Organization of the Human Genome

Biochemistry

Hayder A. Giha

Organization of the human genome

- The term Human Genome stand for human genetic information found in all somatic nucleated cells in a form of DNA (deoxyribonucleic acids).
- The DNA contains the instructions (information) needed for growth, development and maintenance of cellular functions
- During development and growth the DNA need to be copied and passed to the new generations of daughter cells with utmost accuracy, that ensure the preservation of both individual organism and species.
- The cell provides the machinery for the reproduction of the DNA with fidelity

Gene and gene expression

- The gene is a specific sequence of nucleotides in the DNA, which code for specific protein or certain ribonucleic acids (tRNA, rRNA & small RNA).
- Although the genomic DNA sequence (genetic blueprint) is identical in all somatic cells of the same individual, proteins within different cells are different in the same individual e.g. the liver cells are different from bone cells. This is explained by what is known as differential expression of the genes.
- The DNA is transcribed into mRNA and the mRNA is translated into protein, what is known as gene expression (central dogma), the latter is subject to regulation.
- Differential expression: In each tissue certain genes are expressed while others suppressed based on the tissue function.

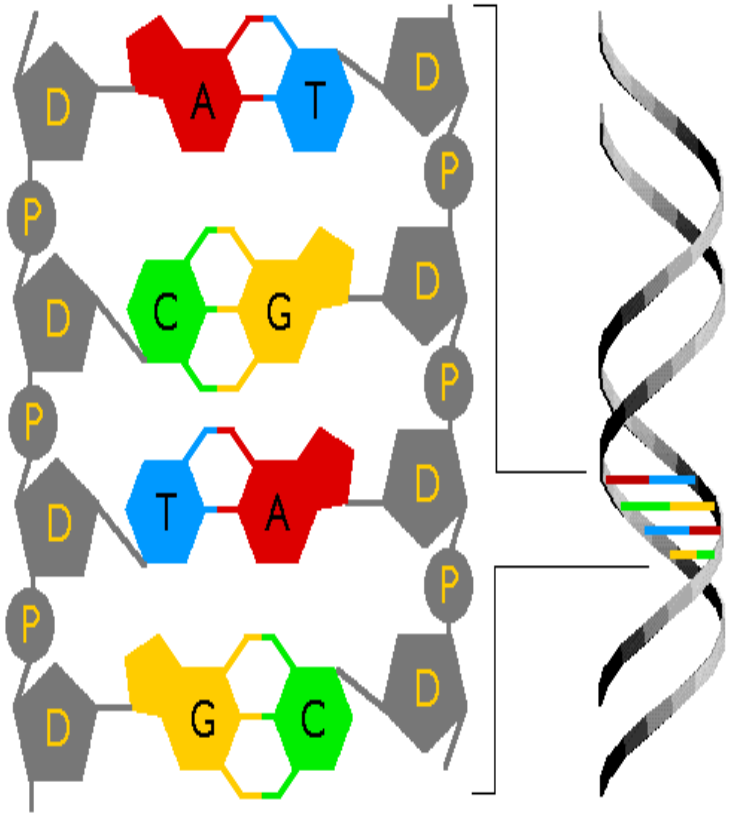
Human genome

Physical organization

- The human genome is found in two different cell compartments, the bulk in the nucleus (20.000 to 25.000 genes, may be up to100.000 genes) and much lesser in the mitochondria (37 genes).
- The nuclear DNA is inherited from both father (paternal) and mother (maternal) while the mitochondrial DNA comes only from the mother (maternal).
- The mitochondrial DNA is used only in the mitochondria while the nuclear DNA needed for all cell functions including the mitochondria.
- The mitochondrial DNA is circular in structure while the nuclear DNA is linear

DNA structure

- The DNA is composed of 2 strands
- Each strand have backbone formed of alternating pentose sugar and P linked by phosphodiester bonds, to the sugar attached the bases (A, G, C & T) by glyosidic bonds.
- The 2 strands are complementary and are linked to each other by H bonds (A with T linked by 2 H bonds and C with G by 3 H bonds).
- The 2 strands are complementary the sequence of nucleotides in one strands determines the sequence of the other strand (base pairing rule).

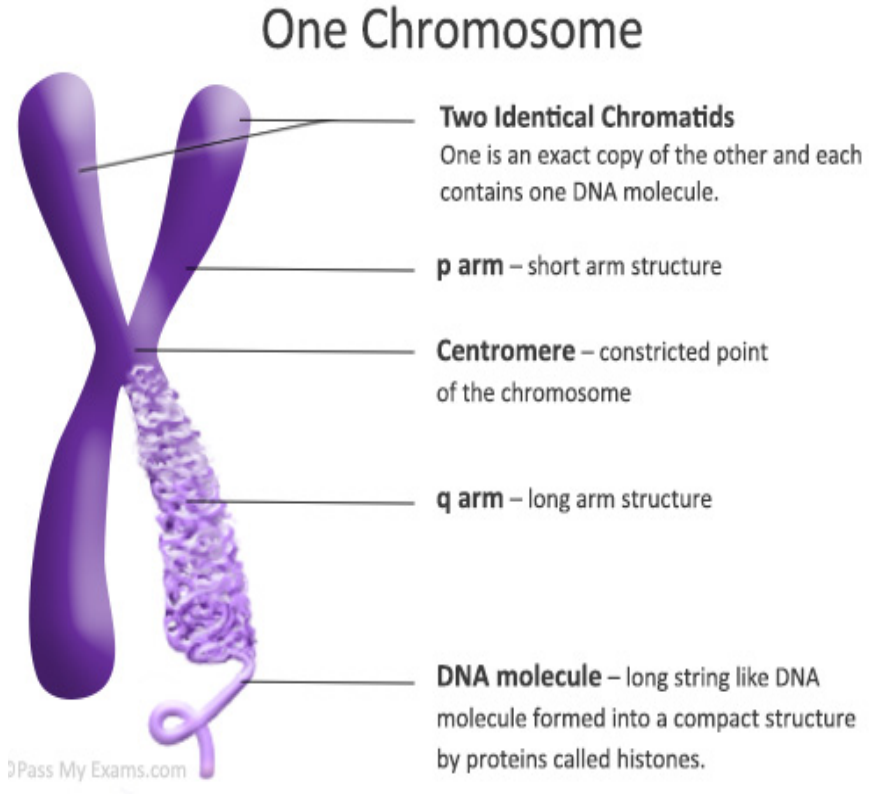


Chromatin

- The chromatin is composed of very long DNA molecules with equal mass of histone proteins and small amount of RNA and non-histone proteins (mostly enzymes and proteins need for replication and transcription).
- The histones are heterogeneous group of small basic proteins rich in the positively charged arginine and lysine amino acids (constituting ¼ of the amino acids of the histone) which help in binding the histones to the negatively charged backbone of the DNA.
- Of the functions of the histones is compaction of the DNA in the nucleus.

DNA packaging and chromosomes

- The chromosome is the condensed chromatin material that appears in the metaphase of the cell cycle
- Each chromosome is composed 2 identical chromatids linked at the centromere
- Each chromatid contain one condensed DNA molecule. The DNA length at the metaphase is about 1.4 μm, while DNA linear length is about 50.000 times it is condensed length.
- Note: the human cell nucleus diameter is 6 μm

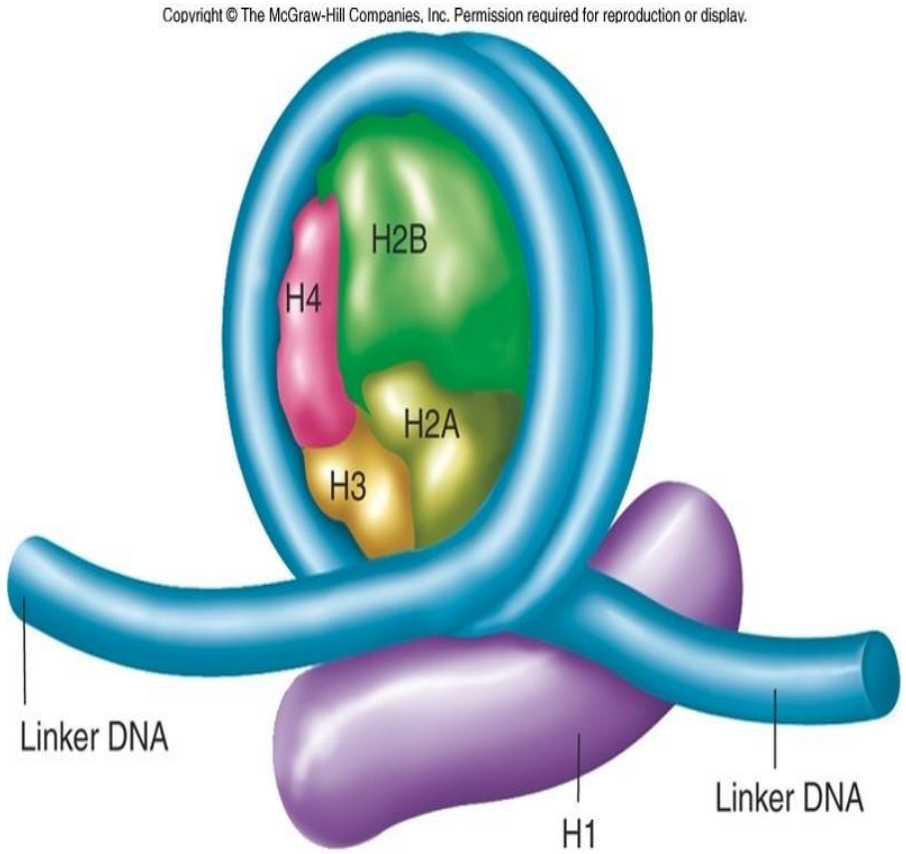


DNA packaging - nucleosomes

- There are at least four levels for DNA packaging in the chromosome:

1. The nucleosome: is basic organization or level of packaging upon which other levels of packaging are built.

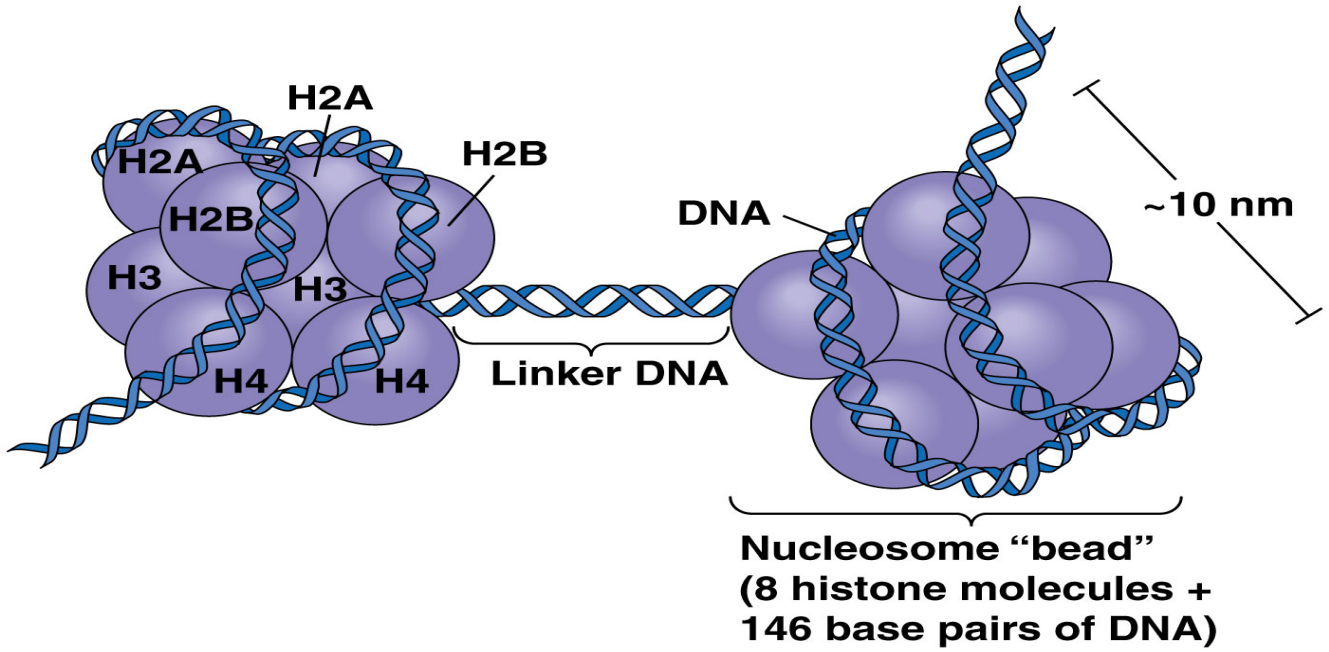
a. Each nucleosome have a core composed of 8 histone protein units, TWO molecules of each type of 4 histones (H): H2A, H2B, H3 & H4, surrounded by a DNA double strand of 146 base pair (bp) in length forming 2 turns around the core histone octamer.



DNA packaging - nucleosomes

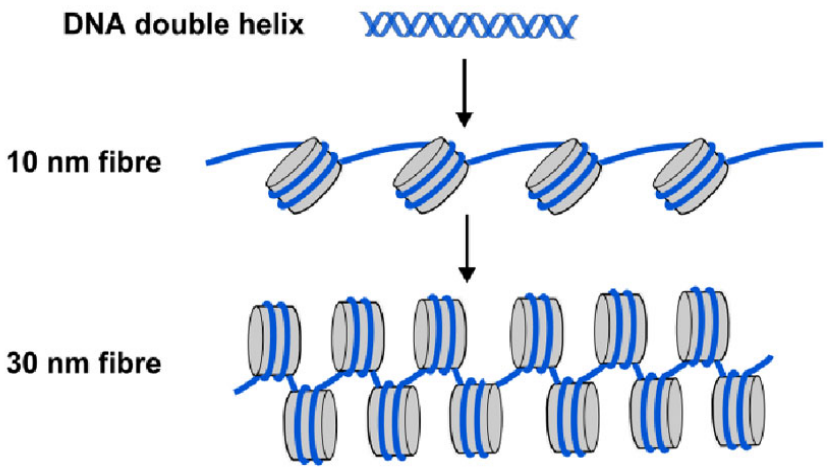
b. The nucleosomes are linked with each other by a DNA linker of 50-70 bp bound to linker histone (H1).

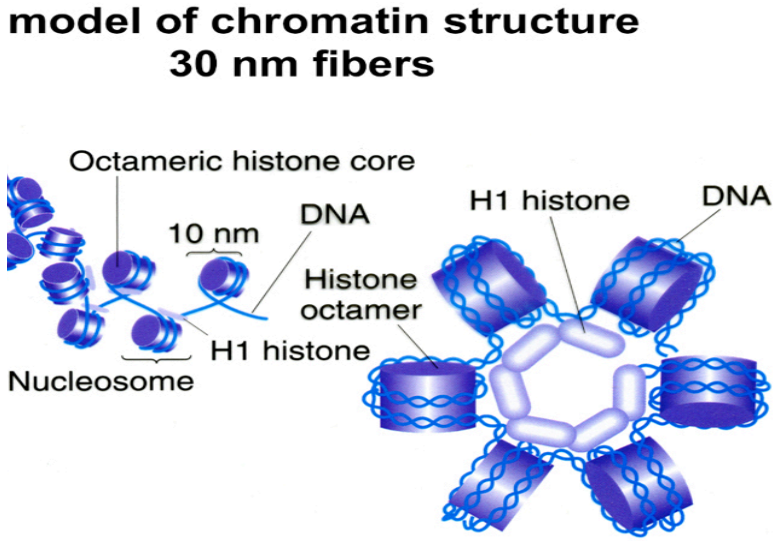
c. In addition to the DNA packaging function the nucleosome are involved in regulation of the gene expression by making the genes in the DNA accessible or not to the enzymes and factors of transcription.



DNA packaging –10nm & 30nm chromatin fibrils

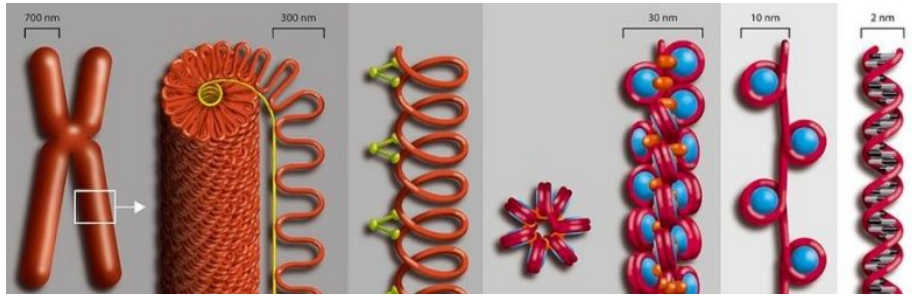
2. 10-nm chromatin fiber (fibril): is the 2nd order of DNA packaging in which the nucleosomes arranged in beads-on-string to form the 10-nm chromatin fibrils
3. 30-nm chromatin fibril: is the 3rd order, where the 10-nm fibrils wounded around a central axis to form the 30-nm fibrils

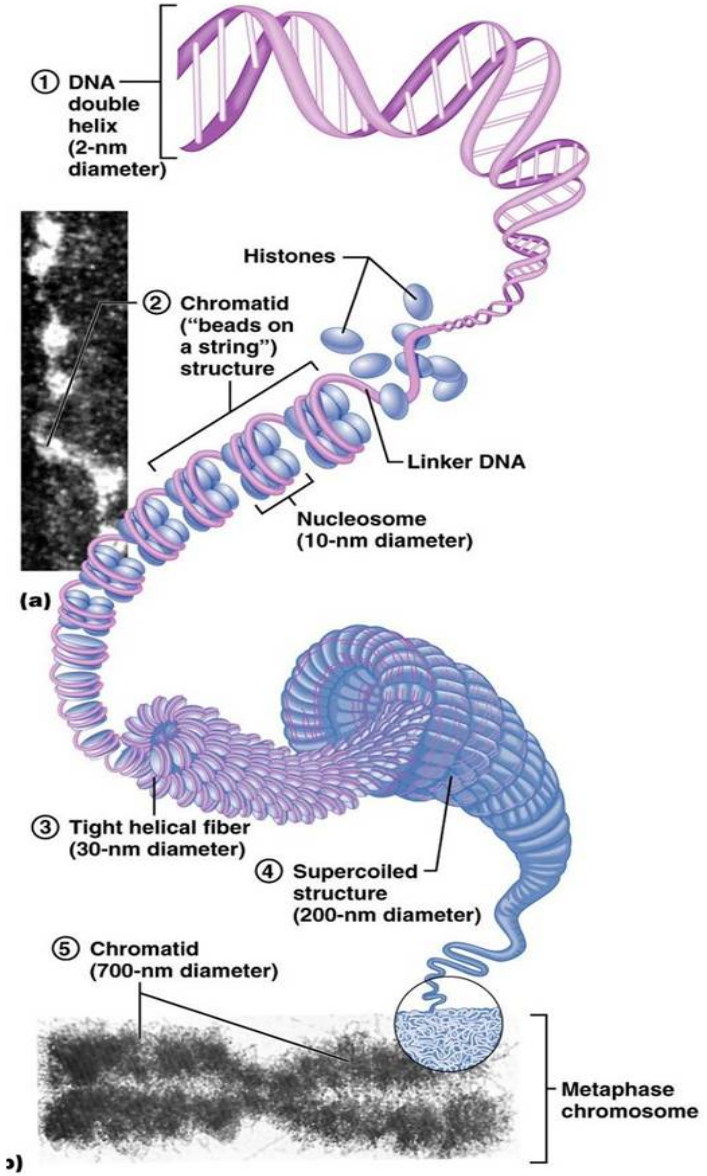




DNA packaging – Chromatin supercoil & chromosomes

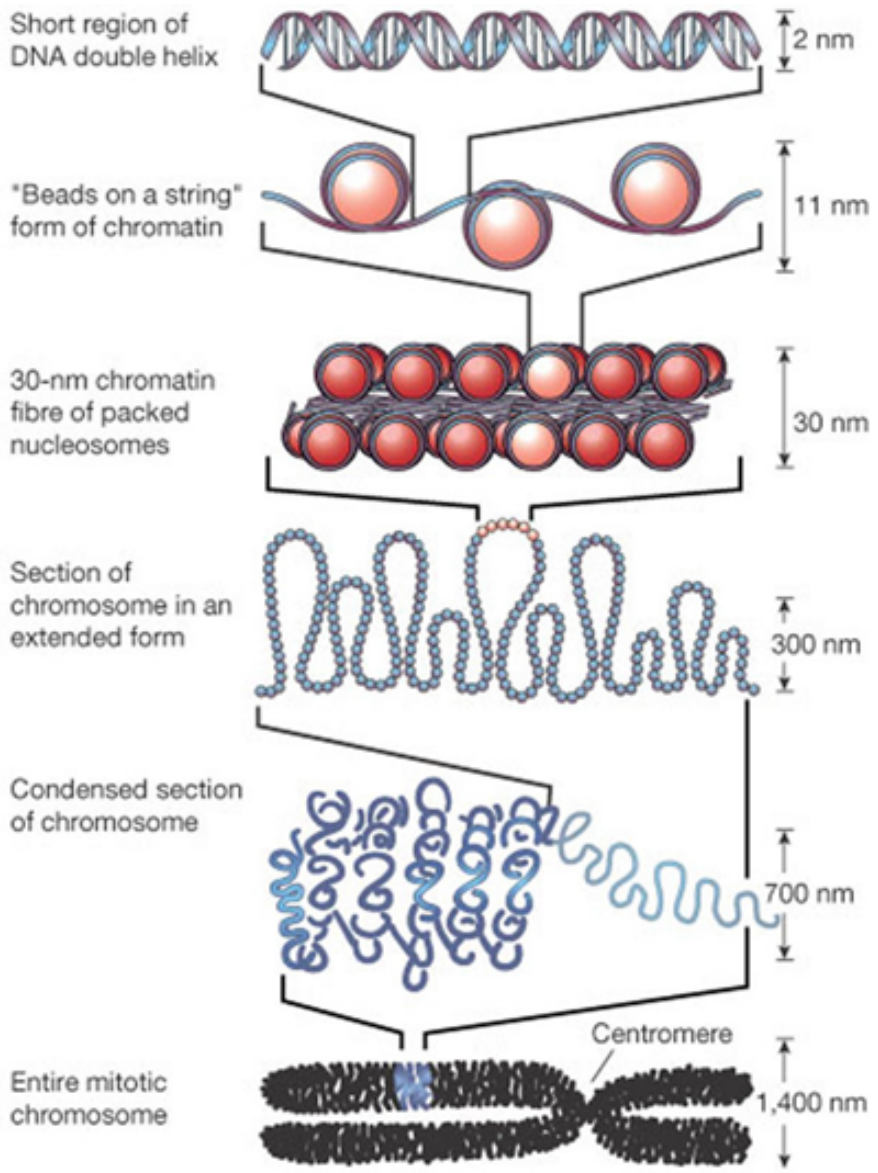
4. Chromatin supercoil: is the 4th order, the 30-nm chromatin fibrils is super-coiled around itself to form the higher order of compaction, that is further folded to form the chromosomes





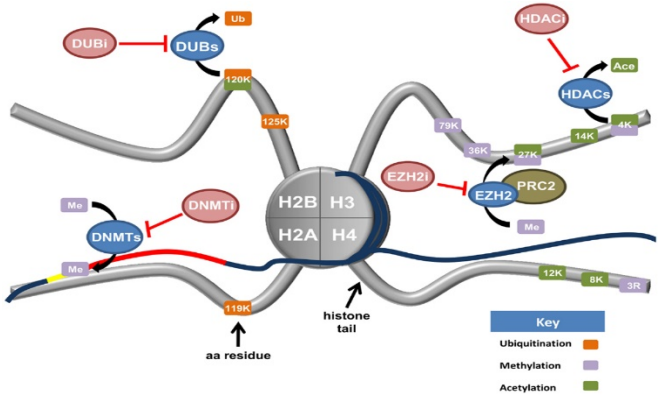
DNA packaging – Chromatin supercoil & chromosomes

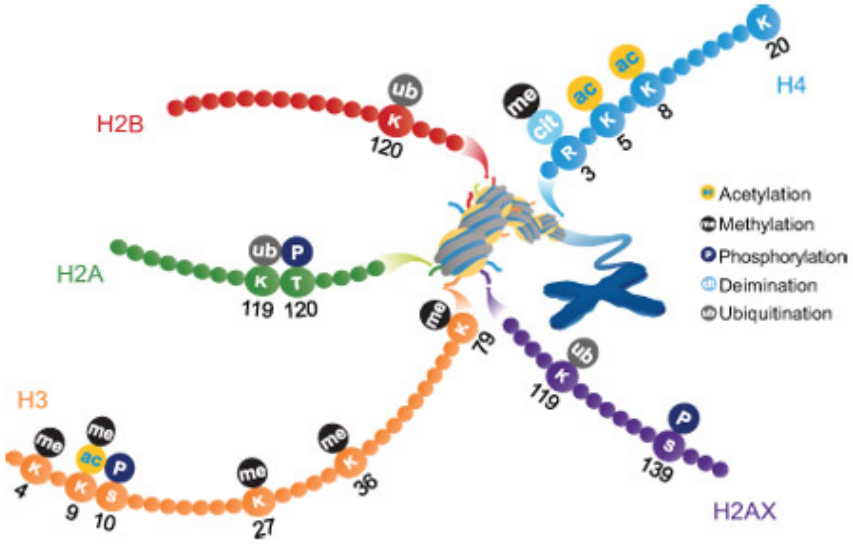
- Chromosomes are found in extended form (300-nm) then the condensed form (700-nm). Finally the paired metaphase chromosomes (1400-nm) occur during Metphase phase of the cell cycle



Histones modifications

- Each of the core histones (H2A, H2B, H3, H4), have a. structured domain and b. unstructured amino-terminal tail of 25-40 amino acid residues.
- The amino-terminal tails are subject to enzymatic modification, e.g. acetylation, methylation, phosphorylation and etc., leading to changes in the histones electric charge and shape, which are reversible modifications.
- They prepare the chromatin DNA for replication and transcription





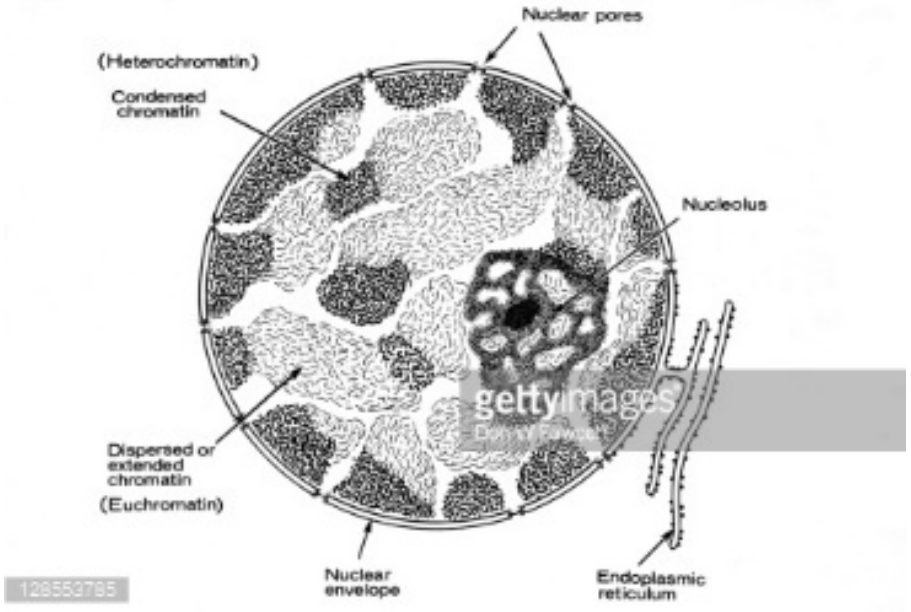
Histones modifications

i. Acetylation and deacetylation of the lysine residues:

- Acetylation of the lysine residues of histones (catalyzed by histone acetyl transferase [HAT] weakens the DNA-histones interaction making the DNA more accessible to the transcription factors (proteins need for transcription), leads to transcription activation.
- On the other hand the histone deacetylation (catalyzed by histone deacetylase [HDAC]) is associated with inactivation of gene expression (gene silencing)

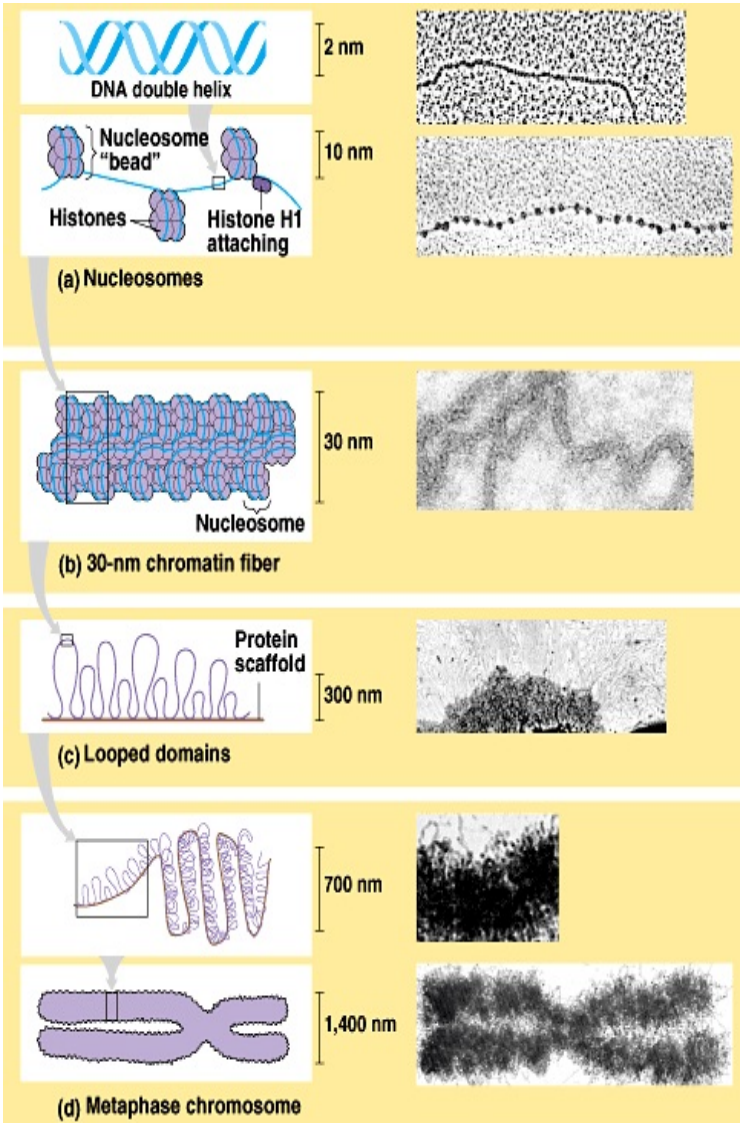
Euchromatin and heterochromatin

a. Densely packed region of chromatin is described as heterochromatin (highly compacted DNA with histone), mostly indicates inactive chromatin i.e. transcription is inhibited
b. Less densely packed region of chromatin is described as euchromatin (loose association between DNA and histones), mostly indicates transcriptionally active chromatin, since the genes in the DNA are accessible to the transcription factors and enzymes.



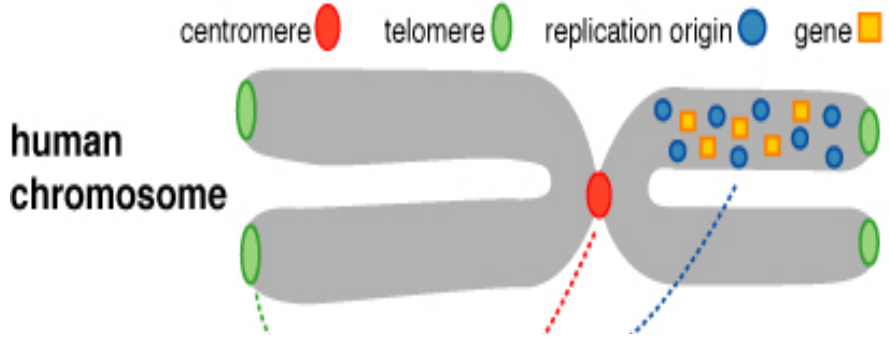
Chromosome structure

- Each chromosome have 2 identical chromatids. Each chromatid is composed of one very long linear duplex of DNA complexes non-covalently with histone proteins
- Chromosome structure varies with the phases of the cell cycle, from loose threadlike appearance in G1 phase to tightly compacted typical chromosome seen during M phase (metaphase).
- The chromosomes as individual units have 3 types of sequences



Chromosomes have 3 types of sequences

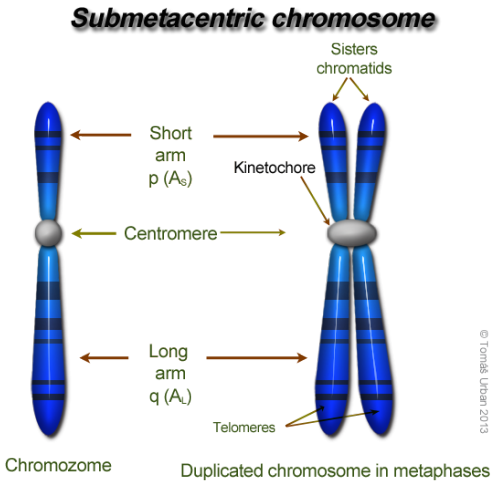
a. Telomeres: are hexameric DNA repeats of TTAGGG, found at the ends of chromosomes protecting the chromosomes from degradation
b. Centromeres: are sequence elements serve as handle for attachment with mitotic spindles during cell division, in a process leading to separation of the two sister chromatids of the metaphase chromosomes.

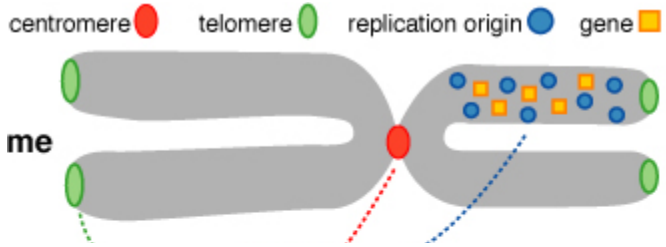


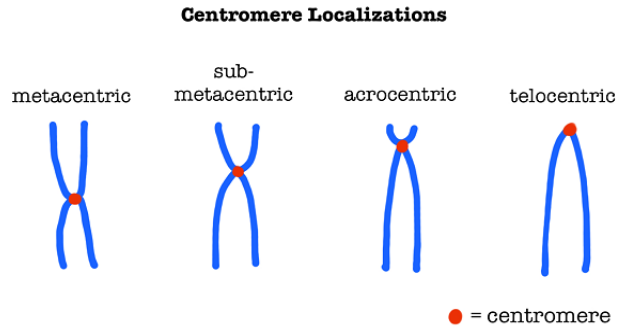
Chromosomes have 3 types of sequences

Centromeres also act as a boundary that separate the 2 arms of the chromosomes, the short p arms (p from petit means small) and long q arms (q follow p in alphabet). The centromeres locations differ between the different chromosome types

c. Origin for replication sequence: several specific sequences dispersed throughout the DNA molecule act as markers for the site for the origin of replication



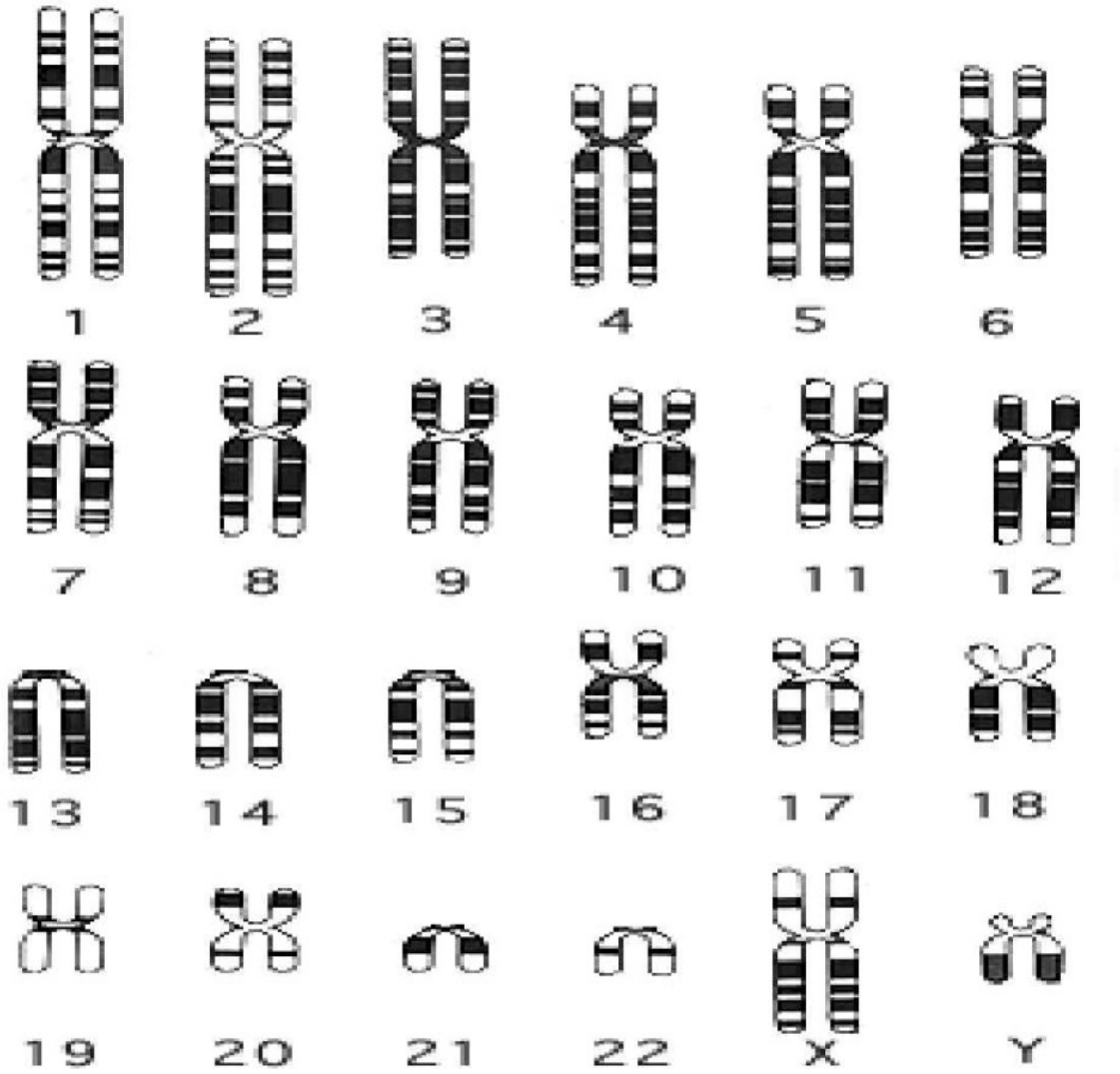


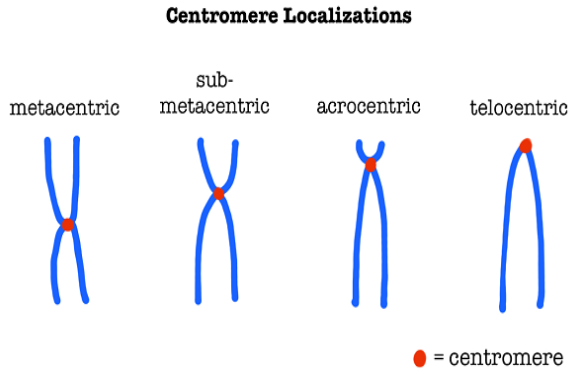


Information organization

- The number of copies of same chromosome in one cell is known as ploidy.
- In human, most of the somatic cells are diploid i.e. each nucleus have 2 copies for each chromosome (known as homologous chromosomes), one inherited from the father and the other from the mother.
- The germ cells (ova and sperm) each contain only one copy of the of each chromosome, so, are described as haploid
- The haploid genome of each human cell consist of 23 chromosomes (22 somatic and 1 sex), containing 3.0×10^9 bp, this DNA material is enough to form 1.5 million gene, however, so far the identified genes between 20.000 to 25.000, other texts mentioned 100.000 genes.

Human chromosomes (haploid)





Human proteome

- The human proteome, indicates the total protein expressed from the human genome (all human genes).
- Previously, it was mentioned that in human, one gene code for one protein, but in fact some genes each can produce more than one protein, by what is known as alternative splicing.

Human genome – DNA

- Eukaryotic DNA can be classified into unique sequence DNA (single copy) and repetitive sequence DNA.

A. Unique DNA sequence: Are single copy genes, each encode for specific protein. There are 20.000 to 25.000 genes in the human genome, grouped into 4 categories. Approx. 5.000 genes involved in the genome maintenance, 5.000 in signal transduction, 4.000 in biochemical functions and 9.000 involved in other activities

B. Repeated sequences: Make up approx. 50% of the human genome. Don't not encode proteins. Have no direct function, but may be important for chromosome structure and dynamics. Grouped into two main classes a. Satellite DNA, b. LINES & SINES

DNA sequence repeats

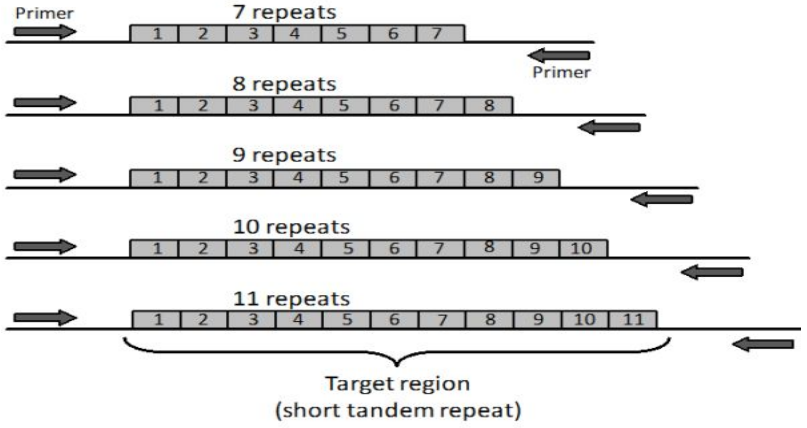
a. Satellite DNA: Are highly repetitive sequences found clustered together in many tandem repeats, found in 1 to 10 million copies per haloid genome. Not expressed into proteins. Also found in the centromeres and telomeres of the chromosomes. Are categorized into subgroups according to the number of bp in each repeat, into:

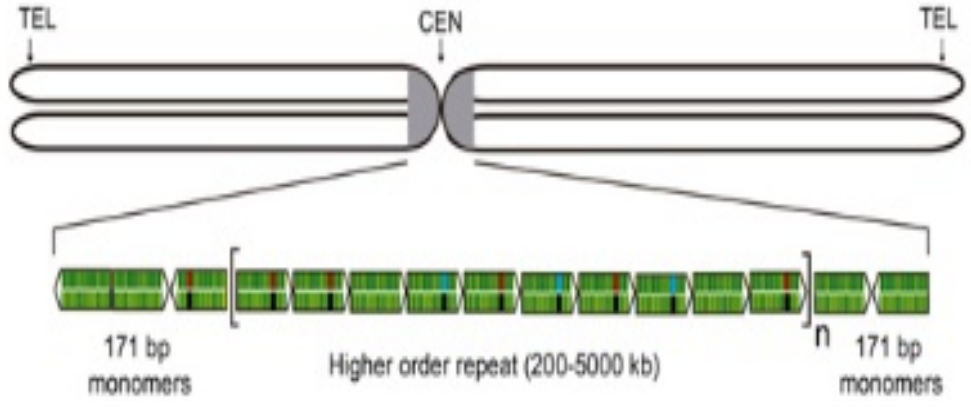
i. Alpha satellite: 171 bp repeats extend for several million bp in length

ii. Minisatellite: 20-70 bp repeats extend for few thousand bp in length

iii. Microsatellite: 2- 4 bp repeats extend to few hundreds bp in length

- Trinucleotides repeats: are microsatellites sequence repeats that are normally found in certain genes, expansion of above normal number can result in human disease.



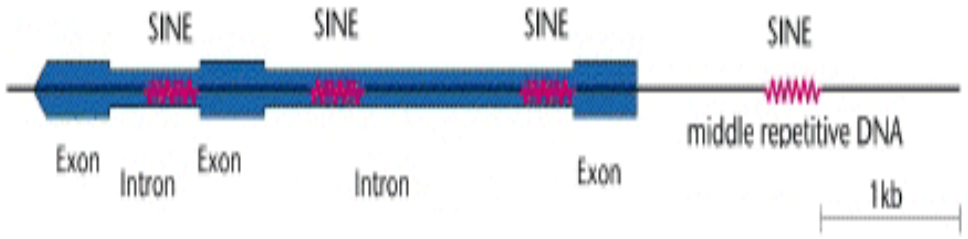


DNA sequence repeats

a. LINES and SINES: Are un-clustered sequences found interspersed between unique sequences. Found at < 1 million copy per haploid genome. Transcribed into RNA. Can be sub-grouped according to their size into:

i. LINES (Long interspersed elements) 7.000 bp (20 – 50.000 copies)

ii. SINES (Short interspersed element) 90 – 500 bp (about 100.000 copies)



Functional organization

- Cell functions are encoded by genes (DNA), found in nucleus and mitochondria. However, not all genes are expressed in all tissues, what is known as differential gene expression i.e. there is a tissue-specific gene expression

a. Gene: is a sequence region in the DNA that generate functional product like RNA and / or protein. It include proper encoding sequence (transcribed sequence), promotors region, regulatory region (DNA elements).

- Only 2% of the human genome encodes for proteins, these genes expands randomly between the non-coding DNA

The End