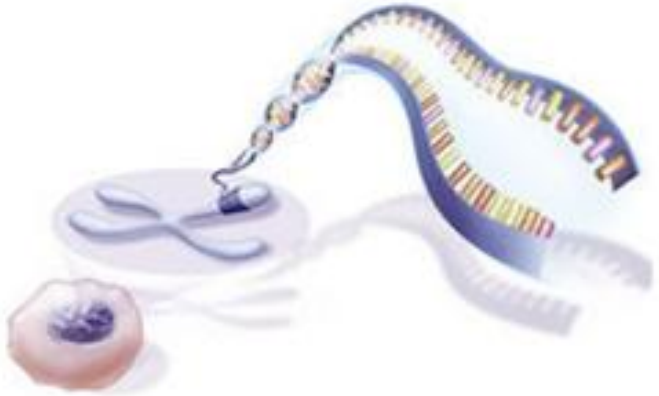


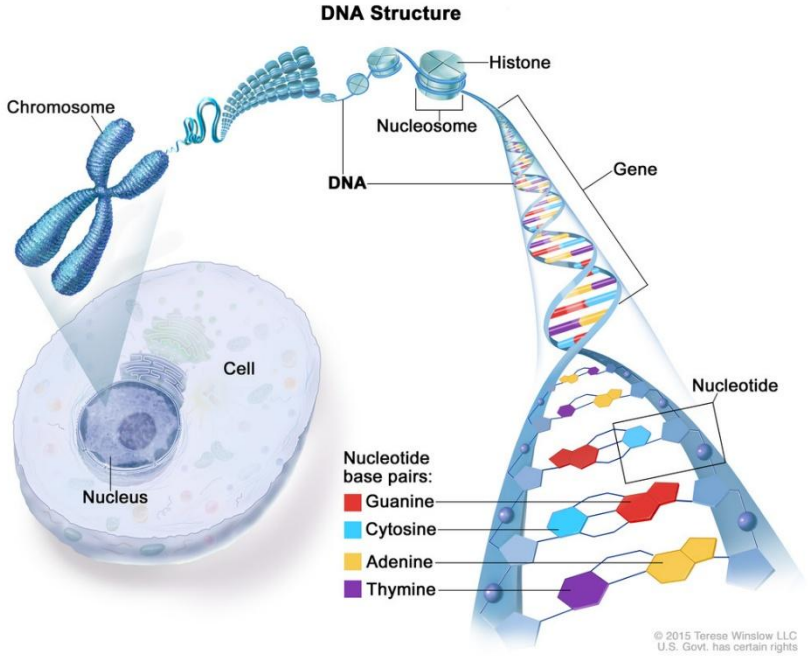
Genome structure, DNA replication, DNA repair, DNA transcription

Monika Ołdak

The structure of the human genome refers to the organization of genetic information in human cells

and includes:

the nuclear genome and the mitochondrial genome



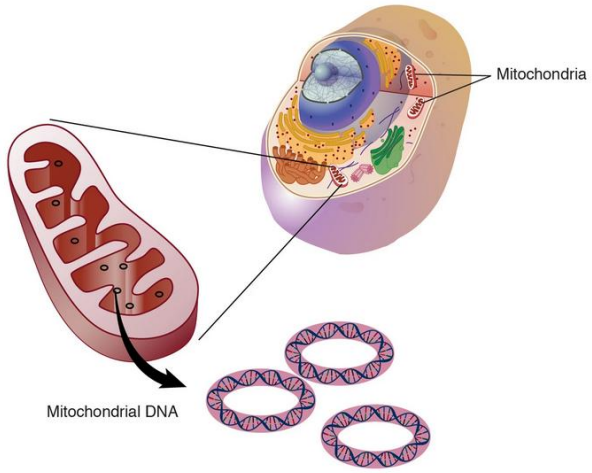
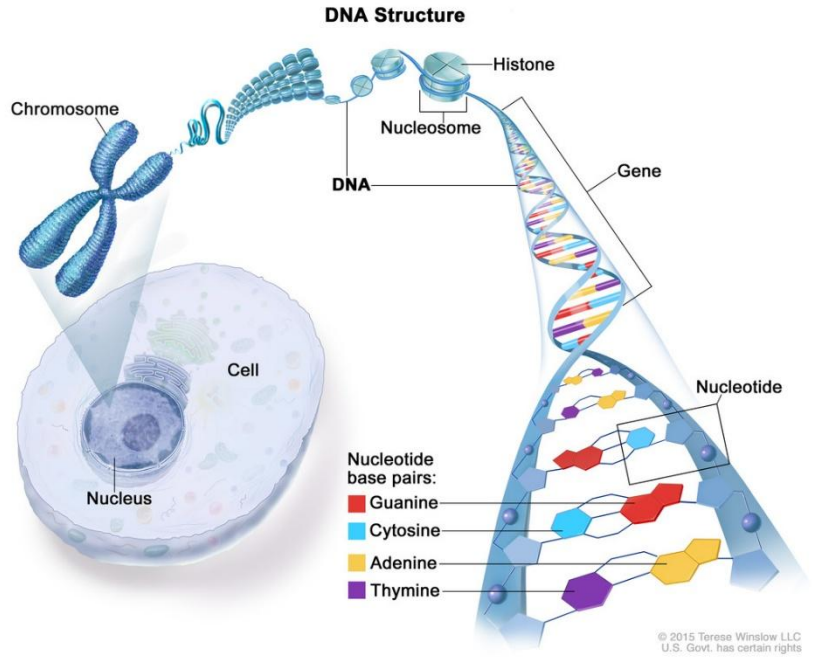


Image by National Human Genome Research Institute (NHGRI)

The structure of the human genome refers to the organization of genetic information in human cells and includes:

the nuclear genome and the mitochondrial genome



<u>Nuclear genome</u>

- Approximately 3.2 billion DNA base pairs
- Organized into 23 pairs of chromosomes (22 autosomes + sex chromosomes)
- Approximately 20,000–22,000 protein-coding genes
- Contains numerous non-coding sequences, including:
  - regulatory regions (promoters, enhancers, repressor sequences)
  - introns, pseudogenes
  - tandem and scattered repeats (LINEs, SINEs, satellites)

The structure of the human genome refers to the organization of genetic information in human cells and includes:

the nuclear genome and **the mitochondrial genome**

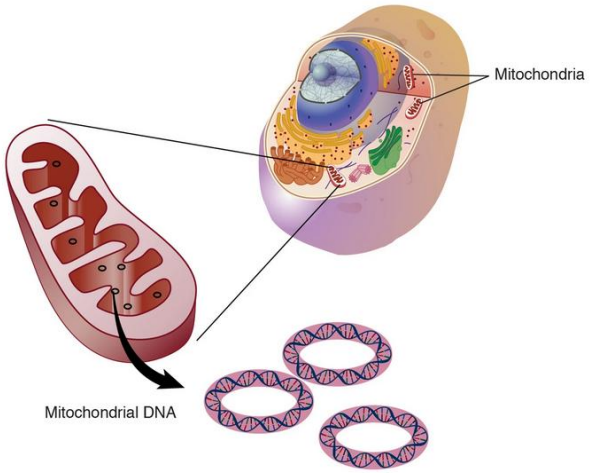


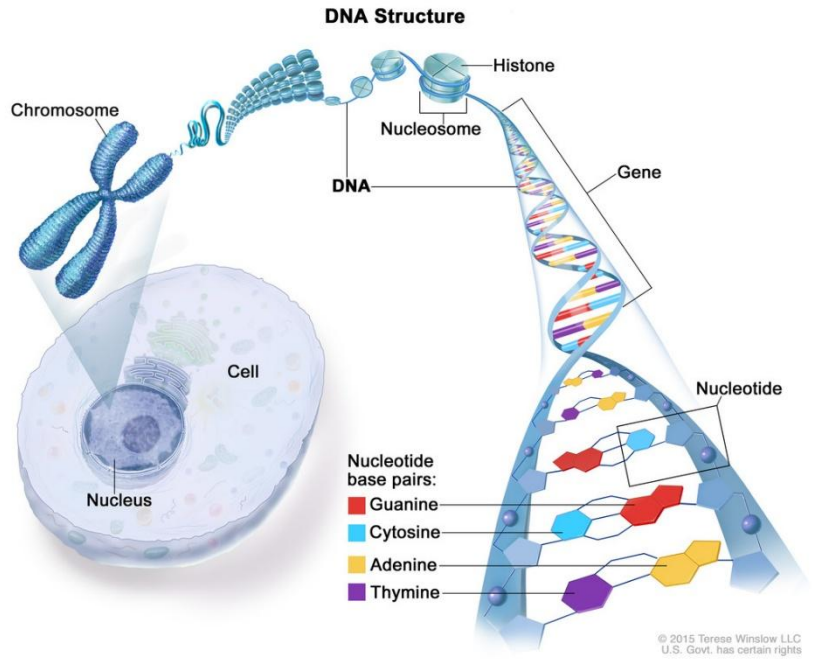
Image by National Human Genome Research Institute (NHGRI)

<u>Mitochondrial genome</u>

- A circular DNA molecule ~16,600 base pairs long
- Approximately 37 genes:
  - 13 respiratory chain protein genes
  - 22 tRNA genes
  - 2 rRNA genes
- Present in multiple copies in each cell

The nuclear and mitochondrial genomes form a complete set of human genetic information

The structure of the genome reflects its physical (chromosomes, mitochondria) and functional (genes and regulatory elements) organization



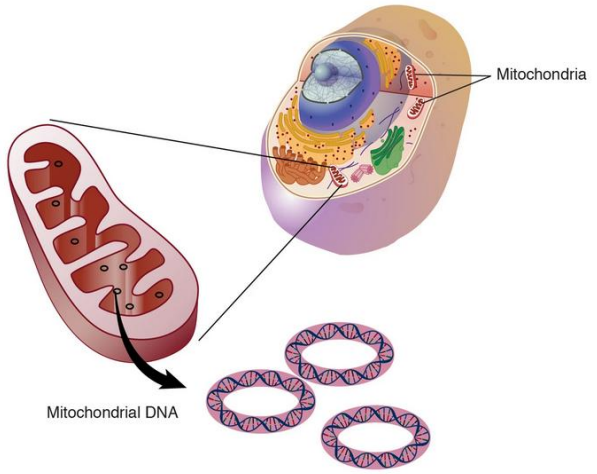
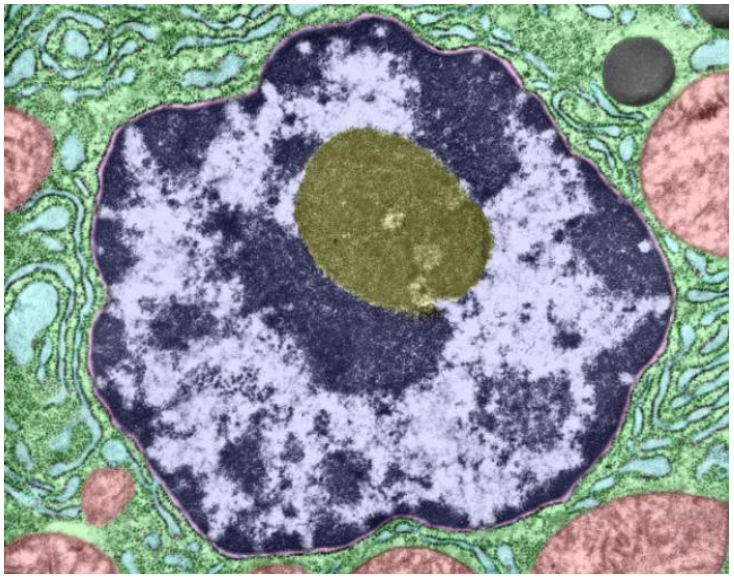


Image by National Human Genome Research Institute (NHGRI)

...it was not until the early 20th century that it was discovered that the nucleus stores genetic information

Structure and function of the cell nucleus and nucleolus

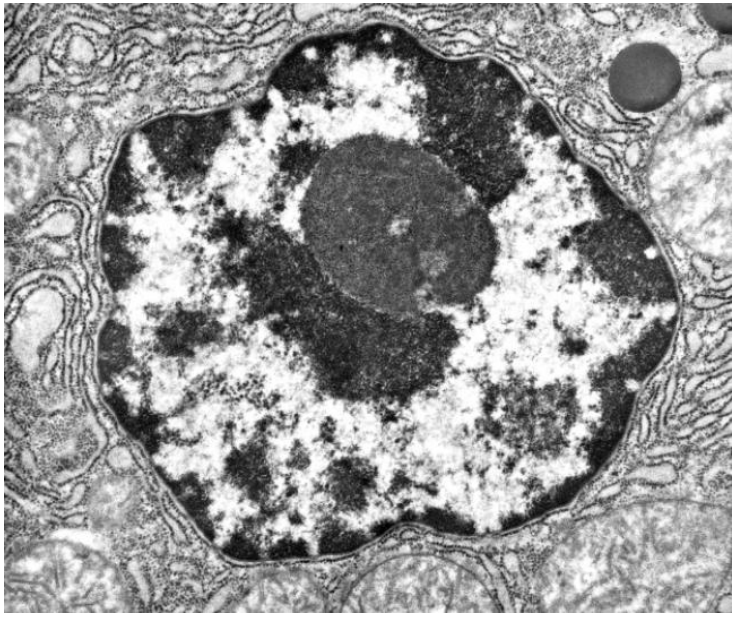


Nucleus: the largest organelle in the cell, contains genetic material encoded in DNA

Function: directs protein synthesis in the cytoplasm via

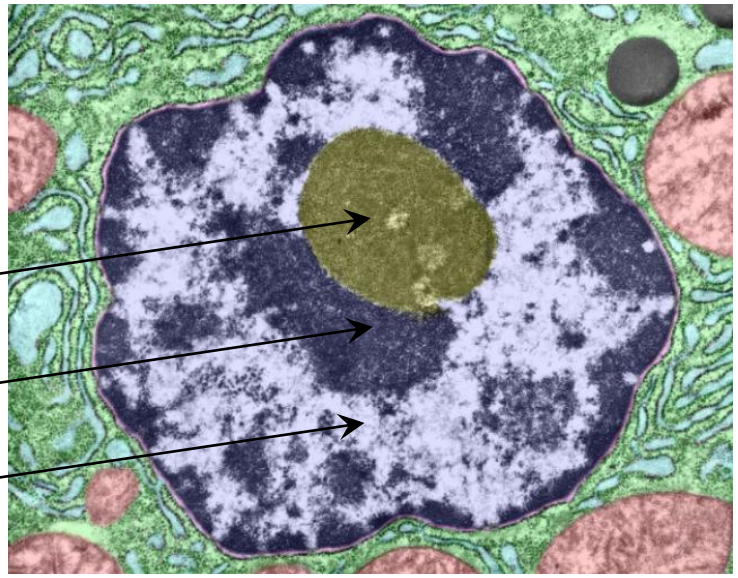
- ribosomal RNA – rRNA
- messenger RNA – mRNA
- transfer RNA – tRNA

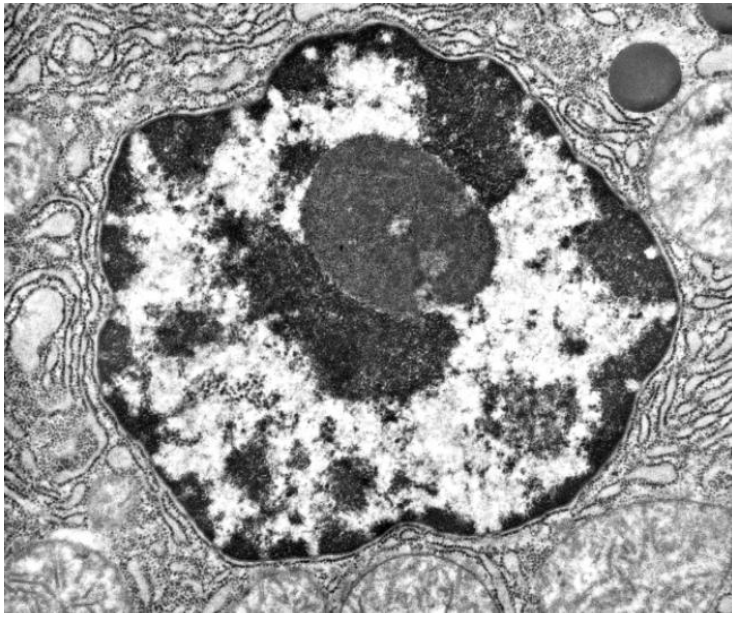
All types of RNA (including regulatory RNA and non-coding RNA) are synthesized in the cell nucleus



http://histologyguide.org/

Structure and functions of the cell nucleus and nucleolus

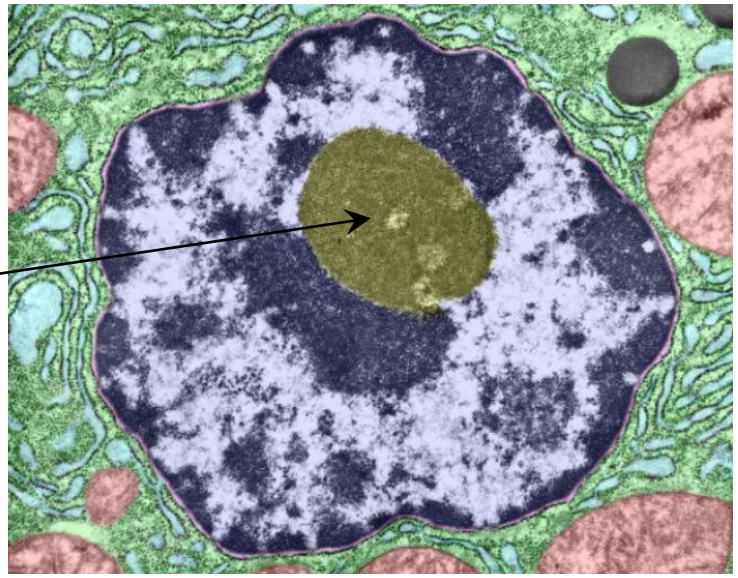




- **Chromatin (blue) -** DNA bound to histones and other proteins
- **Heterochromatin (dark blue) -** condensed chromatin, transcriptionally inactive
- **Euchromatin (light blue) –** dispersed areas of unpacked chromatin (often actively transcribed)
- **Nucleolus (green) -** site of ribosomal RNA (rRNA) transcription, ribosome production

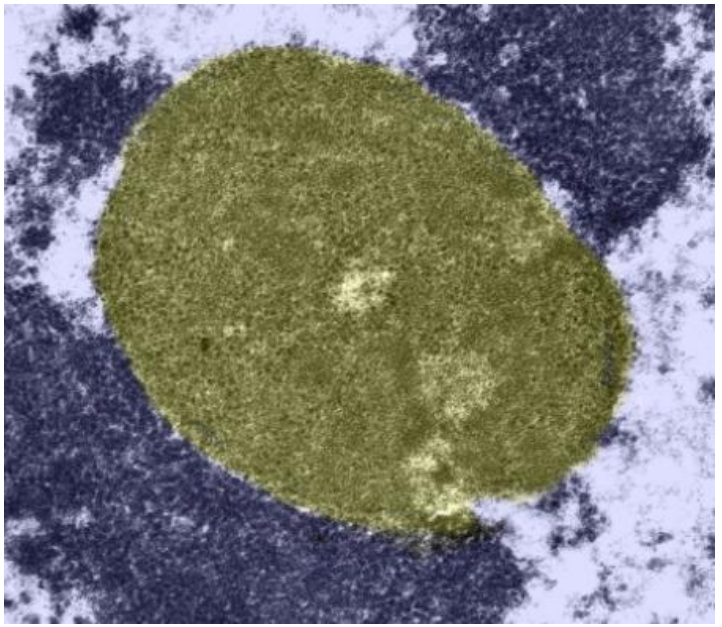
http://histologyguide.org/

Nucleolus



The site of ribosomal RNA transcription and ribosome production, but also the site where certain proteins are stored:

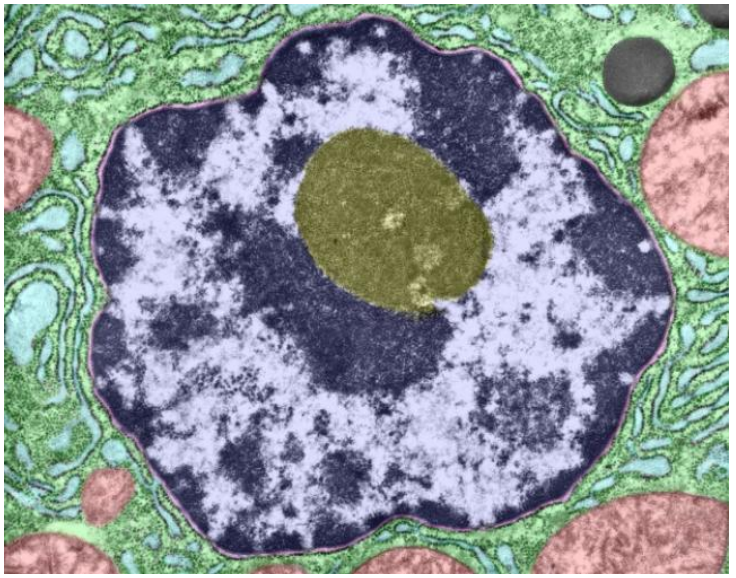
- **nucleolin** – regulates rRNA transcription
- **fibrillarin** – participates in the processing of pre-rRNA in snoRNP complexes
- **telomerase** – not active in the nucleolus
- **nucleostemine** – binds and inactivates p53



http://histologyguide.org/

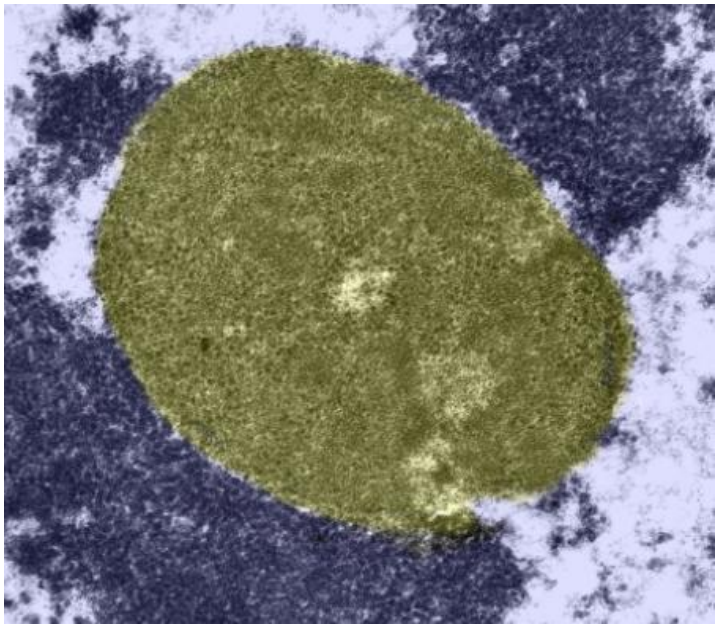
Nucleolus

- There can be 1 to 5 nucleoli in the nucleus



- The nucleolus contains about 500 genes necessary for the production of ribosomes; these genes originate from chromosomes 13, 14, 15, 21, and 22

- The short arms of these chromosomes contain nucleolar organizing regions (NORs), enabling nucleolar reconstruction after cell division



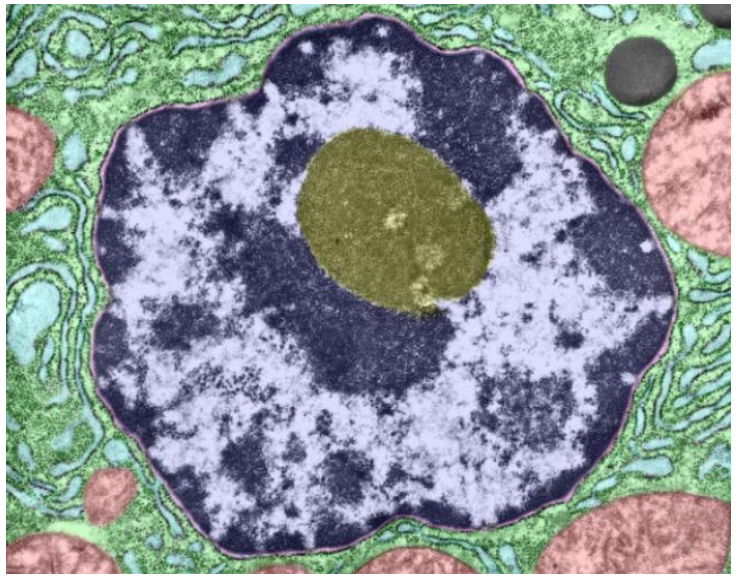
- rRNA genes are transcribed by RNA polymerase I
- rRNA is modified after transcription and assembled with ribosomal proteins (synthesized in the cytoplasm and imported into the nucleus through nuclear pores)

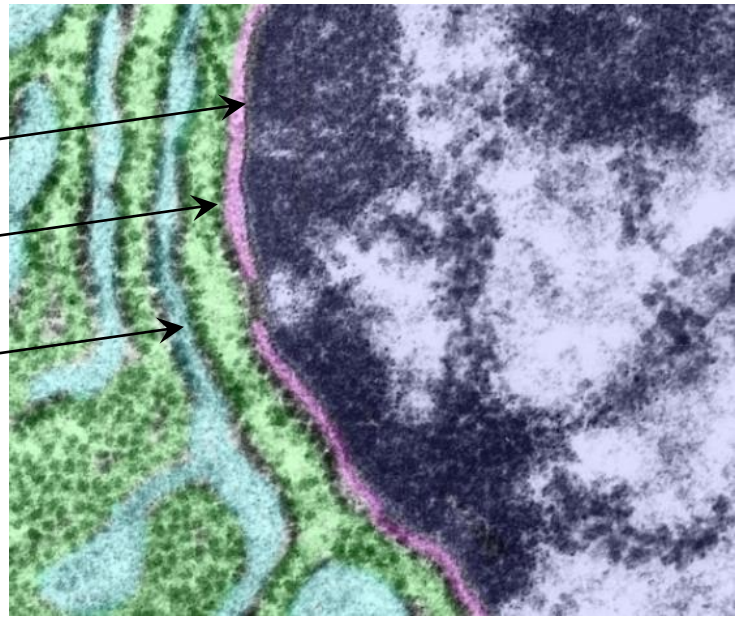
http://histologyguide.org/

Nuclear envelope

Nuclear envelope: consists of two membranes

- **Outer nuclear membrane**:
  - in some places continuous with the lumen of the endoplasmic reticulum
  - surrounded by cytoskeletal elements - vimentin





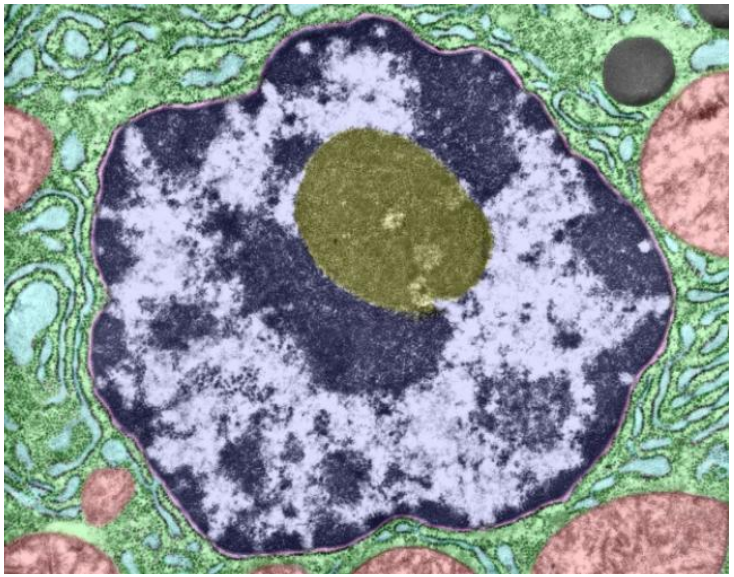
- **Inner nuclear membrane:**
  - separated from chromatin by the nuclear lamina (mainly lamins A, B1, B2, C)

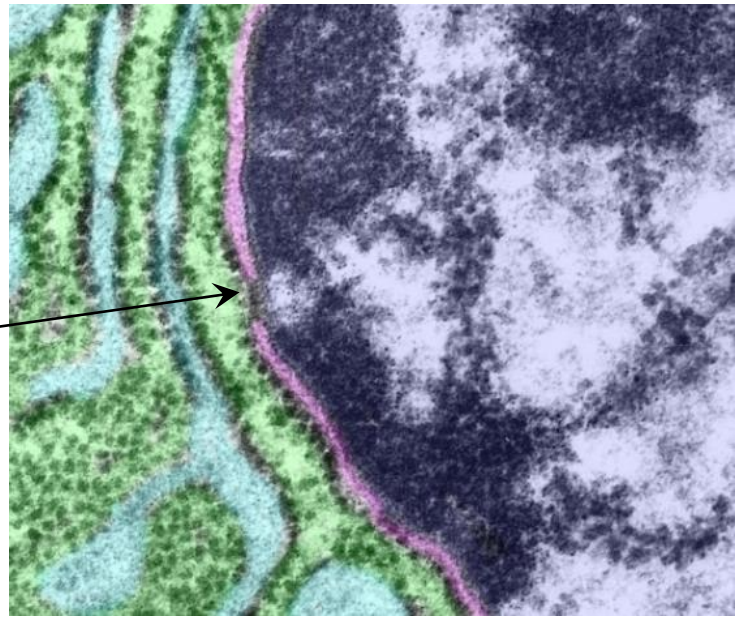
http://histologyguide.org/

Nuclear pores

Nuclear pores: channels in the nuclear membrane

are used for bidirectional transport of cargo (from various types of RNA to various proteins) between the nucleus and the cytoplasm



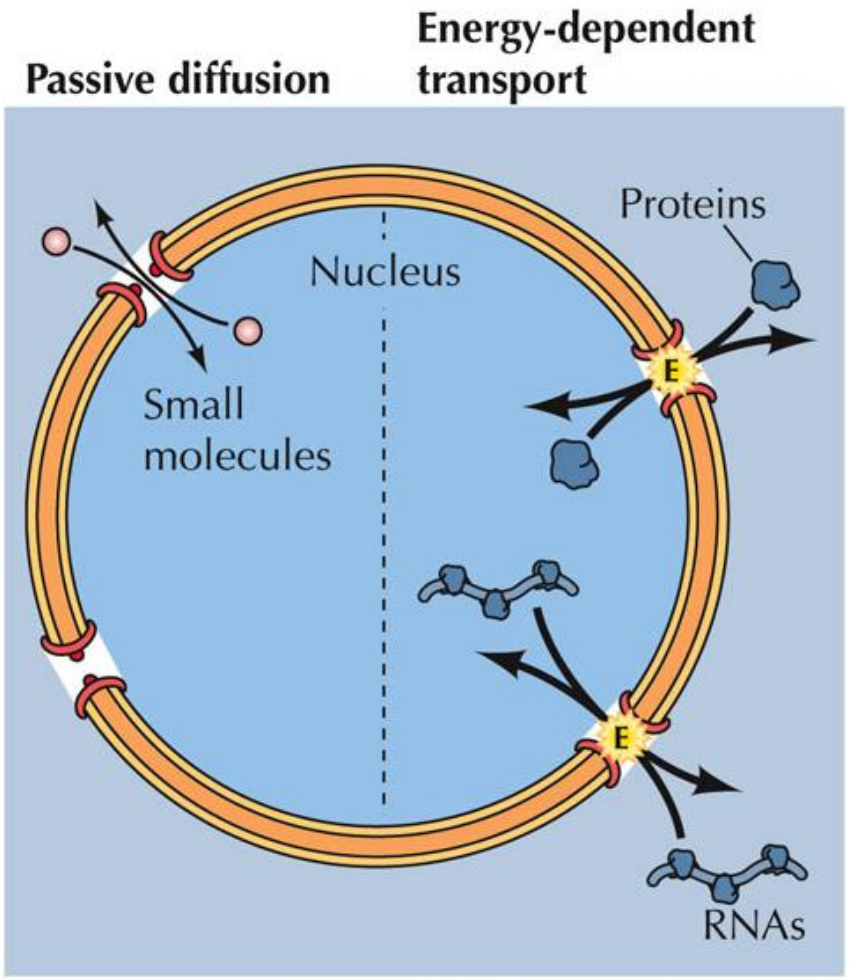


http://histologyguide.org/

Nuclear pores

Smaller molecules (small metabolites or proteins below 40 kDa) – passive transport through NPC

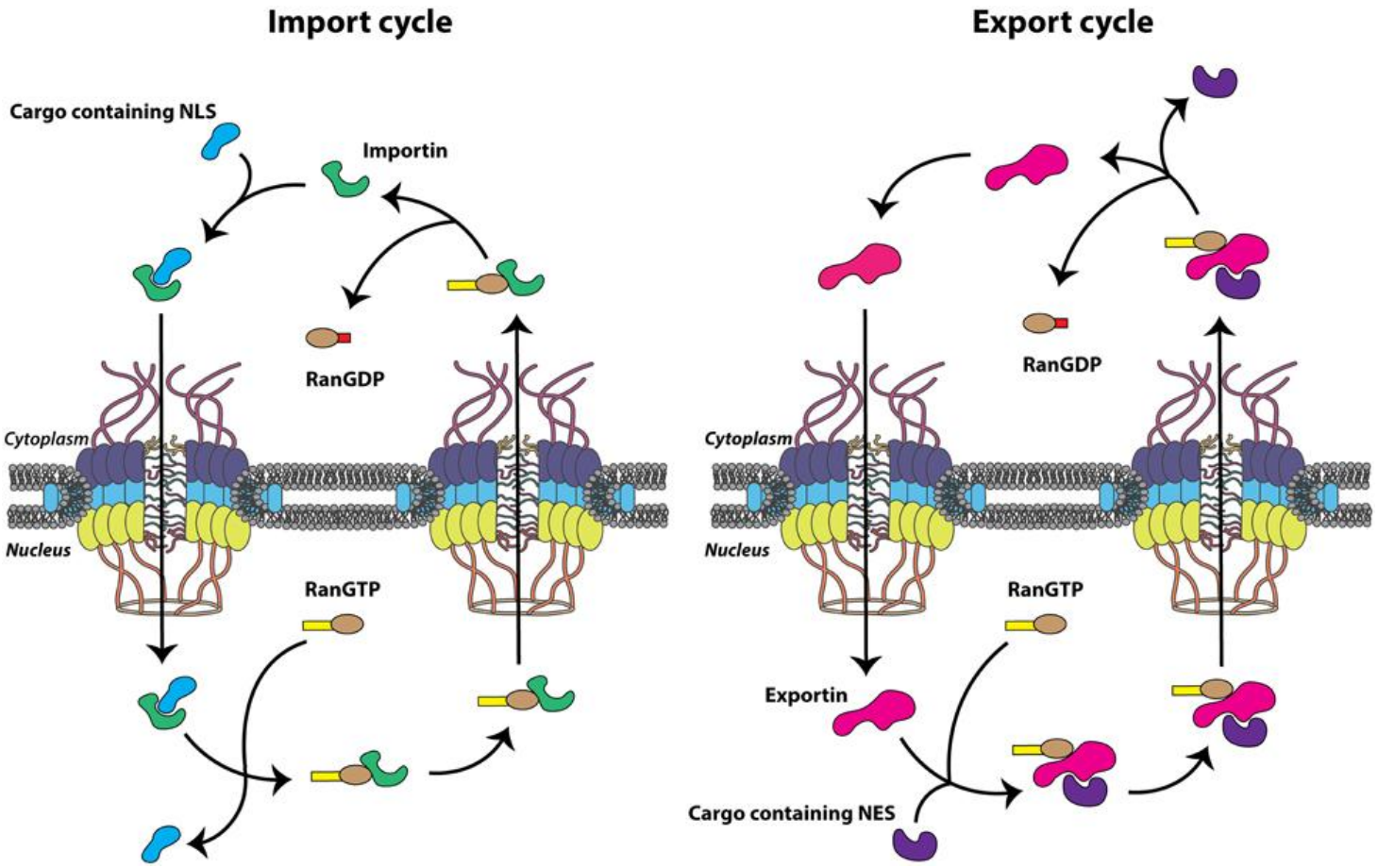
Larger molecules (mRNA, tRNA, ribosomes, and signaling molecules) – active transport



THE CELL, Fourth Edition, Figure 9.6 © 2006 ASM Press and Sinauer Associates, Inc.

Recycling of nuclear transport receptors (importin (green) and exportin (pink)) via the nuclear pore complex - (GTP hydrolysis)

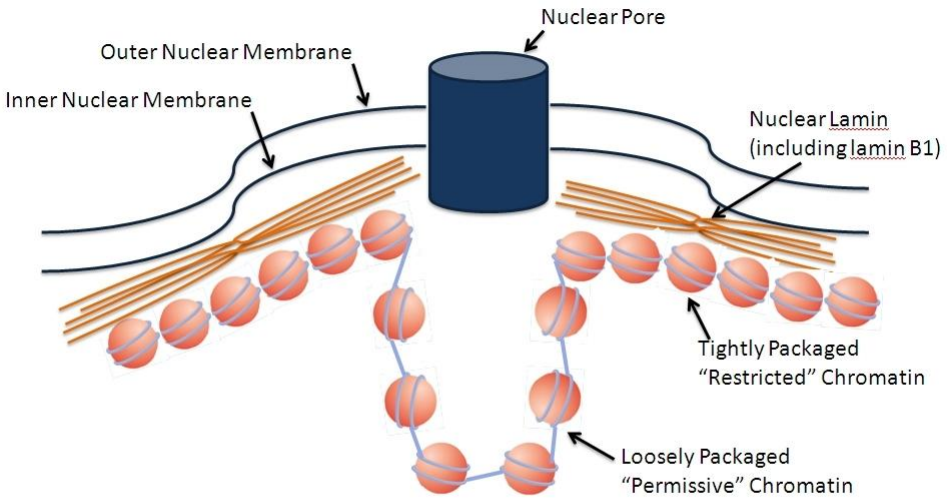
Imported molecules: nuclear localization signal (NLS) (light blue) or nuclear export signal (NES) (purple)



Flather D. et al.; Picornaviruses and nuclear functions: targeting a cellular compartment distinct from the replication site of a positive-strand RNA virus; Front. Microbiol., 2015

Nuclear lamina - laminins

During interphase, chromatin in the nucleus is packed into **chromosome territories** and anchored to the nuclear lamina.



Adapted from Kind and van Steensel, *Curr Opin Cell Biol*

Laminas are regulators of chromatin positioning. Transcriptionally silent regions of the genome, such as centromeres, telomeres, and the inactive X chromosome, are preferentially located under the nuclear lamina.

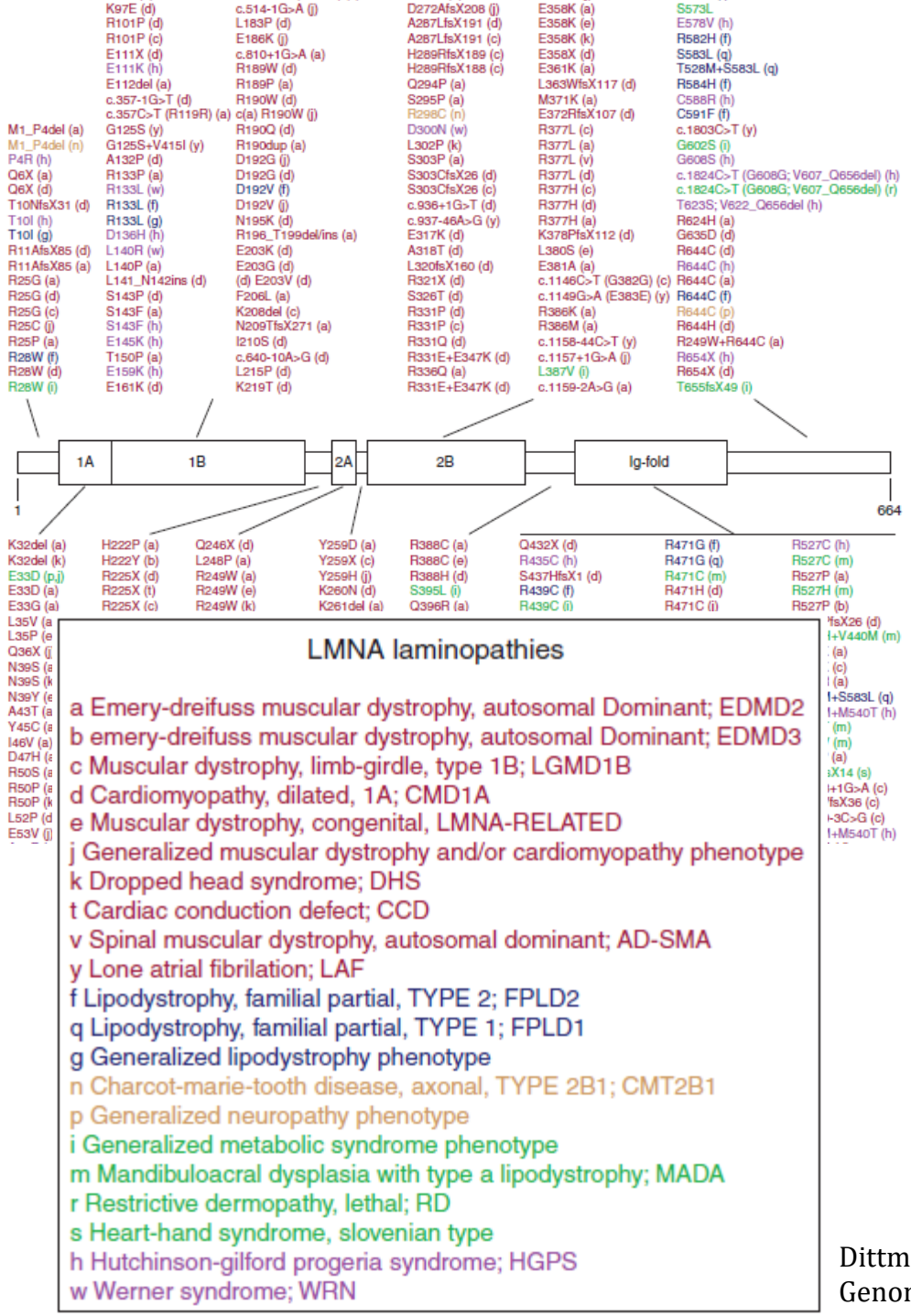
Owen K. et al., Chromatin Structure and Replication Origins: Determinants of Chromosome Replication and Nuclear Organization, In Journal of Molecular Biology, 2014

Bas van Steensel et al.; Genomics tools for unraveling chromosome architecture; Nature Biotechnology; 2010

Laminopathies

***LMNA* mutations**—numerous heritable human diseases

Over 20 distinct diseases - forms of cardiomyopathy, muscular dystrophy, lipodystrophy and aging-related progeria



Red - preferential involvement of skeletal and cardiac muscle

Blue - lipodystrophies;

Yellow - neuropathy disorders (motor and sensory neurons of the peripheral nervous system)

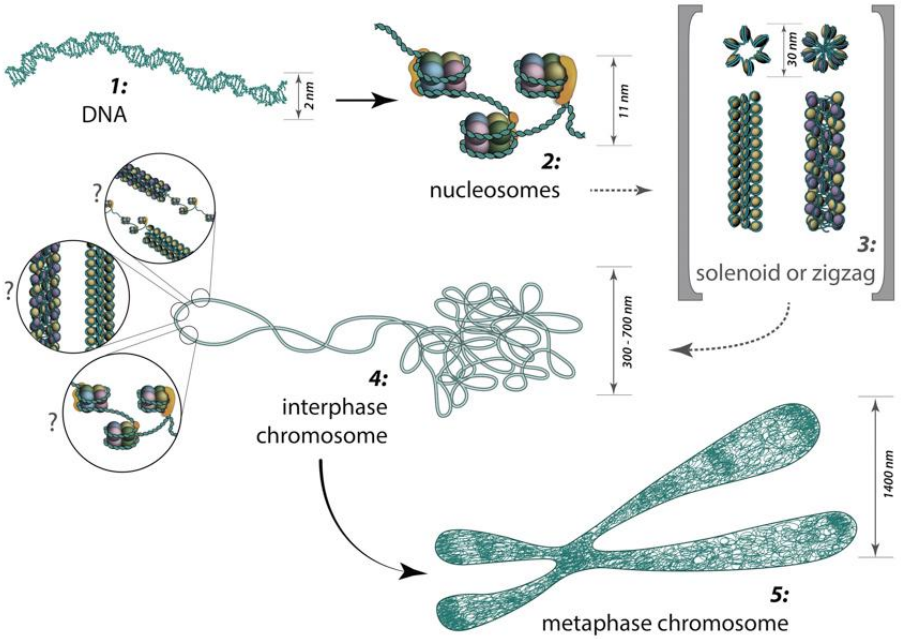
Green - 'systemic' laminopathies- multiple tissue systems

Purple - mutations associated with premature aging disorders.

Dittmer, Travis & Misteli, Tom. (2011). The lamin protein family. Genome biology. 12. 222. 10.1186/gb-2011-12-5-222.

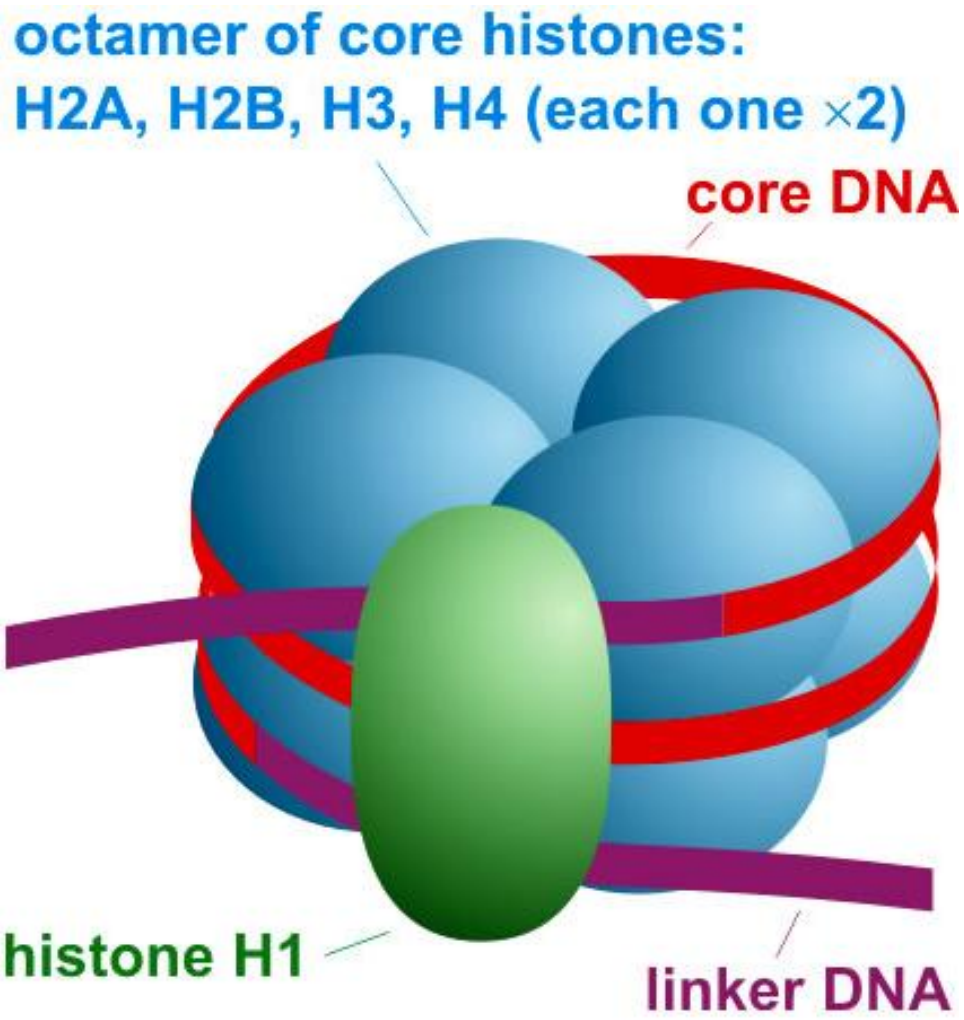
Levels of DNA organization in the cell nucleus

- **Double helix DNA**
- **Nucleosome** – a fragment of DNA (approx. 146 bp) wrapped around a histone octamer
- **10 nm fiber** – so-called "*beads on a string*"
- **30 nm fiber** – a more condensed form, can take two models:
  - solenoid model
  - zigzag model
- **300** nm **fiber** – higher level of condensation, e.g., loops of 30 nm fibers supported by a protein core
- **Metaphase chromosome** (~700–1400 nm) – maximally condensed fiber during mitosis



MBInfo contributors. Interphase and mitotic chromatin. In MBInfo Wiki, Retrieved 10/21/2014 from http://mbinfo.mbi.nus.edu.sg/figure/1389942033388/

Nucleosomes



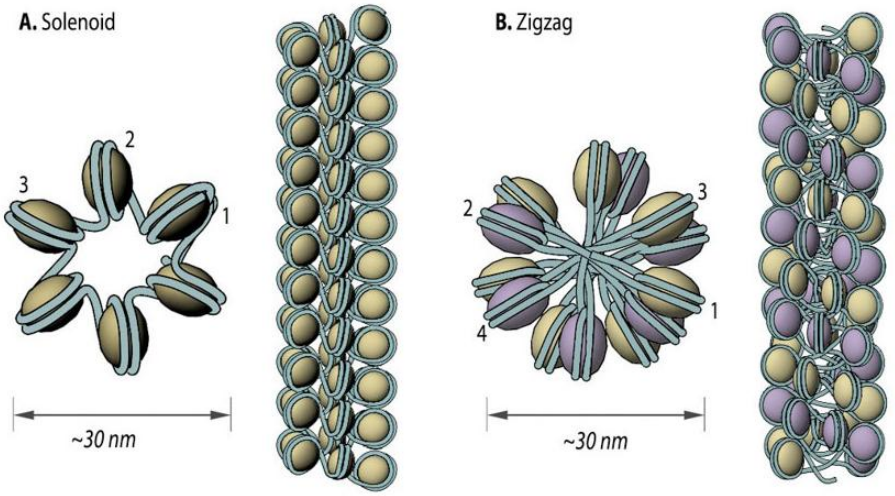
Solenoid/zigzag (30 nm fiber)

Solenoid model

- Nucleosomes arranged spirally in a compact, helical structure
- Each nucleosome contacts neighboring
- Linker DNA is bent inward toward the center of the spiral
- A more compact and stable structure

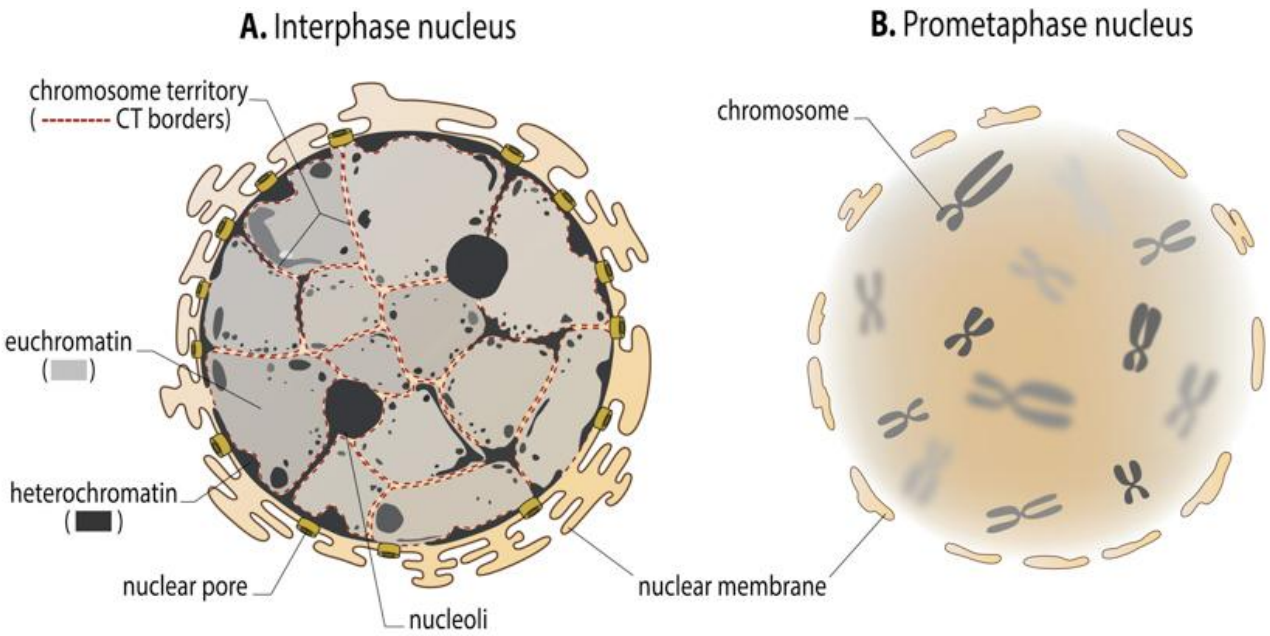
Zigzag model

- Nucleosomes arranged alternately on opposite sides of the fiber axis
- Linker DNA runs in straight segments, connecting every other nucleosome
- Less compact, more flexible structure



MBInfo contributors. Interphase and mitotic chromatin. In MBInfo Wiki, Retrieved 10/21/2014 from http://mbinfo.mbi.nus.edu.sg/figure/1389942033388/

Chromosome territories



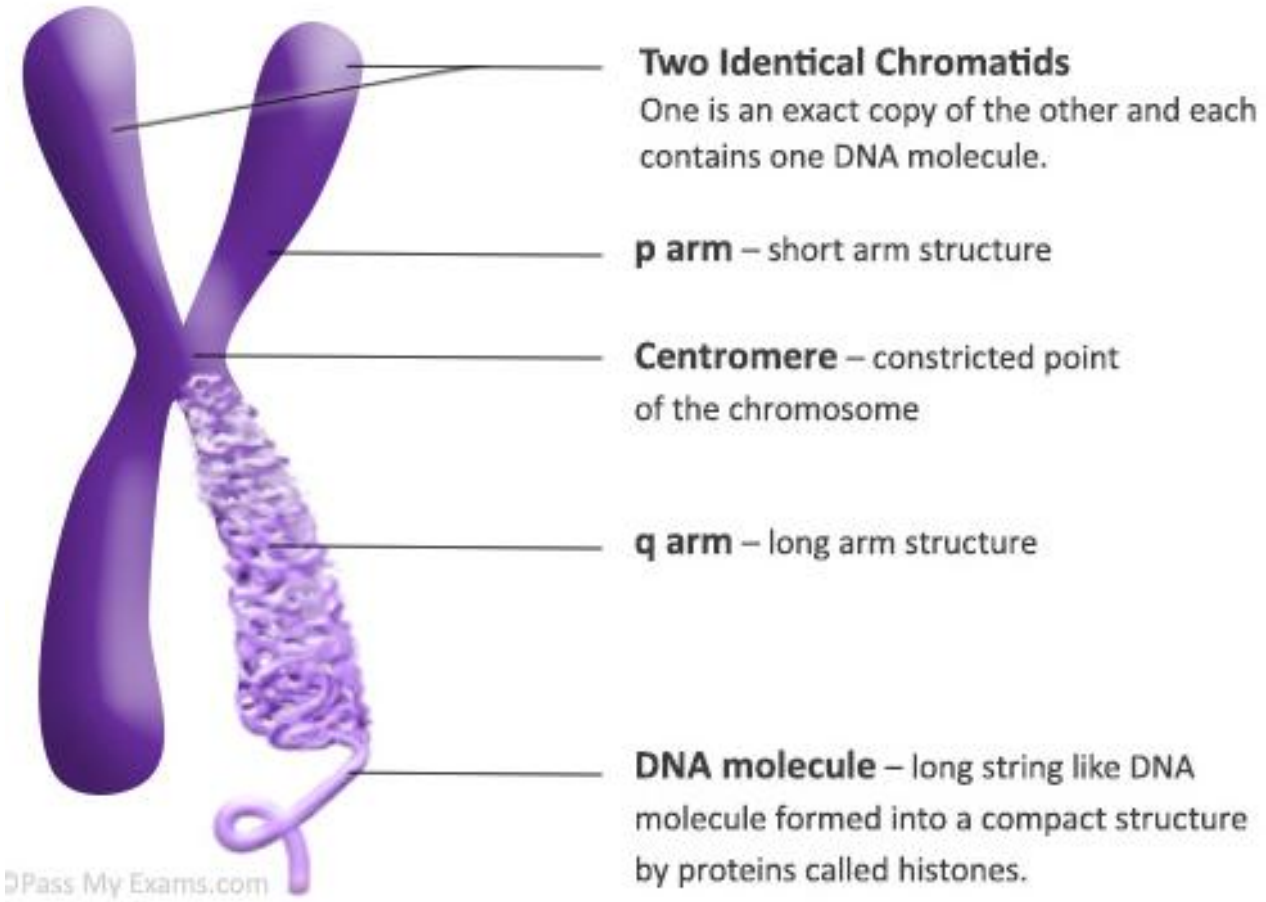
Gene-rich chromosomes - internal positions

Gene-poor chromosomes - peripheries

Genes can move from the periphery toward the interior after being "turned on"

MBInfo contributors. Interphase and mitotic chromatin. In MBInfo Wiki, Retrieved October 21, 2014, from http://mbinfo.mbi.nus.edu.sg/figure/1389942033388/

Chromosome structure

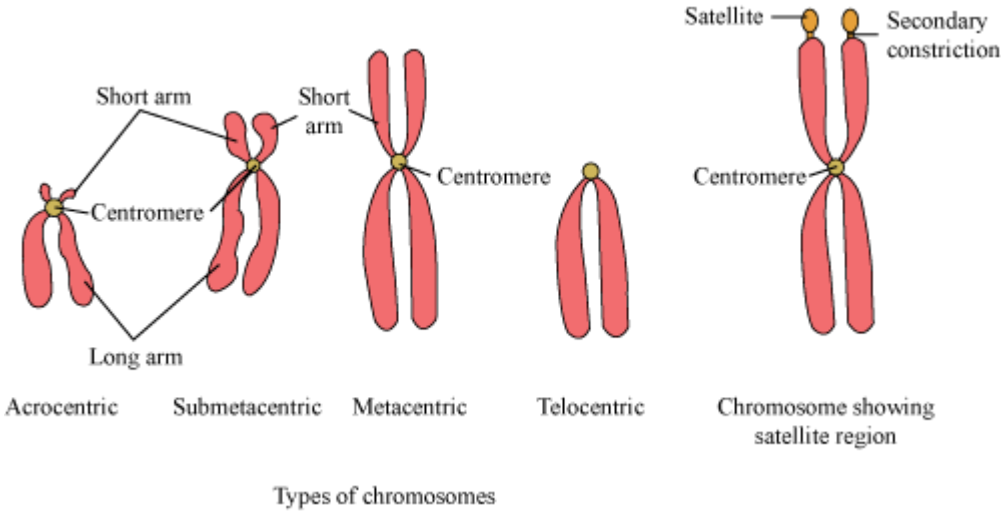


http://www.passmyexams.co.uk/GCSE/biology/what-are-chromosomes.html

Chromosome structure

Metacentric chromosomes
Centromere in the middle, both arms are of equal length. Human chromosomes 1, 3, 16, 19, 20.

Submetacentric chromosomes
Slight asymmetry in the length of the two arms. Human chromosomes 2, 4-12, 17, 18.



Types of chromosomes

Acrocentric chromosomes
One arm is very long, one is greatly shortened. Human chromosomes 13, 14, 15, 21, and 22.

Telocentric chromosomes
Centromere at the very end of the chromosome. Humans do not have telocentric chromosomes, they occur in other species, such as mice.

http://www.passmyexams.co.uk/GCSE/biology/what-are-chromosomes.html

Karyotype

What can we determine based on a karyotype?

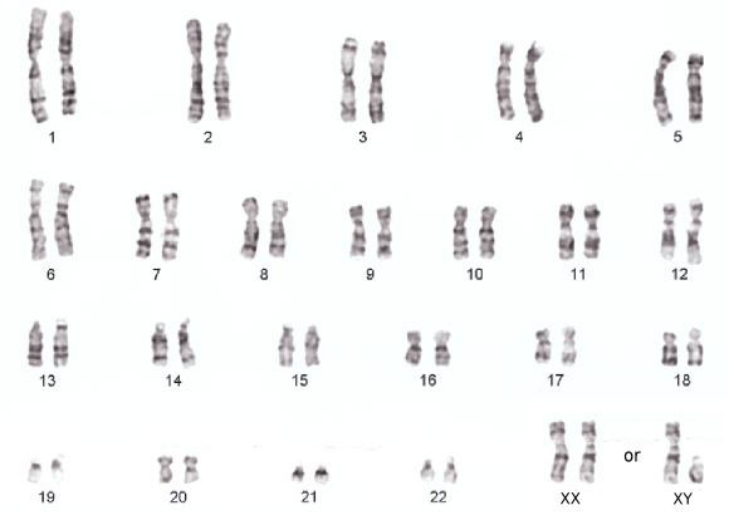
Number of chromosomes

Chromosome size

Centromere location, chromosome arm length

Presence of satellites

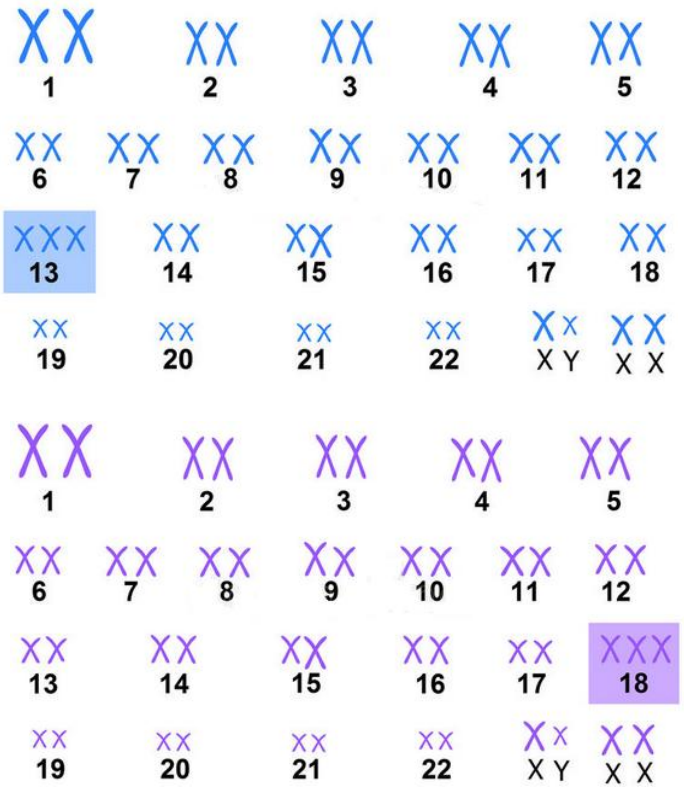
Banding pattern



Humans - a total of 46 chromosomes (23 from each parent), 44 autosomes

http://study.com/academy/lesson/karyotype-definition-disorders-analysis.html

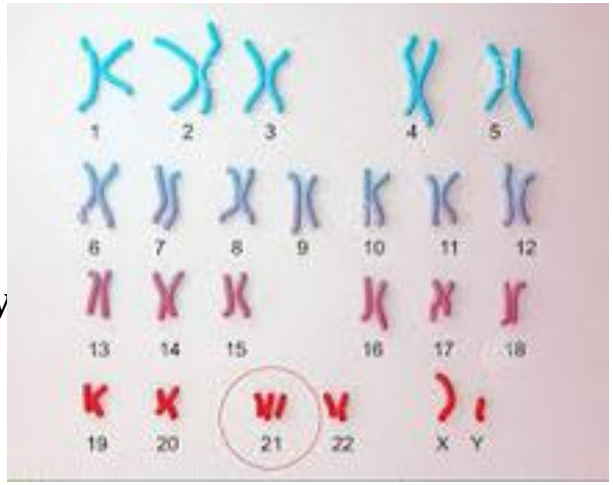
Chromosomal aberrations within autosomes



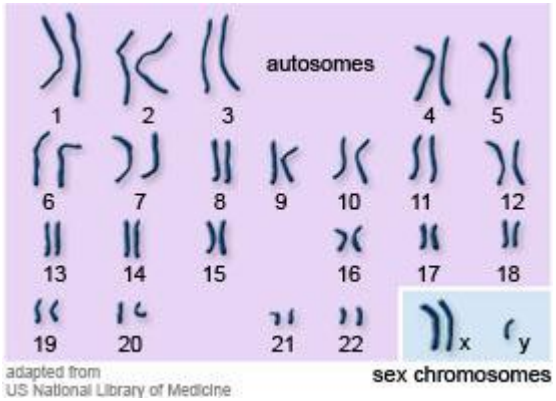
Patau syndrome – severe intellectual disability, multiple malformations, including heart and kidney defects, muscular hypotonia, and craniofacial dysmorphism

Edwards syndrome – multiple organ defects, heart defects, microcephaly, clenched fists with overlapping fingers, and severe intellectual disability

Down syndrome - physical growth retardation, characteristic craniofacial dysmorphia, and mild to moderate intellectual disability

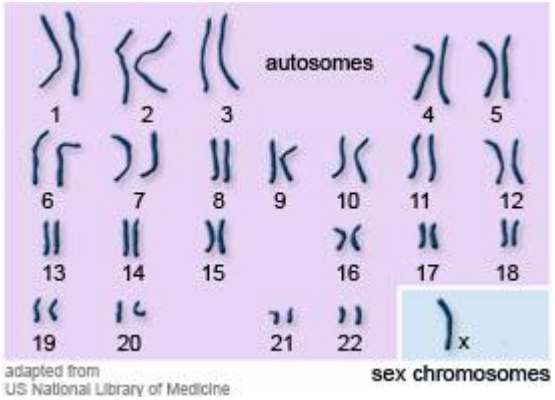


Chromosomal aberrations within sex chromosomes



Klinefelter syndrome - Two X chromosomes and one Y chromosome (47,XXY).

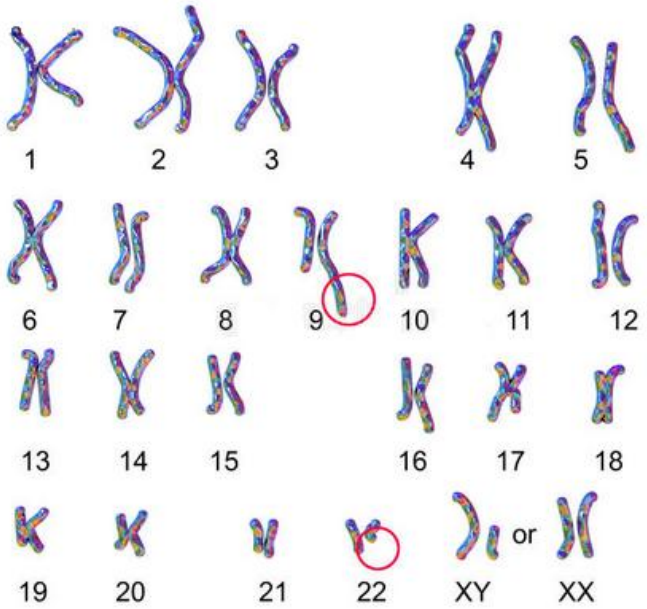
Many men with an extra X chromosome have no obvious clinical symptoms, and the diagnosis may remain undetermined; the most common symptom is infertility.

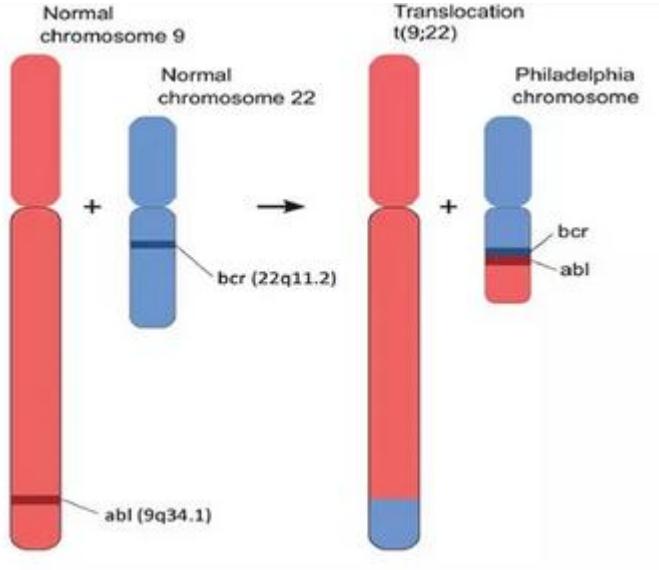


Turner syndrome – monosomy of the X chromosome (45,X), absence of the second sex chromosome, leading to typical phenotypic features and developmental disorders, including short stature and ovarian failure.

Chromosomal aberrations – fusion chromosome

Philadelphia chromosome – a genetic aberration involving translocation t(9;22)(q34;q11), in which genetic material between chromosomes 9 and 22 is exchanged, leading to the formation of a shortened, "defective" chromosome 22; characteristic of chronic myeloid leukemia (CML) and some cases of acute lymphoblastic leukemia.





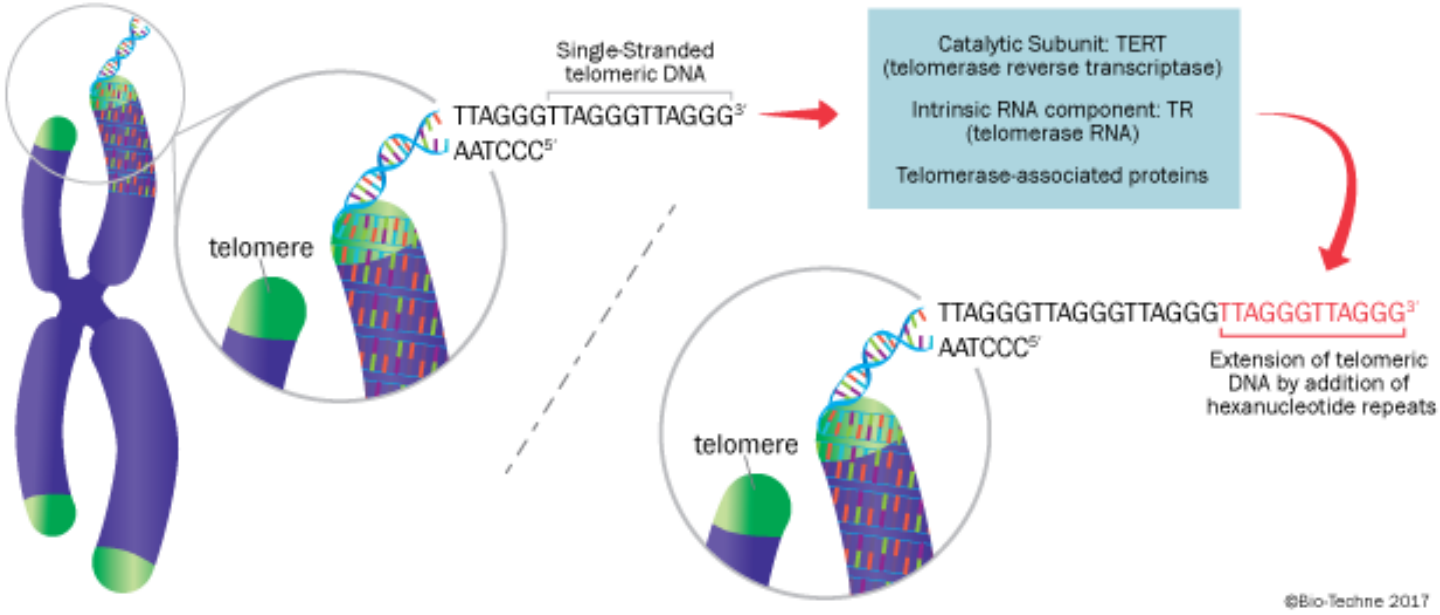
https://www.dreamstime.com/stock-illustration-philadelphia-chromosome-karyotype-male-female-d-illustration-showing-defective-chromosomes-translocational-defect-image79799063

https://en.wikipedia.org/wiki/Philadelphia_chromosome

Telomeres

Hexameric DNA repeats [(TTAGGG)n] at the ends of chromosomes protect chromosomes from degradation and sticking together.

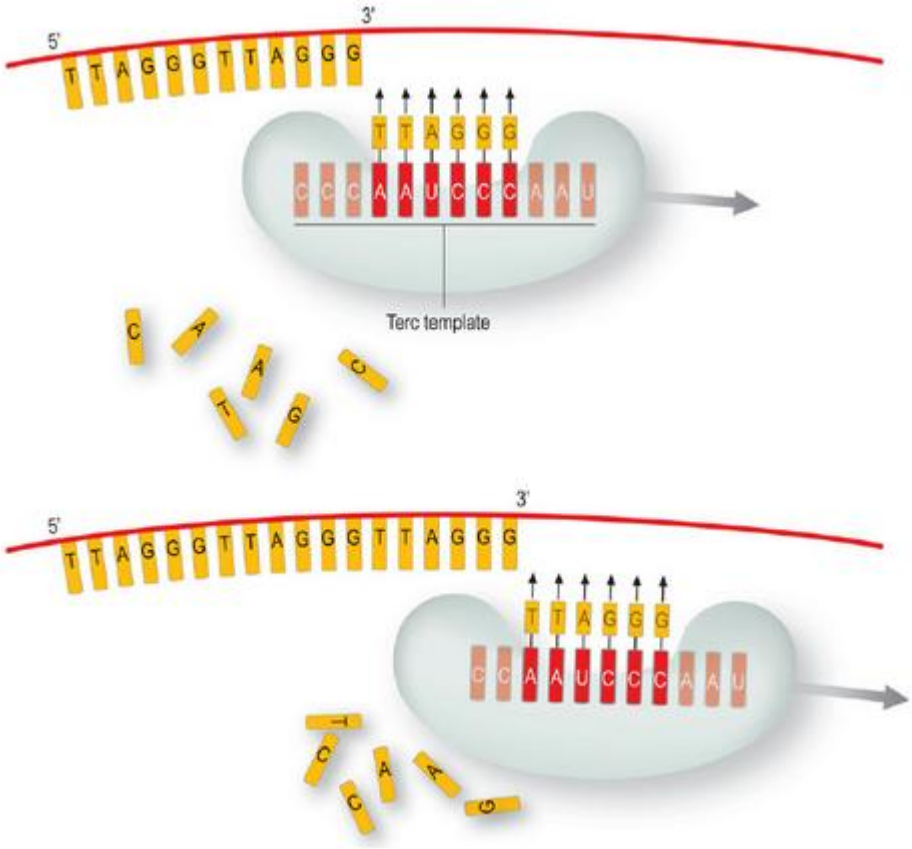
Replicative telomere shortening – the main cause of human aging and age-related diseases



https://www.metabolomicmedicine.com/english/telomere_analysis-na-192.html
https://www.rndsystems.com/resources/articles/telomerase

Telomeres

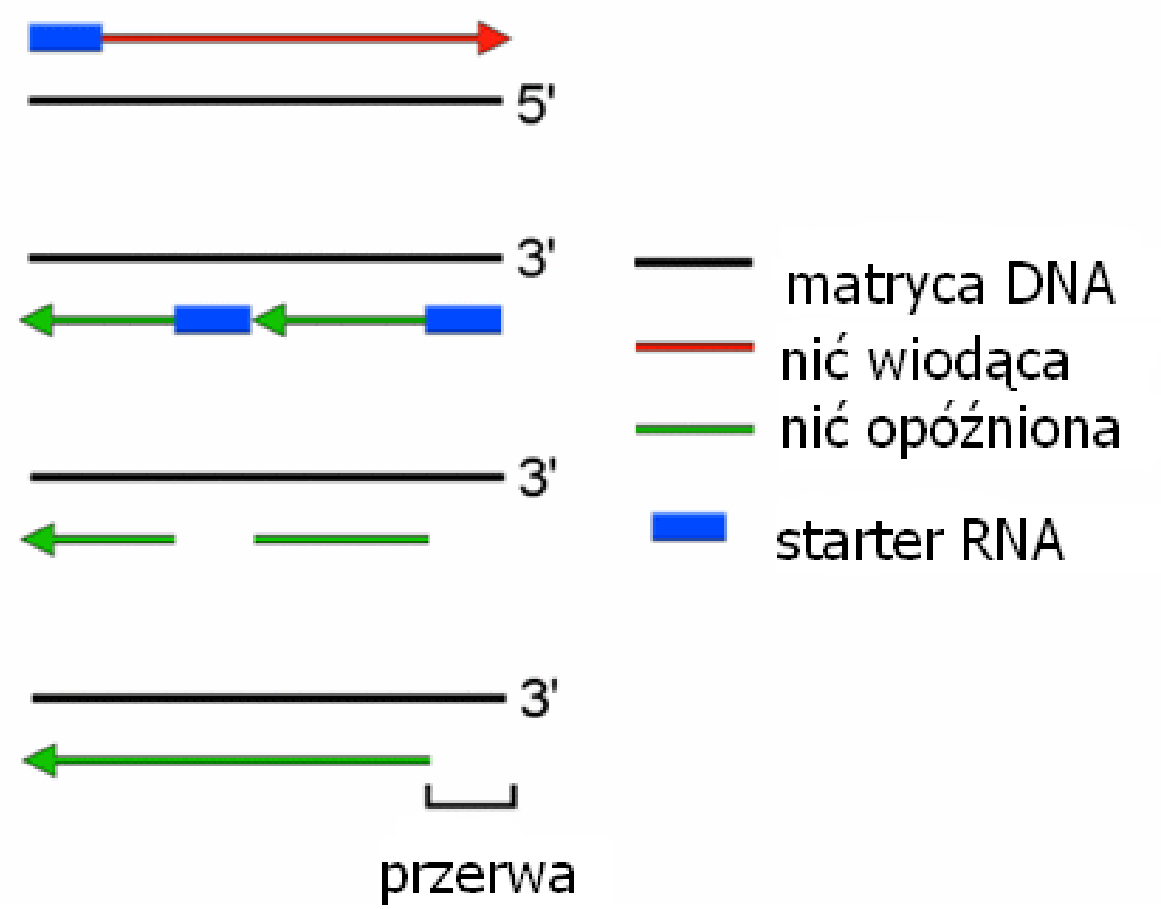
Telomere elongation by telomerase.



Telomerase contains an RNA template that is used to add telomeric nucleotide repeats (in humans TTAGGG) to the 3' ends of telomeres.

S M Wong et al. (2009). Telomere biology in cardiovascular disease: The TERC-/- mouse as a model for heart failure and aging. Cardiovascular research.

Telomeres and the "end-replication problem"



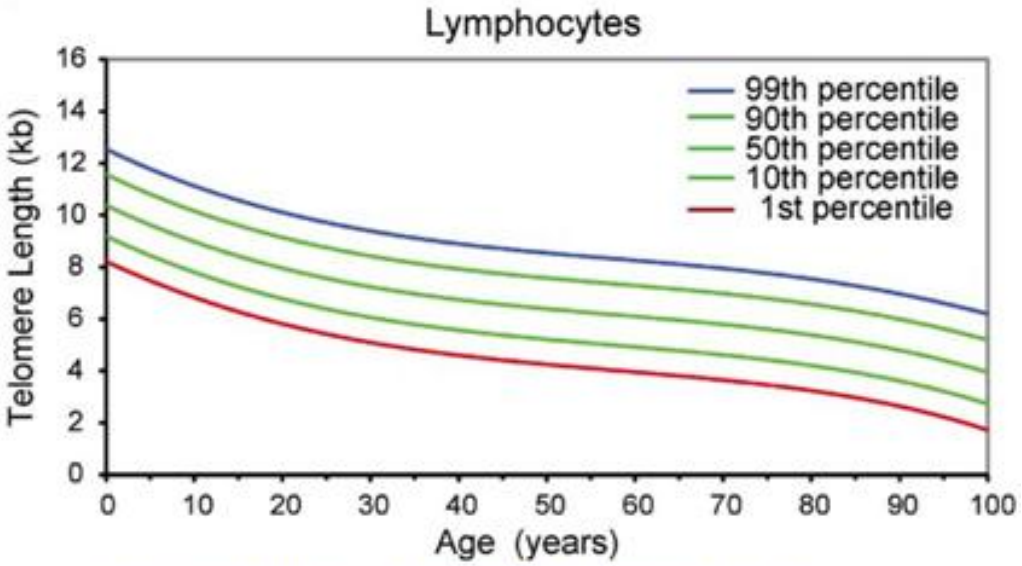
Telomere sequences (TTAGGG)n

Telomeres

The average length of telomeres in human cells at the moment **of fertilization** is 15,000 base pairs.

The average length of telomeres **at birth** is 10,000 base pairs.

Chromosomes become unstable when telomere length is less than 4,500 base pairs



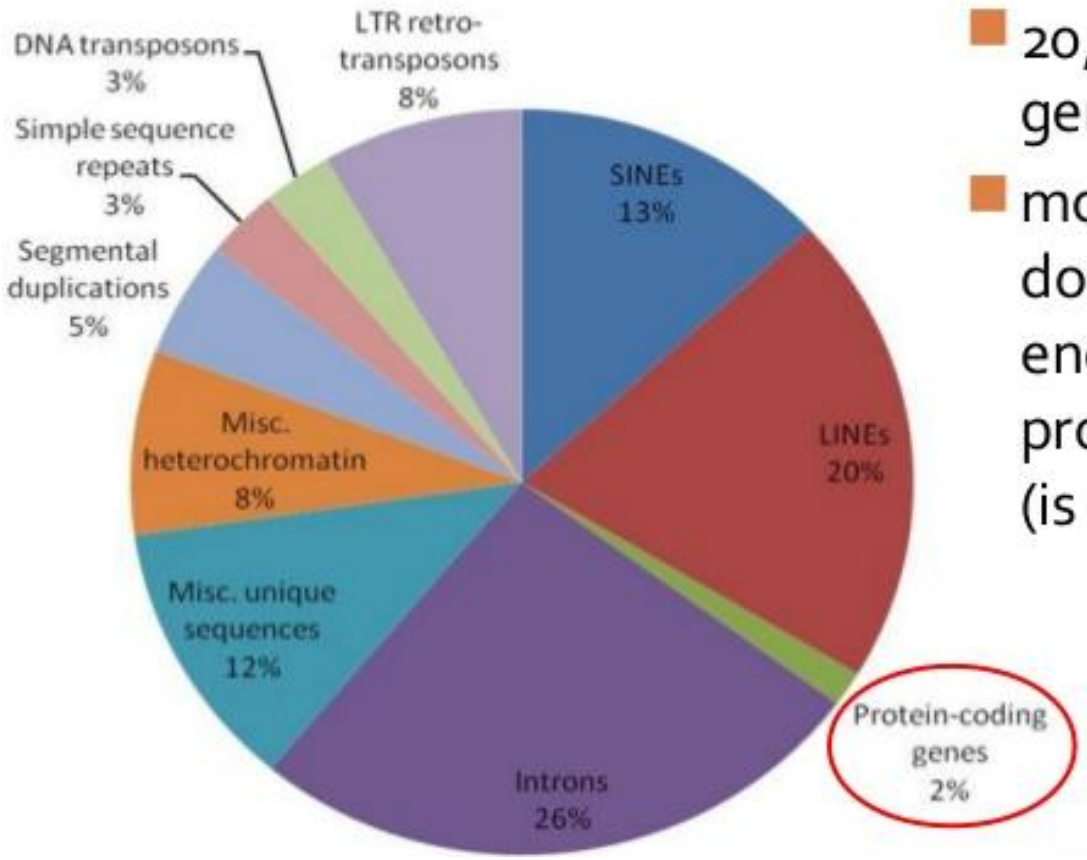
Average telomere length in relation to age.

Modified from J. Exp. Med. 2014 Vol. 211 No. 13.

PIK3R1 mutation augments PI3K signaling in new PID | Lucas et al.

https://www.metabolomicmedicine.com/english/telomere_analysis-na-192.html

Genetic information in humans

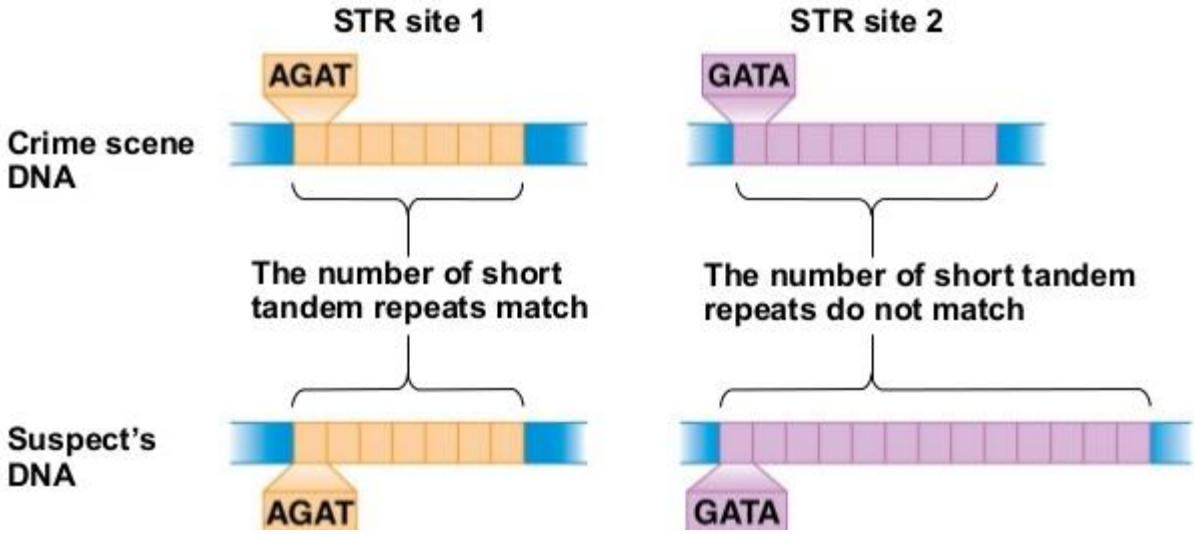


- 20,000-30,000 genes
- most of the DNA does NOT encode for a protein or RNA (is not a gene)

Mini- and micro-satellites

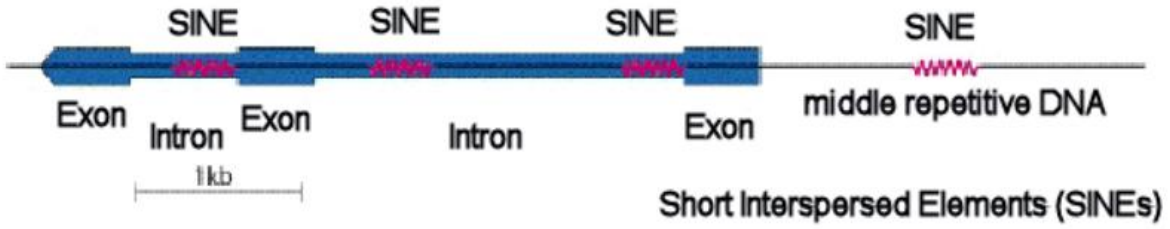
Minisatellites (VNTR – Variable Number Tandem Repeats) – medium-sized tandem repeats (10-100 bp), repeated from a dozen to several hundred times, hypervariable, highly polymorphic, occurring mainly near telomeres; used in DNA fingerprinting (forensic medicine)

Microsatellites or "Short Tandem Repeats" (STR) – small tandem repeats (1-6 bp), repeated several to several dozen times, scattered; most often dinucleotide repeats, used in DNA fingerprinting (forensic medicine)



LINES, SINES, transposons

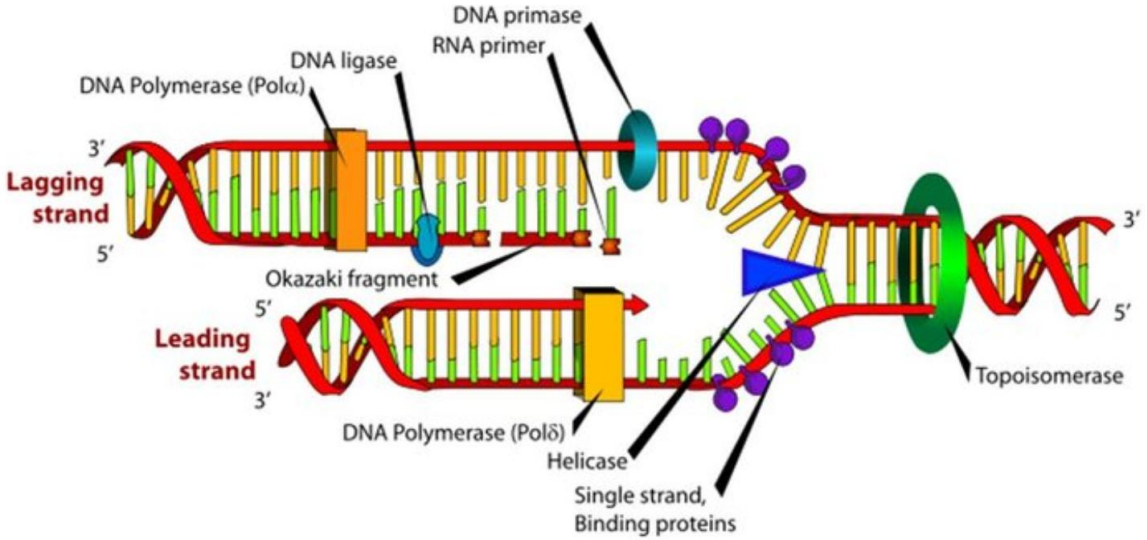
- Transposons – transposable elements migrating within the genome
- **LINE (*Long Interspersed Nuclear Elements*)**
  - Long elements scattered throughout the genome – usually **6–7 thousand base pairs** long
  - The most numerous type is LINE-1 (L1)
  - They constitute approx. 17% of the human genome
  - They originate from ancient retrotransposons — sequences that were able to copy themselves in the genome via RNA
  - Some of them are still **active**, meaning they can insert their copies into new locations in the DNA
- **SINE (Short Interspersed Nuclear Elements)**
  - Shorter sequences — approximately 100–300 base pairs
  - The best known are Alu sequences (about 1.1 million copies in the genome)
  - They constitute about 11% of the human genome
  - They do not have their own enzymes for copying, so they use the LINE-1 mechanism

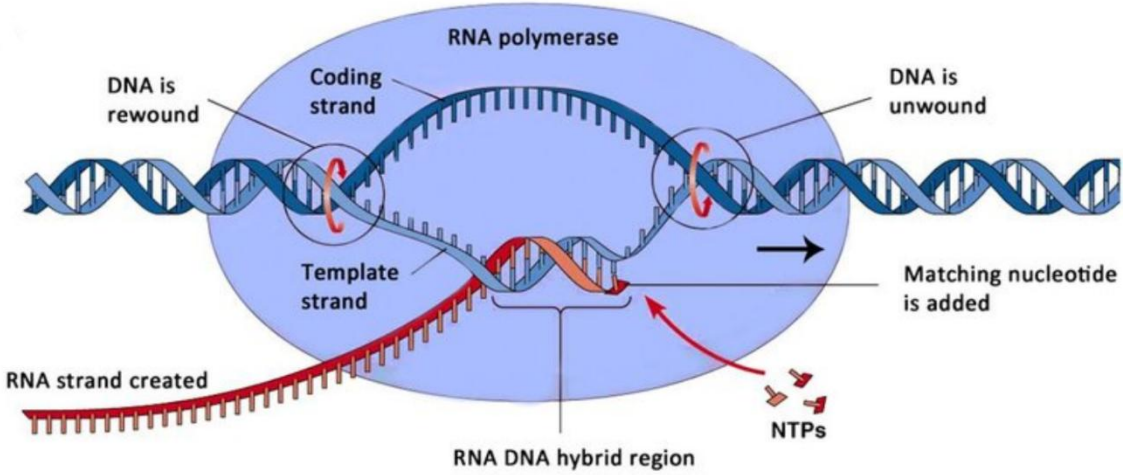


(Klug & Cummings 2000)

DNA

replication and transcription





DNA replication

- Occurs during the S phase of the cell cycle
- It is semi-conservative – two complete double helices are formed from one strand
- Replication origins (approx. 10,000) – rich in A–T pairs (easier to break)

- **Replication forks** – move at a speed of approx. 100 base pairs/s
- **Helicase** – unravels the double helix of DNA.
- **SSB** (*Single-Strand Binding proteins)* – stabilize single strands
- **Topoisomerase** – removes tension in front of the replication forks
- **Primase** – synthesizes an RNA primer from which **DNA polymerase** can begin to build a new strand
- **DNA polymerase** – adds a new strand (5'→3').
- **DNA ligase** – joins Okazaki fragments on the lagging strand
- **Telomerase** – lengthens the ends of chromosomes (telomeres)

Exonuclease activity (3'-5') of DNA polymerase

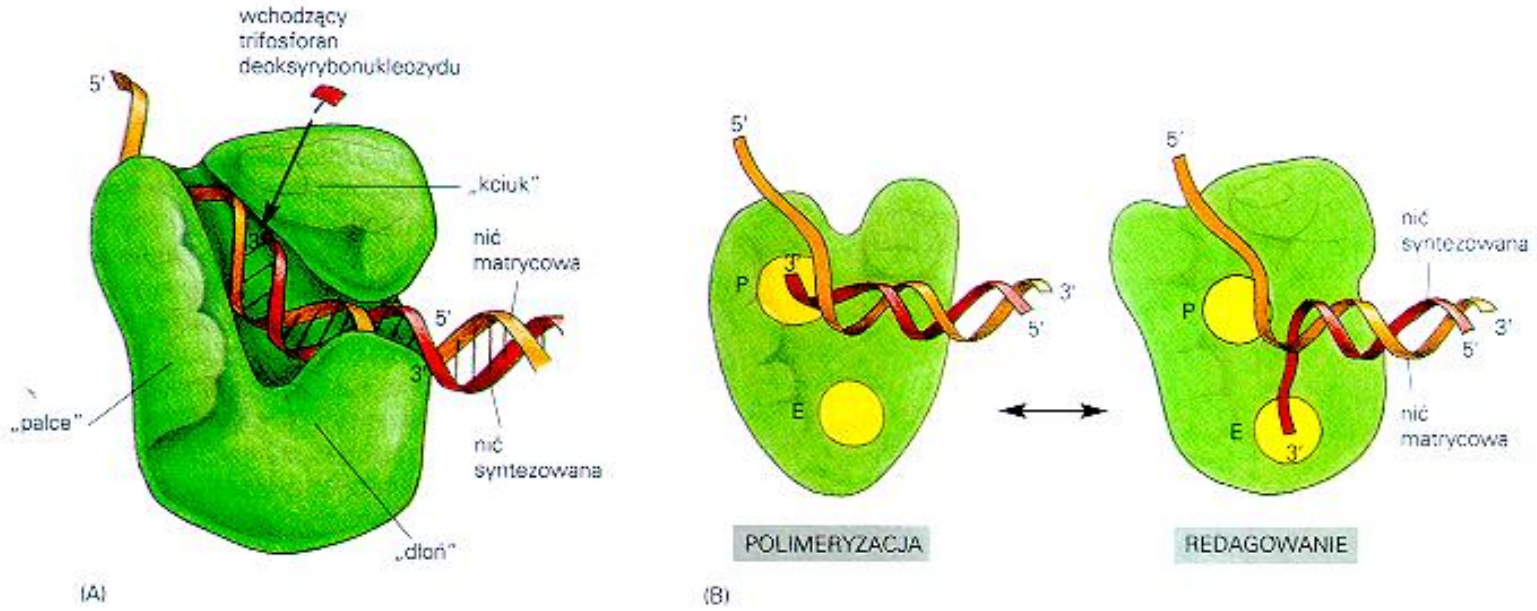
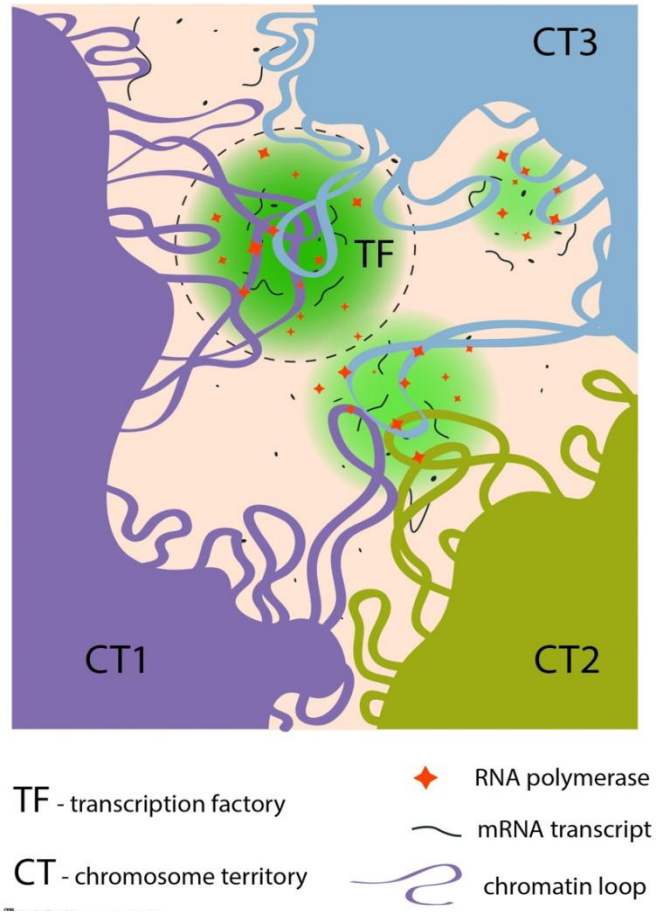


Fig. 6-14 Alberts et al., The Principles of Cell Biology, 2005

Transcription



Transcription foci (transcription factories) - submicron nuclear regions (40-100 nm in diameter) that are enriched in RNA polymerase II (RNApol II) complexes.

RNA polymerase I is restricted to the nucleolus, while RNA polymerases II and III are nucleoplasmic.

MBInfo contributors. Interphase and mitotic chromatin. In MBInfo Wiki, Retrieved October 21, 2014, from http://mbinfo.mbi.nus.edu.sg/figure/1389942033388/

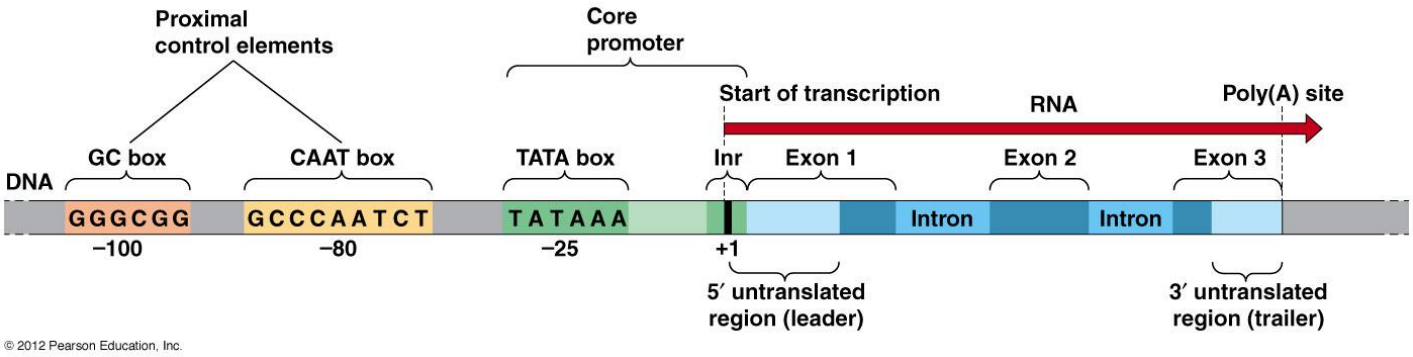
Transcription

In eukaryotes, there are three RNA polymerases (I, II, and III) involved in transcription:

RNA polymerase I (located in the nucleolus) transcribes precursor rRNA molecules.

RNA polymerase II produces most mRNA and miRNA.

RNA polymerase III is responsible for the production of pre-tRNA, 5S rRNA, and other small RNAs.



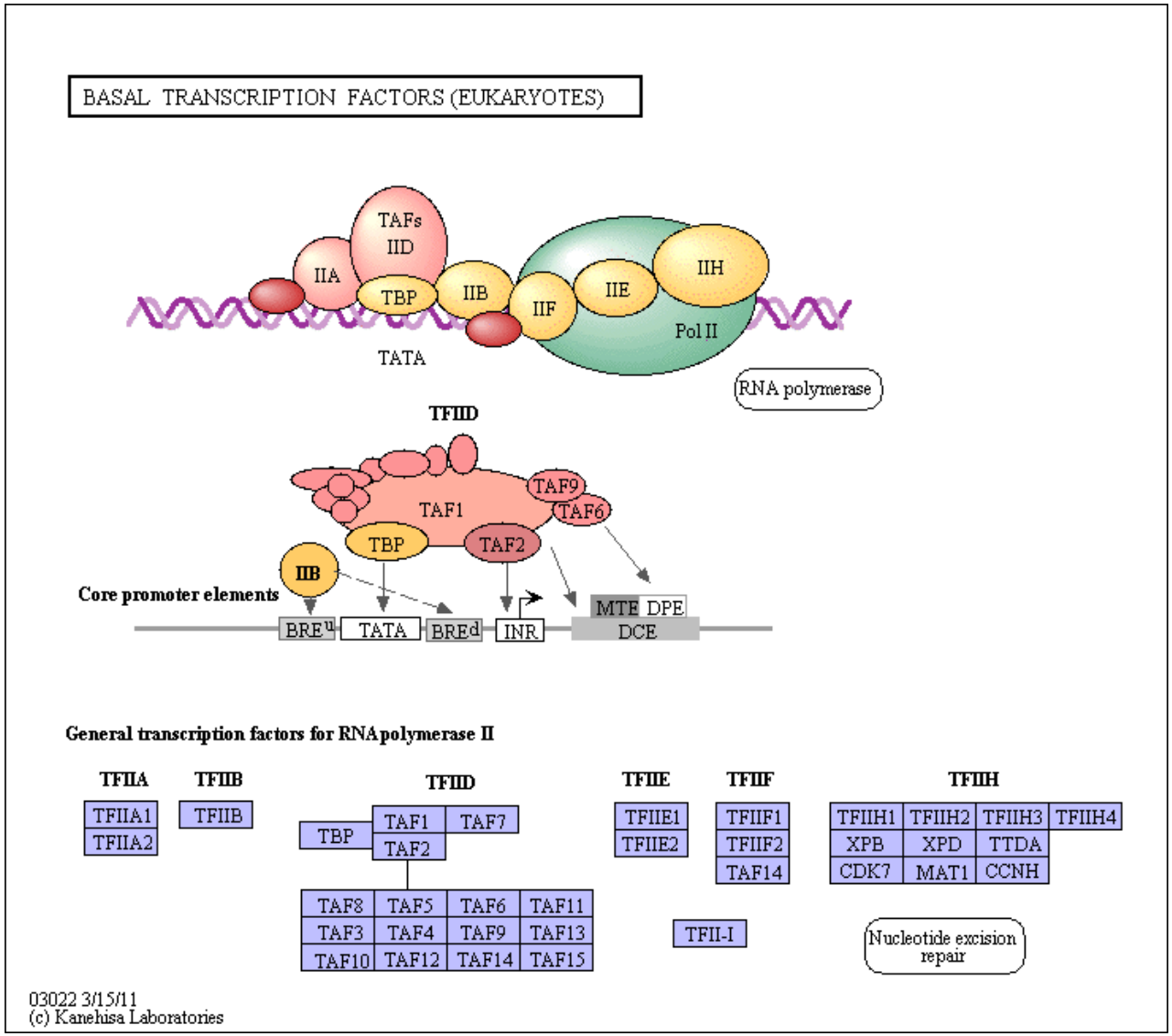
Promoter sequences - usually have two components:

Proximal component - TATA box (directs RNA polymerase II to the correct location)

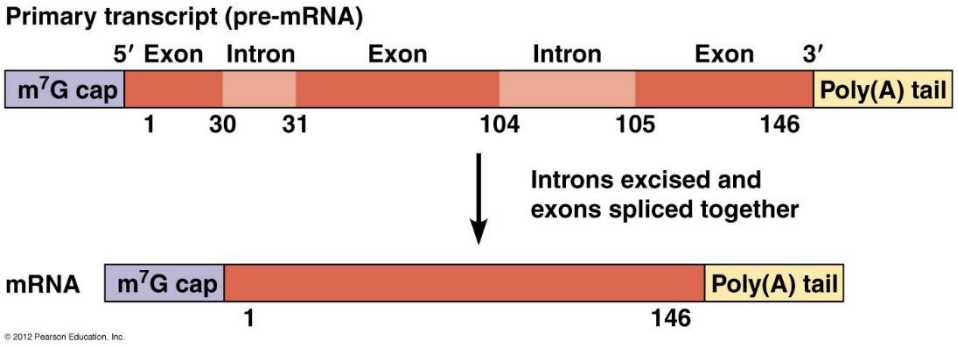
Distal component - CAAT and GC sequences (determines the frequency of transcription initiation)

https://ib-biology2010-12.wikispaces.com/Transcription+and+Translation

General transcription factors



RNA maturation



mRNA is first produced as pre-mRNA (heterogeneous nuclear RNA; hnRNA) and then modified into mature mRNA:

1. Addition of a **5' cap**structure (guanosine nucleotide methylated at position 7) - protects against degradation by 5' exonucleases; helps transcripts bind to ribosomes during protein synthesis

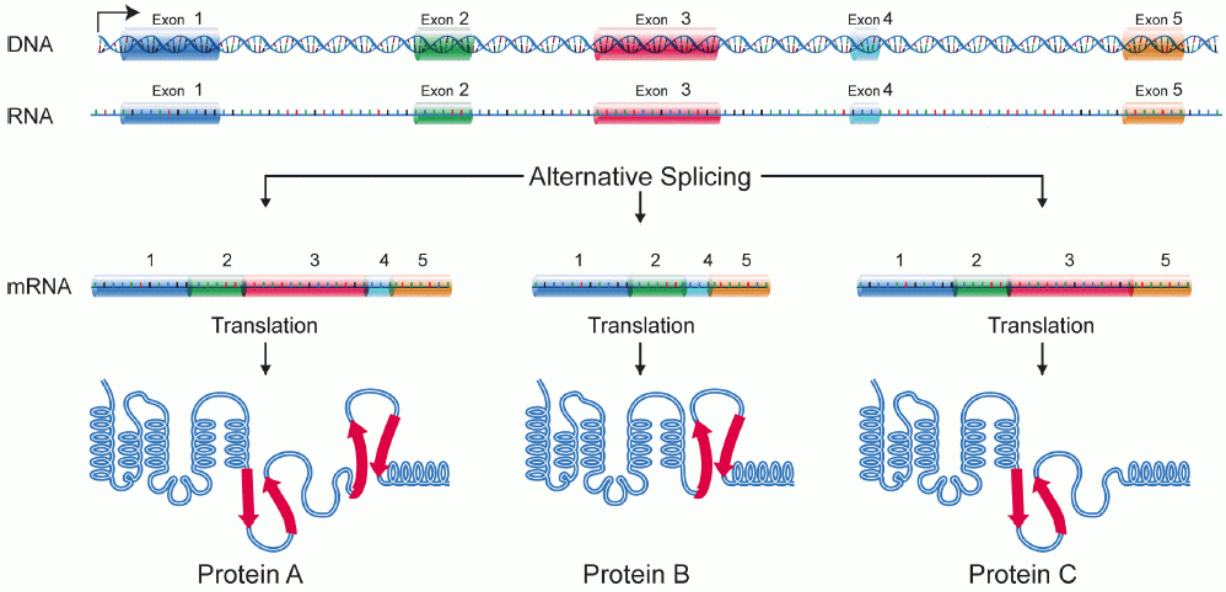
2. Addition of **poly-A tails -** primary transcripts contain a highly conserved AAUAAA sequence known as the polyadenylation signal (near the 3' end); stabilizes the transcript and enables export to the cytoplasm

3. **Splicing** - cutting out introns

https://ib-biology2010-12.wikispaces.com/Transcription+and+Translation

Alternative transcript submission

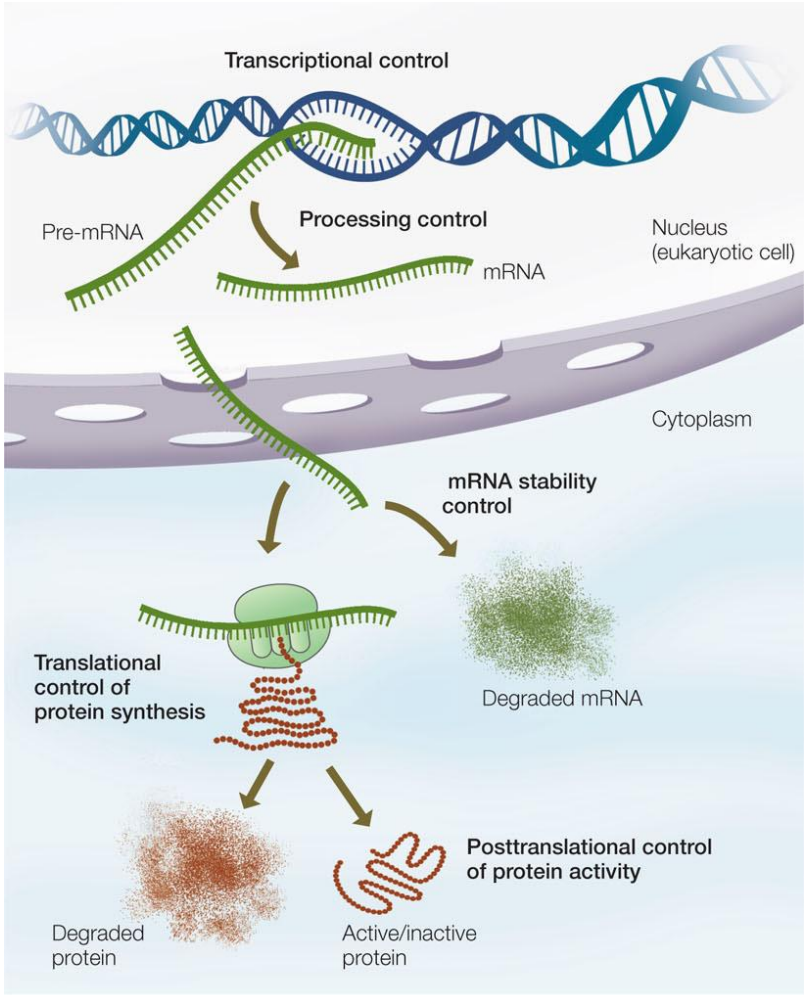
Alternative splicing - alternative forms of mRNA



Proteins produced from alternative mRNAs differ in their amino acid sequence and often also in their function.

Alternative splicing allows for the synthesis of a much larger number of proteins than would be expected from 20,000 protein-coding genes.

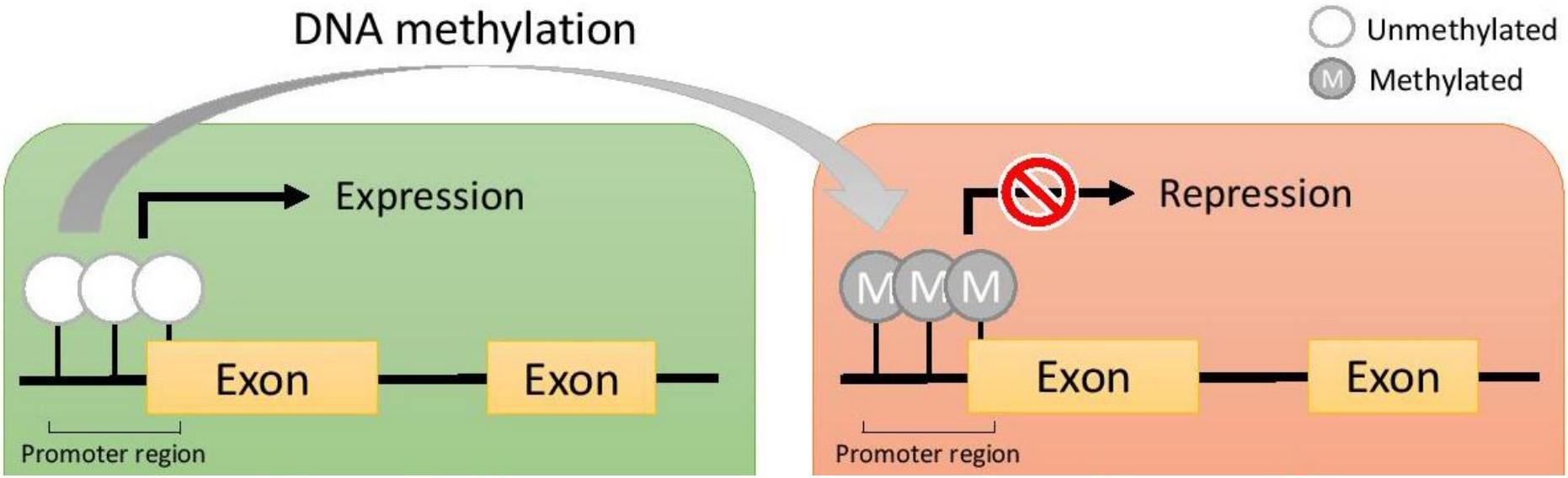
Potential points of gene expression regulation



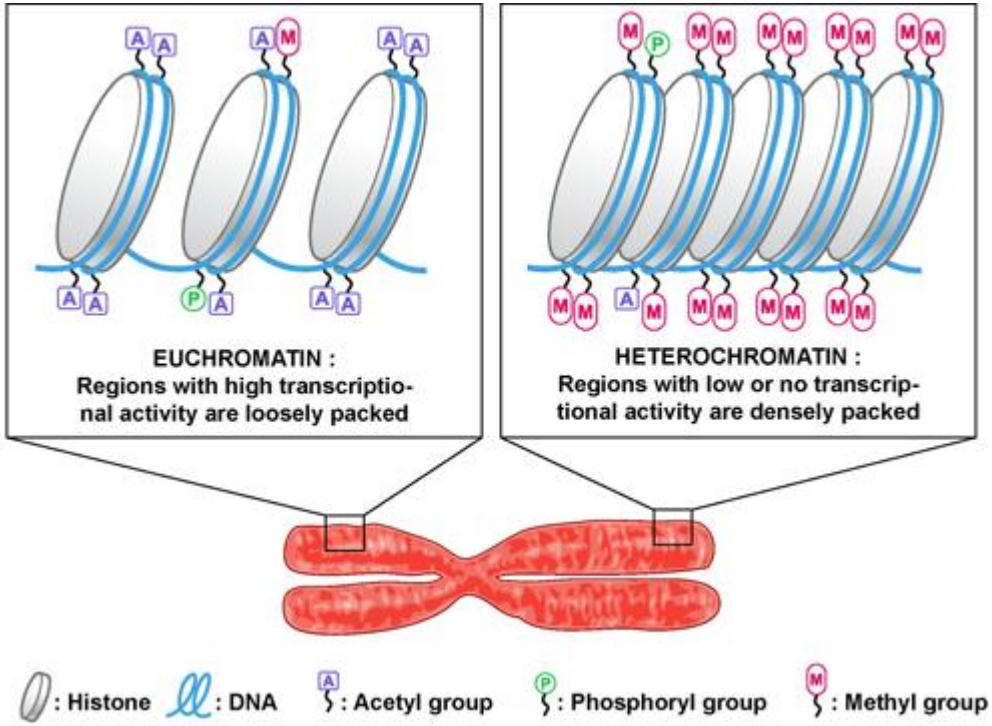
At every stage of the journey from DNA to protein, gene expression can be regulated.

http://www.macmillanhighered.com/BrainHoney/Resource/6716/digital_first_content/trunk/test/hillis2e/hillis2e_ch11_2.html

DNA methylation



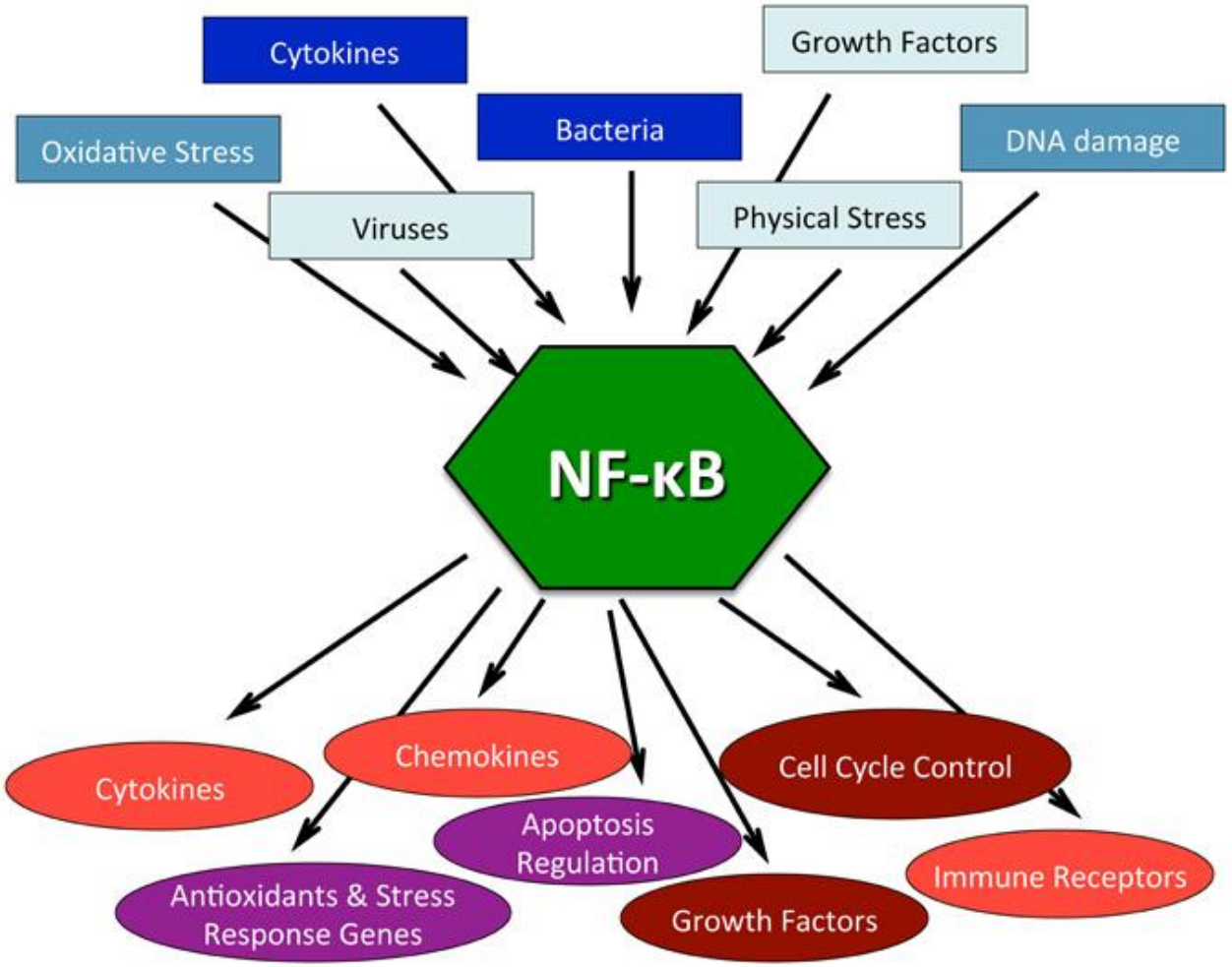
Histone modifications



Post-translational histone modifications – regulate gene expression by organizing the genome into active euchromatin regions, where DNA is accessible for transcription, or inactive heterochromatin regions, where DNA is more compact and less accessible for transcription.

https://www.shmoop.com/dna/dna-packaging.html

Specific transcription factors – NF-kappaB



RNA interference

Wild-type
V26 petunia

Flowers of chalcone synthase
silenced transformants

Craig C. Mello and Andrew Fire's 1998 *Nature* paper reported a potent gene silencing effect after injecting double stranded RNA into *C. elegans*. In investigating the regulation of muscle protein production, they observed that neither mRNA nor antisense RNA injections had an effect on protein production, but double-stranded RNA successfully silenced the targeted gene. As a result of this work, they coined the term *RNAi*.

RNA interference (RNAi) is a natural mechanism of gene regulation in which short RNA molecules—mainly siRNA or miRNA—silence gene expression. This occurs because they bind to complementary mRNA and lead to its destruction or block translation, preventing protein formation. This process plays an important role in defending cells against viruses and transposons and in precisely controlling gene activity.

Intrinsic Damaging Agents
(ROS, replication mistakes, replication stress, alkylating chemicals, & metabolically-derived aldehydes)

Extrinsic Damaging Agents
(radiation, environmental chemicals, & chemotherapeutic agents)

Damaged DNA

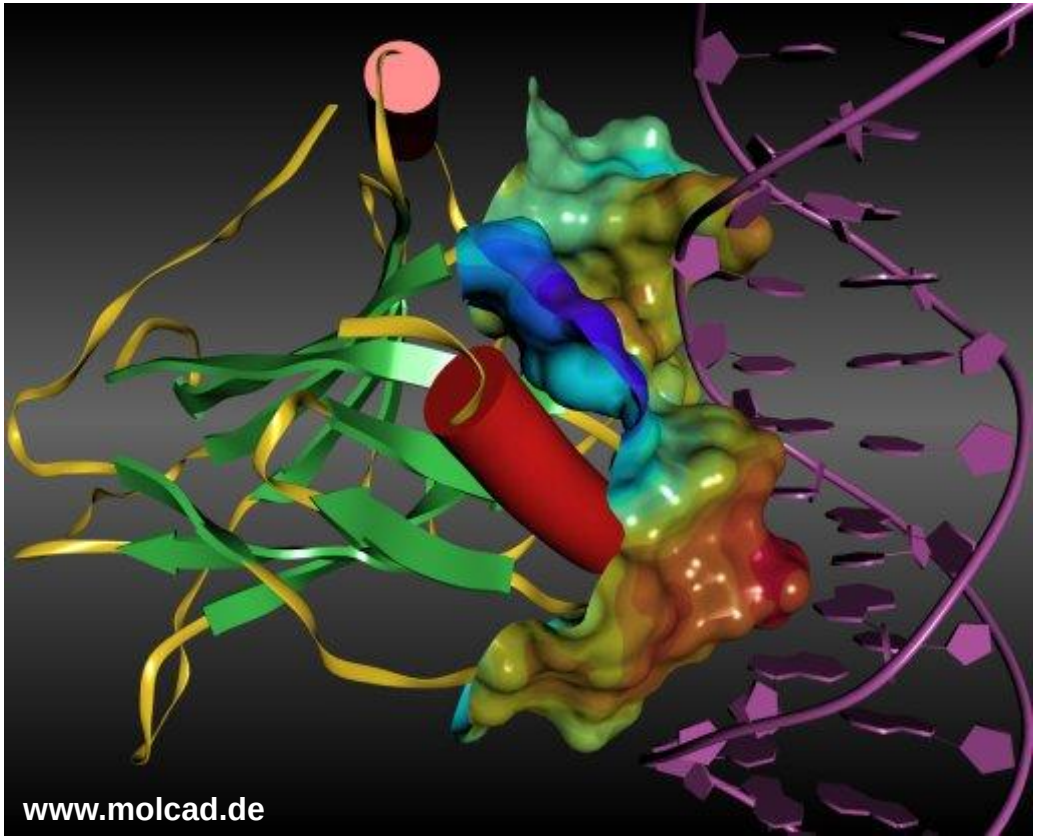
Idle Damage | **Active Damage**

DNA Damage	Mismatch	Base Lesion	"Bulky" Adduct	SSB	DSB
Molecular Effect	Mutagenesis	Mutagenesis	Blocked transcription Blocked replication Mutagenic bypass	Blocked transcription Faithful recombination	Blocked transcription Replication fork collapse Genomic instability
Cellular Consequence	Transformation	Transformation	Transformation Death	Death	Transformation Death
Disease Outcome	Cancer	Cancer Immunodysfunction	Cancer Neurological disease Premature aging	Neurological disease	Cancer Neurological disease Immunodysfunction
Responsible Repair Pathway	MMR	BER	NER	SSBR	DSBR

The American Journal of Human Genetics 105, 237–257, August 1, 2019

p53 guardian of genome integrity

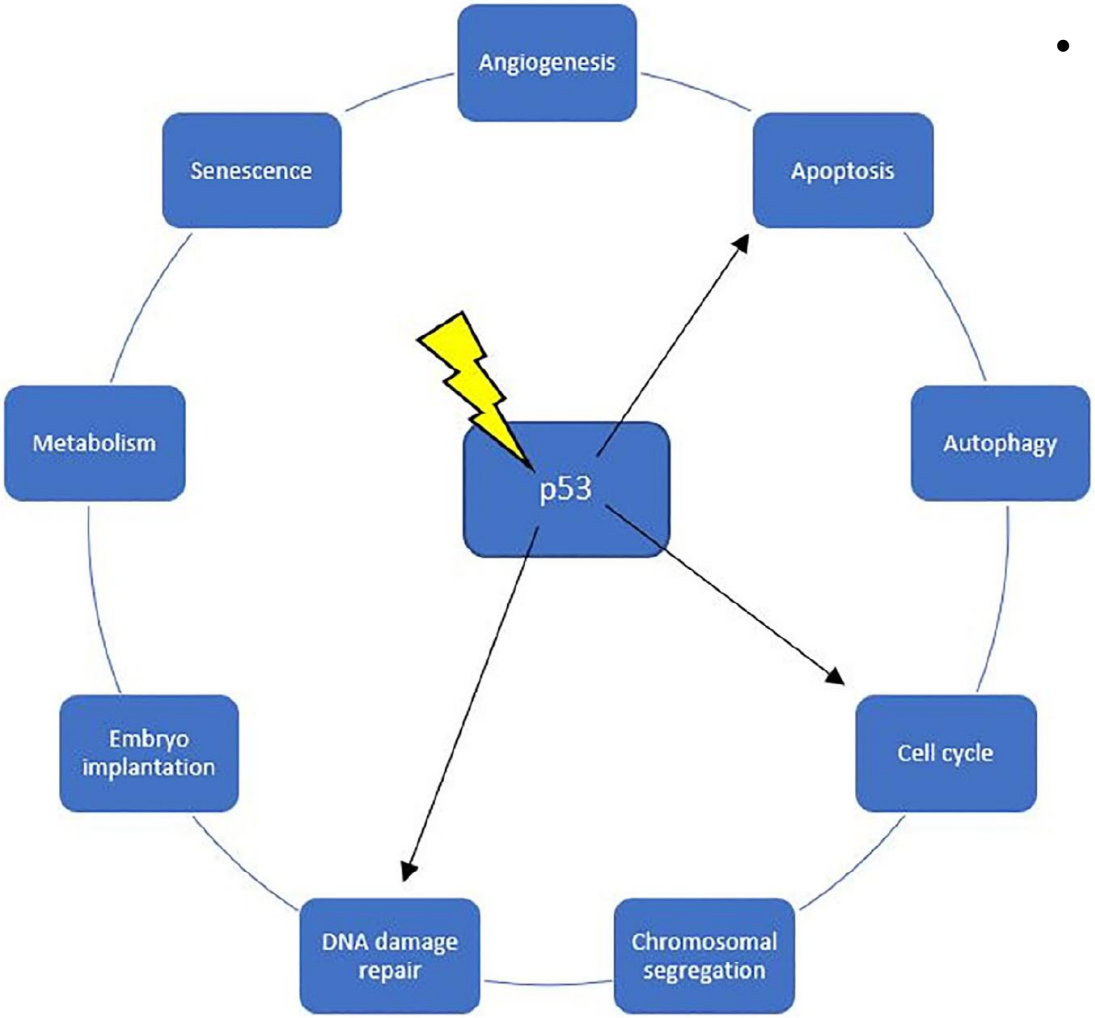
coordinates the cell's response to genotoxic stress and maintains genome integrity



- p53 is a tumor suppressor
- one of the most important genes in cancer biology
- prevents the proliferation of cells with tumorigenic potential
- intensifies apoptosis, cell cycle arrest, and DNA damage repair in normal cells
- directly regulates ~ 500 genes
- the most common mutations of this gene in cancers

p53

involved in a wide range of cellular functions



- The p53 protein is kept at low levels until genotoxic stress occurs - it stimulates the stabilization and accumulation of the protein
  - p53 accumulation occurs in response to DNA damage
  - p53 is regulated in cells not subjected to any stress at the beginning of the S phase to initiate a rapid response when damaged DNA is detected

Cell Biochem Funct. 2020;1–13.

Li Fraumeni Syndrome

Frederick Pei Li and Joseph F. Fraumeni, Jr

- genetically determined syndrome predisposing to cancer
- mutations in the *TP53* gene – autosomal dominant inheritance
- increased risk of cancer in children and young people, particularly soft tissue and bone sarcomas, breast cancer, acute leukemia, brain tumors, adrenal cortex cancer
- other cases include cancers of the gastrointestinal tract, lungs, kidneys, thyroid, skin, and reproductive organs (ovaries, testicles, prostate)