Genetic code

The Central Dogma of Molecular Biology

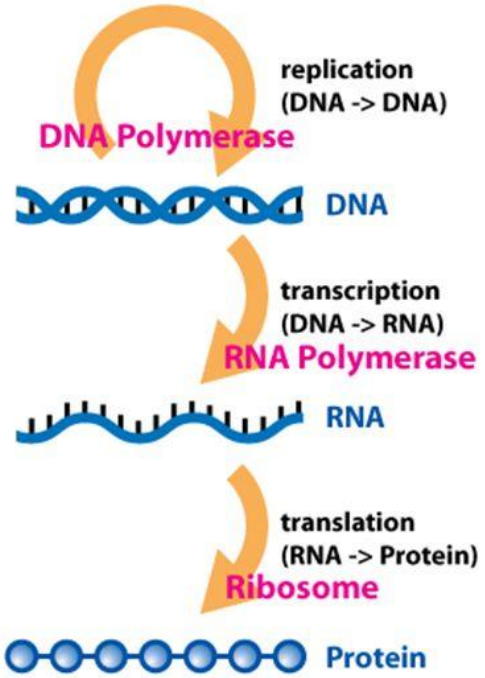
- Information is transferred from DNA to RNA to protein

  DNA -> RNA -> Protein

- Proteins create traits

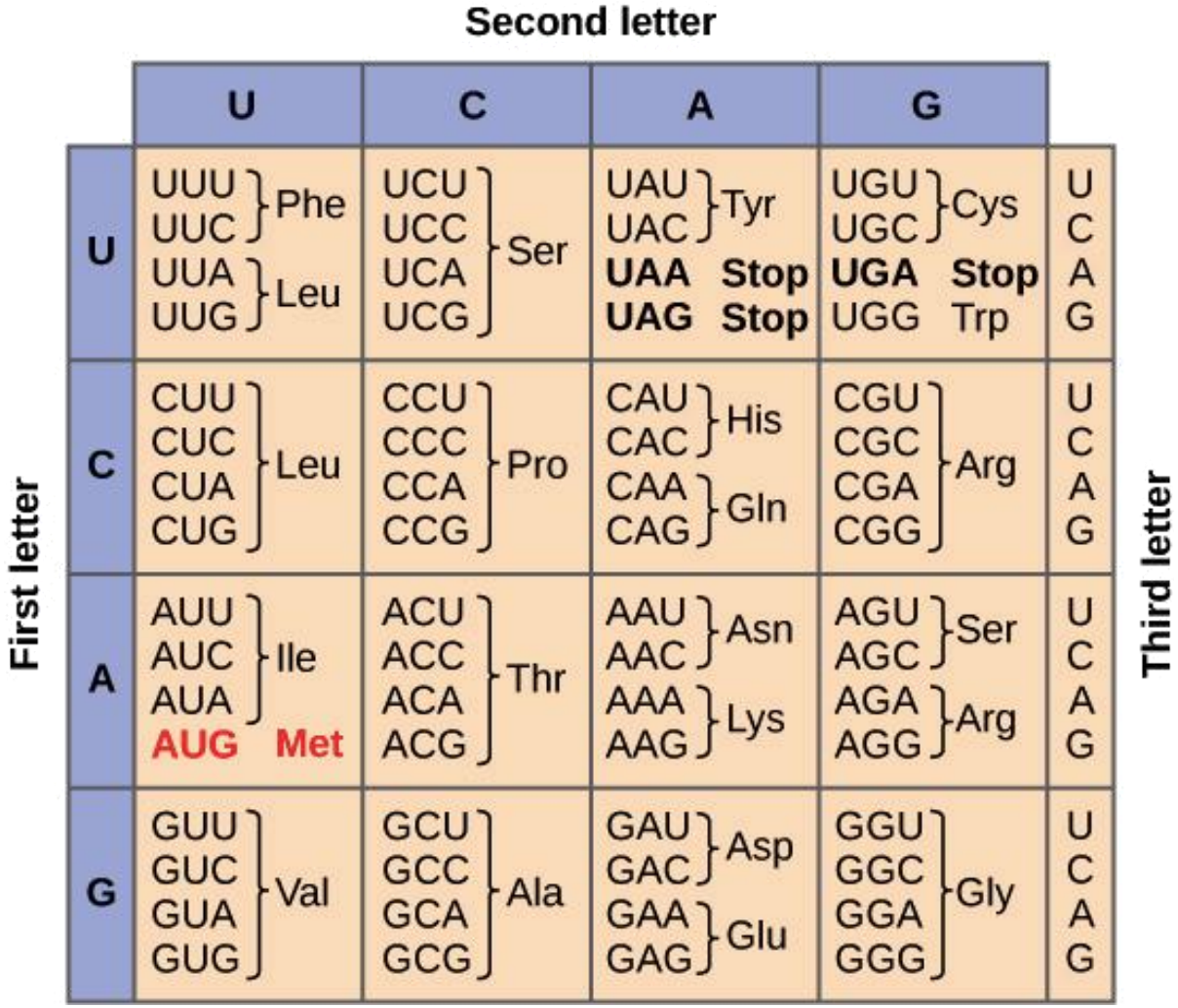
- This is called **gene expression**

- This process is found in all organisms



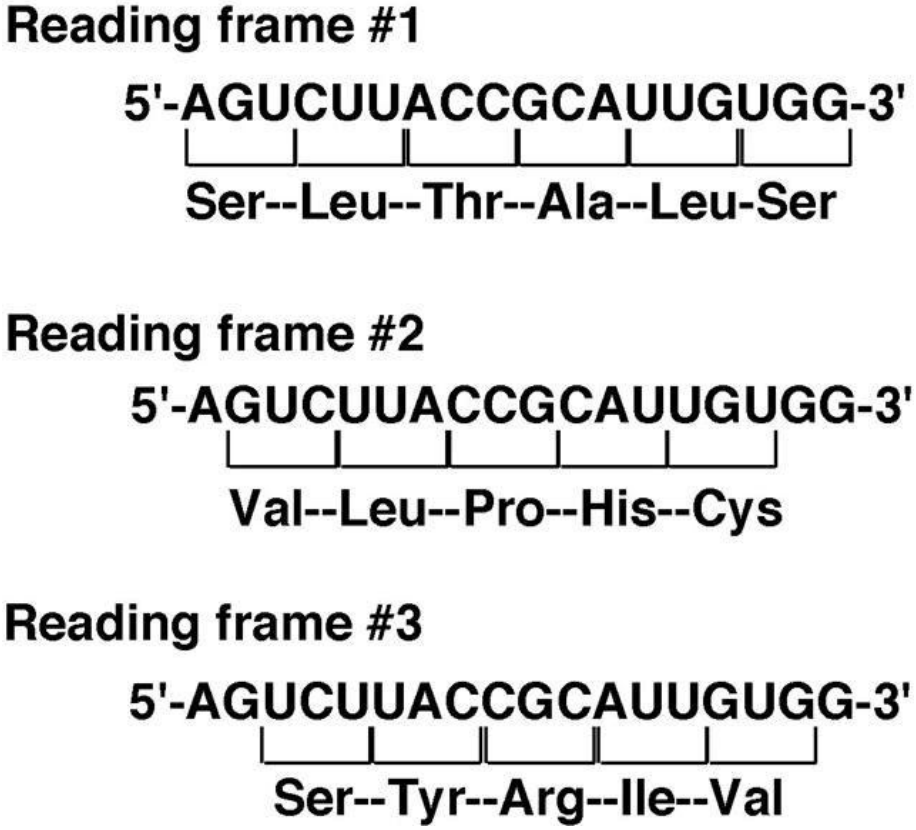
- The set of rules by which information encoded within genetic material (DNA or mRNA sequences) is translated into proteins is called *genetic code*.
- The protein sequence is encoded in mRNA in nucleotide **triplets** called *codons*.
- There are 20 amino acids and 64 possible combinations of 3 consecutive bases. Therefore the genetic code is *degenerative*. For many amino acids there are more then one corresponding codon.
- There are 3 triplets that do not encode for any amino acid. They indicate termination of the translation process. They are called *stop codons*.

Genetic code table



Reading frame

- **a reading frame is a way of dividing the sequence of nucleotides in a nucleic acid (DNA or RNA) molecule into a set of consecutive, non-overlapping triplets.**



- Each strand of a double-stranded DNA (ds-DNA) molecule theoretically encodes for 3 different protein sequences. Therefore, ds-DNA molecule theoretically encodes for 6 different protein sequences.
- The DNA strand from which RNA is transcribed and the frame in which the RNA is translated into the actual protein are chosen through mechanisms of *gene expression regulation.*

Mutations

- DNA replication is a high fidelity process. It ensures the almost precise transmission of traits from parents to offspring.
- Sometimes, however, errors could occur during DNA replication. An alteration of the nucleotide sequence of the genome of an organism is called a *mutation*.
- On average, it was determined that the wrong base is incorporated into the **yeast** genome only once per 10^7 nucleotides. The **human** germline mutation rate is approximately 0.5×10^{-9} per basepair per year.
- Environmental factors, like exposure to certain chemicals called *mutagens,* may lead to an increased mutation rate.

Mutations can cause 2 kinds of alteration of DNA sequence:

- Substitutions of nucleotide sequence with a different sequence. If only one nucleotide is changed the mutation is called a *point mutation*.
- Deletions or insertions of nucleotides

Substitution mutations

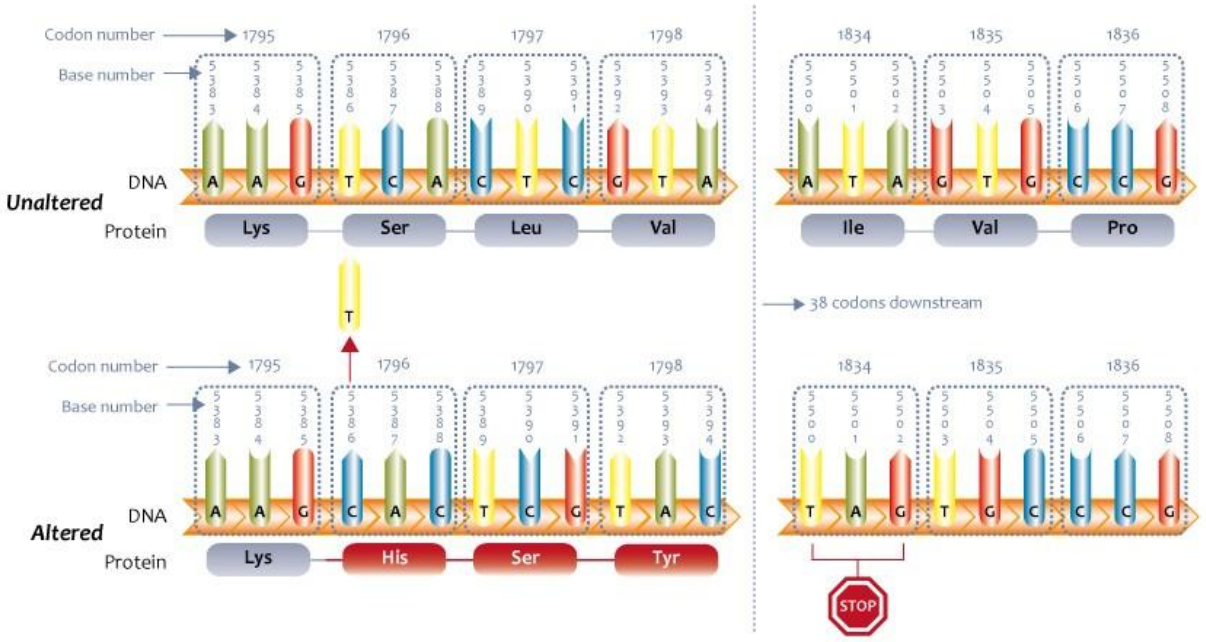
A point substitution mutation results in a change in a single nucleotide and can be either synonymous or nonsynonymous.

- A *synonymous* substitution replaces a codon with another codon that codes for the *same* amino acid, so that the produced amino acid sequence is not modified. Synonymous mutations occur due to the degenerate nature of the genetic code.
- A *nonsynonymous* substitution replaces a codon with another codon that codes for a *different* amino acid, so that the produced amino acid sequence is modified. Nonsynonymous substitutions can be classified as nonsense or missense mutations:
  - A *missense* mutation changes a nucleotide to cause substitution of a different amino acid.
  - A *nonsense* mutation is a point mutation in a sequence of DNA that results in a premature stop codon, or a nonsense codon in the transcribed mRNA, and possibly a truncated, protein product.

- Missense and nonsense mutations can lead to loss or change or loss of the resulting protein function. Sometimes, however, the protein function is fully retained.

Frameshift mutations

- Due to the triplet nature of gene expression by codons, the insertion or deletion can change the reading frame resulting in a completely different translation from the original.



- If a deletion or an insertion consists of 3 nucleotides then the result on the protein product will be a missed or added amino acid. Again, it can either lead to the loss or change of the protein function or have no influence on it. Same could be said of deletions or insertions of nucleotide number that is evenly divisible by three.