Genetic Variation II

Frequency of mutations in human disease

Type of mutation	% disease causing mutations
Nucleotide substitutions	
Missense (amino acid substitution)	50%
Nonsense (premature termination codon)	10%
RNA processing (splice, polyadenylation, etc)	20%
Gene expression regulation (TF binding site, etc)	rare
Deletions & insertions	
Small indels	25%
Large rearrangements (deletion, duplication, inversion, etc)	5%
Insertion of LINE or Alu (interrupting regulation or coding)	rare
Repeat expansion	rare

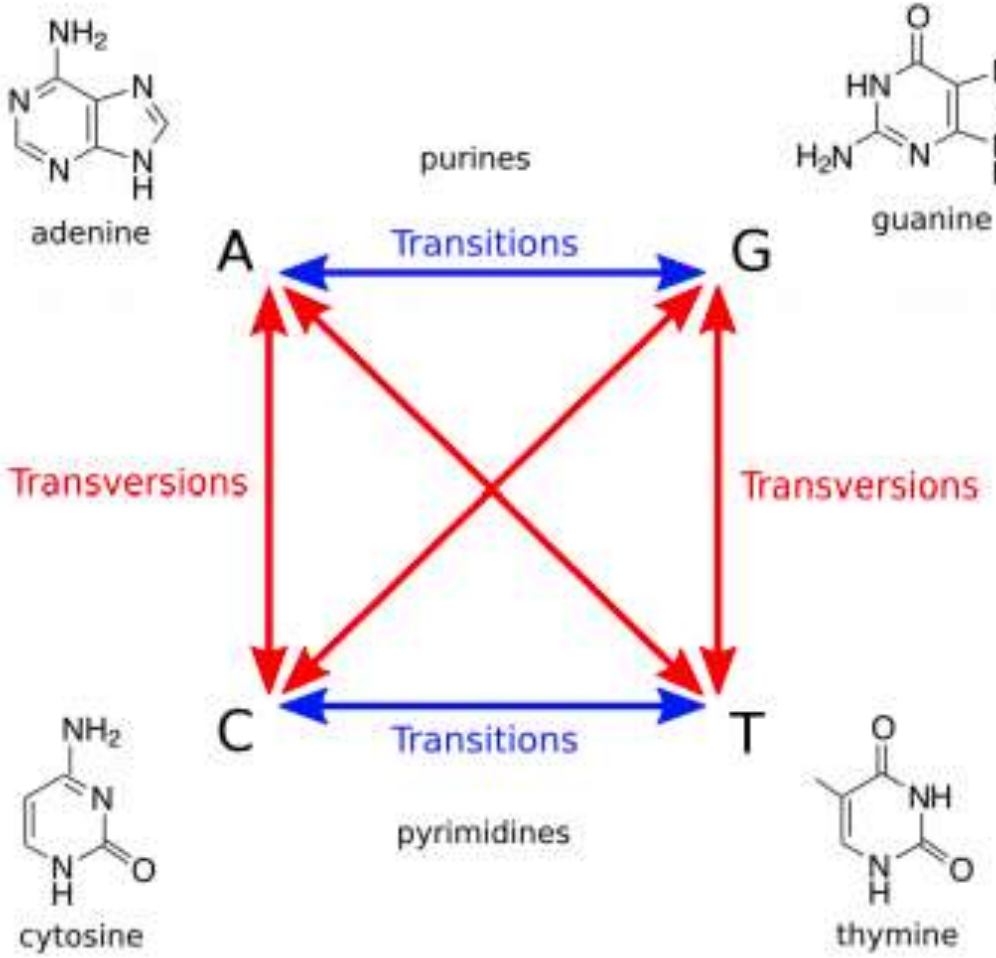
LINE & Alu repeats are non-coding DNA sequences, they can generate copies of themselves that can integrate elsewhere in the genome, if they happen to insert in a clinically important gene, then it might be disease-causing. (nicknamed as the jumping genes)

→ more accurate definition (according to Thompson & Thompson)

Note: These data are changing!

Genetics in Medicine, 8th ed, Thompson&Thompson

Point mutations



Substitutions:

Transition:

purine to purine or

pyrimidine to pyrimidine

i.e: A → G OR C → T

Transversion:

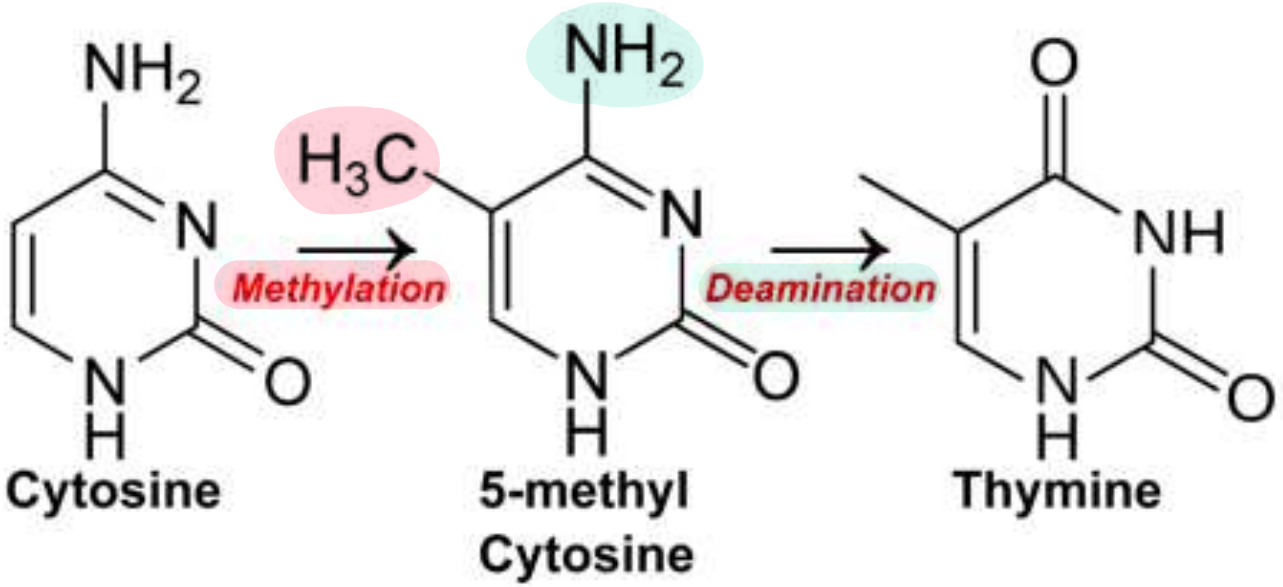
purine to pyrimidine or

pyrimidine to purine

i.e: A → T OR C → G

Most common mutation: C>T transitions

- Most common type of mutation in human genome
- Due to spontaneous deamination of 5-methylcytosine to thymine



Silent (synonymous) mutations

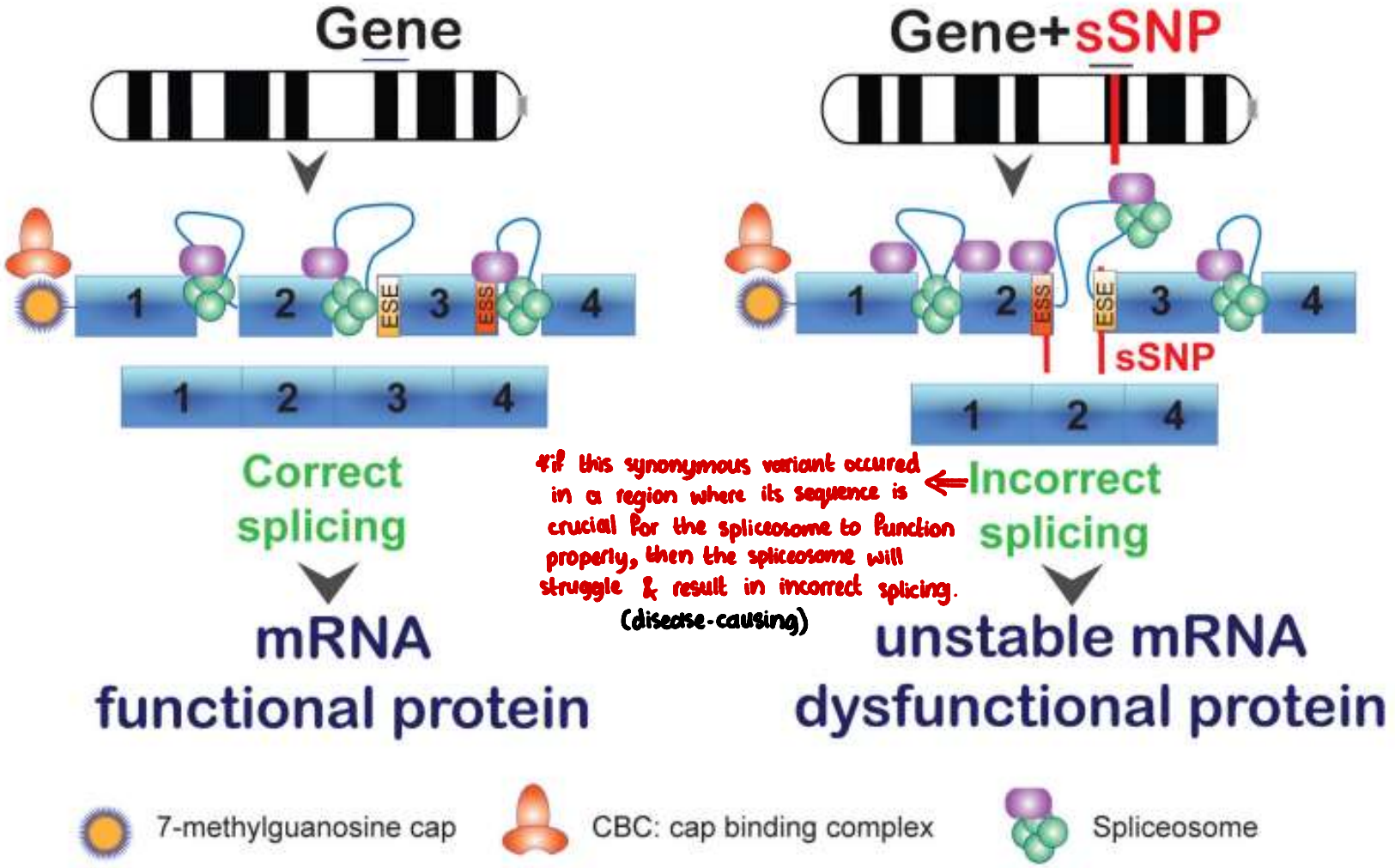
- Do not change the amino acid (p.Ala123Ala)
- Mostly benign, but may impact splicing or RNA secondary structure!

NOT always benign bcz they can have pathogenic implications if they impact splicing OR RNA secondary structure.

mRNA	CAU	CAA	ACG	GGT	GCC	AAC	GGC
Protein	His	Gln	Thr	Gly	Ala	Asn	Gly

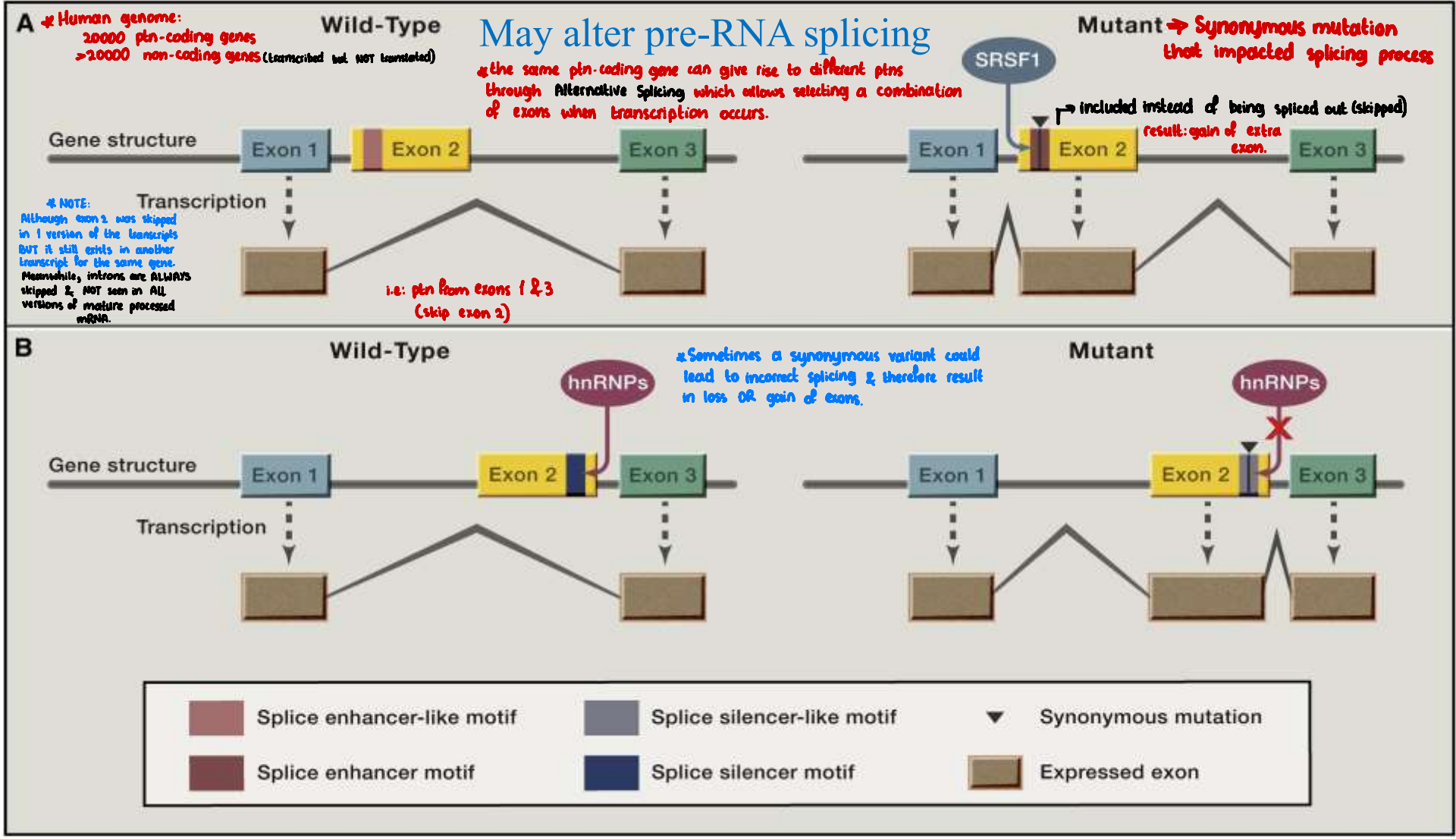
mRNA	CAU	CAA	ACG	GGT	GCU	AAC	GGC
Protein	His	Gln	Thr	Gly	Ala	Asn	Gly

May alter pre-RNA splicing



If this synonymous variant occured in a region where its sequence is crucial for the spliceosome to function properly, then the spliceosome will struggle & result in incorrect splicing. (disease-causing)

Figure 2. The consequence of synonymous mutations in exonic splice regulatory sites
sSNPs may disrupt critical elements necessary for splicing. In the example shown, this results in exon skipping. ESE: exonic splicing enhancer; ESS: exonic splicing suppressor. (For a review concerning pre-mRNA splicing refer to: (Muller-McNicoll and Neugebauer, 2013).



(A) A synonymous mutation leads to the gain of an exonic splicing enhancer motif. Consequently, binding of the splicing regulator SRSF1 is enhanced, resulting in the inclusion of an otherwise skipped exon.

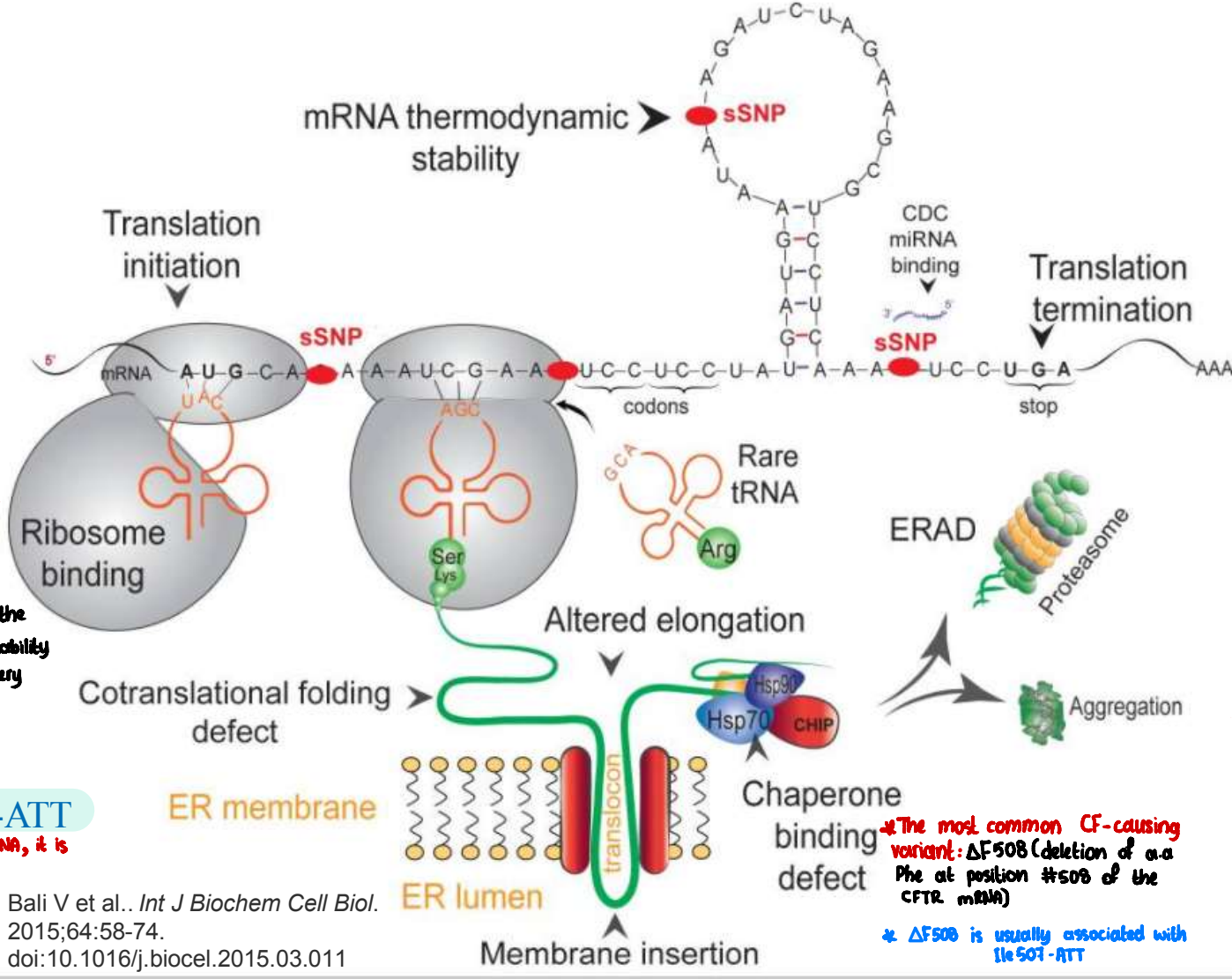
(B) A synonymous mutation deactivates an exonic splicing silencer motif, thereby abolishing the binding of hnRNP splicing regulators

Cell 2014 1561129-1131DOI: (10.1016/j.cell.2014.02.037)

May alter mRNA secondary structure

(details of this slide & figure weren't thoroughly explained except for the handwritten notes, check them just in case)

may alter translation initiation efficiency, translation elongation rate, ribosomal pause rhythm, cotranslational folding or the overall fate of the protein



* Sometimes synonymous variants do NOT impact alternative splicing, rather they impact the thermodynamic stability & mRNA secondary structure.

* Remember: mRNA takes a 3D structure based on the intramolecular H-bonds which play a role in its stability & rates of gene expression & translation machinery

Ile507-ATC and Ile507-ATT

* Ile located at #507 of the CFTR mRNA, it is encoded by either "ATC" or "ATT"

* The most common CF-causing variant: ΔF508 (deletion of a.a Phe at position #508 of the CFTR mRNA)

* ΔF508 is usually associated with Ile507-ATT

Bali V et al.. *Int J Biochem Cell Biol.* 2015;64:58-74. doi:10.1016/j.biocel.2015.03.011

The consequences of a synonymous single nucleotide change on the predicted structure of the mRNA (mfold)

The predicted (mfold) structures of the Ile507-ATC and Ile507-ATT ΔF508 *CFTR* mRNAs. The sequences represent human CFTR mRNA fragments encoding the region of NBD1 near the ΔF508 mutation. The locations of the altered nucleotides (C and U) are highlighted in red.

Missense (Non-synonymous) mutations

- Change the amino acid (substitution)
- Conservative: new amino acid has similar properties as the original (polar to polar, hydrophobic to hydrophobic, etc)
- Non-conservative: new amino acid has different properties than the original (polar to nonpolar, hydrophobic to hydrophilic, etc)
- May be benign or pathogenic

e.g.: Met (hydrophic non-polar) substitution to Arg (hydrophilic, polar positively charged) is expected to be more deleterious than Met substitution to Phe (hydrophic non polar)

Hemoglobin β subunit

hydrophilic ⊖ charged

Example: *HBB* c.17A>T (p.Glu6Val)

hydrophobic

	1	2	3	4	5	6	7	8	9
	Val	His	Leu	Thr	Pro	Glu	Glu	Lys	Ser
NORMAL	GTG	CAT	CTG	ACT	CCT	GAG	GAG	AAG	TCT
SICKLE	GTG	CAT	CTG	ACT	CCT	GTG	GAG	AAG	TCT
	Val	His	Leu	Thr	Pro	Val	Glu	Lys	Ser

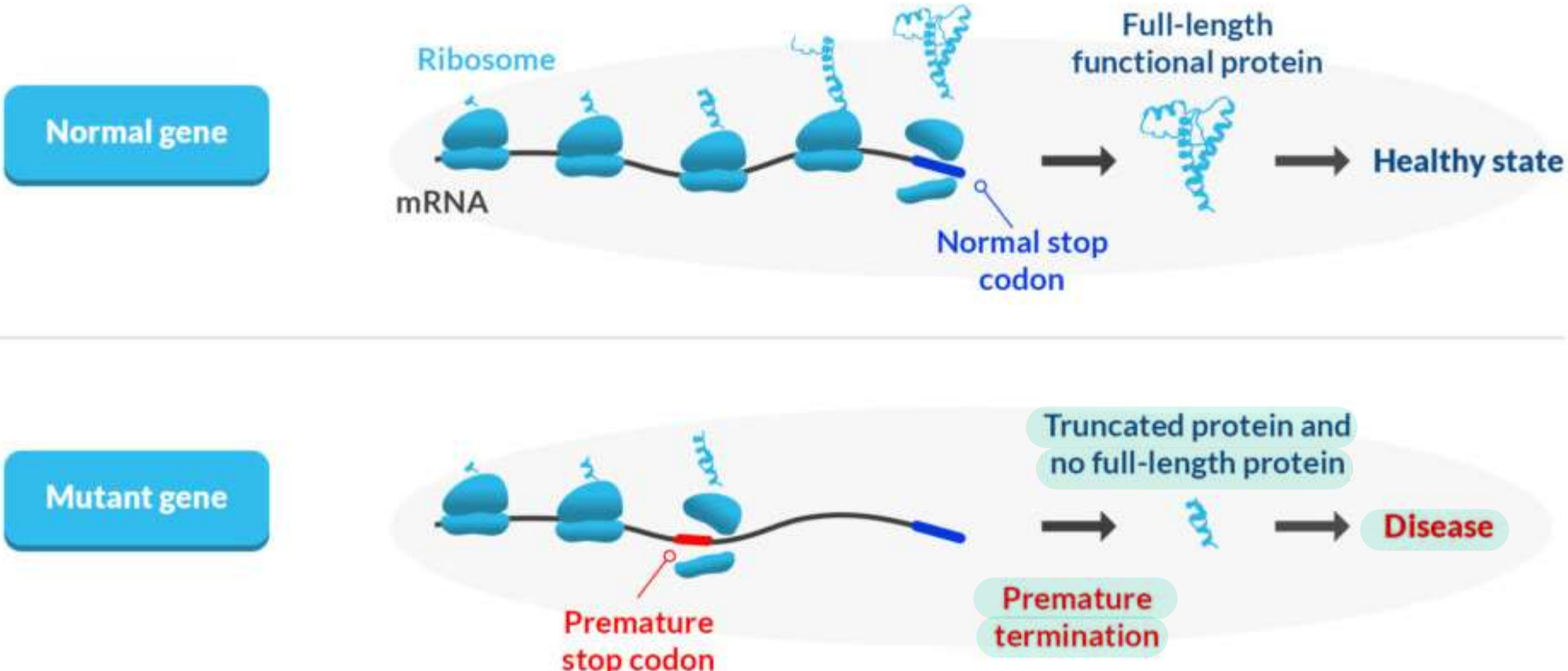
Glutamate change to Valine at position 6 of *HBB* gene encoding β-globin

Leads to β-globin protein aggregates ⟶ Causes Sickle cell anemia

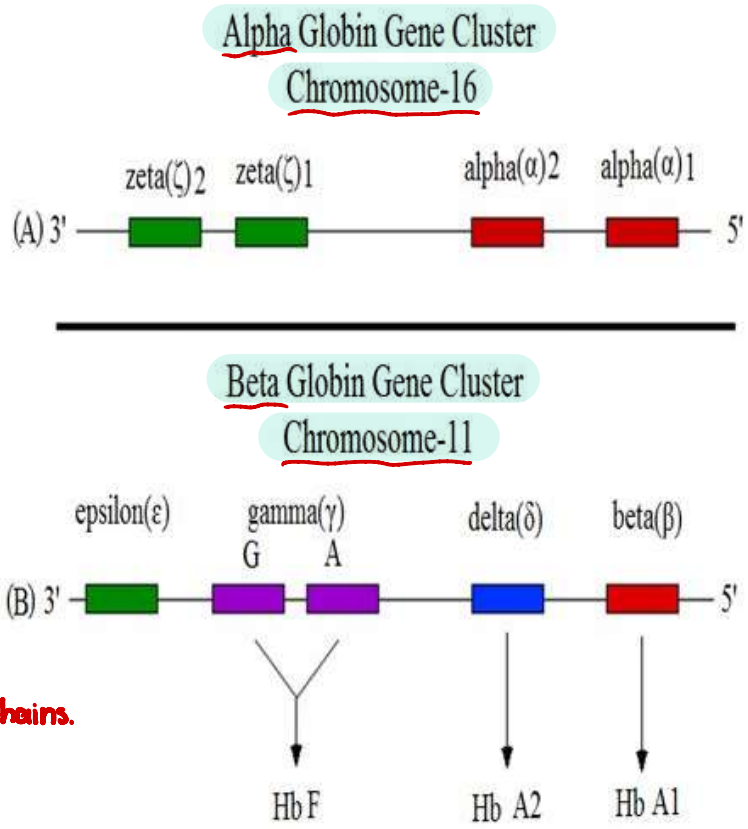
↳ makes HgB sticky

Nonsense mutations

- Cause errors in translation
- Change a codon to a termination codon (UAA, UAG, UGA)
- May result in nonsense mediated decay (NMD), truncated protein, or splicing impact
- Not always pathogenic!



- Type-1: Mutant Alpha(α) globin genes responsible for Alpha(α) thalassemia and
- Type-2: Mutant Beta(β) genes responsible for Beta(β) thalassemia.



※ Remember: HbA is composed of 2 α & 2β globin chains.

Example: *HBB* c.118C>T (p.Gln39*)

	31	32	33	34	35	36	37	38	39
	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln
NORMAL	CTG	CTG	GTG	GTC	TAC	CCT	TGG	ACC	CAG
β^0	CTG	CTG	GTG	GTC	TAC	CCT	TGG	ACC	TAG
	Leu	Leu	Val	Val	Pro	Pro	Trp	Thr	STOP

Creates premature termination codon and leads to NMD

Homozygotes: No β-globin protein ⟶ β-thalassemia

Frameshift mutations

- Cause errors in translation
- Alters the mRNA reading frame
- Often lead to a premature termination codon downstream
- Not always pathogenic!

deletion of G nucleotide at #35 → leads to frameshift

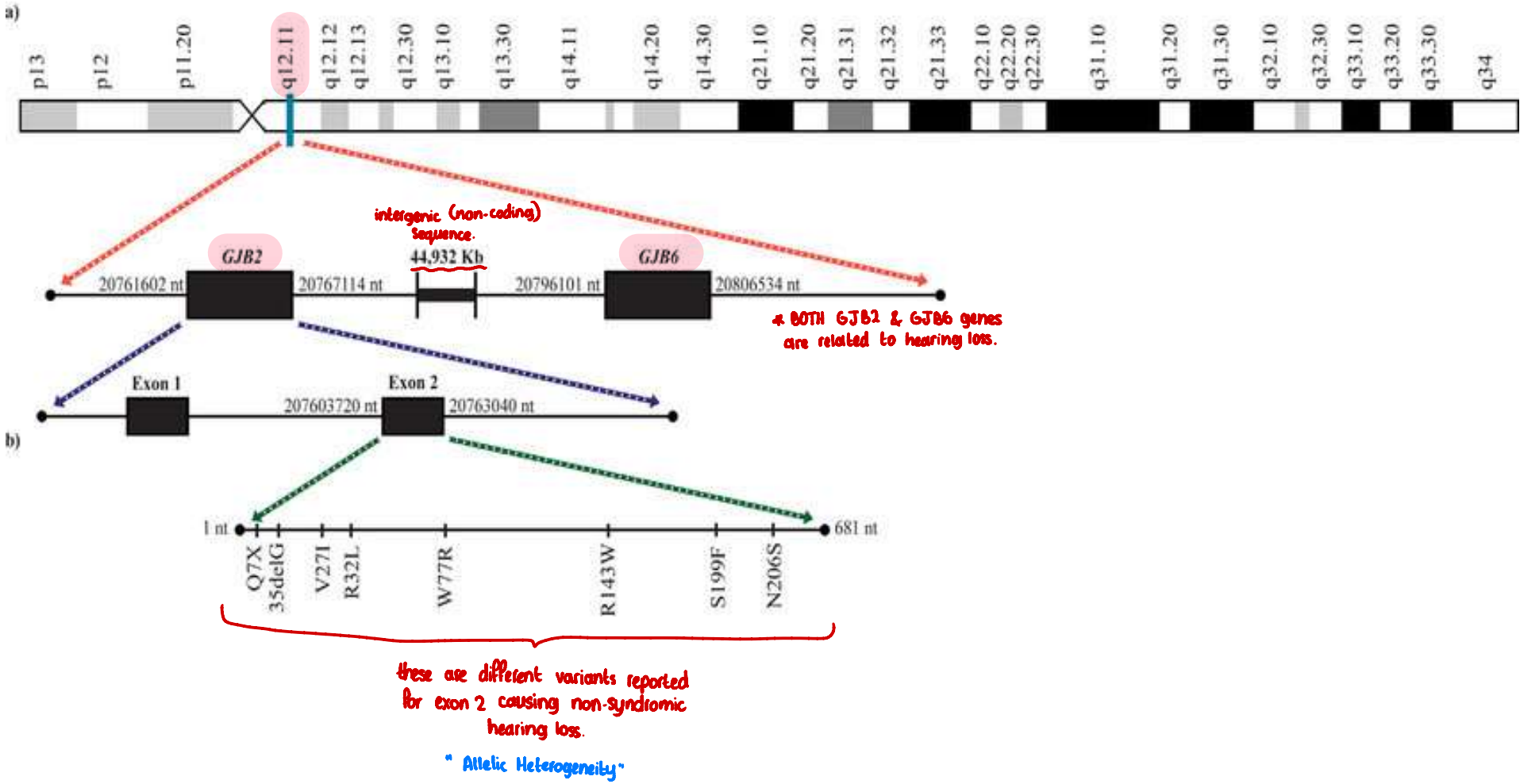
Example: *GJB2* c.35delG (p.Gly12fs)

most common mutated gene causing non-syndromic recessive hearing loss

Changes Glycine at position 12 to a Valine and leads to premature termination codon downstream

Homozygotes: Non-syndromic hearing loss

Chromosome 13

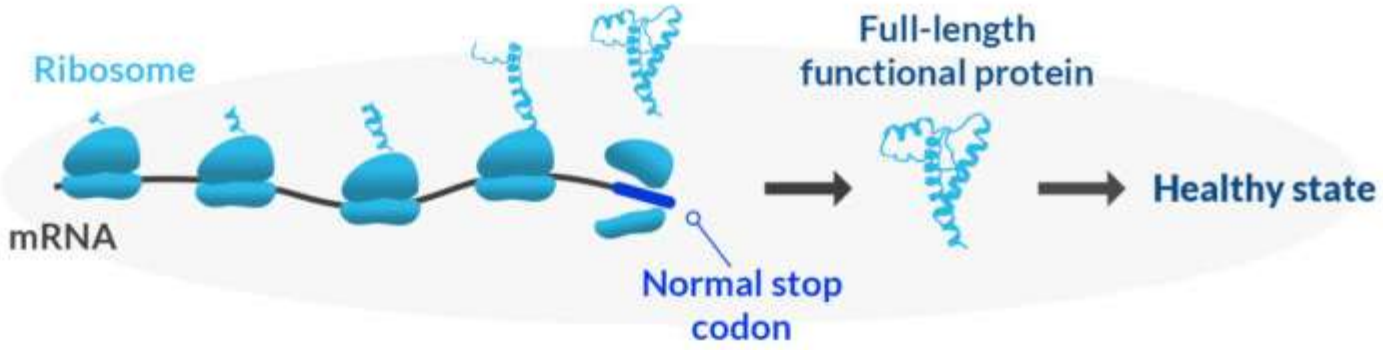


Nonsense mutations

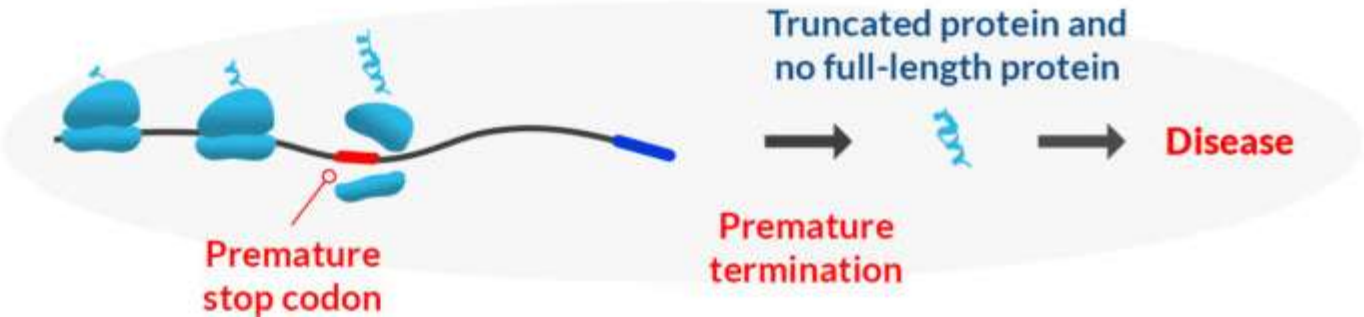
- Cause errors in translation
- Change a codon to a termination codon (UAA, UAG, UGA)
- May result in nonsense mediated decay (NMD), truncated protein, or splicing impact
- Not always pathogenic!

→ premature stop codon results in → truncated ptn → any sequence downstream to the premature stop codon is lost during translation.
OR
→ nonsense mediated decay → abnormal mRNA is degraded even before translation takes place.

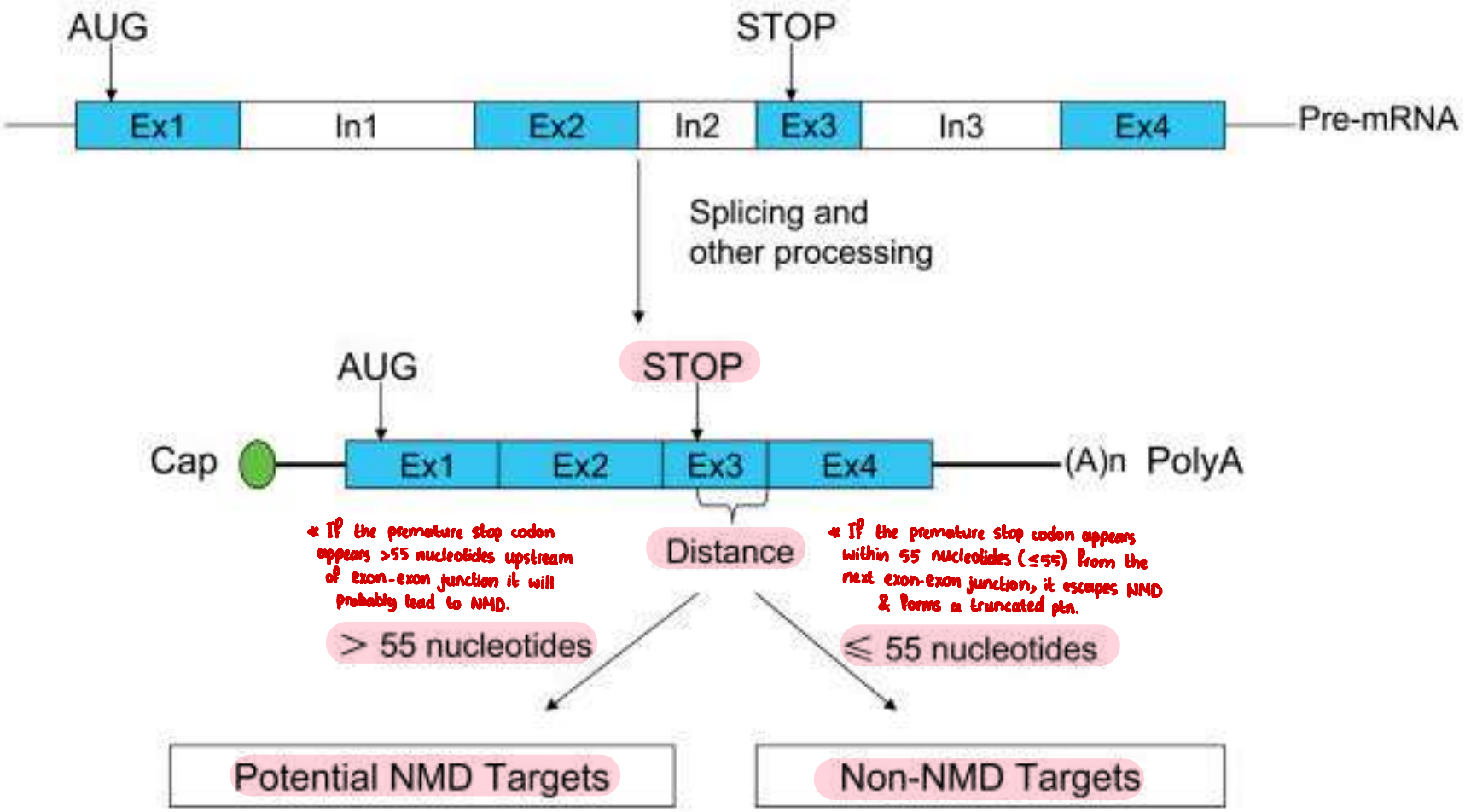
Normal gene



Mutant gene

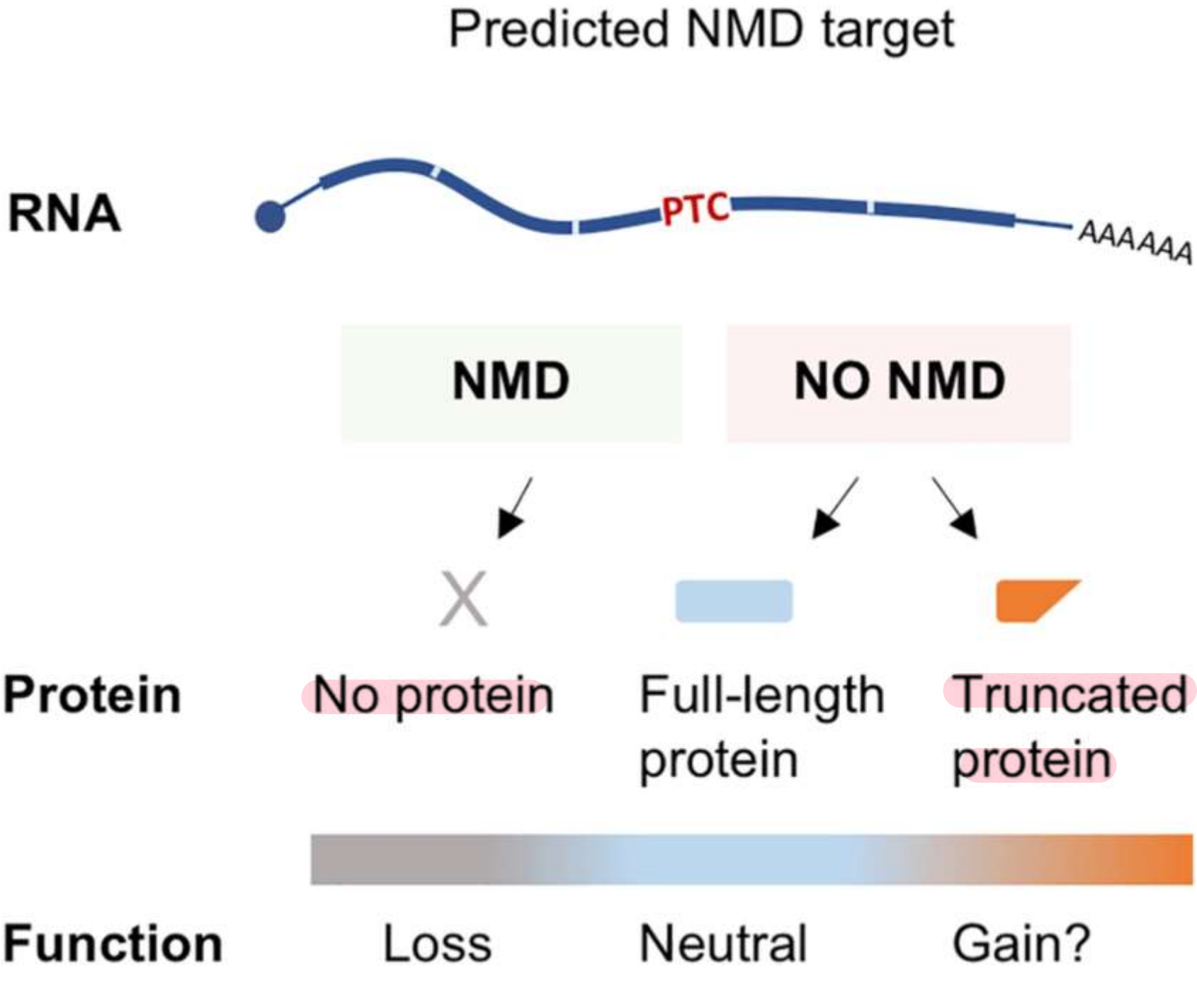


nonsense-mediated mRNA decay (NMD)



Ex*: Exons **In***: Introns **AUG**: start codon **STOP**: termination codon

50 to 55 nucleotides upstream of the 3' most splice-generated exon-exon junction


Predicted NMD target
RNA
PTC
AAAAAA
NMD
NO NMD
Protein
No protein
Full-length protein
Truncated protein
Function
Loss
Neutral
Gain?

In-frame deletions and insertions

- Deletions or insertions of bases in multiples of 3 (3,6,9,…)
- Lead to deletions or insertions of amino acids without altering the reading frame ⇒ a.a sequence is NOT impacted.
- May be benign or pathogenic

Example: CFTR c. (p.Phe508del – ΔF508)

entire codon deletion resulting in loss of Phe at #508

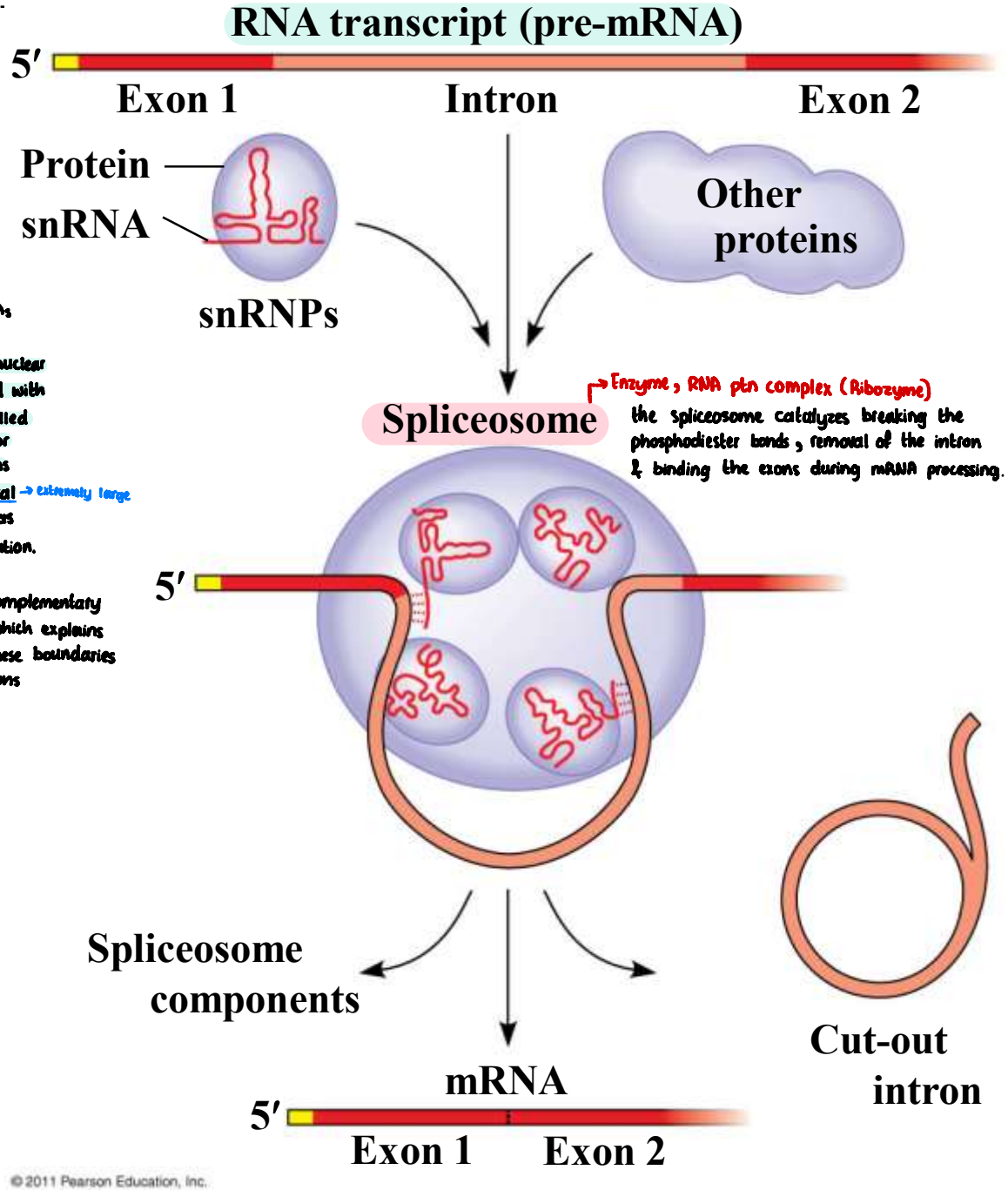
↳ the most common variant causing CF.

Normal	ATC	ATC	TTT	GGT	GTT
	Ile	Ile	**Phe**	Gly	Val
ΔF508	ATC	ATT	GGT	GTT	
	Ile	Ile	Gly	Val	

Leads to deletion of phenylalanine at position 508 of CFTR protein

Block in processing of the protein ⟶ Cystic fibrosis

Figure 17.12



* Remember:

① the ONLY RNA translated to ptn is mRNA BUT there are other RNAs not translated to ptns but have crucial functions such as small nuclear RNA (snRNA) which is then mixed with ptns & gives rise to structures called "small nuclear ribonucleoproteins" or "snRNP", even mixed with other ptns & other snRNPs to form a colossal → extremely large enzyme (Spliceosome) which is as big as RNA responsible for translation.

② the sequence of the snRNA is complementary to the exon-intron boundaries, which explains how the spliceosome recognizes these boundaries & splices out the introns

Figure 1.16 The process of RNA splicing

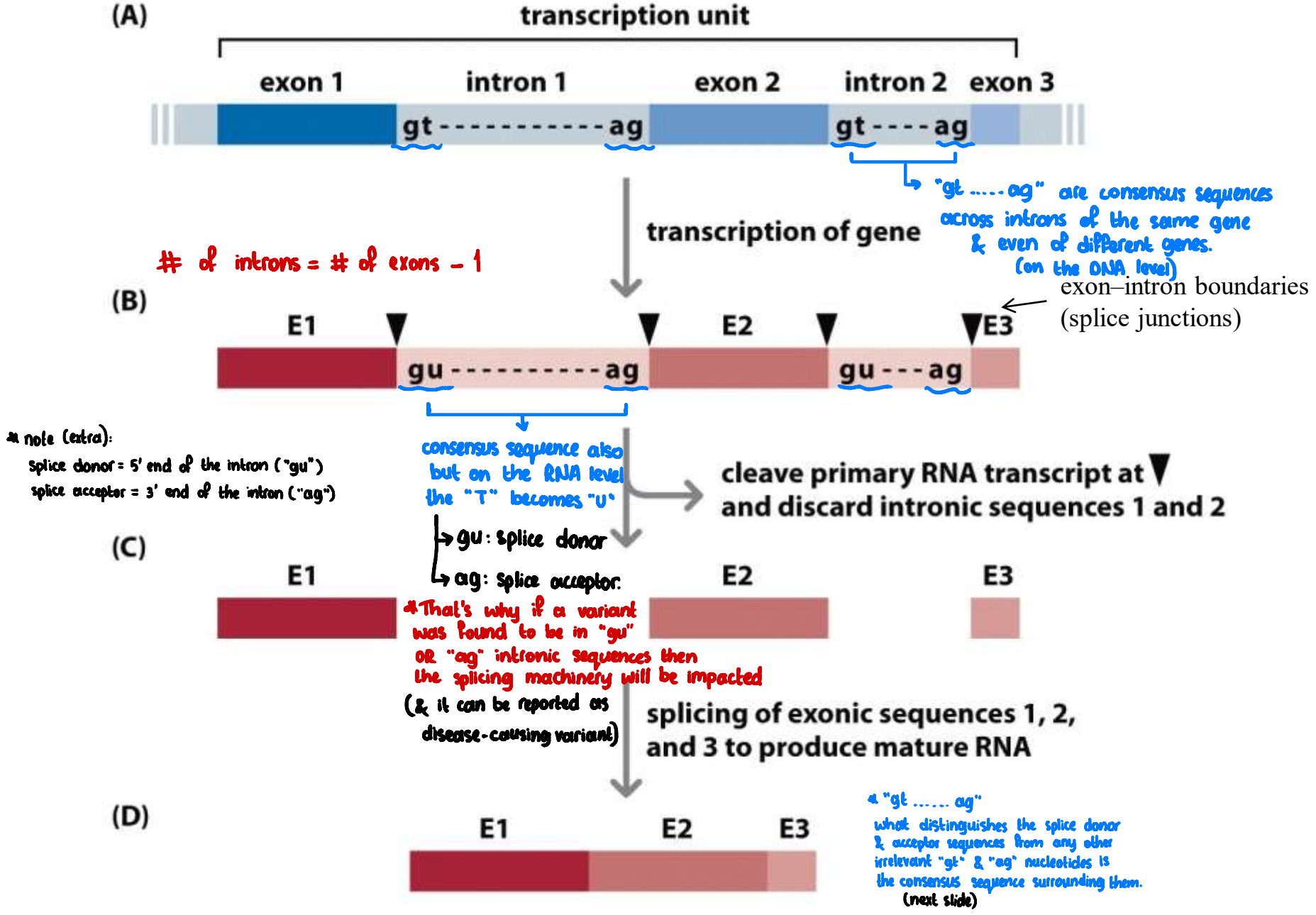


Figure 1.16 Human Molecular Genetics, 4ed. (© Garland Science)

Fig 1.17 3 splice junction **consensus DNA sequences** in introns of complex eukaryotes

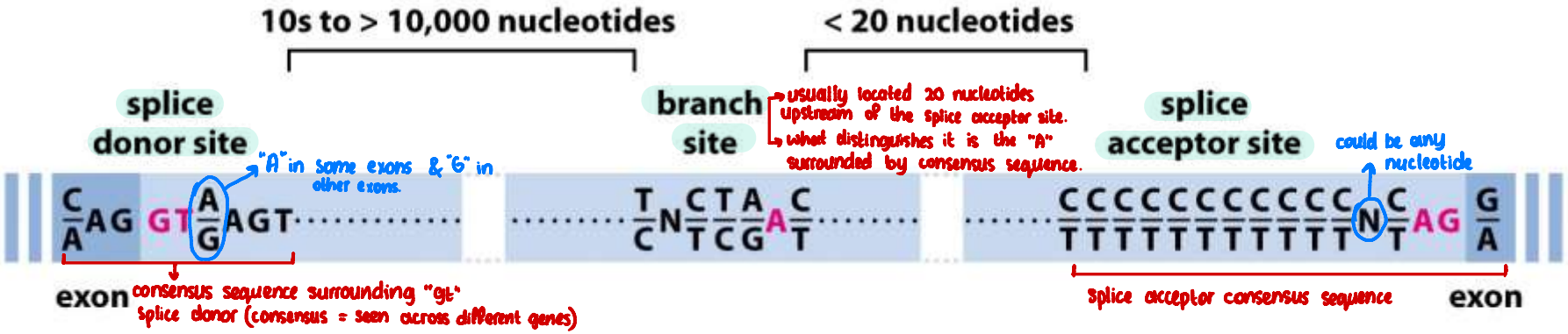


Figure 1.17 Human Molecular Genetics, 4ed. (© Garland Science)

Most introns in eukaryotic genes contain **conserved sequences** that correspond to three functionally important regions:

Two of the regions, the **splice donor** site and the **splice acceptor** site, span the 5′ and 3′ boundaries of the intron

The **branch site** is an additional important region that typically occurs less than **20 nts** upstream of the splice acceptor site

The nucleotides shown in red in these three consensus sequences are almost **invariant**.
The other nucleotides detailed in both the intron and the exons are those **most commonly** found at each position.

In some instances, two nucleotides may be **equally common**, as in the case of **C and T** near the 3′ end of the intron. Where N appears, any of the four nucleotides may occur.

Figure 1.18 The mechanism of RNA splicing

(A) The unprocessed primary RNA transcript with intronic RNA separating sequences E1 and E2 that correspond to exons in DNA

(B) The splicing mechanism involves a **nucleophilic attack** on the **G of the 5′ GU** dinucleotide. This is carried out by the **2′ OH** group on the conserved **A of the branch** site and results in the formation of a **lariat** structure and **cleavage of the splice donor** site

similar to a loop structure.

(C) The **3′ OH** at the 3′ end of the **E1** sequence performs a **nucleophilic attack** on the **splice acceptor** site, causing release of the intronic RNA (as a lariat-shaped structure) and fusion (splicing) of E1 and E2.

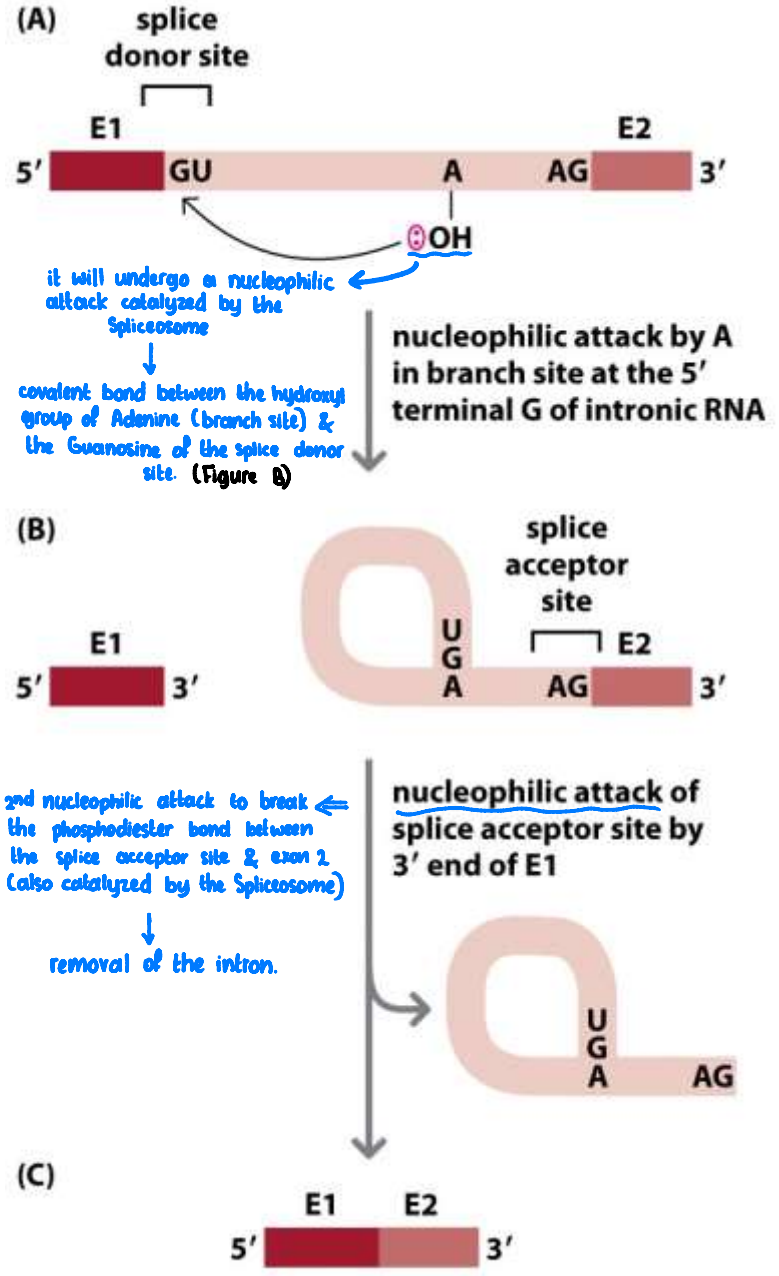


Figure 1.18 Human Molecular Genetics, 4ed. (© Garland Science)

Fig 1.19 Role of small nuclear ribonucleoprotein (snRNPs) in RNA splicing

A) The unprocessed primary RNA transcript

B) Within the spliceosome, part of the **U1** snRNA is **complementary** in sequence to the **splice donor site consensus sequence**. As a result, the U1 snRNA-protein complex (U1 snRNP) binds to the splice junction by **RNA–RNA base pairing**. The **U2** snRNP complex similarly binds to the **branch site** by RNA–RNA base pairing.

C) Interaction between the splice donor and splice acceptor sites is **stabilized** by the binding of a **multi-snRNP** particle that contains the **U4, U5, and U6** snRNAs.

- The **U5** snRNP binds simultaneously to both the splice donor and splice acceptor sites.

- Their cleavage releases the intronic sequence and allows (D) E1 and E2 to be spliced together.

* Since the Spliceosome is a complex, it is composed of subunits: U1, U2, U4, U5, U6.
U1 binds to the splice donor site, U2 binds to the branch site. Once they bind, they recruit other subunits to catalyze the nucleophilic attack to break the phosphodiester bonds in the boundaries between the exon & the intron.

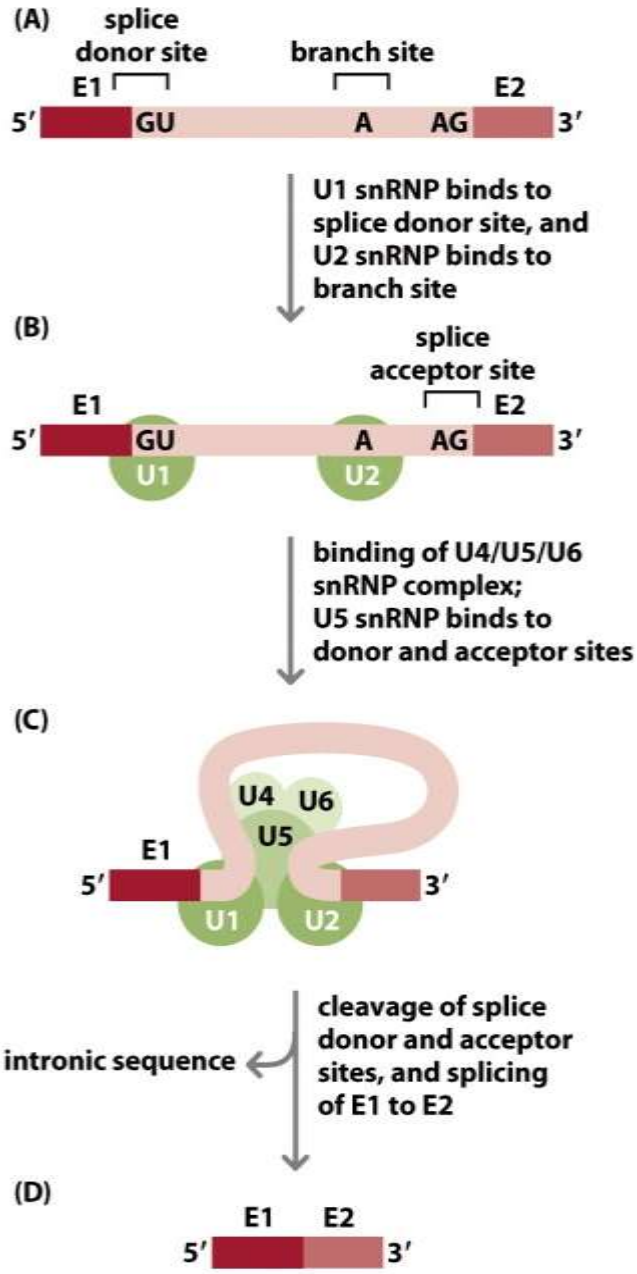


Figure 1.19 Human Molecular Genetics, 4ed. (© Garland Science)

Splice mutations

- Variants that likely impact splicing : +1,2 = donor (gt) / -1,2 = acceptor (ag) } if they undergo a sequence change then it is a red flag.

  Splice donor & acceptor positions (+/- 1,2) → destruction of 5'/3' splice consensus sequence, typically leads to exon skipping

- Variants that may impact splicing:

  Other positions in splice consensus sequence (+/- 15)

  Variants affecting 1st and last 3 bases of an exon

- Other point mutations also have potential to impact splicing