CONCEPT: BIOINFORMATICS

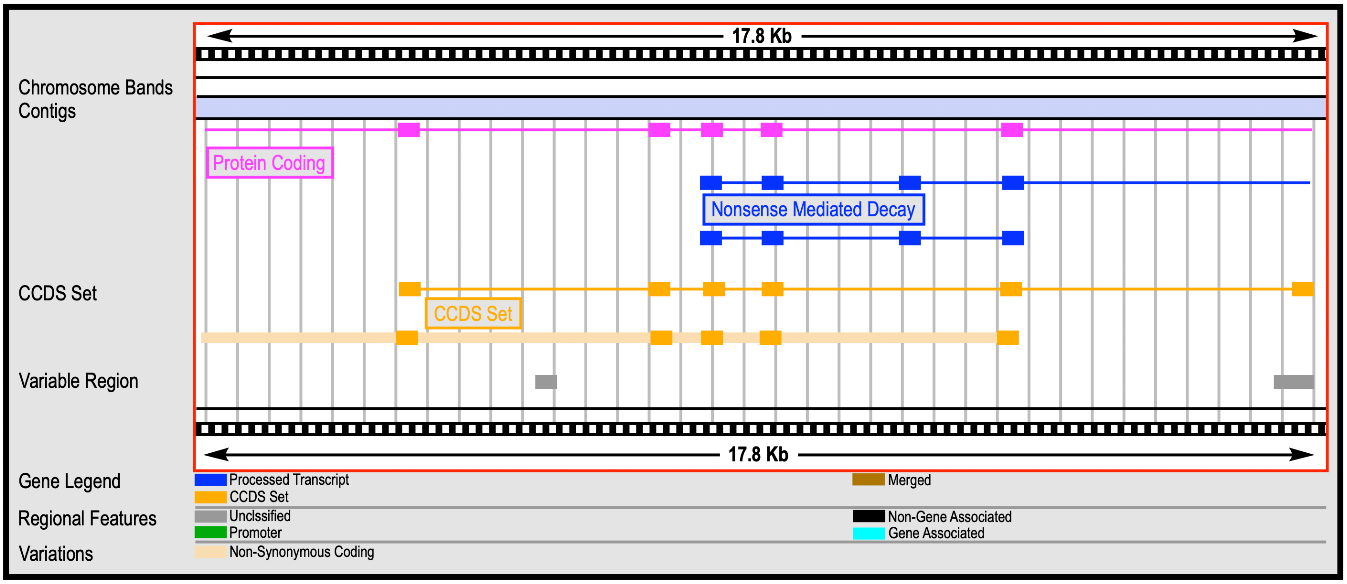
● **Bioinformatics** is the study of the information found within _________________________

□ What information does the genome hold?

- Genes, RNAs, binding sites for proteins, non-coding RNAs, positions for gene regulation

- **Annotation** marks these functional elements of the genome

EXAMPLE:



□ Bioinformatics can be used to determine protein-encoding genes from the genomic sequence:

- **Proteome** is the inventory of all proteins encoded by an organism's genome

- **Open-reading frames (ORFs)** are sequences with characteristics of typical genes

    - 5' and 3' end sequences, introns, exons

- **Codon bias** is when an organism prefers one codon over other codons for the same amino acids

    - *Drosophila* uses UGC 73% and UGU 27% to code for cysteine

□ **cDNA** sequences can be used to __________________- ORFs

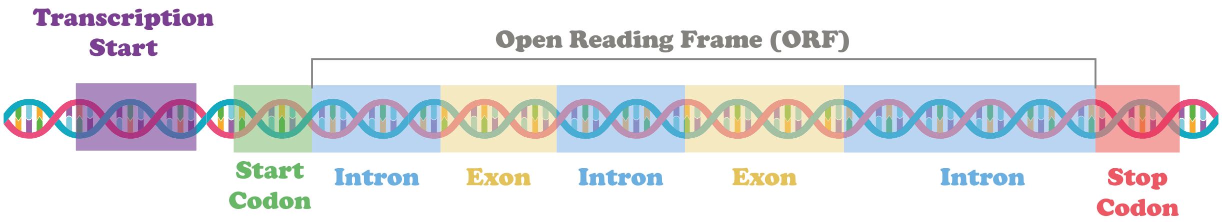
- cDNA is DNA reverse transcribed from RNA

- Introns are removed

- **Expressed sequence tags (ESTs)** are large data sets of short cDNA sequences

    - Made to determine gene boundaries

EXAMPLE:



☐ Bioinformatics can be used to ________________________ DNA binding sites

- Computer software searches the genomic sequence for predicted sequences (promoters, splice sites, etc)

☐ Bioinformatics can be used to study evolution and DNA similarity

- **BLAST** searches can be used to determine if a particular sequence is similar to other known sequences

EXAMPLE:

PRACTICE:

1. Which of the following is NOT a piece of information that bioinformatics can analyze?
    a. Location of DNA-Protein binding sites
    b. Identifying all the proteins expressed in a skin cell
    c. A list of all introns in the genome
    d. The function of one gene

2. Which of the following can be used to identify an open-reading frame?
    a. cDNA sequences
    b. Introns
    c. Enhancer locations
    d. Exons