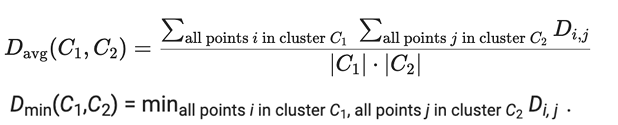
**What is** [**genome**](https://youtu.be/3fr4jBFs25s?si=Lt-QfKw-8cz_ktGz) (*https://youtu.be/3fr4jBFs25s?si=Lt-QfKw-8cz\_ktGz*).

1. humans have almost 20,000 genes.
2. only 2 percent of genes carry genetic information (info to build proteins) and these are called coding gene.
3. remaining 98 percent is called non-coding gene.
4. DNA is mainly located in nucleus and few of is also found in mitochondrial (powerhouse of cell).
5. all the DNA in a cell is called genome.
6. nearly 99.8% of genome is identical in every human being.

**What is** [**replication**](https://youtu.be/TNKWgcFPHqw?si=ZjUN311T1d-xHWj2)(*https://youtu.be/TNKWgcFPHqw?si=ZjUN311T1d-xHWj2*).

1. two ends of strand is called 5' and 3'. both strands run in opposite direction. the corresponding end of 3' in other strand is 5' and vice versa.
2. both strands are separated by a enzyme called helicase.
3. primaze, an enzyme creates RNA part called primer, which marks the starting point of construction of new strand of DNA.
4. a enzyme called dna polymerizer bind to the primer and makes the DNA (the new strand to complement).
5. dna polymerizer can only add in one direction (5' to 3'). so the construction of new strand for the leading strand is done continuous.
6. but the other strand (lagging strand) is in opposite direction (3' to 5'). and dna polymerizer can't work in this direction.
7. so the lagging strand's new strand is make in chunks called okazaki fragments, (a chunk is selected from 3' to 5' and new dna strand is made in 5' to 3' direction, 'reverse direction in that chunk').
8. each chunk is started with primer and then followed by dna polymerizer.
9. once its all complete the enzyme call exonuclease removes all the primers (RNA) from both strands, then these gaps are filled by dna polymerizer.
10. finally, enzyme dna ligase seals up to form continuous strands.

* K centre:
  + Minimize max distance (farthest datapoint from its nearest centre)
* K means:
  + Minimize distortion (squared mean of distance of all datapoints from their nearest centres)
* K means++:
  + Probability of initial centre choosing is proportional to distance between already choosed centres.
* Hierarchical clustering:
  + Single link
    - A pair from both cluseters with minimum distance defines the distance between clusters
    - Average of all points define the distance between clusters



* Primary
  + Seq of amino acids
  + Repeating substructure that forms as a sub structure of overall folded protein
  + Final shape after it has folded and is stable
  + Copies of stable folded protein, interaction with copies of itself to form a single functional or multimer.