KATHOLIEKE UNIVERSITEIT
LEUVEN

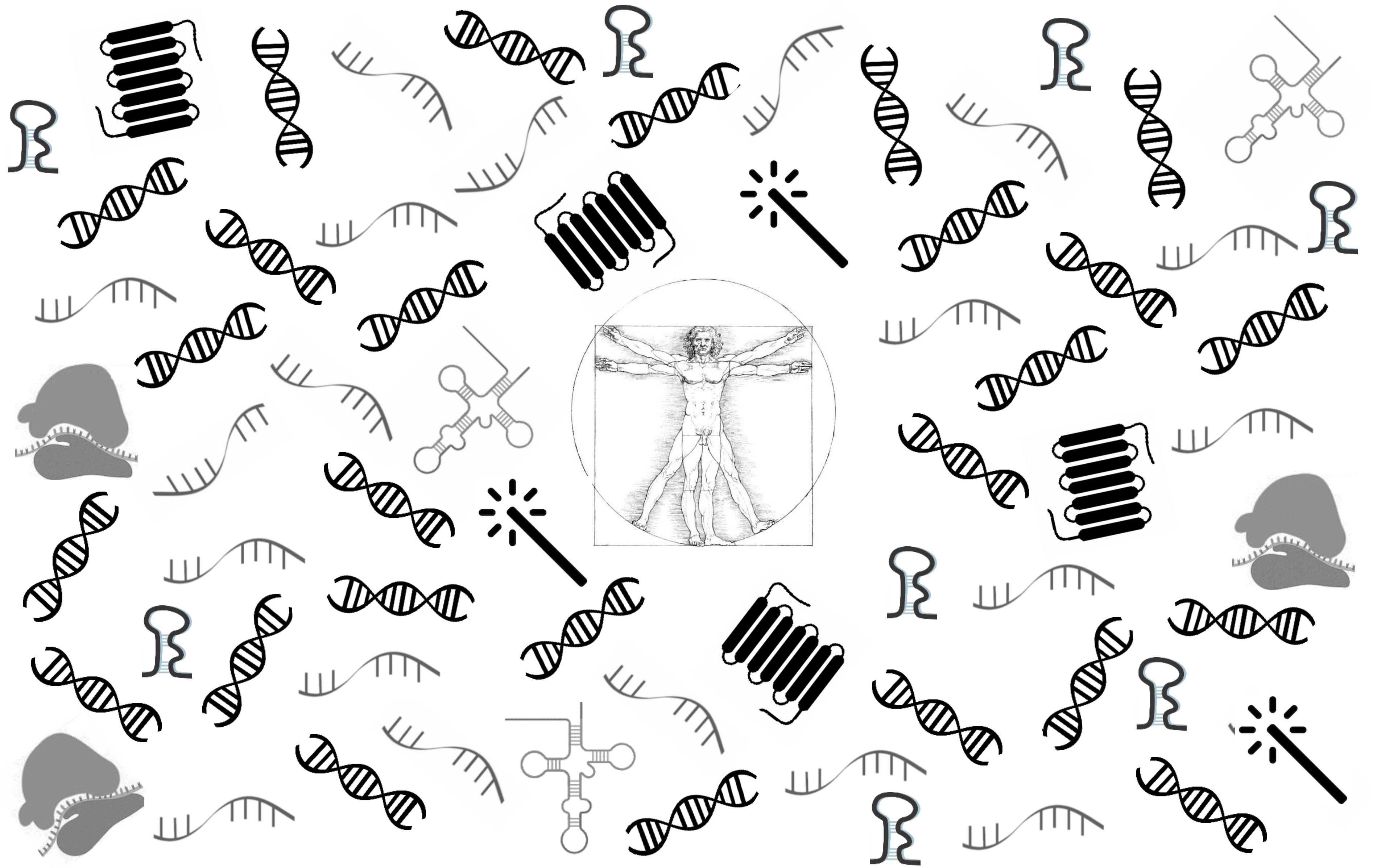
Human genome: gene structure & function

Jeroen Breckpot

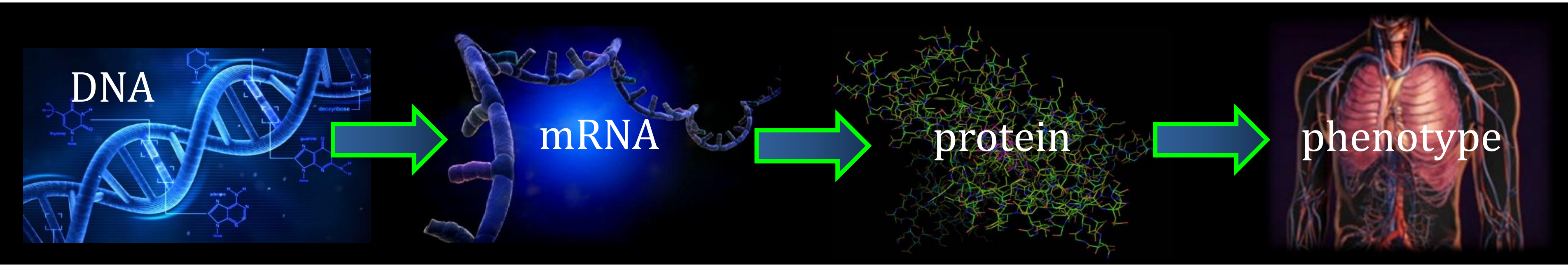
POC Belgian Society of Human Genetics

October 2023

Introduction



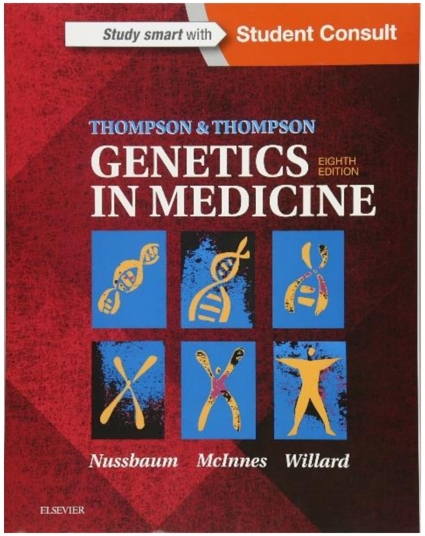
The Central Dogma



nucleus	*nucleus* *-* *cytoplasm*	*cytoplasm,* *organelles,* *extracellular*	*tissues,* *organs,* *organism*

Outline of the presentation

1. Definitions
2. Transcription
3. Translation
4. Regulatory mechanisms

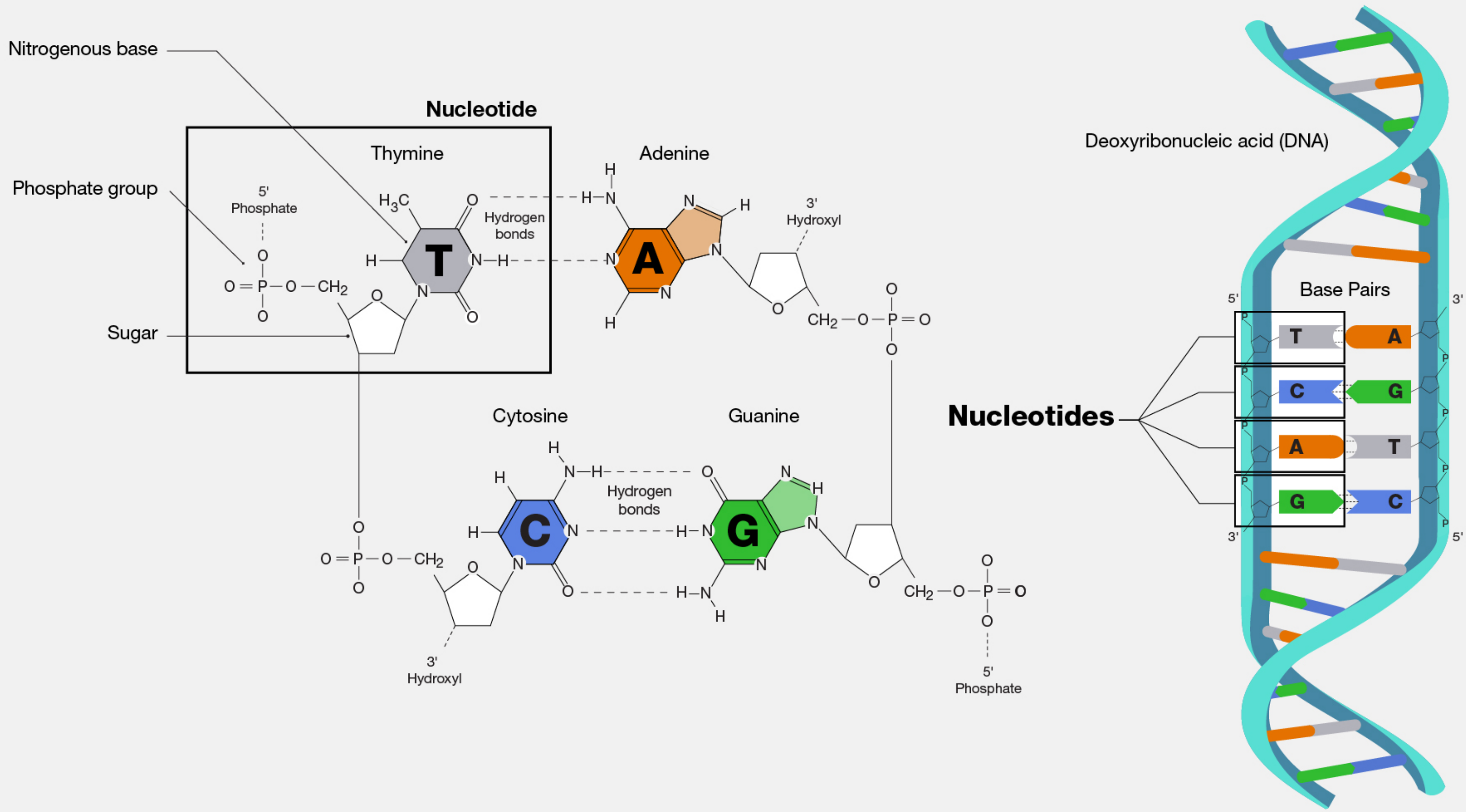


Chapter 2 & 3. Thompson & Thompson Genetics in Medicine

Definitions

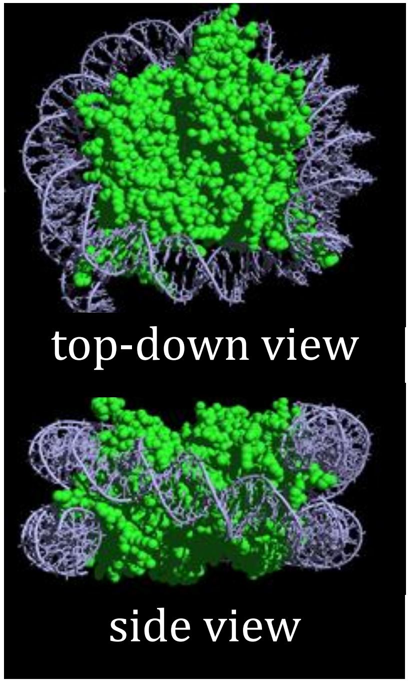
From base pair to chromosome

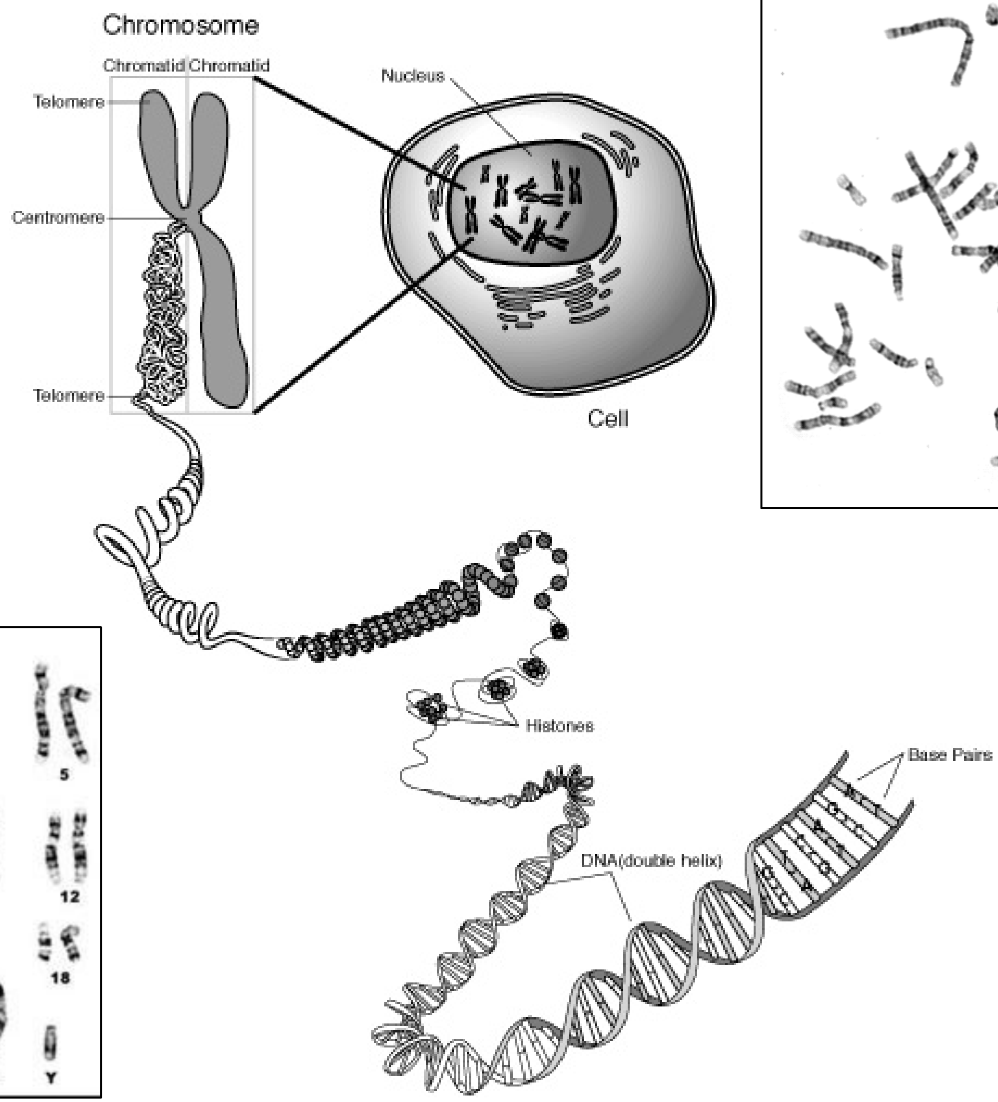
DNA

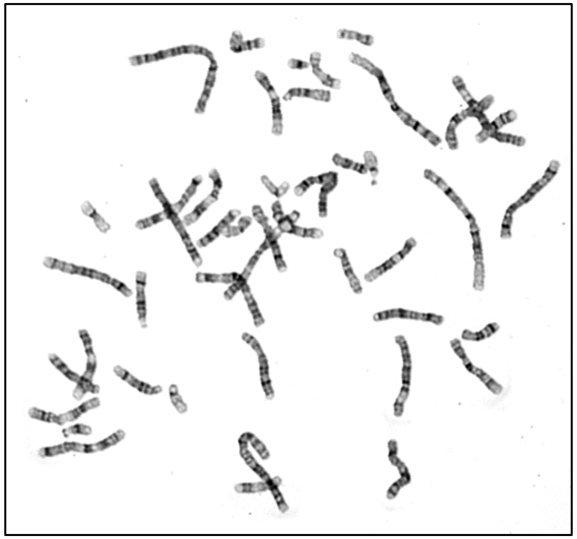


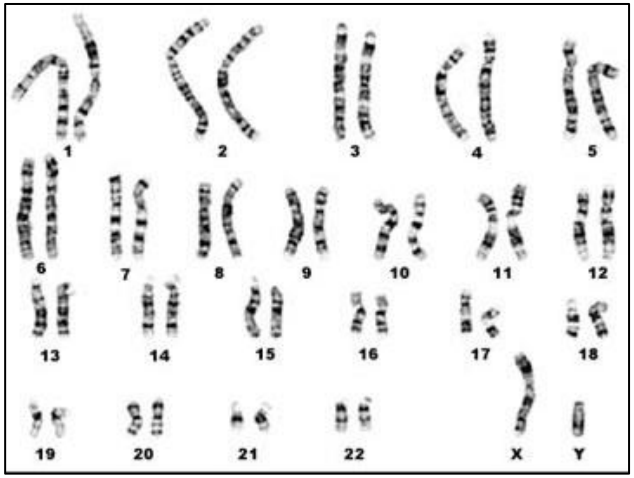
Nitrogenous base
Nucleotide
Thymine
Adenine
Phosphate group
5' Phosphate
Hydrogen bonds
3' Hydroxyl
Sugar
Cytosine
Guanine
Hydrogen bonds
3' Hydroxyl
5' Phosphate
Deoxyribonucleic acid (DNA)
Base Pairs
Nucleotides

From base pair to chromosome





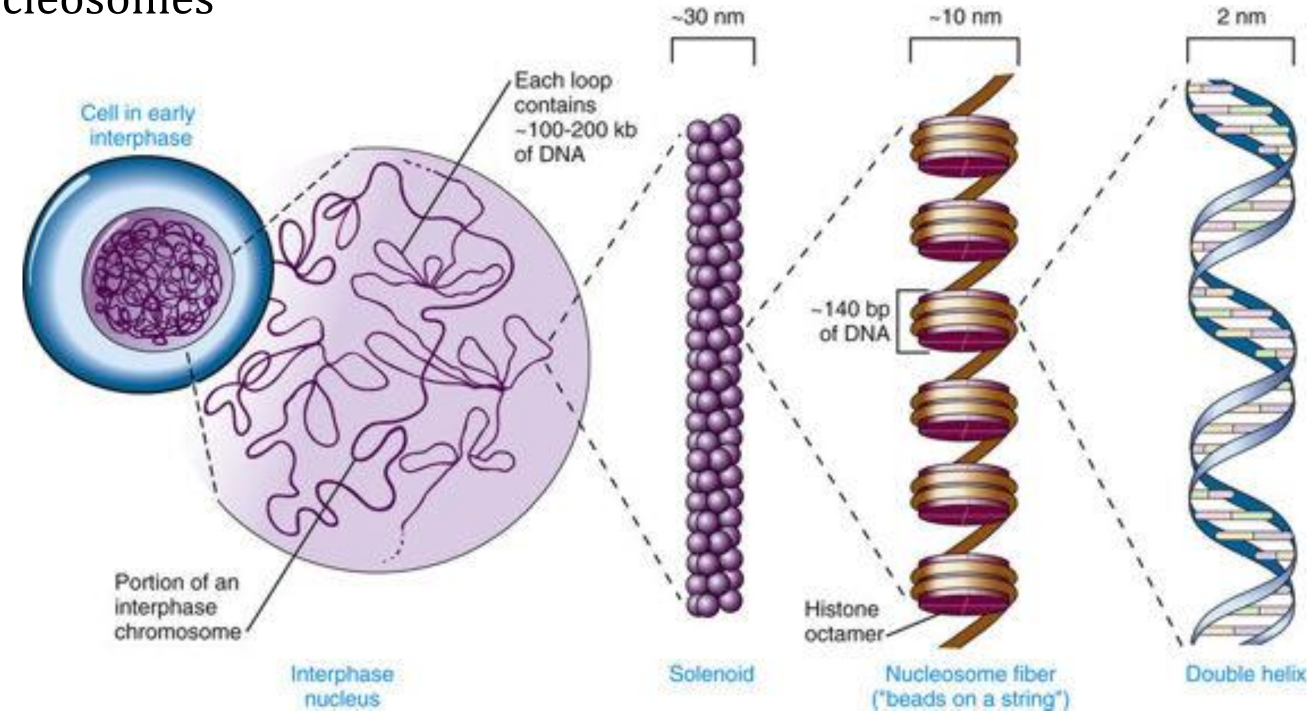




Histones

5 major types of histones play a critical role in the packaging of chromatin:

- ✓ two copies of **H2A, H2B, H3** and **H4** form an octamer around which a DNA segment of about 140 bp is wrapped = **nucleosome**
- ✓ **H1** binds to the 20 to 60 bp 'spacer' segment of DNA between two nucleosomes

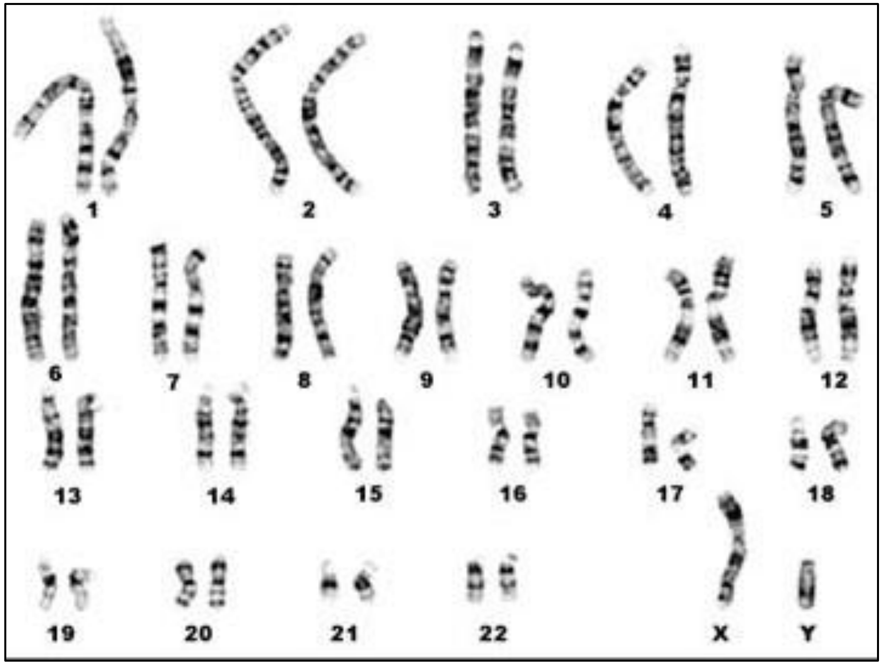


- ✓ H3 and H2A can be substituted by other histone types, or histones can be modified by chemical changes: cfr. regulatory mechanisms

Chromosomes

Human somatic cells: 46 chromosomes:

- 22 pairs of autosomes: 'homologues'
- 1 set of sex chromosomes: XY or XX



✓ Short arm = p ('petit')
Long arm = q
Centromere
Acrocentric (13,14, 15, 21, 22, Y)

✓ Homologous chromosomes typically have the same genes in the same order. However, these genes may be different in sequence: different forms of a gene are called '**alleles**'

✓ Nuclear genome versus mitochondrial genome = circular DNA molecule (16kb)

The Human Genome

Gene Structure and Function

Unique vs Repetitive DNA

Unique versus Repetitive DNA sequences

Unique or Single-copy DNA

DNA whose linear order of specific nucleotides is represented only once around the entire genome

“ALWAYS REMEMBER THAT YOU ARE ABSOLUTELY UNIQUE. JUST LIKE EVERYONE ELSE.

Repetitive DNA

Repeated nucleotide sequences:

✓ **clustered** tandem repeats (‘satellite’)

e.g. short sequence repeats on Y

e.g. 171bp repeats at the centromere

✓ **dispersed** repetitive elements

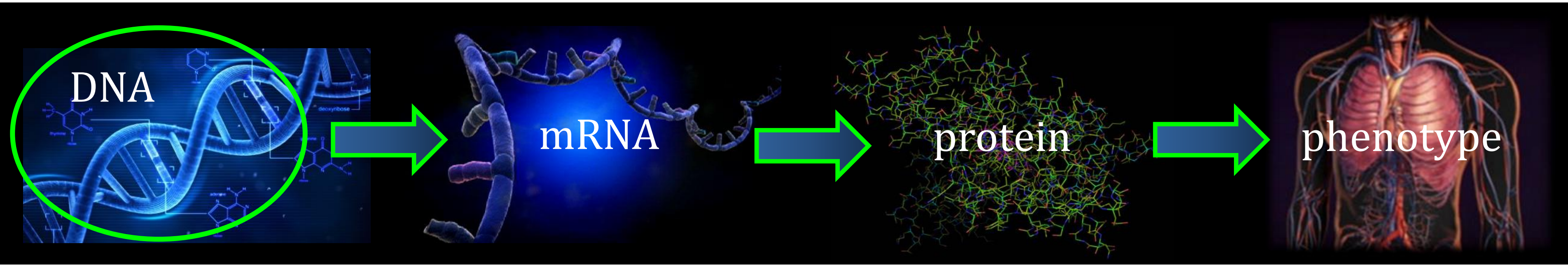
- SINE: e.g. Alu repeats – GT (10%)
- LINE: 6 kb in length – AT (20%)

✓ **segmental duplications:**

- duplicated sequences
- often highly conserved
- > several kb (5%)
- aberrant recombination

Genes

KATHOLIEKE UNIVERSITEIT LEUVEN

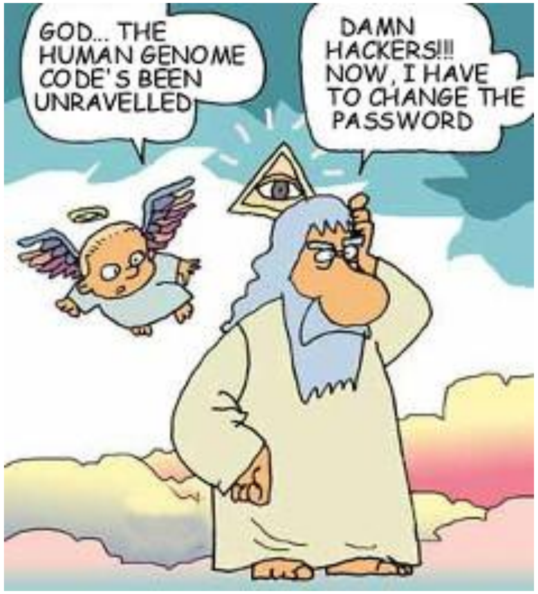


nucleus	*nucleus* - *cytoplasm*	*cytoplasm, organelles, extracellular*	*tissues, organs, organism*

GENOTYPE → PHENOTYPE

DNA sequence
= 3,324,592,091 bp (3 billion) (hg19)

- ✓ coding genes: 20,769
- ✓ short non coding genes: 9,079
- ✓ long non coding genes: 13,564
- ✓ pseudogenes: 14,165



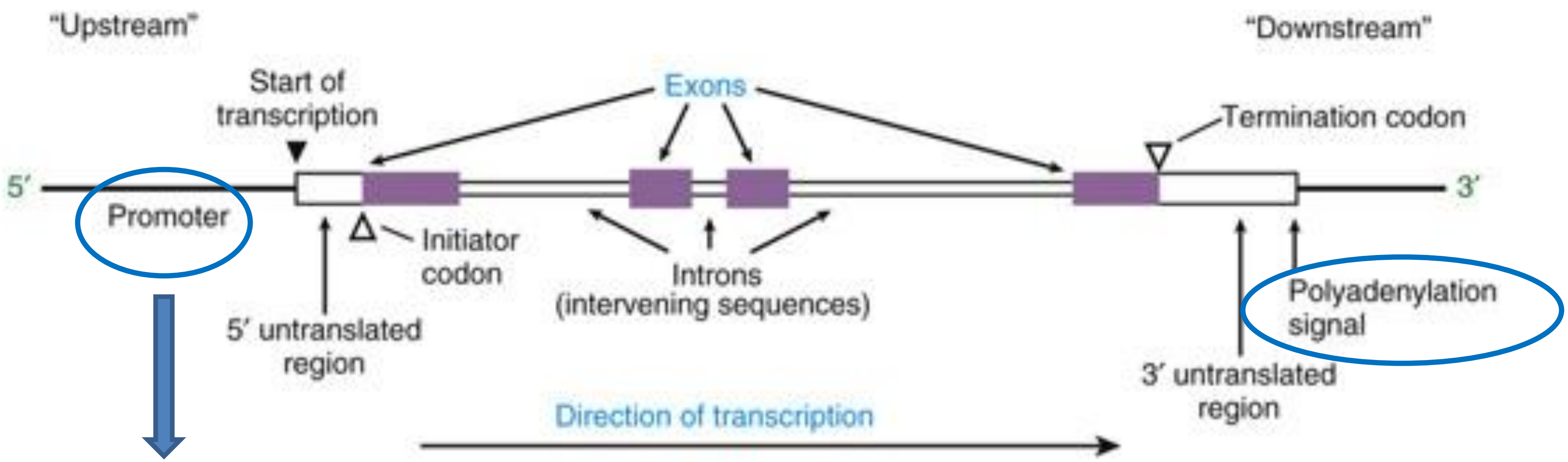
features of a typical human gene

KATHOLIEKE UNIVERSITEIT LEUVEN

DEFINITION:
a gene is a region of DNA that controls a discrete hereditary characteristic, usually corresponding to a single mRNA which will be translated into a protein (coding genes). Some genes encode a functional RNA molecule which is not translated into a polypeptide (non-coding genes)

DOMINANT GENE

Features of a typical coding human gene



- ✓ Promotor region with TF binding site to initiate transcription
- ✓ other regulatory elements (**enhancers, silencers and locus control regions**) can lie at the 5' or 3' of a gene, or can be intronic, and sometimes lie a significant distance away from the coding portion of a gene: 'genomic environment'
- ✓ in eukaryotes, genes have their coding sequences (exons) interrupted by non-coding sequences (introns)
- ✓ poly adenylation signal at the 3' UTR .

features of non-coding genes

KATHOLIEKE UNIVERSITEIT LEUVEN

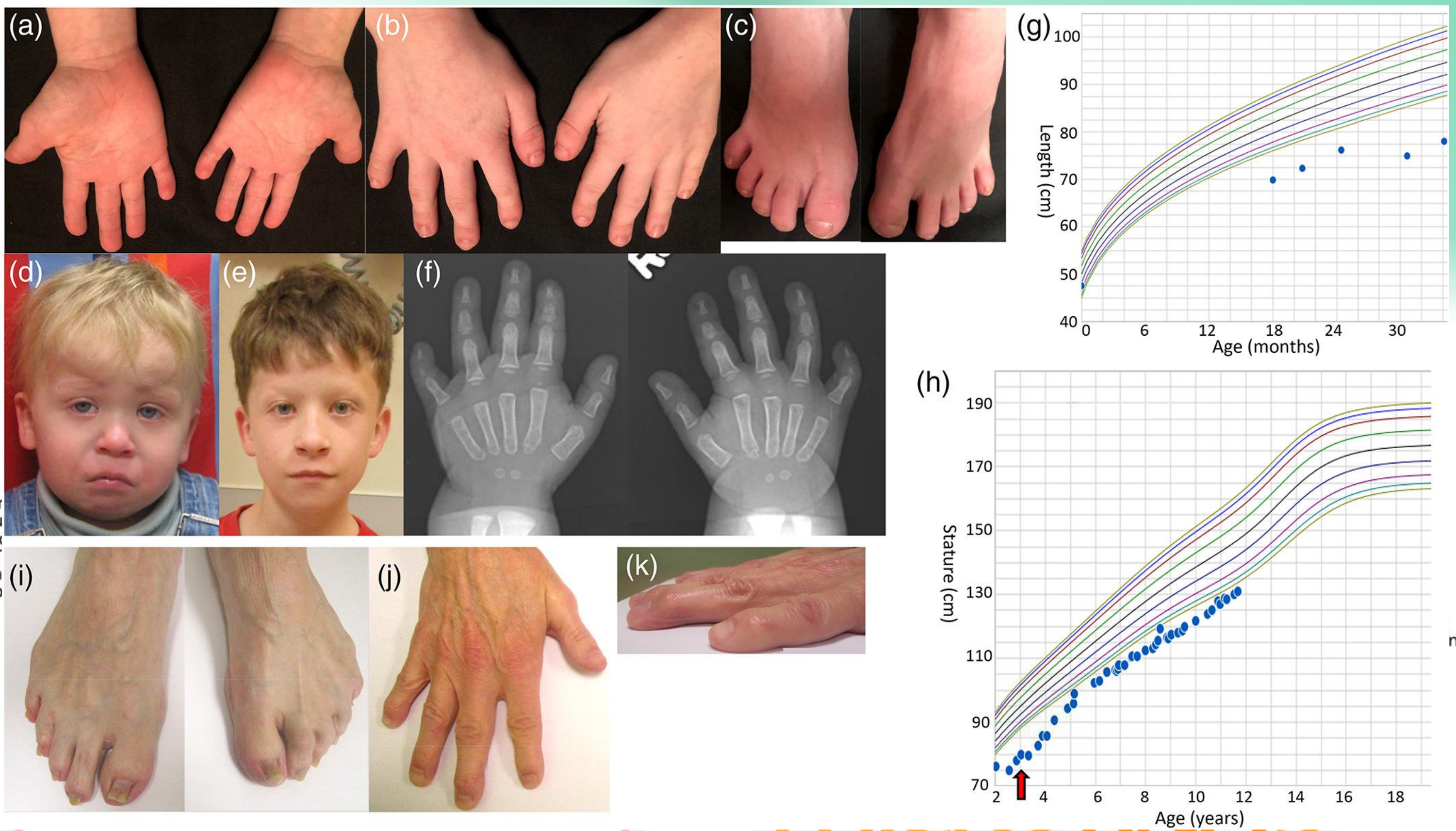
DEFINITION

DNA sequences that encode an untranslated functional RNA product

✓	tRNA	transfer RNA	translation
✓	rRNA	ribosomal RNA	translation
✓	snoRNA	small nucleolar RNA	modification of rRNA
✓	lncRNA	long non-coding RNA	gene regulation & silencing
✓	miRNA	microRNA	mRNA binding

miRNA

miRNA



Feingold syndrome caused by deletion of *MIR17HG* on 13q31

Muriello *et al.* Am J Med Genet A 2019

features of a pseudogene

DEFINITION

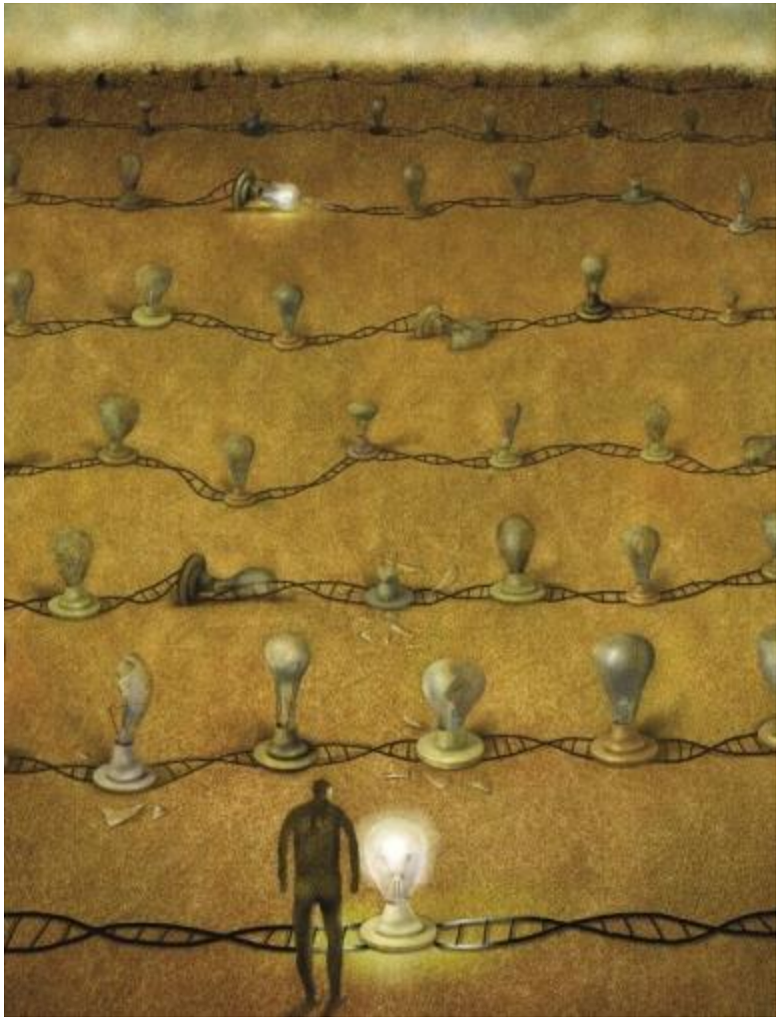
DNA sequences that closely resemble known genes but are afunctional

- ✓ nonprocessed pseudogenes

  'dead' genes: 'duplicates' which were inactivated by mutations in coding or regulatory sequences

- ✓ processed pseudogenes

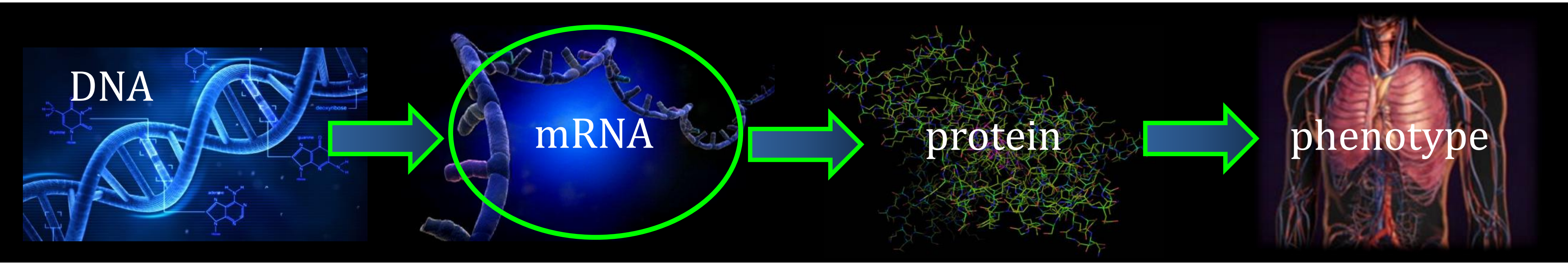
  formed by retrotransposition: reverse transcription of RNA followed by integration in genome → lack of introns



The Human Genome

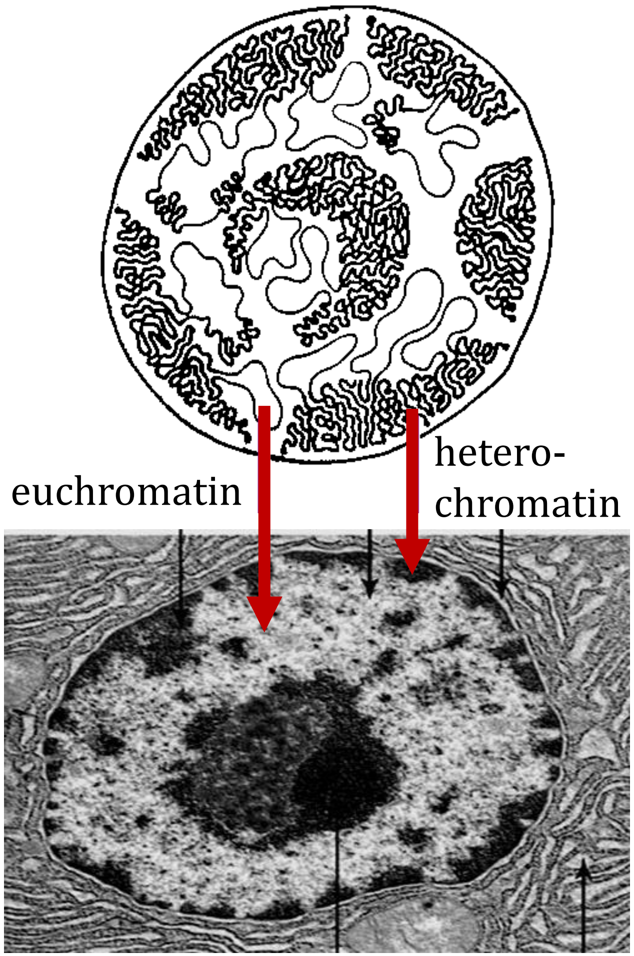
Basic Principles of Transcription

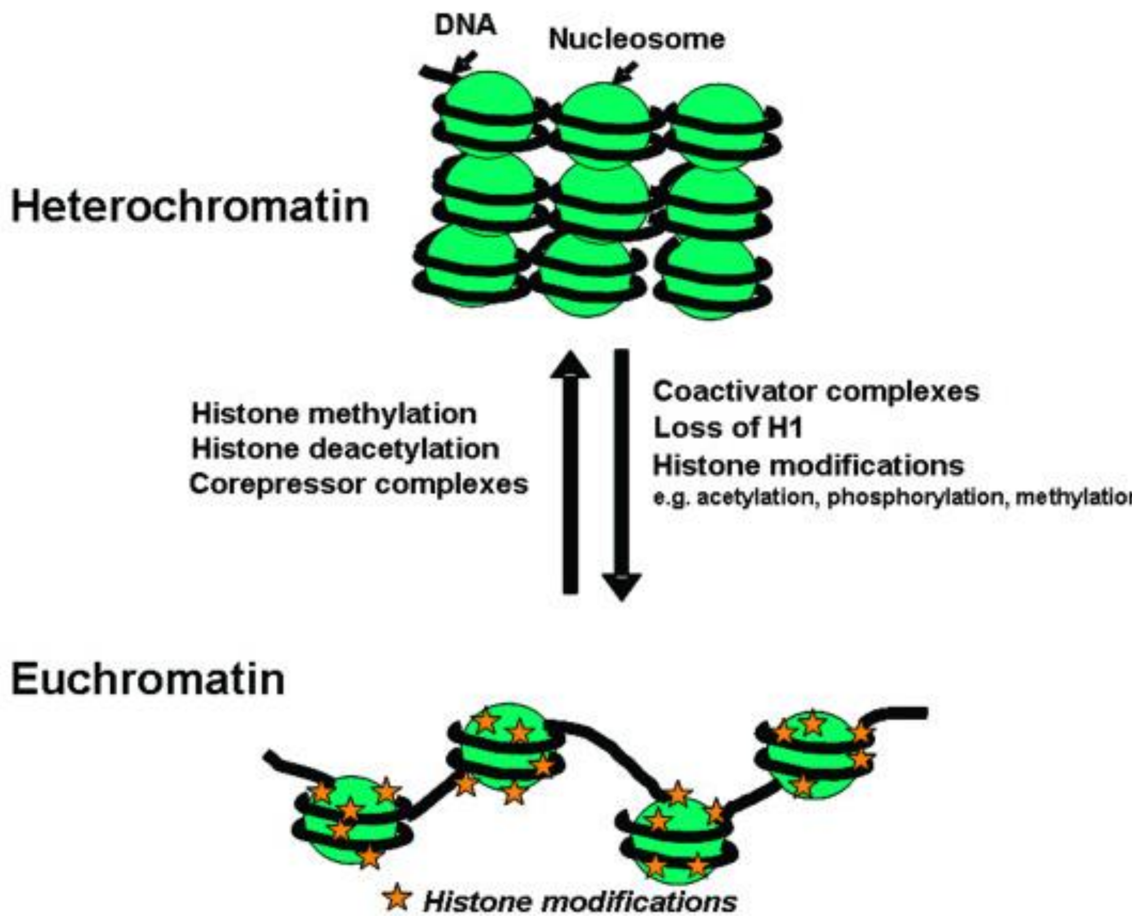
Initiation of Transcription



Initiation of transcription

- Chromatin remodeling required to make the DNA accessible to transcription factors





Initiation of transcription

KATHOLIEKE UNIVERSITEIT LEUVEN

THE GENETIC BASIS OF THE REDUCED EXPRESSION OF BILIRUBIN UDP-GLUCURONOSYLTRANSFERASE 1 IN GILBERT'S SYNDROME

PITER J. BOSMA, PH.D., JAYANTA ROY CHOWDHURY, M.D., CONNY BAKKER, SHAILAJA GANTLA, PH.D., ANITA DE BOER, BEN A. OOSTRA, PH.D., DICK LINDHOUT, PH.D., GUIDO N.J. TYTGAT, M.D., PETER L.M. JANSEN, M.D., PH.D., RONALD P.J. OUDE ELFERINK, PH.D., AND NAMITA ROY CHOWDHURY, PH.D.

Abstract *Background.* People with Gilbert's syndrome have mild, chronic unconjugated hyperbilirubinemia in the absence of liver disease or overt hemolysis. Hepatic glucuronidating activity, essential for efficient biliary excretion of bilirubin, is reduced to about 30 percent of normal.

Methods. We sequenced the coding and promoter regions of the gene for bilirubin UDP-glucuronosyltransferase 1 (bilirubin/uridine diphosphoglucuronate-glucuronosyltransferase 1) — the only enzyme that contributes substantially to bilirubin glucuronidation — in 10 unrelated patients with Gilbert's syndrome, 16 members of a kindred with a history of Crigler–Najjar syndrome type II, and 55 normal subjects.

Results. The coding region of the gene for the enzyme was normal in the 10 patients with Gilbert's syndrome. These patients were homozygous for two extra bases (TA) in the TATAA element of the 5′ promoter region of the gene ($A(TA)_7TAA$ rather than the normal $A(TA)_6TAA$). The presence of the longer TATAA element resulted in the reduced expression of a reporter gene, encoding firefly luciferase, in a human hepatoma cell line. The frequency of the abnormal allele was 40 percent among the normal subjects. The 3 men in the control group who were homozygous for the longer TATAA element had significantly higher serum bilirubin levels than the other 52 normal subjects ($P = 0.009$). Among the kindred with a history of Crigler–Najjar syndrome type II, only the six heterozygous carriers who had a longer TATAA element on the structurally normal allele had mild hyperbilirubinemia, characteristic of Gilbert's syndrome.

Conclusions. Reduced expression of bilirubin UDP-glucuronosyltransferase 1 due to an abnormality in the promoter region of the gene for this enzyme appears to be necessary for Gilbert's syndrome but not sufficient for the complete manifestation of the syndrome. (N Engl J Med 1995;333:1171-5.)

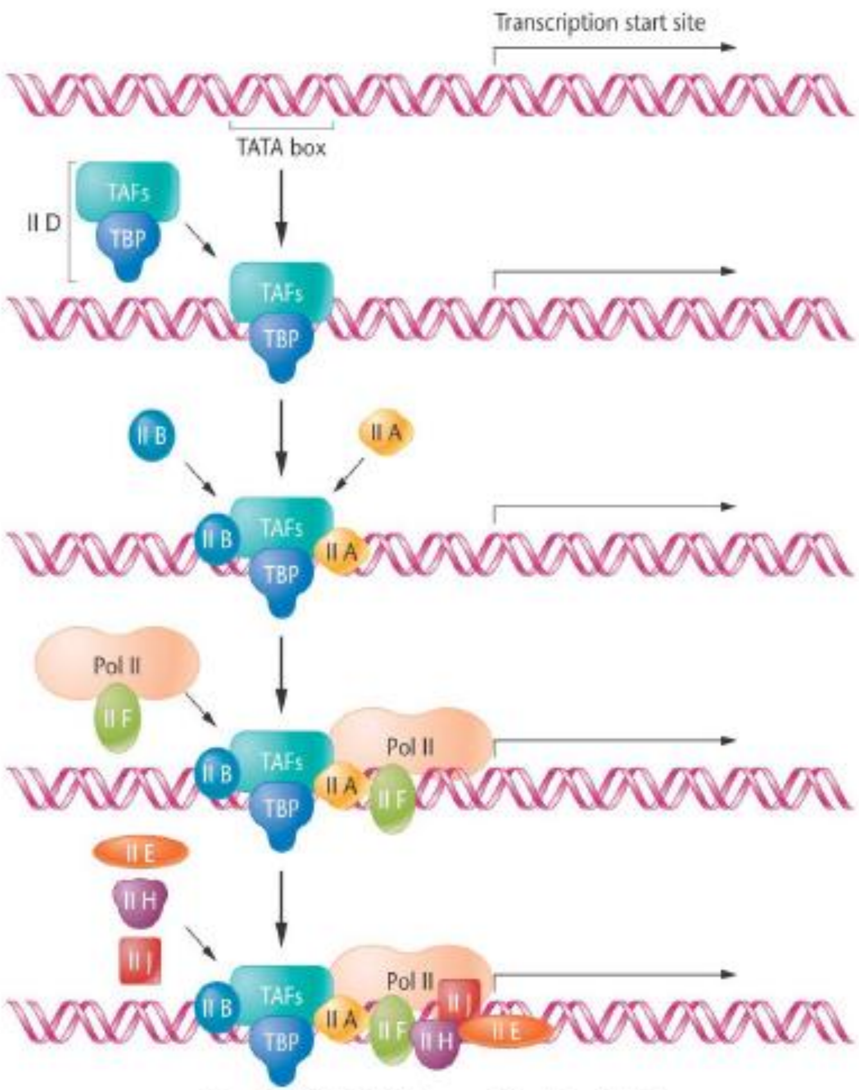
Initiation of transcription

assembling basal initiation complex

RNA polymerase + TF

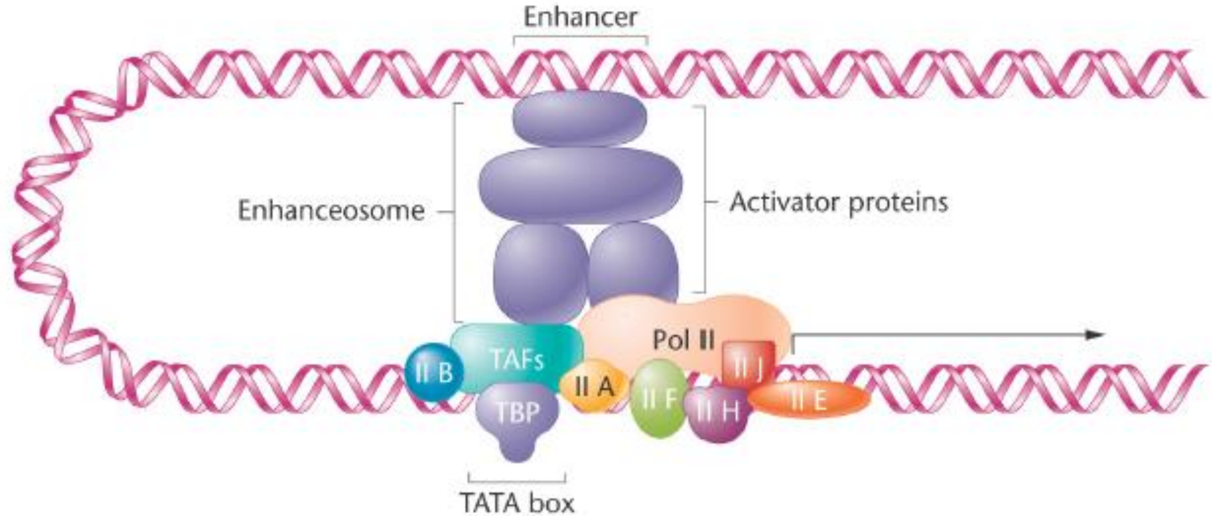
- ✓ RNA polymerase I: rRNA
- ✓ RNA polymerase II: mRNA, miRNA, snRNA, siRNA
- ✓ RNA polymerase III: tRNA, 5S rRNA,...

Each RNA polymerase has its own promotor characteristics and transcription factors (some are shared)

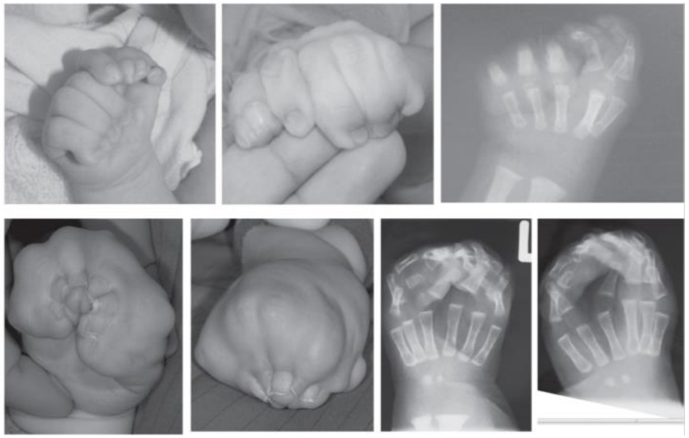


Enhanceosome

- ✓ Transcription factors often have two domains: DNA binding domain and protein binding domain (interaction with other TF or RNAP)
- ✓ Most transcription factors have multiple targets, and most promotor regions are targeted by multiple transcription factors
- ✓ Activator proteins can interact with regulatory elements (**silencers or enhancers**) that are at a distance from the promotor region (5', 3' or intronic): '**chromosomal loops**' -> topologically associating domains



Enhanceosome

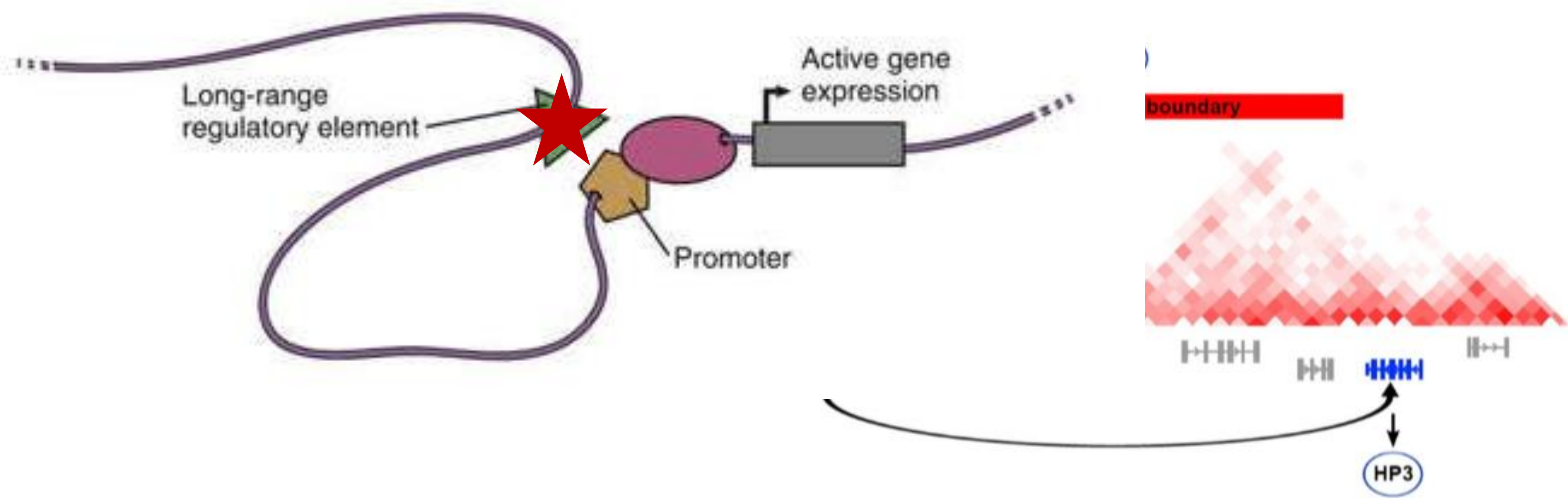


Clin Genet 2014: 86: 318–325

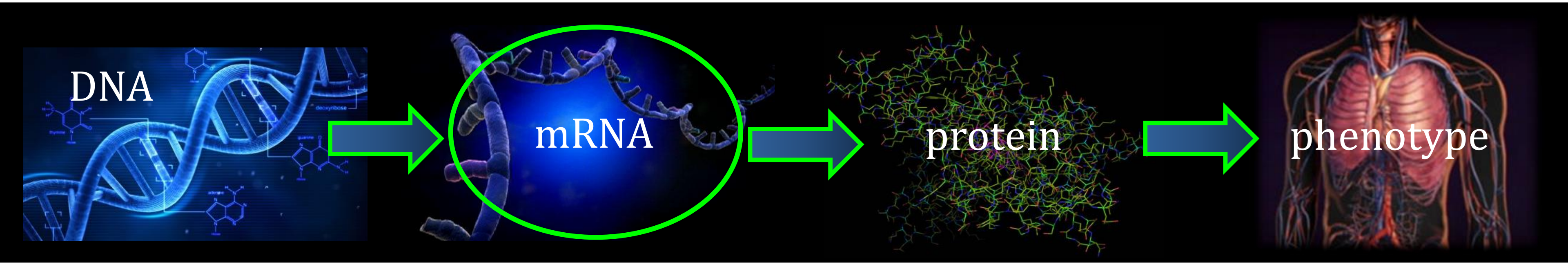
CLINICAL GENETICS
doi: 10.1111/cge.12352

Original Article

Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome

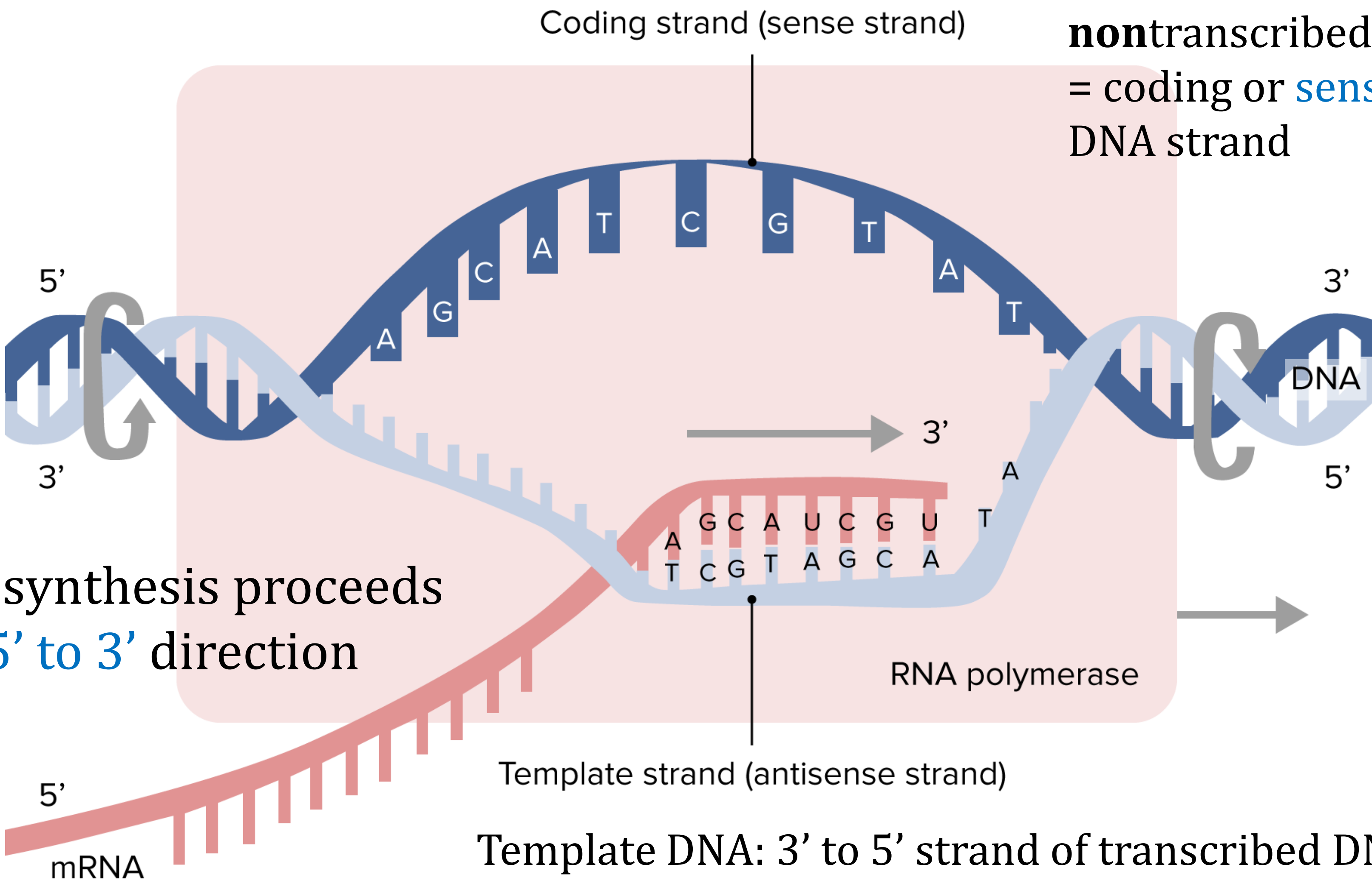


mRNA synthesis



mRNA synthesis

5’ to 3’ strand of **non**transcribed DNA = coding or sense DNA strand



RNA synthesis proceeds in a 5’ to 3’ direction

Template DNA: 3’ to 5’ strand of transcribed DNA = non-coding or antisense strand

mRNA synthesis

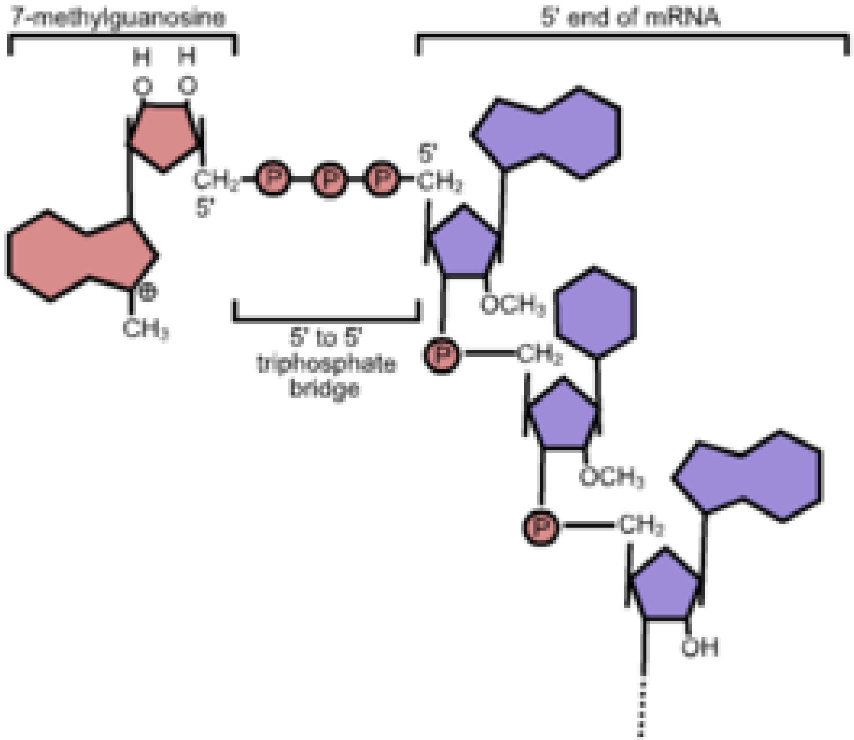
Differences between replication and transcription

	replication	transcription
template	double strands	single strand
substrate	dNTP	NTP
primer	yes	no
Enzyme	DNA polymerase	RNA polymerase
product	dsDNA	ssRNA
base pair	A-T, G-C	A-U, T-A, G-C

no proof reading

mRNA processing

1. **5' cap**: after 20-30 nucleotides have been synthesized, the 5'cap of the mRNA is capped.
   - ✓ Guanine is connected to the 5' of mRNA by 5' to 5' triphosphate linkage.
   - ✓ The guanosine is methylated at the 7 position: m7G (7-methylguanylate)

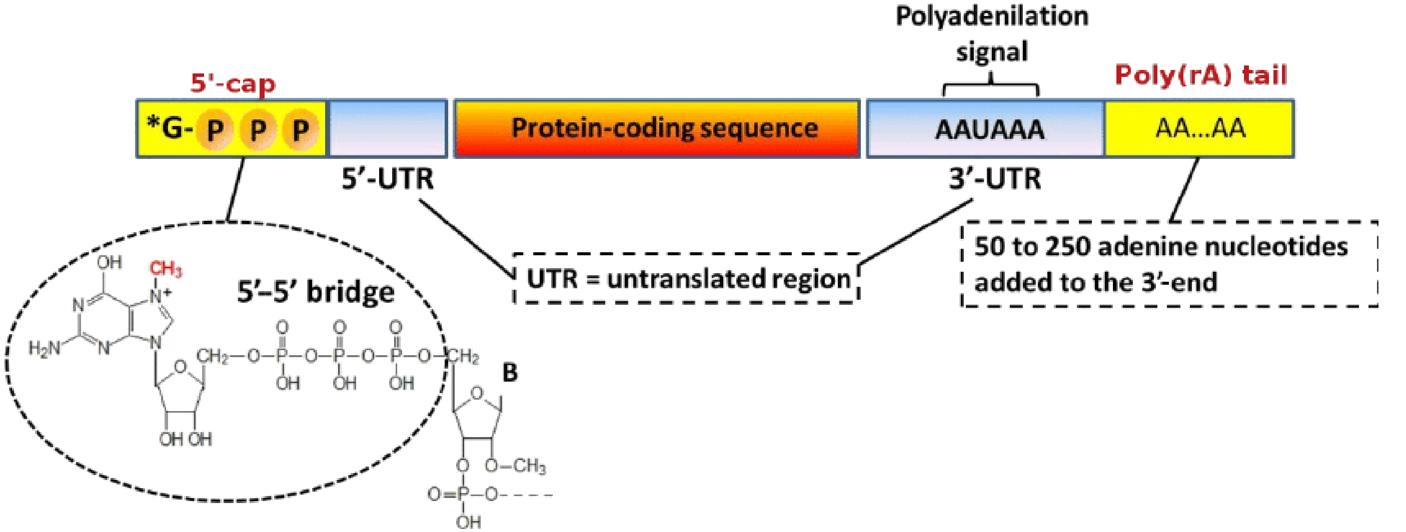


Function of the 5' cap:

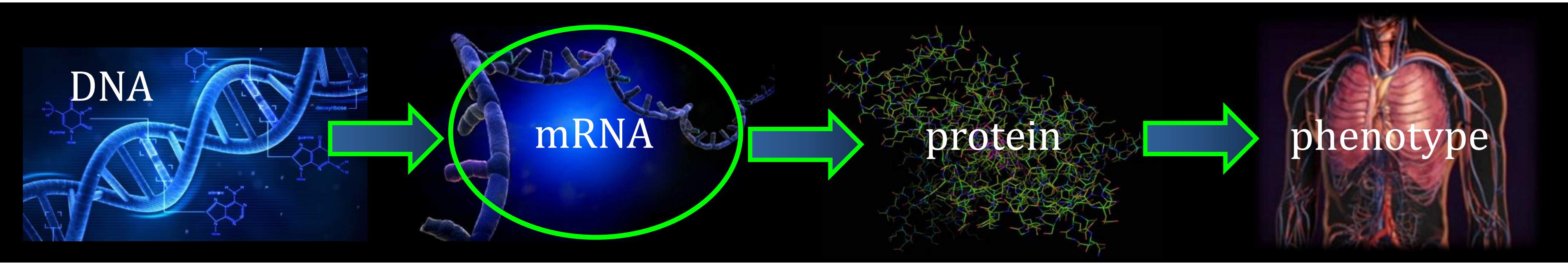
1. Regulation of nuclear export
2. Prevention of degradation by exonucleases
3. Promotion of translation (interaction with ribosome)
4. Promotion of 5' proximal intron excision

mRNA processing

2. **Poly (A) tail:**
   - ✓ 50-250 adenine nucleotides are added to the 3' end of mRNA
   - ✓ poly(A)-tail is not coded by DNA, but is added by poly(A)-polymerase in a complex enzymatic reaction, initiated by detection of the polyadenylation signal (5'…AAUAAA…3').
   - ✓ stabilizes mRNA and is involved in transcription termination and nuclear export
   - ✓ mature forms of long ncRNAs have a poly(A) tail as well, whereas small RNAs, such as miRNA, don't.

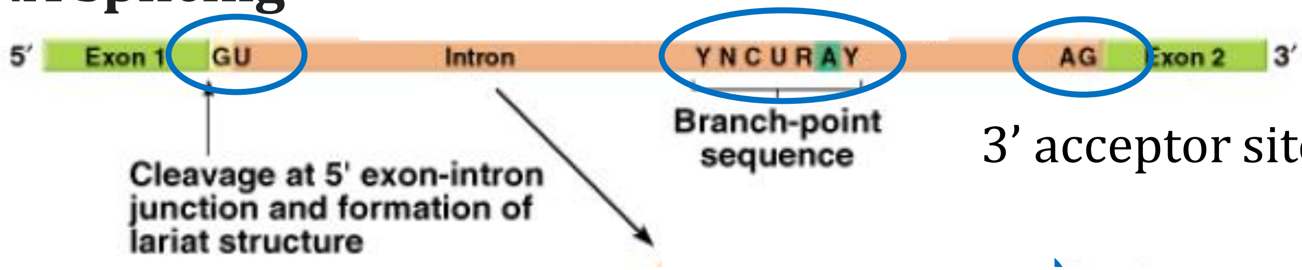


mRNA splicing



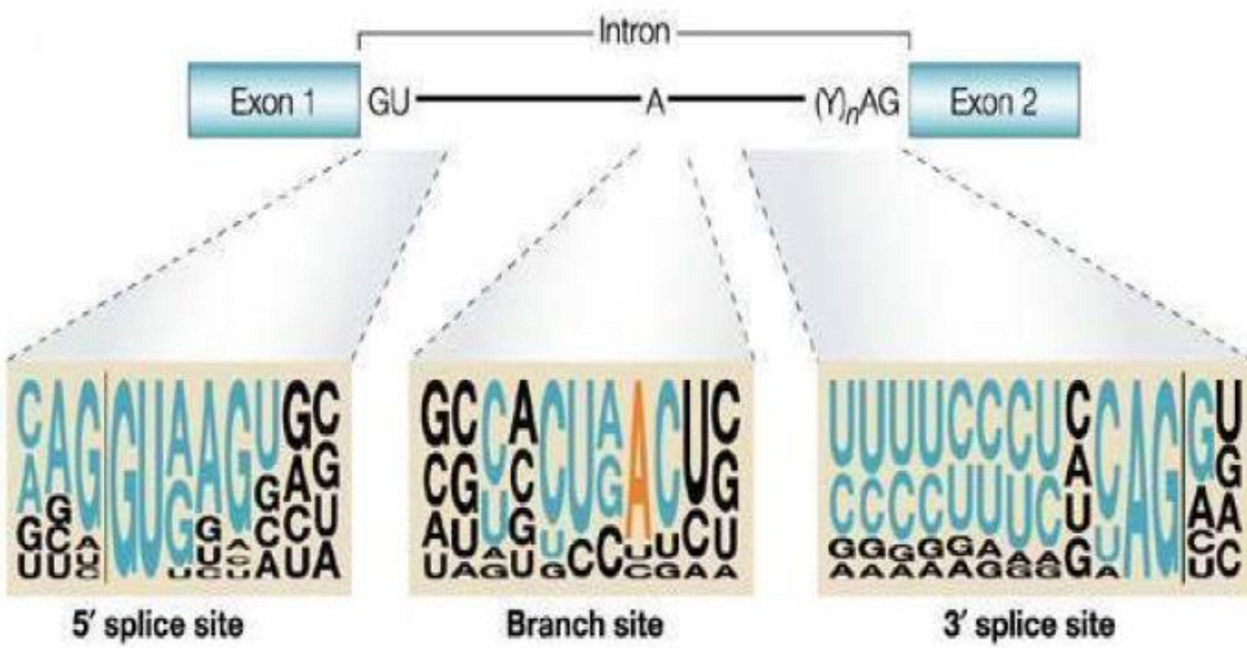
mRNA splicing

mRNA splicing



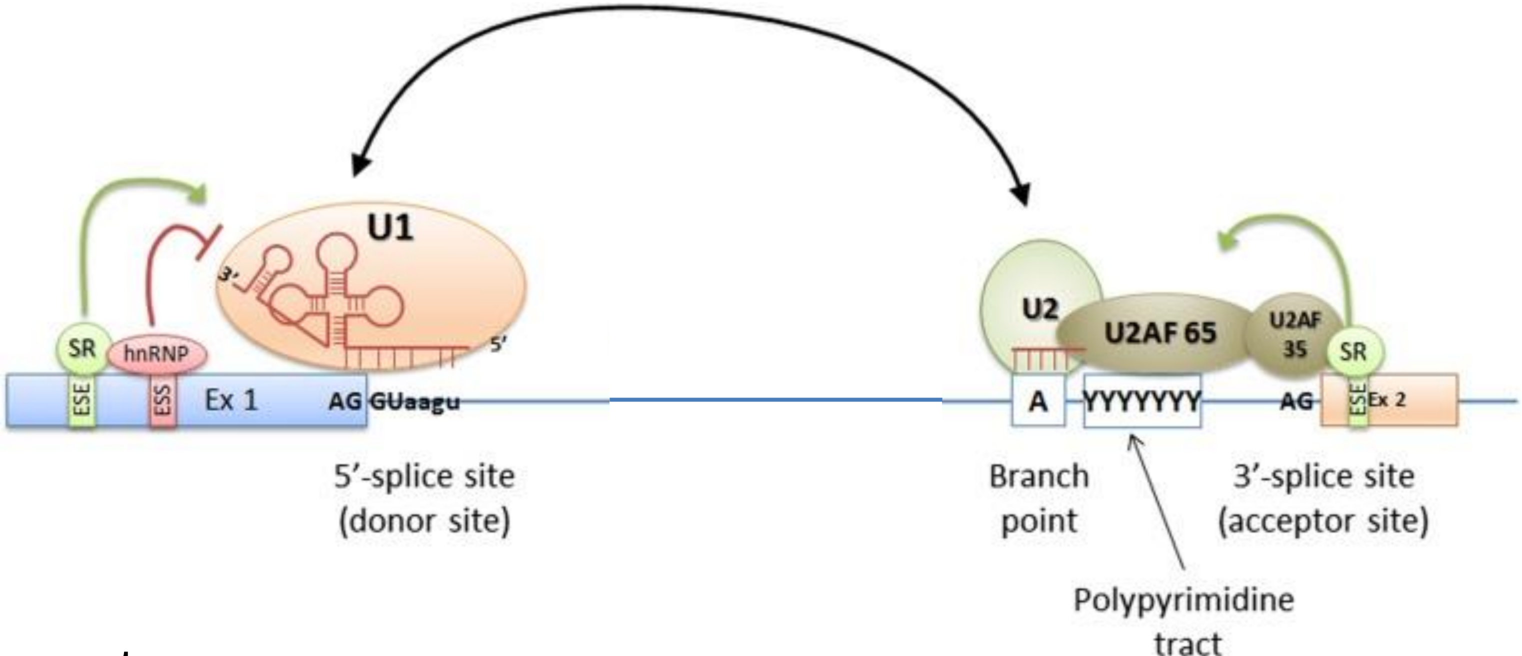
3’ acceptor site

5’ donor site



consensus sequence for an intron

Spliceosome



Cis elements:

- ✓ donor and acceptor sites, branch point and polypyrimidine tract
- ✓ splicing silencers and enhancers (DNA sequence)

Trans-acting elements:

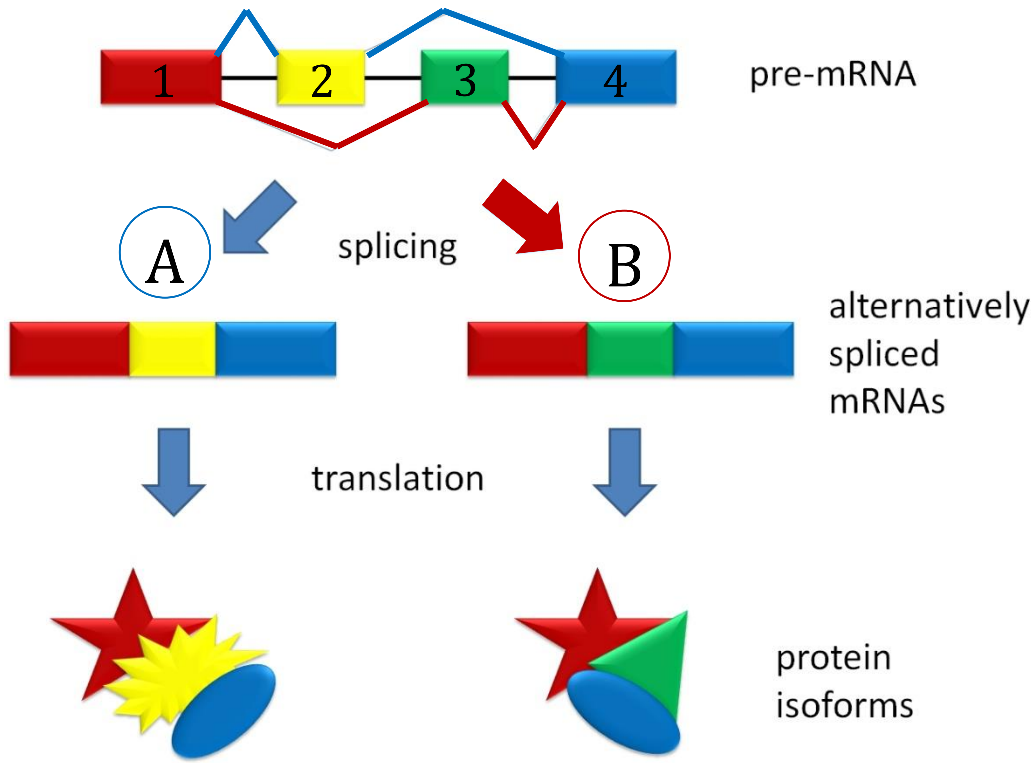
- ✓ spliceosome proteins
- ✓ splicing repressors and activators (SR proteins)

Adapted from Abramowicz and Gos, J Appl Genet 2018

alternative splicing

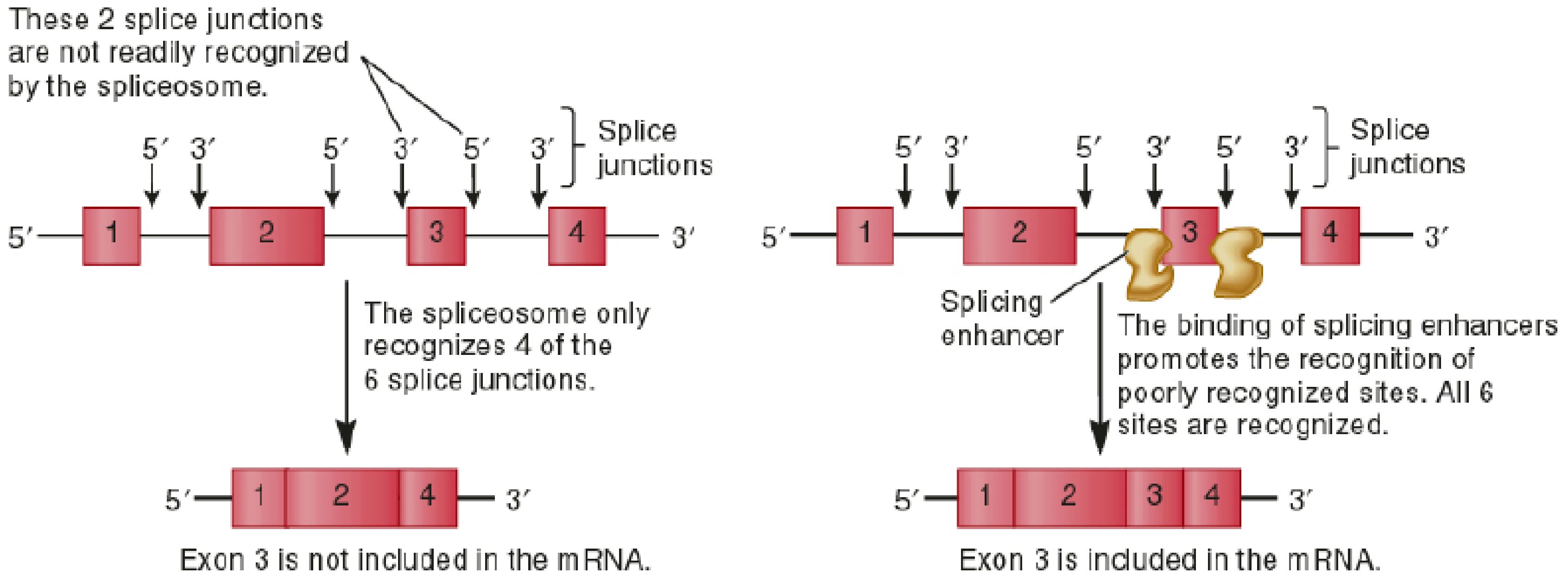
DEFINITION

the splicing process can create a range of unique proteins by varying the exon composition of the same mRNA.



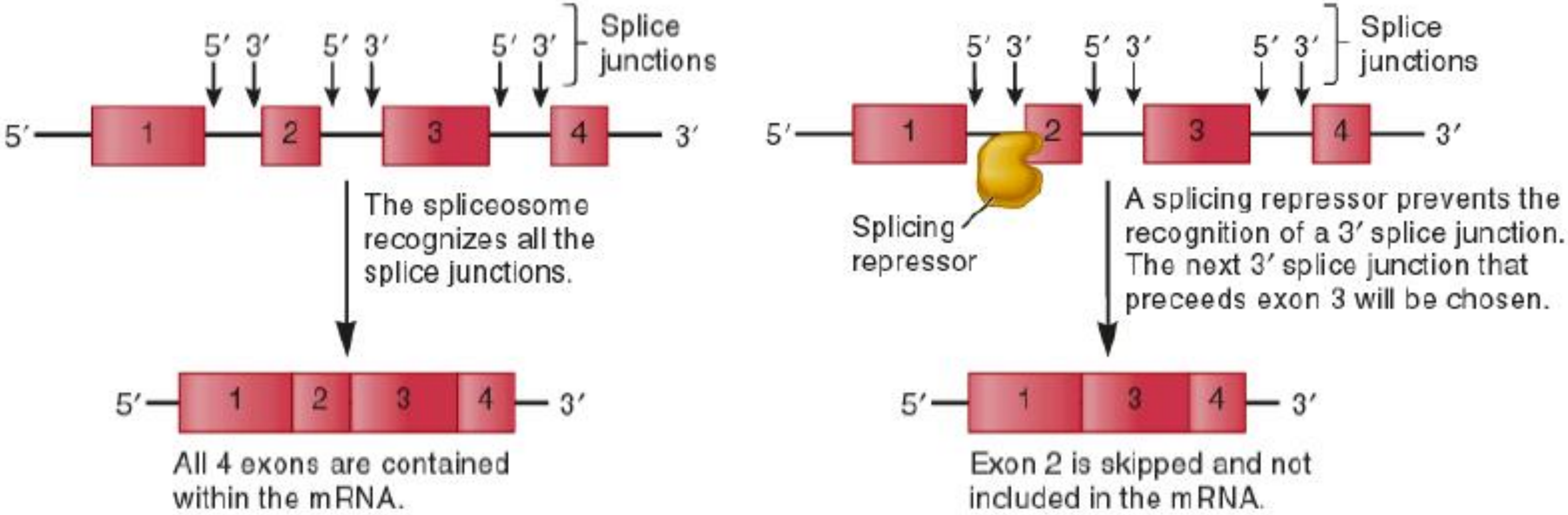
alternative splicing: enhancers

Splicing enhancers



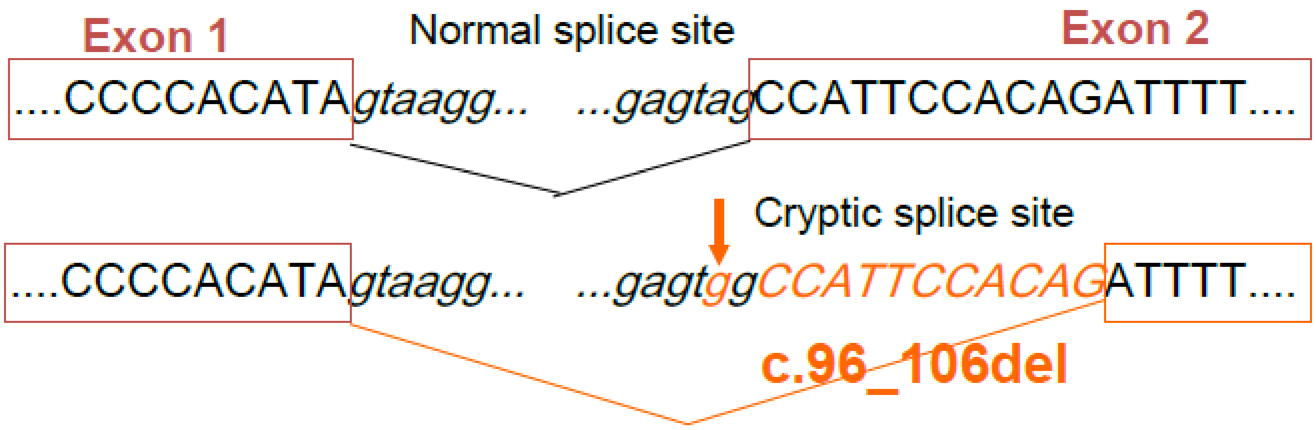
alternative splicing: repressors

Splicing repressors



splice site mutations

splice site mutations can activate a cryptic splice site in part of the transcript that usually is not spliced



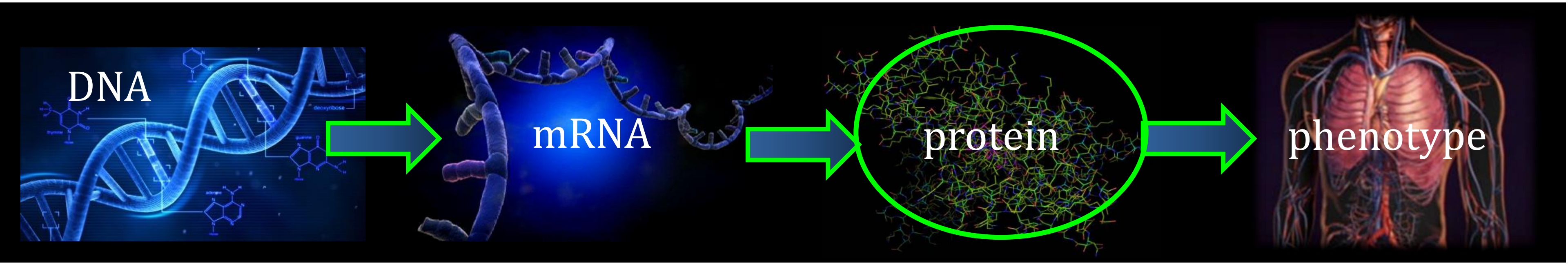
This results in a mature mRNA with a missing section of an exon.

The most classical mutations affect +1 and +2 residues at the 5′ donor splice site and −1 and −2 residues at the 3′ acceptor site.

The Human Genome

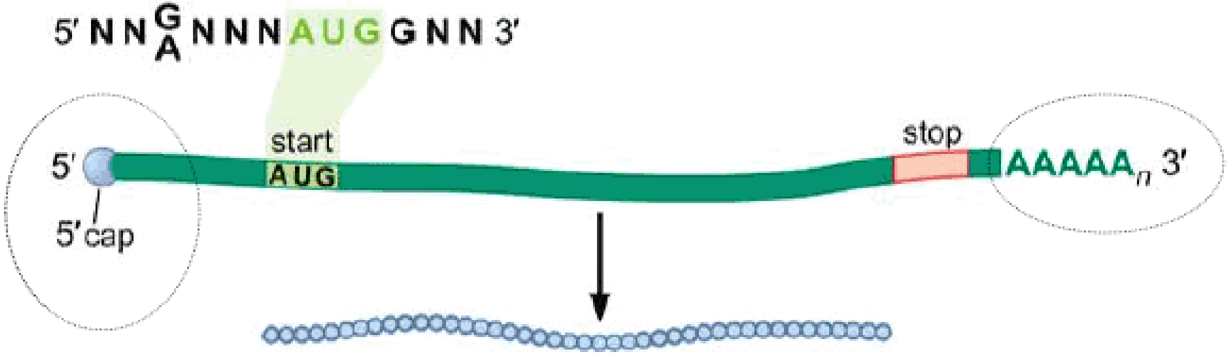
Basic Principles of Translation

Initiation of Translation

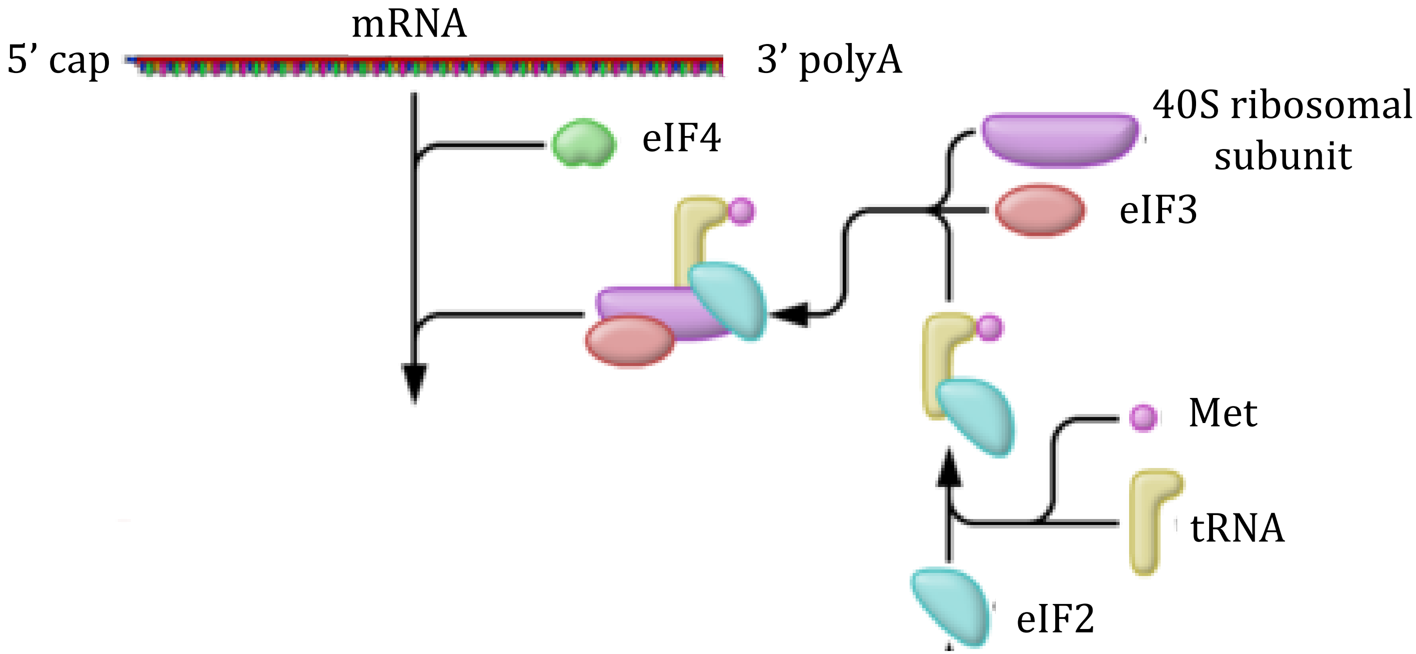


Initiation of translation

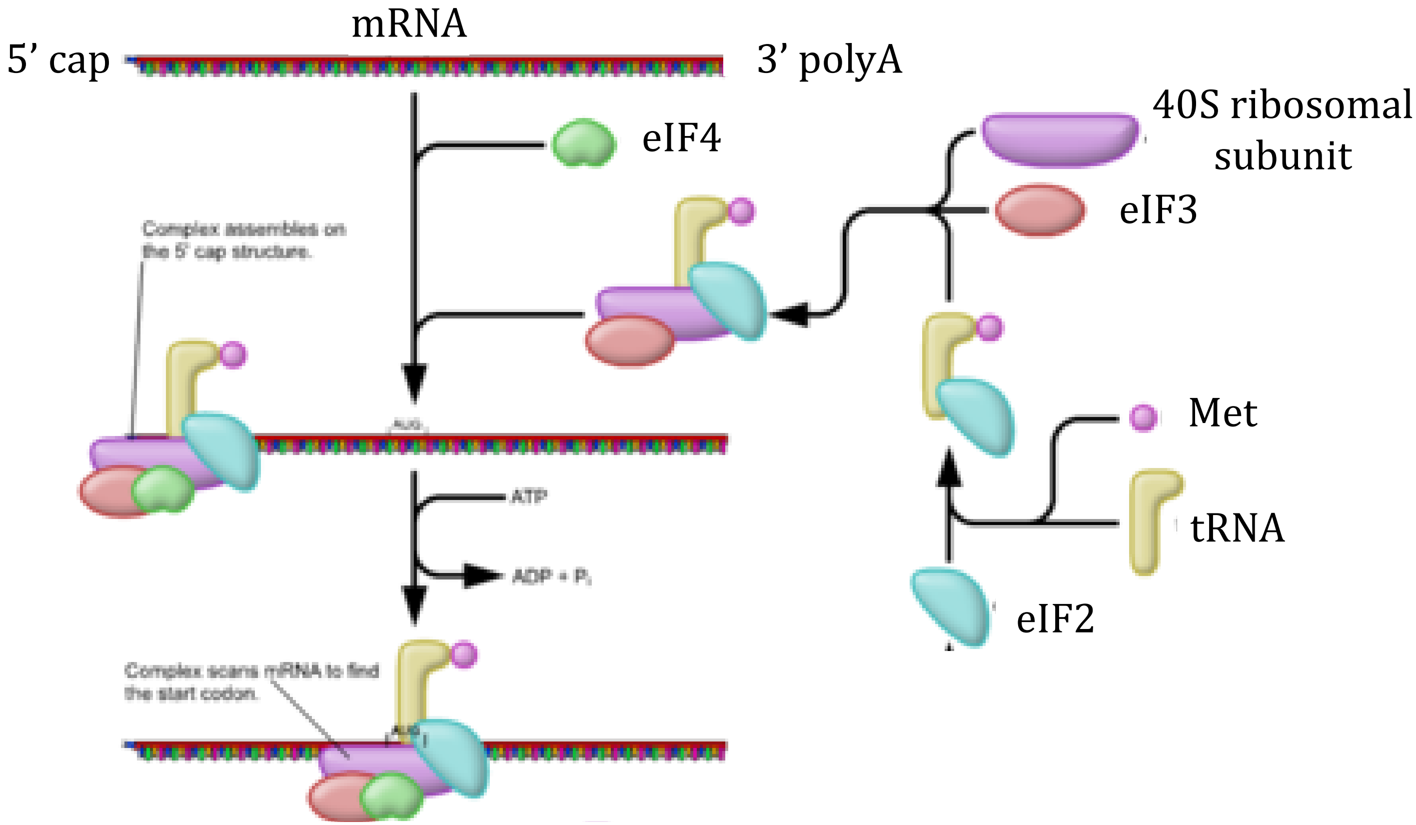
Eukaryotic mRNAs possess a 5'end cap and are polyadenylated.



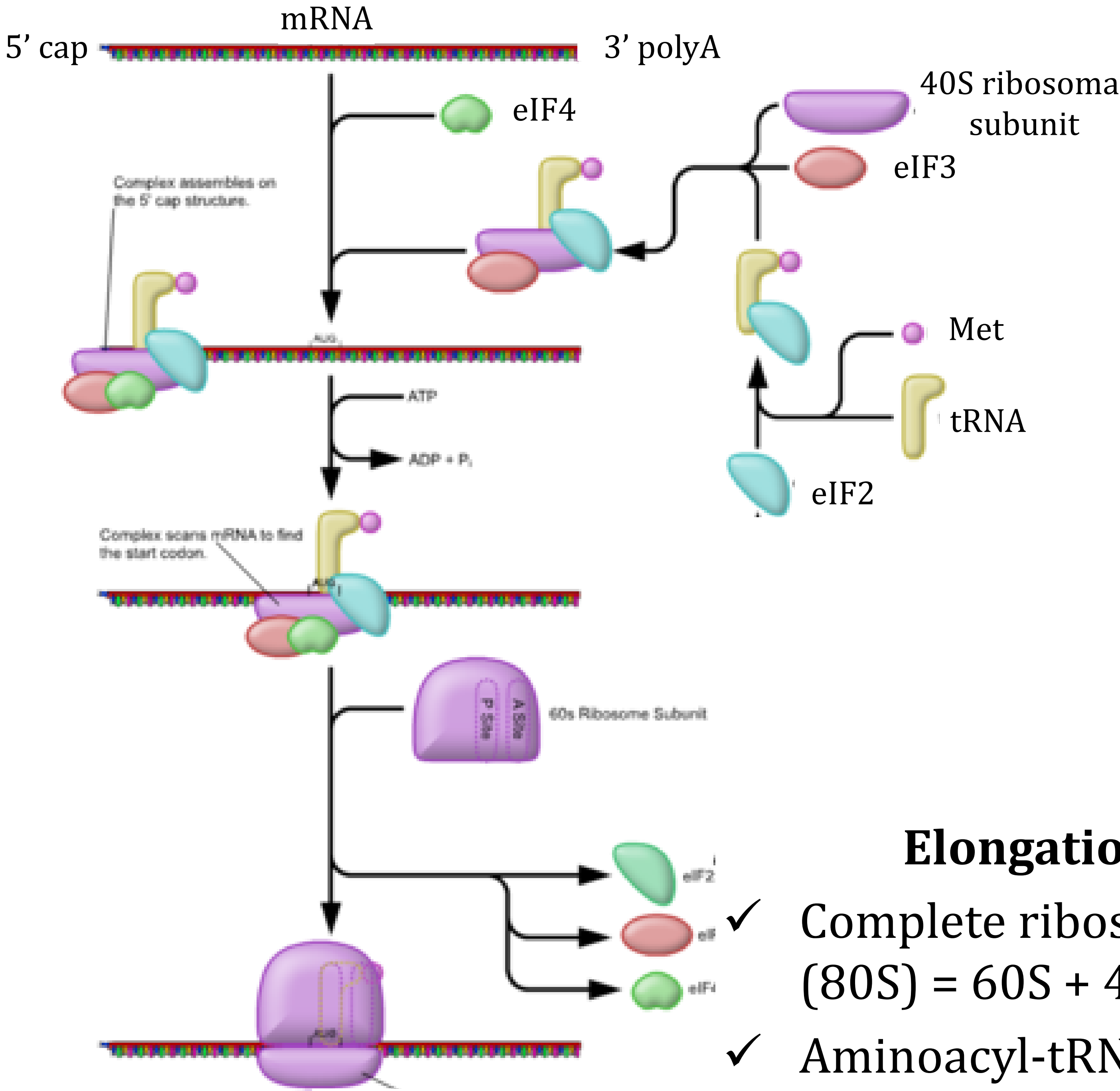
1. The 5' cap interacts with the translation initiation complex.
2. mRNA is translated starting from codon AUG (Methionin)
3. mRNA strand is read in direction from 5' to 3'



1. 5’cap and UTR interact with pre-initiation complex:
   - ✓ 40S ribosomal subunit
   - ✓ eIF3 eukaryotic initiation factor 3
   - ✓ eIF2 + initiator tRNA with Methionine (start codon AUG)
   - ✓ eIF4 (eIF4A, eIF4E, eIF4F, eIF4G)



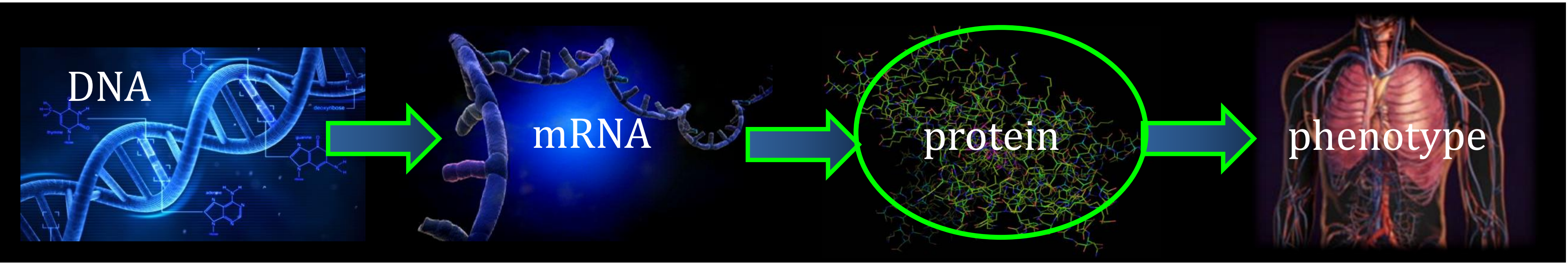
2. Eukaryotic initiation factor eIF4 scans along mRNA from 5’ cap to **find the start codon AUG**: bases around the initiating AUG influence the efficiency of initiation (Kozak consensus sequence)



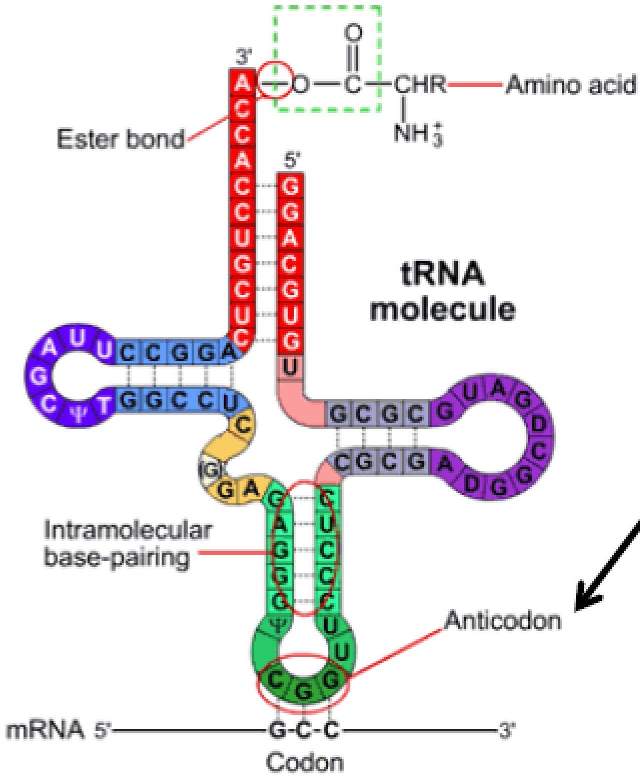
Elongation:

- ✓ Complete ribosomal unit (80S) = 60S + 40S
- ✓ Aminoacyl-tRNA

Elongation of Translation



Elongation: protein synthesis



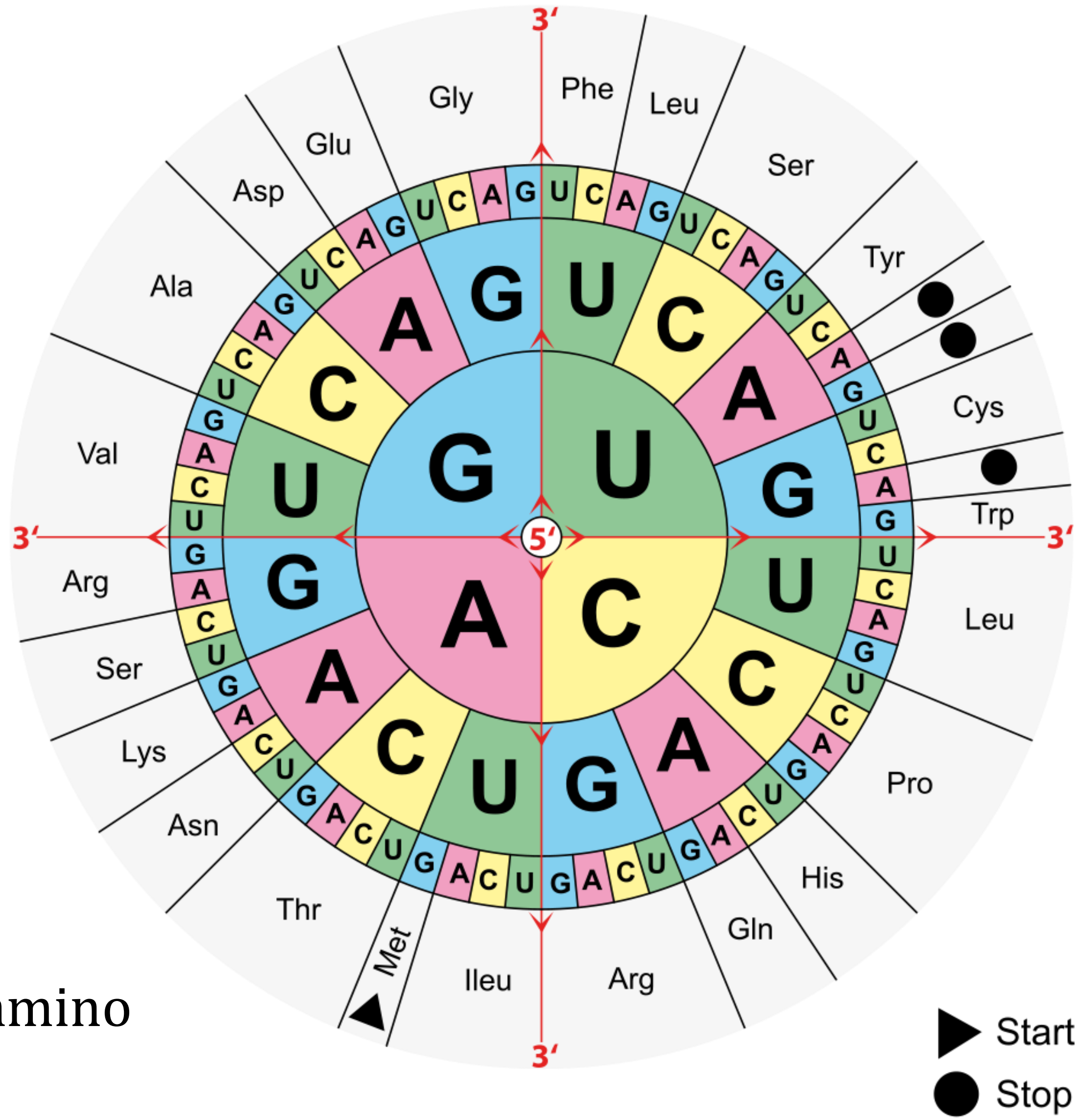
- Transfer RNA
- Bound to one amino acid on one end
- Anticodon on the other end complements mRNA codon

Elongation: protein synthesis

KATHOLIEKE UNIVERSITEIT LEUVEN

Codon: $4^3 = 64$ codons

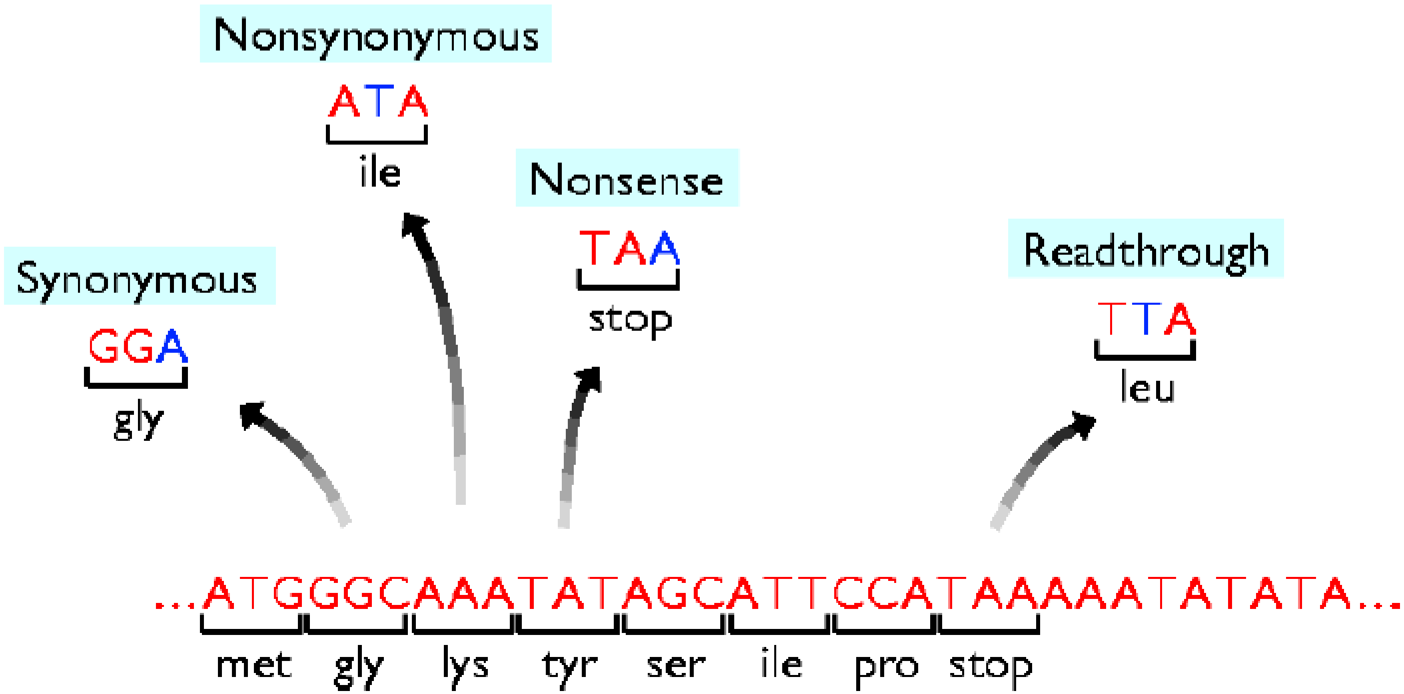
- Start codon: AUG (Met)
- Stop codon:
  - ✓ UGA
  - ✓ UAA
  - ✓ UAG



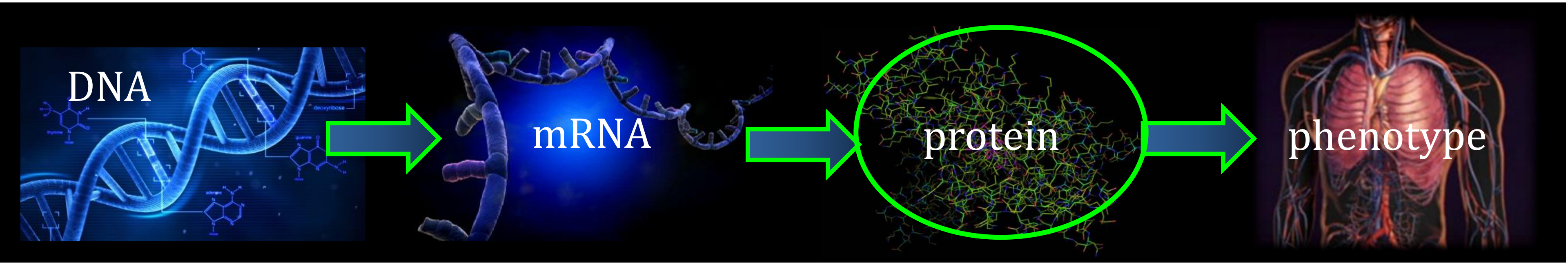
20 naturally occurring amino acids:

- ✓ Met = AUG (start)
- ✓ 60 codons for 19 other amino acids

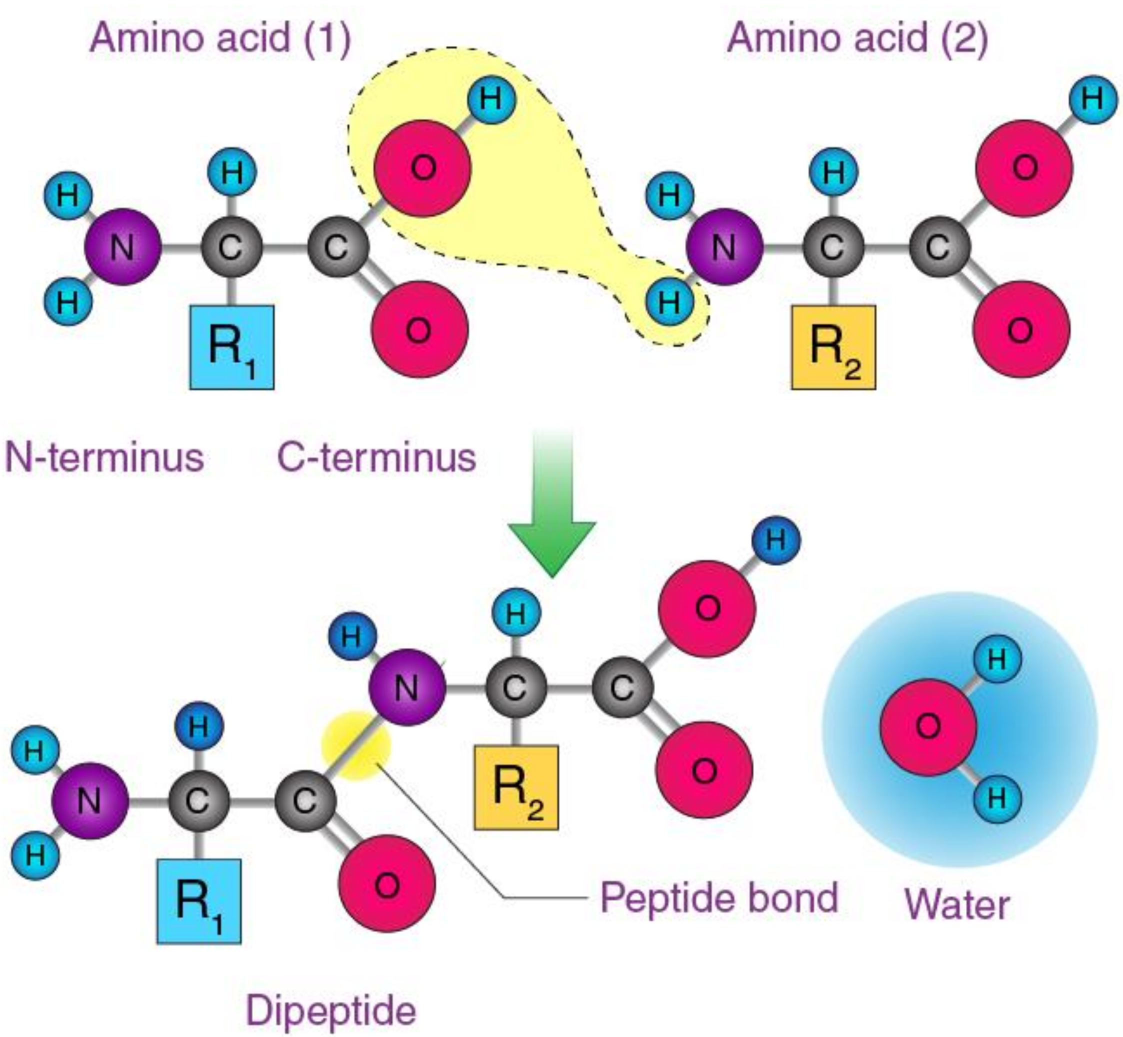
DNA variants



Protein synthesis

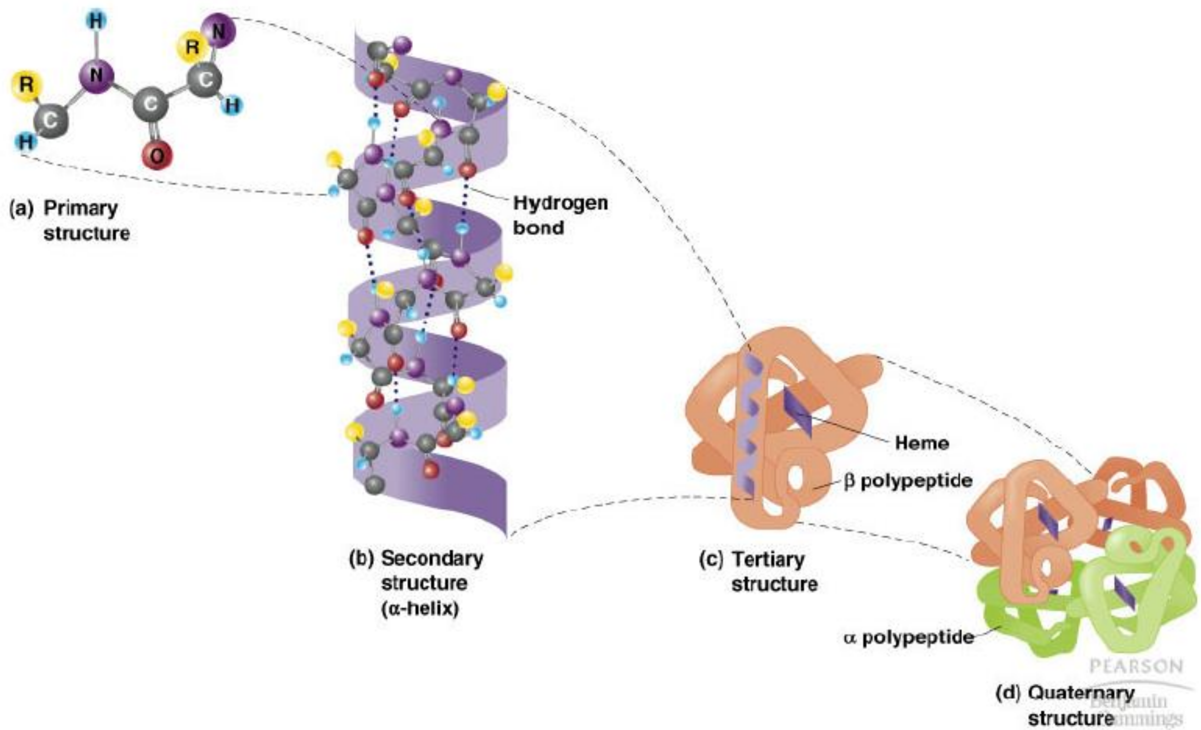


Protein synthesis



Protein synthesis

- A protein is a linear polymer of amino acids linked together by peptide bonds.
  - protein functions: structure, catalysis of reactions, …
  - quaternary structural levels
  - glycosylation, methylation, phosphorylation,…



The Human Genome

Basic Principles of Regulation of Gene Expression

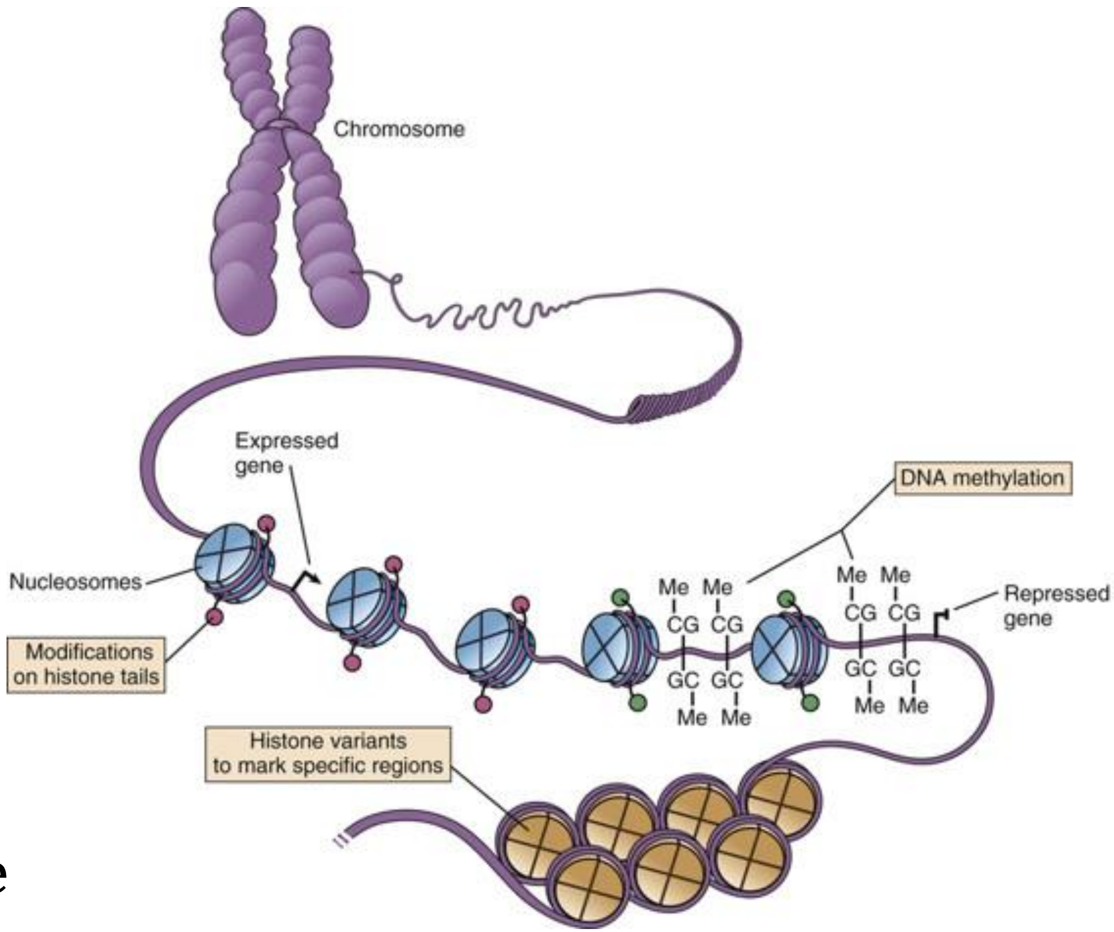
Levels of Control

Regulation of transcription by chromatin changes due to:

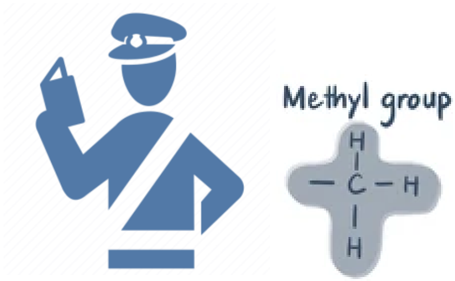
1. DNA methylation
2. Histone modification
3. Histone variants

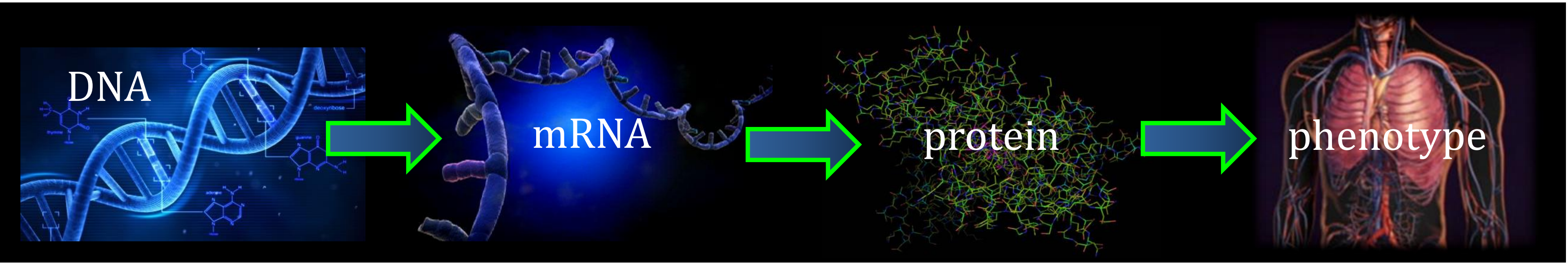
***Epi**genetic changes:*

- ✓ do not alter the underlying DNA sequence
- ✓ transient or long lasting



DNA methylation



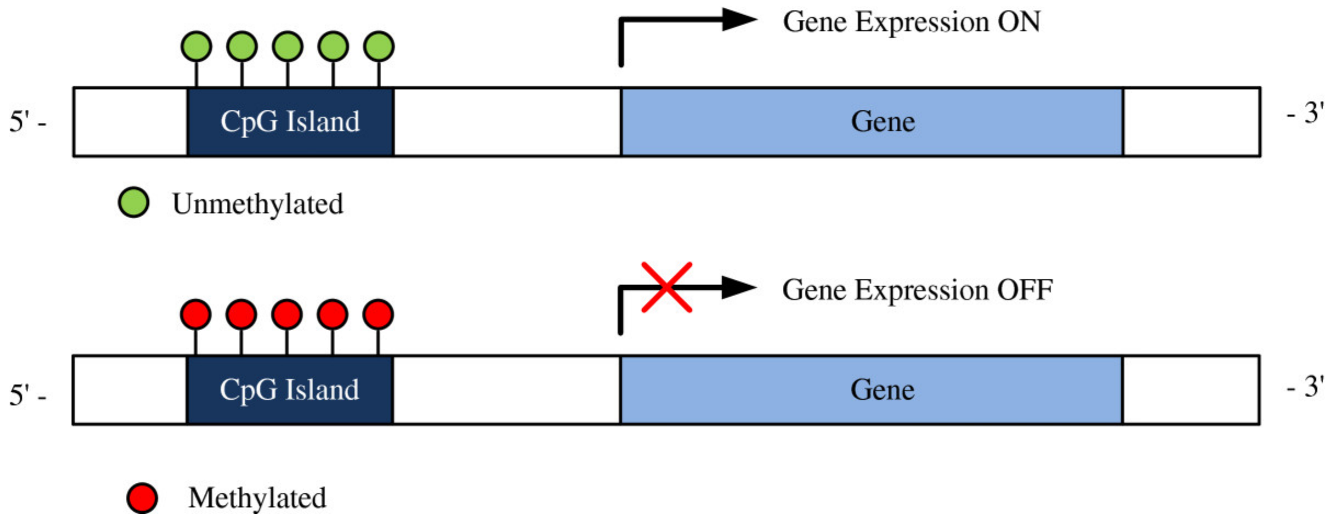


CpG islands

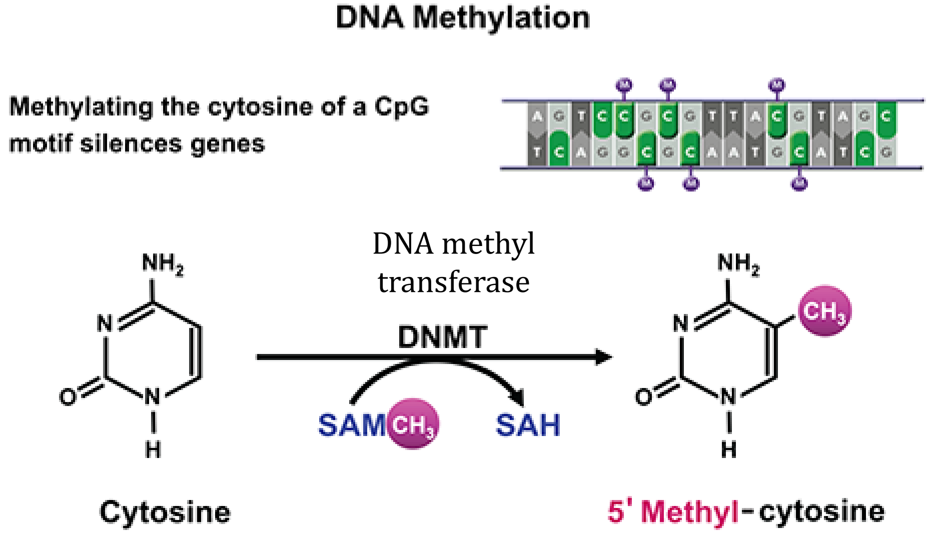
Promotors of constitutively and ubiquitously expressed genes ('house-keeping genes') have a high proportion of G and C in relation to the surrounding DNA: **CpG islands** (5'-CpG-3'):

- ✓ binding sites for TF
- ✓ targets for DNA methylation:

repression of gene transcription



Levels of Control: DNA methylation



abnormal hypermethylation of CpG islands can cause cancer, e.g. transcriptional silencing of tumor suppressor genes: target for gene therapy?

DNA methylation occurs mainly at the C5 position of CpG dinucleotides:

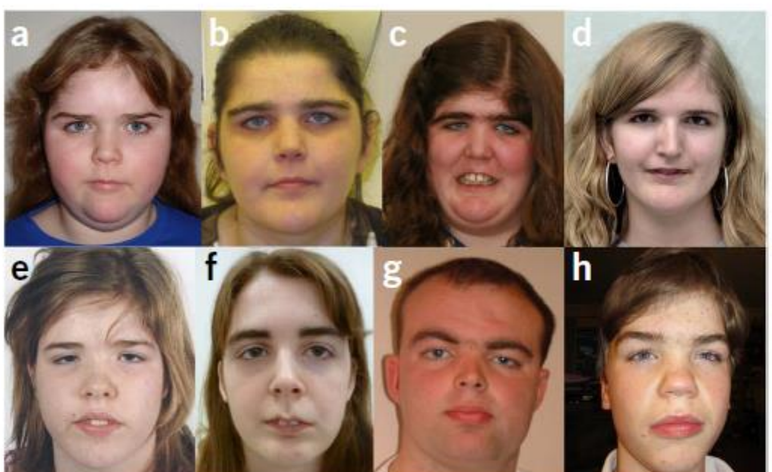
- ✓ *de novo* methylation: installing methylation patterns early in development DNMT3a and DNMT3b:
  - DNA methylation can stably alter the expression of genes in cells during cell division and differentiate from embryonic stem cells into specific tissues.
  - DNA methylation is typically removed during zygote formation and re-established through successive cell divisions during development.
- ✓ maintenance methylation activity is necessary to preserve DNA methylation after every cellular DNA replication cycle: DNMT1.

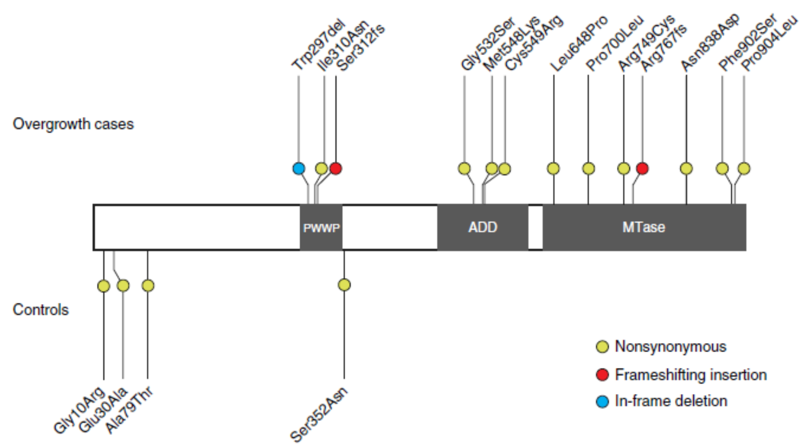
DNA methylation related disease

nature genetics

Mutations in the DNA methyltransferase gene *DNMT3A* cause an overgrowth syndrome with intellectual disability

Katrina Tatton-Brown[1–3], Sheila Seal[1], Elise Ruark[1], Jenny Harmer[4], Emma Ramsay[1], Silvana del Vecchio Duarte[1], Anna Zachariou[1], Sandra Hanks[1], Eleanor O'Brien[1], Lise Aksglaede[5], Diana Baralle[6], Tabib Dabir[7], Blanca Gener[8], David Goudie[9], Tessa Homfray[3], Ajith Kumar[10], Daniela T Pilz[11], Angelo Selicorni[12], I Karen Temple[6], Lionel Van Maldergem[13], Naomi Yachelevich[14], Childhood Overgrowth Consortium[15], Robert van Montfort[4] & Nazneen Rahman[1,2]





DNA methylation related disease

nature COMMUNICATIONS

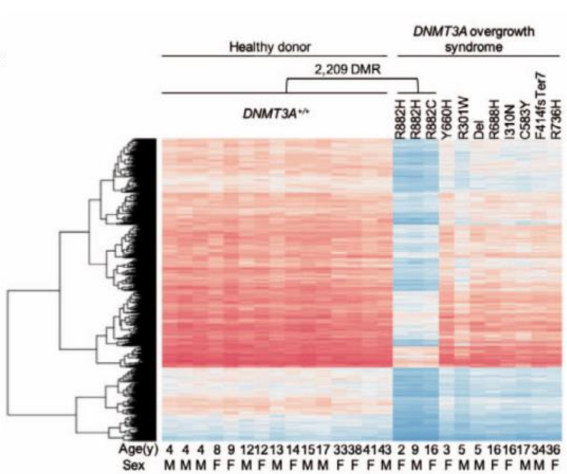
ARTICLE

Check for updates

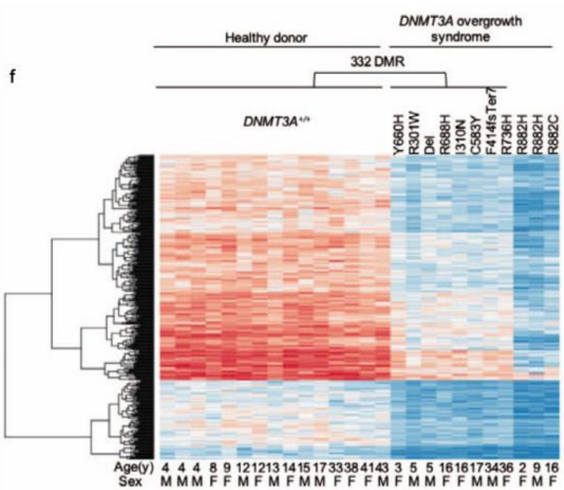
https://doi.org/10.1038/s41467-021-24800-7 OPEN

Functional and epigenetic phenotypes of humans and mice with DNMT3A Overgrowth Syndrome

Amanda M. Smith[1], Taylor A. LaValle[1], Marwan Shinawi[2], Sai M. Ramakrishnan[1], Haley J. Abel[1], Cheryl A. Hill[3], Nicole M. Kirkland[3], Michael P. Rettig[1], Nichole M. Helton[1], Sharon E. Heath[1], Francesca Ferraro[1], David Y. Chen[4], Sangeeta Adak[5], Clay F. Semenkovich[5], Diana L. Christian[6], Jenna R. Martin[6], Harrison W. Gabel[6], Christopher A. Miller[1] & Timothy J. Ley[1]✉



f



DNA methylation related disease

KATHOLIEKE UNIVERSITEIT LEUVEN

FRAGILE X SYNDROME

CGG trinucleotide expansion > 200 ('full mutation') in *FMR1* gene (on X chromosome) causes hypermethylation of the *FMR1* promotor: inactivation of FMR1 expression

Phenotype depends on # CGG repeats & methylation status

- ✓ Moderate to severe intellectual disability in affected males
- ✓ Males with full mutation:
  - large head
  - long face
  - prominent forehead and chin
  - protruding ears
  - large testes after puberty
  - behavioral abnormalities & autism.

- ✓ Females can have ID, FXTAS, POF

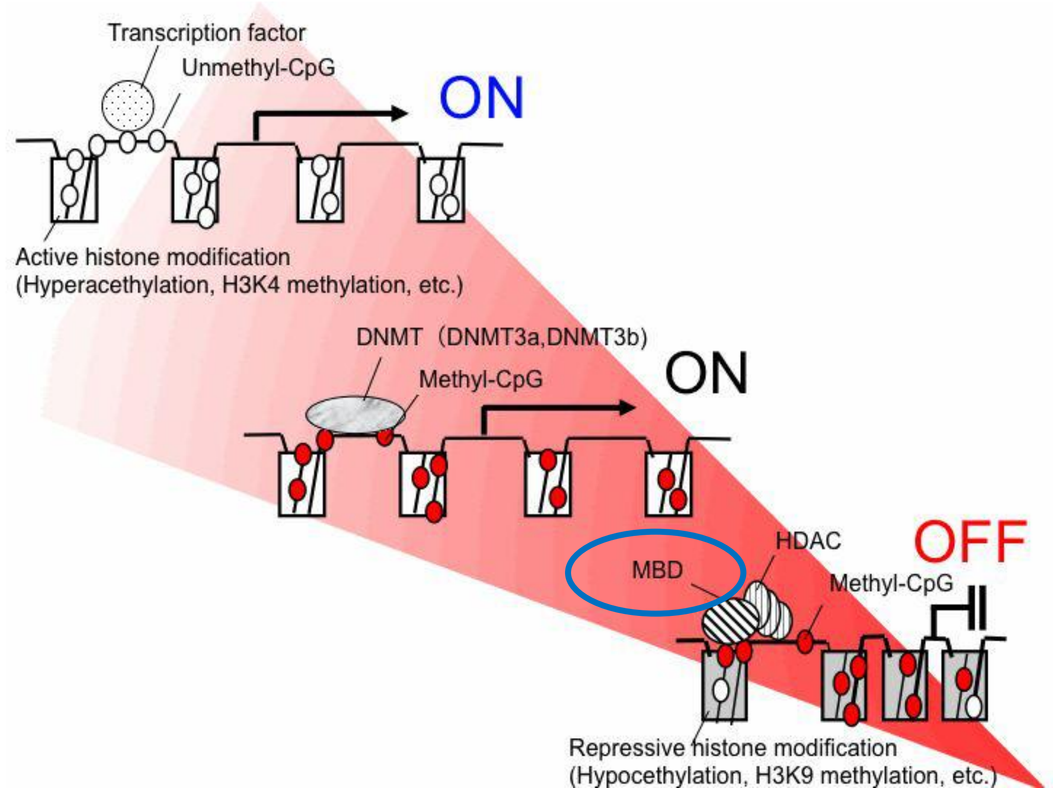
DNA methylation related disease

FRAGILE X SYNDROME

Levels of Control: DNA methylation

Effect of DNA methylation on gene transcription:

- ✓ the methylation of DNA itself physically impede the binding of transcriptional factors to the gene
- ✓ methylated DNA may be bound by methyl-CpG-binding domain proteins: MBDs.
  - MBD proteins recruit histone deacetylases and other chromatin remodeling proteins > histone modification: forming compact, inactive heterochromatin.



MBD proteins related disease

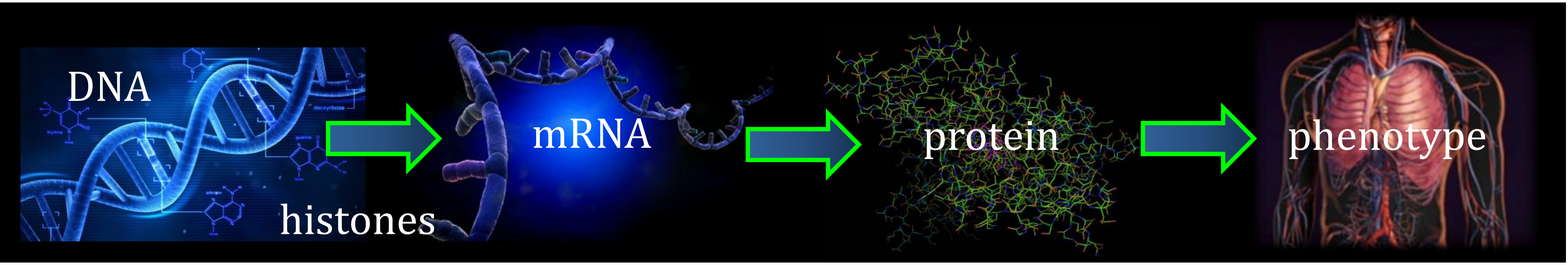
RETT SYNDROME

- ✓ Developmental regression: onset 6 to 18 months
- ✓ Severe ID & autism
- ✓ Epilepsy
- ✓ Ataxia
- ✓ Behavioral problems
- ✓ Stereotyped hand movements
- ✓ Acquired microcephaly

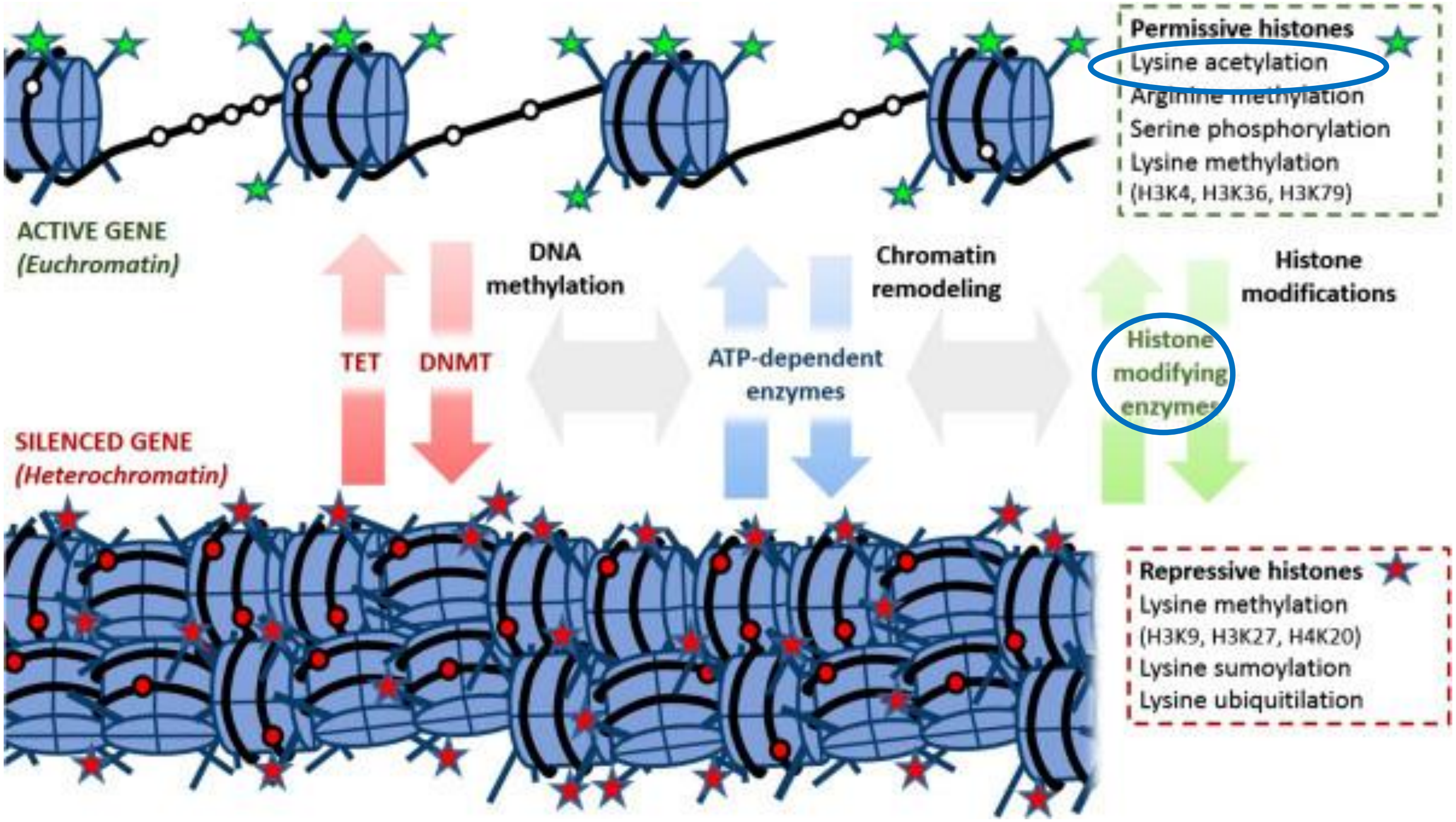
Loss of methyl-CpG-binding protein 2 (MeCP2) (on X chromosome) has been implicated in girls with Rett syndrome. MECP2 is an MBD protein, which can act as a transcriptional repressor.

MECP2 duplications cause severe ID in boys.

Chromatin remodeling



Levels of Control: chromatin remodeling



Lysine acetylation by Histone acetyl transferase (HAT): reduces electrostatic attraction between the histone and the negatively charged DNA backbone, loosening the chromatin structure = EUCHROMATIN (<> HDAC)

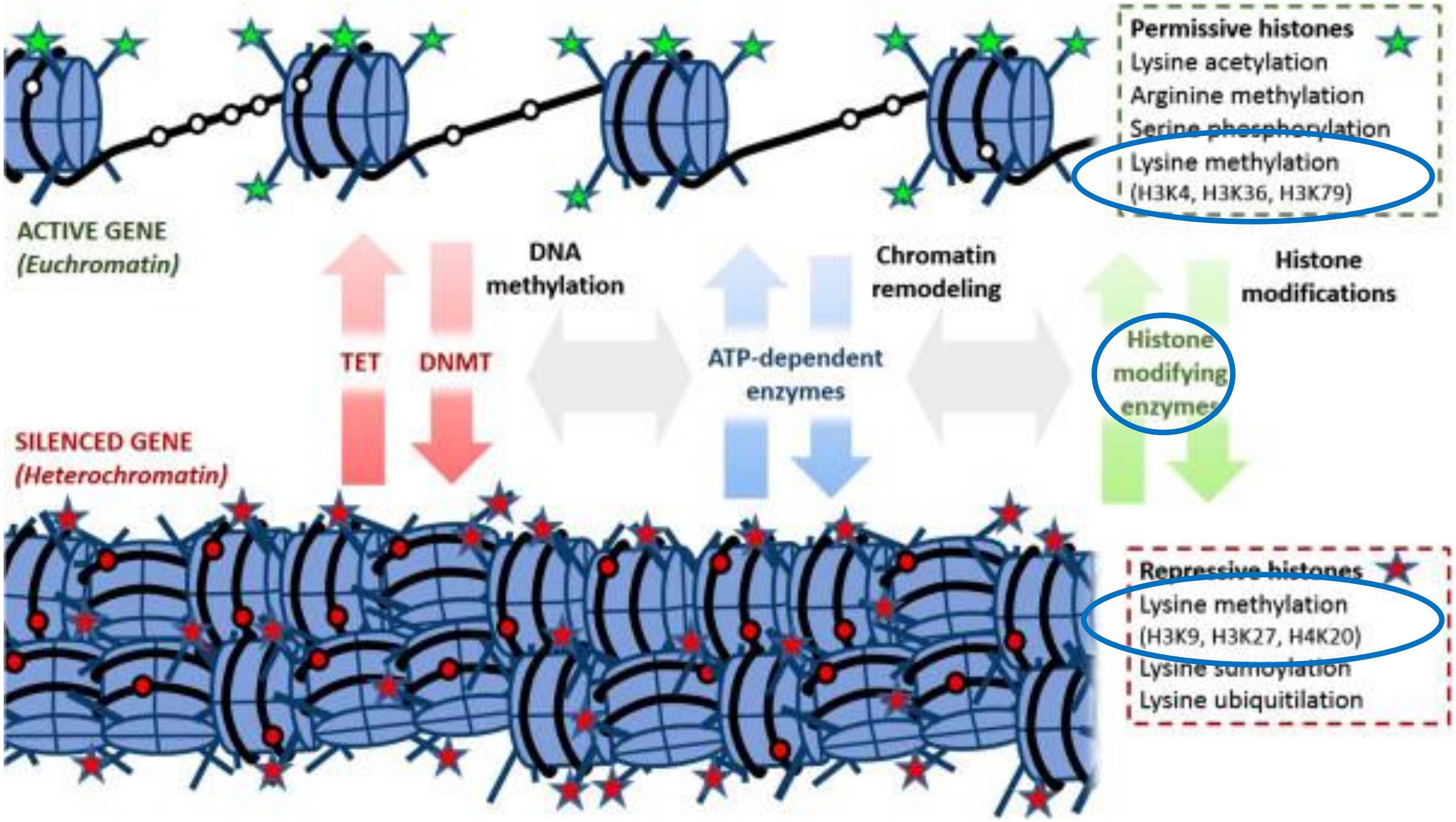
Histone acetylation related disease

REPORT

De Novo Nonsense Mutations in *KAT6A*, a Lysine Acetyl-Transferase Gene, Cause a Syndrome Including Microcephaly and Global Developmental Delay

Valerie A. Arboleda,[1] Hane Lee,[1] Naghmeh Dorrani,[2] Neda Zadeh,[3,4] Mary Willis,[5] Colleen Forsyth Macmurdo,[6] Melanie A. Manning,[6,7] Andrea Kwan,[6,8] Louanne Hudgins,[6] Florian Barthelemy,[9] M. Carrie Miceli,[9] Fabiola Quintero-Rivera,[1] Sibel Kantarci,[1] Samuel P. Strom,[1] Joshua L. Deignan,[1] UCLA Clinical Genomics Center,[1] Wayne W. Grody,[1,2,10] Eric Vilain,[2,10] and Stanley F. Nelson[1,10,*]

Levels of Control: chromatin remodeling

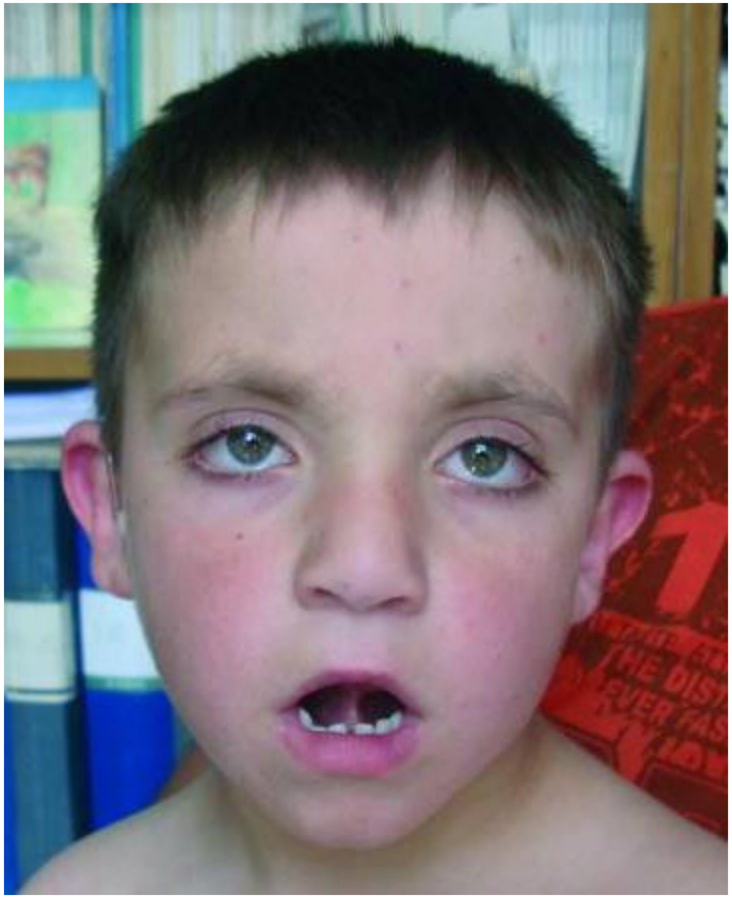


Lysine methylation by histone methyl transferase:

- ✓ induces euchromatin: H3K4, H3K36, H3K79
- ✓ induces heterochromatin: H3K9, H3K27, H4K20

Histone methylation related disease

KATHOLIEKE UNIVERSITEIT LEUVEN



Kabuki syndrome
KMT2D loss-of-function mutations in 50-70% of KS patients

- facial gestalt
- short stature
- microcephaly
- feeding problems
- oligodontia
- high/cleft palate
- fetal pads
- lax joints
- cardiac defects
- renal defects
- ID
- hypotonia
- frequent infections

KMT2D is a histone methyltransferase that targets lysine 4 of histone H3 (H3K4) to promote an open chromatin state.

Closing remark