# Genetic Variation II

# Frequency of mutations in human disease

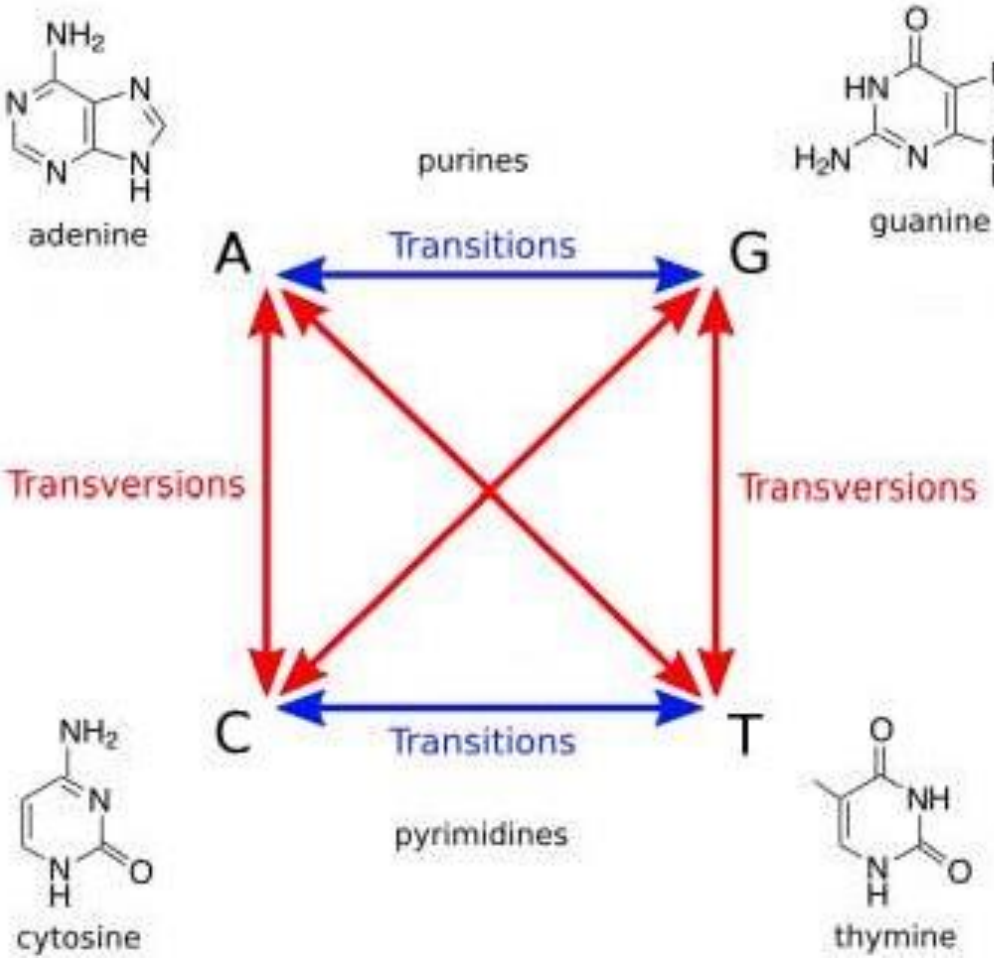
| Type of mutation | % disease causing mutations |
|---|---|
| **Nucleotide substitutions** | |
| Missense (amino acid substitution) | 50% |
| Nonsense (premature termination codon) | 10% |
| RNA processing (splice, polyadenylation, etc) | 20% |
| Gene expression regulation (TF binding site, etc) | rare |
| **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 |

Note: These data are changing!

Genetics in Medicine, 8th ed, Thompson&Thompson

*These are non-coding DNA sequences present in the genome that can move from one location to another ("jumping genes") through cut-and-paste mechanisms. If they insert themselves into a coding region or a clinically relevant gene, they can disrupt its function and potentially cause disease.

# Point mutations



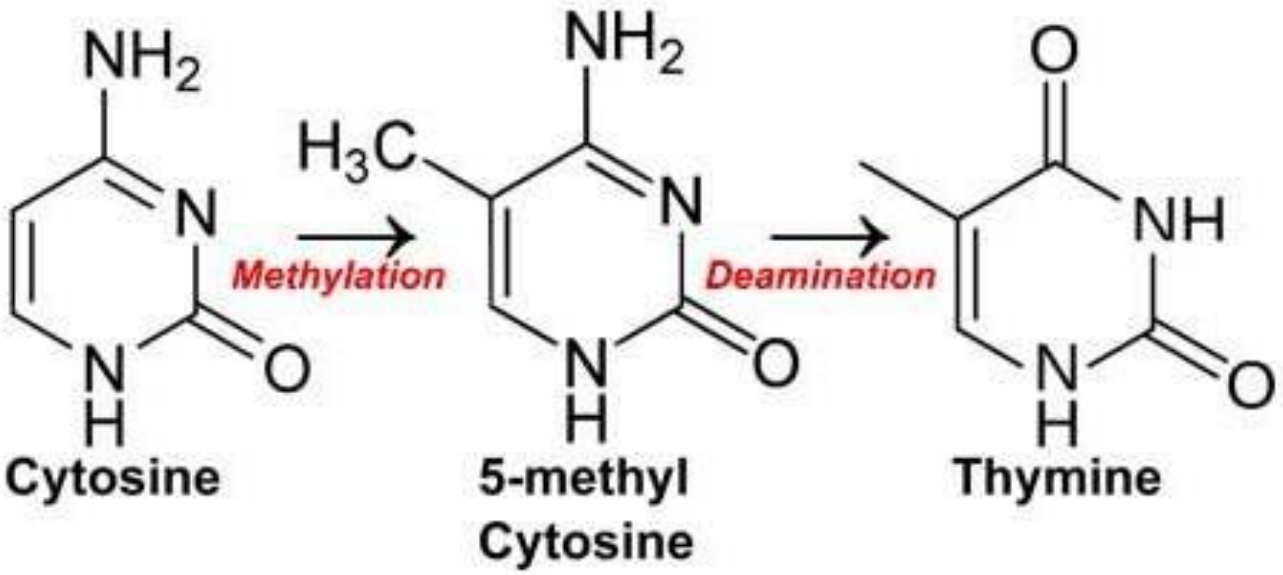

**Transition:** Within the same group
purine to purine or
pyrimidine to pyrimidine

**Transversion:**
purine to pyrimidine or
pyrimidine to purine

# Most common mutation: C>T transitions

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

Removal of NH2



Different codon → Same Amino Acid

# Silent (synonymous) mutations

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

Because the amino acid sequence is not altered, and therefore the protein structure is not impacted.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 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

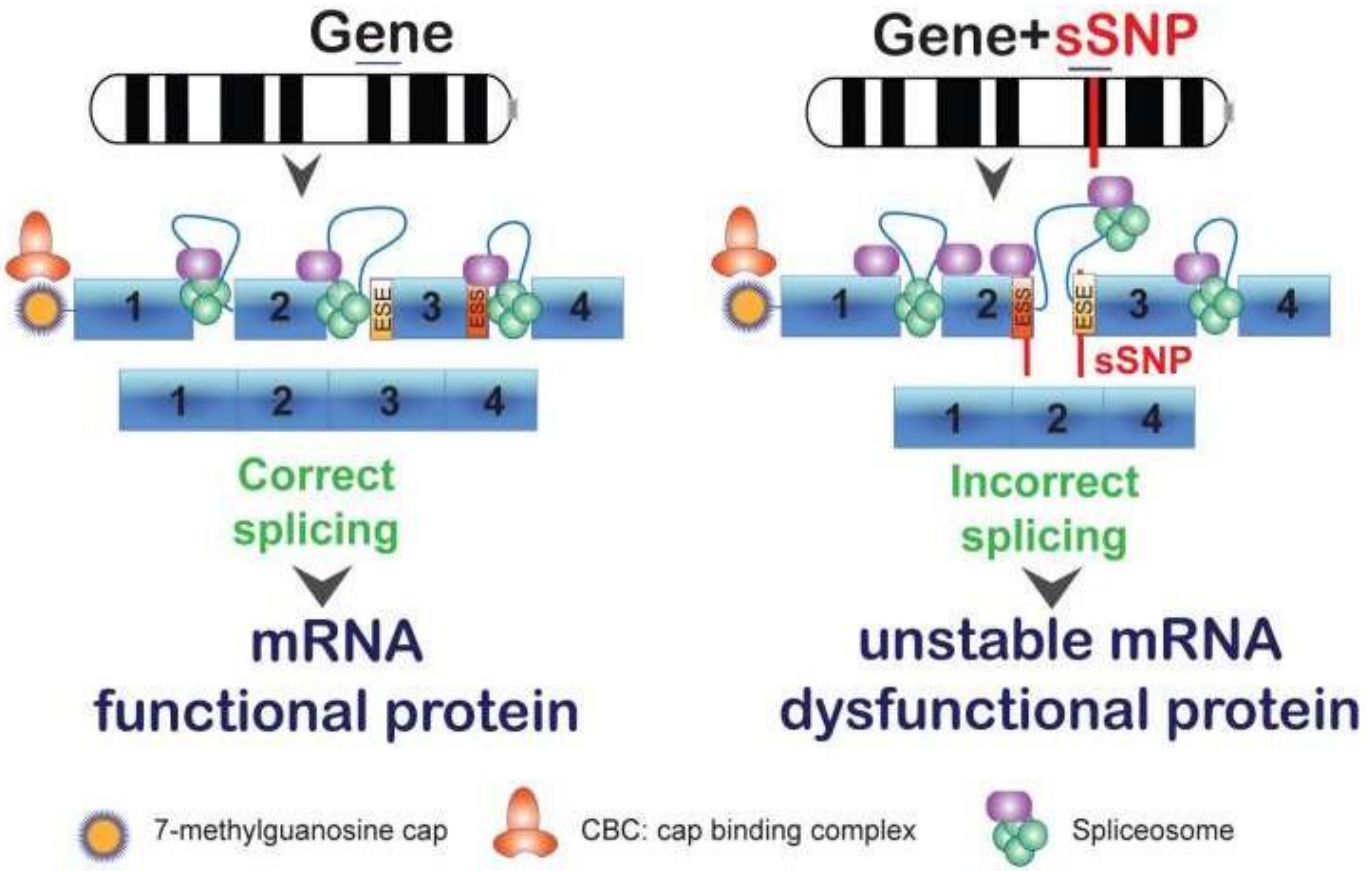
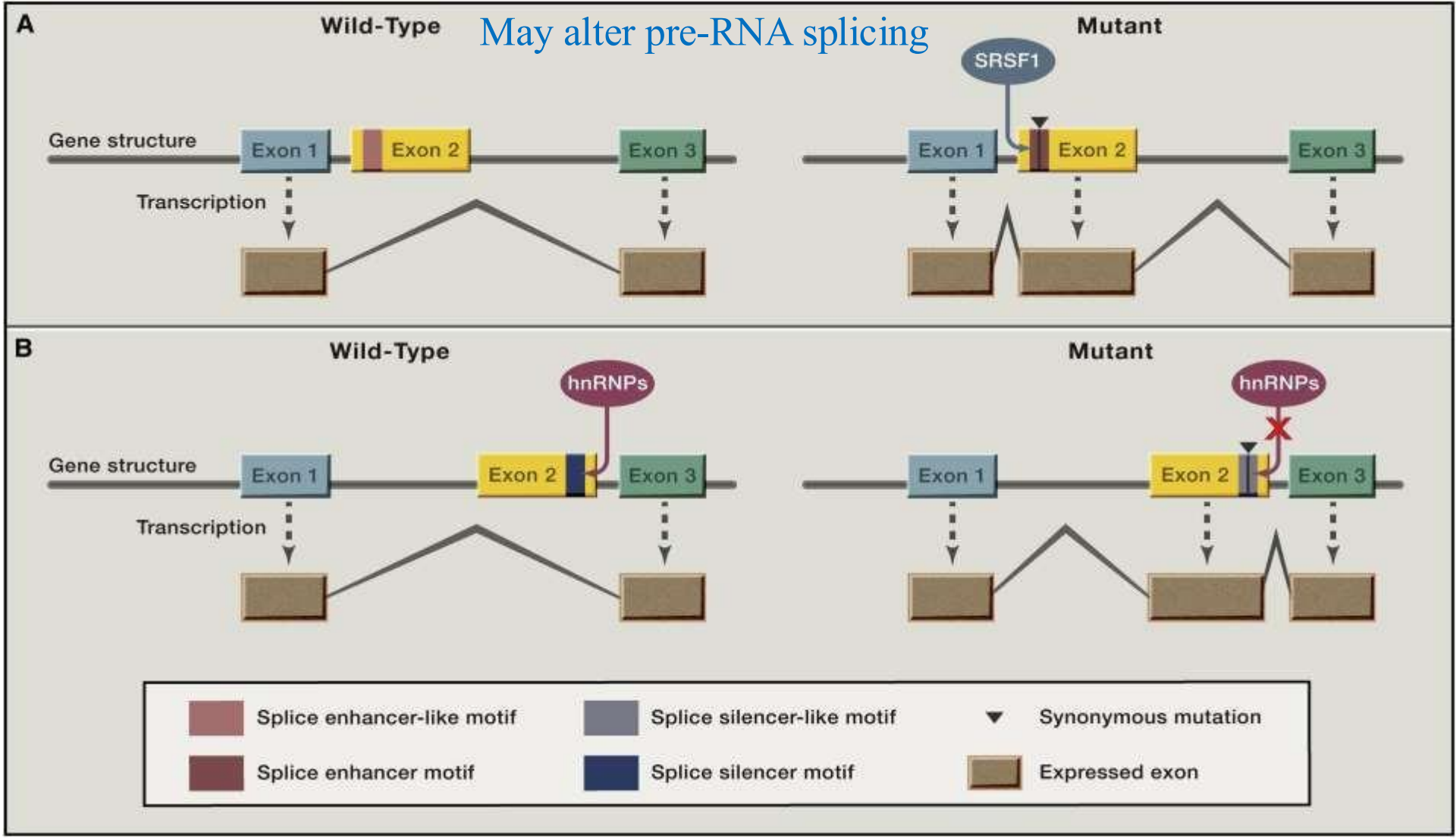



**Figure 2. The consequence of synonymous mutations in exonic splice regulatory sites**
sSNPs may disrupt critical elements necessary for splicing. In the example shown, this results in exon skipping. ESE: exonic splicing enhancer; ESS: exonic splicing suppressor. (For a review concerning pre-mRNA splicing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

- If a synonymous variant occurs in a region where the spliceosome requires a specific sequence to function properly, it may interfere with normal splicing. As a result, the spliceosome may perform incorrect splicing, meaning that even synonymous variants can sometimes be disease-causing.




(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)

*See the next slide*

- **Alternative splicing** allows the same gene to produce different proteins by selecting different combinations of exons during the transcription. Although humans have about 20,000 protein-coding genes, they can produce a much larger number of proteins through alternative splicing.

- For example, a gene may consist of three exons. One mRNA may contain exons 1, 2, and 3, producing one protein. Another mRNA from the same gene may contain only exons 1 and 3, with exon 2 skipped, producing a **different but related (similar) protein**.

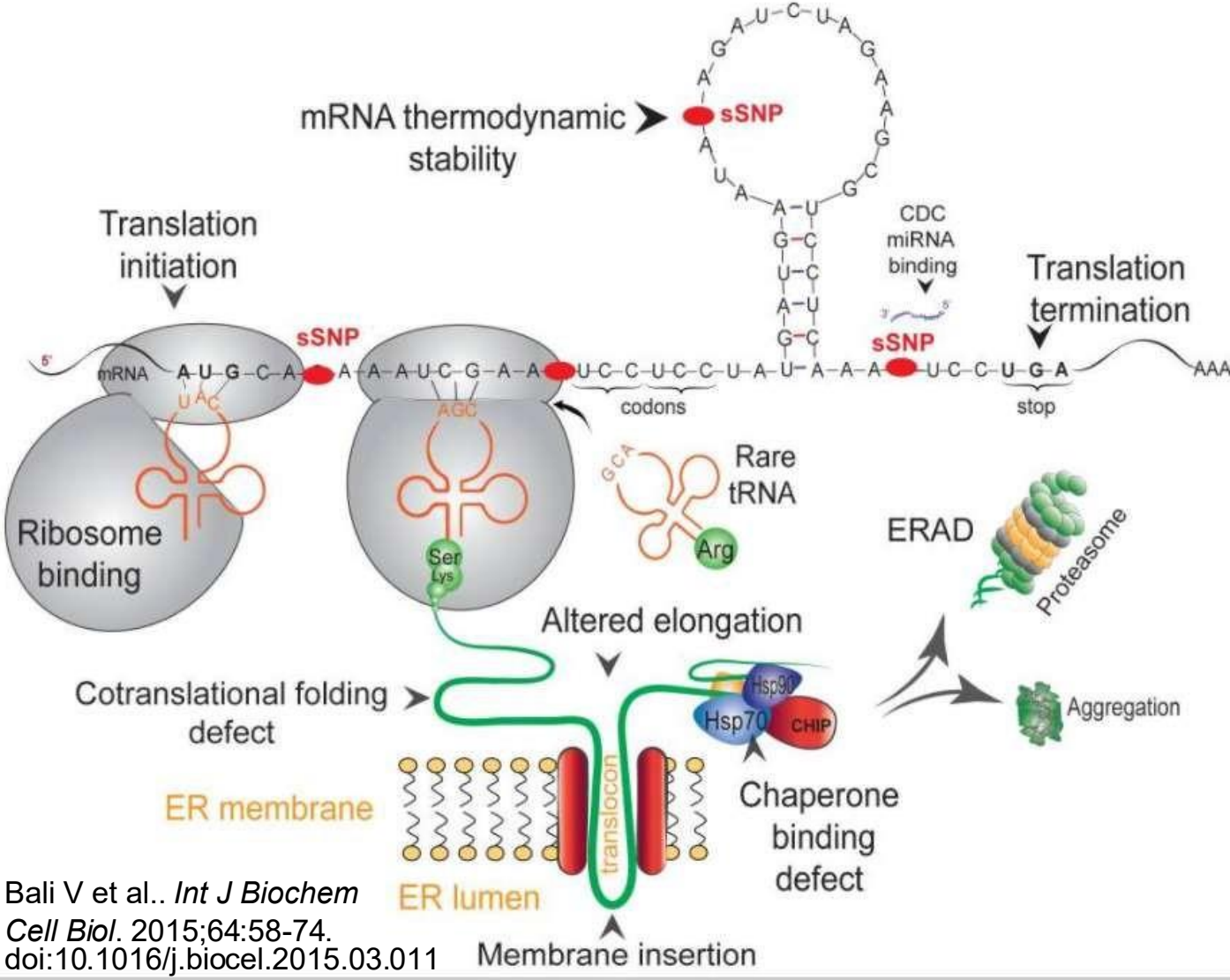
- It is important to note that ***exon*** skipping in alternative splicing does NOT mean that the exon is permanently removed from the gene. An exon may **be present in one transcript and absent in another transcript** from the same gene. In contrast, ***introns*** are always **removed during mRNA processing** and are not present in mature mRNA.

- In this scenario, a silent (synonymous) mutation has affected the splicing process. *Although the amino acid sequence encoded by the altered codon remains unchanged, the variant disrupts the normal function of the spliceosome*. As a result, exon 2, which would normally be spliced out, is incorrectly retained in the mature mRNA. This illustrates how a synonymous variant can still be **disease-causing by affecting splicing rather than altering the amino acid sequence.**

- The opposite can also occur: an exon that should normally be included may be skipped and lost from the mature mRNA. In either case, the splicing machinery is disrupted, leading to abnormal mRNA processing and potentially abnormal protein function.

# May alter mRNA secondary structure

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




Ile507-ATC and 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.

*See the next slide*

- Sometimes a synonymous variant does not affect splicing. Instead, it may influence the **thermodynamic stability or secondary structure of the mRNA**. Remember that mRNA has a three-dimensional structure maintained by hydrogen bonds, and changes in this structure can affect mRNA stability, translation efficiency, and overall gene expression levels. This can ultimately contribute to disease development by altering the amount or efficiency of protein production.

- For example, in **cystic fibrosis**, mutations in the **CFTR** gene can cause disease. One of the most common pathogenic variants is the **ΔF508** deletion, which involves the deletion of phenylalanine at position 508 of the CFTR protein. At amino acid position 507, the codon can still encode the same amino acid (for example isoleucine encoded by ATT or ATC). Actually, ΔF508 is associated with isoleucine codons such as ATT. This is showing you that even a synonymous mutation could be associated with disease. It can impact disease by affecting mRNA thermodynamic stability, RNA structure, or gene expression.

# 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

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

| | 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

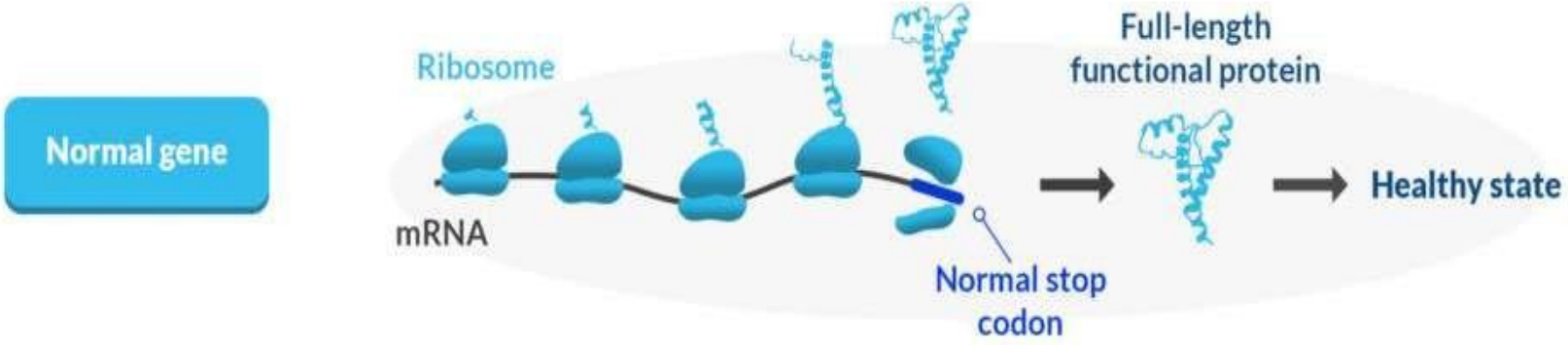
- If methionine is replaced by arginine, arginine is positively charged and hydrophilic while methionine is hydrophobic and non-polar. this is a non-conservative substitution and is predicted to be more deleterious. while replacing methionine with phenylalanine is more conservative because both are hydrophobic, so the effect is usually less severe, as non-conservative changes are more likely to affect original protein properties.

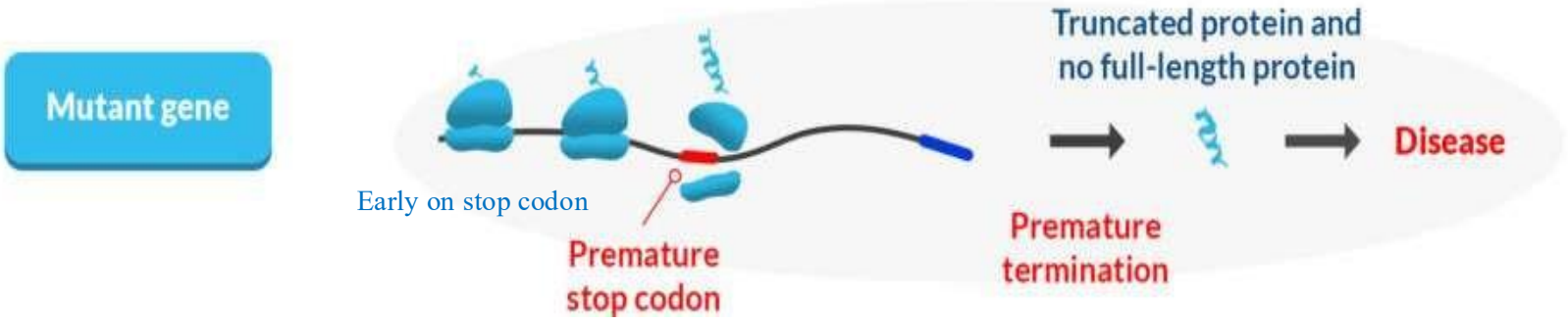
- In hemoglobin beta (HBB), there is a disease-causing mutation at c.17 A>T, resulting in a change from glutamic acid to valine. Glutamic acid is acidic and negatively charged, while valine is hydrophobic and non-polar. This is a non-conservative substitution. As a result, hemoglobin beta molecules polymerize under low oxygen conditions, leading to sticky sickle-shaped red blood cells seen in sickle cell disease.

# Nonsense mutations

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

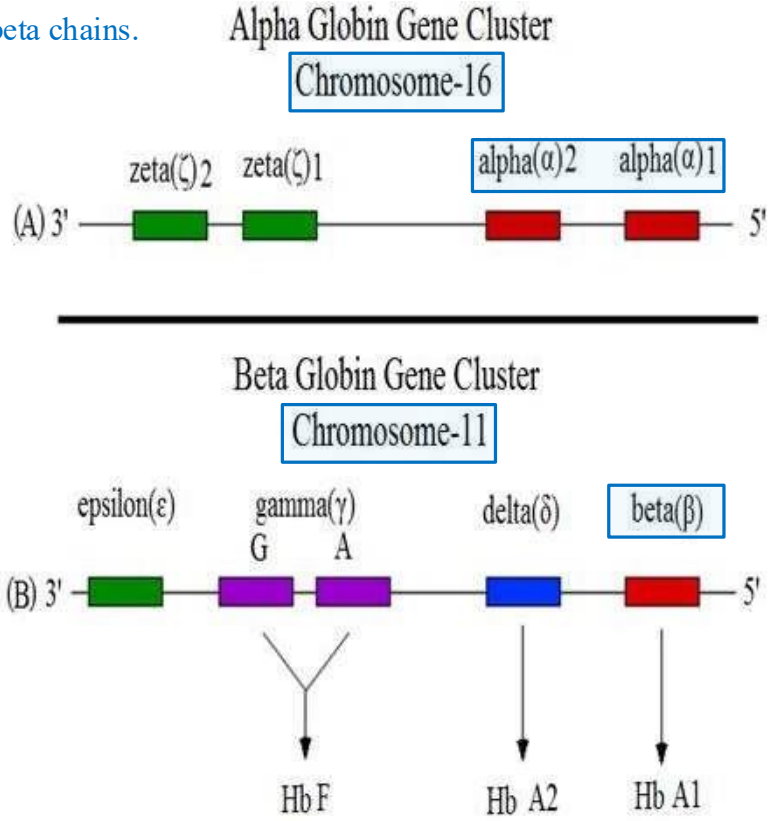
  - *If a stop codon is interrupted, translation will not terminate at the normal position. Instead, additional amino acids will be added, resulting in a longer protein. Therefore, the normal protein length is altered, which can affect protein function.*



Remember, hemoglobin consists of two alpha chains and two beta chains.

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




## 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 |
| $\beta^0$ | CTG | CTG | GTG | GTC | TAC | CCT | TGG | ACC | TAG |
| | Leu | Leu | Val | Val | Pro | Pro | Trp | Thr | STOP |

*The premature stop codon in the β-globin gene leads to early termination of translation, affecting the β protein.*

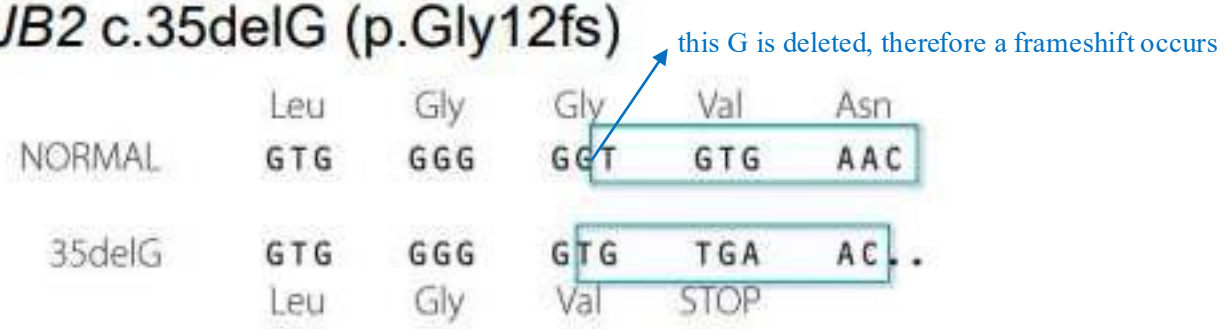
Creates premature termination codon and leads to NMD

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!

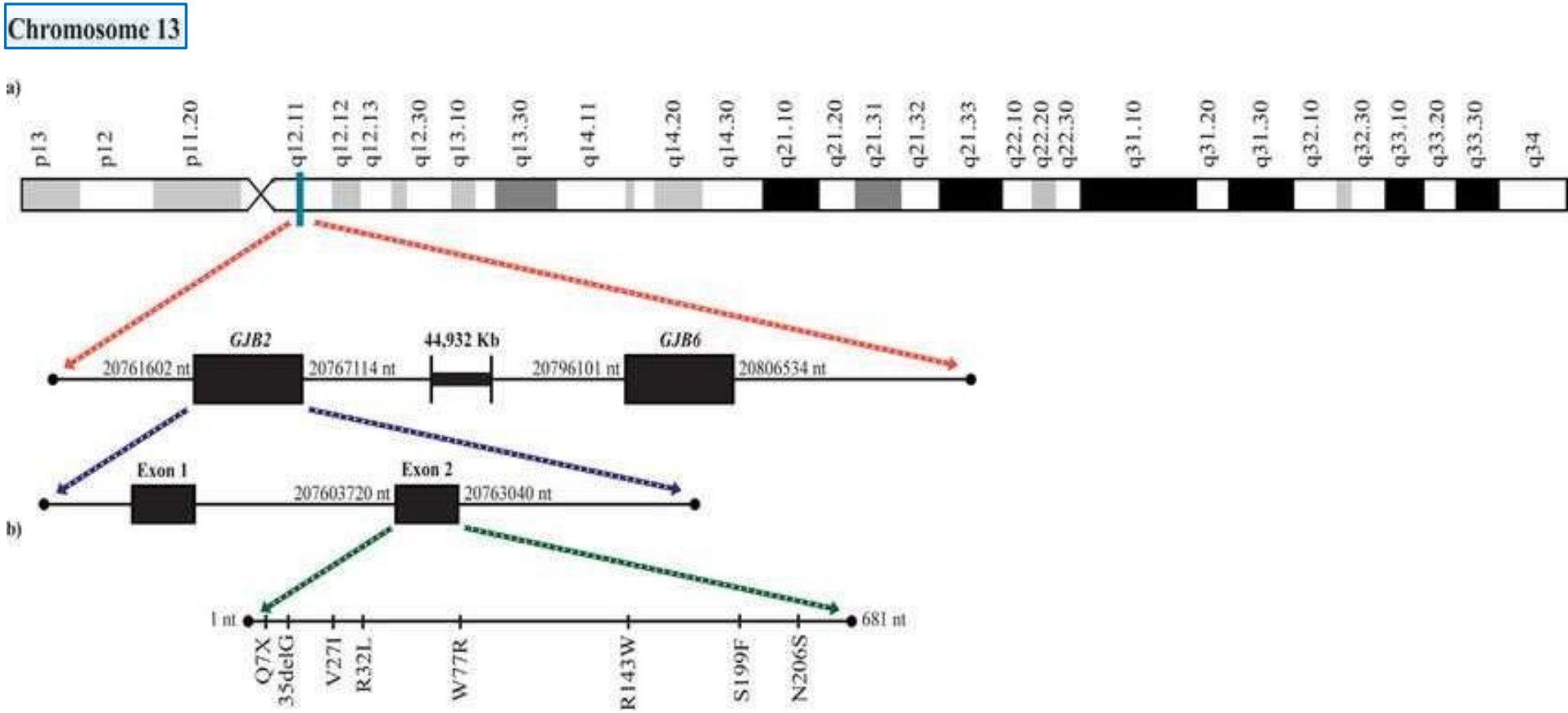
Example: *GJB2* c.35delG (p.Gly12fs)




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

Homozygotes: Non-syndromic hearing loss

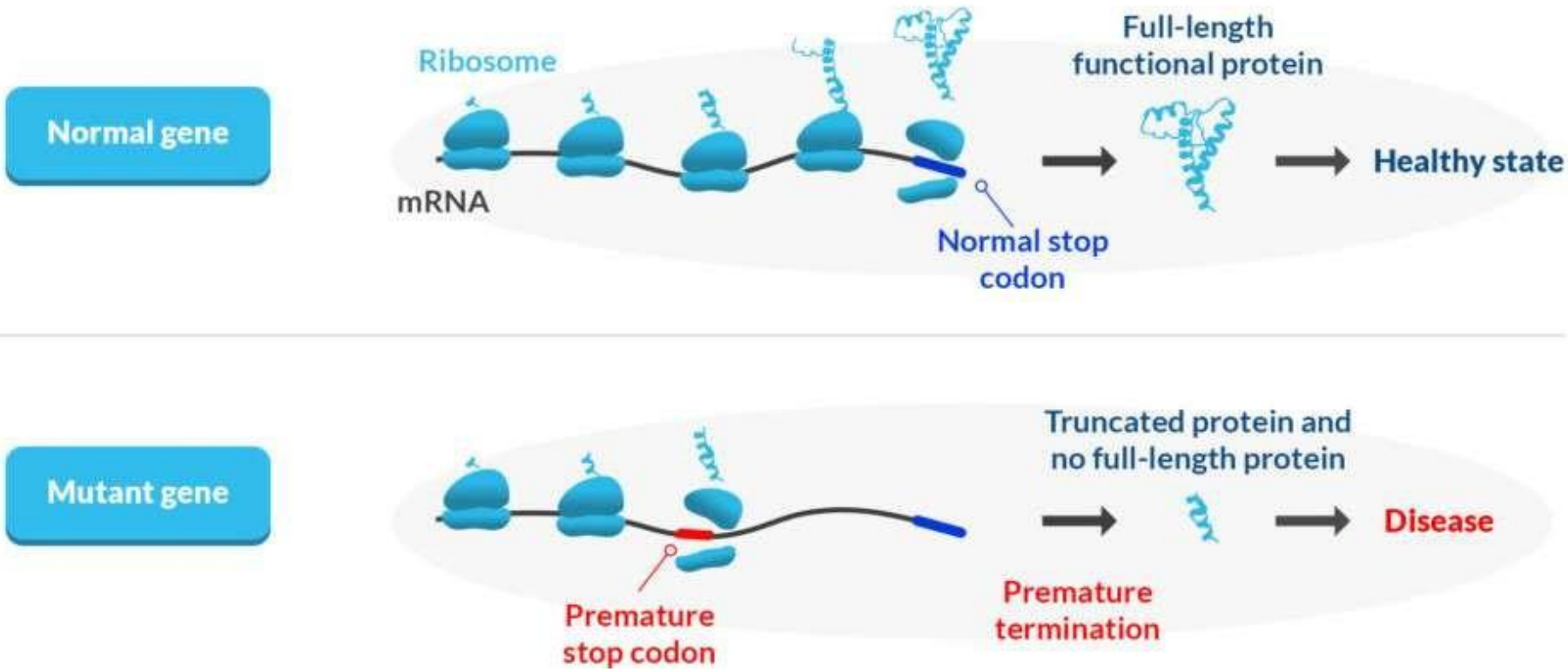
- In non-syndromic hearing loss, a frameshift mutation can occur in the GJB2 gene due to a small nucleotide deletion. For example, the c.35delG mutation involves deletion of a single G nucleotide, leading to a frameshift. As a result, the reading frame is altered downstream (e.g., GTG, which encodes valine), producing an abnormal protein sequence. This frameshift mutation disrupts normal protein function and is associated with recessive non-syndromic hearing loss.




- This is a zoom-in view of chromosome 13, specifically the q arm, region 1, sub-region 2.11. In this region, the genes GJB2 and GJB6 are located, along with intergenic and non-coding sequences spanning several megabases. When zooming into GJB2, which is associated with non-syndromic hearing loss, and GJB6, which is also linked to hearing loss, different types of variants can be observed (allelic heterogeneity), including deletions in exon 1 or exon 2. These variants include nucleotide substitutions, deletions, and mutations that may introduce stop codons or lead to amino acid changes, all of which can affect the resulting protein function.

# 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!

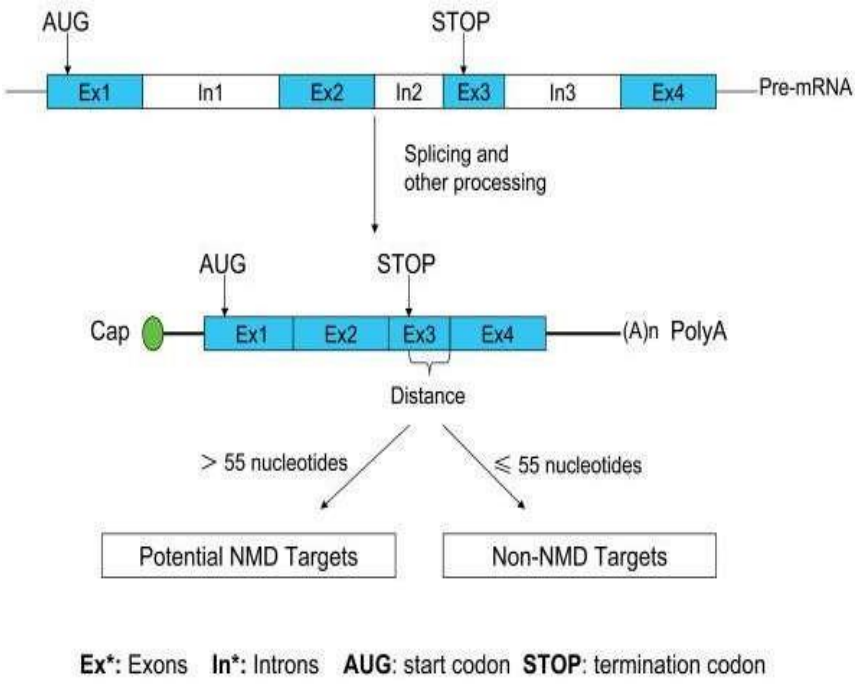


# Nonsense-mediated mRNA decay (NMD)

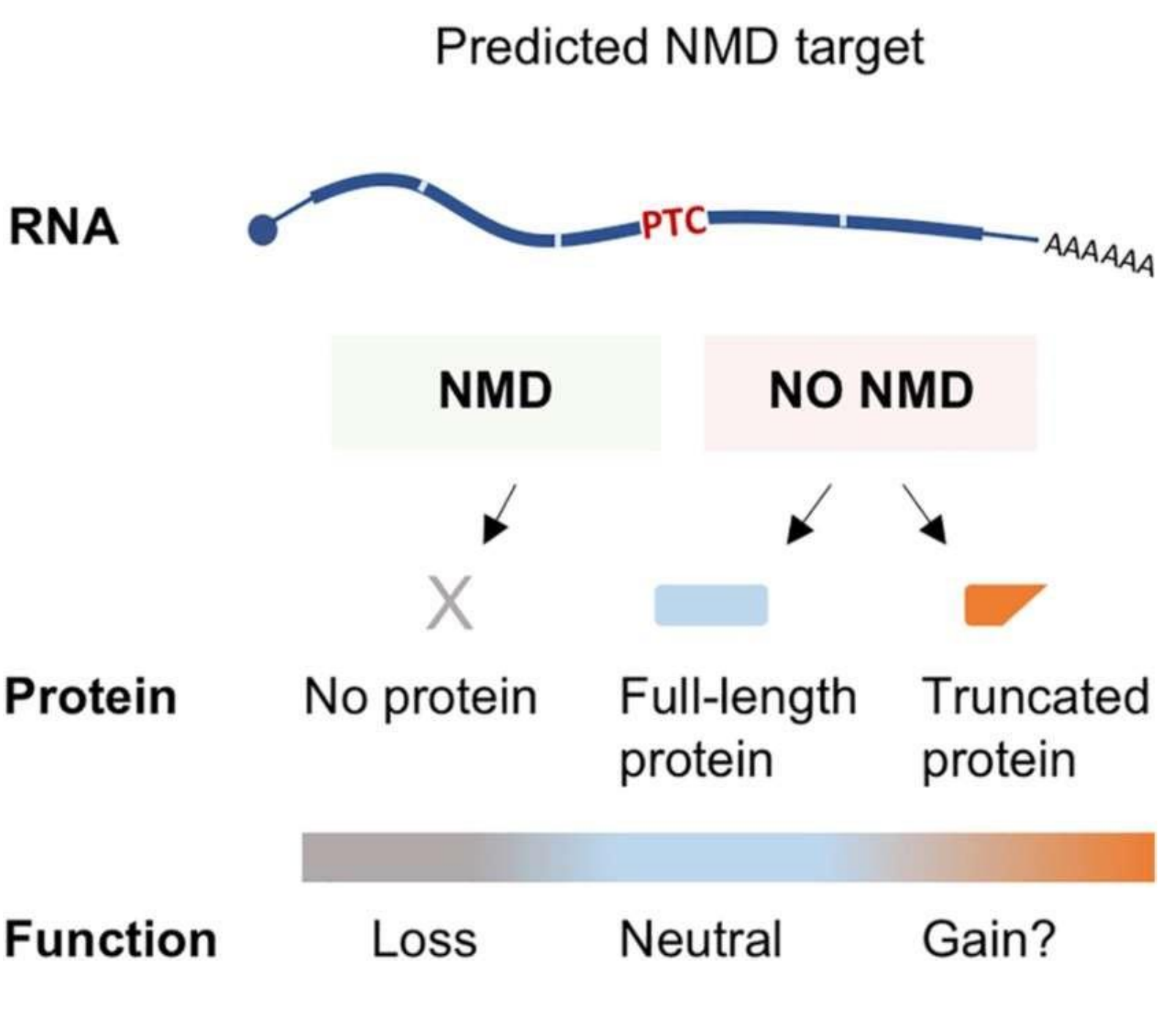
A premature stop codon gives two possible outcomes. **The first** is formation of a truncated protein, meaning everything downstream of this premature stop codon is lost. The upper panel in the figure shows the full protein, while the lower panel shows the truncated protein. This truncated protein loses its function and could lead to disease.

**The second** possibility is a mechanism called nonsense-mediated decay. In this mechanism, the mRNA is degraded before translation happens. So when there is a premature stop codon, either a truncated protein is produced or the mRNA is degraded.

What determines which outcome happens is the position of the stop codon. If the stop codon appears more than 55 nucleotides upstream of the exon junction, this leads to nonsense-mediated decay. If it is 55 nucleotides or less from the next exon–exon junction, nonsense-mediated decay does not happen.




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
- May be benign or pathogenic

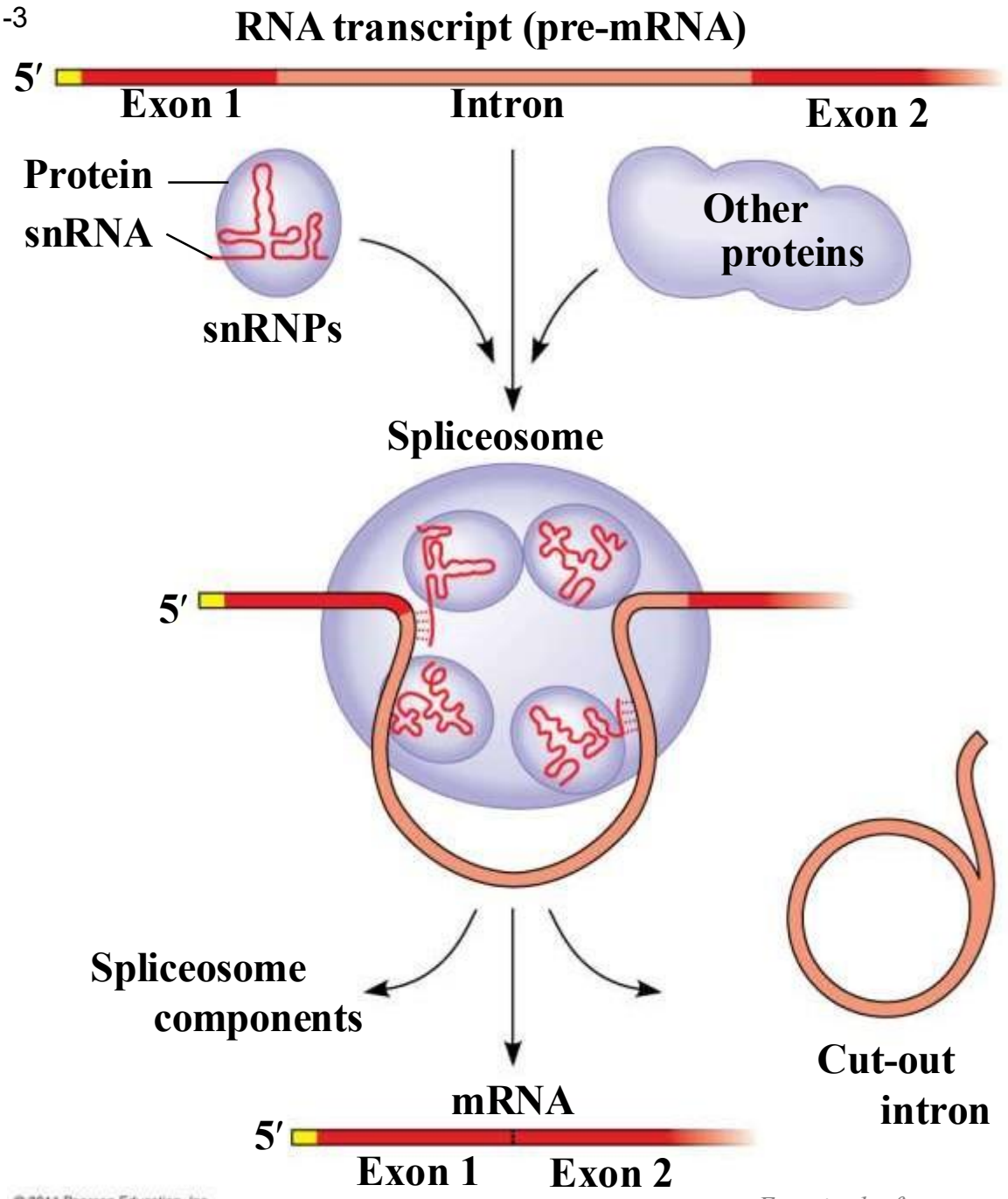
Example: CFTR c. (p.Phe508del – ΔF508)

| | | | | | |
|---|---|---|---|---|---|
| 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 17.12-3






*Examin the figure carefully then see the next slide*

# Spliceosome / Further explanation

- We have **exon 1**, then an **intron**, then **exon 2** in the **pre-mRNA**. During RNA processing, a large enzyme complex called the **spliceosome** carries out splicing. The spliceosome is an **RNA–protein complex**.

- The only RNA that is translated into protein is **mRNA**, but there are other types of RNA that are not translated and still have crucial functions. One of these is **small nuclear RNA (snRNA)**.

- **snRNA** combines with proteins to form **small nuclear ribonucleoproteins (snRNPs)**. These **snRNPs**, together with other proteins, assemble to form the **spliceosome**, which is a very large complex, almost as big as a **ribosome**.

- The sequence of the **snRNA** within the spliceosome is complementary to the **exon–intron boundaries**. This is how the **spliceosome** recognizes where the **exon** ends and the **intron** begins. It then catalyzes breaking the **phosphodiester bond**, removes the **intron**, and joins the **exons** together during **RNA processing**.

# Figure 1.16 The process of RNA splicing

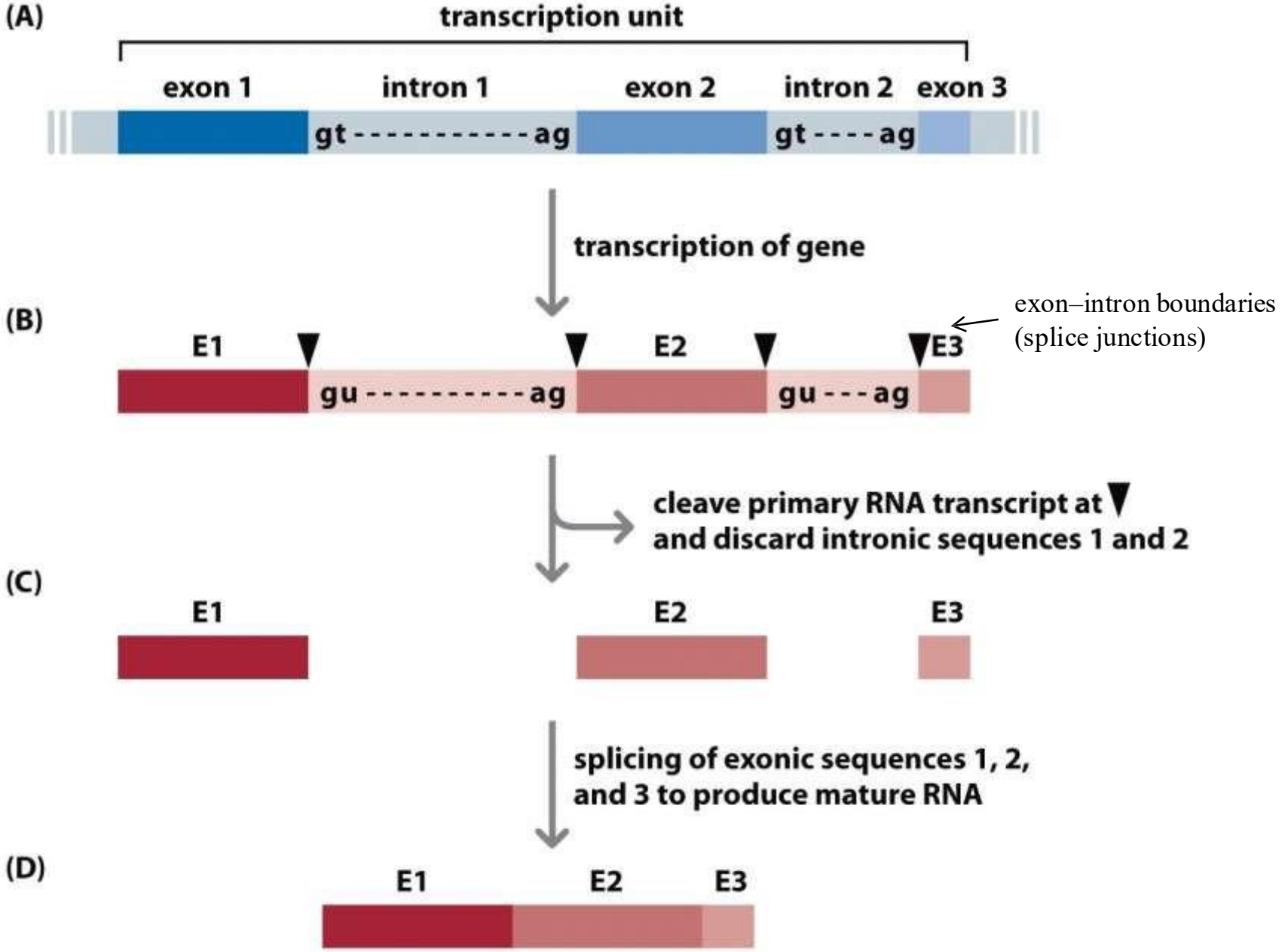



Figure 1.16 Human Molecular Genetics, 4ed. (© Garland Science)

*Examin the figure carefully then see the next slide*

# The process of RNA splicing

- If you look here, we have **three exons**, and the number of **introns** is always the number of exons minus one.
- At the beginning of the intron, we usually find **GT** at positions **c.+1 and c.+2**, and at the end of the intron we find **AG**. The same pattern is present in every intron. These are **consensus sequences**, and they can be found in introns of the same gene or in different genes at the **DNA level**.
- At the **RNA level**, **T** is replaced by **U**, so **GT** becomes **GU**, while **AG** remains **AG**.
- The **spliceosome** cuts before **GU** and after **AG** to remove the intron. Therefore, **GU** is called the **splice donor site**, and **AG** is called the **splice acceptor site**.
- That is why if these intronic sequences (**GT** or **AG**) are mutated, the **spliceosome** will not be able to recognize the splice site correctly.

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

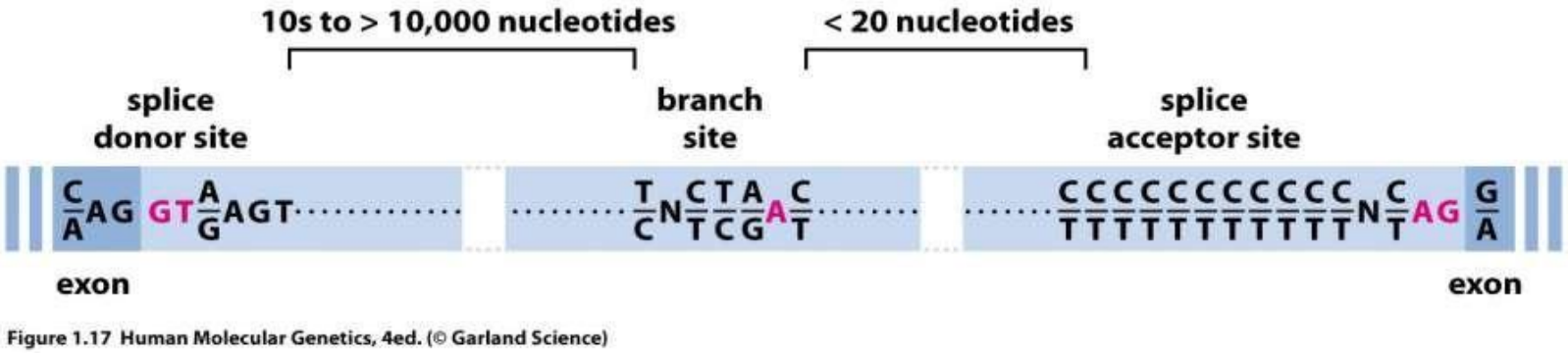



Figure 1.17 Human Molecular Genetics, 4ed. (© Garland Science)

- 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.

# The process of RNA splicing

- Are all **GT** or **AG** sequences within an intron splice sites? **No.** Only specific ones function as splice sites, and what makes them special is the **sequence surrounding them**.

- The **splice donor site**, highlighted in pink, contains **GT**. This **GT** is surrounded by a **consensus sequence**. A consensus sequence means it is **conserved** across genes—similar, but not necessarily identical. If you look at the Fig 1.17 3 , you will often find **GT**, but the surrounding nucleotides follow a recognizable pattern, such as **C or A then AG before GT**, and **A or G followed by AGT after it**.

- The same applies to the **splice acceptor site**. The **AG** lies within a **splice acceptor consensus sequence**, which may include a stretch of **C's or T's**, followed by **N** (any nucleotide), then **C or T**, then the **AG** splice acceptor sequence, followed by **A or G**. Again, it is the surrounding sequence that makes this **AG** recognizable.

- There is another important region located about **20 nucleotides upstream** of the **splice acceptor site** called the **branch site**. What distinguishes this region is a specific **adenine (A) nucleotide**. This adenine is special because of the **sequence surrounding it**, which allows it to be identified during the splicing process.

# 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

(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.

- Therefore, when there is a **variant** in any of these regions—such as the **splice donor site**, **splice acceptor site**, or **branch site**—it is considered a **red flag** and can be reported as a **disease-causing variant**.

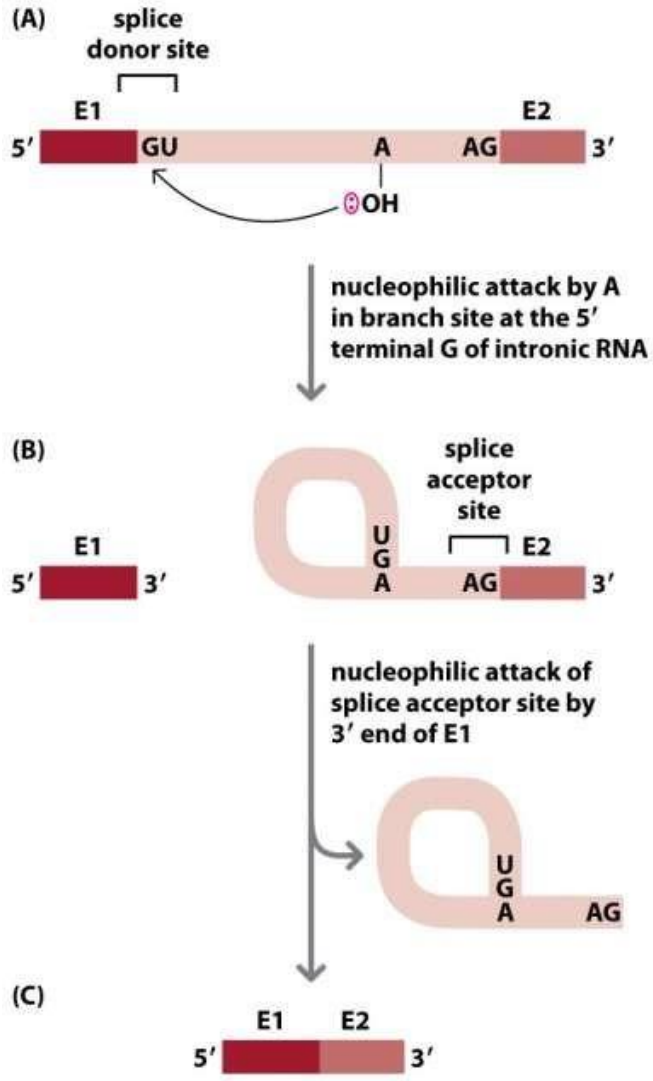



Figure 1.18 Human Molecular Genetics, 4ed. (© Garland Science)

*Examin the figure carefully then see the next slide*

# Role of small nuclear ribonucleoprotein (snRNPs) in RNA splicing

- The **spliceosome** is a large complex made up of **U1, U2, U4, U5, and U6** subunits.

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 E1 and E2 to be spliced together.

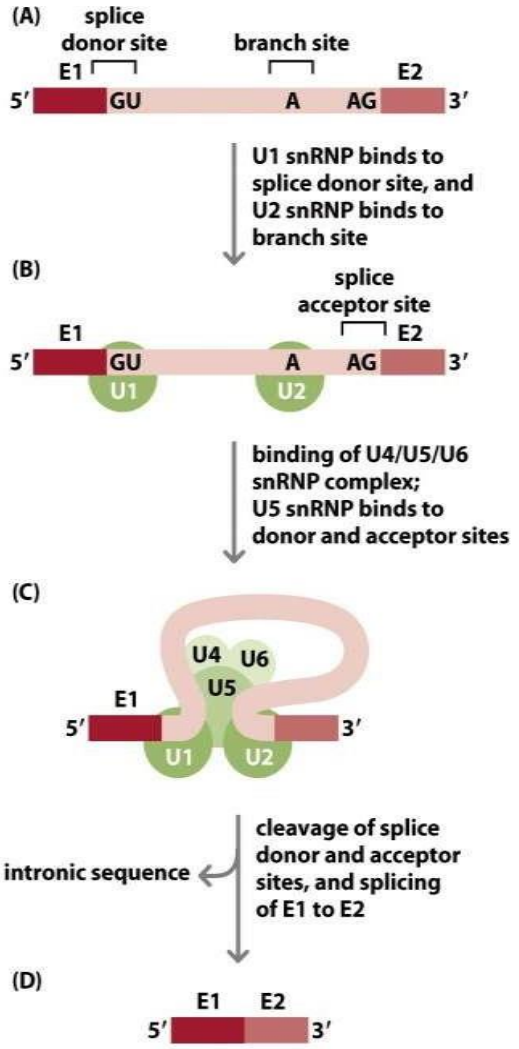



Figure 1.19 Human Molecular Genetics, 4ed. (© Garland Science)

# Splice mutations

+(1,2) is the GT (splice donor site)
-(1,2) is the AG (splice acceptor site)

- Variants that likely impact splicing :

  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