

Gene structure, organization, Genomes and genetic code

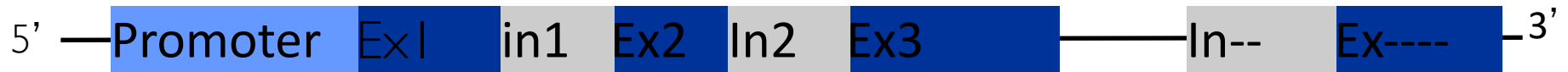
AHMAD BABALGHITH

Objectives

- To understand the structure of the gene.
- To understand how the genetic code is used.

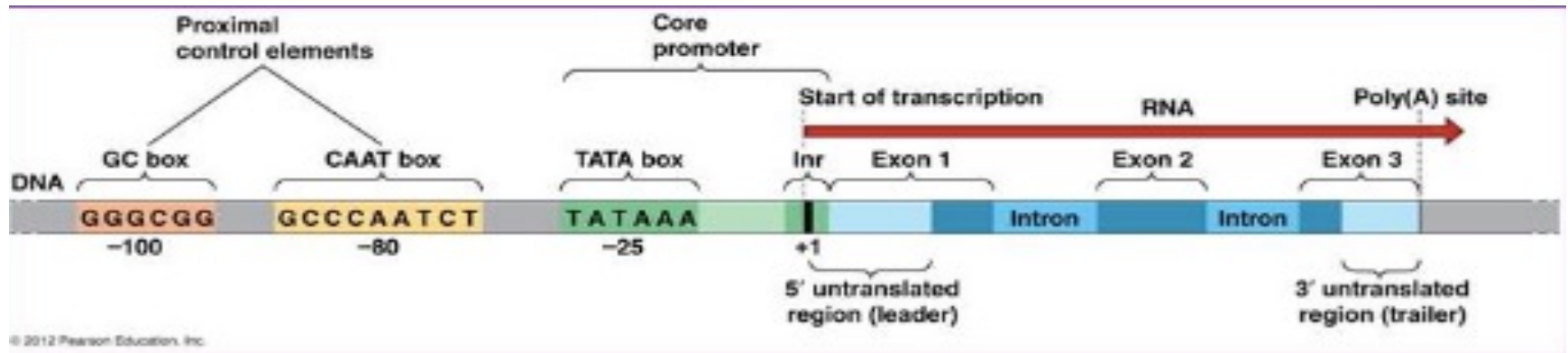
Gene Structure

The Gene



- A gene: DNA sequence that is needed to encode amino acid sequence of a protein
- all genes are Composed of exons, introns and different control elements

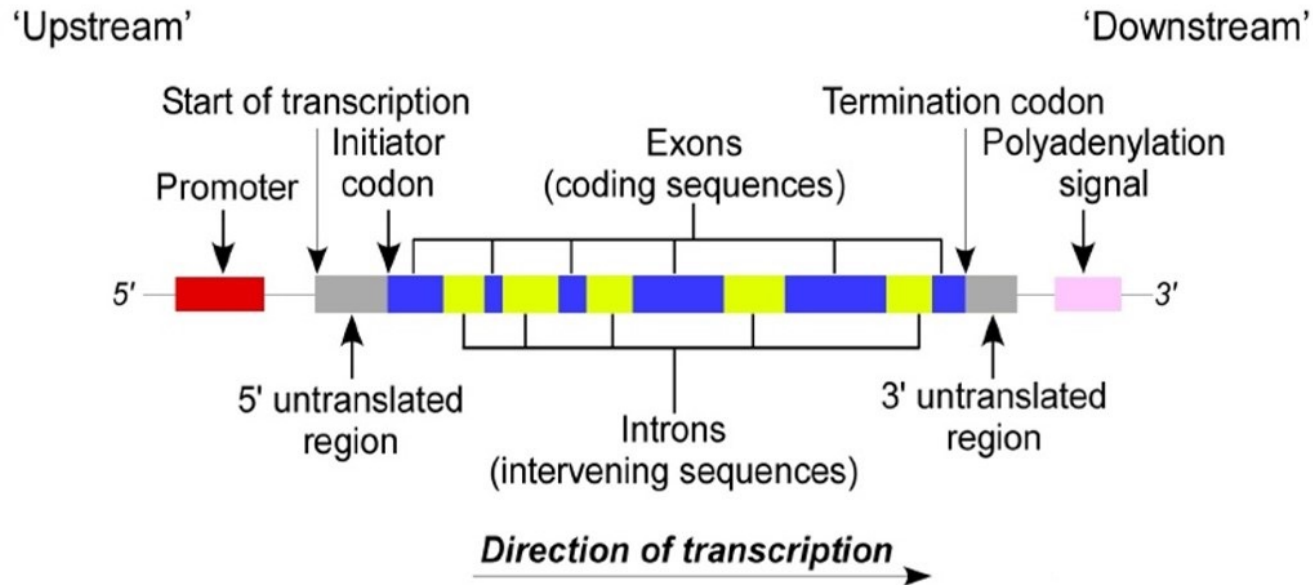
The Gene



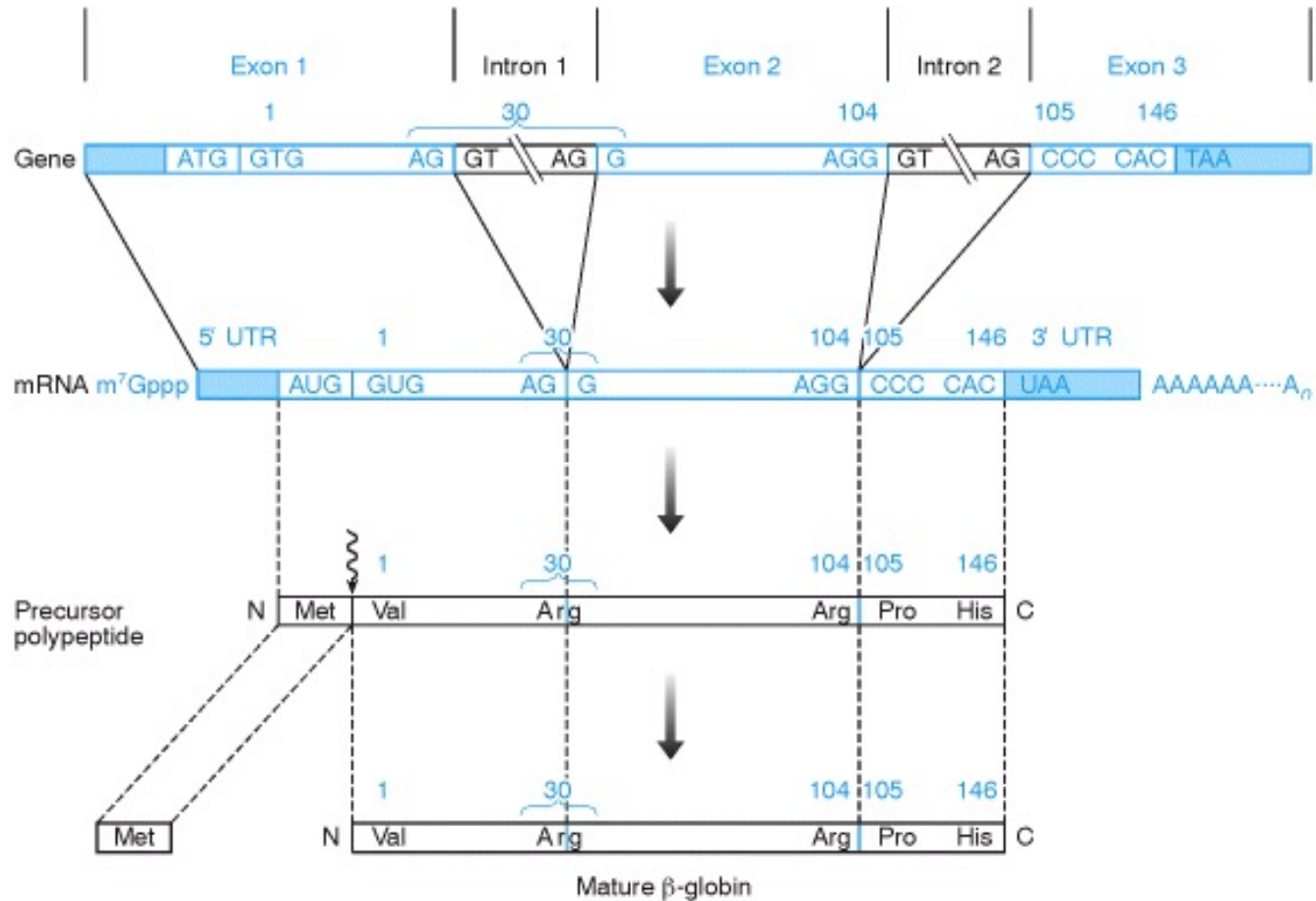
- **Exons:** are the functional portions of gene sequences that code for proteins.
- **Introns:** are the noncoding sequences which separate the coding sequence (exons).
- **The open reading frame:** a sequence with variable length that does not contain stop codons and therefore can be translated. The sequence beginning with ATG which exist at the 5' end of genes.
- **TATA boxes:** These regions are about 20-30 bases to the 5' end (left) of the open reading frame (ATG). TATA boxes direct important enzymes to the correct initiation site for transcription.
- **Termination codon:** the end of translation is signified by a termination codon at the 3' end of genes. The termination codon could be TAA, TAG, or TGA.

Organisation Of Genes

Structure of a typical eukaryotic gene



The Gene



Multigene Families

Many genes have similar functions making up what are known as multigene families.

Multigene families can be split into two types:

1- **Classical gene families** which show a high degree of sequence **homology**.

2- **Gene superfamilies** which have limited sequence **homology** but are functionally related

Genetic code

Genetic code

- The genetic code describes how base sequences are converted into amino acid sequences during protein synthesis.
- The DNA sequence of gene is divided into a series of units of **three bases**.
- Each set of three bases is called a **codon** and specifies a particular amino acid
- The four bases in DNA and RNA can combine as a total of $4^3=64$ **codons** which specify the **20 amino acid** found in proteins.

Genetic code

- Genetic code is the set of biological rules by which DNA nucleotide base pair is translated into corresponding amino acids.
- Genetic code is triplet code (i.e. contain three nucleotides)
- Start codon (ATG or AUG)
- Stop codon (TAA, TAG, and TGA)
- Genetic code is universal between all organisms with few exceptions.

- The direction of the code is 5 to 3
- Several amino acids have more than one genetic code (redundancy of the code).
- Only two amino acids (methionine and tryptophan) are encoded by single genetic code. On the other hand, Leucine and serine by 6 codon

Genetic code

	U	C	A	G	
U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	U
	UUC } Leu	UCC } Ser	UAC } Tyr	UGC } Cys	C
	UUA } Leu	UCA } Ser	UAA } Stop	UGA } Stop	A
	UUG } Leu	UCG } Ser	UAG } Stop	UGG } Trp	G
C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg	U
	CUC } Leu	CCC } Pro	CAC } His	CGC } Arg	C
	CUA } Leu	CCA } Pro	CAA } Gln	CGA } Arg	A
	CUG } Leu	CCG } Pro	CAG } Gln	CGG } Arg	G
A	AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser	U
	AUC } Ile	ACC } Thr	AAC } Asn	AGC } Ser	C
	AUA } Ile	ACA } Thr	AAA } Lys	AGA } Arg	A
	AUG } Met	ACG } Thr	AAG } Lys	AGG } Arg	G
G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly	U
	GUC } Val	GCC } Ala	GAC } Asp	GGC } Gly	C
	GUA } Val	GCA } Ala	GAA } Glu	GGA } Gly	A
	GUG } Val	GCG } Ala	GAG } Glu	GGG } Gly	G

First position (5' end)

Third position (3' end)

Amino acid names:

- | | | | |
|------------------|------------------|---------------------|------------------|
| Ala = alanine | Gln = glutamine | Leu = leucine | Ser = serine |
| Arg = arginine | Glu = glutamate | Lys = lysine | Thr = threonine |
| Asn = asparagine | Gly = glycine | Met = methionine | Trp = tryptophan |
| Asp = aspartate | His = histidine | Phe = phenylalanine | Tyr = Tyrosine |
| Cys = cysteine | Ile = Isoleucine | Pro = proline | Val = valine |