

POST-TRANSCRIPTIONAL MODIFICATION

Capping, Splicing, Polyadenylation, RNA Editing, and the Maturation of tRNA and rRNA

1. Introduction

The RNA molecule produced directly by RNA polymerase — the **primary transcript** — is rarely the final, functional RNA used by the cell. Before a eukaryotic pre-mRNA can be translated, it must undergo a series of **post-transcriptional modifications** that convert it into a stable, exportable, translatable mature mRNA. Non-coding RNAs such as tRNA and rRNA are likewise extensively processed and chemically modified after being transcribed.

- The three major processing events for a eukaryotic protein-coding transcript are **5' capping**, **splicing** (removal of introns), and **3' cleavage and polyadenylation** — all of which occur largely co-transcriptionally, while RNA polymerase II is still elongating the transcript.
 - Additional processing includes **alternative splicing** (generating multiple protein products from one gene), **RNA editing** (altering individual bases after transcription), and extensive **chemical modification** of tRNA and rRNA.
 - Because bacterial mRNA is translated as it is transcribed and generally requires little or no processing, most of these mechanisms are hallmarks of **eukaryotic** gene expression, reflecting the physical separation of transcription (nucleus) from translation (cytoplasm).
-

2. Processing the Two Ends: 5' Capping and 3' Polyadenylation

Both ends of a eukaryotic pre-mRNA are chemically modified to protect the transcript and mark it as ready for export and translation.

Processing the Two Ends of a Pre-mRNA: Capping and Polyadenylation

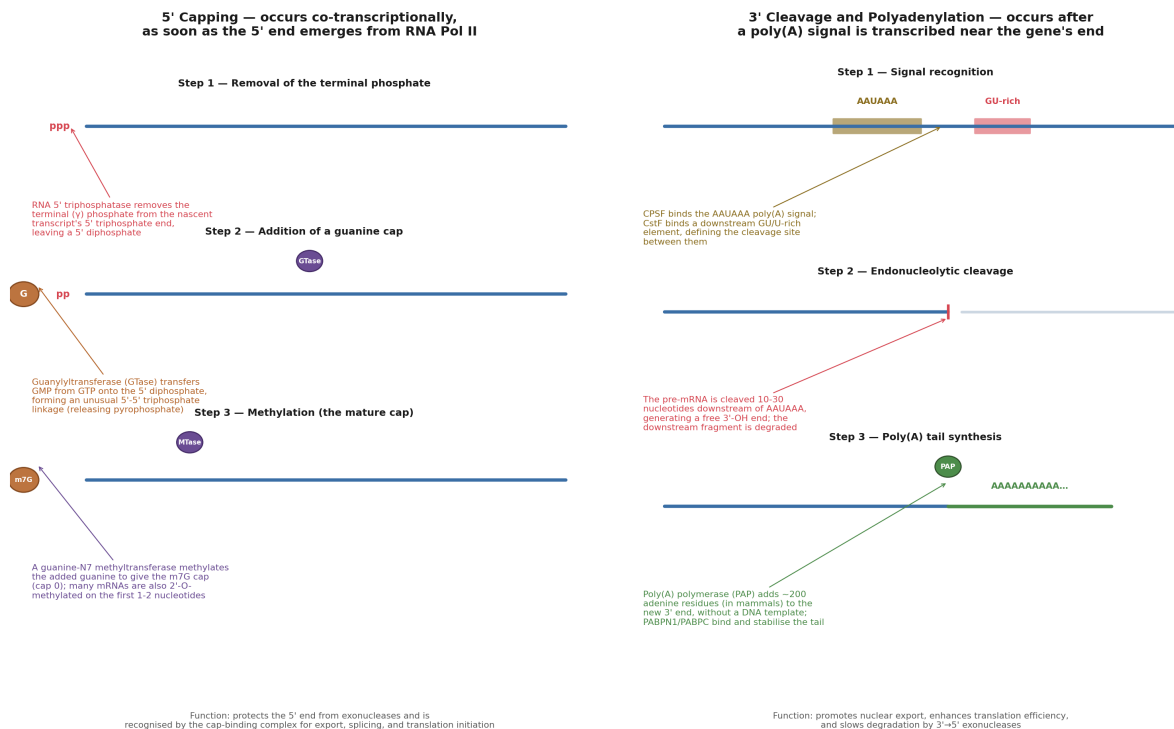


Fig. 1 — The 5' end is capped with a modified guanine nucleotide within seconds of transcription initiation, while the 3' end is cleaved and polyadenylated once a poly(A) signal is transcribed.

(a) 5' Capping

- Occurs almost immediately after RNA Pol II synthesises the first 20-30 nucleotides, catalysed by enzymes physically tethered to the phosphorylated C-terminal domain (CTD) of Pol II itself.
- Three enzymatic steps: **RNA triphosphatase** removes the 5' γ -phosphate; **guanylyltransferase** adds a GMP in an unusual 5'-5' triphosphate linkage; **guanine-N7 methyltransferase** methylates the added guanine to form the **7-methylguanosine (m⁷G) cap**.
- The cap protects the transcript from 5'→3' exonucleases, and is recognised by the nuclear cap-binding complex and, later, eIF4E in the cytoplasm — required for efficient splicing, nuclear export, and translation initiation.

(b) 3' Cleavage and Polyadenylation

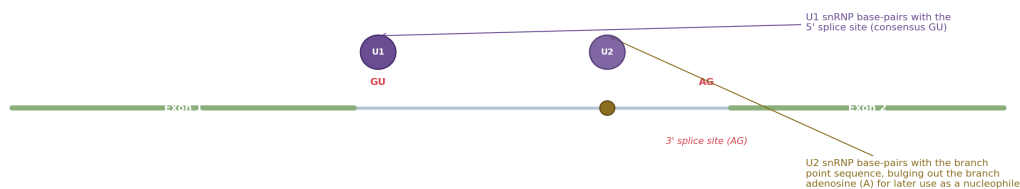
- Occurs once RNA polymerase transcribes a **poly(A) signal** (consensus **AAUAAA**) near the end of the gene; **CPSF** binds this signal while **CstF** binds a downstream GU/U-rich element, defining the cleavage site between them.
- An endonuclease cleaves the transcript 10-30 nucleotides downstream of AAUAAA, generating a free 3'-OH; the downstream RNA fragment (still attached to elongating Pol II) is degraded, a signal that also helps trigger transcription termination.
- Poly(A) polymerase (PAP)** then adds roughly 200 adenine residues (in mammals) to the new 3' end without using a DNA template; poly(A)-binding proteins coat the tail and protect it.
- The poly(A) tail promotes nuclear export, enhances translational efficiency (partly via interactions with eIF4G that circularise the mRNA), and slows degradation by 3'→5' exonucleases; its gradual shortening (deadenylation) is usually the first step in normal mRNA turnover.

3. Pre-mRNA Splicing

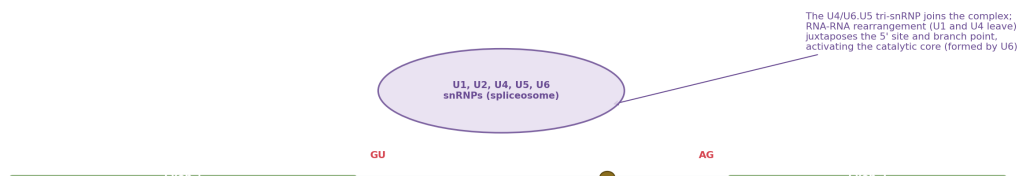
Eukaryotic genes are typically interrupted by non-coding **introns** that must be precisely removed, joining the flanking coding **exons** together. Splicing is carried out by the **spliceosome**, a large, dynamic ribonucleoprotein machine built from five small nuclear ribonucleoproteins (**snRNPs**: U1, U2, U4, U5, U6) plus many associated proteins.

Pre-mRNA Splicing by the Spliceosome (the Lariat Pathway)

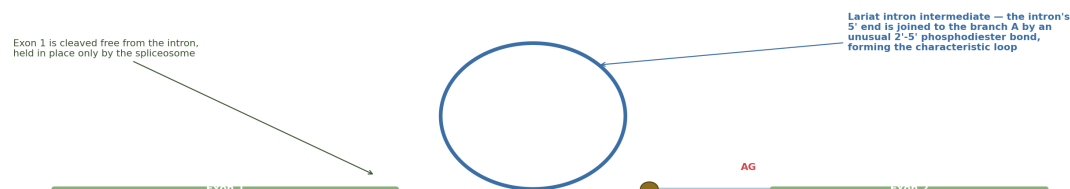
1. Splice-site recognition — U1 snRNP binds the 5' splice site, U2 snRNP binds the branch point



2. Spliceosome assembly — the U4/U6.U5 tri-snRNP joins to form the catalytically active spliceosome



3. First transesterification — the branch A's 2'-OH attacks the 5' splice site, forming a lariat



4. Second transesterification — exon 1's 3'-OH attacks the 3' splice site, joining the exons

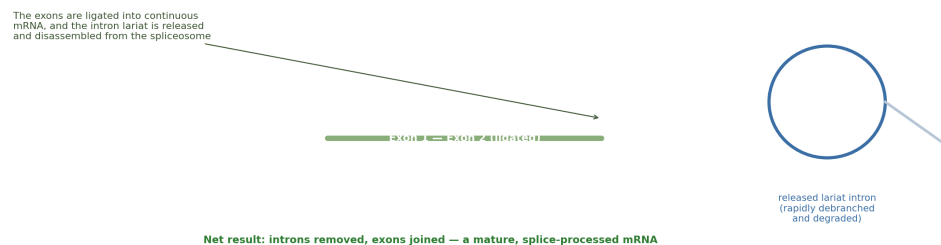


Fig. 2 — Splicing proceeds through two sequential transesterification reactions, forming and then resolving a branched lariat intermediate, without any net energy loss from the reactions themselves.

- Three conserved sequence elements define an intron: the **5' splice site** (consensus GU at the intron's start), the **branch point** (an adenosine roughly 18-40 nucleotides upstream of the 3' end, within a conserved sequence), and the **3' splice site** (consensus AG at the intron's end).

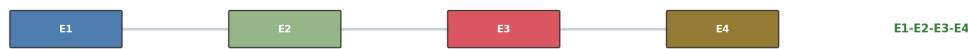
- **U1 snRNP** base-pairs with the 5' splice site and **U2 snRNP** base-pairs with the branch point sequence, bulging out the branch adenosine; the **U4/U6.U5 tri-snRNP** then joins, and extensive rearrangement (releasing U1 and U4) activates the catalytic core formed by U6.
 - In the **first transesterification**, the branch point adenosine's 2'-OH attacks the 5' splice site, cleaving the RNA there and joining the intron's 5' end to the branch point by an unusual 2'-5' phosphodiester bond — creating the characteristic **lariat** (loop-and-tail) intermediate, with exon 1 now cleaved free.
 - In the **second transesterification**, the newly freed 3'-OH of exon 1 attacks the 3' splice site, simultaneously ligating the two exons together and releasing the intron as a free lariat, which is rapidly debranched and degraded.
 - Because both steps are transesterifications (a phosphodiester bond is broken and a new one formed in the same step), splicing does not consume ATP directly, although ATP is required for the extensive spliceosome remodelling that occurs throughout the cycle.
-

4. Alternative Splicing

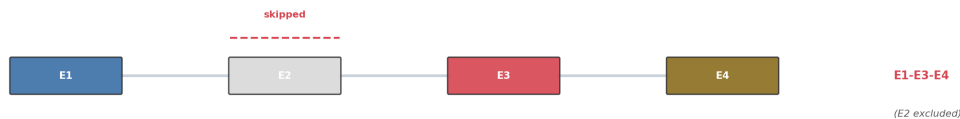
Most human genes undergo **alternative splicing**: the same pre-mRNA can be spliced in more than one way, producing different mature mRNAs — and therefore different protein isoforms — from a single gene. This is a major reason the human proteome is far more diverse than the roughly 20,000 protein-coding genes would suggest on their own.

Patterns of Alternative Splicing

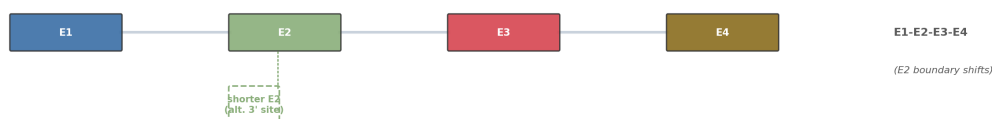
Reference — pre-mRNA with four exons, spliced constitutively (all exons retained, in order)



1. Exon skipping (cassette exon) — an internal exon is excluded from the mature mRNA



2. Alternative splice-site selection — a different 5' or 3' splice site shortens/lengthens an exon



3. Intron retention — an intron is left unspliced and remains in the mature mRNA



4. Mutually exclusive exons — only one of two alternative exons is ever included

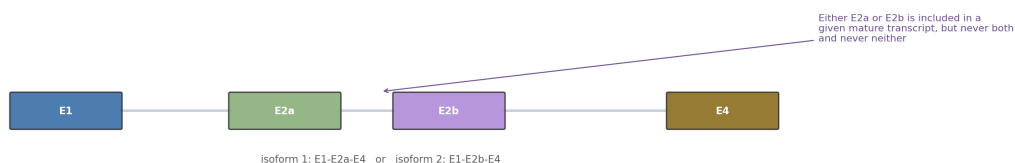


Fig. 3 — The main patterns of alternative splicing: exon skipping, alternative splice-site selection, intron retention, and mutually exclusive exons.

- **Exon skipping (cassette exons)** — an internal exon is sometimes included and sometimes spliced out entirely; the most common form of alternative splicing in mammals.
- **Alternative 5' or 3' splice-site selection** — use of a different splice site shortens or lengthens one end of an exon.
- **Intron retention** — an intron is left unspliced and remains part of the mature mRNA, which can introduce a premature stop codon or alter the protein's C-terminus.
- **Mutually exclusive exons** — of two (or more) alternative exons, exactly one is always included and the others always excluded.

- Alternative splicing is often regulated by tissue-specific or developmentally-regulated splicing factors (e.g. SR proteins and hnRNPs) that bind exonic/intronic splicing enhancers or silencers, and is itself frequently disrupted in disease — mutations affecting splice sites or splicing regulators cause conditions such as spinal muscular atrophy and are found in many cancers.

5. RNA Editing

RNA editing changes the nucleotide sequence of an RNA after transcription, independent of splicing, through enzymatic modification of individual bases.

- C-to-U editing** — the enzyme **APOBEC1** deaminates a specific cytosine to uracil in the apolipoprotein B (apoB) mRNA in the human intestine, converting a glutamine codon into a premature stop codon and producing a shorter protein isoform (apoB-48) than in the liver, where the mRNA is left unedited (apoB-100).
- A-to-I editing** — enzymes of the **ADAR** family deaminate adenosine to inosine (read by the ribosome as guanosine) at specific double-stranded RNA sites; this is especially important in the nervous system, where it can alter amino acid coding, splicing, or microRNA targeting, and in innate immunity, where it marks the cell's own double-stranded RNA as "self".

6. Processing of tRNA and rRNA

(a) tRNA Processing

- The 5' leader sequence of a pre-tRNA is removed by the ribozyme **RNase P**; the 3' trailer is removed by other nucleases.
- A non-templated **CCA sequence** is added (or repaired) at the 3' end by CCA-adding enzyme — the site to which the amino acid is later attached.
- Many nucleotides are chemically modified (e.g. to pseudouridine, inosine, or dihydrouridine), and rare tRNA introns (removed by a dedicated splicing endonuclease and ligase, unrelated to the spliceosome) are excised in some tRNA genes.

(b) rRNA Processing

- A single long precursor (pre-rRNA, 45S in humans) is transcribed by RNA Pol I in the nucleolus and cleaved into the mature 18S, 5.8S, and 28S rRNAs; 5S rRNA is transcribed separately by RNA Pol III.
- Small nucleolar RNAs (snoRNAs)** guide site-specific chemical modification of the pre-rRNA — chiefly **2'-O-methylation** and conversion of uridine to **pseudouridine** — which fine-tunes ribosome structure and function.

7. Prokaryotic vs Eukaryotic mRNA Processing

Because bacterial transcription and translation are coupled in the same cellular compartment, bacterial mRNA generally undergoes little to no processing compared with the extensive maturation required by eukaryotic transcripts.

Feature	Prokaryotic mRNA	Eukaryotic mRNA
5' cap	Absent — retains a 5' triphosphate	m ⁷ G cap added co-transcriptionally

Feature	Prokaryotic mRNA	Eukaryotic mRNA
Splicing	Rare (occasional self-splicing group I/II introns)	Extensive; carried out by the spliceosome
3' end	Often unmodified, or a short poly(A) tail added as a degradation signal	Cleaved and polyadenylated (~200 A residues), stabilising the mRNA
Translation timing	Begins on the nascent transcript, before transcription finishes	Only after processing is complete and mRNA is exported to the cytoplasm
mRNA lifespan	Typically minutes	Typically hours or longer

8. Quick Revision Summary

Process	Key point
5' Capping	m ⁷ G cap added co-transcriptionally; protects mRNA and aids export, splicing, translation
3' Polyadenylation	CPSF/CstF define cleavage site at AAUAAA; PAP adds poly(A) tail; aids stability and export
Splicing	Spliceosome (U1, U2, U4, U5, U6 snRNPs) removes introns via two transesterifications, forming a lariat
Alternative splicing	Exon skipping, alternative splice sites, intron retention, mutually exclusive exons — expands proteome diversity
RNA editing	APOBEC1 (C→U) and ADAR (A→I) alter individual bases post-transcriptionally
tRNA/rRNA processing	Nucleolytic trimming, CCA addition, and snoRNA-guided methylation/pseudouridylation
Prokaryotes vs eukaryotes	Bacterial mRNA is minimally processed; eukaryotic mRNA requires capping, splicing, and polyadenylation before translation

Compiled study note — molecular biology reference. Figures are original schematic diagrams prepared for this note and are not to scale.