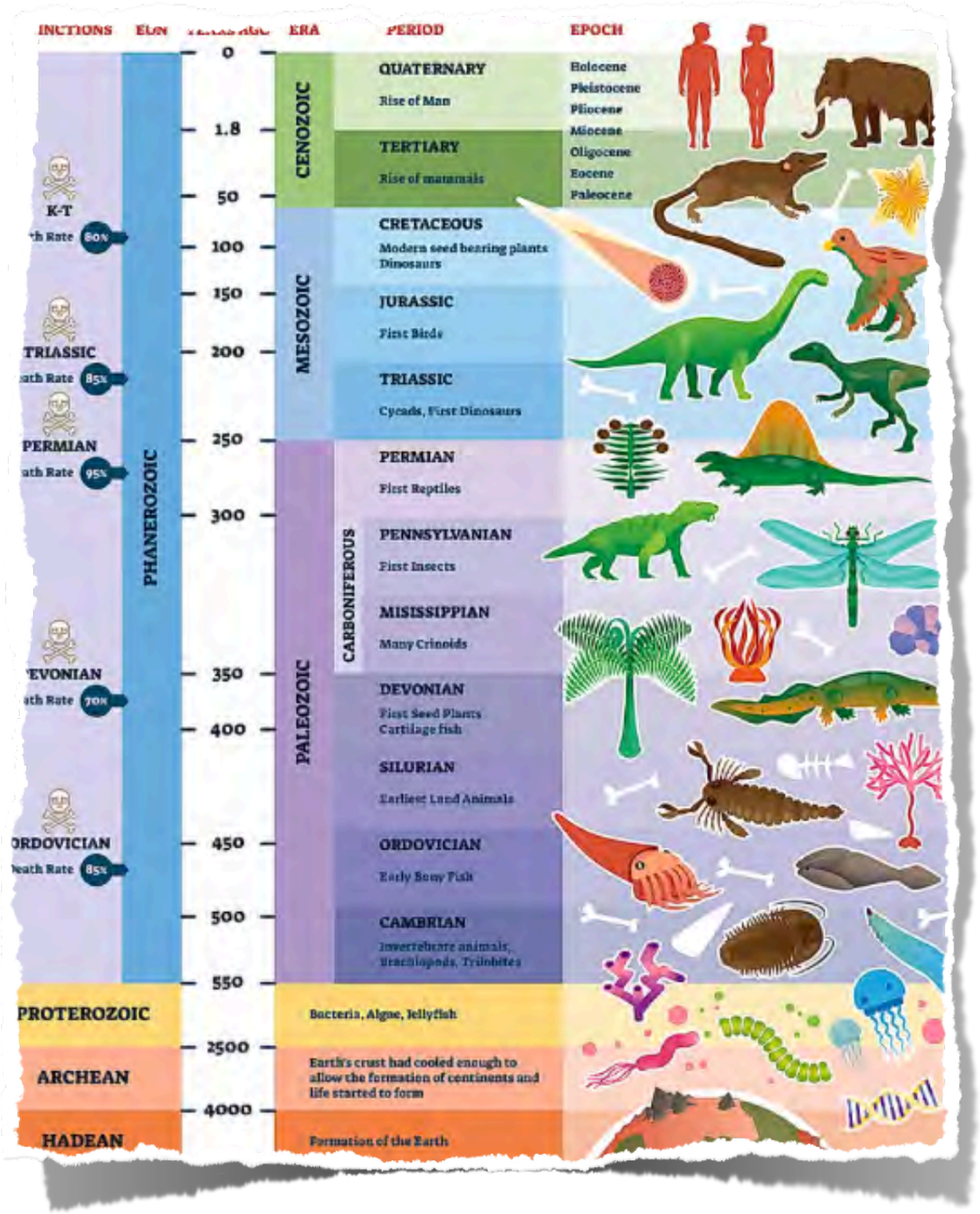


Have you ever wondered how life evolved, evolves, sustains and replicates ?

Biology and psychiatry: Epigenetics for beginners

Dr. Ashutosh Shah D.N.B.
Consultant Psychiatrist

Only basics will be covered.

Clinical translation unlikely in immediate future.

WARNINGS

Only basics will be covered.

Clinical translation unlikely in immediate future.

WARNINGS

AGENDA

- Evolution of life.
- Cell.
- Genetics.
- Epigenetics.

AGENDA

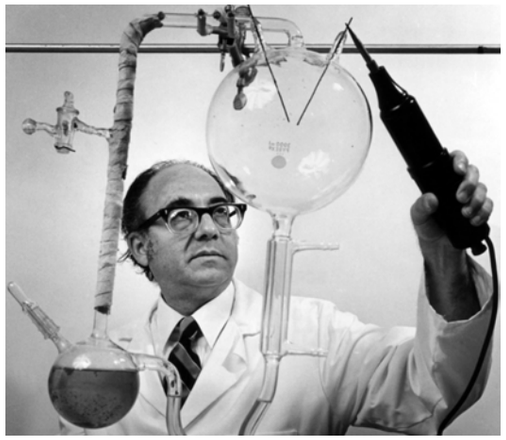
- Evolution of life.
- Cell.
- Genetics.
- Epigenetics.

Evolution of life

Primitive Earth atmosphere had water, electrical storms, molecular hydrogen, nitrogen and carbon. Molecular oxygen was absent.

Nitrogen, water, methane, ammonia, carbon dioxide and carbon monoxide were the raw materials for the basic chemistry of life. Energy for chemical reaction was provided by the sun's ultraviolet radiation plus energy from electrical storms and volcanic heat.

This hypothesis was first tested experimentally in 1953 by Stanley Miller: after a week, contents of the flask revealed a number of organic compounds, including amino acids.



Similar experiments showed other biologically important molecules could be synthesised from cyanides and aldehydes. Nitrogen bases, an important component of nucleic acids, can be produced from hydrogen cyanide and ammonia.

Evolution of life

The first living organism was likely a large nucleic acid with a cell-like structure. These first cell-like organisms depended on the soup of organic compounds for their own structure and to provide the food for energy.

Once the soup was used up, a new source of energy in form photosynthesis evolved wherein energy was captured from sunlight and a high-energy phosphate bond in form of ATP was produced.

When ATP released the energy from the phosphate it got converted to ADP, and the released energy helped manufacture sugars from carbon dioxide and hydrogen and the end product was oxygen gas.

Most of the oxygen in our present atmosphere came from photosynthetic reactions. Once oxygen had accumulated in the atmosphere, a more complex form of cell evolved, that used oxygen to help break down fats and sugars to provide energy.

This was a more advanced and efficient way of providing energy. Once this more complex and efficient form of cell developed, then it paved the way for evolution of the complex forms of life that exist today.

Evolution of life

One of the properties of life is its organisation. Any living organism is a coordinated organisation of structural units - the cell.

A cell is composed of large polymers which have an inherent capacity to organise themselves into a structure and to replicate themselves.

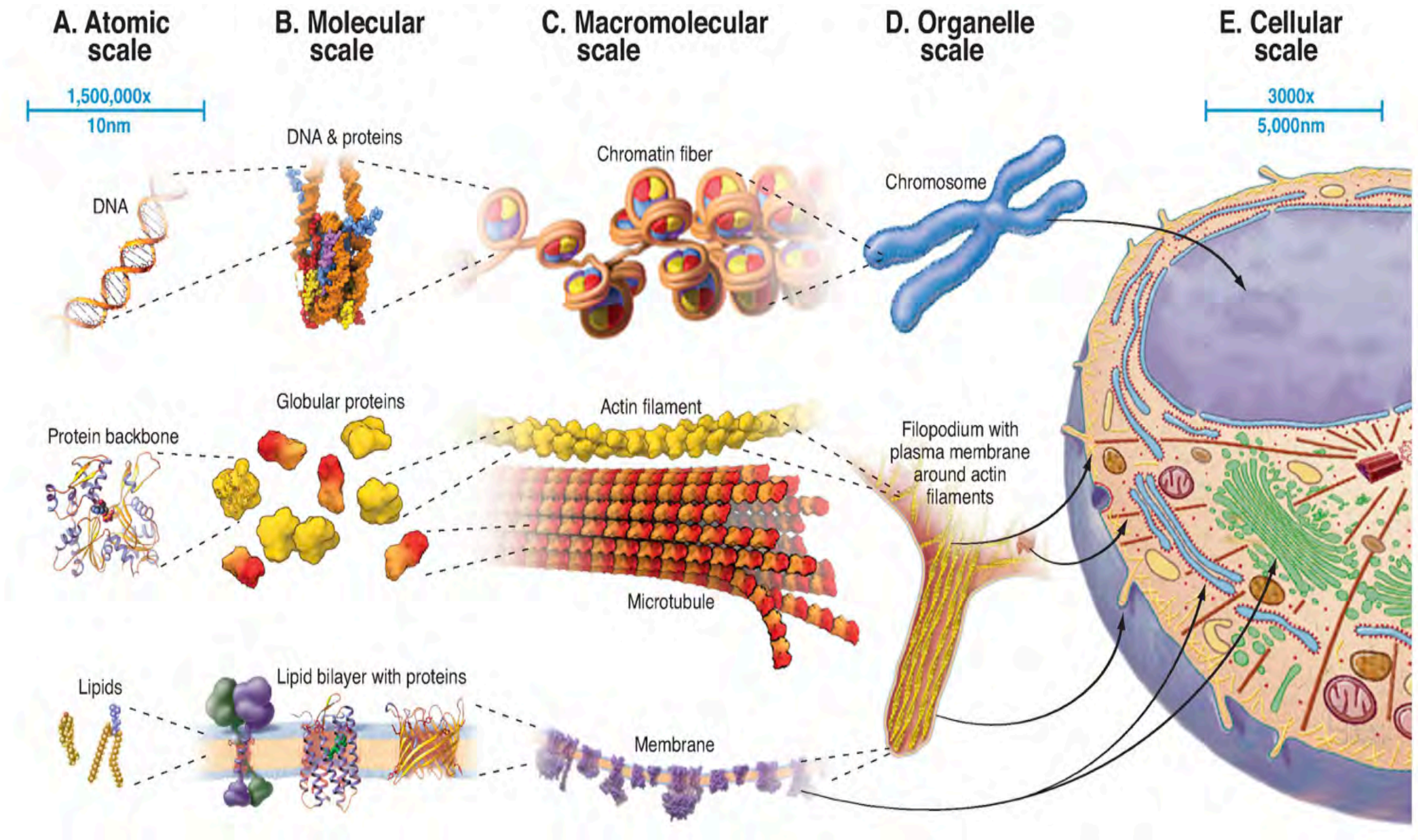
These polymers are composed of small organic molecules that arose originally about 4.5 billion years ago from simple combinations of hydrogen, oxygen, nitrogen, and carbon.

The unique characteristics of carbon atoms enable them to bond to each other successively to form long stable chains. Because of these inherent properties of the carbon atom, the evolution of life occurred.

AGENDA

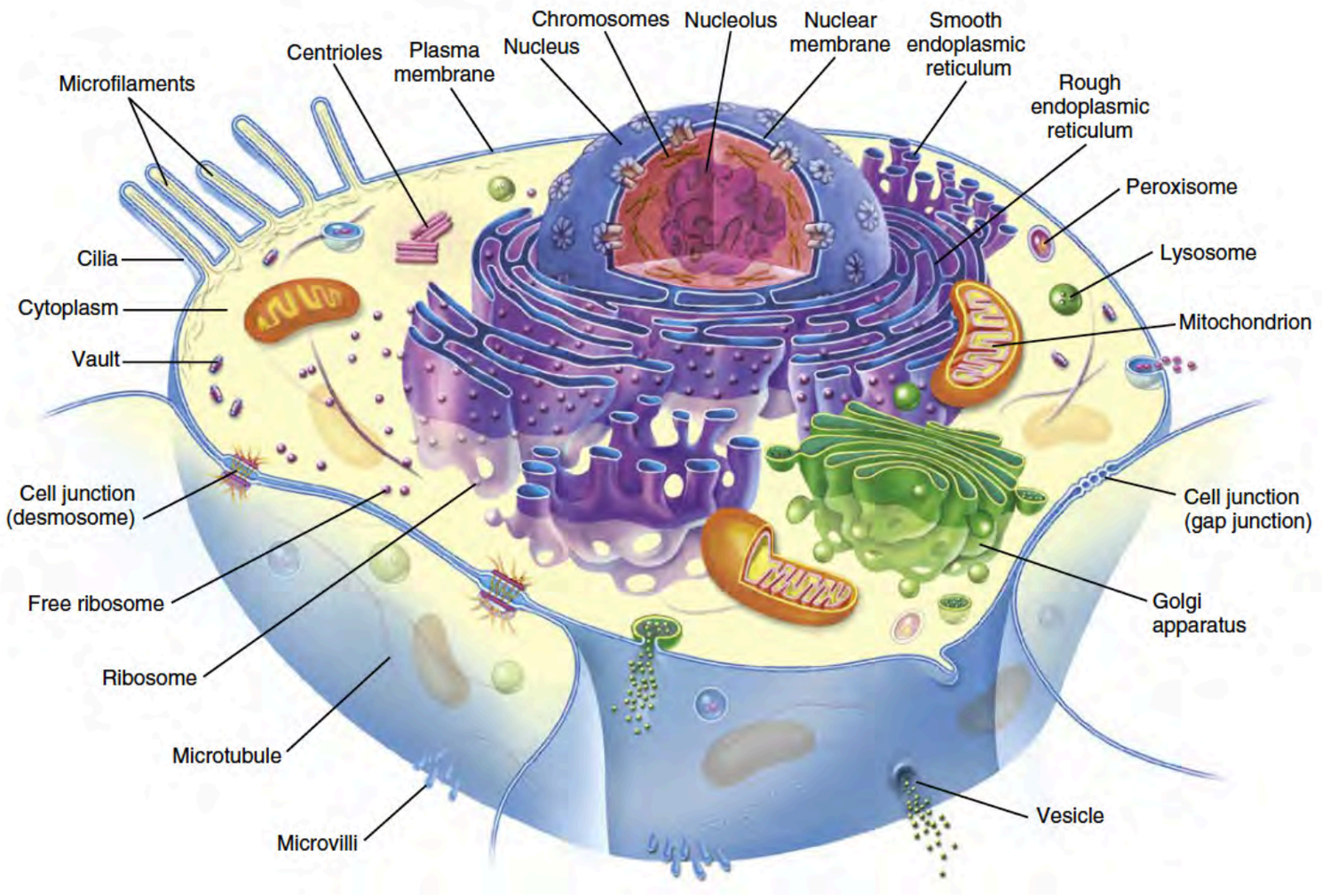
- Evolution of life.
- Cell.
- Genetics.
- Epigenetics.

Cell - the basic unit of life.



Genetic disease involves defects at the cell level.

Cell - the basic unit of life.



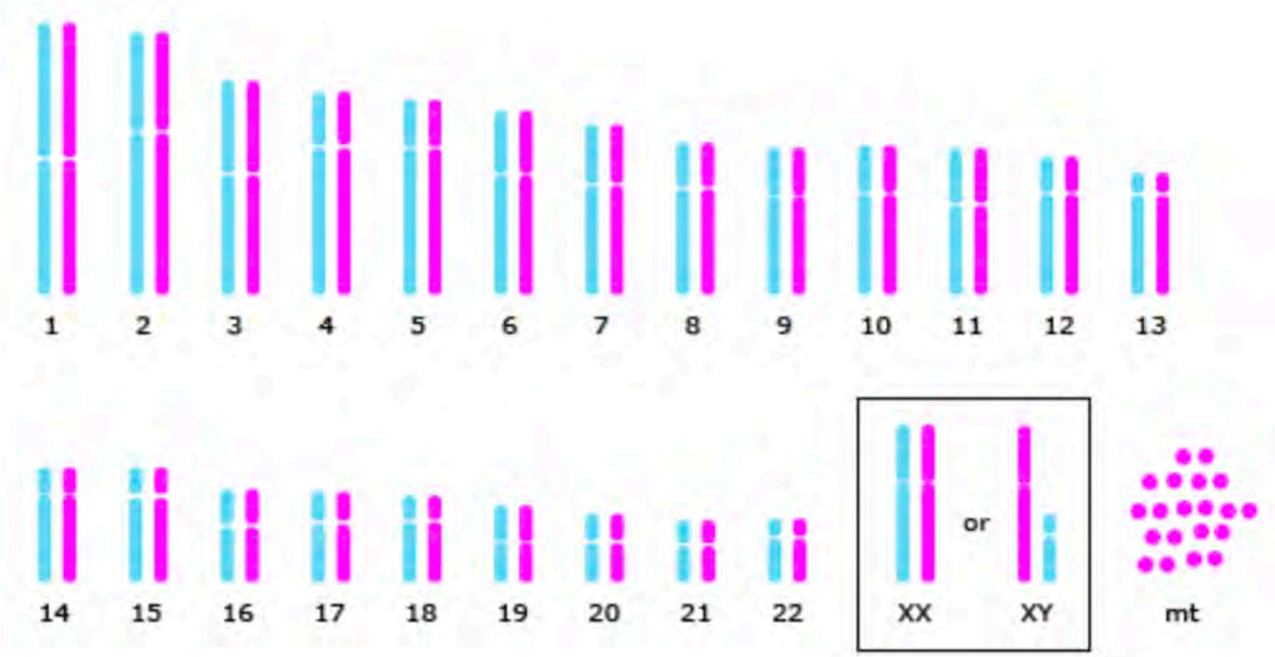
The body is a social order of about 100 trillion cells organised into various functional structures, the largest of which are called organs.

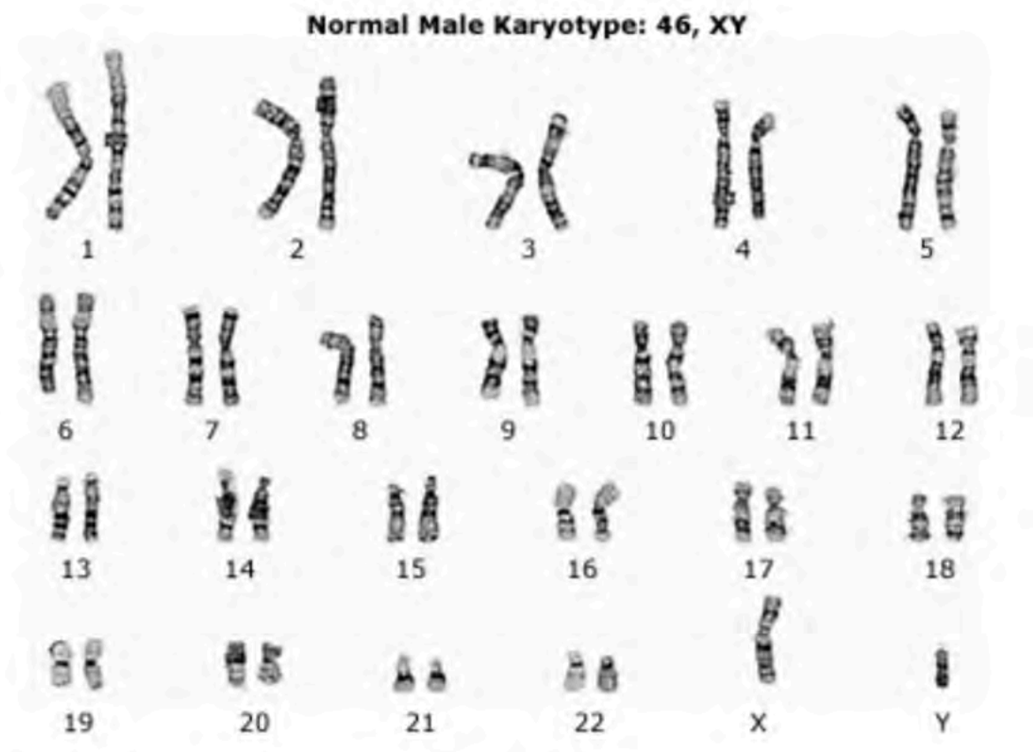
AGENDA

- Evolution of life.
- Cell.
- Genetics.
- Epigenetics.

HUMAN GENOME

- Human genome is DNA organised as 23 chromosomes: 22 autosomes (1-22), and 1 sex chromosome (X or Y).
- Human somatic cells are diploid, consisting of two sets of 23 chromosomes, one paternally inherited (blue) and one maternally inherited (pink).
- Y chromosome is paternally inherited.
- The mitochondrial genome (mt) is derived solely from mitochondria in the ova and therefore exhibits exclusive matrilineal inheritance (controversial).

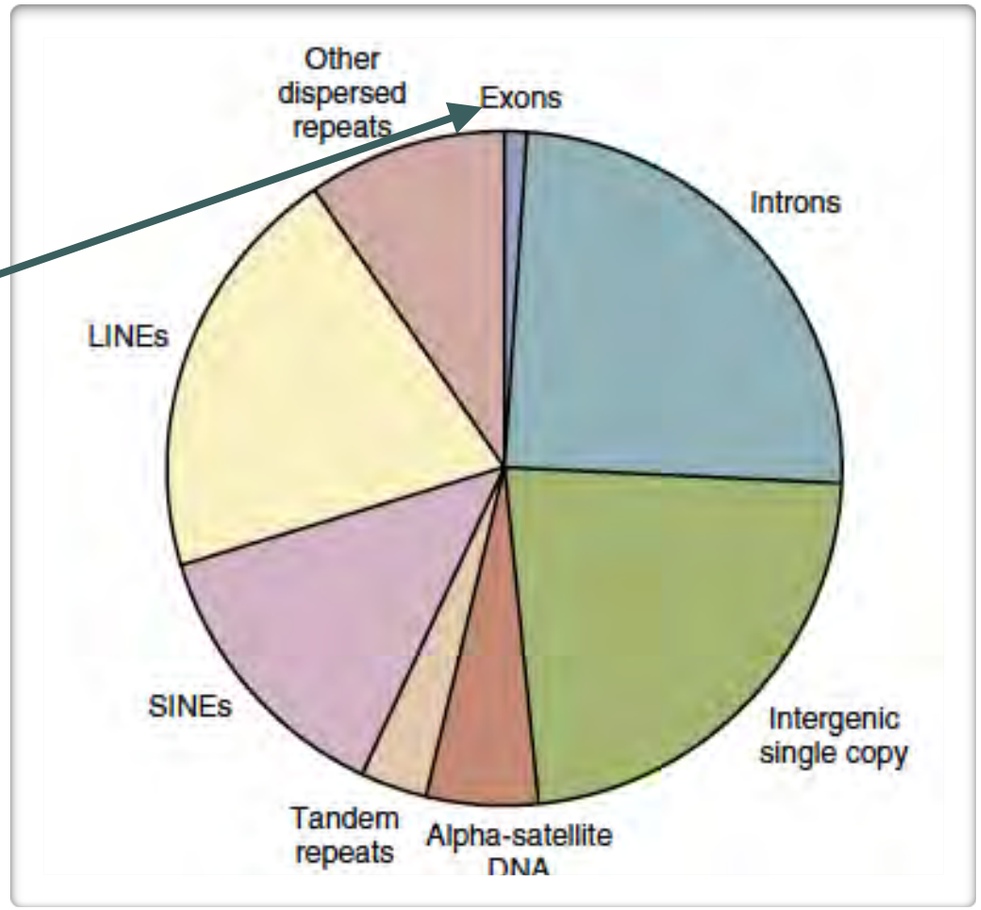




STRUCTURAL COMPOSITION OF HUMAN GENOME

Nuclear genome

- Blueprint for the human genome in every nucleated body cell. Consists of almost 3 billion base pairs.
- Around 1% to 2 % of the nuclear genome encodes for genes that provide instructions for making proteins.
- The total combination of parts of the genomes that encode for genes is known as the exome.

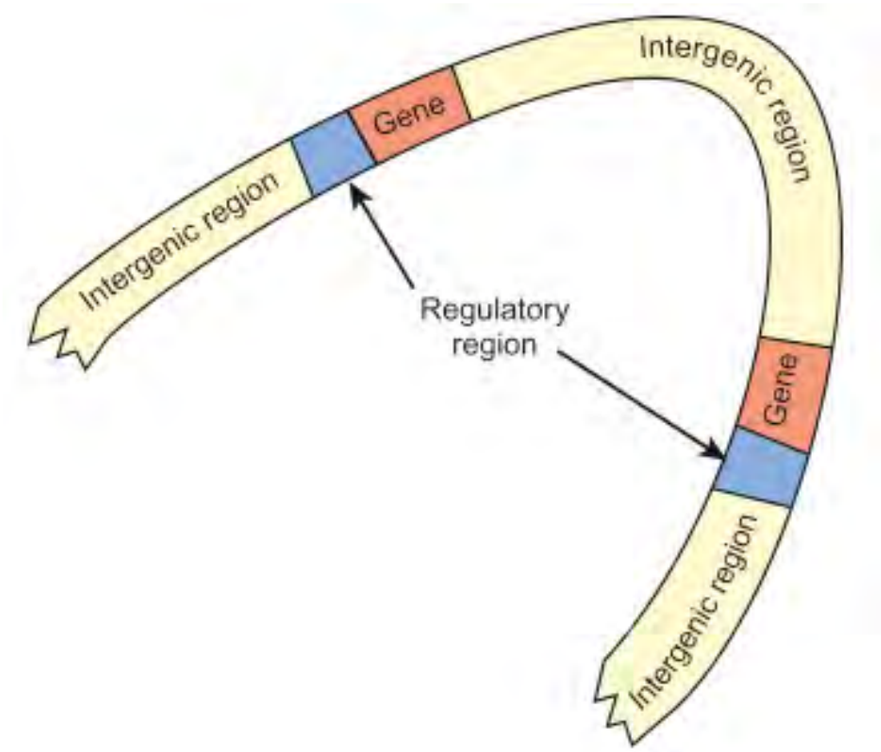


Structural composition of the human genome. (figures are an approximate.)

STRUCTURAL COMPOSITION OF HUMAN GENOME

Nuclear genome

- Blueprint for the human genome in every nucleated body cell. Consists of almost 3 billion base pairs.
- Around 1% to 2 % of the nuclear genome encodes for genes that provide instructions for making proteins.
- The total combination of parts of the genomes that encode for genes is known as the exome.



The General Pattern of Information on a Chromosome

STRUCTURE OF DNA

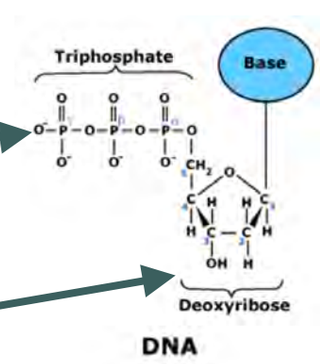
- DNA are composed of three basic building blocks:

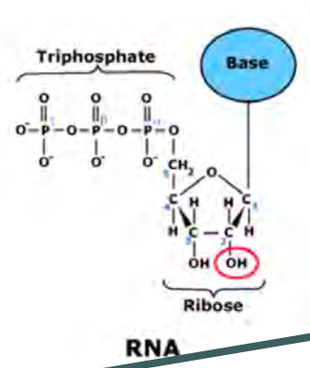
1.phosphoric acid

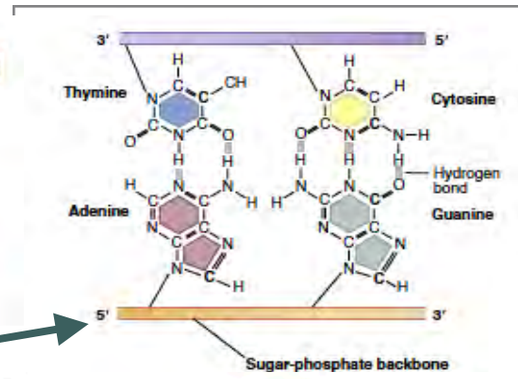
2.deoxyribose (a sugar)

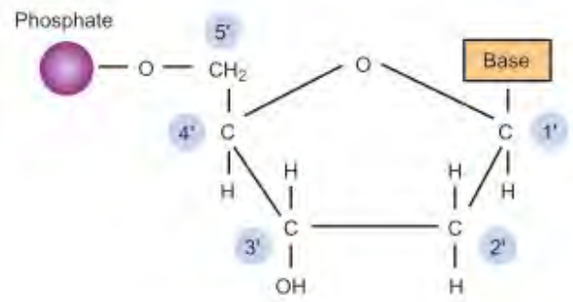
3.four nitrogenous bases: two purines (adenine and guanine) and two pyrimidines (thymine and cytosine).

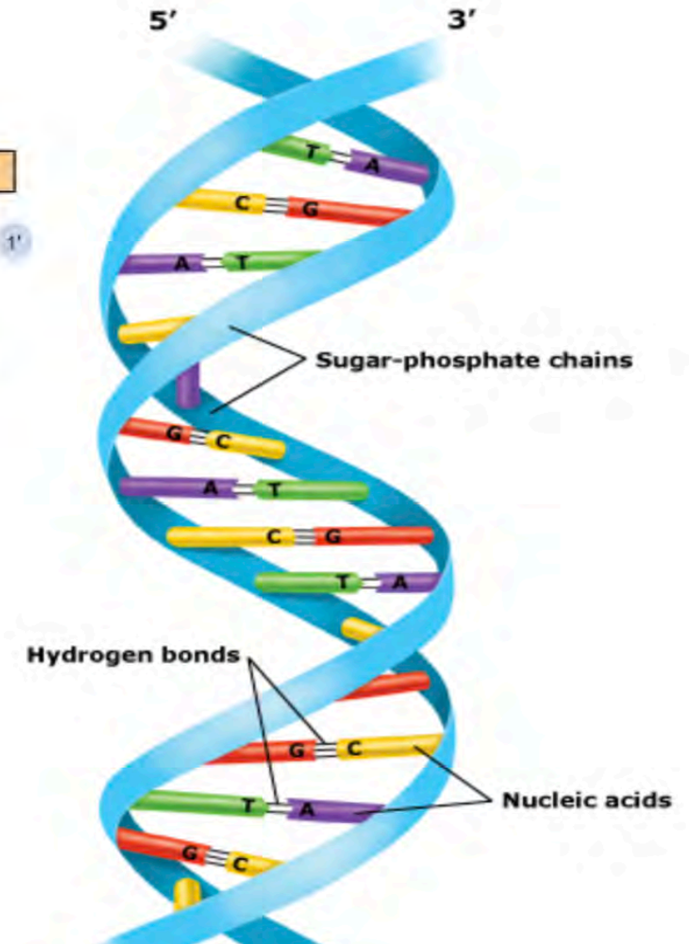
- The four bases are represented by their first letters: C, T, A, and G.









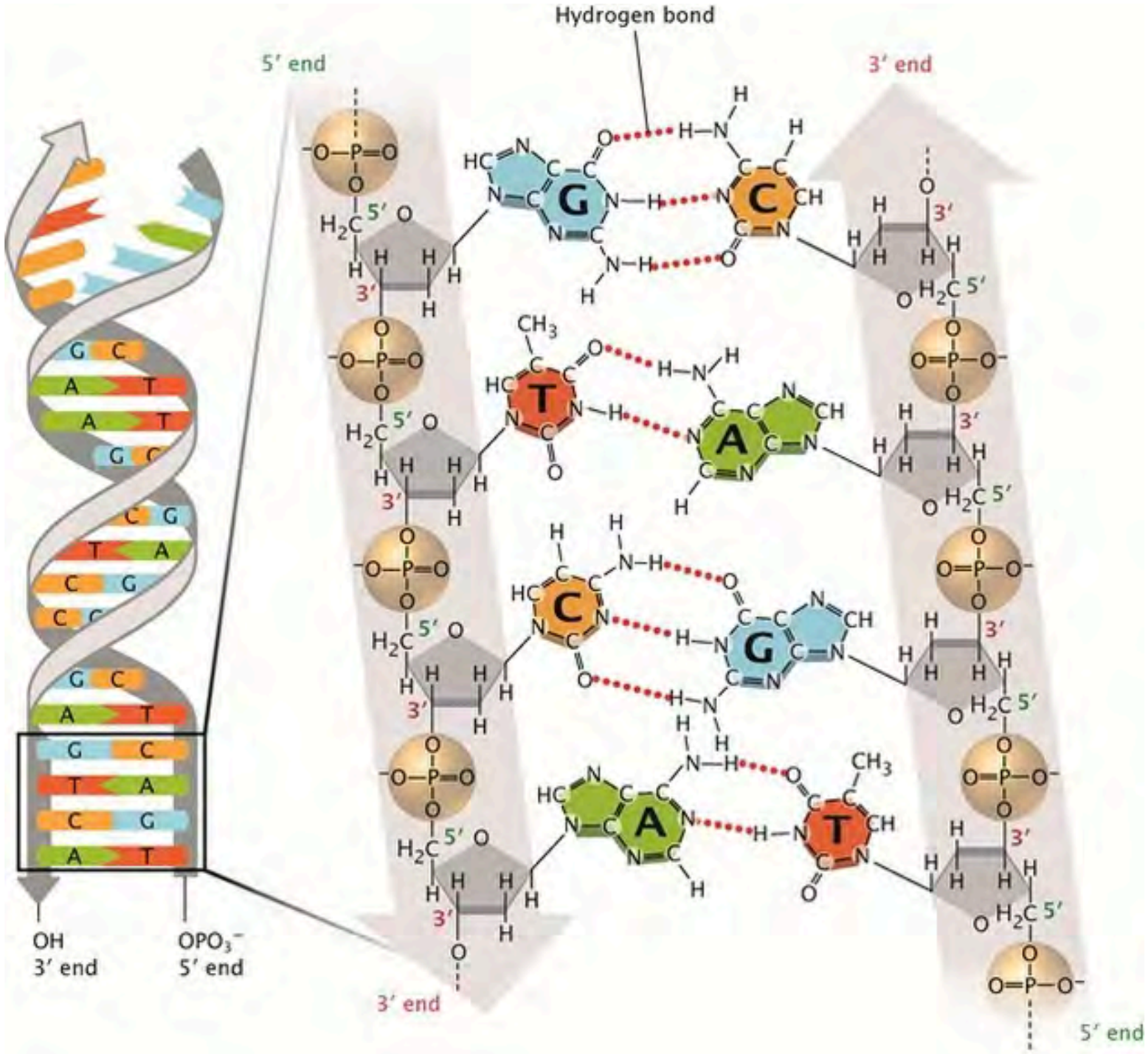


STRUCTURE OF DNA

Two hydrogen bonds connect T to A.

Three hydrogen bonds connect G to C.

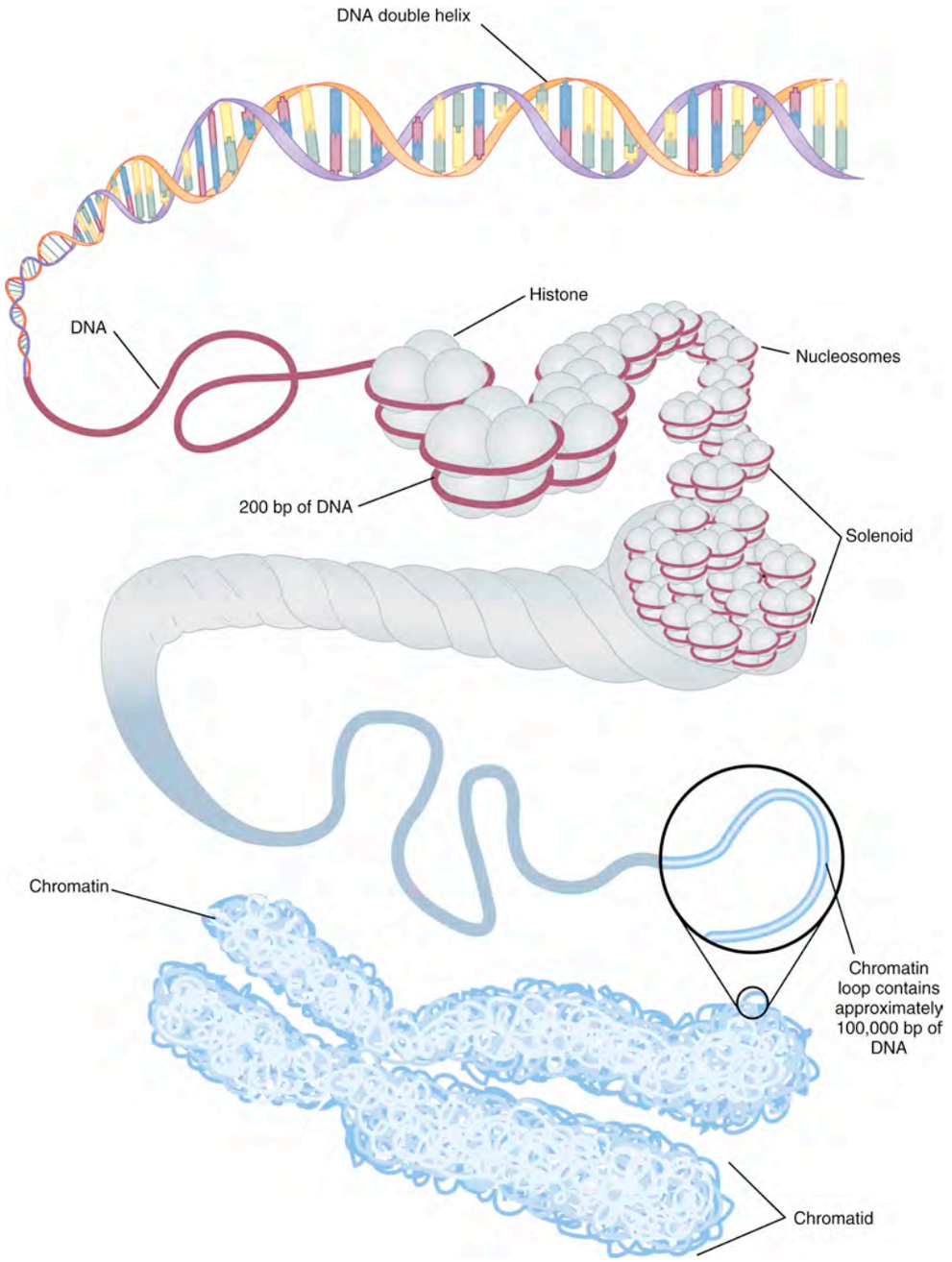
The sugar-phosphate backbones (grey) run antiparallel to each other, so that the 3′ and 5′ ends of the two strands are aligned.



DNA COILING

- > 2 metres of DNA is tightly packed to fit in the nucleus of each human cell.

- End result of coiling and looping makes the DNA, condensed to about 1/10000.



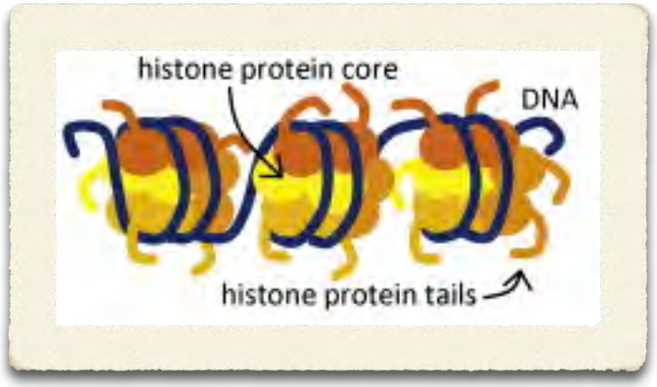
CHROMATIN

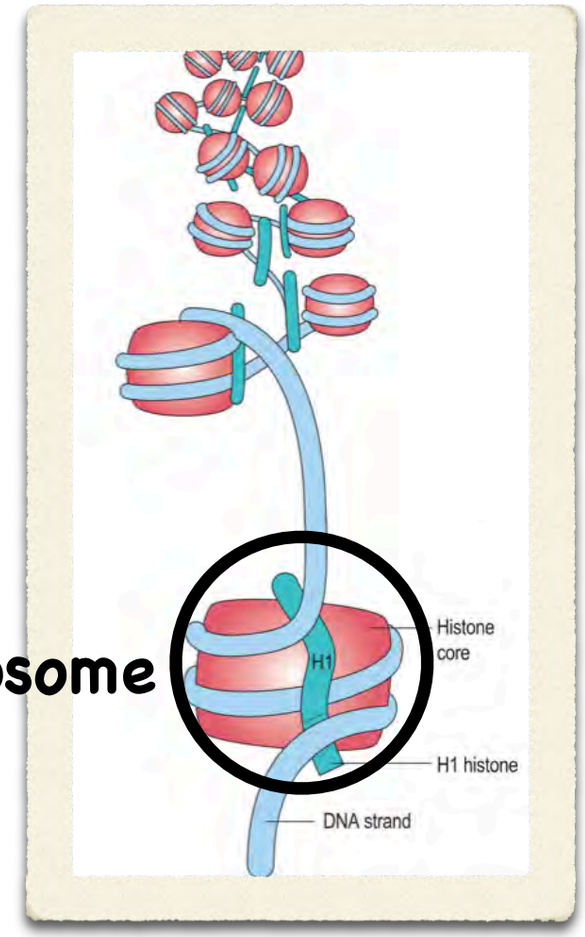
- Native chromosome has DNA mixed with RNA and an equal mass of protein. DNA–RNA–protein complexes are termed **chromatin**.

- The majority of the proteins in chromatin are histones. There are **five classes of histones,** termed H1, H2A, H2B, H3, and H4.

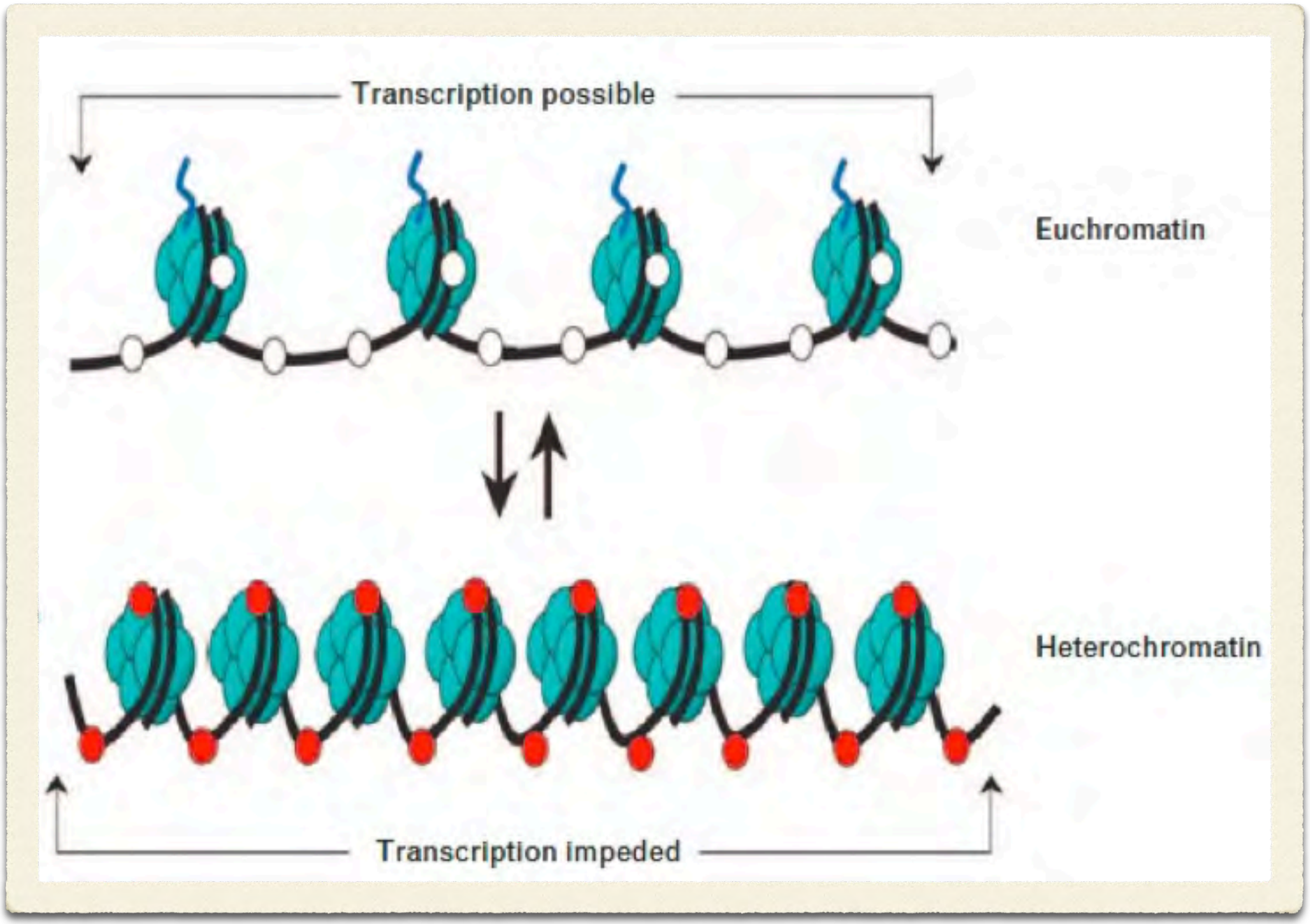
- All histones are positively charged and are rich in basic amino acids (lysine and arginine).

- These positive charges interact with the negatively charged, acidic phosphate groups of the DNA strands to reduce electrostatic repulsion and permit tighter DNA packing.

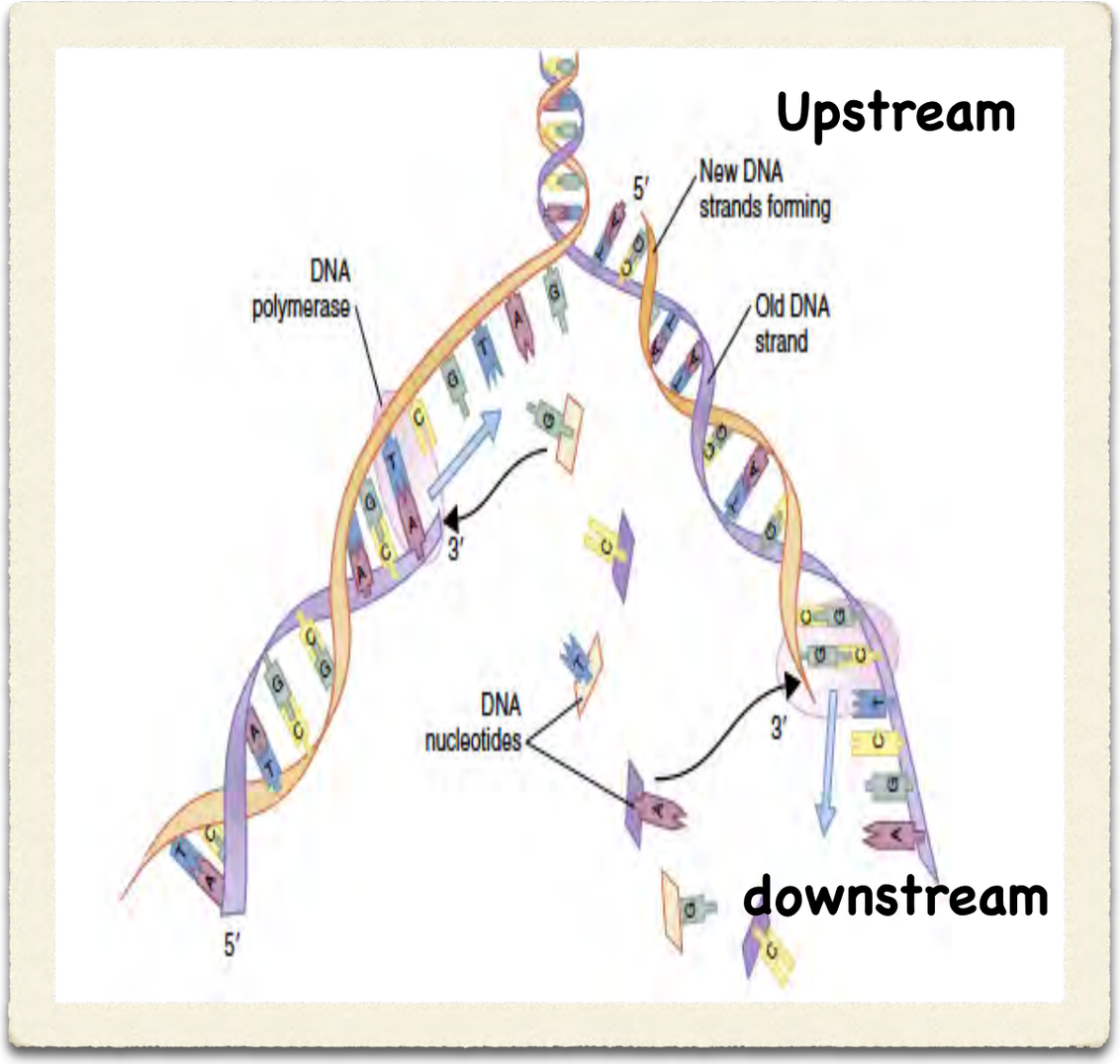




EU/HETERO CHROMATIN

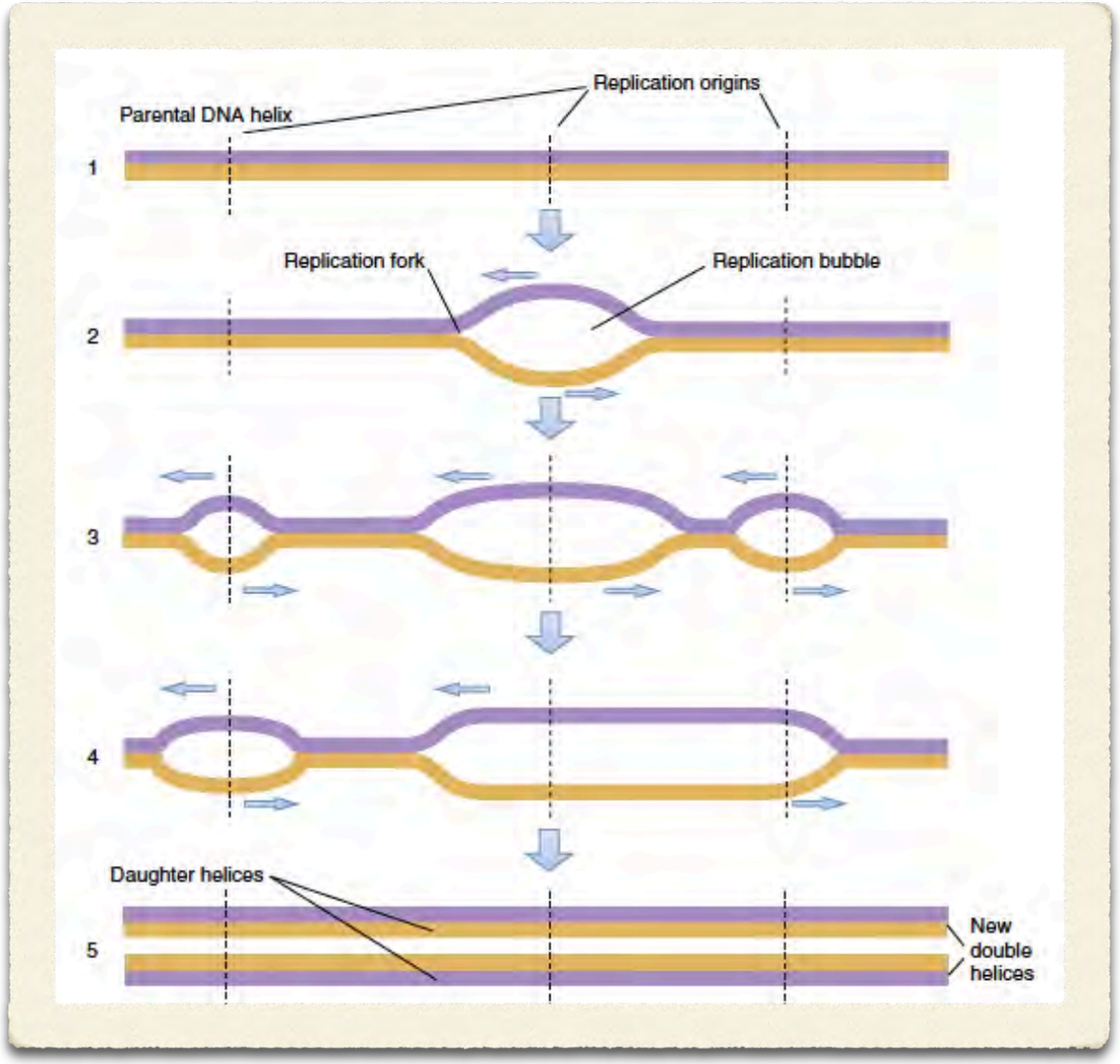


DNA REPLICATION



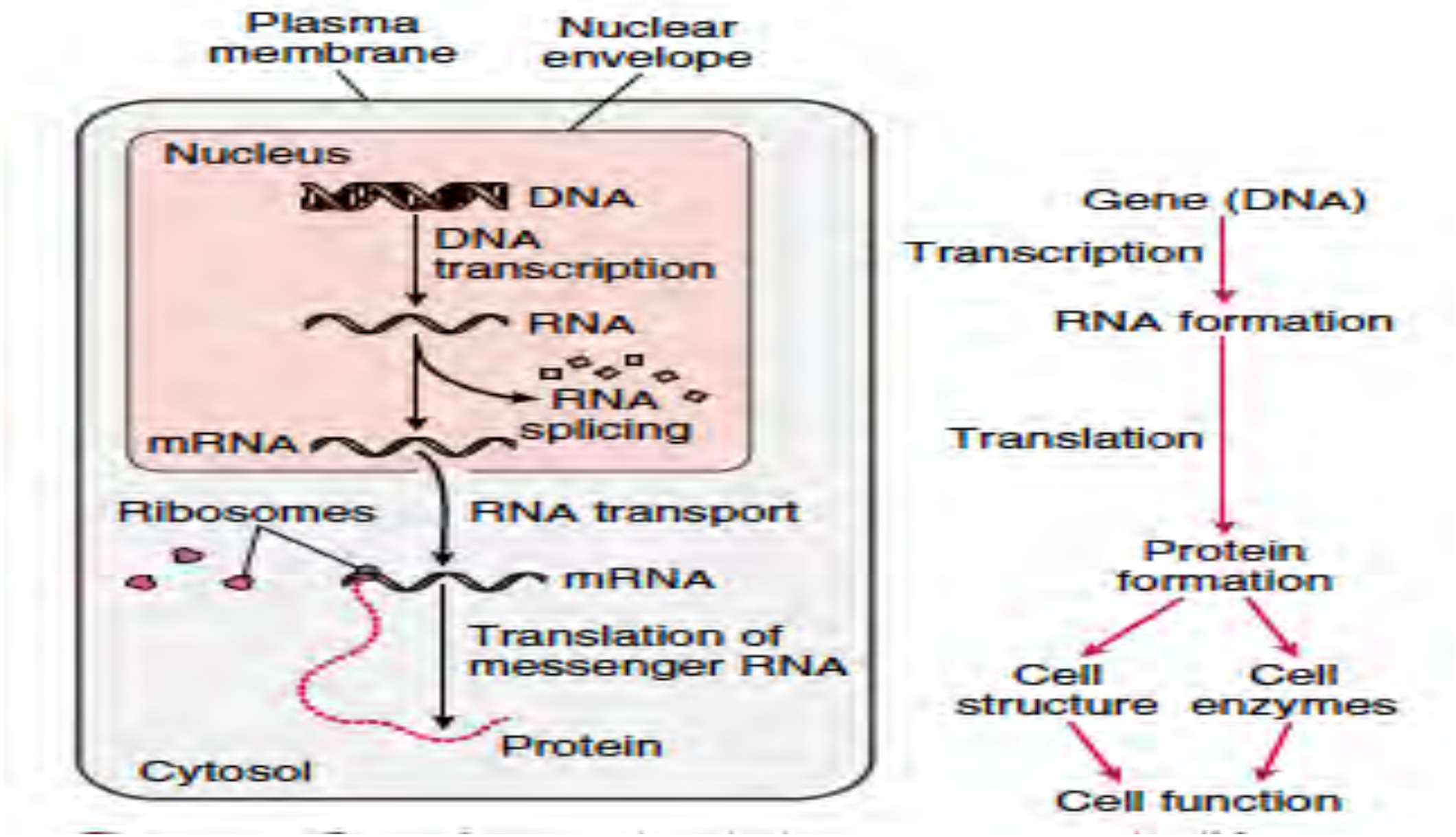
- Complementary base pairing is the key to accurate replication (consistent pairing of A-T and G-C.)
- Several different enzymes are involved in DNA replication:
  - A. one unwinds the double helix,
  - B. another holds the strands apart,
  - C. DNA polymerase adds free nucleotides and performs part of a proofreading procedure .

DNA REPLICATION

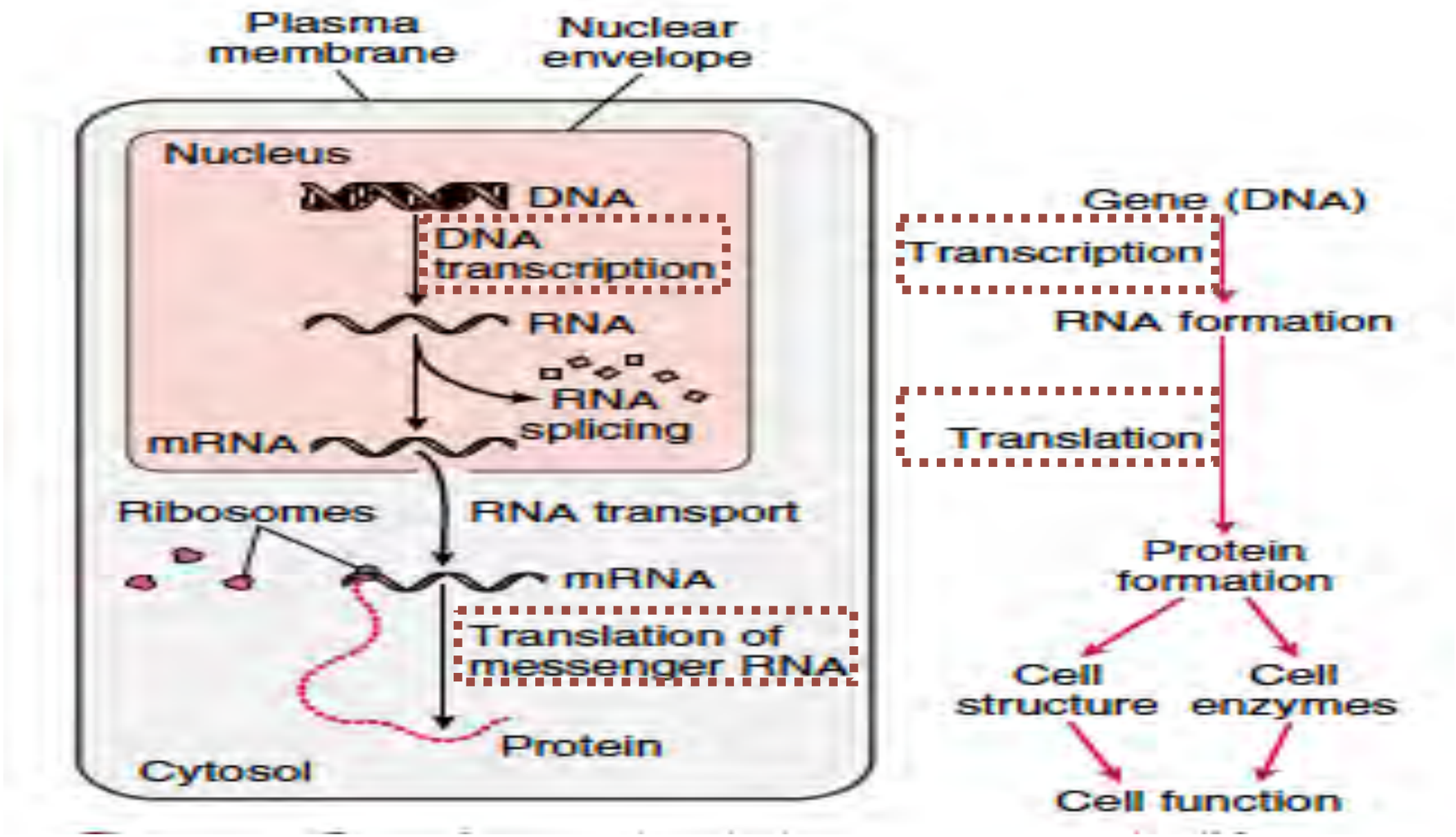


- Humans: rate of DNA replication ~ 40-50 nucleotides/second (relatively slow).
- Bacteria: 500-1000 nucleotides/second !
- Replication in humans would take 2 months if it occurred linearly (~250 million bases).
- Replication begins at several points along a chromosome, termed replication origins.
- The resulting multiple separations of the DNA strands are called replication bubbles.
- Greatly speeds up the replication process.

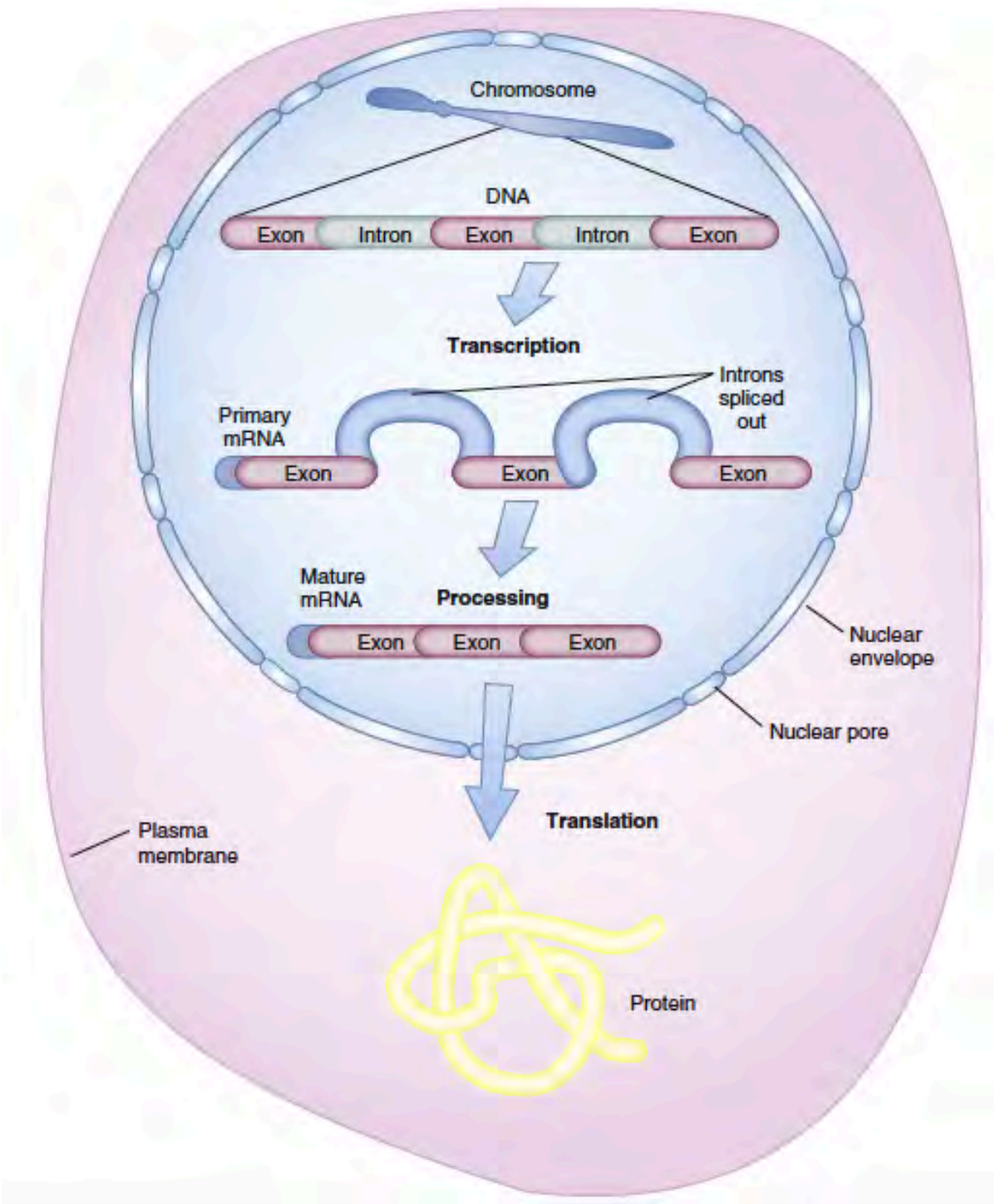
General schema: how genes control cell function



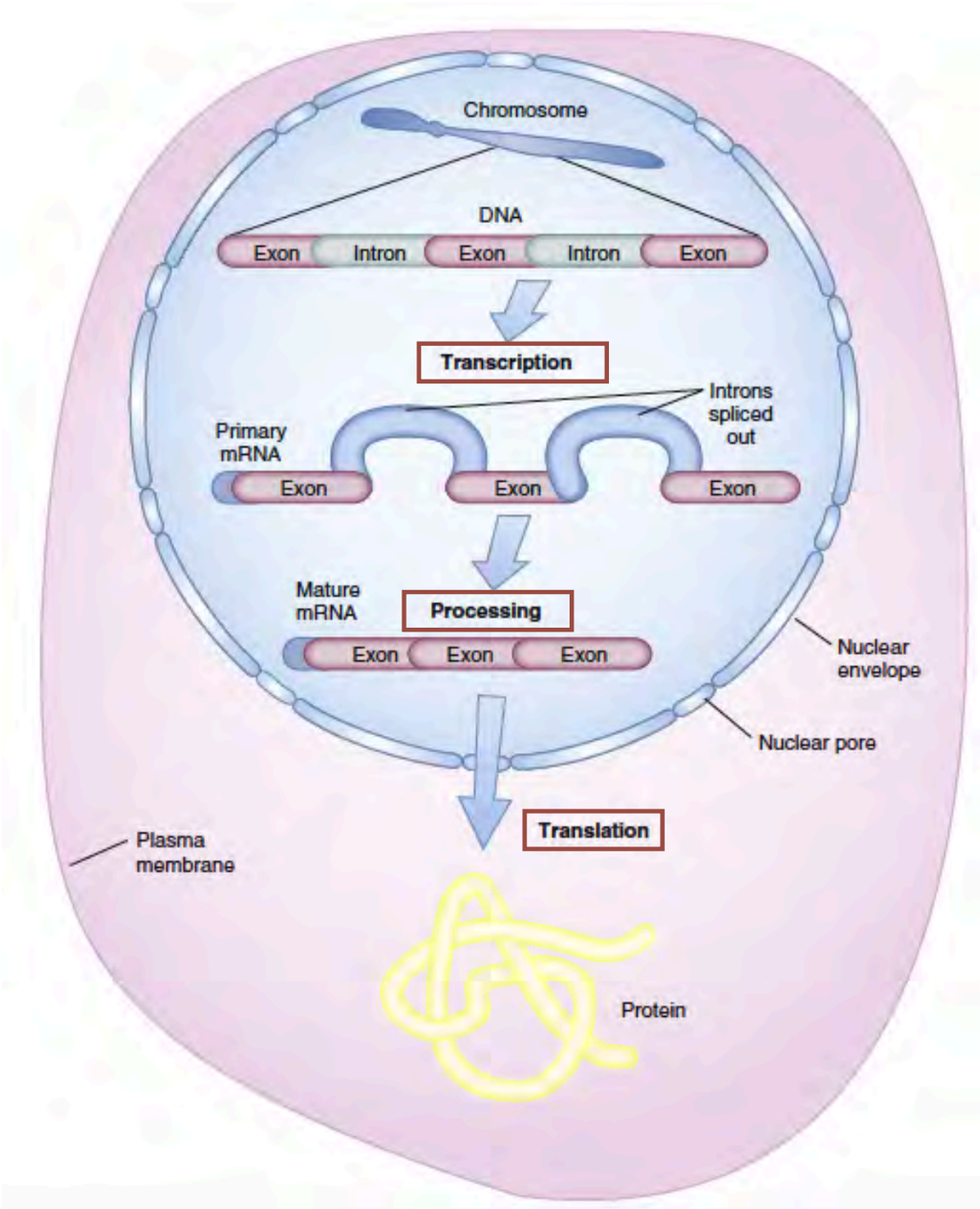
General schema: how genes control cell function



From genes to protein

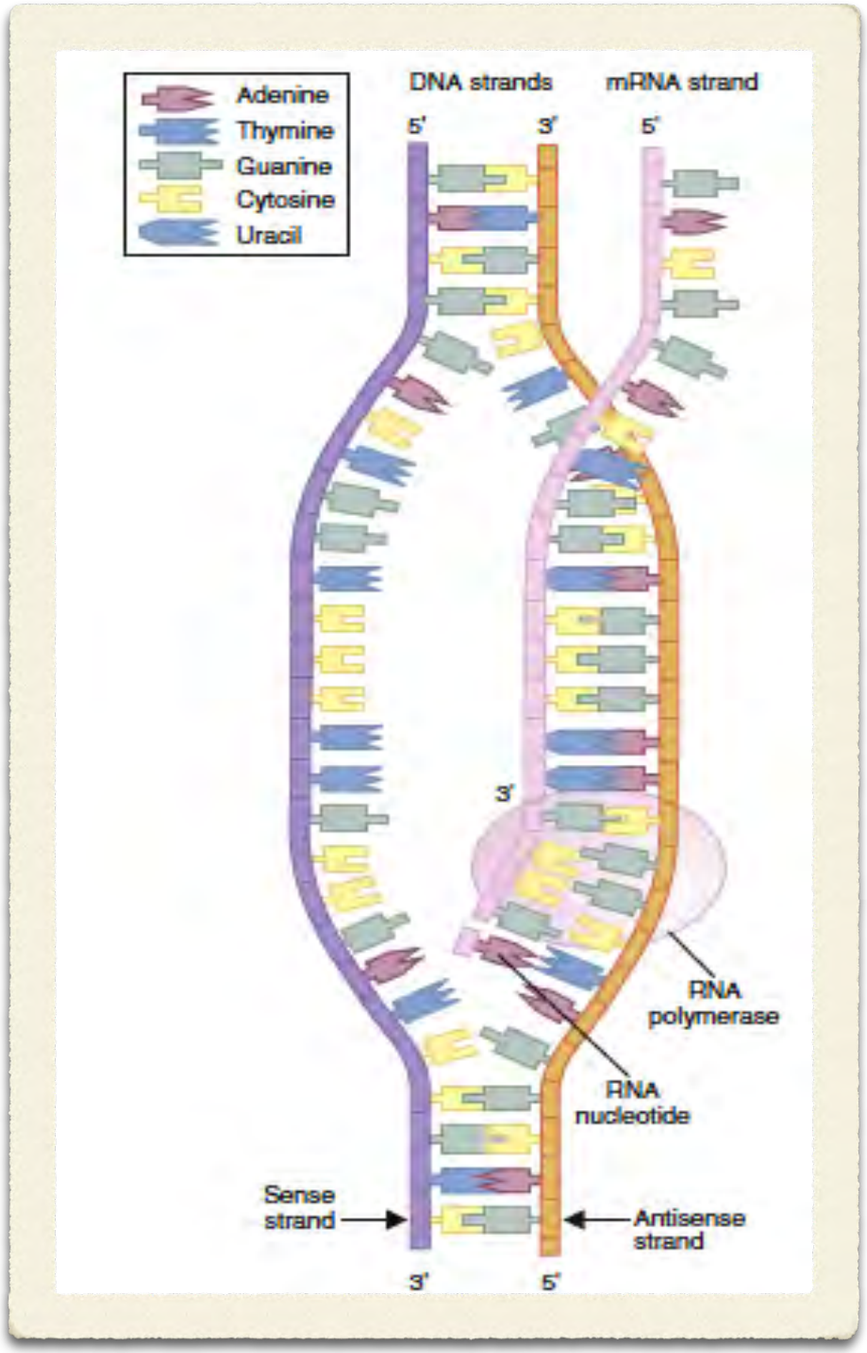


From genes to protein



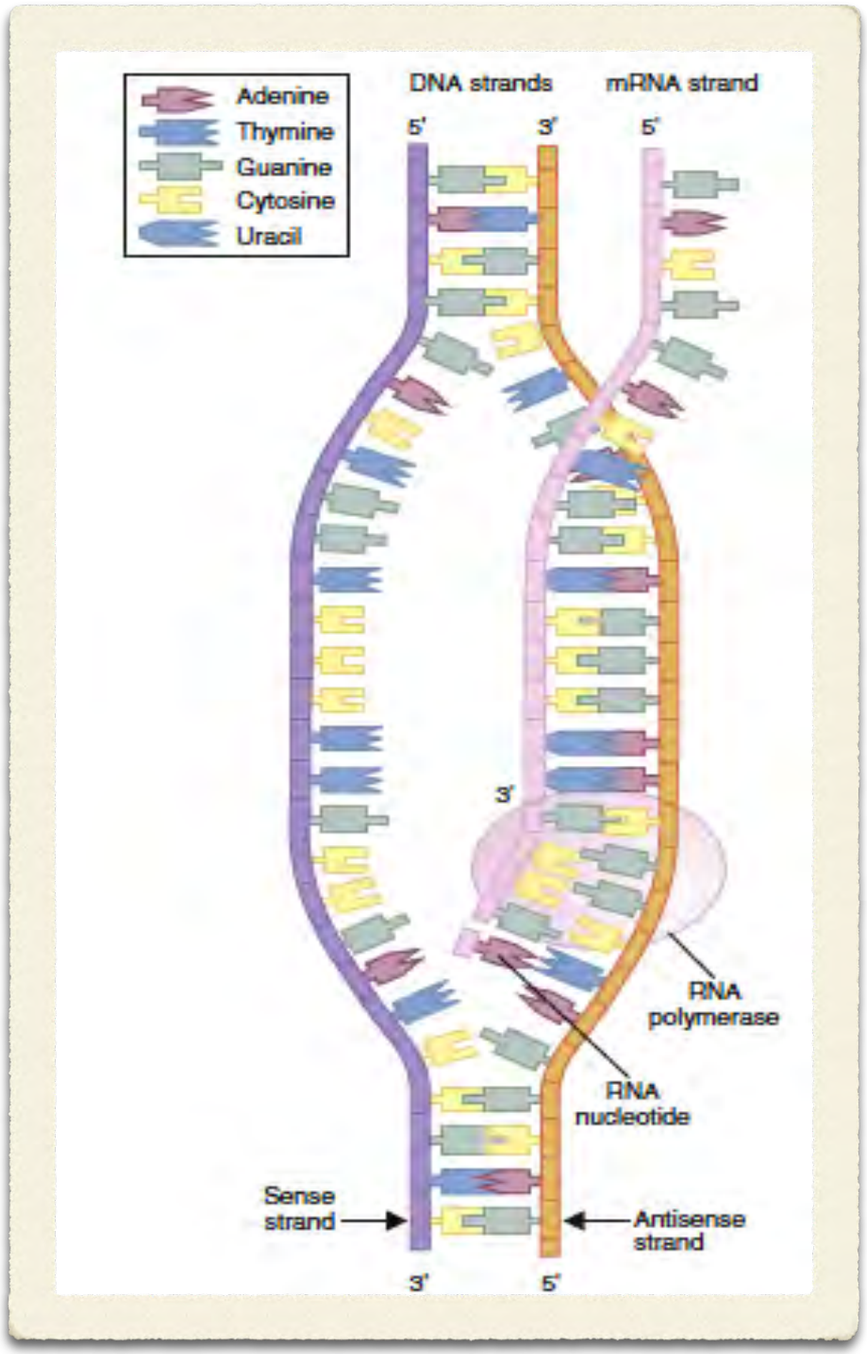
TRANSCRIPTION

- Process in which an RNA sequence is formed from a DNA template.
- Type of RNA produced by the transcription process is messenger RNA (mRNA).



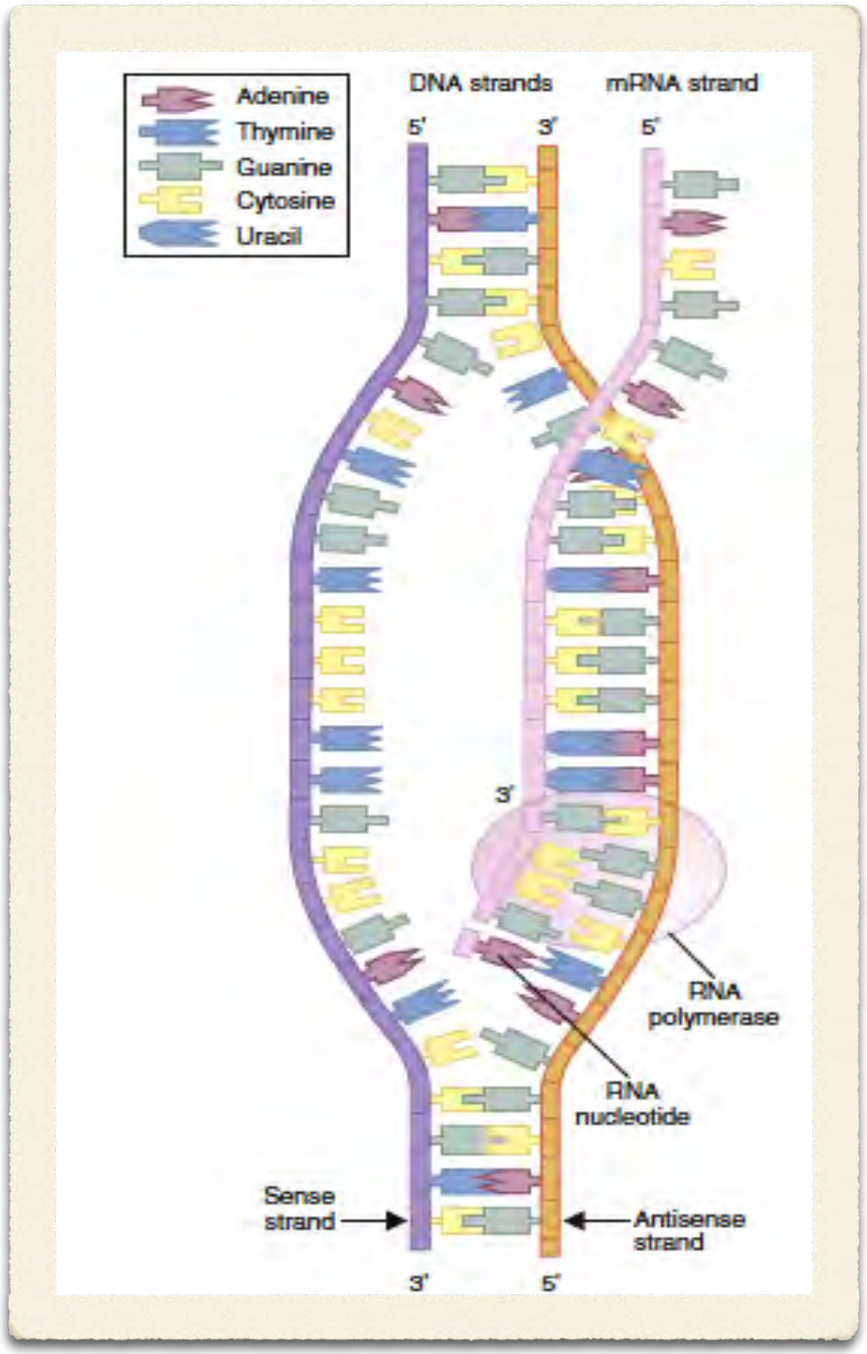
TRANSCRIPTION

- To initiate mRNA transcription, one of the **RNA polymerase** enzymes (RNA polymerase II) binds to a **promoter** site on the DNA.
- A promoter is a nucleotide sequence that lies just upstream of a gene.
- The RNA polymerase then pulls a portion of the DNA strands apart from each other, exposing unattached DNA bases.



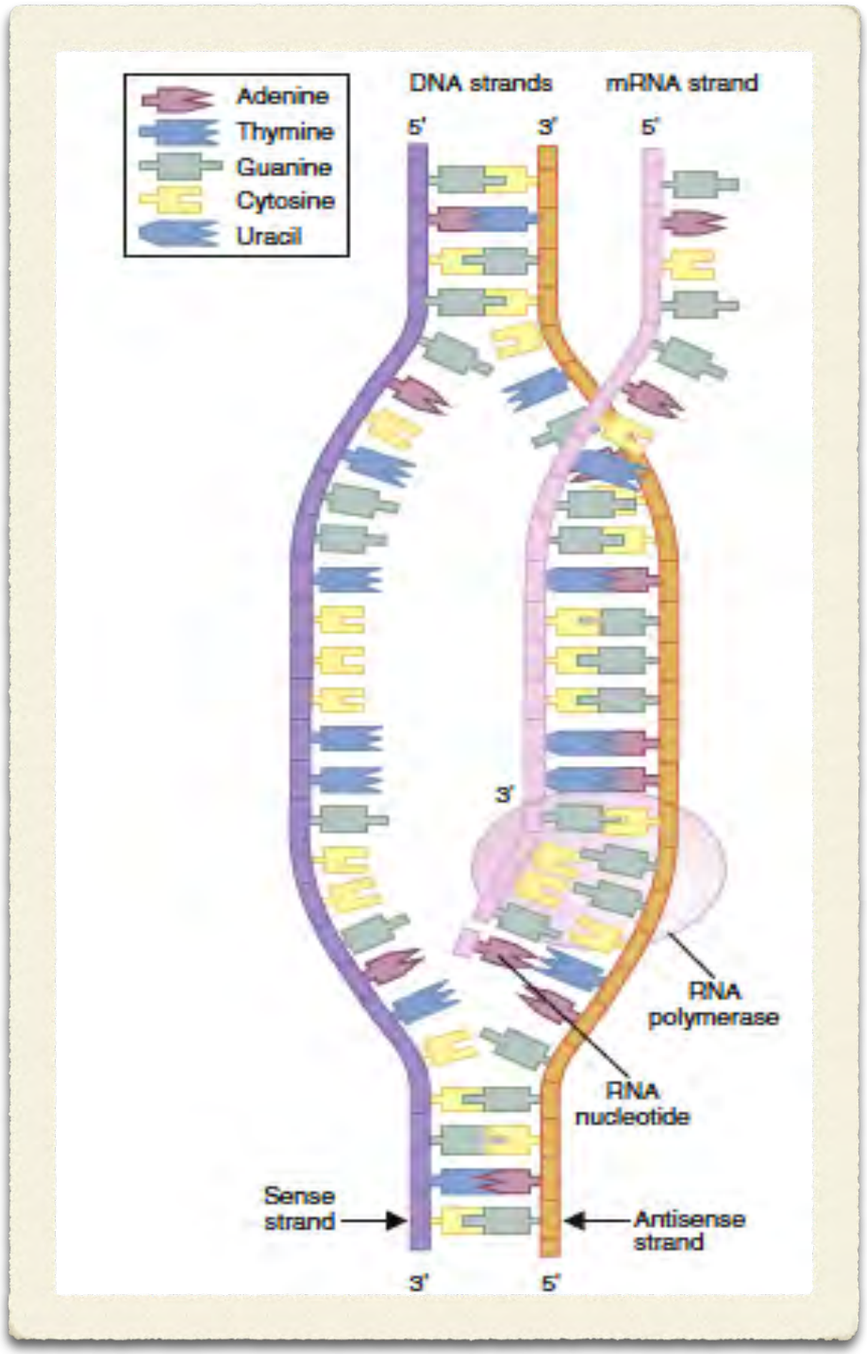
TRANSCRIPTION

- mRNA molecule can be synthesised only in the 5'to 3'direction.
- The promoter specifies directionality and determines which DNA strand serves as the template.
- This template DNA strand is known as the antisense strand.
- RNA polymerase moves in the 3'to 5'direction along the DNA template strand, assembling the complementary mRNA strand from 5'to 3'.



TRANSCRIPTION

- Because of complementary base pairing, the mRNA nucleotide sequence is identical to that of the DNA strand that does not serve as the template—the sense strand—except for the substitution of uracil for thymine.



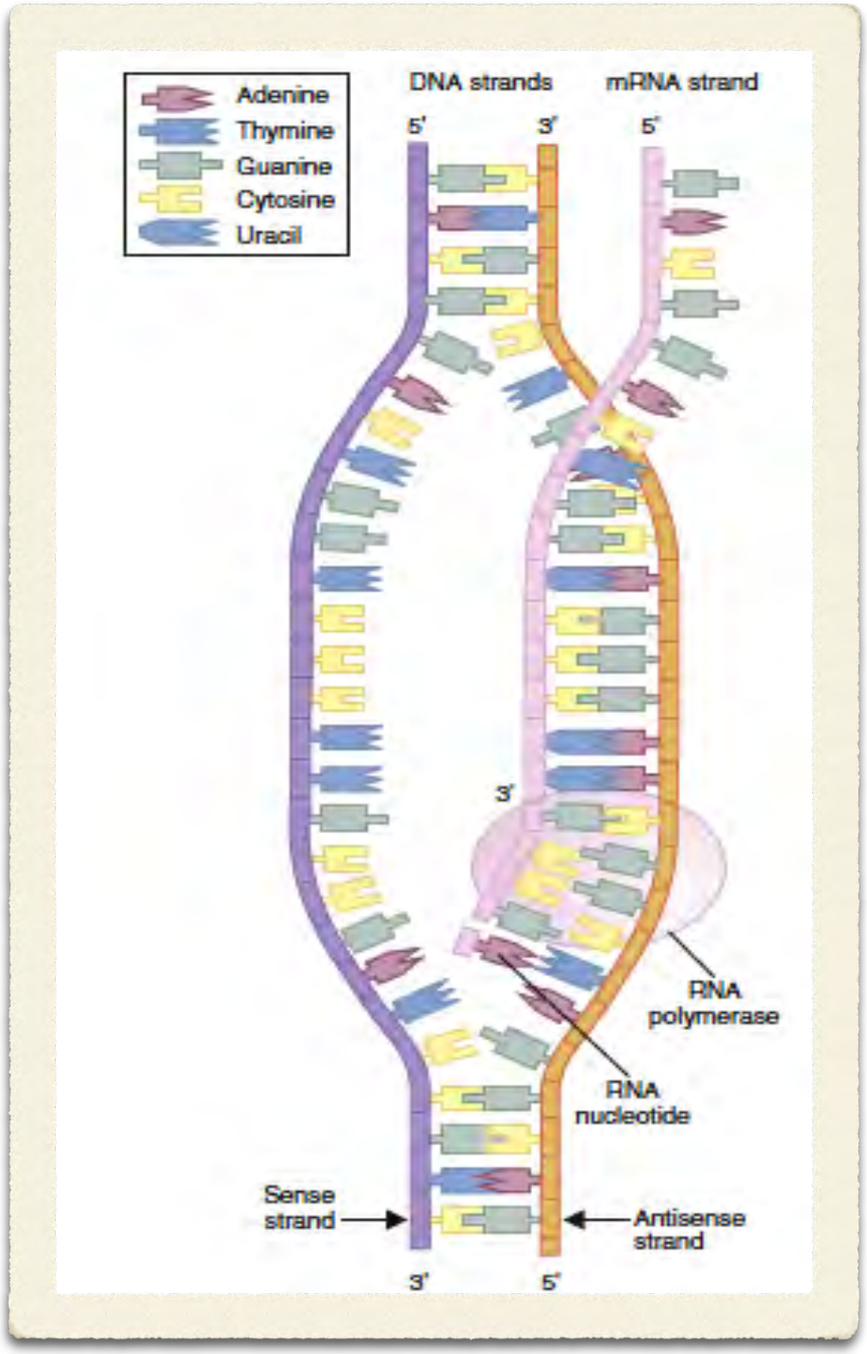
TRANSCRIPTION

- 5'end is capped by the addition of a chemically modified guanine nucleotide which prevents degradation and indicates starting position for translation of the mRNA molecule into protein.

- Transcription continues until a group of bases called a termination sequence is reached.

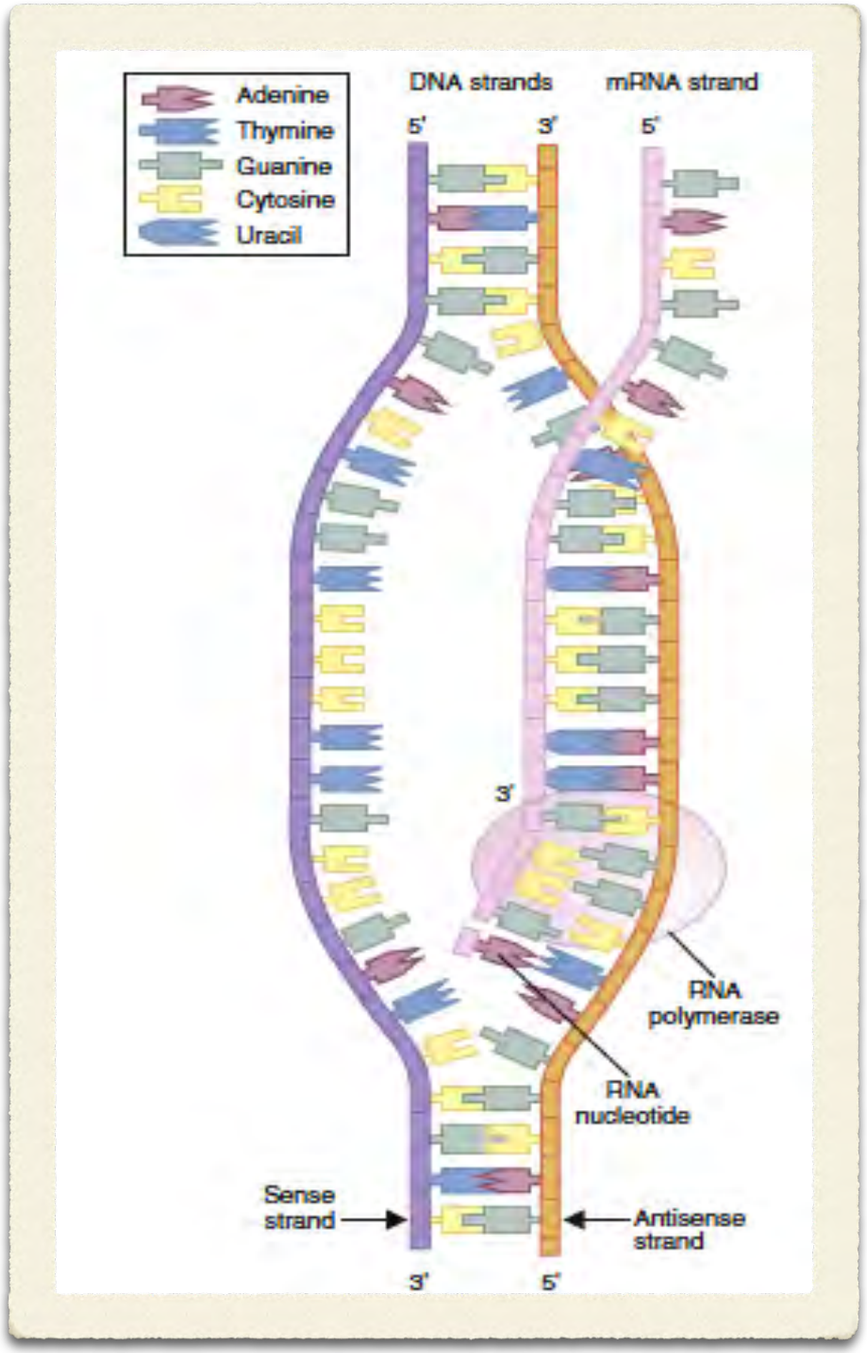
- A series of 100 to 200 adenine bases are added to the 3'end of the RNA molecule -> known as the **poly-A tail** that stabilises mRNA molecule to prevent degradation.

- The DNA strands and the RNA polymerase separate from the RNA strand, leaving a transcribed single mRNA strand termed the **primary transcript.**



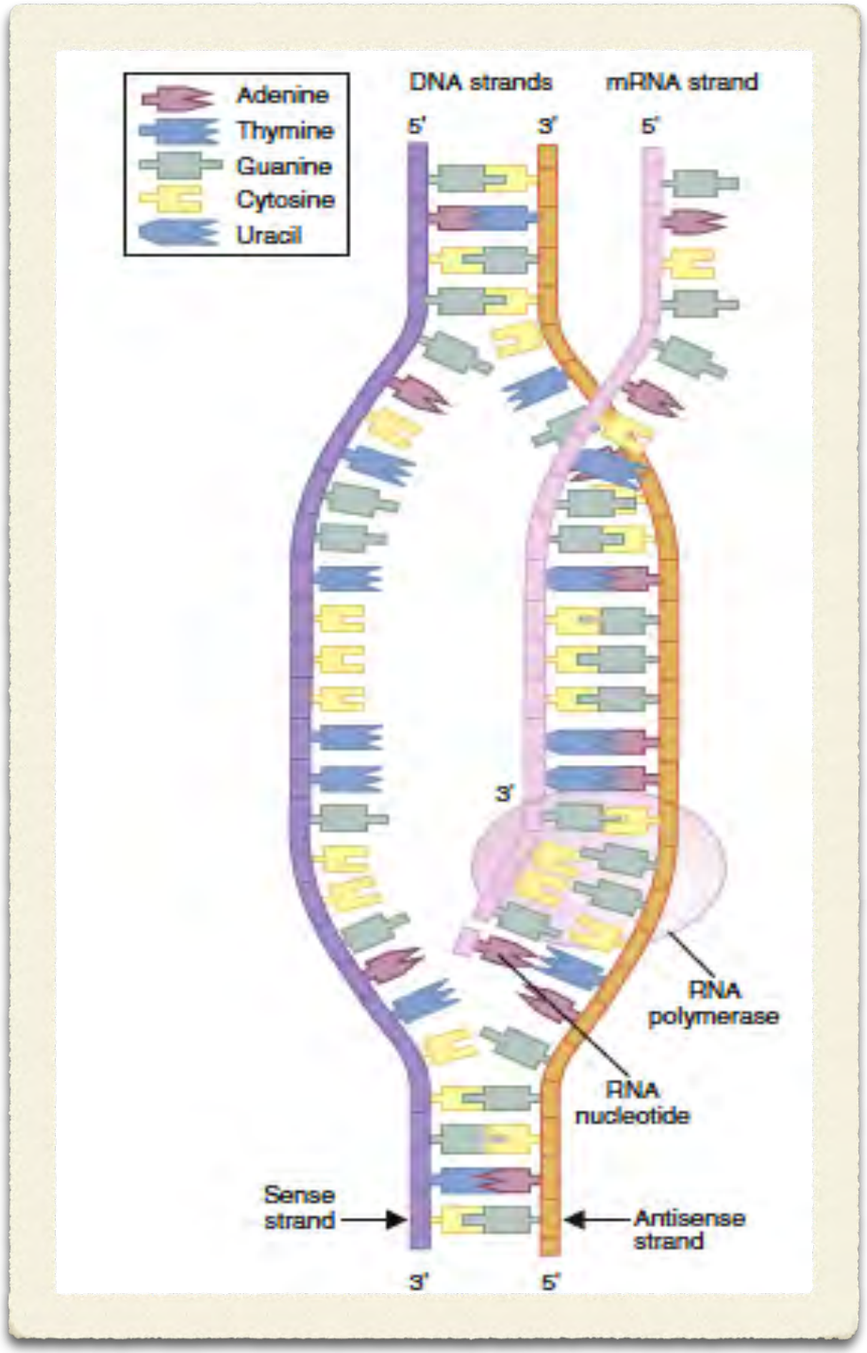
TRANSCRIPTION

- **Housekeeping genes** encode products that are required for a cell's maintenance and metabolism and are transcribed in all cells of the body.
- Most genes are transcribed only in specific tissues at specific points in time.
- Therefore in most cells only a small fraction of genes are actively transcribed.
- This specificity explains why there is a large variety of different cell types making different protein products, even though almost all cells have exactly the same DNA sequence.

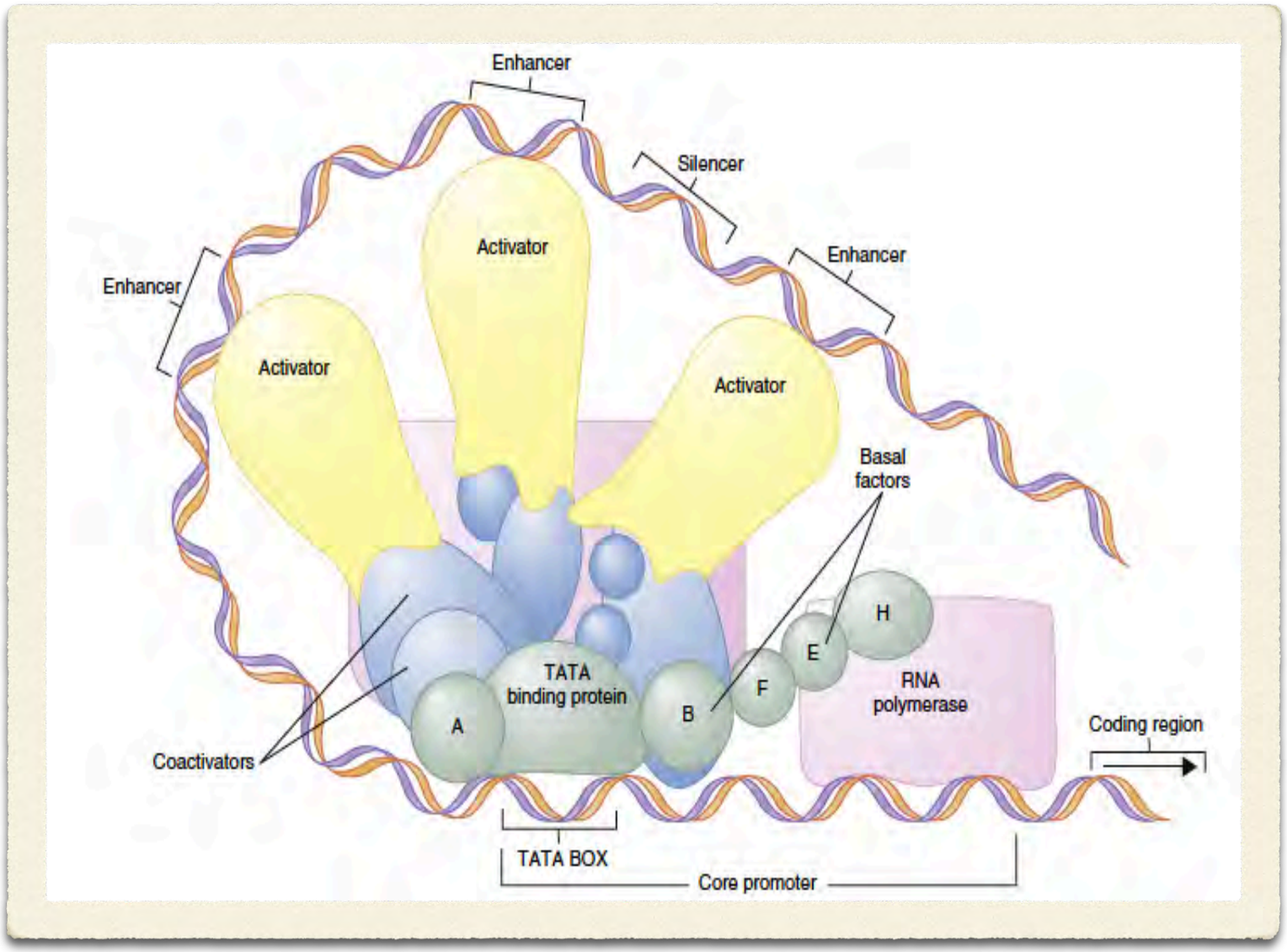


TRANSCRIPTION

- Many proteins participate in the process of transcription.
- Some are required for the transcription of all genes, and these are termed general **transcription factors**.
- Others, labeled specific transcription factors, have more specialised roles, activating only certain genes at certain stages of development.

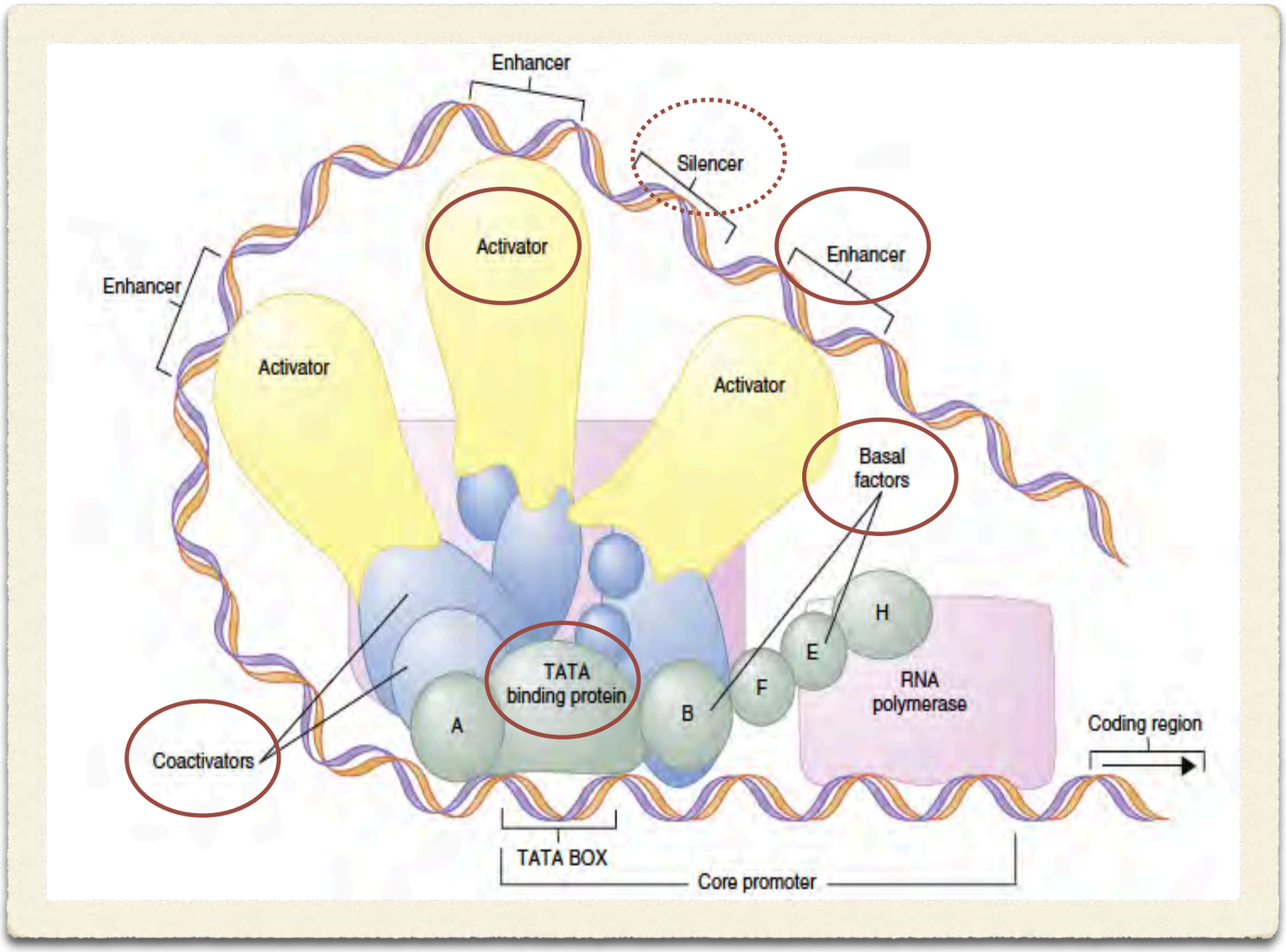


TRANSCRIPTION



- RNA polymerase II is vital in initiating transcription but it can't locate promoter region.
- RNA polymerase II is incapable of producing significant quantities of mRNA.
- Effective transcription requires the interaction of a large complex of approximately 50 different proteins.

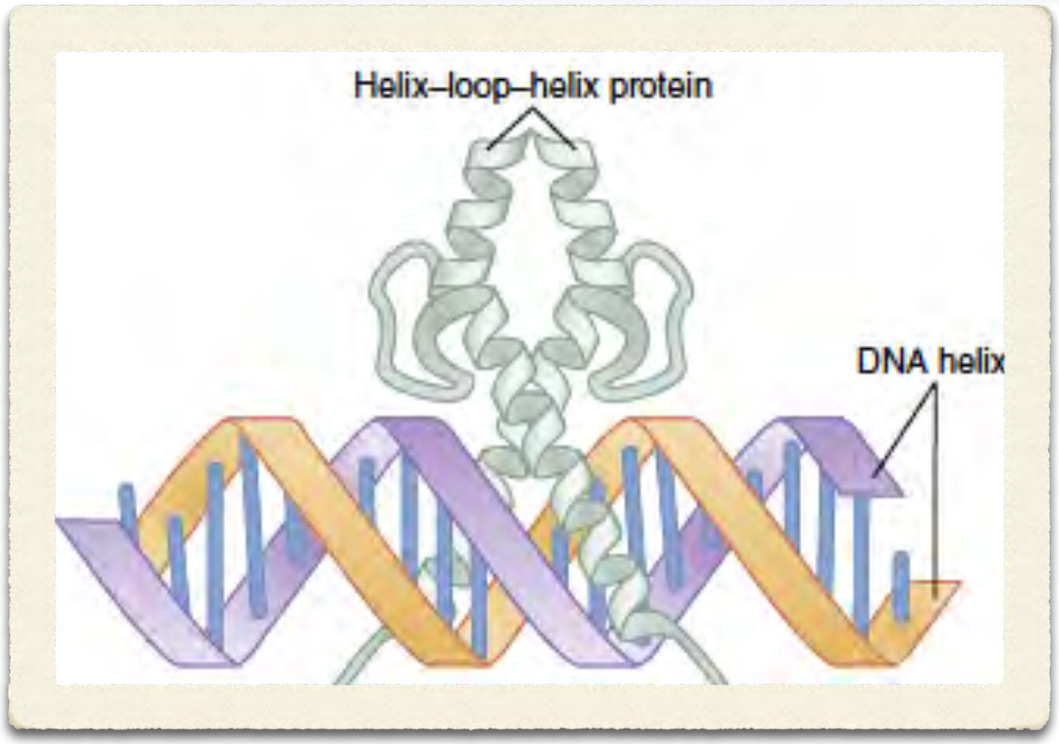
TRANSCRIPTION



- RNA polymerase II is vital in initiating transcription but it can't locate promoter region.
- RNA polymerase II is incapable of producing significant quantities of mRNA.
- Effective transcription requires the interaction of a large complex of approximately 50 different proteins.

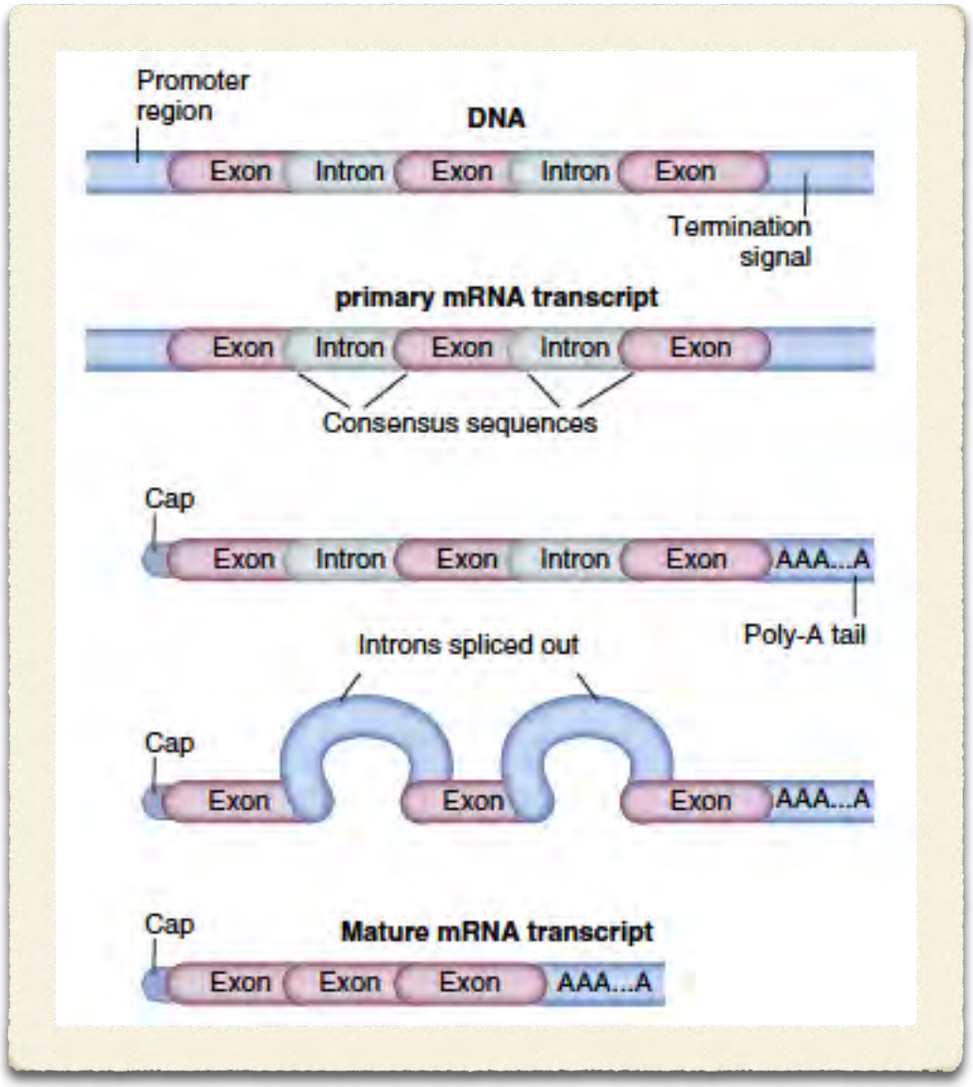
TRANSCRIPTION

Motif	Description	Human Disease Examples
Helix–turn–helix	Two α helices are connected by a short chain of amino acids, which constitute the turn. The carboxyl-terminal helix is a recognition helix that binds to the DNA major groove.	Homeodomain proteins (HOX): mutations in human *HOXD13* and *HOXA13* cause synpolydactyly and hand–foot–genital syndrome, respectively.
Helix–loop–helix	Two α helices (one short and one long) are connected by a flexible loop. The loop allows the two helices to fold back and interact with one another. The helices can bind to DNA or to other helix–loop–helix structures.	Mutations in the *TWIST* gene cause Saethre–Chotzen syndrome (acrocephalosyndactyly type III)
Zinc finger	Zinc molecules are used to stabilize amino acid structures (e.g., α helices, β sheets), with binding of the α helix to the DNA major groove.	*BRCA1* (breast cancer gene); *WT1* (Wilms tumor gene); *GL13* (Greig syndrome gene); vitamin D receptor gene (mutations cause rickets)
Leucine zipper	Two leucine-rich α helices are held together by amino acid side chains. The α helices form a Y-shaped structure whose side chains bind to the DNA major groove.	*RB1* (retinoblastoma gene); *JUN* and *FOS* oncogenes
β Sheets	Side chains extend from the two-stranded β sheet to form contacts with the DNA helix.	TBX family of genes: *TBX5* (Holt–Oram syndrome); *TBX3* (ulnar–mammary syndrome)



- DNA-binding motifs allow transcription factors to locate specific DNA sequences.
- In some cases, they bend DNA so that distant enhancer sequences can interact with target genes.

TRANSCRIPTION



- Sections of the RNA are removed by nuclear enzymes, and the remaining sections **spliced** together to form the functional mRNA.

- The excised sequences = introns, and the sequences that are left to code for proteins = **exons**.

- Some genes contain alternative splice sites –> allow the same primary transcript to be spliced in different ways –> producing different protein products from the same gene.

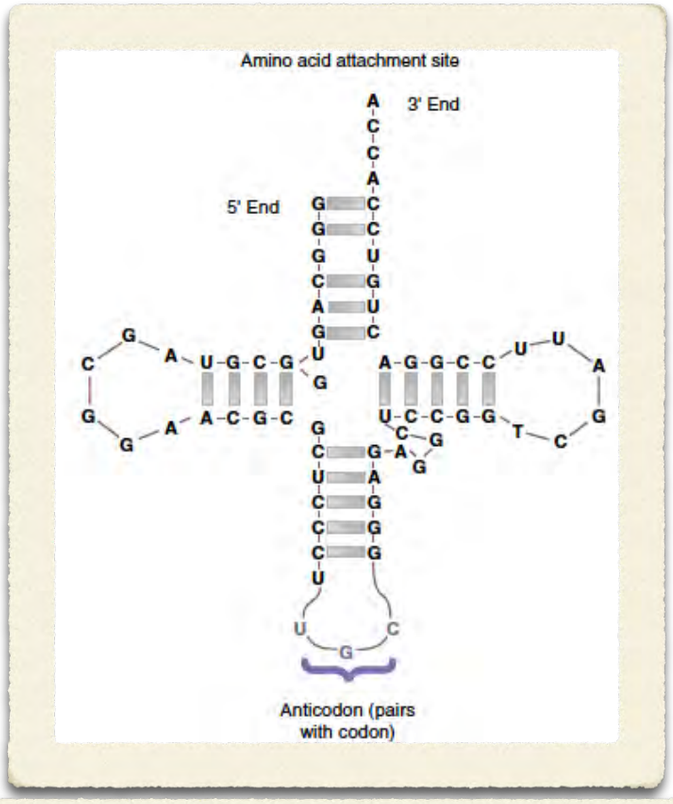
Genetic code

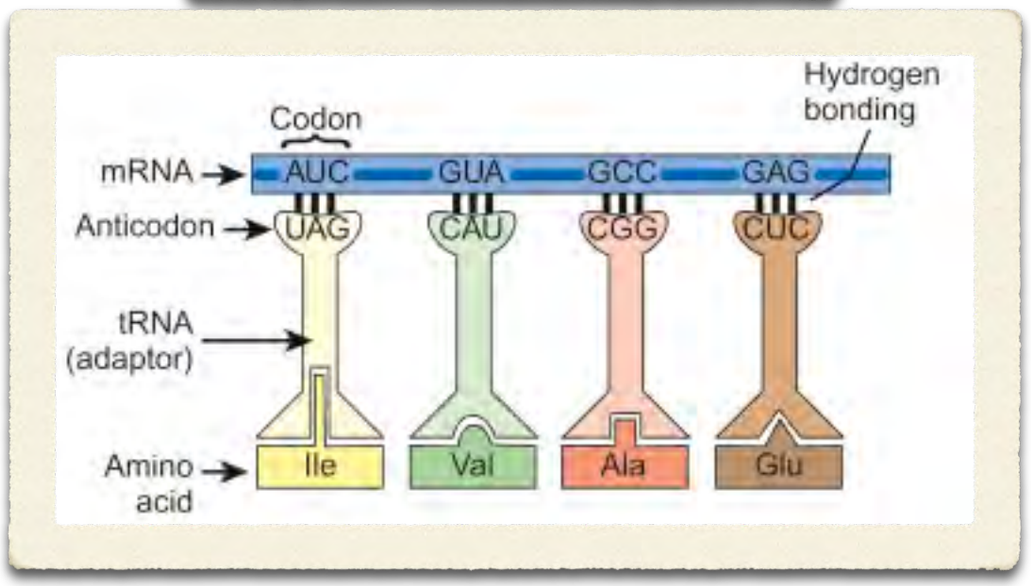
First Position (5′ END)	Second Position				Third Position (3′ END)
↓	U	C	A	G	↓
U	Phe	Ser	Tyr	Cys	U
U	Phe	Ser	Tyr	Cys	C
U	Leu	Ser	STOP	STOP	A
U	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
C	Leu	Pro	His	Arg	C
C	Leu	Pro	Gln	Arg	A
C	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
A	Ile	Thr	Asn	Ser	C
A	Ile	Thr	Lys	Arg	A
A	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
G	Val	Ala	Asp	Gly	C
G	Val	Ala	Glu	Gly	A
G	Val	Ala	Glu	Gly	G

Ala, Alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine.

TRANSLATION

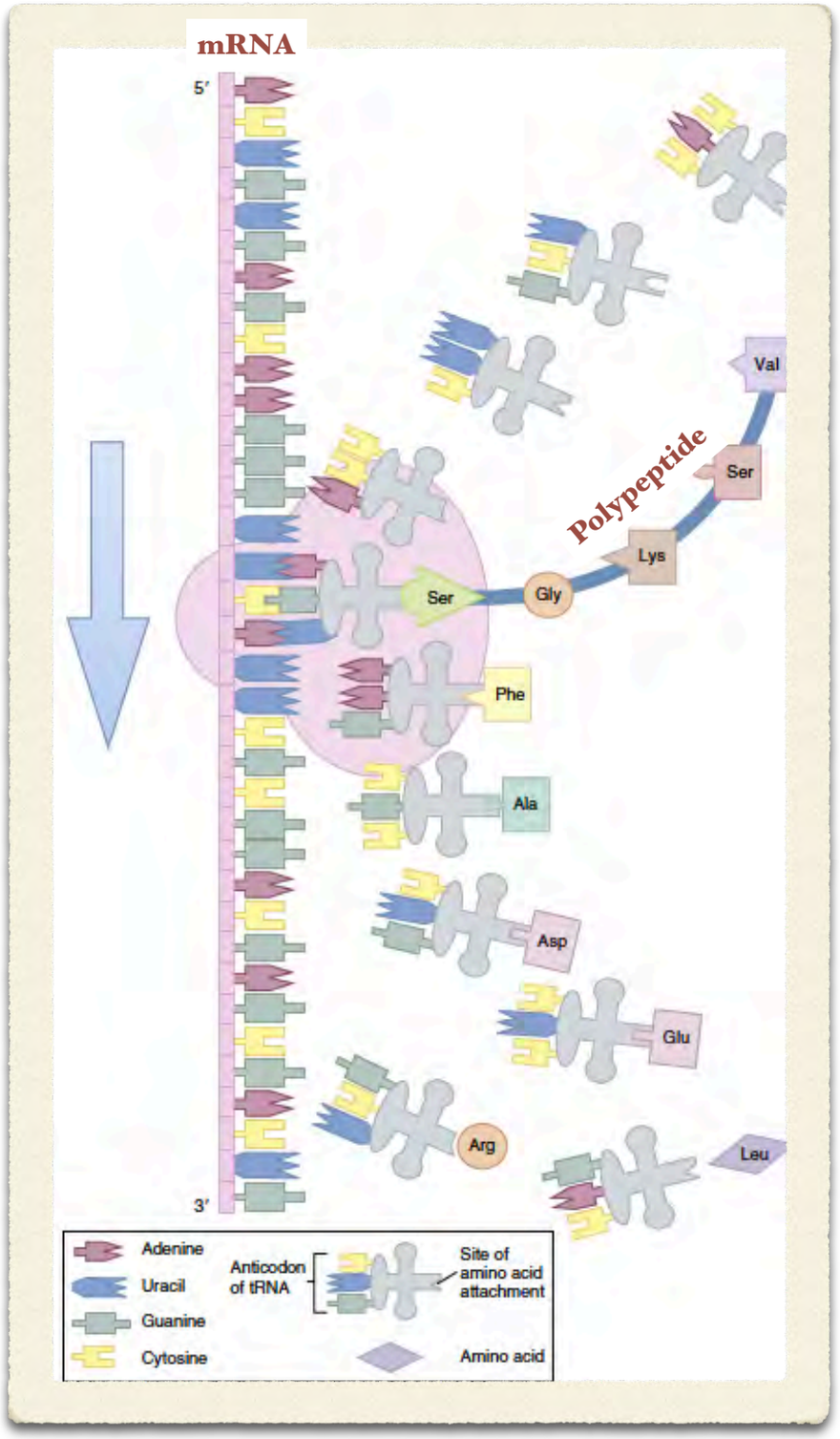
- Process in which mRNA provides a template for the synthesis of a polypeptide.
- mRNA interacts with molecules of transfer RNA (tRNA; cloverleaf-shaped RNA strands of about 80 nucleotides.)





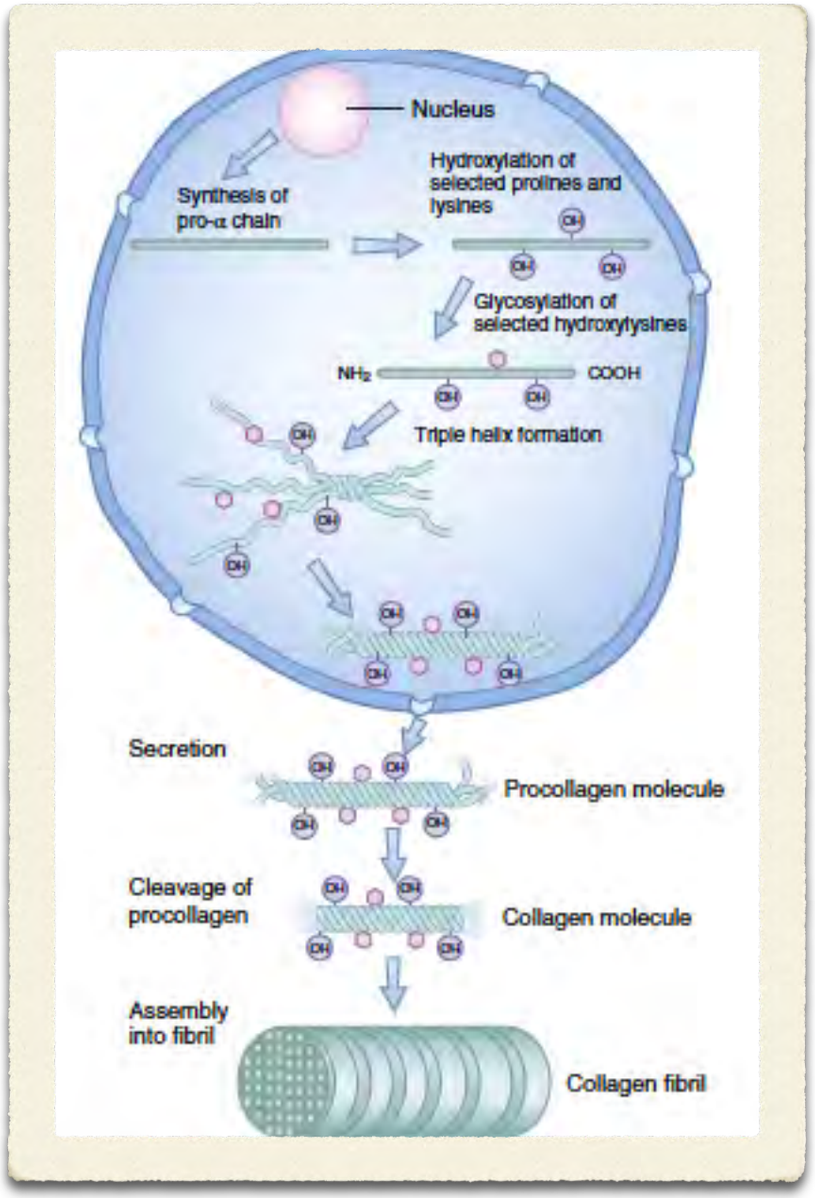
TRANSLATION

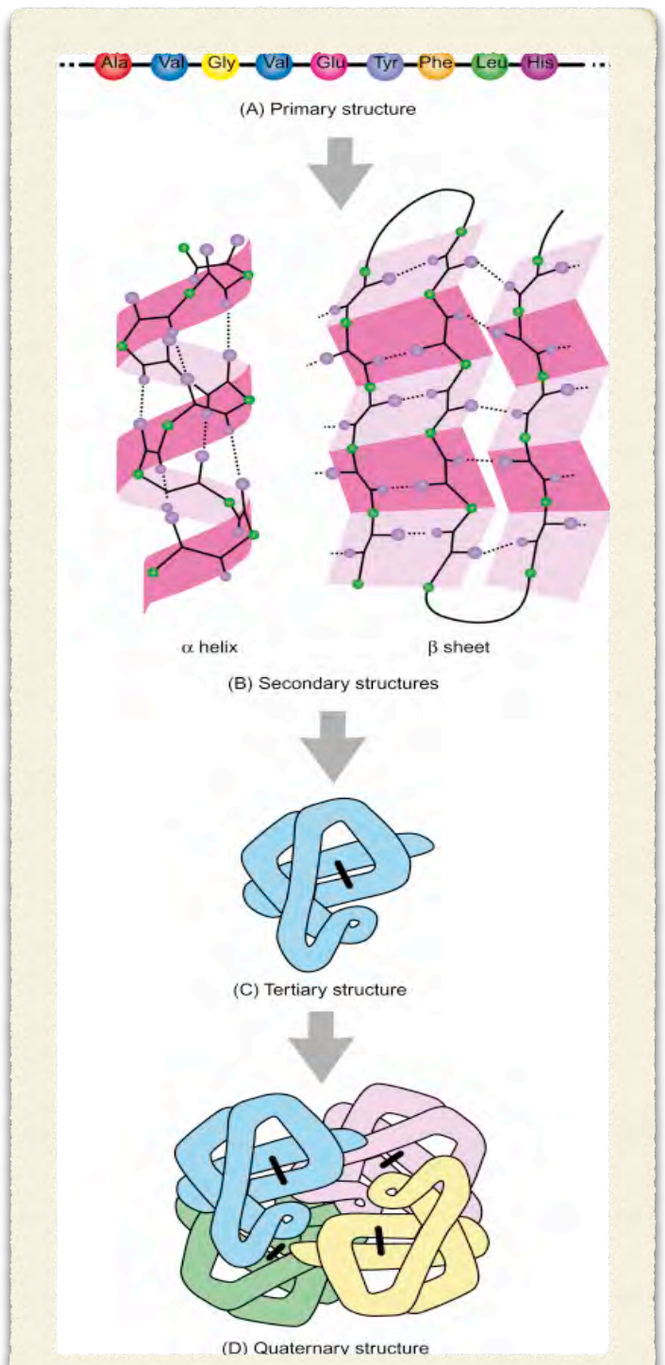
- Protein synthesis occurs in the ribosome.
- Ribosome consists of 50:50 enzymatic proteins:ribosomal RNA (rRNA).
- The function of rRNA is to help bind mRNA and tRNA to the ribosome.



TRANSLATION

- Post-translational modification consists of various chemical changes that occur in proteins shortly after they are translated.





The process of collagen fibril formation. After the proα polypeptide chain is formed, a series of post-translational modifications takes place, including hydroxylation and glycosylation. Three polypeptide chains assemble into a triple helix, which is secreted outside the cell. Portions of each end of the pro-collagen molecule are cleaved, resulting in the mature collagen molecule. These molecules then assemble into collagen fibrils.

TRANSLATION

Ribosomal RNA (rRNA)	comprises major portion of ribosome and is involved in synthesis of polypeptide chains
Transfer RNA (tRNA)	carries amino acids to ribosome and recognizes codons on mRNA
Small nuclear RNA (snRNA)	involved in the processing of mRNA molecules in the nucleus of eukaryotic cells
microRNA (miRNA)	small RNAs encoded by the genome and used to regulate gene expression
Short interfering RNA (siRNA)	short RNA created by enzymatic cleavage of a larger double-stranded RNA and used in defense against viruses
Guide RNA	involved in processing of RNA or DNA in some organisms
Regulatory RNA	functions in the regulation of gene expression by binding to proteins or DNA or to other RNA molecules
Antisense RNA	functions in regulating gene expression by base pairing to mRNA
Recognition RNA	part of a few enzymes (e.g., telomerase); enables them to recognize certain short DNA sequences
Ribozymes	enzymatically-active RNA molecules
U RNA	nUclear RNA; RNAs that function in the nucleus, including snRNAs
Y RNA	cYtoplasmic RNA; used in the regulation of RNA stability, the cellular stress response, and required for the initiation of DNA replication in eukaryotic cells

Various classes of RNA

Several classes of RNA exist that carry out a variety of roles in addition to carrying information for protein synthesis.

GENE EXPRESSION

Proteins are fundamental for nearly all functions of the cell.

Proteins serve as enzymes to catalyse chemical reactions in the cell.

Proteins are the chief components of the physical structures in the cell.

Genes control protein synthesis in the cell i.e. control cell function.

Gene is a double-stranded, helical molecule of deoxyribonucleic acid (DNA) that controls formation of ribonucleic acid (RNA).

RNA spreads in the cells to control the formation of a specific protein.

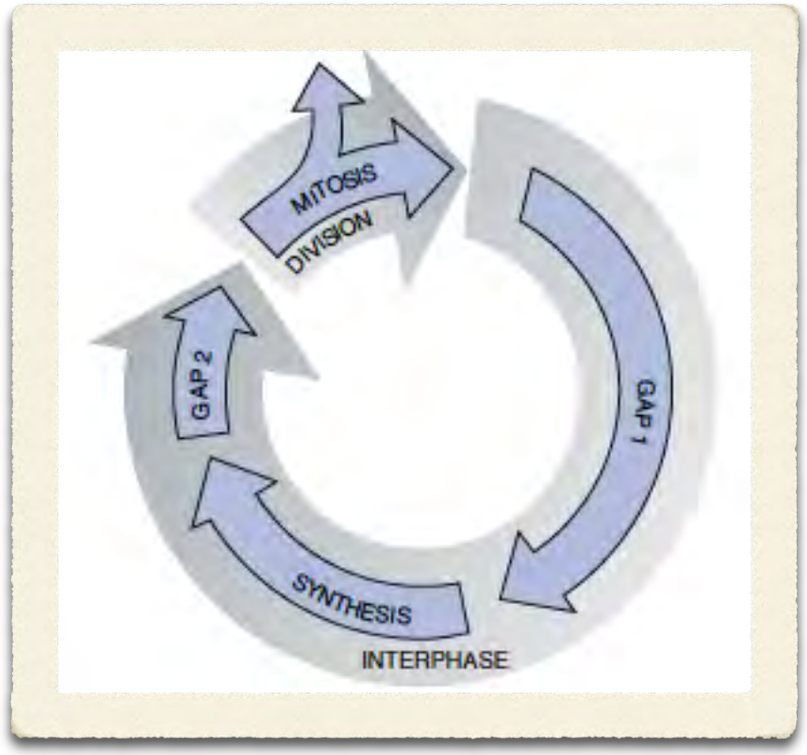
Gene expression = all the processes right from transcription of the genetic code in the cell nucleus to translation of the RNA code and formation of proteins in the cell cytoplasm.

CELL CYCLE

Processes responsible for the creation of new diploid cells are mitosis (nuclear division) and cytokinesis (cytoplasmic division).

A cell duplicates its contents including its DNA during interphase.

Alternation of mitosis and interphase is referred to as the cell cycle.



Length of the cell cycle varies between cell types:

A. Epithelial tissue (in intestines and lungs), cycle is completed in < 10 hours.
B. Liver cells may divide only once each year.
C. Skeletal muscle cells and neurons, largely lose their ability to divide and replicate in adults.

AGENDA

- Evolution of life.
- Cell.
- Genetics.
- Epigenetics.

Epigenetics

- Each cell in the body of an individual has an almost identical genome sequence. So how does the body generate so many distinct cell types?

- The answer is that addition of chemical modifications to the genome sequence and to the proteins that package the genome inside of cells - the epigenome.

- The genome is largely constant between different cells of an individual, while the epigenome can vary drastically between cell types and can change over time and in response to the environment.

Epigenetics

Historical origin:

- Around 1940s, Conrad Waddington and Ernst Hadorn, understood that genetics and developmental biology were related.

- Waddington is the founder of epigenetics in the 1940s. He coined the term epigenetics - a neologism derived by combining the words "epigenesis" (embryonic development) and "genetics".

- Epigenetics as per Waddington is "the branch of biology which studies the causal interaction between genes and their products which brings the phenotype into being".

- Epigenetics is defined as modifications of the genome, heritable during cell division, that do not involve a change in DNA sequence.

Epigenetics

Mechanisms:

- Epigenome - two primary components viz.

  1.chemical modifications of individual nucleotides making the genome

  2.chemical modifications of histone proteins packaging the genome

  "epigenetic marks"

- Nucleotide modifications - methylation and hydroxymethylation of cytosine.
- Histone modifications are diverse - methylation, phosphorylation, acetylation, sumoylation and ubiquitination. In neurobehavioral disorders, acetylation is the most studied.
- Non-coding RNAs (out of scope for this presentation)

Direct causality between epigenetic marks and genome function is not well established.

NUCLEOTIDE MODIFICATION

Cytosine methylation occurs when followed by a guanine in the genome: CpG dinucleotide ("p" = phosphodiester bond between adjacent nucleotides).

5mC first discovered in heterochromatin. Genes are silenced in heterochromatin, therefore 5mC is considered an epigenetic mark that silences gene expression.

The mechanisms involved in silencing of gene expression by DNA methylation include:

- direct interference in the binding of transcription factors to recognition elements that contain a CG dinucleotide.

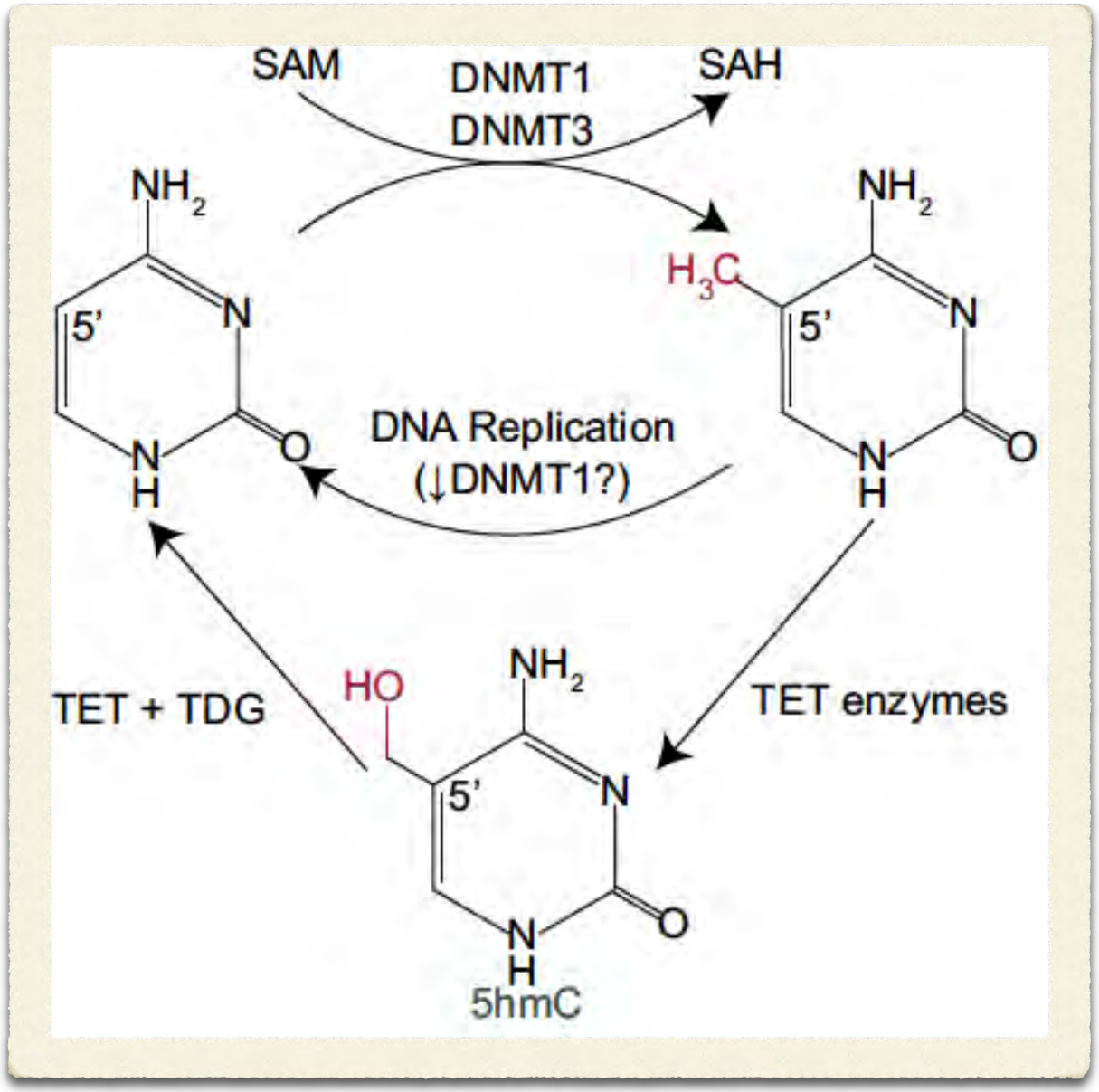
- recruitment of methylated DNA binding factors like methylated DNA binding domain (MBD) containing proteins that in turn attract chromatin-inactivating complexes including histone deacetylases and histone methyltransferases.

NUCLEOTIDE MODIFICATION

Schematic of the methylation and demethylation of cytosine in the human genome:

DNMT1 and DNMT3 enzymes catalyze the transfer of a methyl group from SAM to the 5` carbon of cysteine, resulting in 5mC and S-adenosyl-l-homocysteine (SAH).

The mechanisms by which 5mC is converted back to cytosine are less clear.

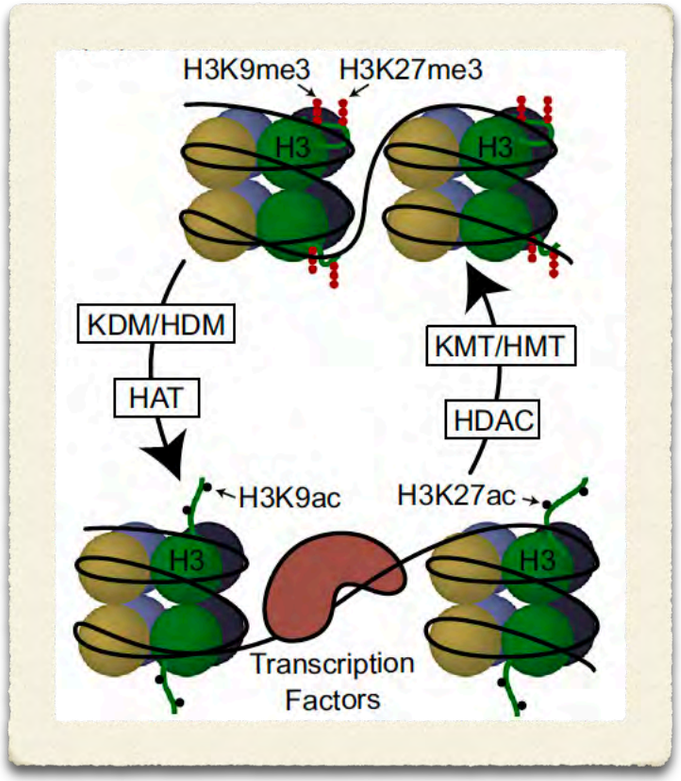


HISTONE MODIFICATION

Histone - octamer with 2 units each H2A,H2B, H3 & H4.

Within each histone protein, there are numerous sites where the amino acids can be modified by the covalent addition of small molecule groups.

The most well-understood histone modifications: addition of between one and three methyl groups or an acetyl group to specific lysine residues on the N-terminal tail of histone H3.



HISTONE MODIFICATION

- In general, acetylation of histone tails by histone acetyltransferases (HATs) decreases the electrostatic affinity between DNA and histone, resulting in an open-chromatin structure that leads to transcriptional upregulation of genes.

- Histone deacetylation by histone deacetylases (HDACs) induces transcriptional suppression.

- HDACs can be classified into four distinct groups (classes I to IV). Valproic acid is a class I and II HDAC inhibitor.

HISTONE MODIFICATION

The modifications are typically written using a shorthand notation that specifies (1) the histone, (2) the amino acid, and (3) the covalent modification.

Example, the acetylation of histone H3 lysine 27 is expressed as H3K27ac.

H3K4me3 means that lysine 4 on histone H3 has three methyl groups.

Transcriptional effect	Modification	Histone protein	Residue
Transcriptional repression	Arginine methylation	H3	Arg2
	Lysine methylation	H3	Lys9, Lys27
Transcriptional activation	Lysine methylation	H3	Lys4, Lys36, Lys79
		H4	Lys20
		H2B	Lys5
	Lysine acetylation	H3	Lys4, Lys14, Lys18, Lys23, Lys27
		H4	Lys5, Lys8, Lys12, Lys16, Lys20, Lys91
		H2A	Lys5, Lys9
		H2B	Lys5, Lys12, Lys20, Lys120
	Serine phosphorylation	H3	Ser10
		H4	Ser1

EPIGENETIC MECHANISMS OF DISEASES

- Mutations in methyl CpG binding protein 2 (MeCP2) occurs in Rett syndrome a neurodevelopmental disorder.

- Discovery of hypermethylation of Reelin in brain of patients with 'schizophrenia' that was associated with an increase in DNMT in GABAergic cortical neurons.

GWS of differentially methylated CGs in frontal cortex and germline associated with 'schizophrenia' and 'bipolar disorder' showed extensive changes in DNA methylation of many functional gene networks for mitochondrial function, brain development and stress response.

Unclear whether the DNA methylation changes are a cause or downstream effect of the pathology or whether they are involved in progression of the disease.

EPIGENETIC MECHANISMS OF DISEASES

- Defective histone-modifying enzymes can directly cause diseases though extremely low prevalence. Examples:

    1. Weaver syndrome - combination of skeletal and cognitive abnormalities, causal mutation is an enzyme EZH2 that is responsible for the methylation of H3K27.

    2. Brachydactyly-mental retardation syndrome - mutations that abolish function of histone deacetylase HDAC4, characterised by developmental delays and skeletal and craniofacial abnormalities.

- Mutations in other genes that lead indirectly to altered histone modifications. Example: Huntington's disease.

    A. Genetics: expansion of trinucleotide CAG repeats in the DNA sequence of the huntingtin (HTT) gene.

    B. Repeats lead to polyglutamines in HTT protein, finally causing progressive neurodegeneration.

    C. Epigenetic abnormalities now known includes loss of histone acetylation (mechanism under investigation).

    D. Evidence of HDAC inhibitors improving HD-associated symptoms in animal models; human trials not yet performed.

SUMMARY

1. Life on Earth evolved / is evolving since 3.5-4 billion yrs. Cell is the basic unit of life. DNA controls cells functions by coding for various proteins (>100,000 in humans). Proteins are essential for structure and function of cells.

2. Proteins are manufactured by processes of transcription, processing in nucleus and translation and post translation modifications in cytoplasm.

3. Various proteins (RNApolymerase II and 50+ proteins - promoters, activators, TATA box, basal factors, co-activators, enhancers, silencers) are involved in initiating transcription to generate the primary transcript mRNA. DNA binding motifs help transcription factors to identify the specific DNA sequence to be transcribed.

4. The primary mRNA transcript undergoes processing by splicing in the nucleus to give rise to mature mRNA transcript containing the exons. This is then transported to cytoplasm. The genetic code codes for 20 different amino acids - the building blocks for polypeptides and protein.

5. mRNA binds to tRNA to generate the polypeptide chain, which generates a protein. The protein undergoes post translational modifications.

6. Gene expression are all the processes right from transcription of the genetic code in the cell nucleus to translation of the RNA code and formation of proteins in the cell cytoplasm.

7. Epigenetics is defined as modifications of the genome, heritable during cell division, that do not involve a change in DNA sequence.

8. Two chief mechanisms of epigenetic regulation are: chemical modifications of individual nucleotides making the genome and chemical modifications of histone proteins packaging the genome. Third mechanism is by non-coding RNA which is not discussed in this basic view.

9. DNA methylation generally suppresses gene expression and demethylation promotes gene expression. Example - Rett syndrome, Reelin gene hypermethylation in psychosis/'schizophrenia'.

10. Histone acetylation generally promotes gene expression and deactylation suppresses gene expression. Example - Weaver syndrome, indirect histone modification in HD.

"The field of epigenetics is susceptible to over-interpretation, unrealistic hype and even deliberate misrepresentation.

Experts and the public alike therefore need to exercise a degree of caution when it comes to claims about epigenetics."

Introducing epigenetics: Eds. C Ennis and O Pugh. Publisher: Icon books London 2017.