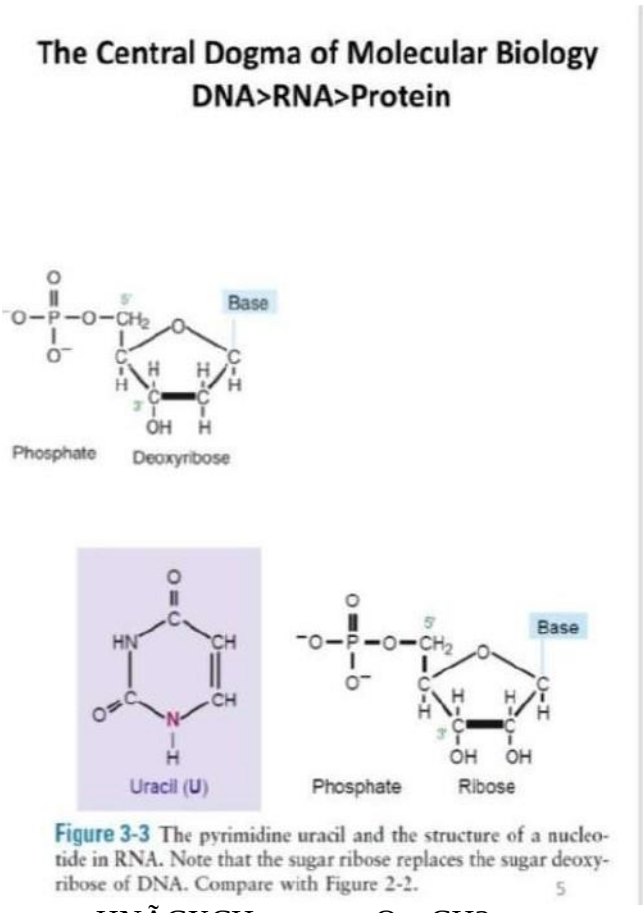


Figure 3-3 The pyrimidine uracil and the structure of a nucleotide in RNA. Note that the sugar ribose replaces the sugar deoxyribose of DNA. Compare with Figure 2-2.

5

HNÃCXCH -o—p —O—CH2

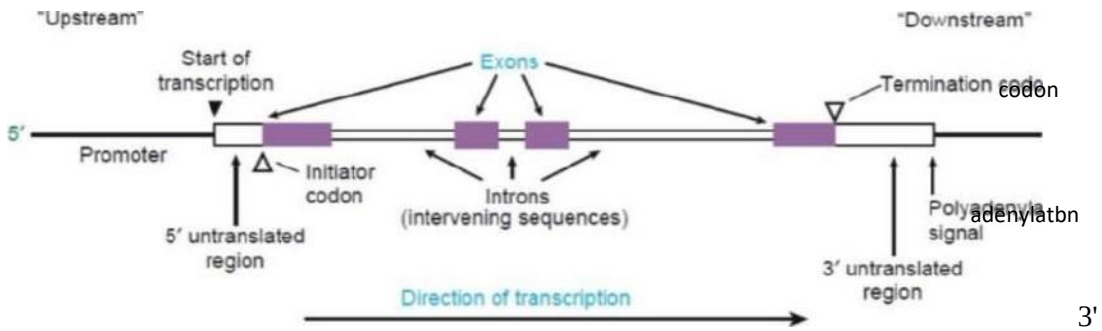
Uracil (U) Pt•osphate Ribose

Figure 3-3 The pyrimidine uracil and the structure of a nucleotide in RNA. Note that the sugar ribose replaces the sugar deoxy-ribose of DNA. Compare With figure 2-2.

5:58 AM

	DNA	RNA
Type of nucleotides	A, T, G, C	A, U, G, C
Type of sugar	Deoxyribose	Ribose
Number of strands	Double stranded	Single stranded

General structure of Typical human gene



At 5' end we have

- Promoter: a site where RNA polymerase and transcription factors bind.
- 5' UTR

At 3' end we have

- 3' UTR
- Polyadenylation signal (AAUAA), after 10-35 nucleotides the transcription will stop.

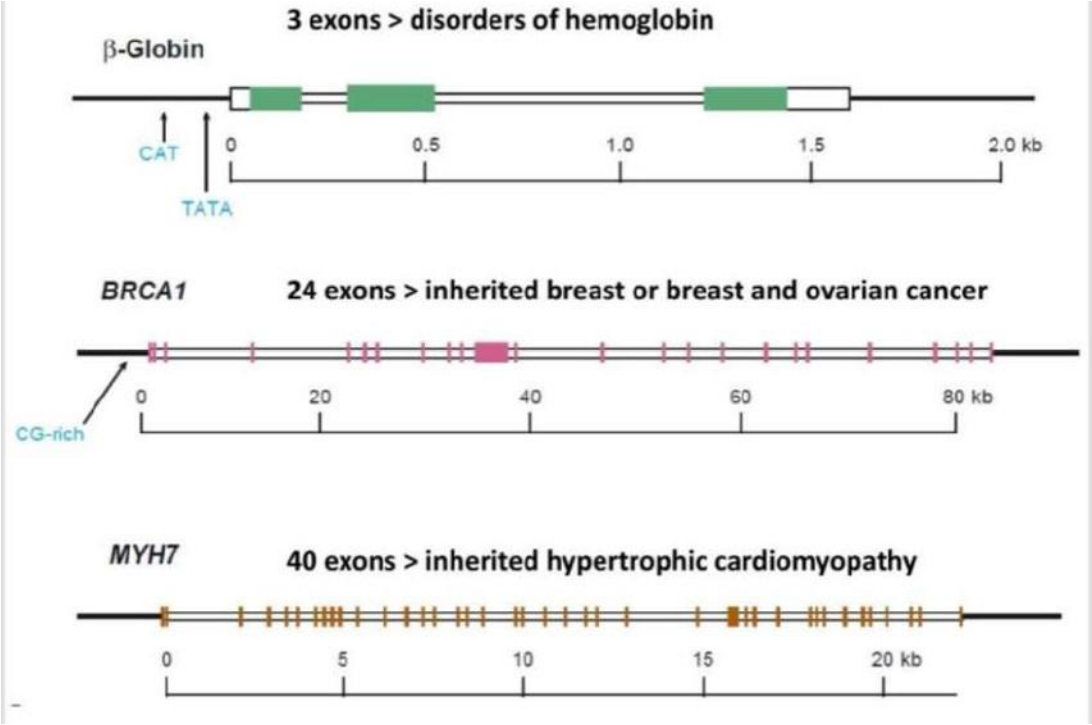
  Initiator codon(ATG): a site where translation will start.

  Termination codon (TAA, TAG, TGA): a site where the translation will have stopped.

  Transcription occur from 5' to 3'

  You should know the exons is not only the coding sequence, UTR regions consider

exons, so the exons are the sequences remain after RNA splicing.



From this figure, we conclude the size of gene is not related with number of exons

Pseudogenes

Structures resemble gene but Are not the genes, it might functional in the past but now not functional

There are two types of Pseudogenes

1. **Non processed Pseudogenes**

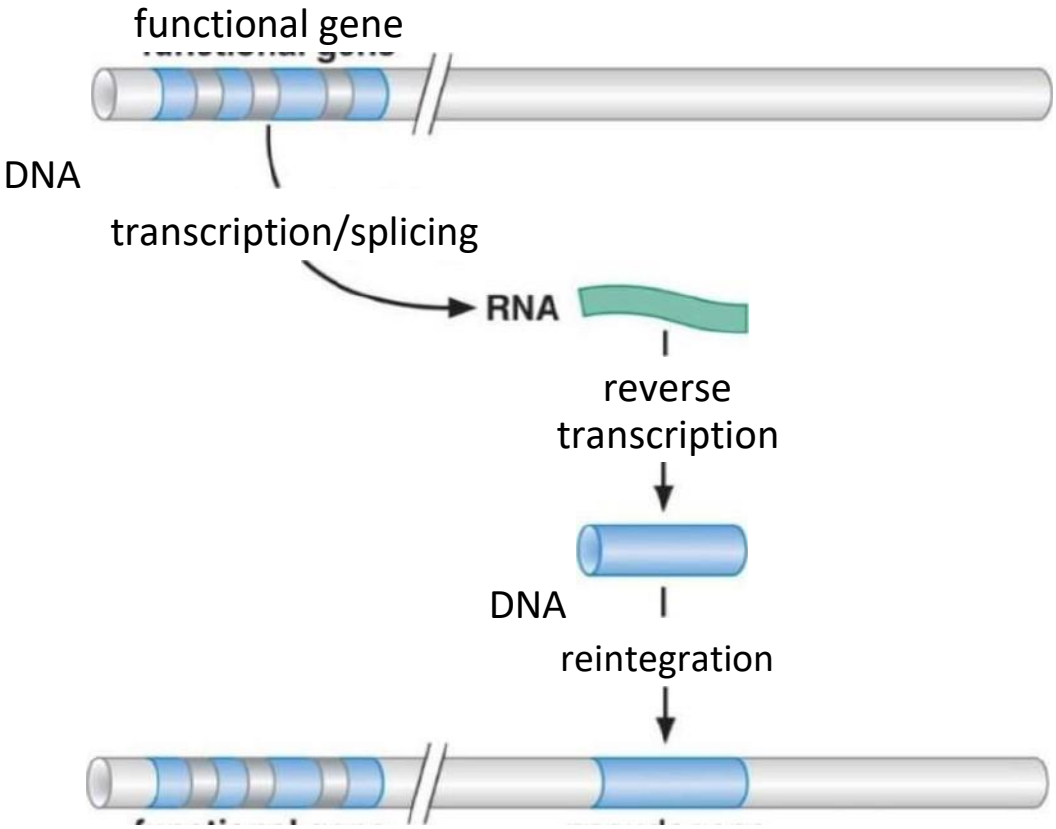
Accumulation of mutations in critical coding sites or regulatory sequences, non-processed word come from processing RNA which include remove of intervening sequencing (introns), no processed Pseudogenes contain introns and exons.

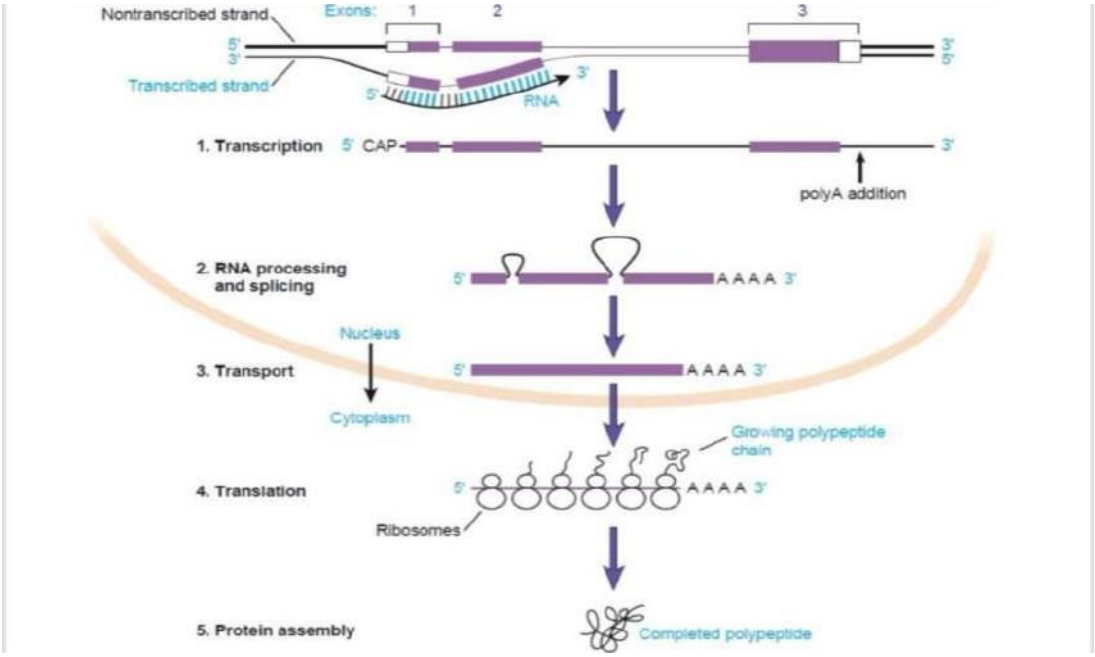
2. **Processed pseudogenes**

Result by retro-transposition

Retro-viruses Have Special Polymerase enzyme called reverse transcriptase, in the case of virus infection the virus enter the host cells, inside the cells there are m-RNA molecules, Reverse transcriptase use a m-RNA molecule as templet to synthesis CDNA, the problem here the mature m-RNA was lost regulator sequences, cDNA is lacked Regulatory elements that essential for gene expression so it become nonfunctional.

Processed pseudogenes arise from integration of reverse-transcribed RNAs





Over view in gene expression in eukaryotic cells

One of the double stranded DNA called Templet stand /non sense or non-coding is used to synthesis pre-mRNA, pre-mRNA undergo RNA splicing which include removing of introns, add 5' cap and poly A tail, and join exons to gather, now we have mature m-RNA, RNA processing occur inside nucleus, mature RNA migrate from nucleus to Cytoplasm, a site where translation occur, Ribosomes using information stored in mature m-RNA to synthesis polypeptide, after that polypeptide will transport to ER where the polypeptide undergo Specific modifications to become functional polypeptide we call it protein.

VIP note

AUG : initiation codon a site where translation start

UAA, **UGA** and **UAG** : terminal codon, site where translation will stop and poly-peptide cut.

gt→Splicing Donor, **ag**→splicing acceptor.

Both essential for splicing, a sites where the cutting occur, mutation in these sites will lead in proper splicing missing part of introns and cutting parts of exons might occur.

AAUAA: Polyadenylation signal

Enhancer : located at 5' end.

Cell differentiation

Multi cellular organism has different types of cells each is responsible for specific function, let's take example human, human has different types of cells, bone cells, heart cells, skin cells, eye cells, and so... As we know human organism originate from one cell called Zygote and this cell divide by mitosis to give

multi cellular human which contain billions of cells, if the cells originated from one cell by mitosis why the all cells are not identical to each other and have same function.

All somatic cells have the same Genetic material except lymphocytes (T and B cells)

So the heart cells and skin have the same genetic material, the difference is the genes that activated in Heart cells, and the genes that activated in skin cells.

At embryonic stages, certain modification occur which allow activate some genes and muted other genes.

Depend which genes are expressed and genes are muted, the function of the cell determined.

So not all genes express at same time in every cell.

Some genes are expressed in all cells; we call them housekeeping genes such as microtubules which is very important for cell division.

DNA which exist in our cells is not naked, it associated with proteins like histone and other proteins, these proteins are important to packing DNA for many purposes.

When the DNA is Packaged it has less accessibility for transcription, which mean the low or no gene expression, and it's a natural process to regulate gene expression.

Certain modification occur at histone tails could be change the chromatin structure and become less condense which allow gene expression.

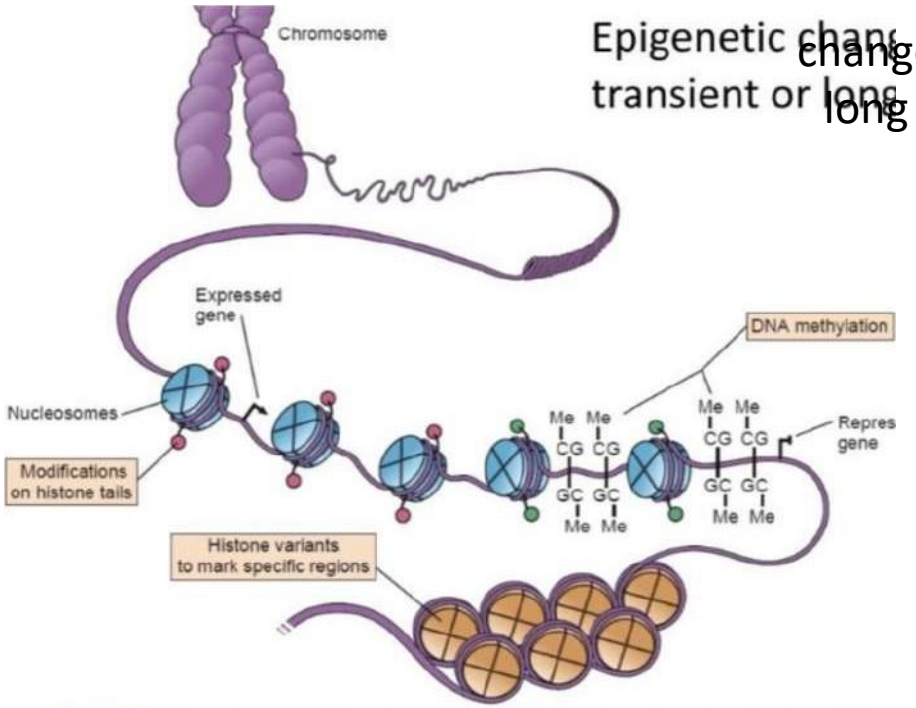


Figure 3-8 Schematic representation of chromatin and three major epigenetic mechanisms: DNA Repressed methylanon at CpG dtnucleotIdes, assocuted with gene repression; vanous modifications undicated by different colorsi on histone tails. with either gene expression or represston; ahåiWå*ous histone variants that mark specific regions of the genome. associated with specific functions required for chromosome stability or Fnornc integnty. Not to scale.

Epigenetics: reversible change in Chromatin structure not in DNA sequence.

Epigenetics change could be transient or long term.

DNA methylation: occur at C nucleotide, but not any C, C must be followed with G or precede with G (CpG) →GC or CG

Usually DNA methylation associated with stop gene expression (repress Gene expression)

Extensive de methylation occurs in in Germ cell development and early stage of embryonic development.

De methylation not occur directly; hydroxyl group will add to methyl group which to formation intermediate called hydroxymethyl Cytosine.

Epigenetics

1 - Histone modification: at N terminal end which include **methylation** (usually associated with Repress gene expression)

Acetylation (usually associated with induce gene expression)

Phosphorylation: usually associated with metabolism

2- Histone variants: product of other genes, and their amino acid distinct from histones

CENP-A similar to H3 and found in centromere.

3- Chromatin Architecture: each chromosome organizes in highly order structure inside nucleus and this arrangement Dynamic.

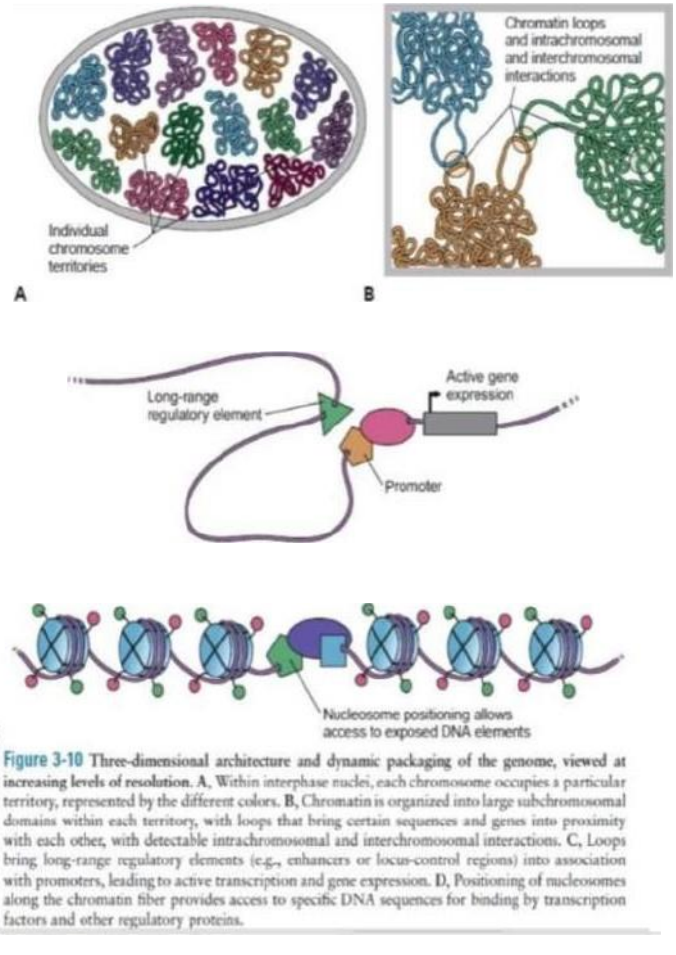


Figure 3-10 Three-dimensional architecture and dynamic packaging of the genome, viewed at increasing levels of resolution. A, Within interphase nuclei, each chromosome occupies a particular territory, represented by the different colors. B, Chromatin is organized into large subchromosomal domains within each territory, with loops that bring certain sequences and genes into proximity with each other, with detectable intrachromosomal and interchromosomal interactions. C, Loops bring long-range regulatory elements (e.g., enhancers or locus-control regions) into association with promoters, leading to active transcription and gene expression. D, Positioning of nucleosomes along the chromatin fiber provides access to specific DNA sequences for binding by transcription factors and other regulatory proteins.

Chromosomes arrangement are changed depend one many factors,

First: cell stage

In prophase chromosomes have looped structure, in metaphase chromosomes have X shape structure.

Second: interaction between genes

Many genes are interacting to each other to give certain trait, for example gene X and Gene Y

Gene X exist on chromosomes 11 and gene Y exist on chromosome 14, chromosomes 11 and 14 each of them change his structure and become approximate to each other to allow **inter chromosomal** interaction, which lead to interaction between gene X and Gene Y.

Also some times, at the **same chromosome** gene and Regulatory factor both exist, but far from each other, so chromosome start change it structure to allow gene and regulatory factor become approximate to each other (**intra chromosomal interactions**)

Chromosomes are organized in the 3D space of the nucleus to fulfill gene regulation.

Chromosomes are spatially segregating in A- and B type genomic compartments

- A genomic compartment

  (Chromatin=open chromatin) **active domains**

  - B genomic compartment

  Heterochromatin (repressive chromatin) **inactive domains**.

*Genomic compartment divide into sub compartments

- Nucleolus: it site contain hundreds of ribosomal genes of acrocentric chromosomes (13, 14, 15,21 ,22)
- Splicing speckles are nuclear domains enriched for splicing machinery.
- Promyelocytic leukemia (PML) bodies: are involved in cell cycle processes and DNA repair.

Topological associated domains

(TADS)

Cluster of genomic interaction within genomic compartments.

As we said previously, sometimes, gene and regulatory factors for this gene present at same chromosome but far from each other, TADs formation allow the gene and regulatory factors become approximately to each other to start gene expression.

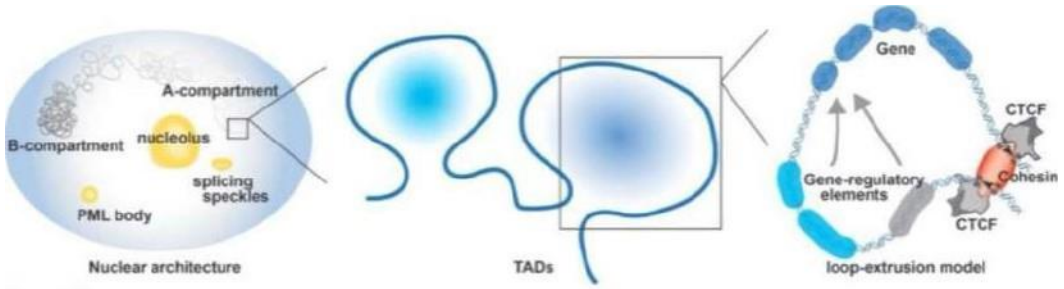


figure 3.11 Topol%ically associating domains(TADs).The is organized in the architecture of the nucleus, with A and B compartments representing open chromaun and hetenxhromaun, respectively. Further functional nuclear subdomains are the nucleolus, splicing spccklcs. and promyclocytic lcukrrnia (PML) bodies. Thr next organization lcvcl of chromatin involves TADs that are formed by rhe loop-exrrusion with CCC.TC-bindjng factor (CICH and cohesion 10 facilitate spatial proomity ofgene-regulatory

As we see in figure non coding DNA play important role in TAD formation, so if the mutation occurs at these sequence this affect the TAD formation

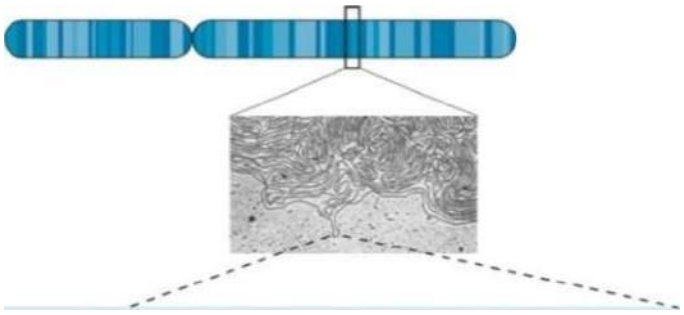
Reorganization of TAD architecture by chromosomal rearrangements can alter gene expression and may cause clinically apparent disease phenotypes.

Genomic deletions: can lead to TADs fusing

Duplications: may form neo-TADs

Inversions: reshuffle TADs

Translocations :could alter inter chromosomal contacts between non-homologous chromosomes



Double Helix Y...GGATTTCTAGGTAACTCAGTCGA... 3
T...CCTAAAGATCCATTGAGTCAGCT ... 5

Reference ...GGATTTCTAGGTAACTCAGTCGA...

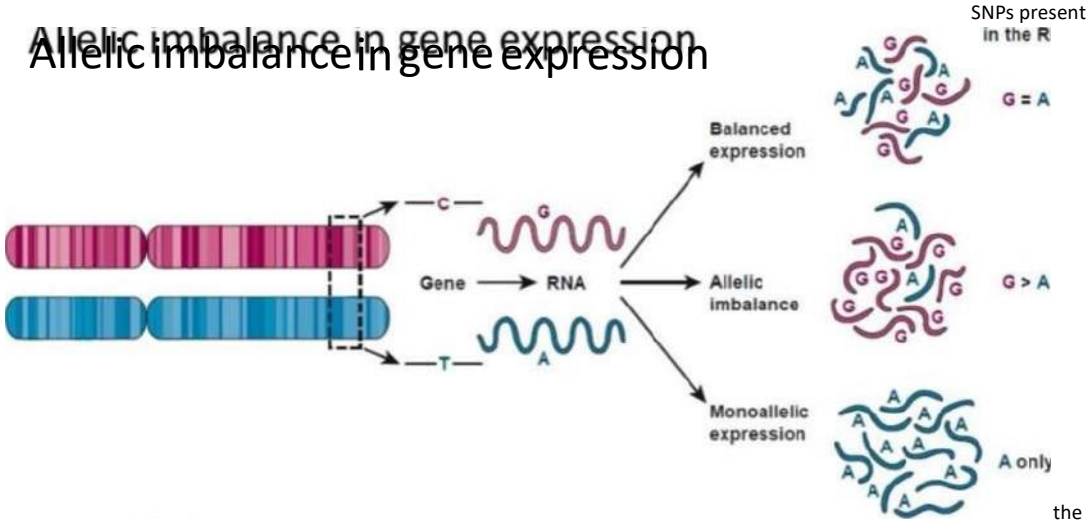
1 ...GGATTTCTAGGTAACTCAGTCGA...
Individual2...GGATTTC@AGGTAACTCAGTCGA...
Individual3...GGATTTC©AGGTAACTCAGTCGA...
Individual4...GGATTTCTAGGTAACTCAGTaGA...
Individual5...GGAT"CTAGGTAACTCAGTCGA...

Reference sequence: common sequence without changes in certain population.

Change in reference sequence lead to variations between plasma organisms.

Changing in reference sequence not necessary cause Disease.

Also individuals in certain population have sequence similar to reference sequence.



the RNA

Figure 3-11 Allelic expression patterns for a gene sequence with a transcri•ed DNA variant (here. a C or a to distinguish the alleles. As descrlbed tn the text, the relative abundance of RNA transcripts from the two alleles (here. carrying a G or an A) demonstrates whether the gene shows balanced expression (top). allelic imbalance (center). or exclusively monoallclic expression (bottom). Different underlying mechanisms for allelic imbalance are compared in Table 3-2. SNR Single nucleotide polymorphism.

I-Most of genes show balanced (bi allelic) expression, two allele give equal amount of m-RNA

2- 5-20 % of our genes show imbalance gene expression, different in regulatory elements give advantage to other allele for expansion more than the other (prefer one allele) it necessary abnormal

3-mono allelic only one allele express not both, and this could be occurring normal (lymphocytes) or by mutations.

TABLE 3-2 Allelic Imbalance in Gene Expression

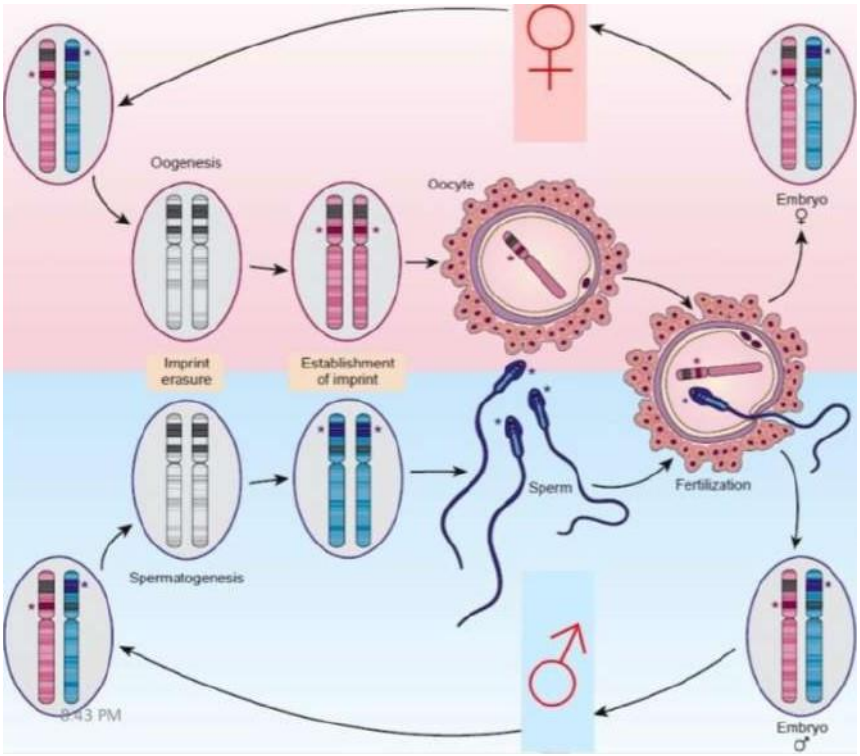
TABLE 3-2 Allelic Imbalance in Gene Expression

Type	Characteristics	Genes Affected	Basis	Developmental Origin
Unbalanced expression	Unequal RNA abundance from two alleles due to DNA variants and associated epigenetic changes; usually < twofold difference in expression	5%-20% of autosomal genes	Sequence variants cause different levels of expression at the two alleles	Early embryogenesis
Monoallelic expression				
• Somatic rearrangement	Changes in DNA organization to produce functional gene at one allele, but not other	Immunoglobulin genes, T-cell receptor genes	Random choice of one allele	B- and T-cell lineages
• Random allelic silencing or activation	Expression from only one allele at a locus, due to differential epigenetic packaging at locus	Olfactory receptor genes in sensory neurons; other chemosensory or immune system genes; up to 10% of all genes in other cell types	Random choice of one allele	Specific cell types
• Genomic imprinting	Epigenetic silencing of allele(s) in imprinted region	>100 genes with functions in development	Imprinted region marked epigenetically according to parent of origin	Parental germline
• X chromosome inactivation	Epigenetic silencing of alleles on one X chromosome in females	Most X-linked genes in females	Random choice of one X chromosome	Early embryogenesis

Genomic imprinting

Some genes only express at paternal chromosome and some genes only express at maternal chromosome

Inside our cells, for each gene we have two copies, one at paternal chromosome and the other cope on maternal chromosome, if gene X only express at paternal chromosome the other copy on maternal chromosome will be Epigenetically modified to become inactive this called **genomic imprinting**



In our cells we have two chromosomes, paternal and maternal, and each of them has imprinted genes

For example, male, in gametogenesis chromosomes inside Germ line cells become paternal chromosomes, which mean the location of imprinting will change for the chromosomes that inherited from the mother, and both chromosomes now consider paternal chromosomes.

Same thing for females, during gametogenesis the chromosome that was inherited from the father become maternal and the locations of imprinting will change.

X inactivation

Scientists were found that genes on X chromosome are related with intellectual ability, female has two X chromosomes where the male has one X.

To make balance between males and females, one of the X chromosomes randomly inactivated

During first weak (1000 cell stage) one of the two X chromosomes randomly inactivated, and once the cell decided one of X chromosome inactivation, all the Descendant cells have the same choice.

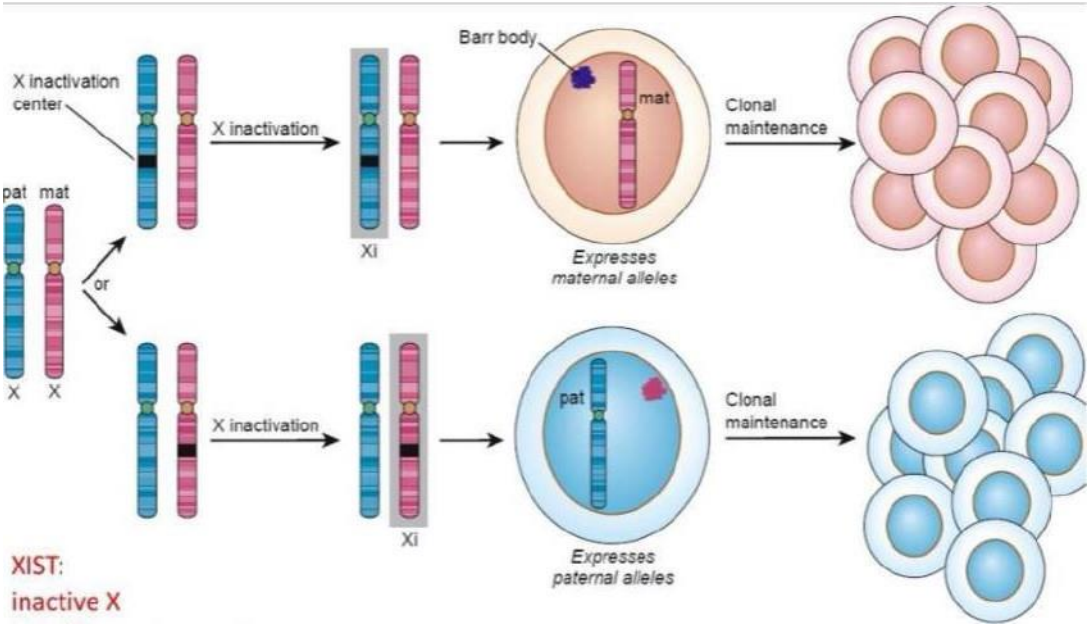
The Probability is 50% to 50%, but not necessary half the cells have Xl and other half have, it similar when you drop coin the Probability to get head is 50% and for tail 50%, but we could drop a coin two times and get head and head.

Xi= inactivated X chromosome

Not all genes in Xi are inactivated, some genes still active at regions called pseudo autosomal, also present at Y chromosome.

XSIT gene: present at Xi, which produce non coding RNA cover the majority of Xi genes which lead to stop gene expression.

Xi also called bar body



XIST:
inactive X

specific Figure 3-13 Random X chromosome inactivation early in female development. Shortly after transcripts conception of a female embrvo, borh rhe paternally and maternally inherited X chromosomes (par and mat, respectively) are active. Within rhe first week of embryogenesis. one or the other X is IncRNAl chosen at random to become the future inactive X, through a series of events involving the X inactivation center /black box)- That X hen becomesthe inactive X (Xi. indicated by the shading) in that cell and its pr%eny and forms the Barr body in interphase nuclei. The resulting female embryo is Thus a clonal mosaic of rwo epigenetically determined cell t)pes: one expresses alleles from the maternal X ipmk cells'. whereas other expresses alleles from the paternal X (blue

8:43 PM cells). The ratio of the two cell types is determined randomly but varies among normal females 26 and among females whoarc carriers of X-Iinked disease alleles isec Chapters 6 and 7K