I. Properties of the universal genetic code

Yanovsky studied mutations in *E. coli trpA* and found that positions of amino acid alterations were **co-linear** with the genetic map of the corresponding mutations:

5' – nucleotide sequence – 3'

↓

H_2N – amino acid sequence – COOH

This type of experiment also shows that the genetic code is **non-overlapping**: most mutations alter only a single amino acid.

How many bases specify each amino acid? Crick et al. studied many different mutations due to addition of a single base pair (+1) or deletion of a single base pair (-1) . They then made new alleles that were combinations of these mutations.

+1 combined with –1 sometimes restores a normal phenotype
+1 combined with +1 never restores a normal phenotype
–1 combined with –1 never restores a normal phenotype
combinations of three +1 mutations sometimes restores a normal phenotype
combinations of three –1 mutations sometimes restores a normal phenotype

They concluded that the code must be made of **triplets**. Each of the 64 possible triplets is called a **codon**.

II. Deciphering the genetic code

Nirenberg synthesized polyU and mixed it with *E. coli* extract; poly-phenylalanine was produced, so he concluded UUU is the codon for phenylalanine. By synthesizing RNA with precise mixtures of nucleotides (*e.g.*, 2U:1C), the entire code was determined.

Brenner was the first to determine a **stop codon** (UAG, often called amber codon), UGA and UAA are also stop codons. There are no spaces or commas in the code.

AUG is the start codon. It encodes methionine, so all translation begins with met (which is often removed from the polypeptide after synthesis). AUG can also occur *within* a translated sequence, so met can also occur within a polypeptide.

There are 64 codons, but only 20 amino acids – most amino acids are encoded by more than one different codon. This is referred to as **degeneracy** in the genetic code. Degeneracy is usually in the third position; Crick proposed that this position is less critical during translation, a proposal known as the **wobble hypothesis**.

Note that codons refer only to the RNA sequences that are translated into amino acid sequence (or signal the stop). The part of the mRNA 5' to the initiating AUG is called **5' untranslated** region (5' UTR or 5' UT); the part of the mRNA 3' to the stop codon is the **3' UTR** (or 3' UT). Promoters do not have codons.

III. Mutations that affect coding sequences

Mutations are sequence changes in the DNA. Mutations in protein-coding regions may affect amino acid sequence.

AUGCCUCAAUUGUAG
met pro gln leu STOP

missense – a single nucleotide substitution that alters a single amino acid; may or may not affect protein function:

AUGC**A**UCAAUUGUAG
met **his** gln leu STOP

silent – a single nucleotide substitution that does not alter any amino acid, due to degeneracy of the code; no effect on function:

AUGCC**G**CAAUUGUAG
met pro gln leu STOP

nonsense – a single nucleotide substitution that changes a codon for an amino acid to a stop codon; usually obliterates protein function:

AUGCCU**U**AAUUGUAG
met pro **STOP**

frameshift – an insertion or deletion of one or more nucleotides (but not a multiple of three) that changes the reading frame; usually annihilates protein function:

AUGCCC**C**UCAAUUGUAG
met pro **ser** **ile** **val**