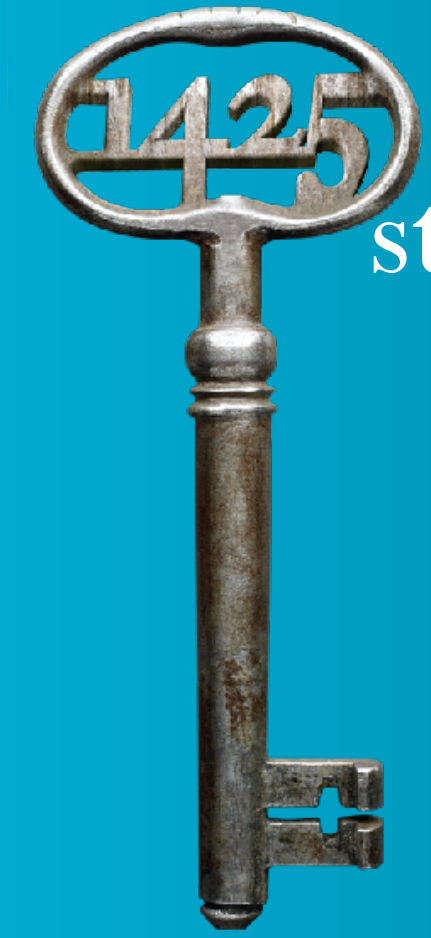
KATHOLIEKE UNIVERSITEIT
LEUVEN

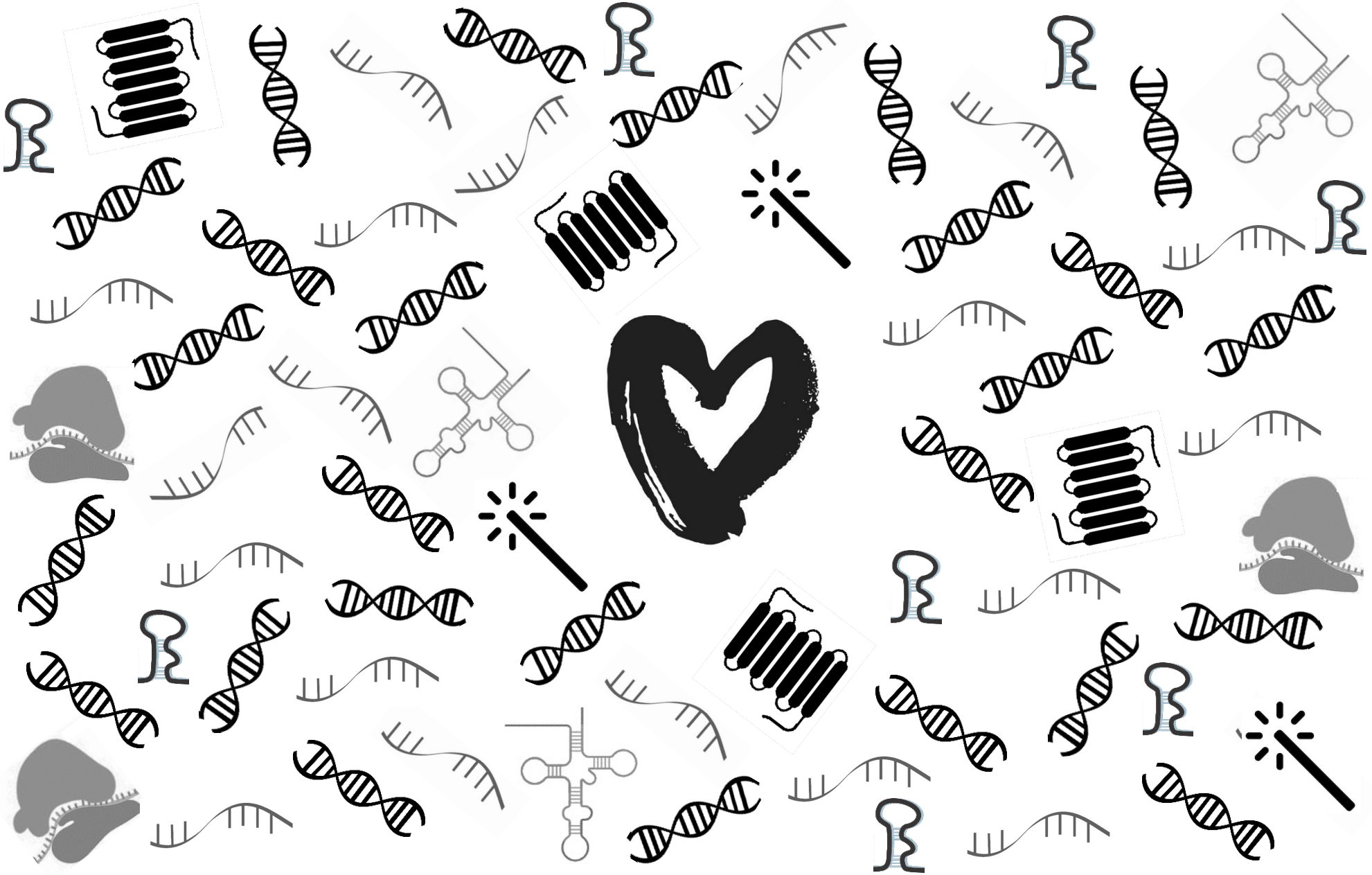


Human genome: gene structure, function & regulation

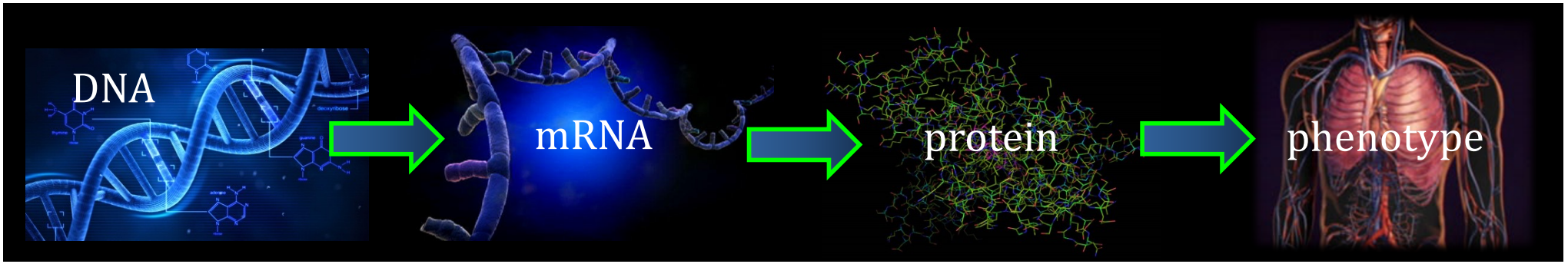
Jeroen Breckpot

Interuniversity course in
human genetics
October 2021

Introduction



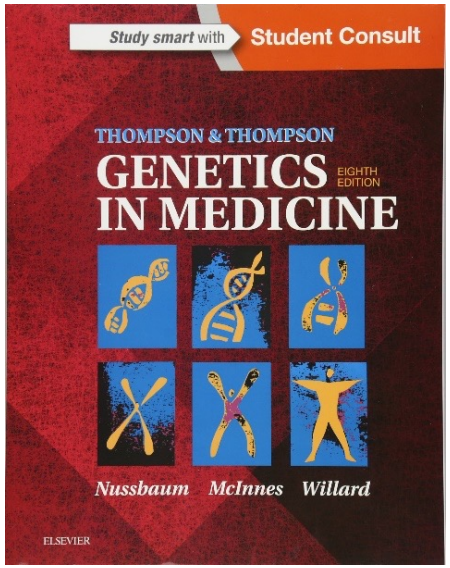
The Central Dogma



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

Outline of the presentation

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

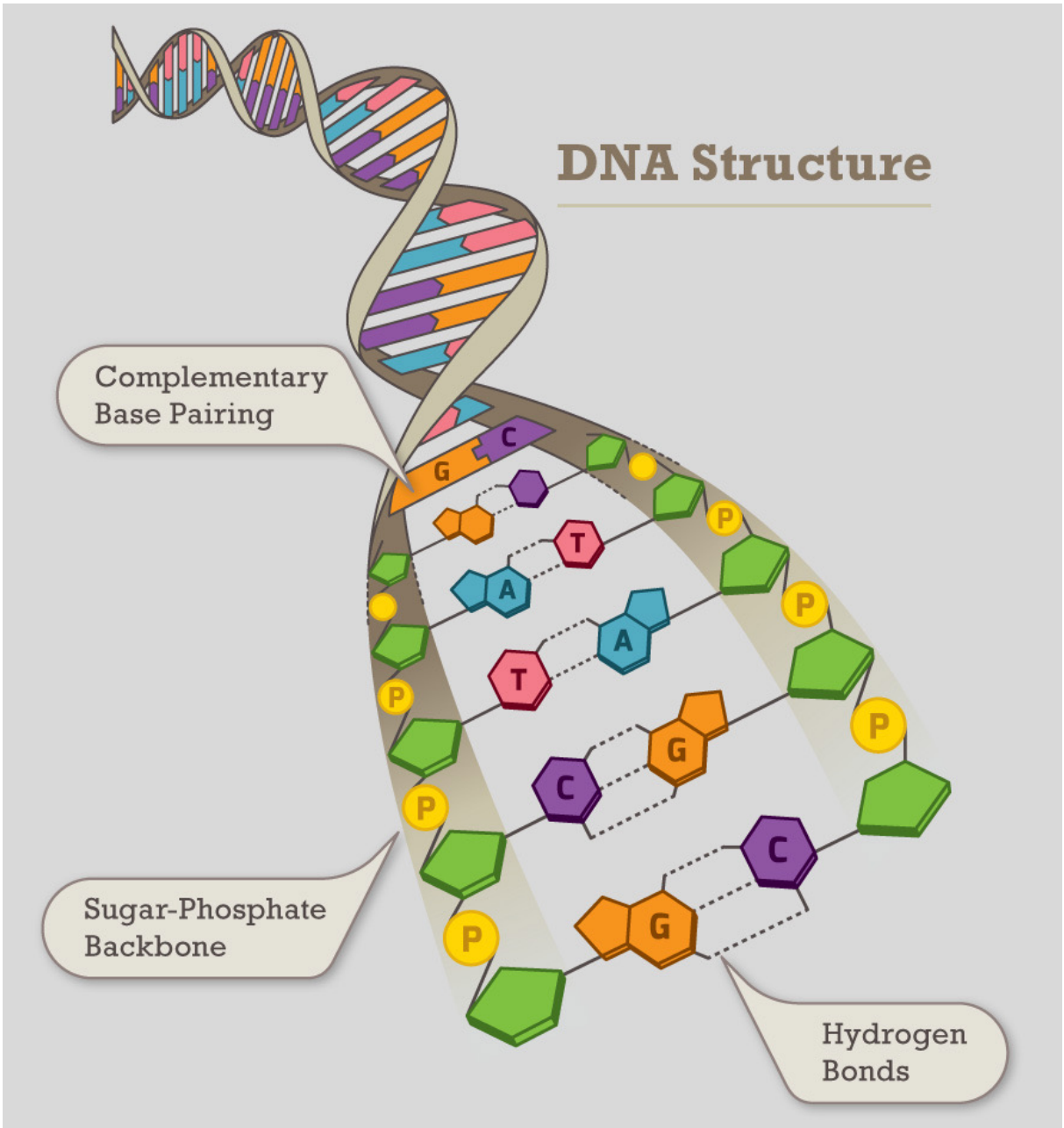


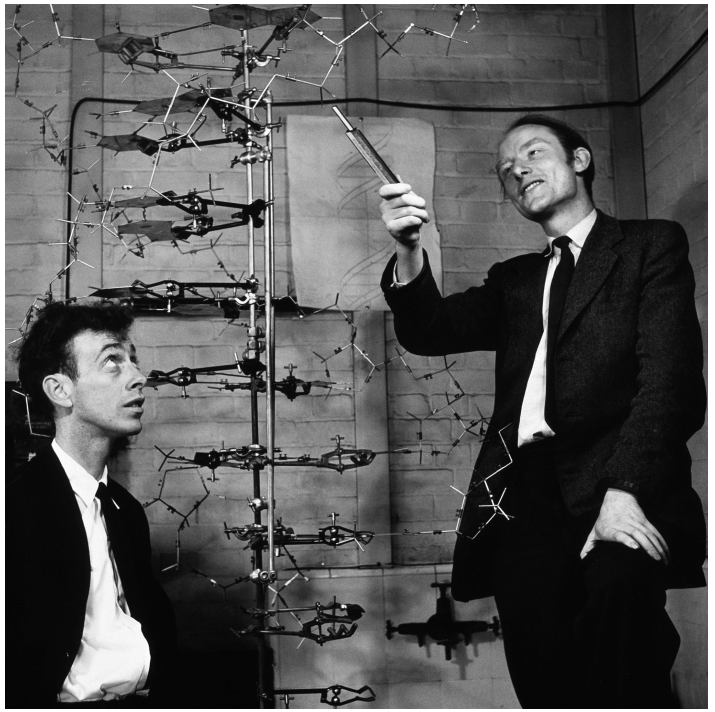
Chapter 2 & 3. Thompson & Thompson Genetics in Medicine

Definitions

From base pair to chromosome

DNA

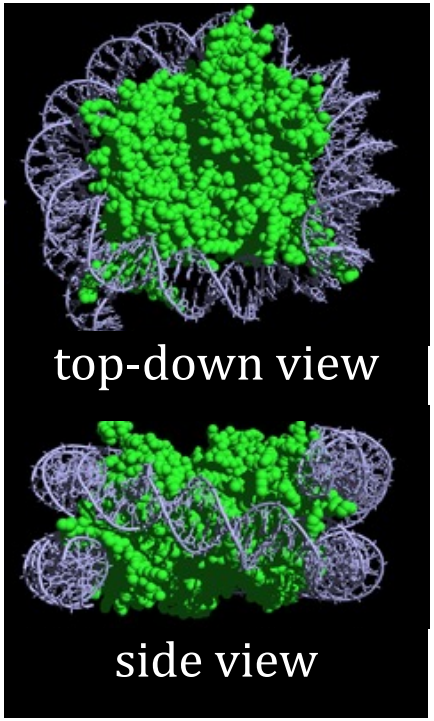


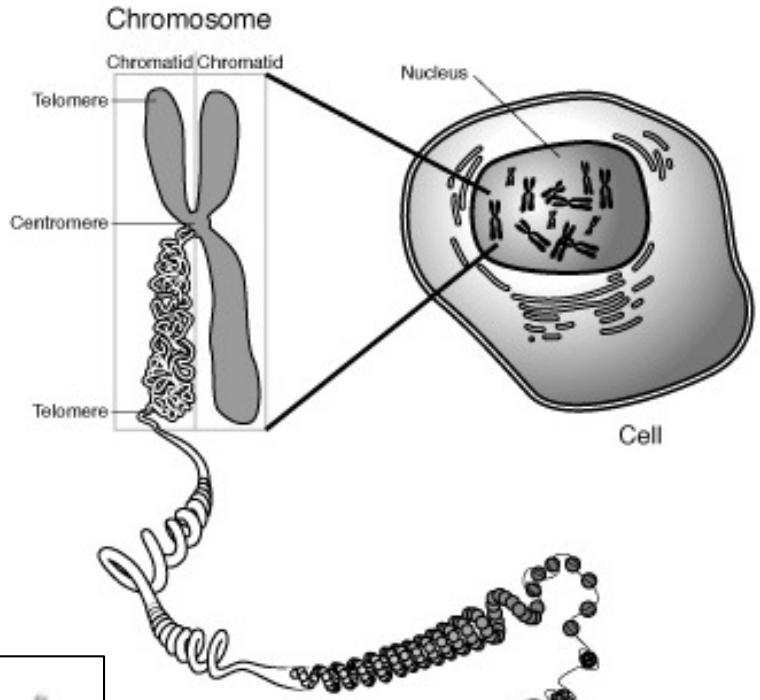


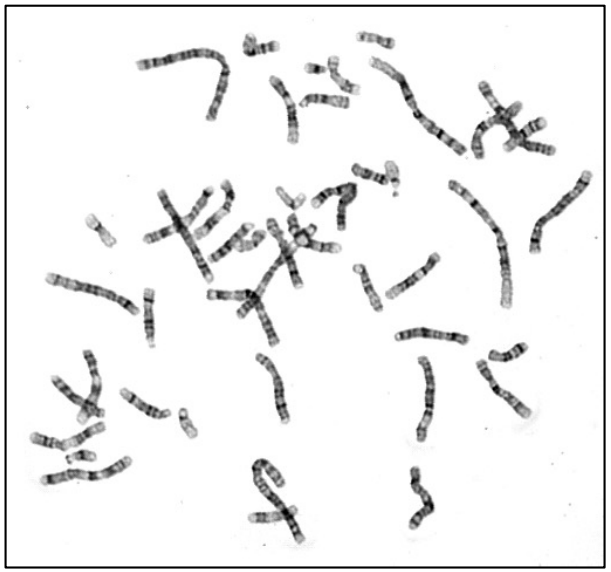
DNA structure discovered by Watson, Crick, Wilkins and Franklin in 1953

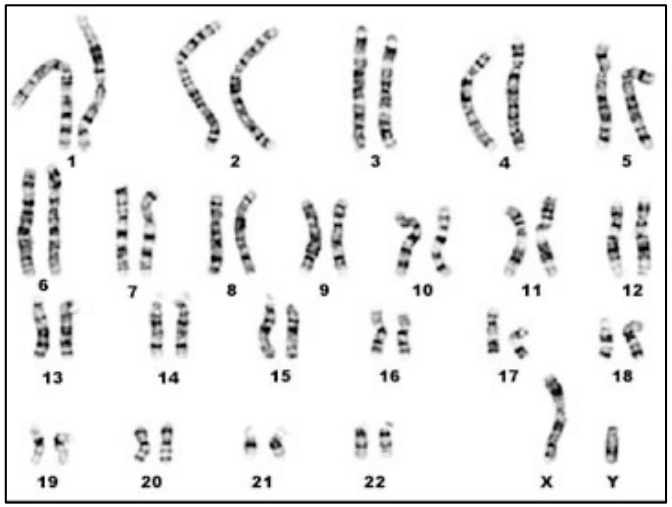
Chapter 2. Thompson & Thompson Genetics in Medicine

From base pair to chromosome





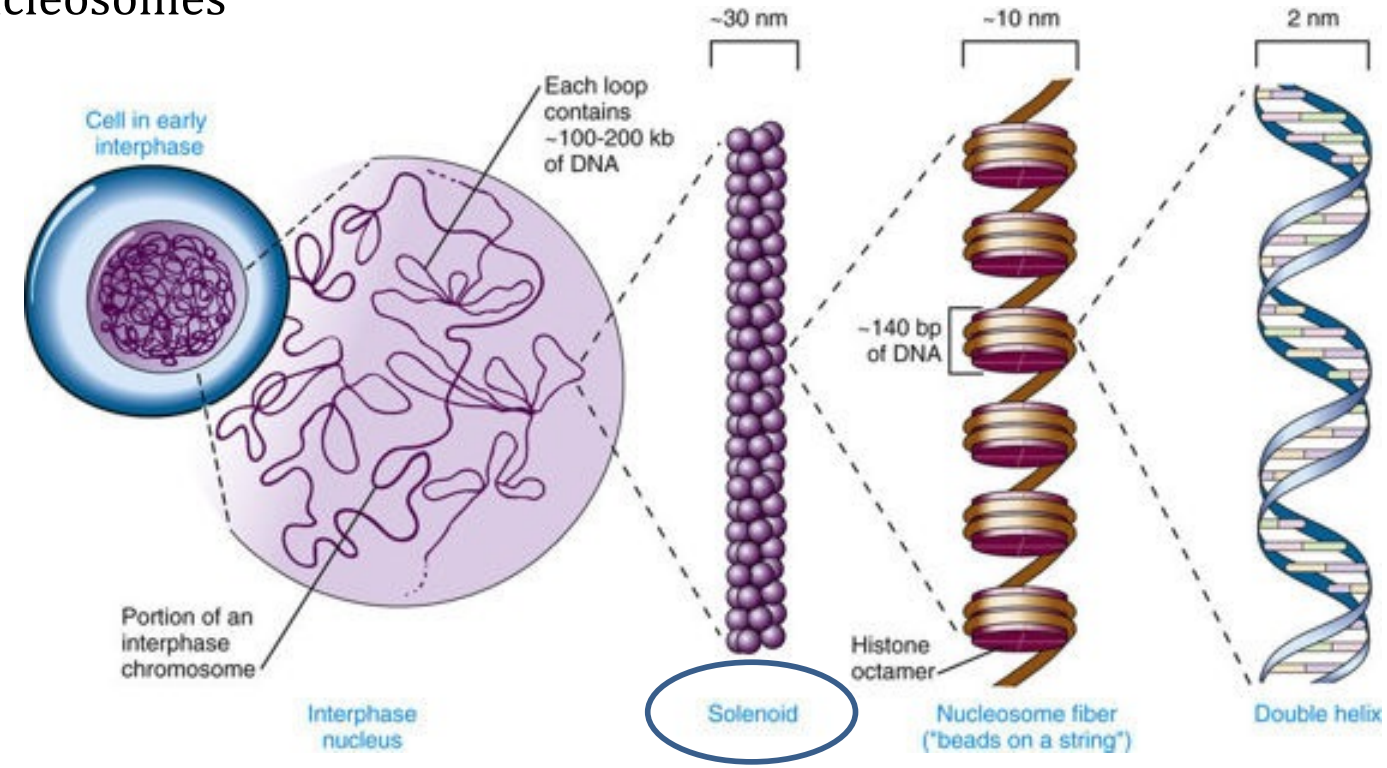




Histones

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

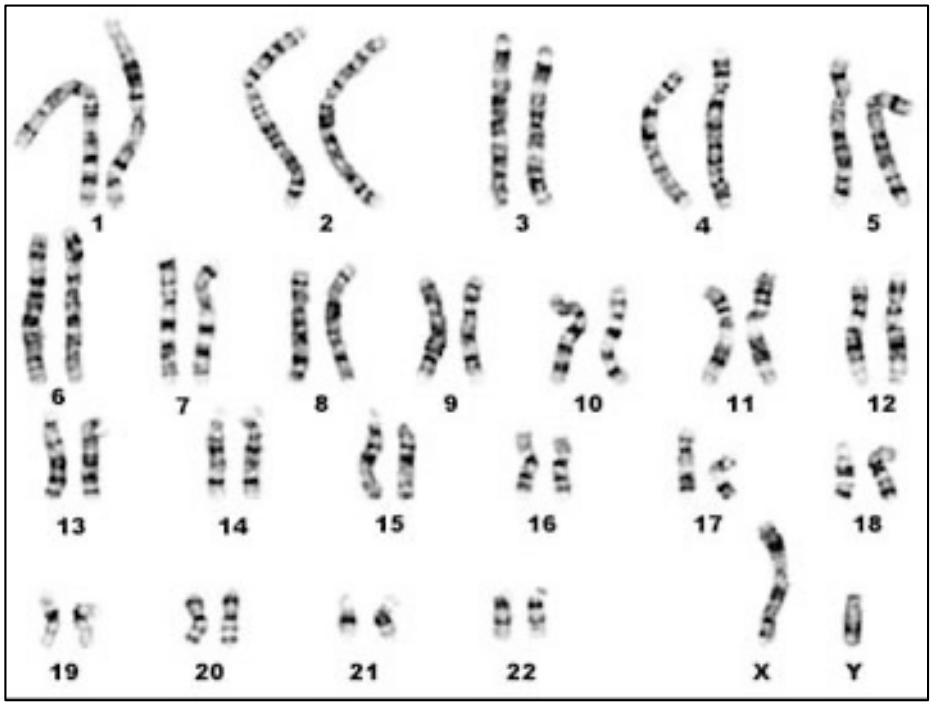


- ✓ H3 and H2A can be substituted by other histone types, or histones can be modified by chemical changes: cfr. regulatory mechanisms

Chromosomes

Human somatic cells: 46 chromosomes:

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



✓ 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)

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

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

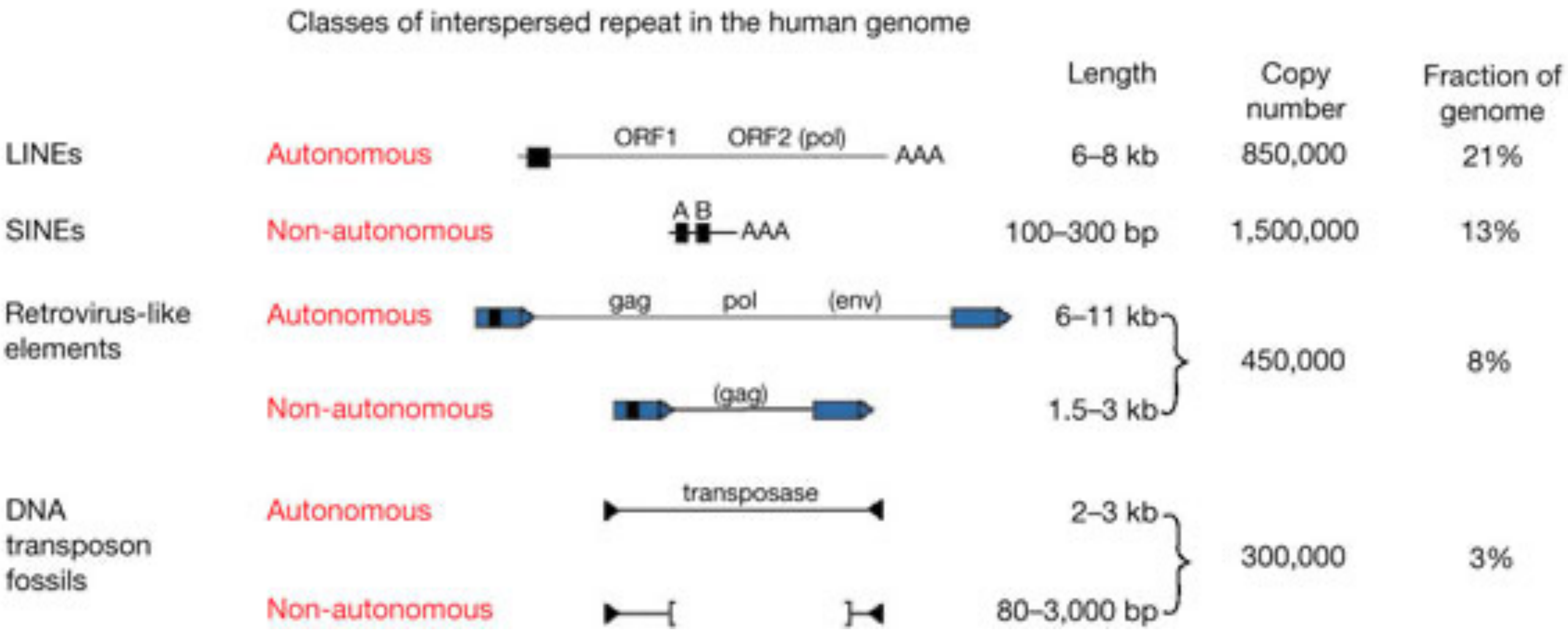
- SINE: e.g. Alu repeats – GT (10%)
- LINE: 6 kb in length – AT (20%)

✓ **segmental duplications:**

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

LINE and SINE

- LINE and SINE: "copy and paste"
  - LINE begins with an UTR, followed by 2 ORF, one encoding an RNA binding protein, the other encoding an endonuclease and a reverse transcriptase: machinery for retrotransposition.
  - during the past 50 million years: decrease seems to have occurred in the rate of accumulation of repeats
  - SINE do not encode a functional reverse transcriptase protein: they integrate at chromosomal breaks e.g. Alu repeats



Alu repeats

FISH with Alu sonde marks gene-rich regions in human metaphase chromosomes

Molecular Genetics & Genomic Medicine

Open Access

ORIGINAL ARTICLE

Identification of an *Alu* element-mediated deletion in the promoter region of *GNE* in siblings with GNE myopathy

Jennifer Garland[1,2,a], Joshi Stephen[1,a], Bradley Class[2,a], Angela Gruber[3], Carla Ciccone[1], Aaron Poliak[1], Christina P. Hayes[4], Vandana Singhal[2], Christina Slota[2], John Perreault[2,5], Ralitza Gavrilova[6], Joseph A. Shrader[7], Prashant Chittiboina[4], Galen Joe[7], John Heiss[4], William A. Gahl[1,8,9], Marjan Huizing[1], Nuria Carrillo[1,2,a] & May Christine V. Malicdan[1,8,9,a]

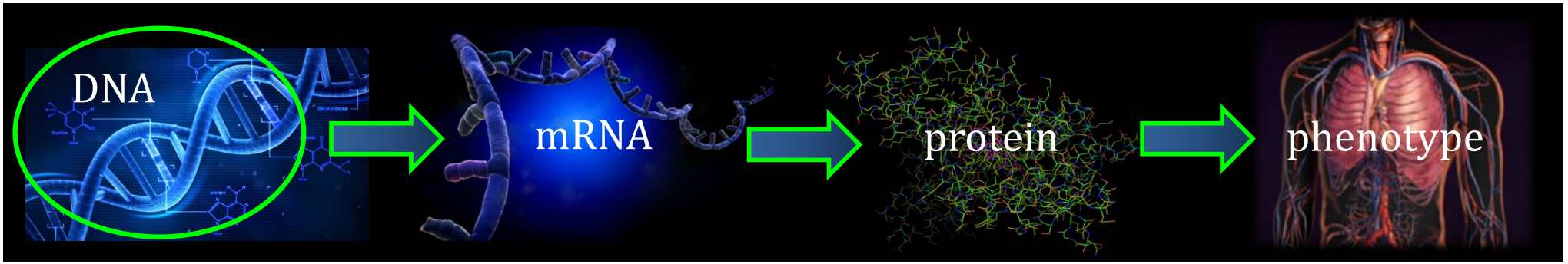
Alu sequences enriched in GT rich regions, which often correspond to gene-dense regions

Alu sequence structure: 5'Part A- A5TACA6 -Part B - PolyA Tail - 3‘
Part A and Part B are similar nucleotide sequences.

DNA counterstain = red

Alu sequences = green

Genes

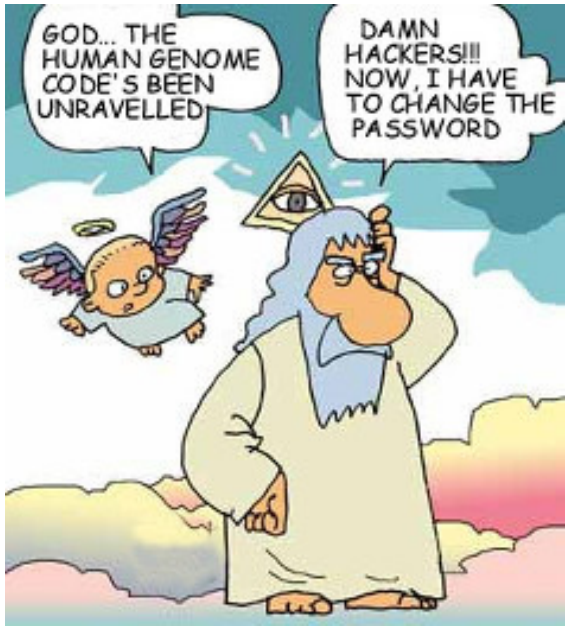


nucleus	*nucleus - cytoplasm*	*cytoplasm, organelles, extracellular*	*tissues, organs, organism*

GENOTYPE ⟶ *PHENOTYPE*

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

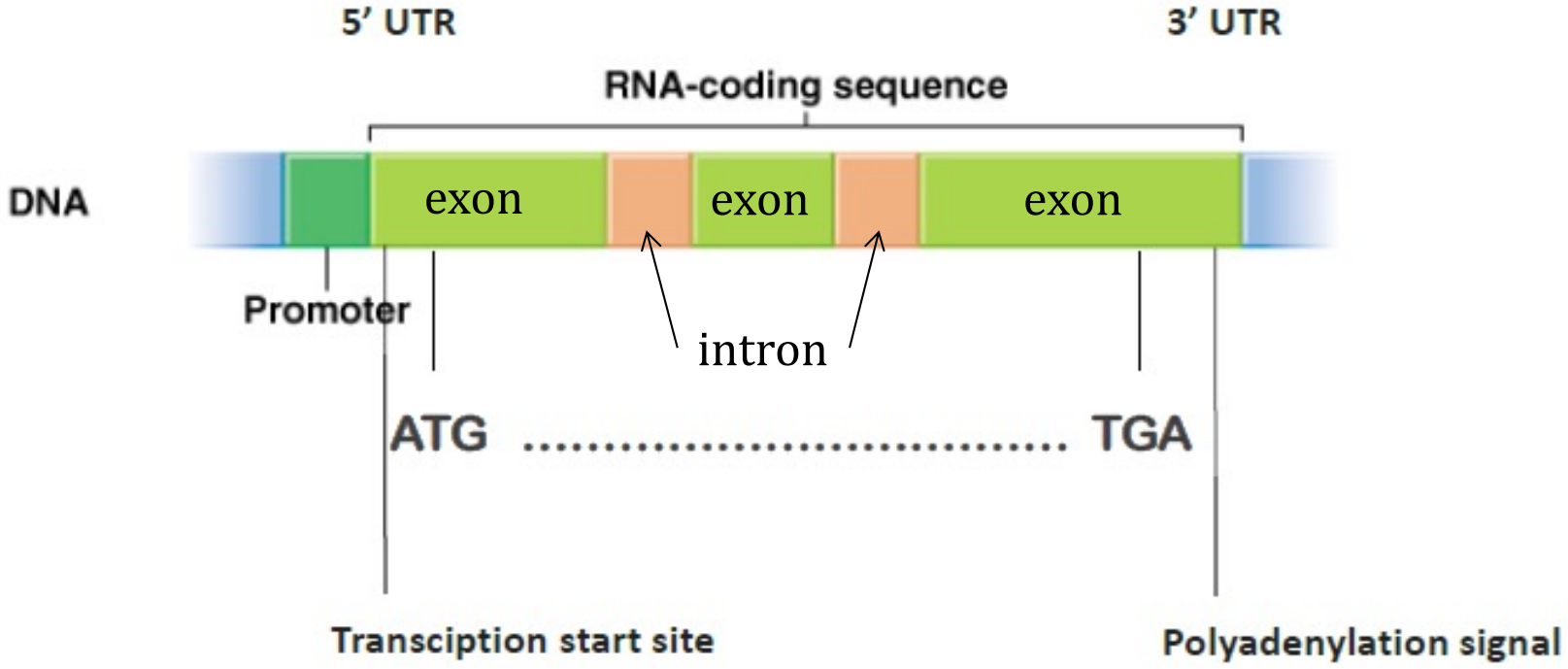
- ✓ coding genes: 20,769
- ✓ short non coding genes: 9,079
- ✓ long non coding genes: 13,564
- ✓ pseudogenes: 14,165



features of a typical human gene

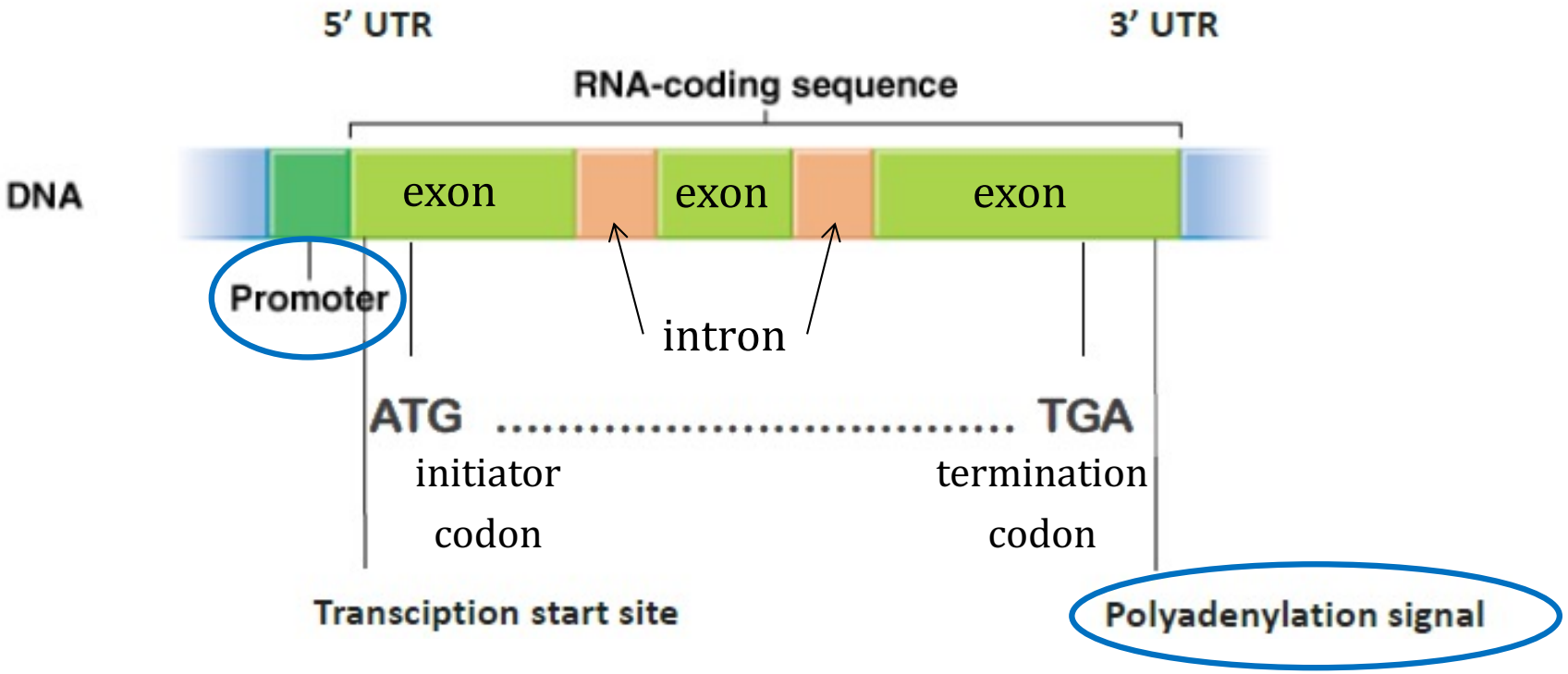
DEFINITION:

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)



features of a typical human gene

a gene includes:
- ✓ actual coding sequences
- ✓ adjacent nucleotide sequences for proper expression of the gene



in eukaryotes, the genes have their coding sequences (exons) interrupted by non-coding sequences (introns)

features of a pseudogene

DEFINITION

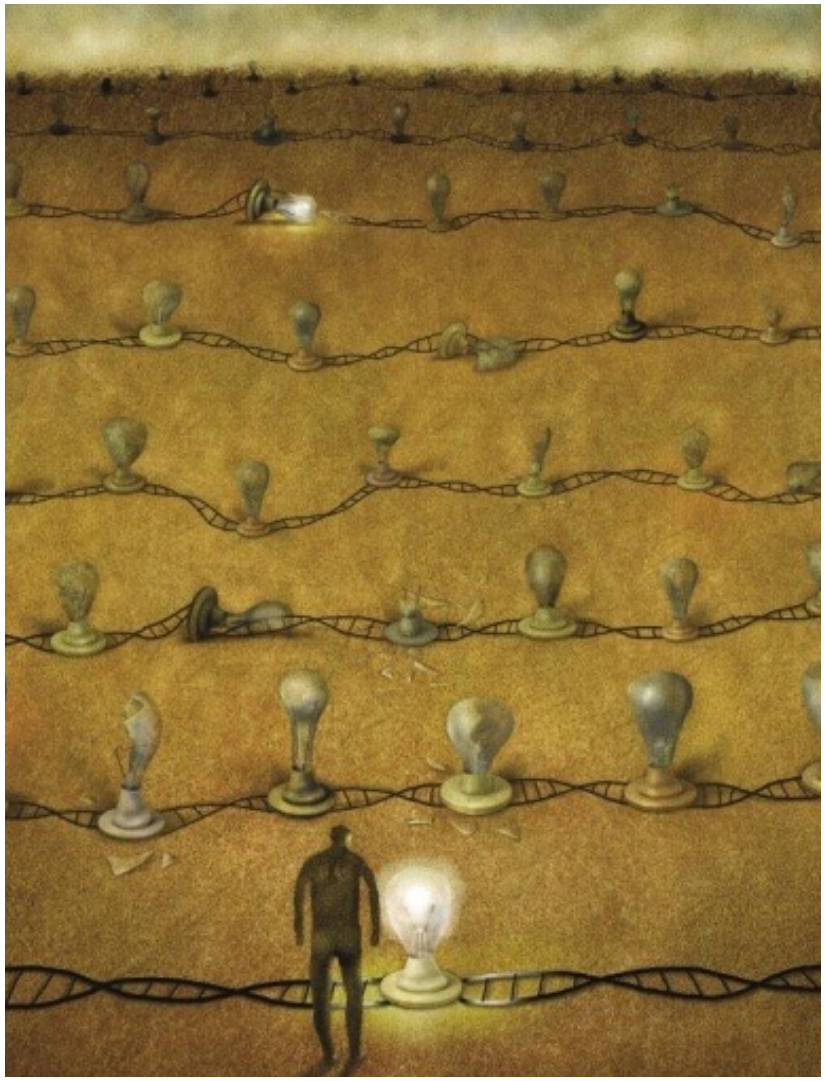
DNA sequences that closely resemble known genes but are afunctional

- ✓ 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

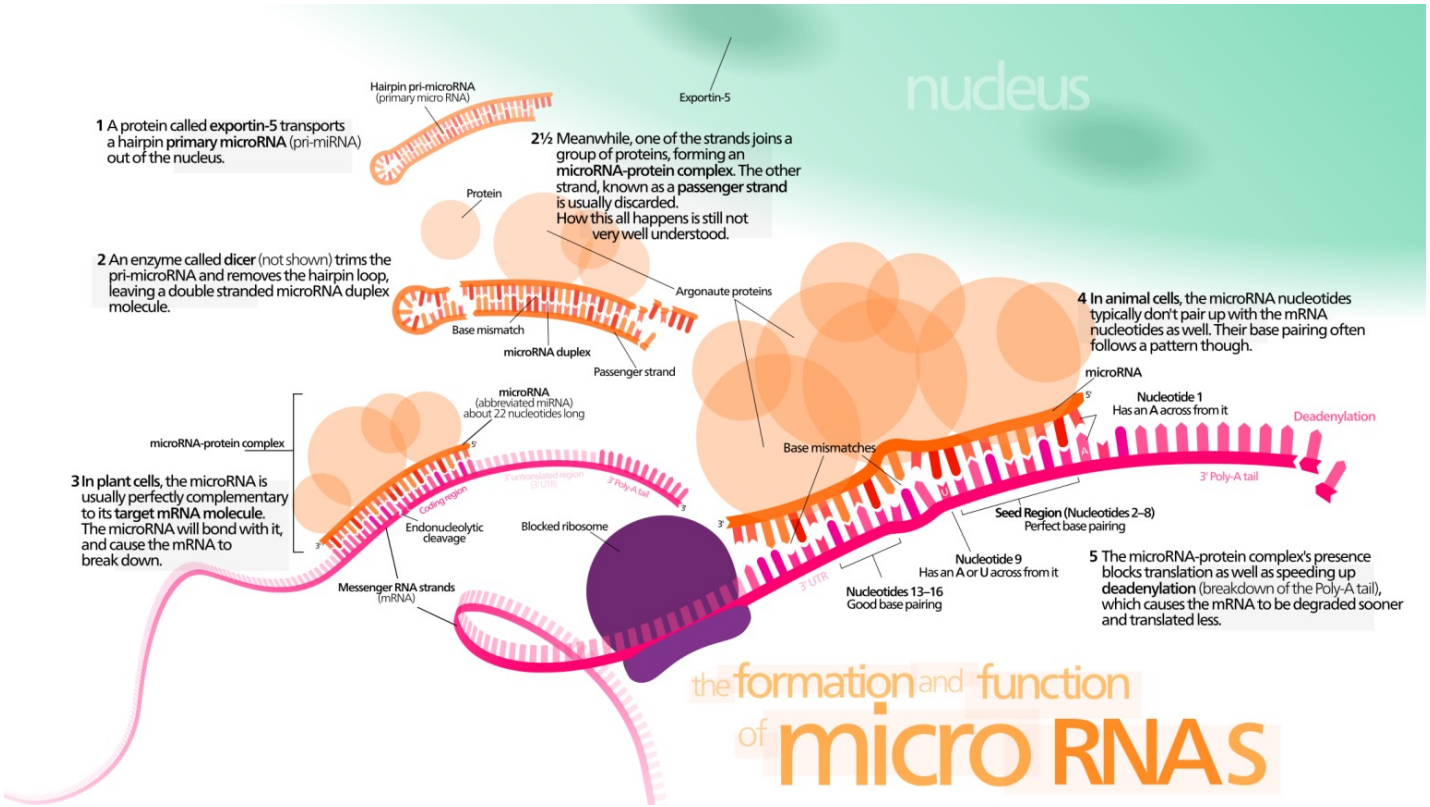


features of non-coding genes

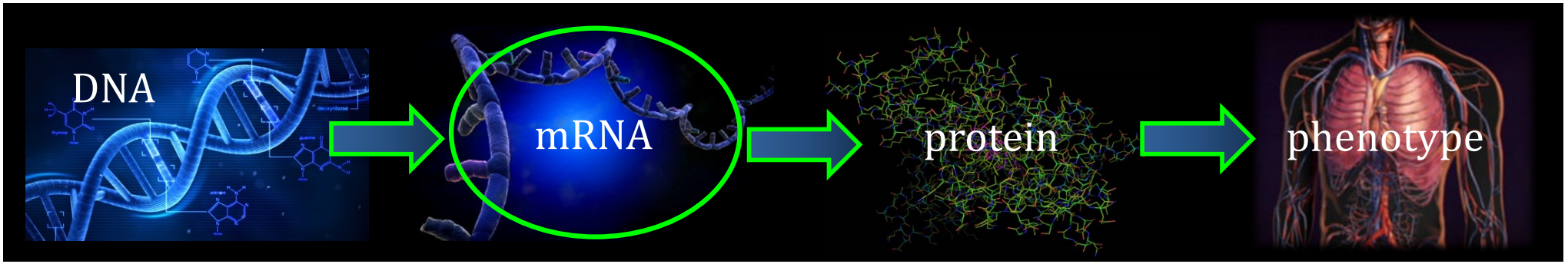
DEFINITION

DNA sequences that encode an untranslated functional RNA product

miRNA

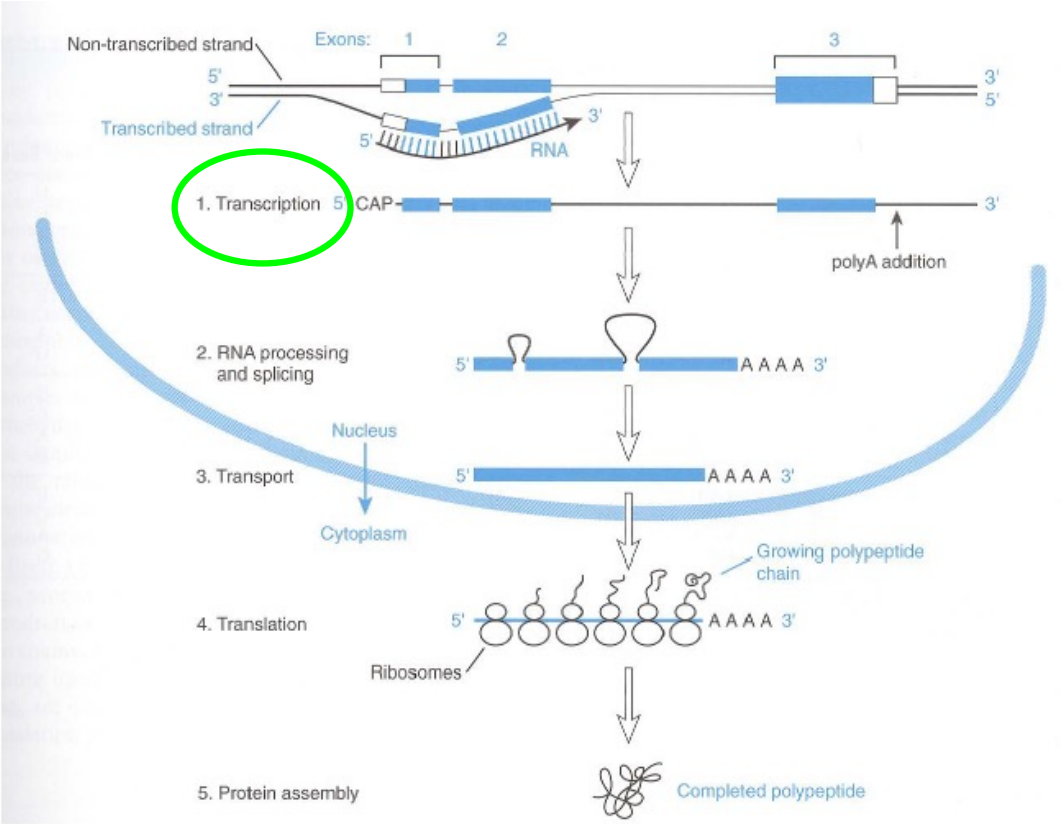


The central dogma



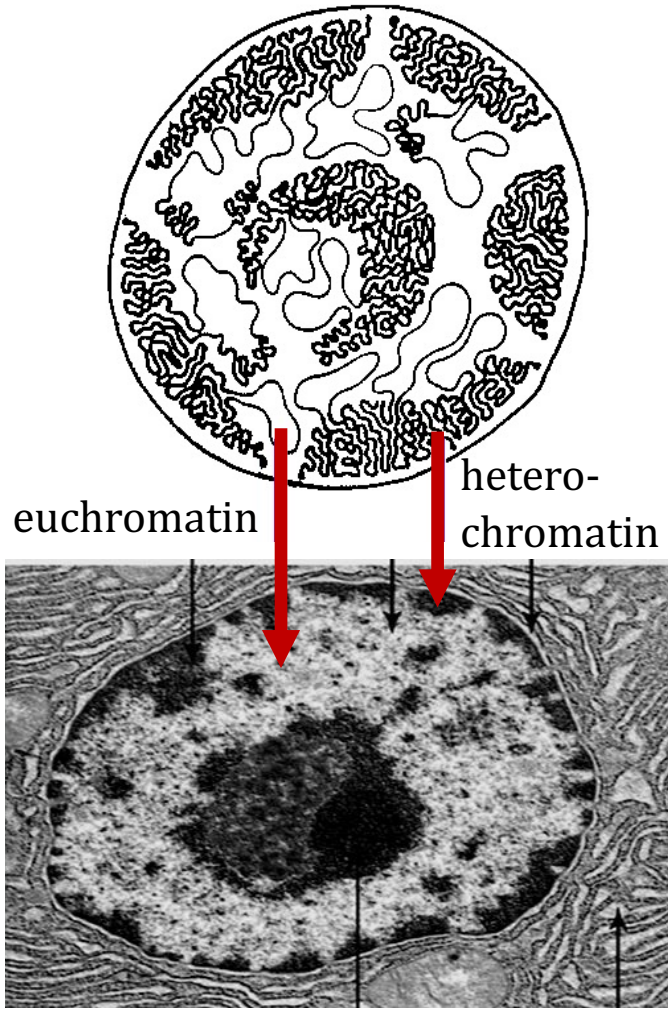
Transcription

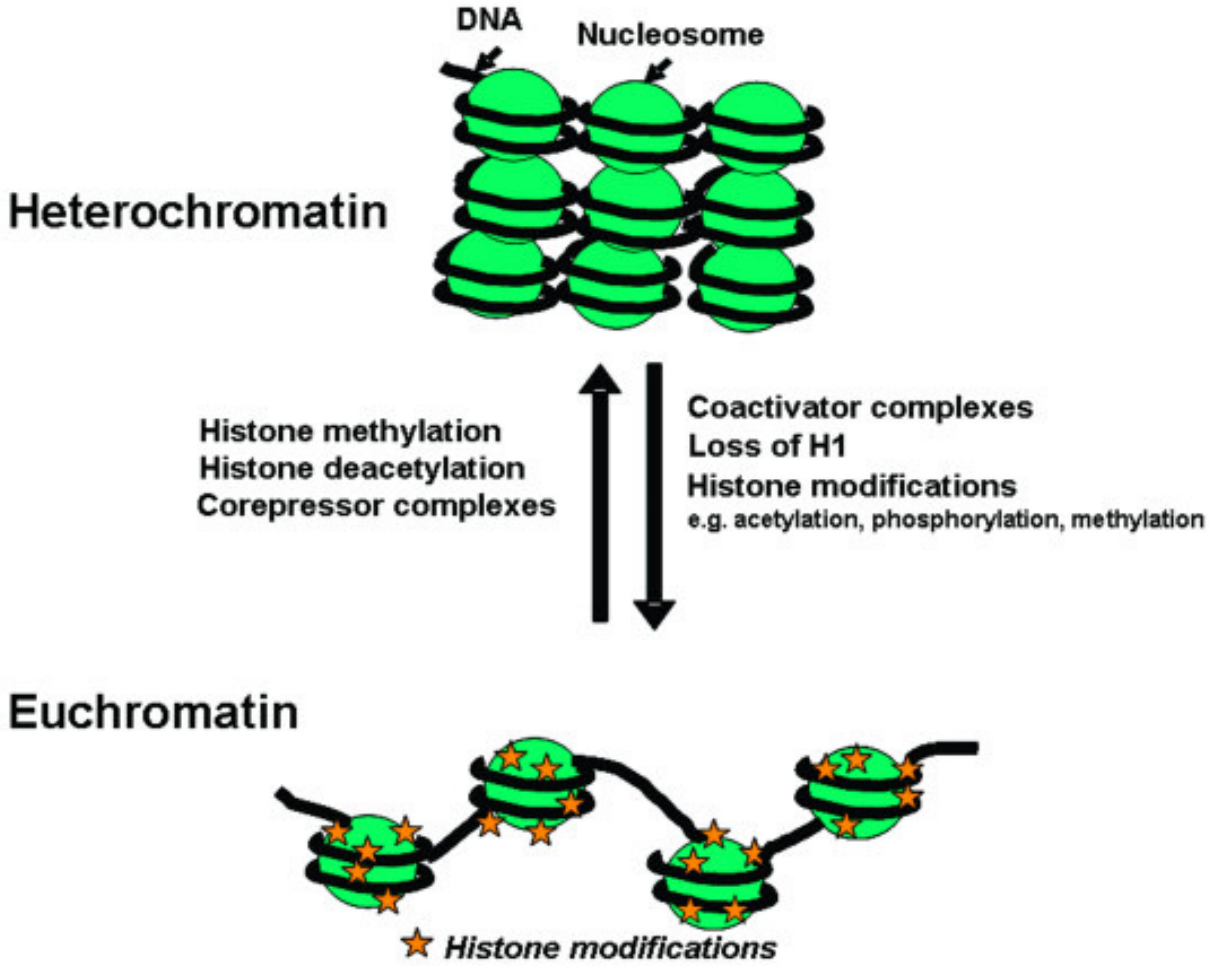
1. Initiation of transcription
   - ✓ basal initiation complex
   - ✓ transcription factors
2. mRNA synthesis
3. mRNA processing



Initiation of transcription

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





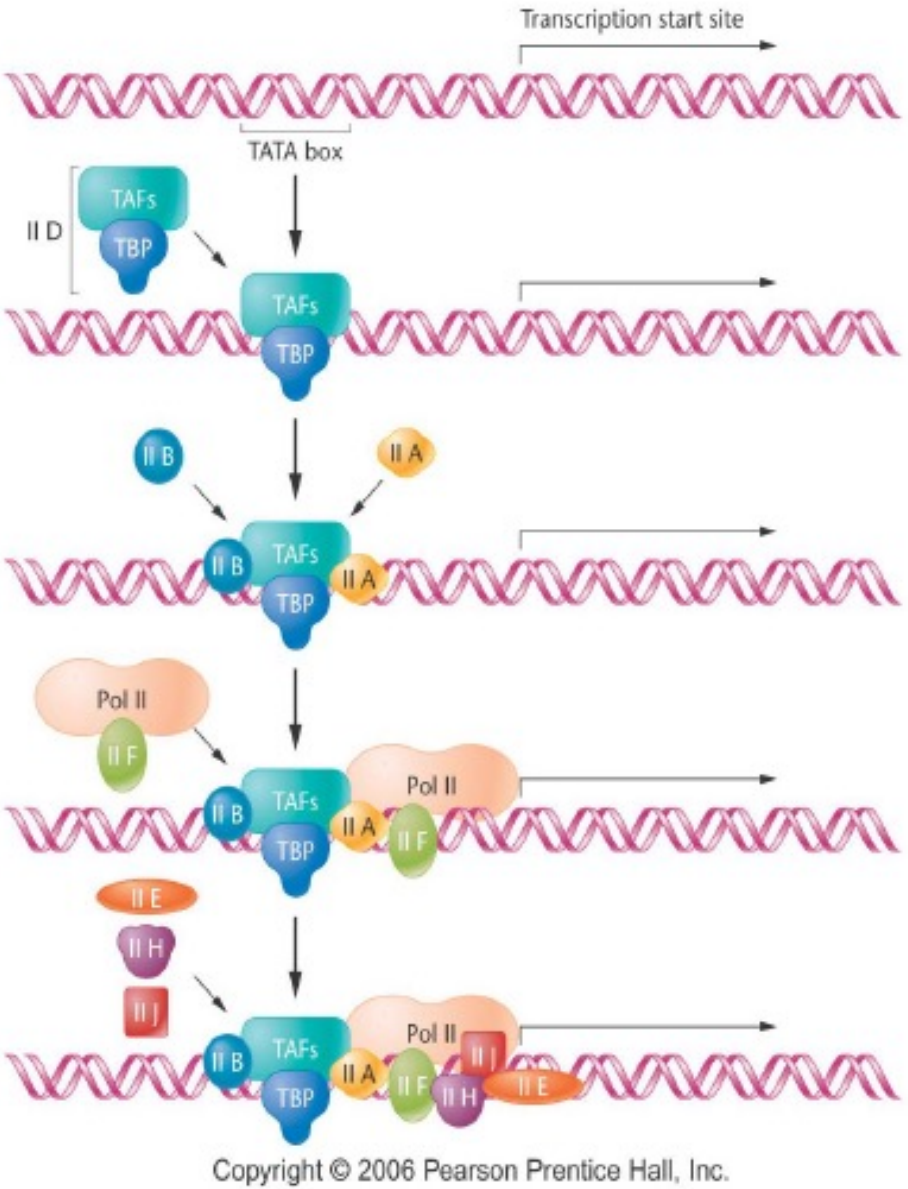
Initiation of transcription

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)



Initiation of transcription

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.

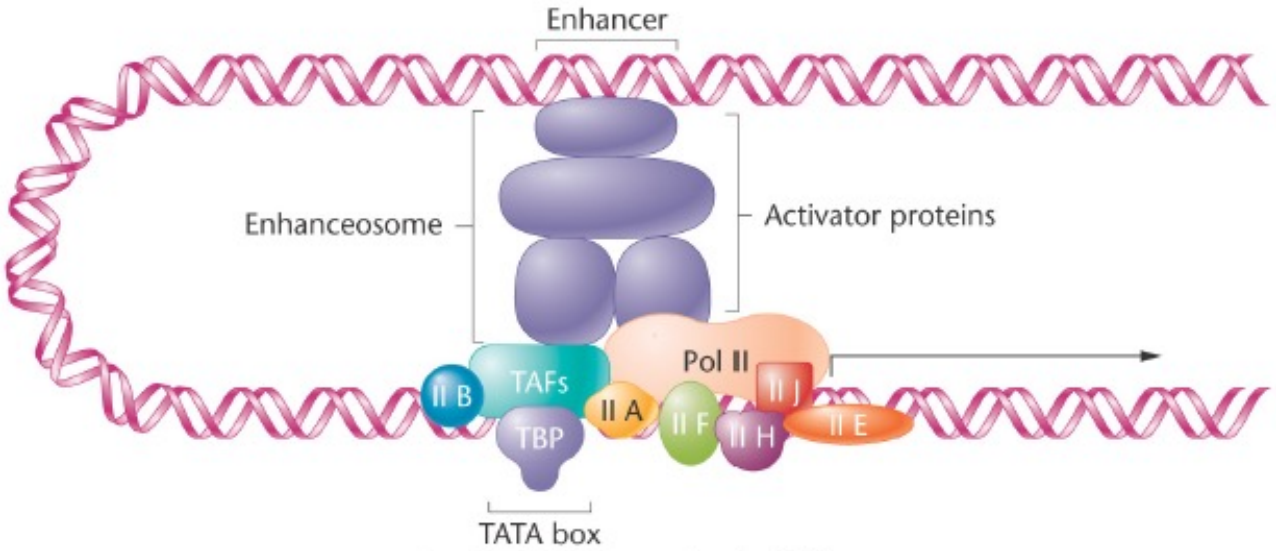
Results. The coding region of the gene for the enzyme was normal in the 10 patients with Gilbert's syndrome. These patients were homozygous for two extra bases (TA) in the TATAA element of the 5′ promoter region of the gene ($A(TA)_7TAA$ rather than the normal $A(TA)_6TAA$). The presence of the longer 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 the enhanceome

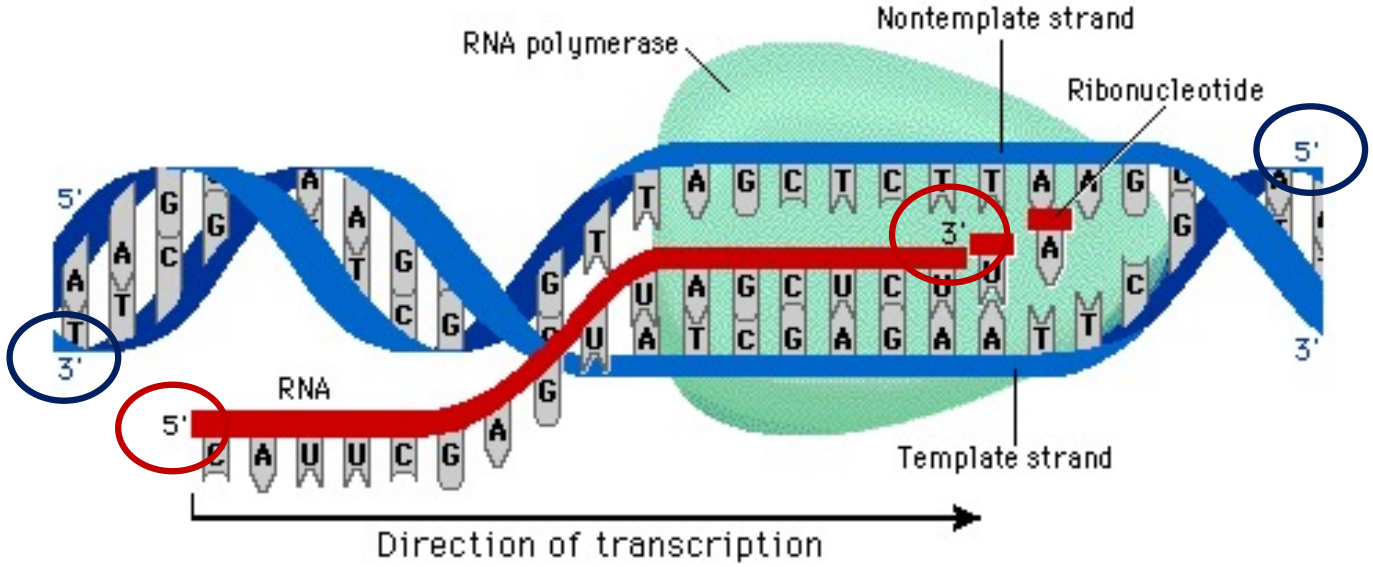
- Transcription factors bound to enhancer elements interact with basal complex
  - Factors have two domains
    - DNA binding domain for element interaction
    - Protein binding domain to interact with other transcription factors or RNAP
  - Domains are composed of motifs



mRNA synthesis

Transcription by RNAP II is initiated at the transcriptional start point at the 5'end:

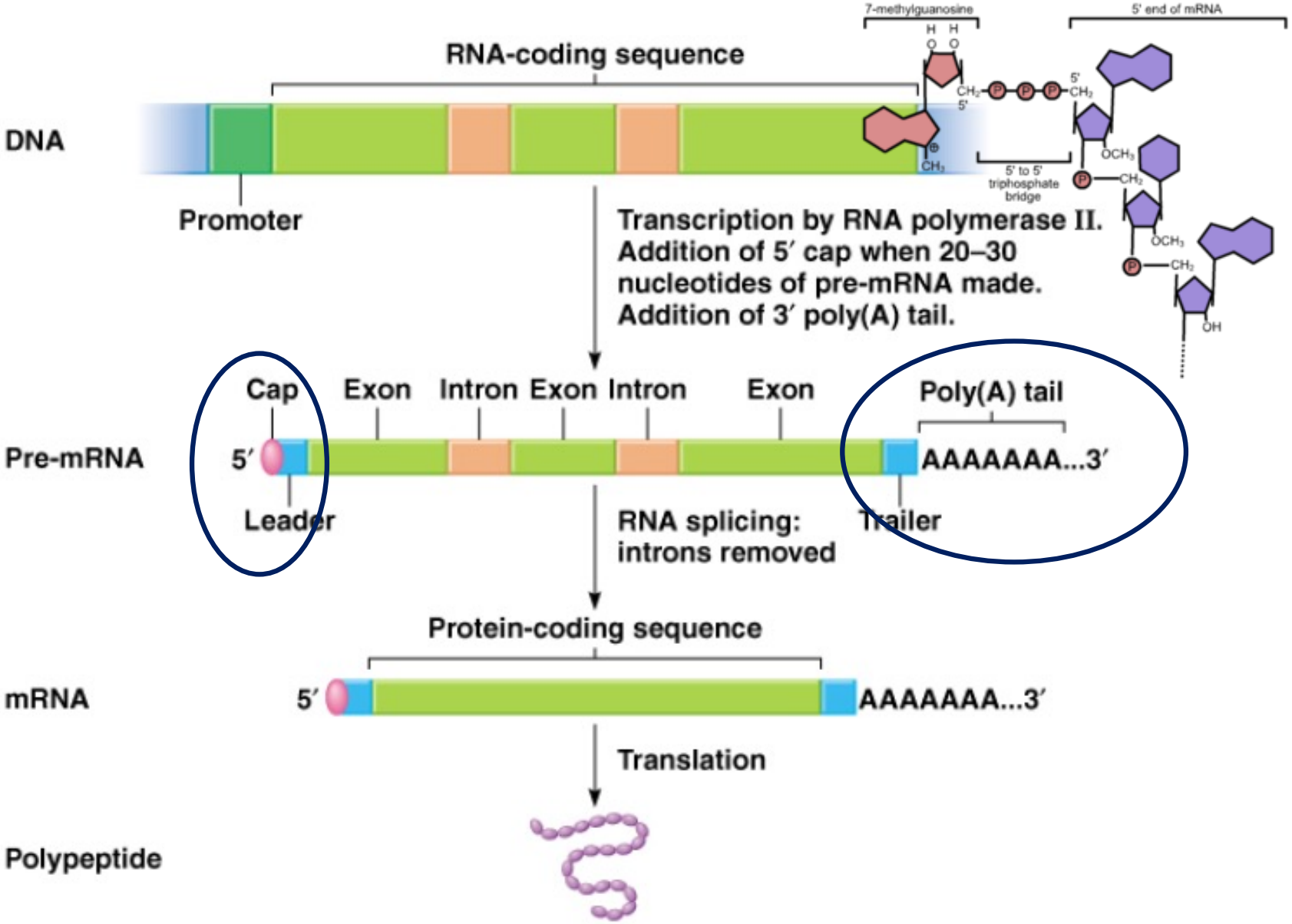
- ✓ RNA synthesis proceeds in a 5' to 3' direction
- ✓ Template DNA strand is read from 3' tot 5':
  - ✓ 5' to 3' strand of **non**transcribed DNA
    = coding or sense DNA strand
  - ✓ 3' to 5' strand of transcribed DNA
    = non-coding or antisense strand



Similar to DNA syntesis, except:

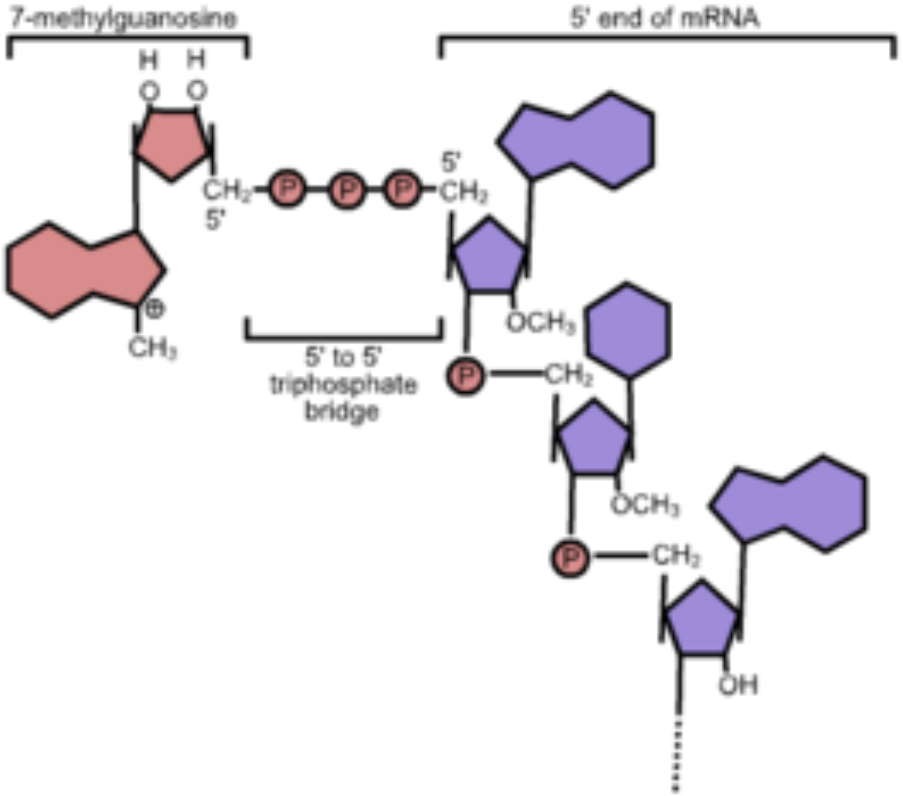
- ✓
- ✓
- ✓
- ✓

mRNA processing



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)



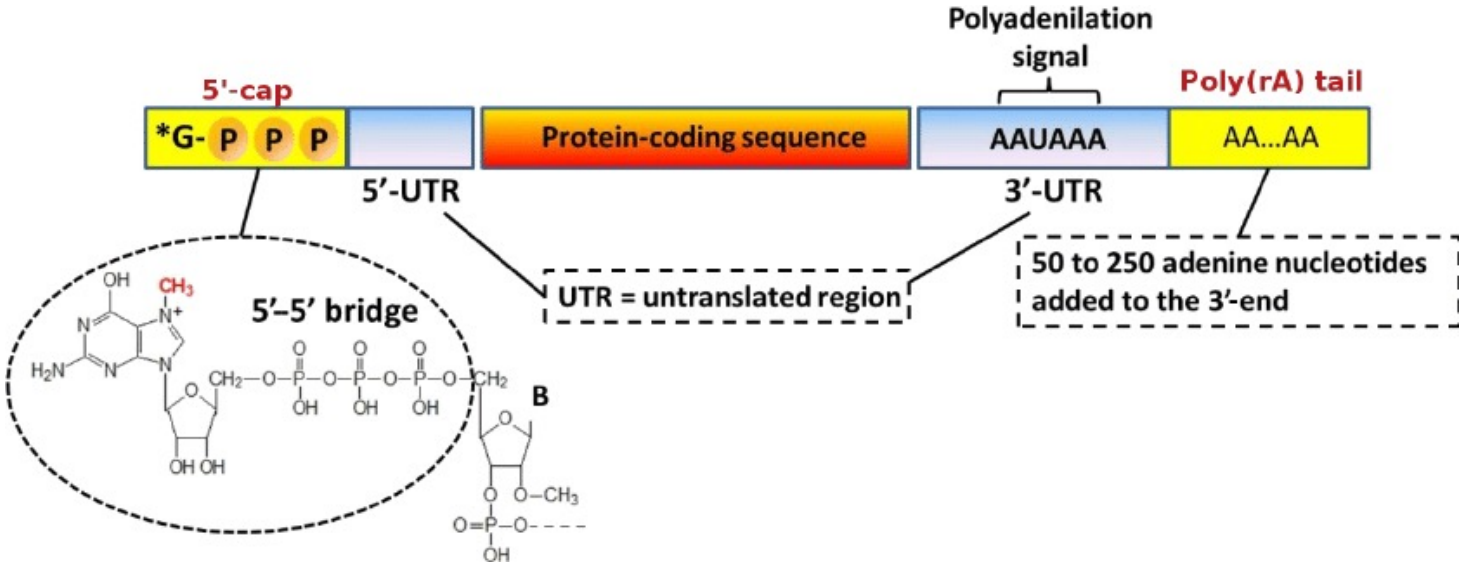
Function of the 5' cap:

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

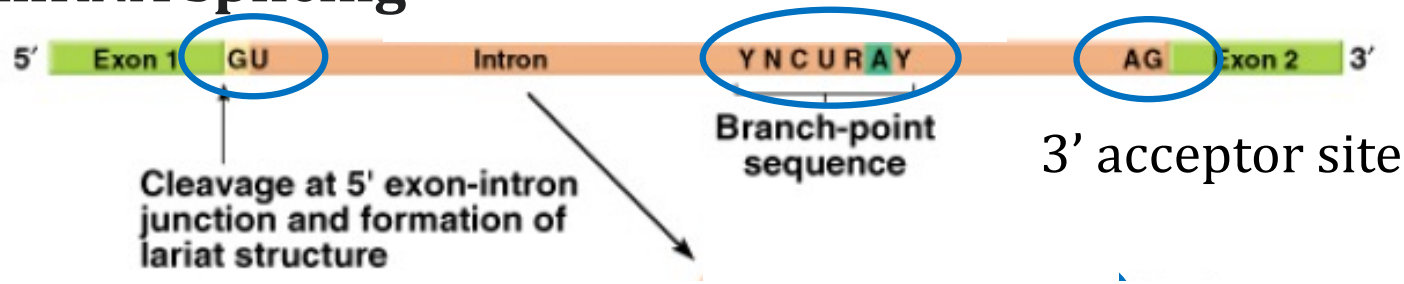
2. Poly (A) tail:

- ✓ 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.



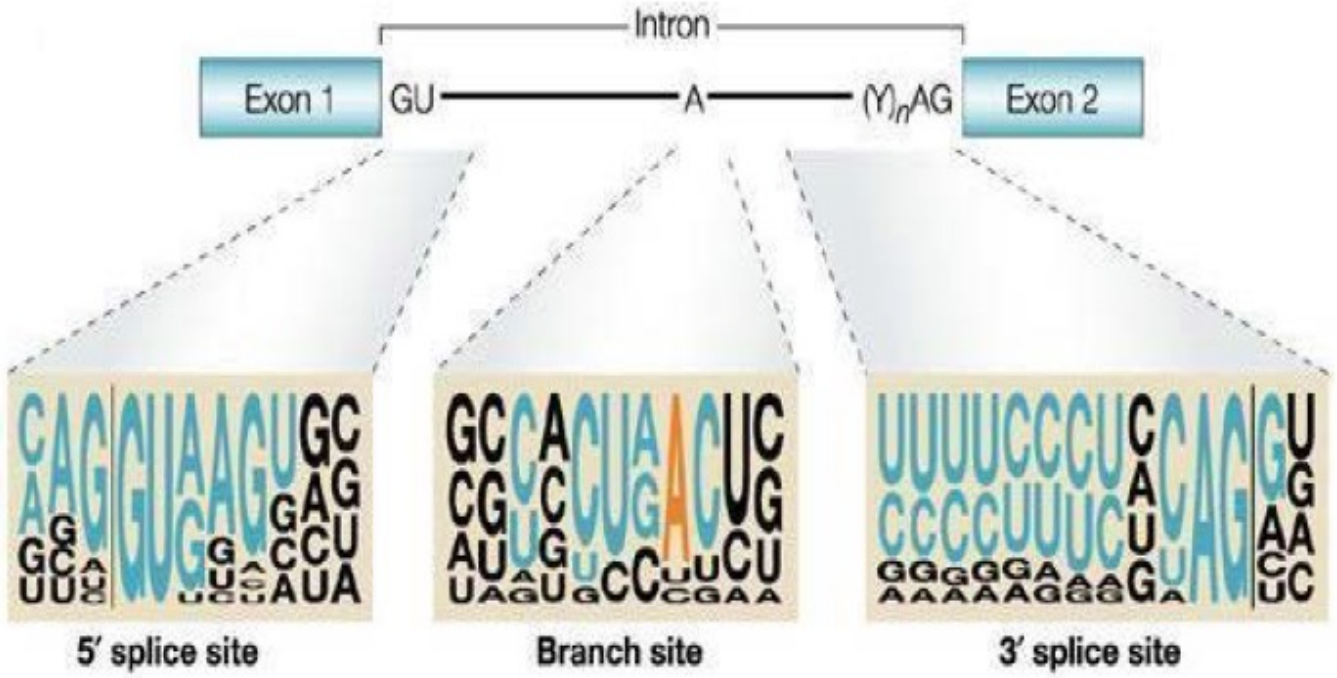
mRNA splicing

3. mRNA splicing



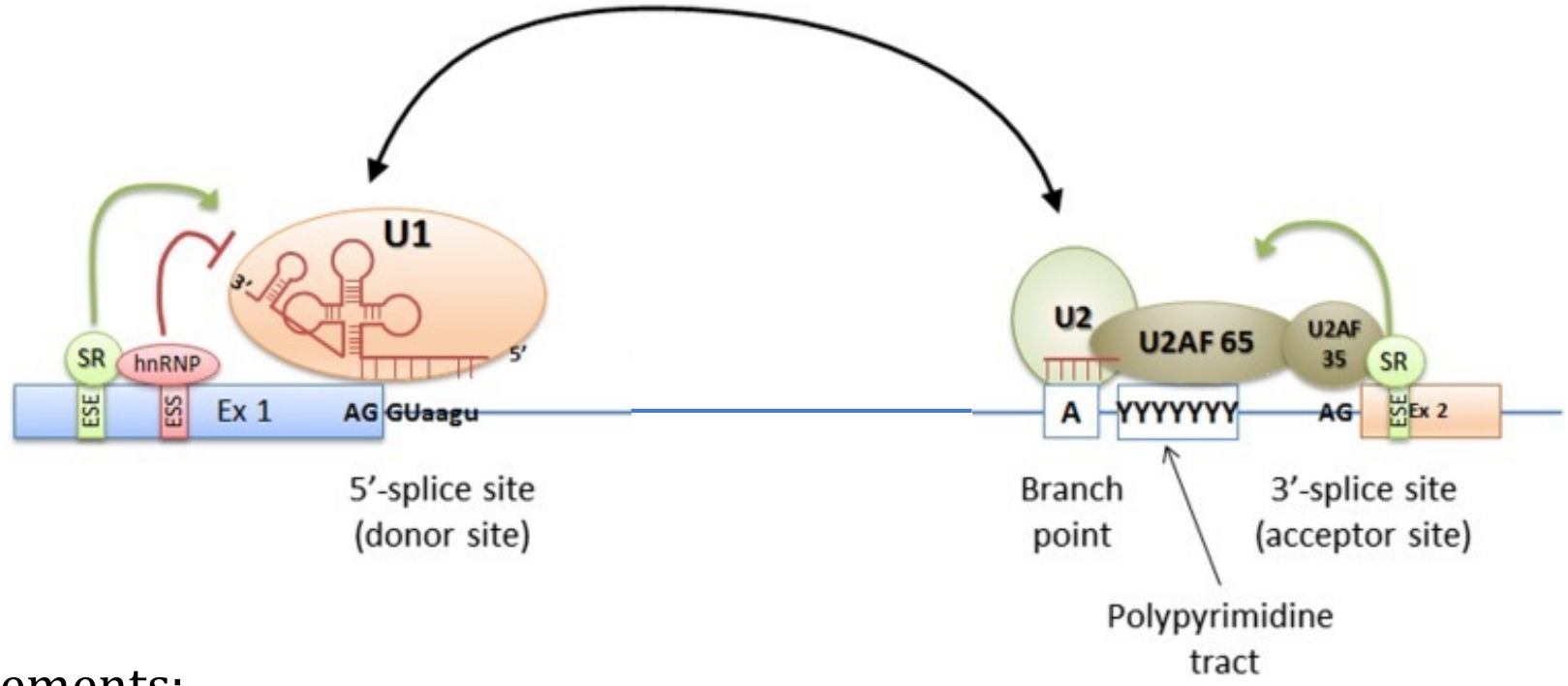
3’ acceptor site

5’ donor site



consensus sequence for an intron

mRNA splicing



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

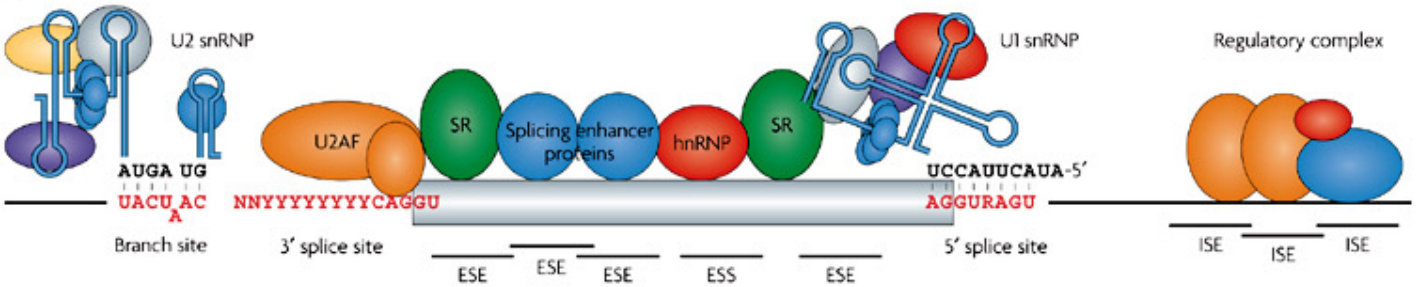
Adapted from Abramowicz and Gos, J Appl Genet 2018

alternative splicing

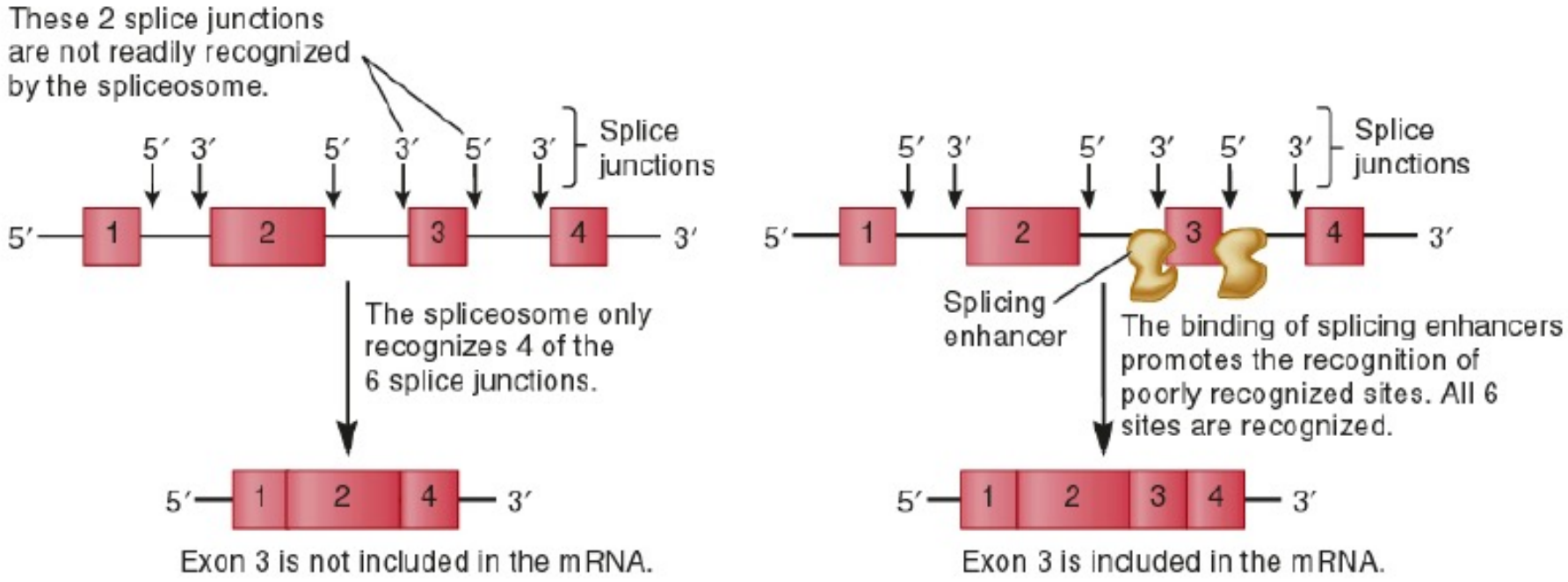
DEFINITION

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

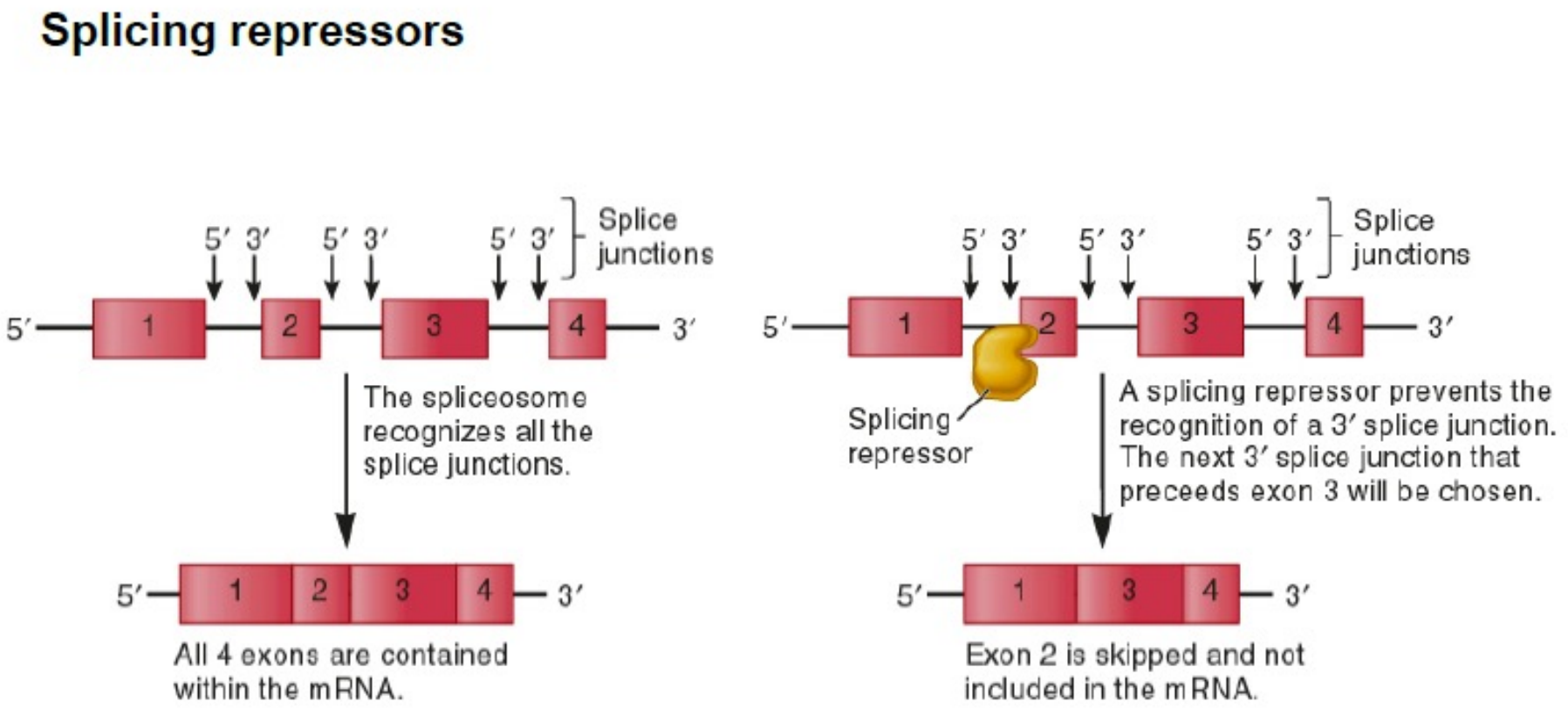
- Alternative splicing is not a random event
  - ✓ The specific pattern of splicing is regulated in a given cell
- It involves proteins known as splicing factors
  - ✓ These play a key role in the choice of splice sites
  - ✓ Spliceosome
- One example of splicing factors is the SR proteins



alternative splicing: enhancers

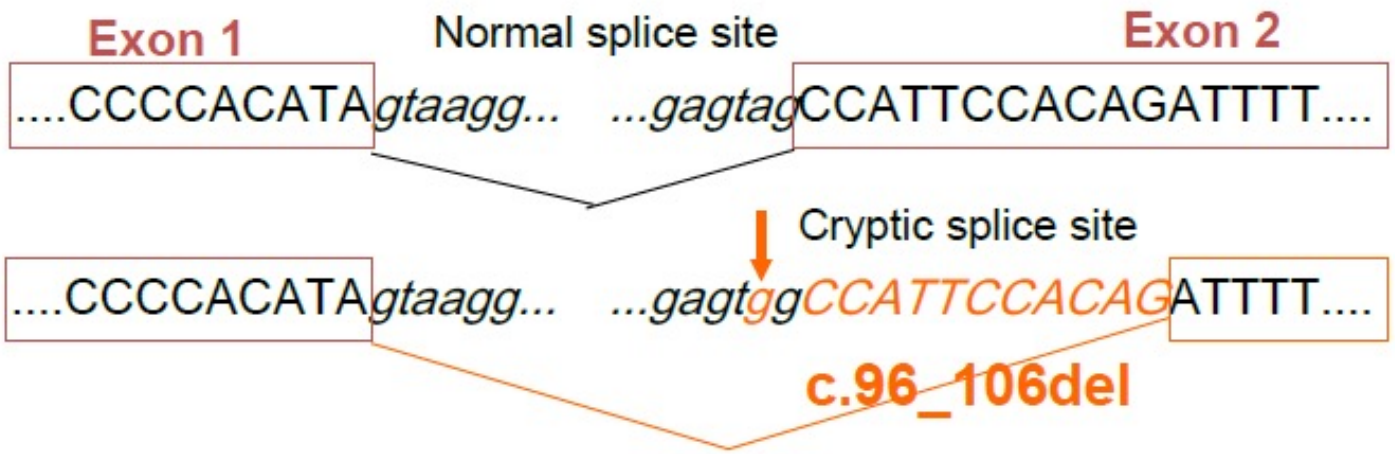


alternative splicing: repressors



splice site mutations

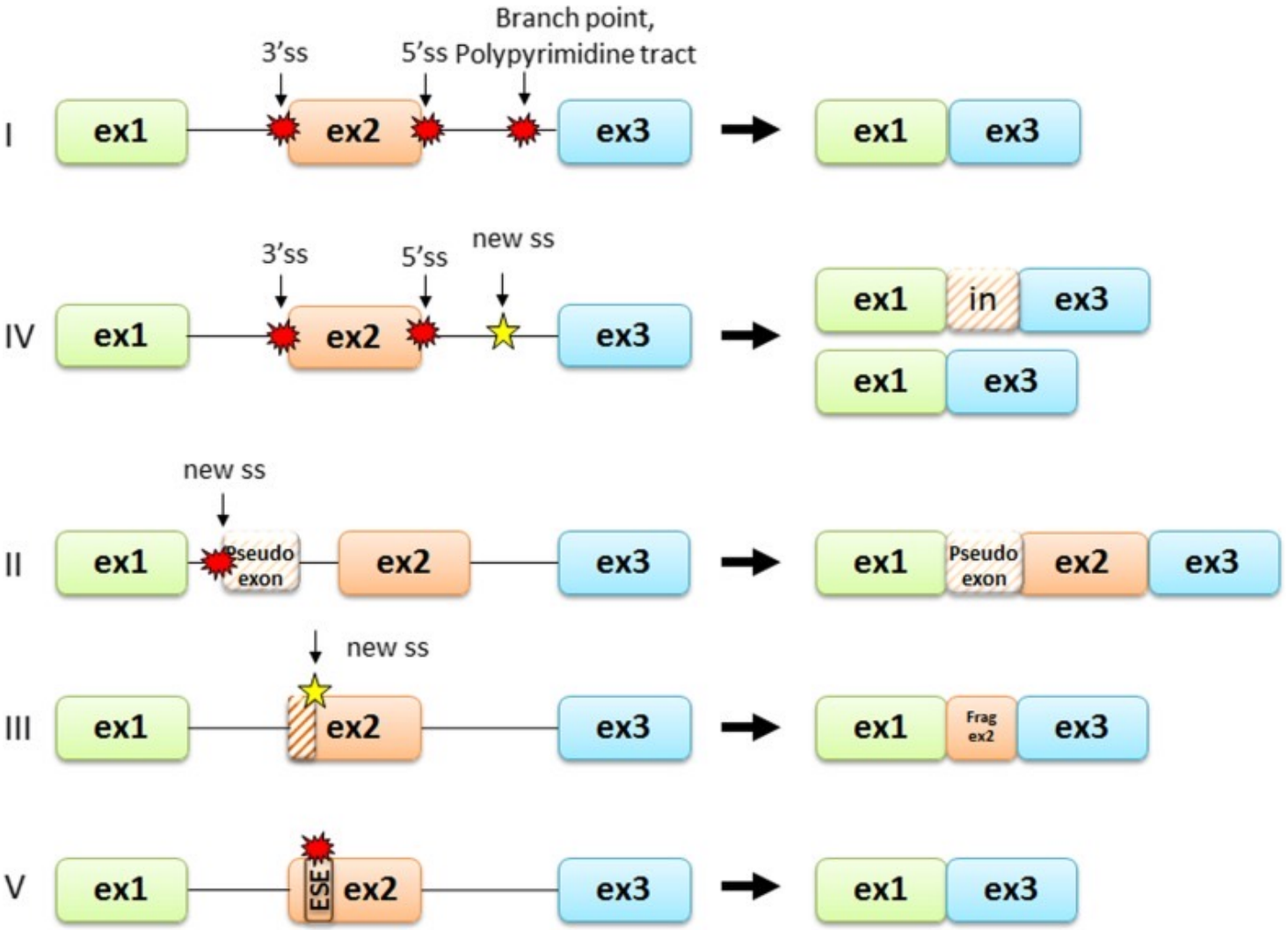
splice site mutations can activate a cryptic splice site in part of the transcript that usually is not spliced



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.

splice site mutations



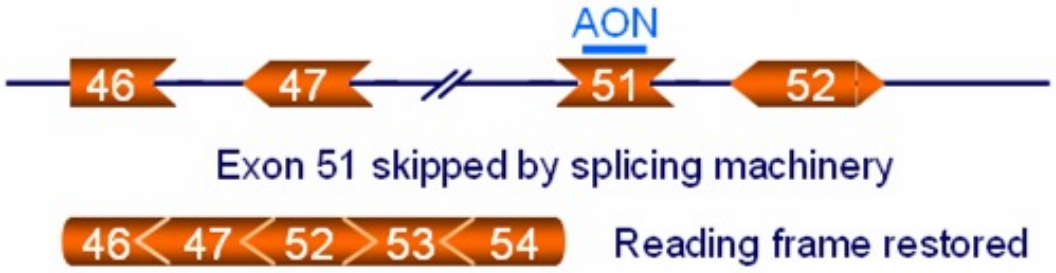
Abramowicz and Gos, J Appl Genet 2018

The Status of Exon Skipping as a Therapeutic Approach to Duchenne Muscular Dystrophy

Qi-Long Lu[1], Toshifumi Yokota[2], Shin'ichi Takeda[3], Luis Garcia[4], Francesco Muntoni[5] and Terence Partridge[2]

Duchenne: Open reading frame disrupted
Truncated, non-functional dystrophin

AON treatment: Exon skipped from pre-mRNA
Reading frame restored
Internally deleted, partly functional dystrophin
Becker-like phenotype



Dzierlega et al. Gene Therapy 2020

post-transcriptional mRNA modification

1. Adds or deletes nucleotides from a pre-RNA, or chemically alters the bases, so the mRNA bases do not match the DNA sequence.

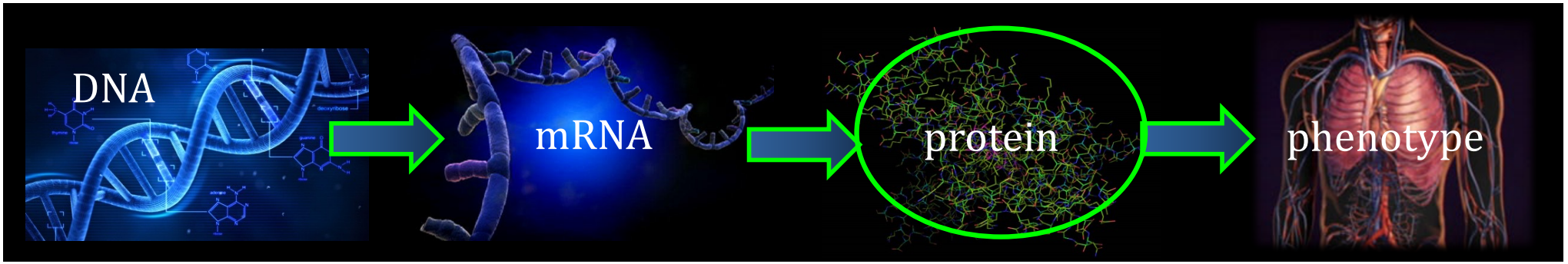
2. Can results in the substitution, addition, or deletion of amino acids (relative to the DNA template).

3. Generally cell or tissue specific.

4. Examples occur in protozoa, slime molds, plant organelles, and mammals.

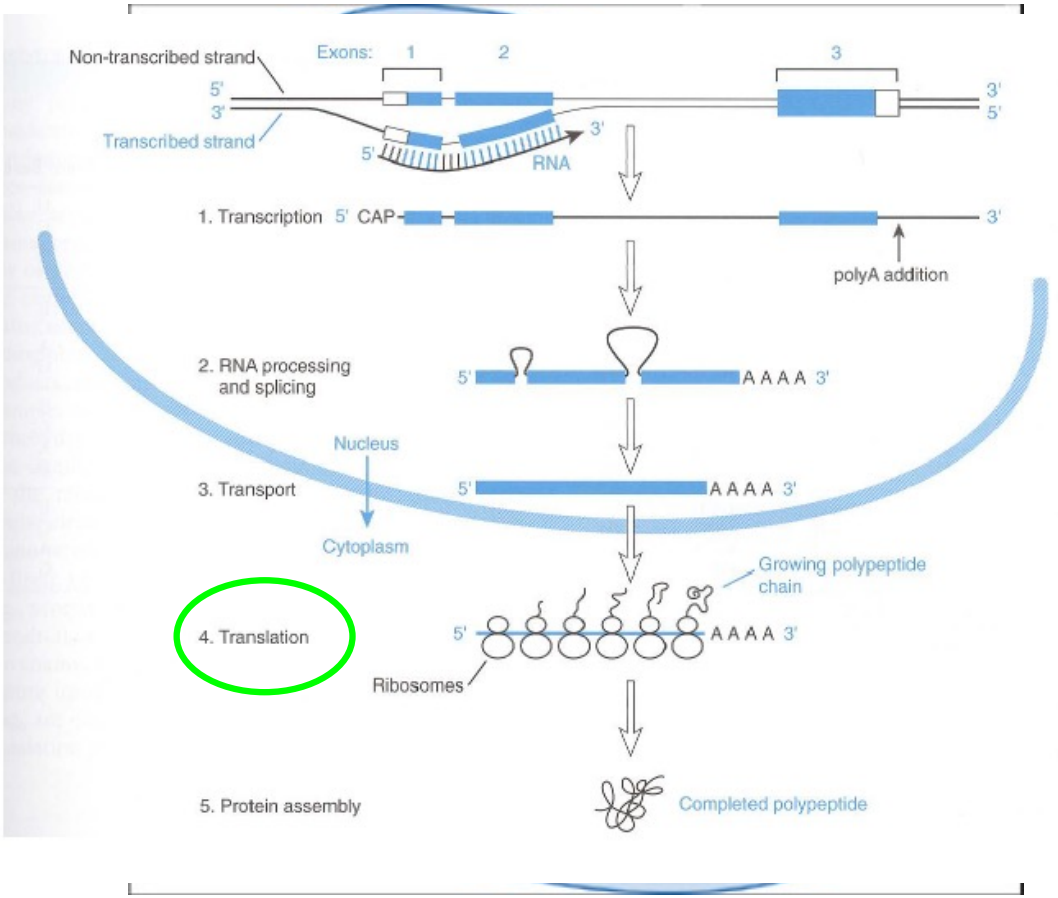
Am. J. Hum. Genet. 60:305–312, 1997

The central dogma



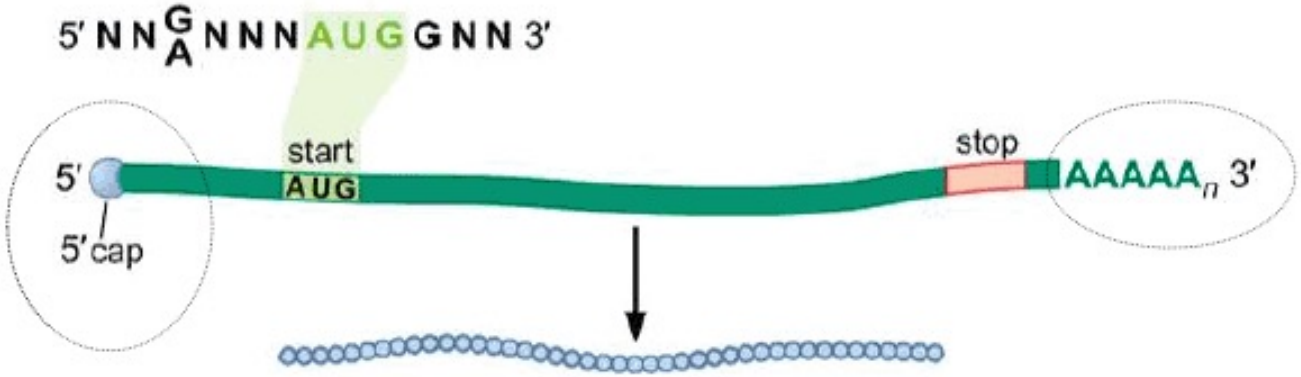
Translation

1. Initiation of translation
2. protein synthesis
3. protein modifications

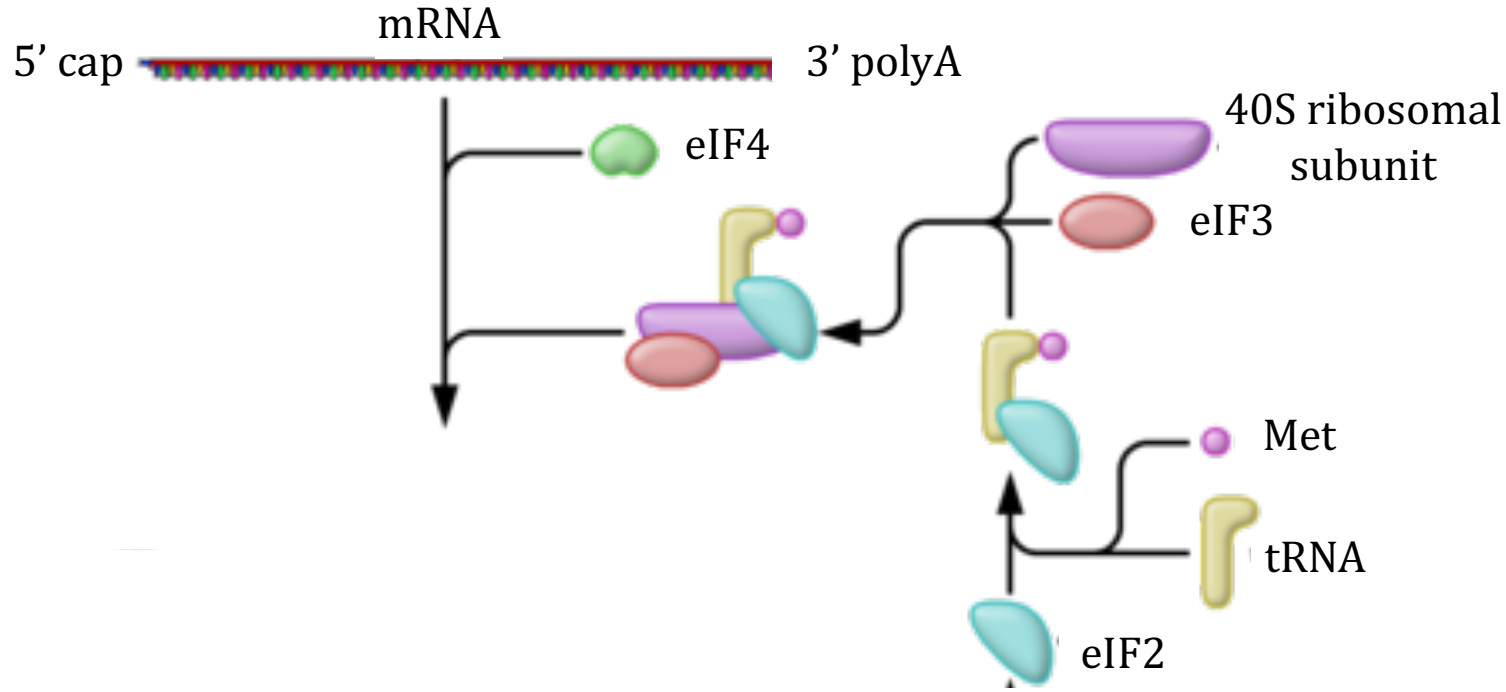


Initiation of translation

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



1. The 5' cap interacts with the 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) consensus sequence)
   - ✓ Complete ribosomal unit (80S) = 60S + 40S
   - ✓ Aminoacyl-tRNA

eIF4 mutations & Parkinson disease 18

KATHOLIEKE UNIVERSITEIT LEUVEN

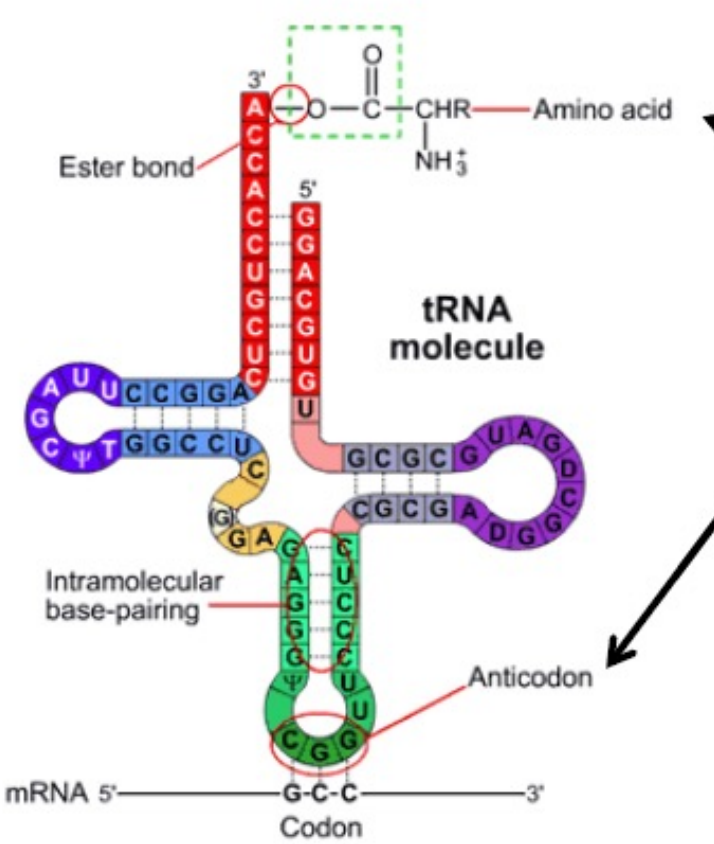
ARTICLE

Translation Initiator *EIF4G1* Mutations in Familial Parkinson Disease

Marie-Christine Chartier-Harlin,[1,2,24] Justus C. Dachsel,[3,24] Carles Vilariño-Güell,[4,24] Sarah J. Lincoln,[3] Frédéric Leprêtre,[1,2] Mary M. Hulihan,[3] Jennifer Kachergus,[3] Austen J. Milnerwood,[4] Lucia Tapia,[4] Mee-Sook Song,[4] Emilie Le Rhun,[5] Eugénie Mutez,[1,2,5] Lydie Larvor,[1,2] Aurélie Duflot,[1,2] Christel Vanbesien-Mailliot,[1,2,6] Alexandre Kreisler,[1,2,5] Owen A. Ross,[3] Kenya Nishioka,[3] Alexandra I. Soto-Ortolaza,[3] Stephanie A. Cobb,[3] Heather L. Melrose,[3] Bahareh Behrouz,[3] Brett H. Keeling,[3] Justin A. Bacon,[3] Emna Hentati,[3] Lindsey Williams,[3] Akiko Yanagiya,[7] Nahum Sonenberg,[7] Paul J. Lockhart,[8] Abba C. Zubair,[9] Ryan J. Uitti,[3] Jan O. Aasly,[10] Anna Krygowska-Wajs,[11] Grzegorz Opala,[12] Zbigniew K. Wszolek,[3] Roberta Frigerio,[13] Demetrius M. Maraganore,[13] David Gosal,[14] Tim Lynch,[14,15] Michael Hutchinson,[16] Anna Rita Bentivoglio,[17] Enza Maria Valente,[18,19] William C. Nichols,[20] Nathan Pankratz,[21] Tatiana Foroud,[21] Rachel A. Gibson,[22] Faycal Hentati,[23] Dennis W. Dickson,[3] Alain Destée,[1,2,5,25] and Matthew J. Farrer[3,4,25,*]

Genome-wide analysis of a multi-incident family with autosomal-dominant parkinsonism has implicated a locus on chromosomal region 3q26-q28. Linkage and disease segregation is explained by a missense mutation c.3614G>A (p.Arg1205His) in eukaryotic translation initiation factor 4-gamma (*EIF4G1*). Subsequent sequence and genotype analysis identified *EIF4G1* c.1505C>T (p.Ala502Val), c.2056G>T (p.Gly686Cys), c.3490A>C (p.Ser1164Arg), c.3589C>T (p.Arg1197Trp) and c.3614G>A (p.Arg1205His) substitutions in affected subjects with familial parkinsonism and idiopathic Lewy body disease but not in control subjects. Despite different countries of origin, persons with *EIF4G1* c.1505C>T (p.Ala502Val) or c.3614G>A (p.Arg1205His) mutations appear to share haplotypes consistent with ancestral founders. eIF4G1 p.Ala502Val and p.Arg1205His disrupt eIF4E or eIF3e binding, although the wild-type protein does not, and render mutant cells more vulnerable to reactive oxidative species. *EIF4G1* mutations implicate mRNA translation initiation in familial parkinsonism and highlight a convergent pathway for monogenic, toxin and perhaps virally-induced Parkinson disease.

Elongation: protein synthesis

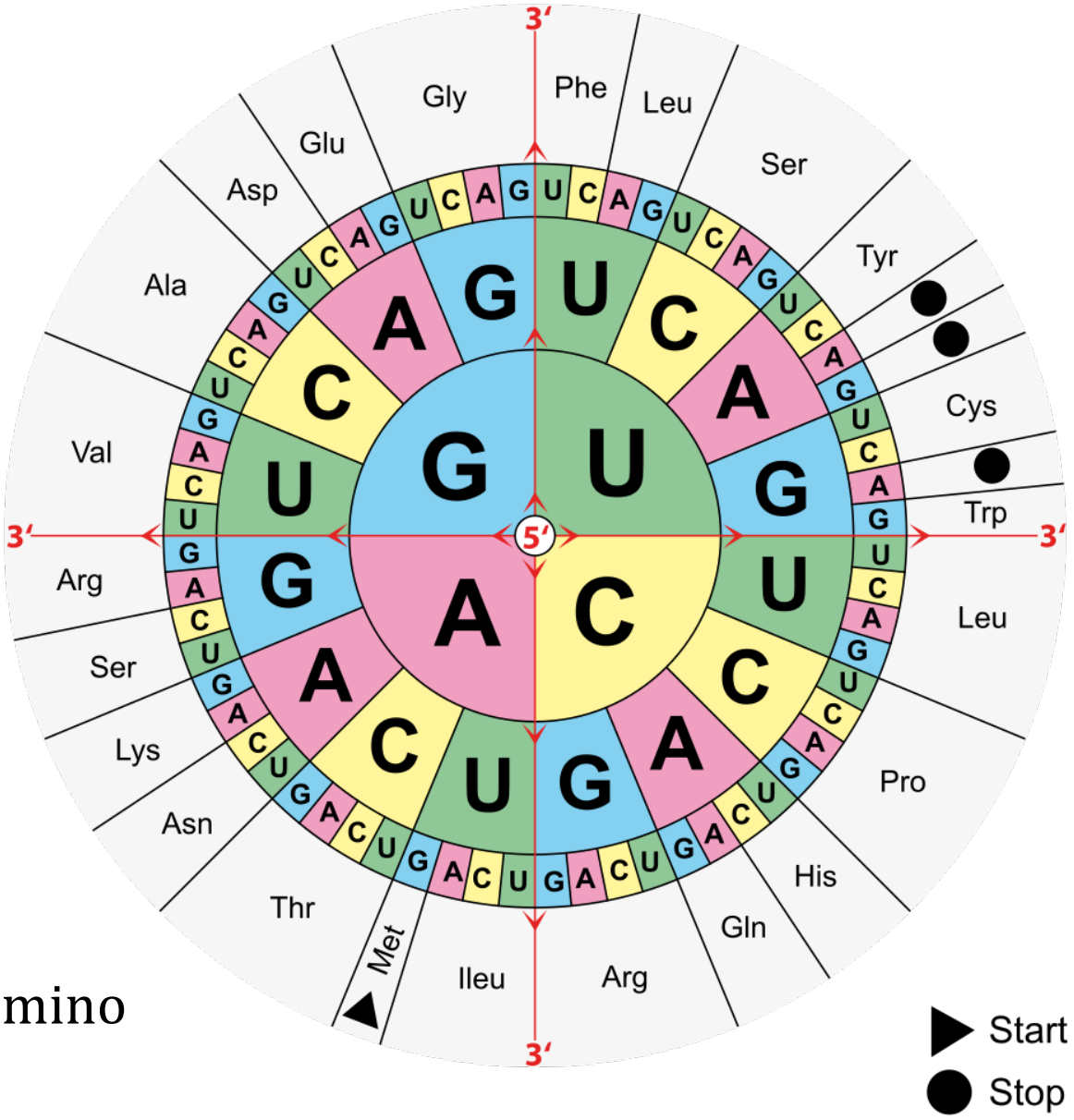


- Transfer RNA
- Bound to one amino acid on one end
- Anticodon on the other end complements mRNA codon

Elongation: protein synthesis

Codon: $4^3 = 64$ codons

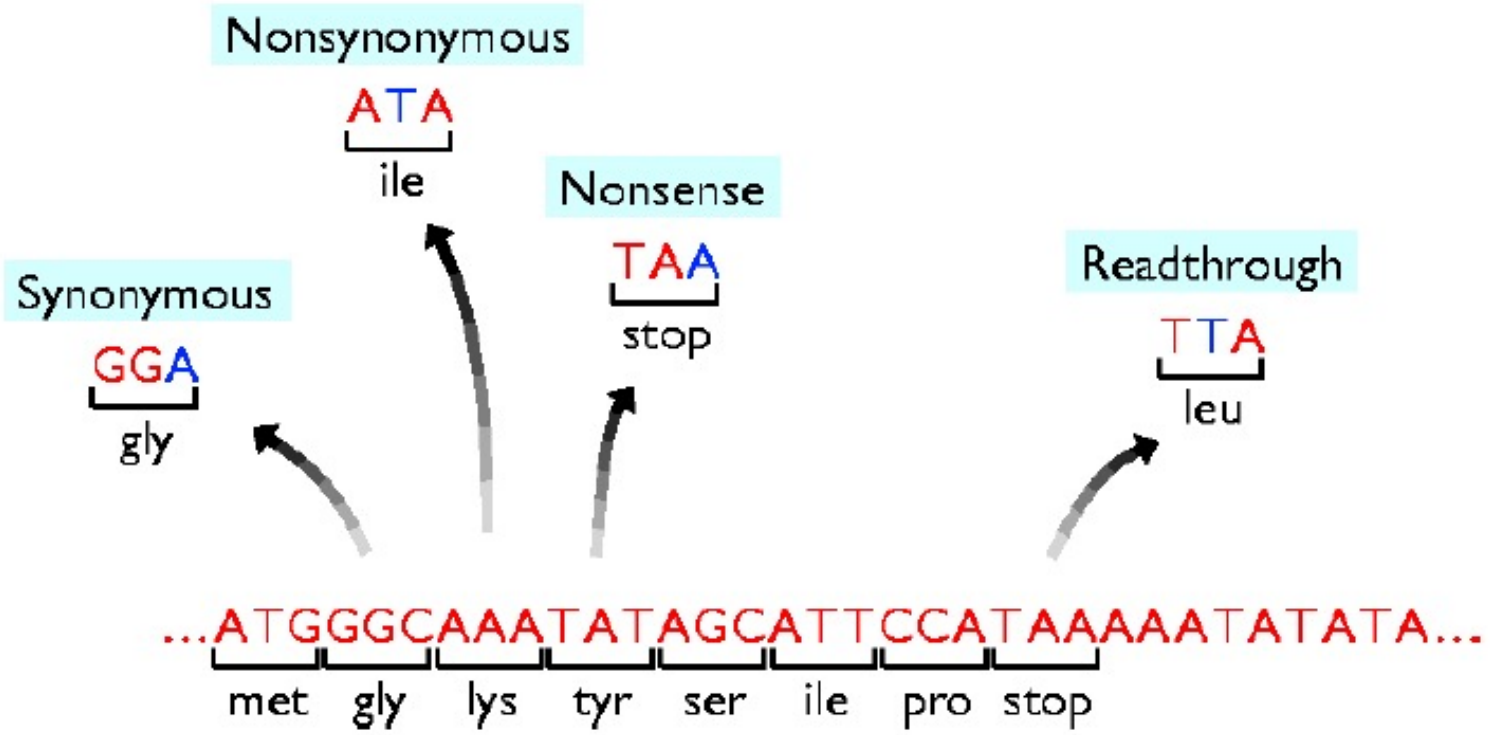
- Start codon: AUG (Met)
- Stop codon:
  - ✓ UGA
  - ✓ UAA
  - ✓ UAG



20 naturally occurring amino acids:

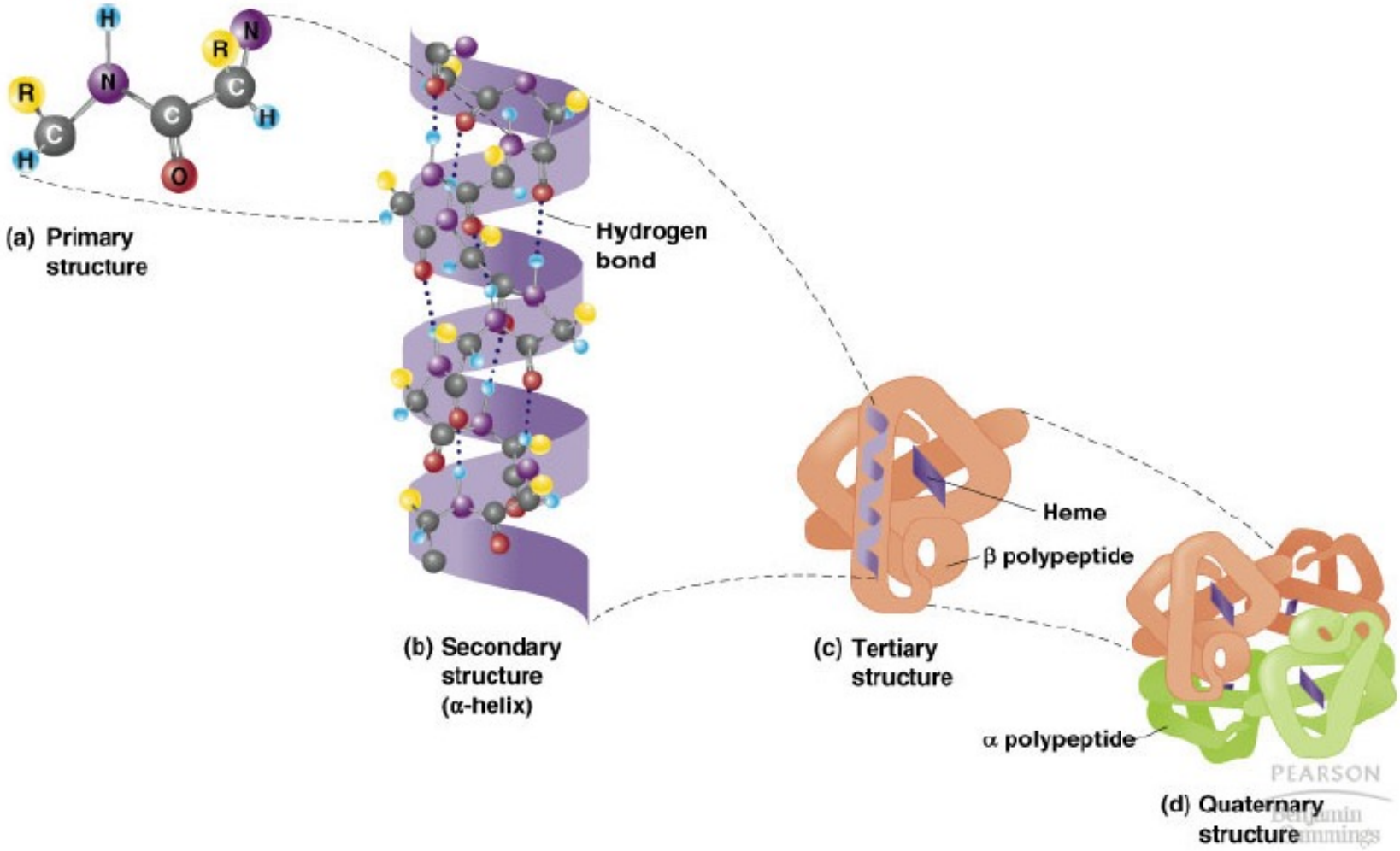
- ✓ Met = AUG (start)
- ✓ 60 codons for 19 other amino acids

DNA variants



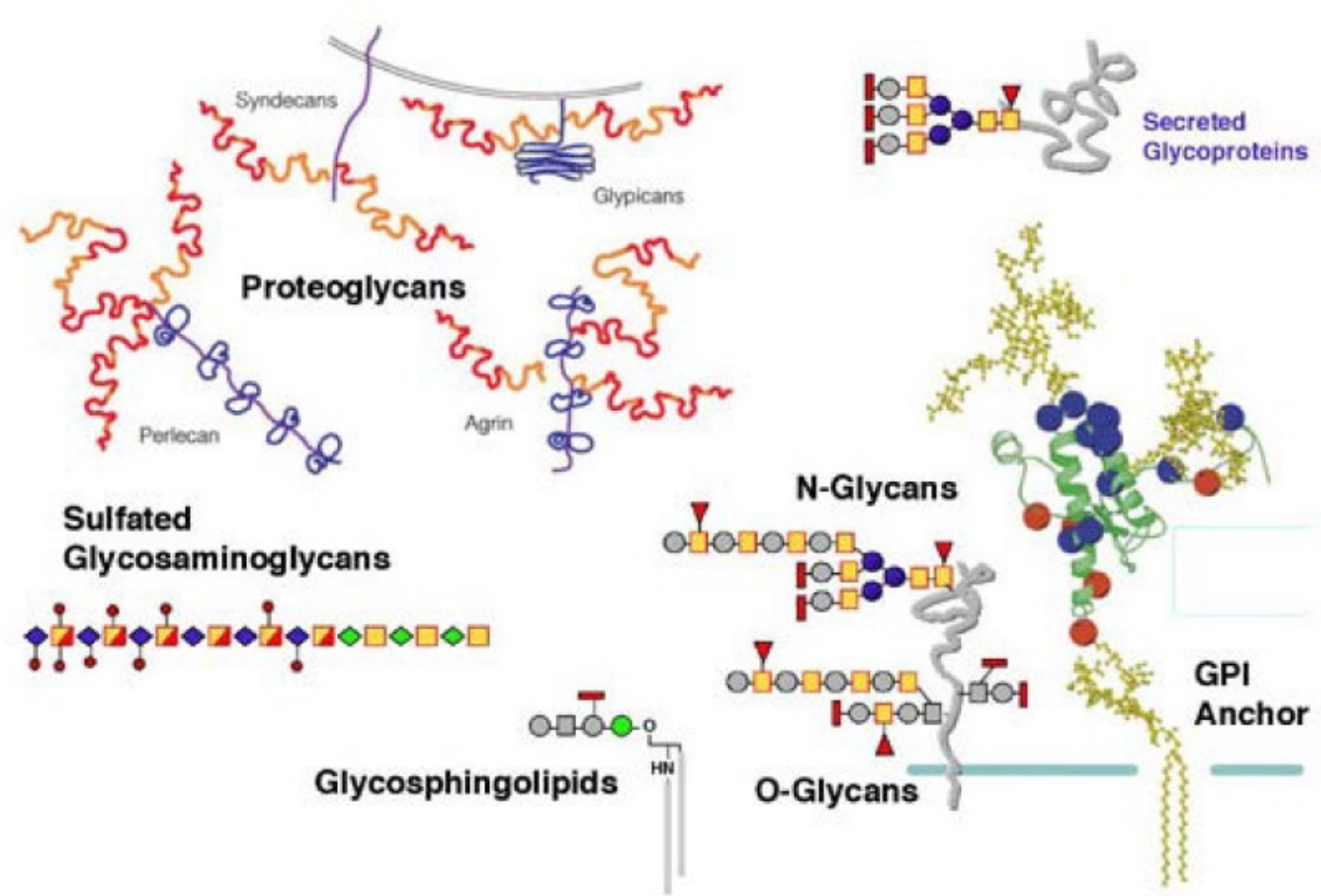
Protein synthesis

- A protein is a linear polymer of amino acids linked together by peptide bonds.



MW: 114.11 MW: 128.14 MW: 137.14, $pK_a = 6.04$ MW: 128.17, $pK_a = 10.79$ MW: 156.19, $pK_a = 12.48$

Protein modifications

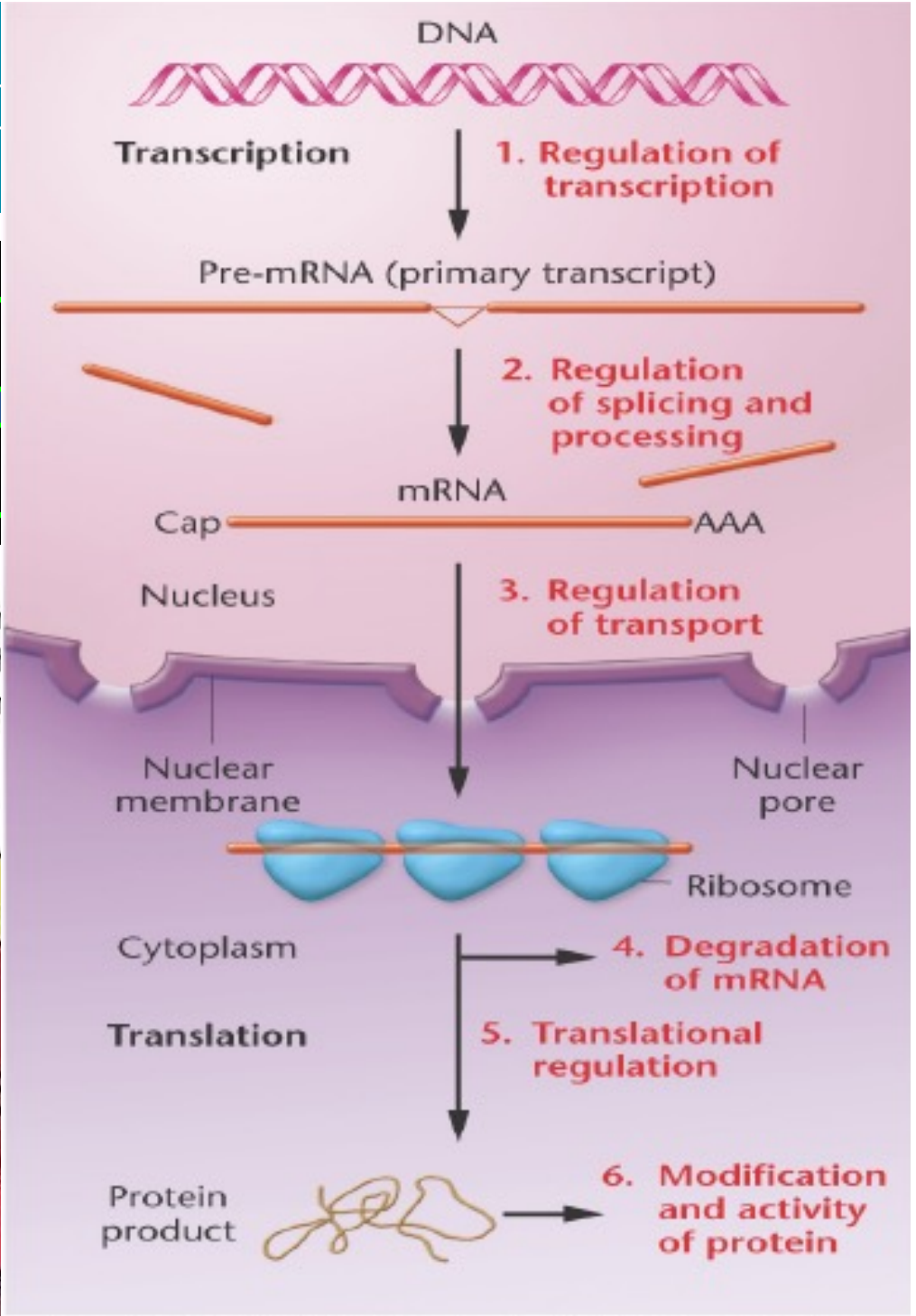


- Glycosylation, methylation, phosphorylation,…

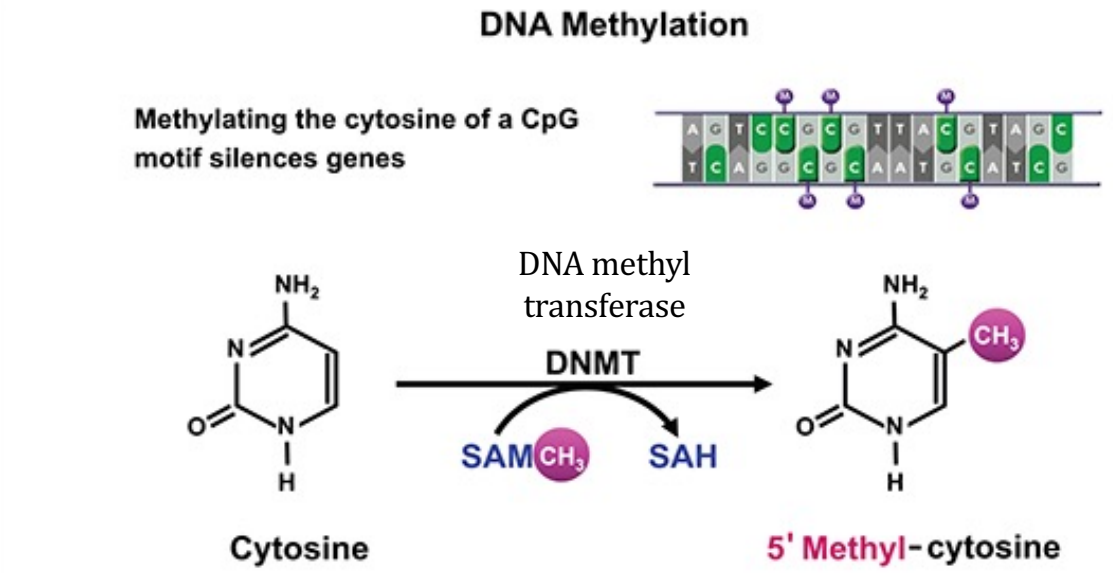
Levels o

Regulation of transcription:

1. DNA methylation
2. Histone modification



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:

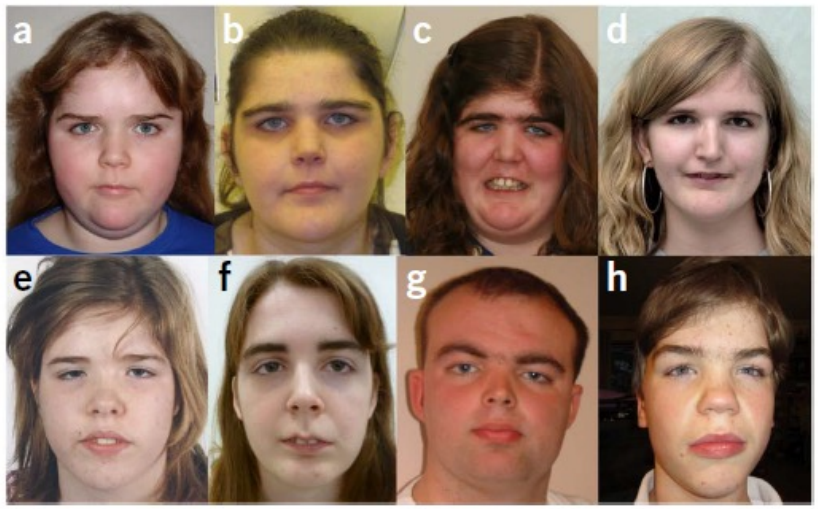
- ✓ *de novo* methylation: installing methylation patterns early in development DNMT3a and DNMT3b:
  - DNA methylation can stably alter the expression of genes in cells during cell division and differentiate from embryonic stem cells into specific tissues.
  - DNA methylation is typically removed during 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.

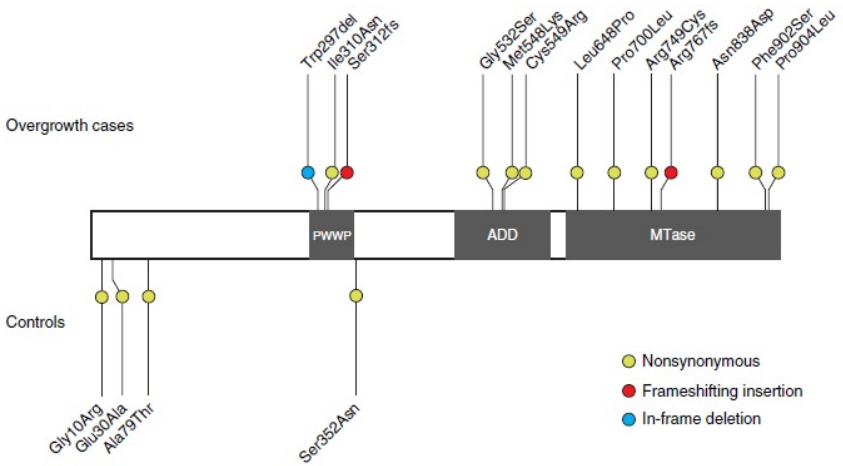
DNA methylation related disease

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

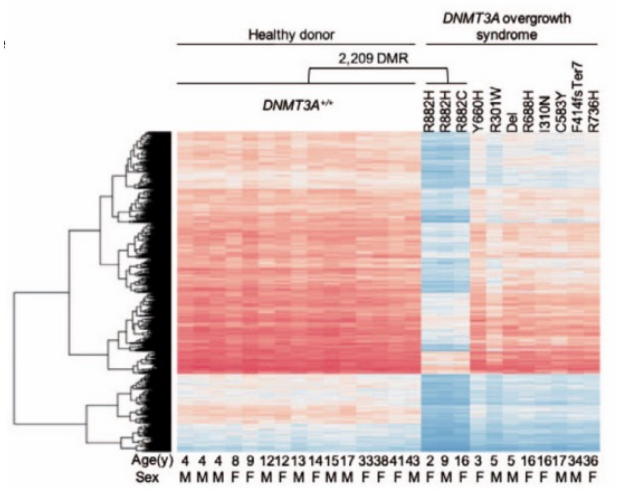
nature COMMUNICATIONS

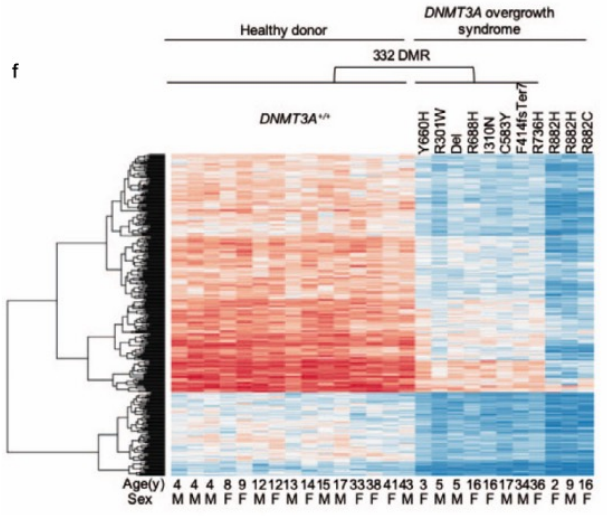
ARTICLE

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

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

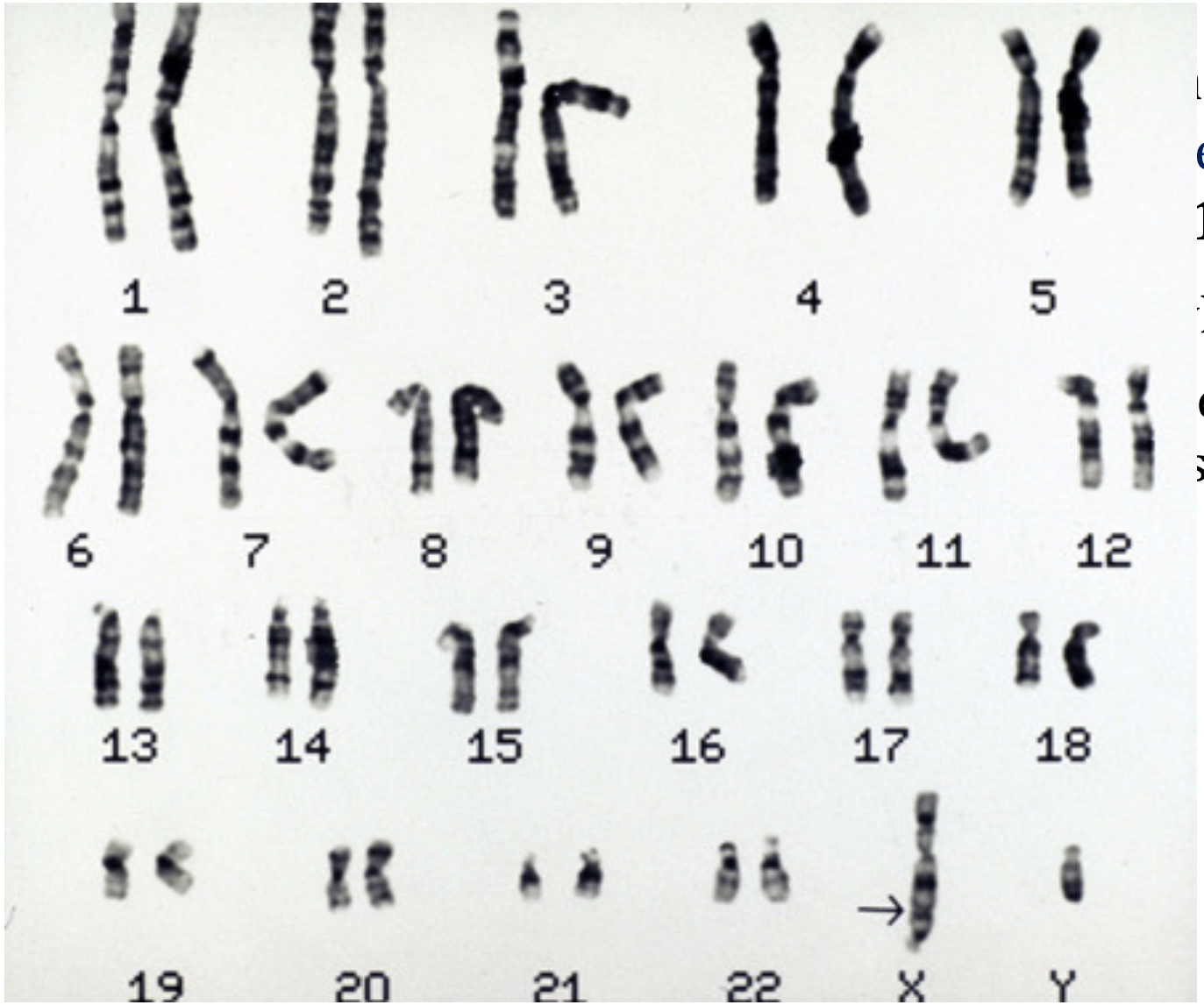
Amanda M. Smith[1], Taylor A. LaValle[1], Marwan Shinawi[2], Sai M. Ramakrishnan[1], Haley J. Abel[1], Cheryl A. Hill[3], Nicole M. Kirkland[3], Michael P. Rettig[1], Nichole M. Helton[1], Sharon E. Heath[1], Francesca Ferraro[1], David Y. Chen[4], Sangeeta Adak[5], Clay F. Semenkovich[5], Diana L. Christian[6],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





DNA methylation related disease

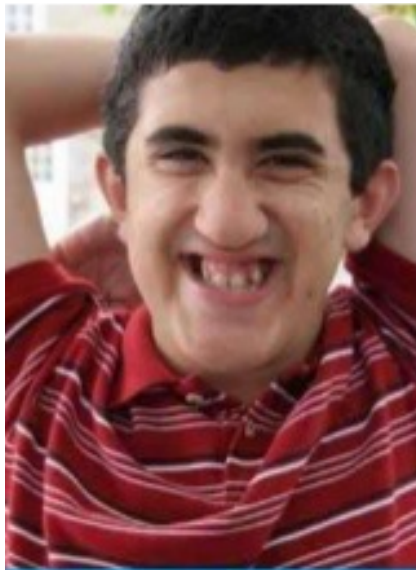
FRAGILE X SYNDROME



utation') in *FMR1*
ethylation of the
1 expression

lation status

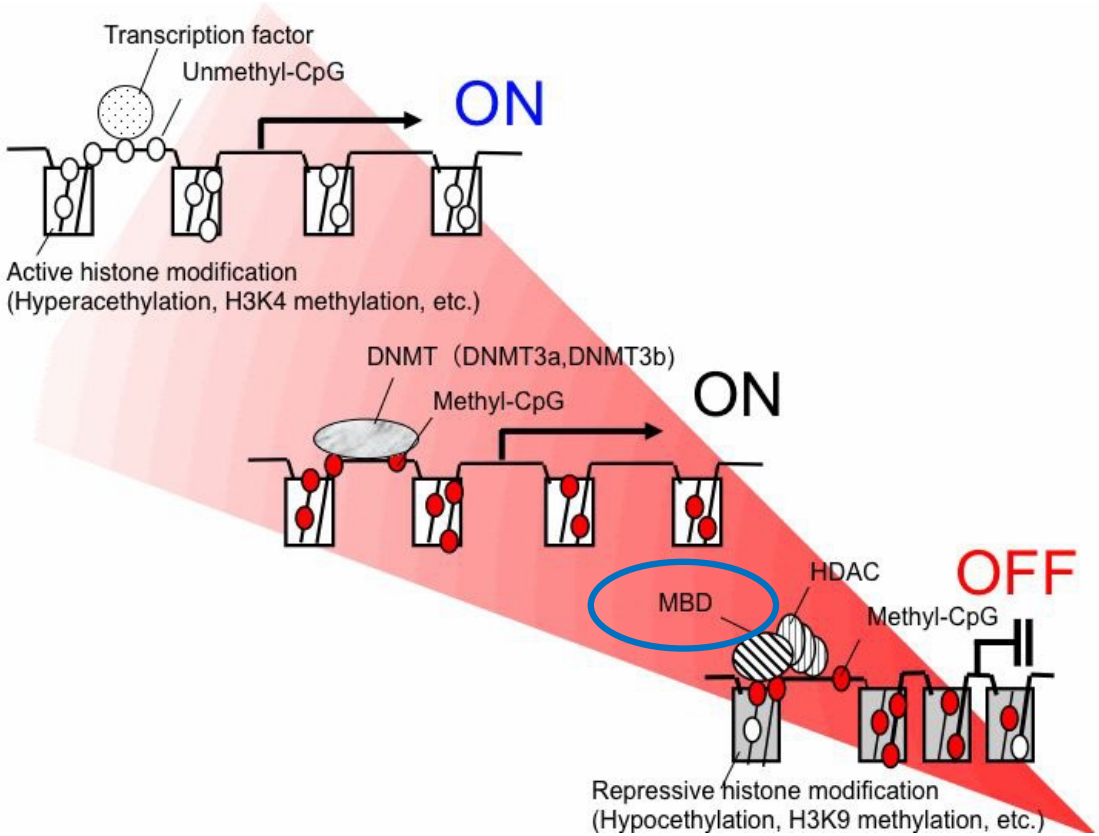
ed males and mild
s



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.



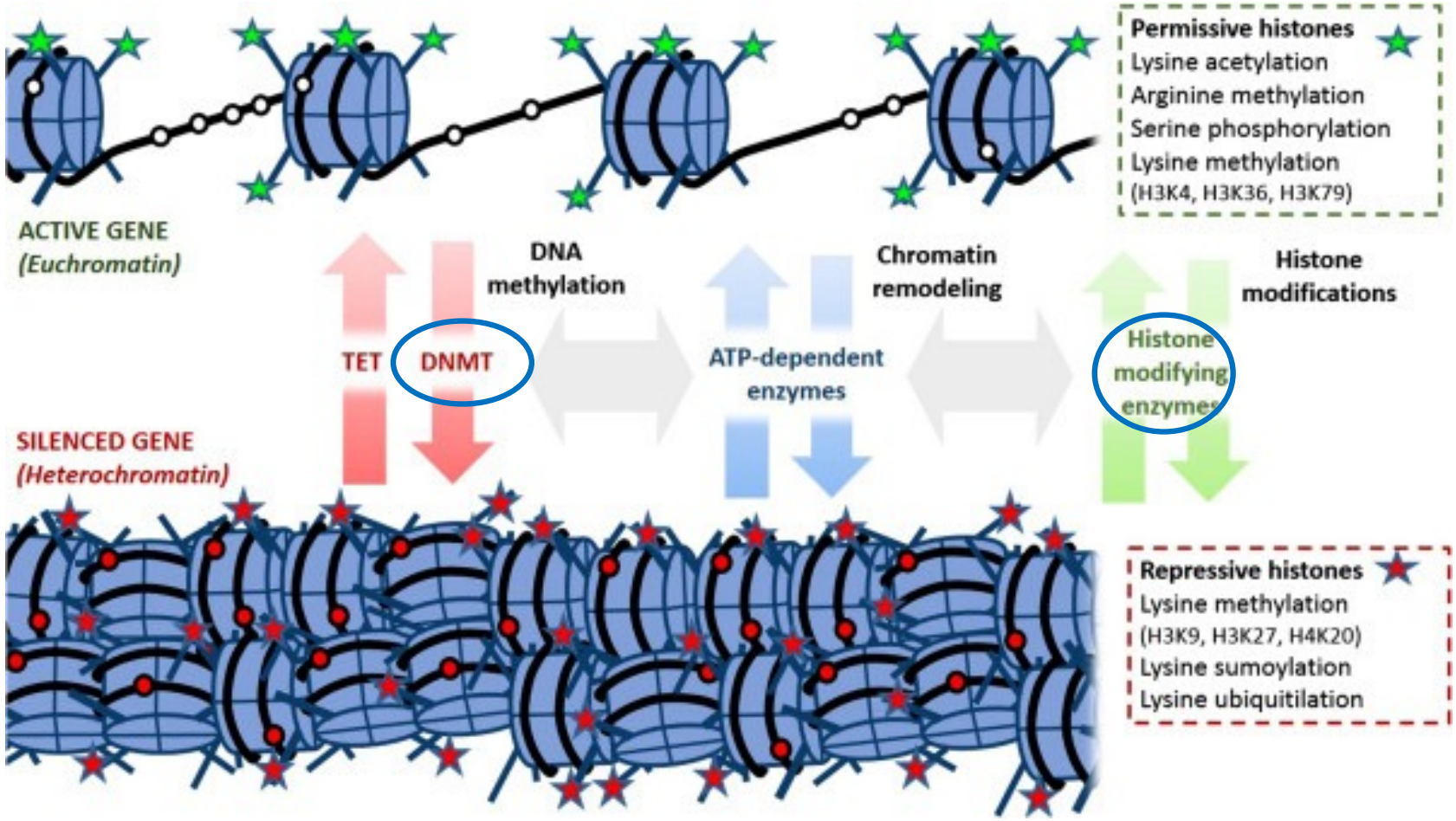
MBD proteins related disease

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 of MBD protein, which can act as a transcriptional repressor.
MECP2 duplications cause severe ID in boys.

Levels of Control: chromatin remodeling



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

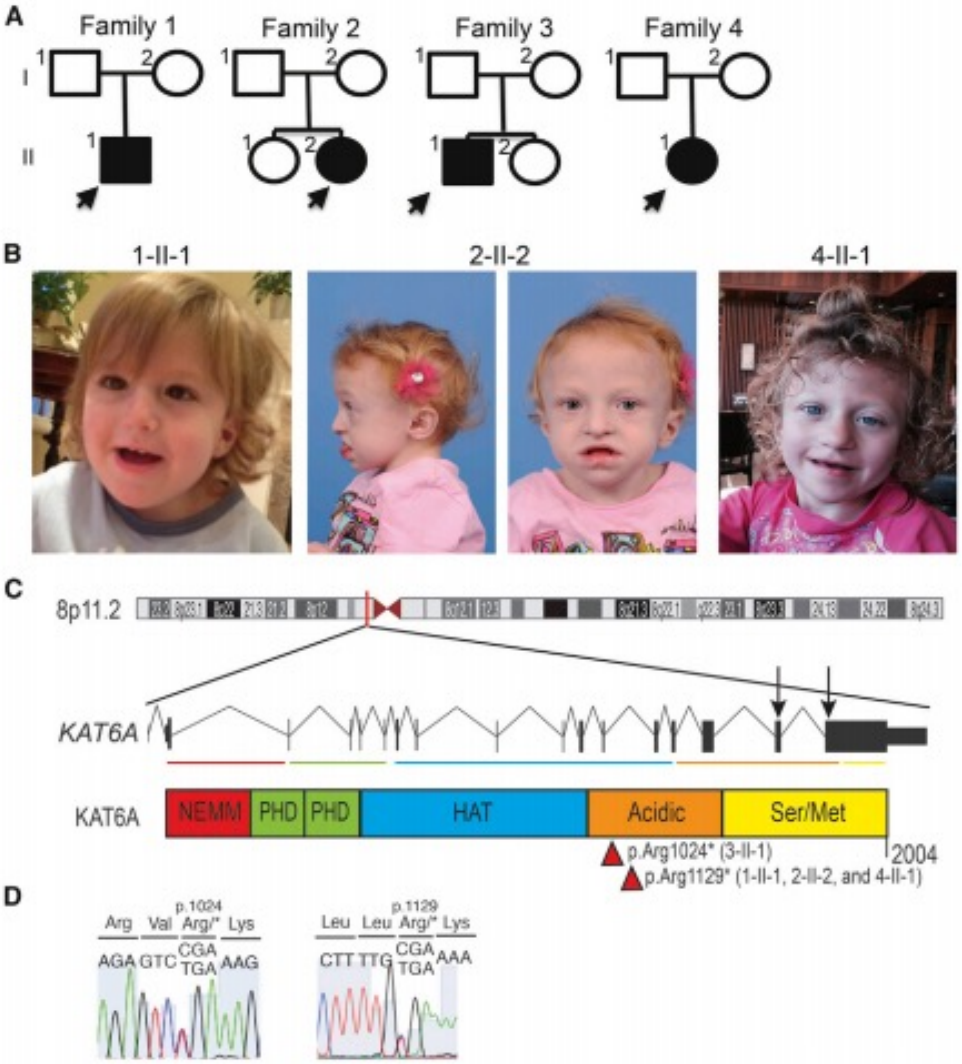
Histone acetylation related disease

REPORT

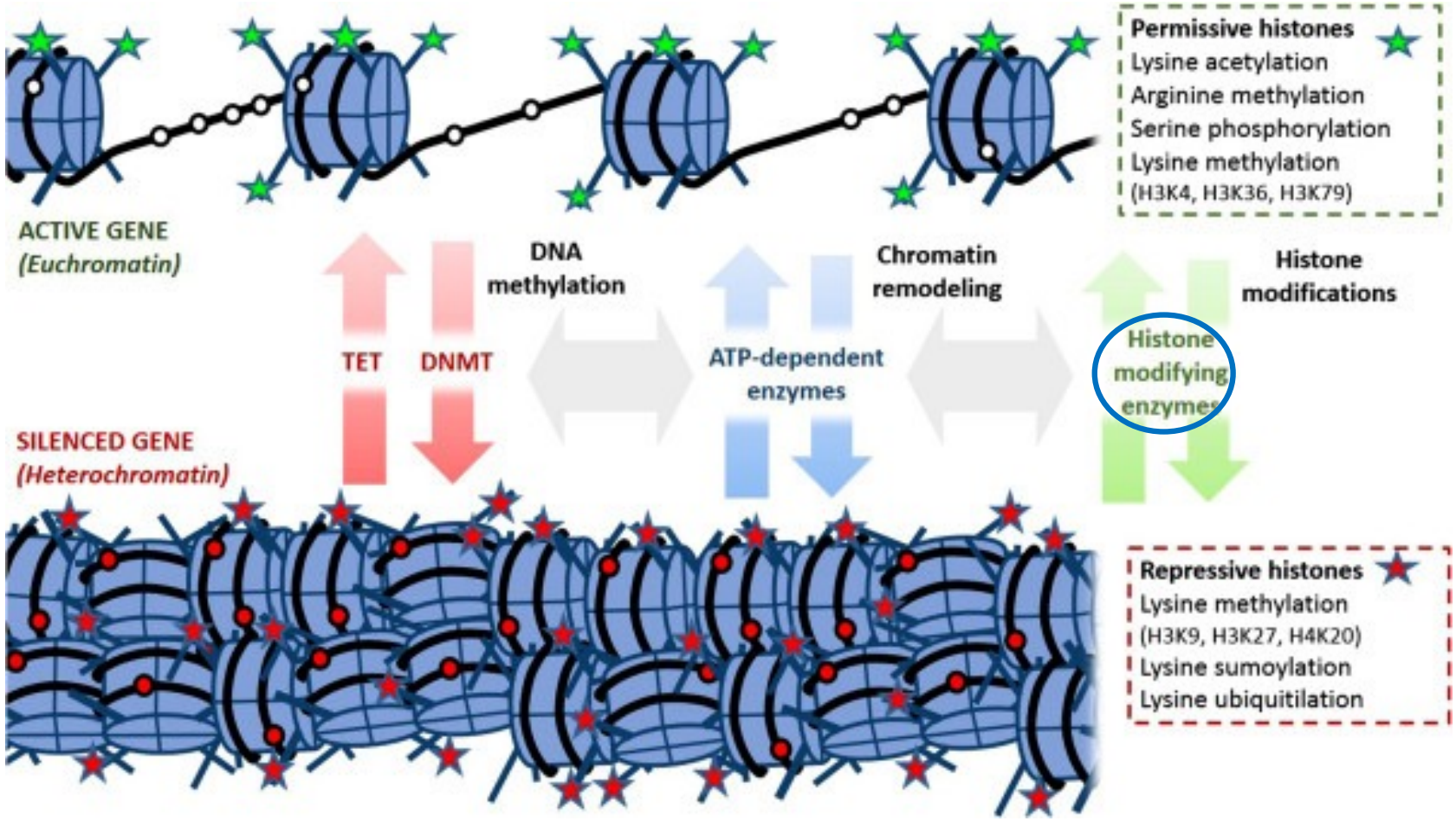
De Novo Nonsense Mutations in *K* Acetyl-Transferase Gene, Cause a S Including Microcephaly and Globa

Valerie A. Arboleda,[1] Hane Lee,[1] Naghmeh Dorran Colleen Forsyth Macmurdo,[6] Melanie A. Manning, Florian Barthelemy,[9] M. Carrie Miceli,[9] Fabiola Qu Joshua L. Deignan,[1] UCLA Clinical Genomics Cen and Stanley F. Nelson[1,10,*]

Chromatin remodeling through histone acetyltransferase (HAT) ar processes including the cell-cycle, cell differentiation, metabolism, tone acetylation and deacetylation result in multiple congenital an delay, microcephaly and dysmorphism. Here, we report a syndro (a.k.a., MOZ, MYST3) identified by clinical exome sequencing (CES (c.3385C>T [p.Arg1129*]) was observed in three individuals, an (c.3070C>T [p.Arg1024*]). Neither of these variants was present i among all four probands include primary microcephaly, global dev dysmorphism, as well as more varied features such as feeding diffic that *KAT6A* mutations result in dysregulation of H3K9 and H3K18 tone acetylation, KAT6A affects multiple cellular processes and illus disease.



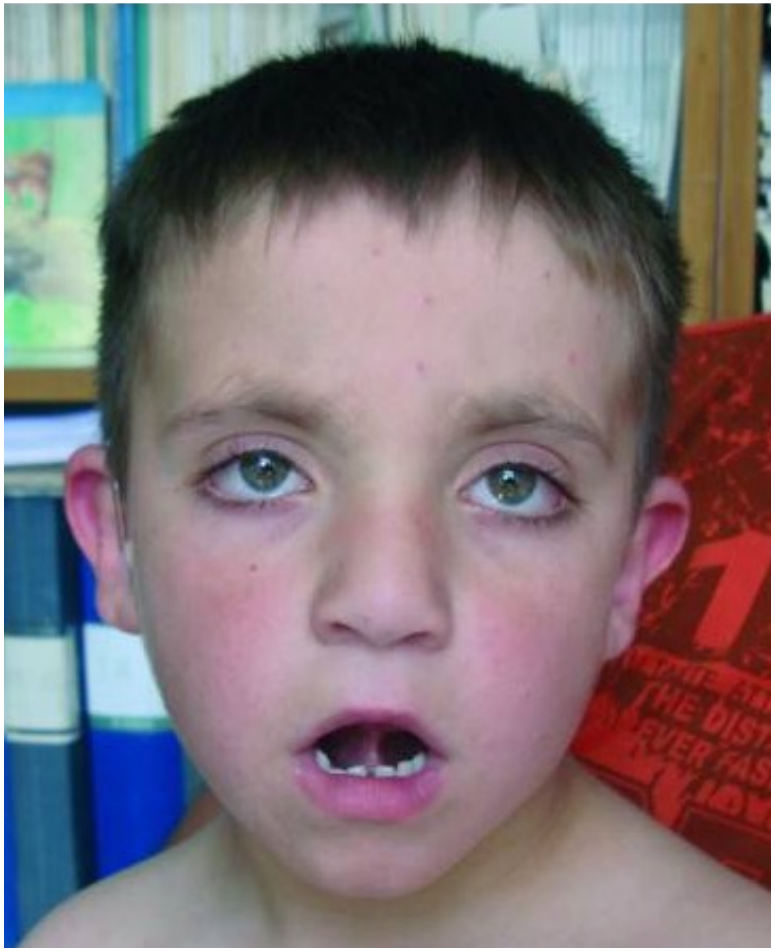
Levels of Control: chromatin remodeling



Lysine methylation by histone methyl transferase:

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

Histone methylation related disease



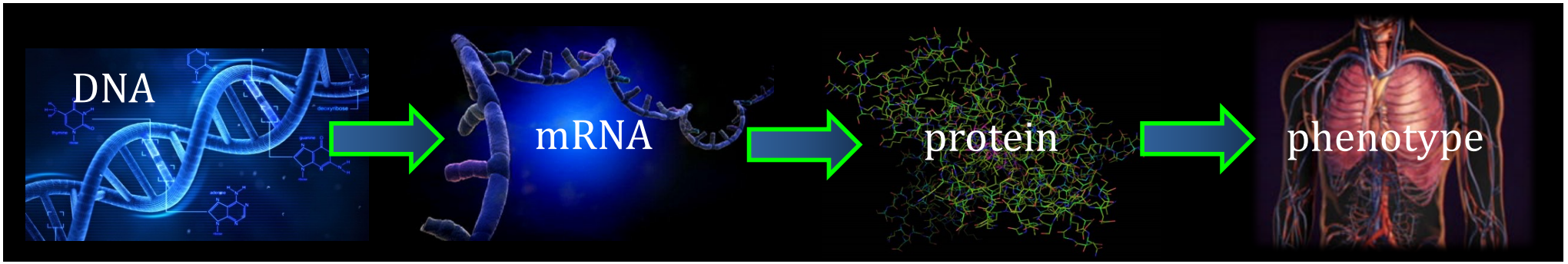
Kabuki syndrome

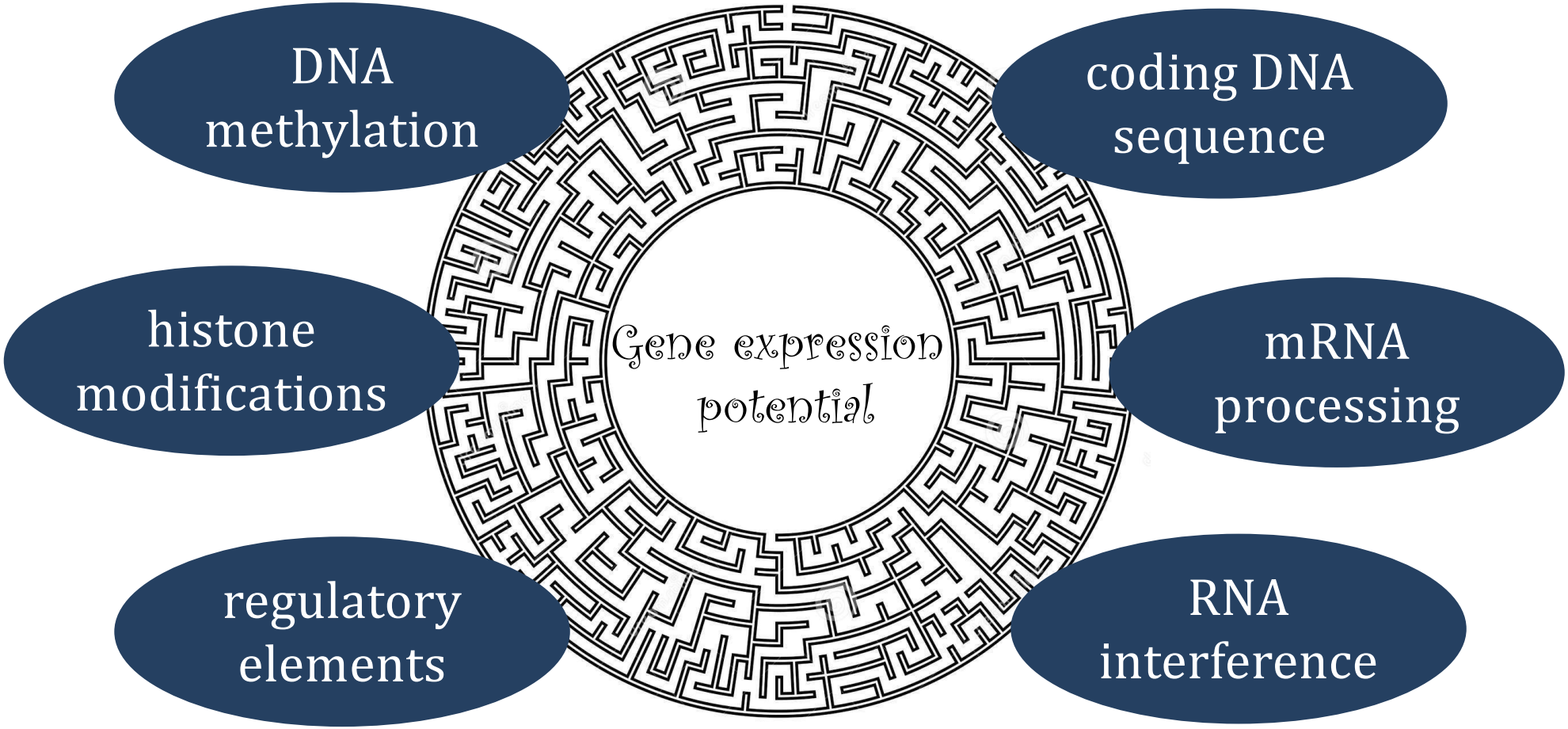
MLL2 (KMT2D) loss-of-function mutations in 50-70% of KS patients

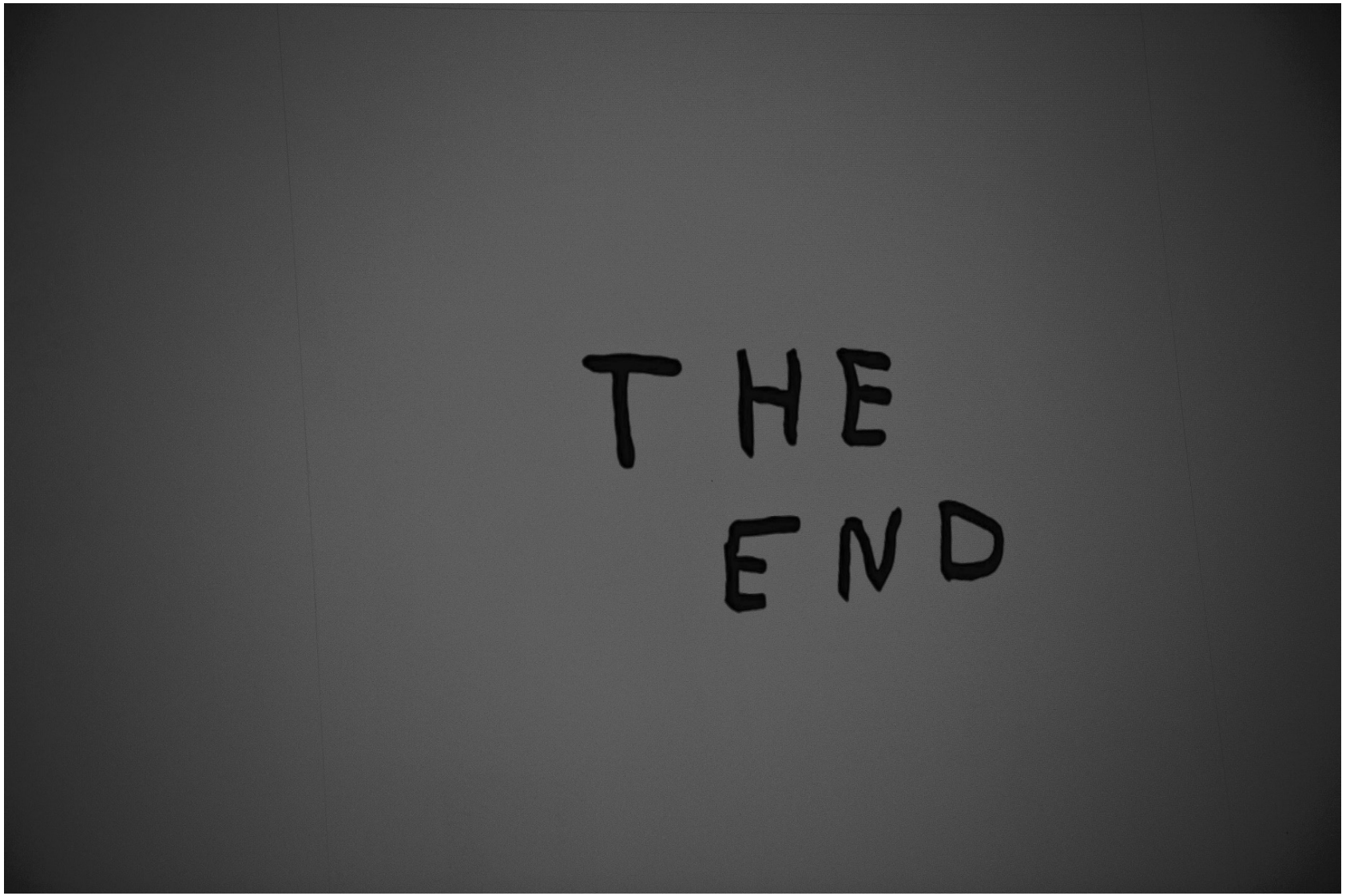
facial gestalt
short stature
microcephaly
feeding problems
oligodontia
high/cleft palate
fetal pads
lax joints
cardiac defects
renal defects
ID
hypotonia
frequent infections

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

Closing remark






THE
END

KATHOLIEKE UNIVERSITEIT
LEUVEN