# Genetic code/ Triplet code

- There are 4 bases A,U,G, C
- For a triplet code, how many combinations are possible?

- Protein is made of 20 different amino acids

# Types of mutations

- Substitution- Exchanges one base for another

- Insertion- Insertion of extra base pairs

- Deletion- Deletion of a base pair or sections of DNA

- Frame shift- Insertions or deletions resulting in altered proteins due to a shift in the frames.

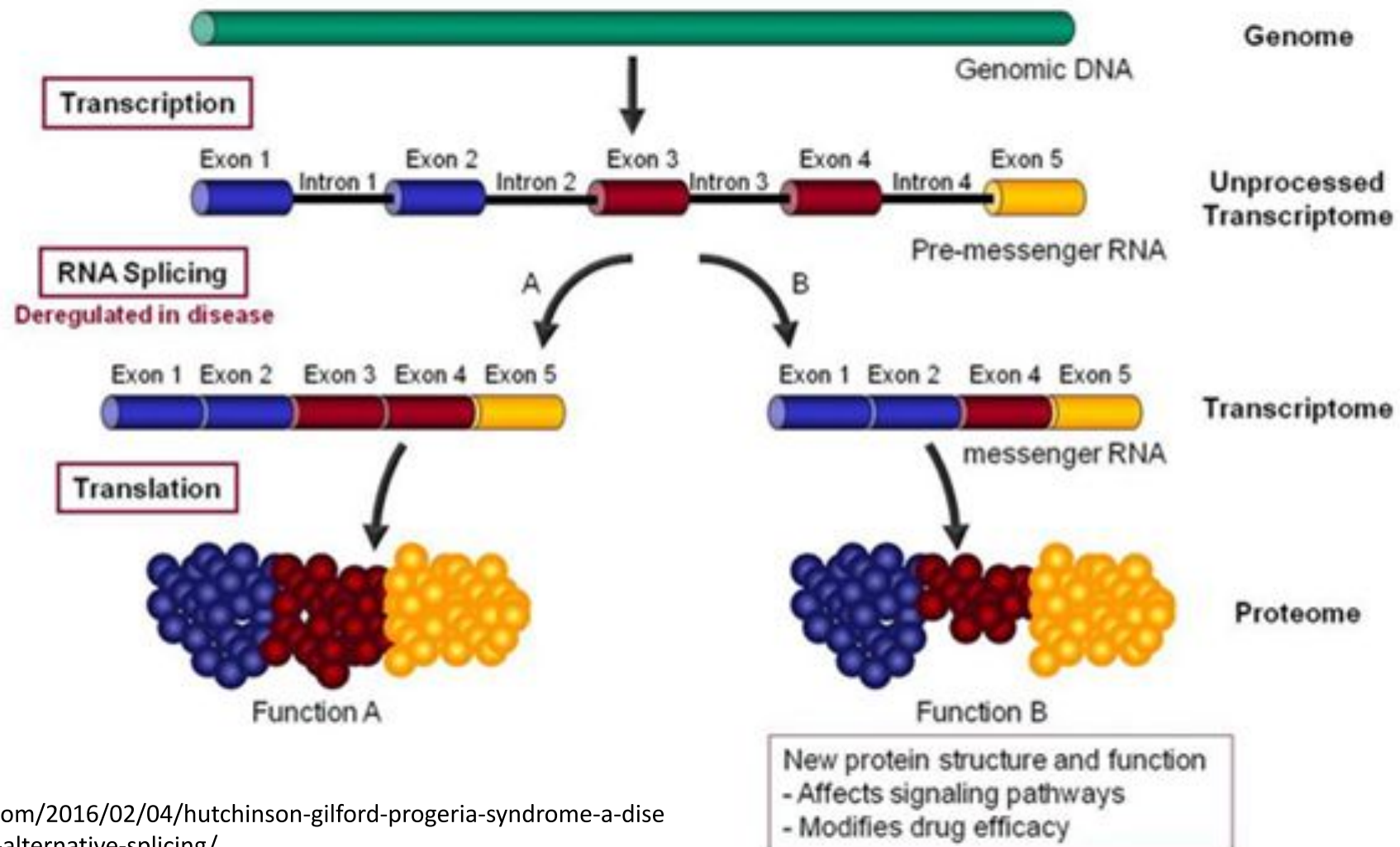- There are other types but the above ones are the basic.

# Causes of mutation

- DNA fails to copy accurately

- External influences can create mutations- harmful chemicals and radiation
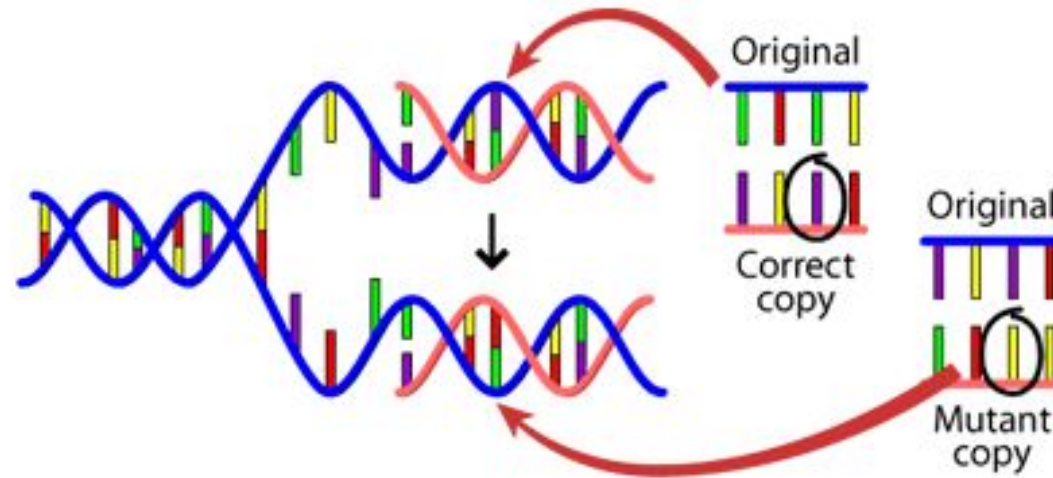


- Cell can repair- But is not perfect

# Hutchinson-Gilford Progeria Syndrome – a disease of accelerated aging due to Alternative Splicing

"HGPS is a major clue to solving the "puzzle of aging" and the molecular mechanisms here are relevant to normal aging."

# Databases of Biological Data

**General Repositories**

      **GenBank - USA**

      **EMBL - Europe: http://www.ebi.ac.uk/embl/**

**Specialized by data type**

**NCBI** Trace Archive - raw reads from sequencing machines

 **SwissProt** - curated protein information: http://www.expasy.org

**KEGG** - metabolic pathways: http://www.genome.jp/kegg/

**Gene Expression Omnibus (GEO)** - gene expression

 **PDB** - protein structures

**Specialized by organism**

      **ZFIN** - zebrafish

      **SGD** - yeast

      **WormBase** - c. elegans

      **FlyBase** - fruit fly

# Primary Database

In bioinformatics, databases are often categorised as primary or secondary .
**Primary databases** are populated with
experimentally derived data such as nucleotide sequence, protein sequence or macromolecular structure.
Experimental results are submitted directly into the database by researchers, and the data are essentially archival in nature.

Once given a database *accession* number, the data in primary databases are never changed: they form part of the scientific record.

# Secondary Database

- Comprise data derived from the results of analysing primary data.

- Secondary databases often draw upon information from numerous sources, including other databases (primary and secondary), controlled vocabularies and the scientific literature. They are highly curated, often using a complex combination of computational algorithms and manual analysis and interpretation to derive new knowledge from the public record of science.

- Secondary databases have become the molecular biologist's reference library over the past decade or so, providing a wealth of (often daunting) information on just about any gene or gene product that has been investigated by the research community. The potential for mining this information to make new discoveries is vast.

# Genbank

https://www.ncbi.nlm.nih.gov/genbank/statistics/

First genome sequenced in 1995
(the bacteria *H. influenzae* with a
genome of 1,830,140 letters).



Growth of GenBank (1982 - 2005)
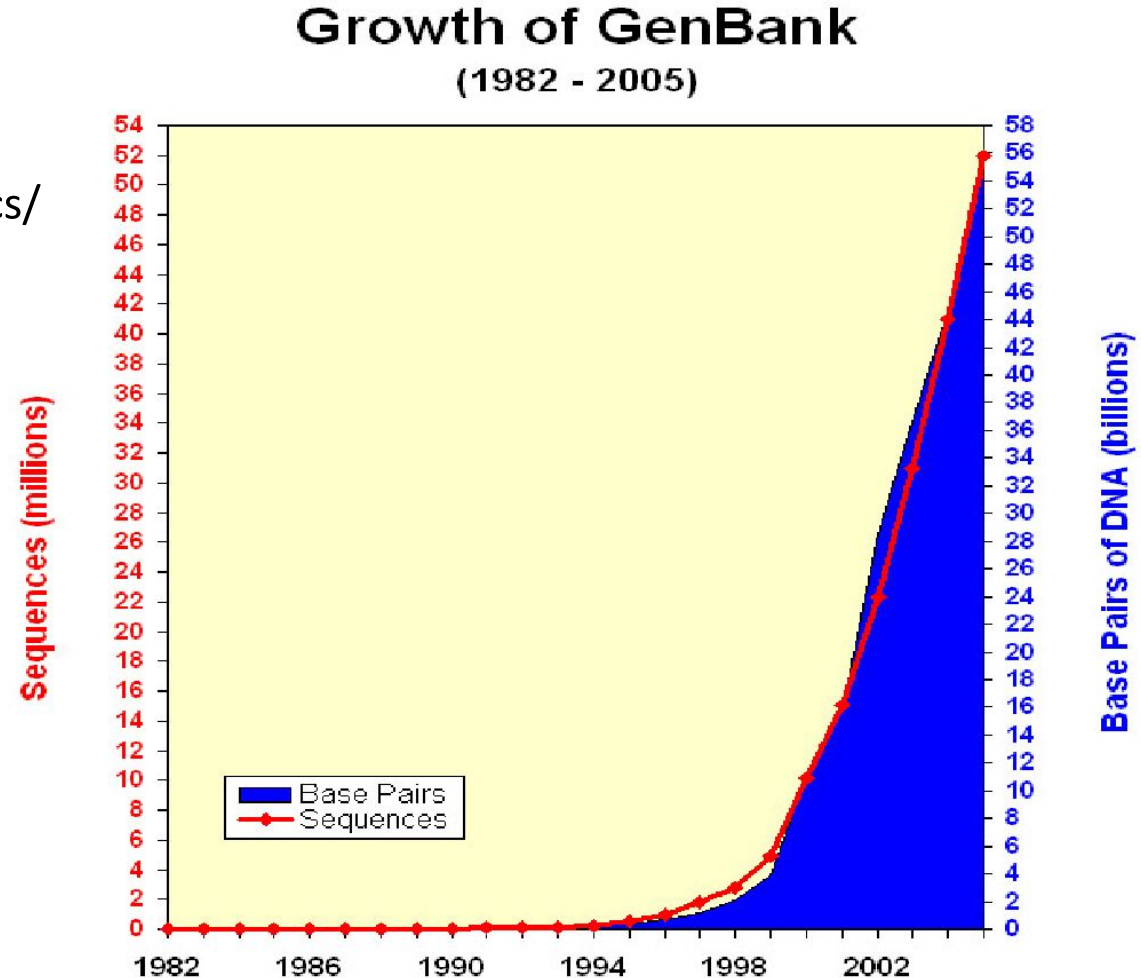
# Genbank website

# Worksheet - example

Databases