Molecular Biology

2025-2024

DR.Ahmad Al Qawasmi

ARKAN ACADEMY

Human Genome

- Human genome project (1990) multinational project led by USA costed 3 billion dollars and 13 years
  - ➢ It led to publishing a draft in 2001 and then 92% of genome is sequenced in 2004
  - ➢ Its major outcomes were determination of the number of human genes, development of major technologies and bioinformatic tools, completed sequences of other genomes and open discussion of legal and ethical issues
    - ✓ The mouse has a genome of 40 chromosome doublets with 20,000 gene compared to human having 46 chromosome doublets with 20,000 genes also
    - ✓ The number of nucleotides of a genome do not indicates its complexity
    - ✓ Mitochondrial genome (encode 13 protein) is similar to the bacterial genome (single circular)

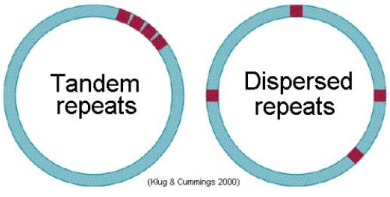
- ENCODE project (2003): Encyclopedia of DNA Elements
  - ➢ *~75% of the entire human genome is relevant* (either transcribed, binds to regulatory proteins, or is associated with some other biochemical activity)
  - ➢ The last chromosome to be sequenced is *Y chromosome* due to the large number of repetitive sequences, which is finally done in 23 august 2023

Summary of ENCODE Results	
Protein-coding genes	20,687
Short noncoding RNAs	8801
Long noncoding RNAs	9640
Pseudogenes	11,224
Percentage of genome transcribed into RNA	74.7%
Percentage of genome-binding transcription factors	8.1%

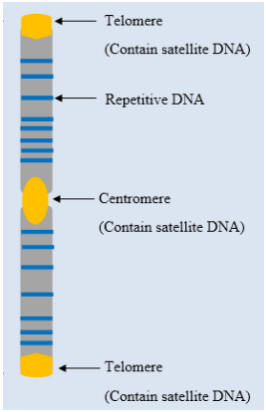
- Human Genome consists of *nuclear* (3300 Mb) and *Mitochondrial* (16.6 Kb) genomes
- Nuclear genome consists of:
  - ➢ ***Genes*** and gene related sequences (***30 %***)
    - ✓ *Non-Coding Sequences (30 %)* which do not give proteins, including:
      - Introns (30 %)
      - Pseudogenes and other sequences such as untranslated regions
    - ✓ *Protein-Coding Sequences (1.2 %)*

~5% of the genome contains sequences of noncoding DNA that are highly conserved (critical to survival).

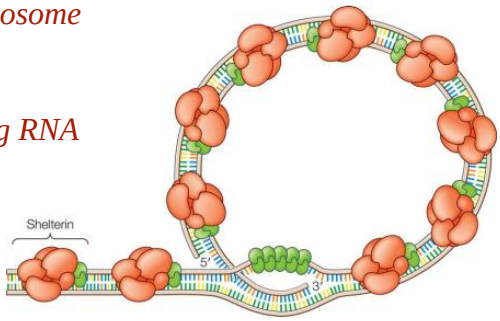
  - ➢ ***Extragenic and Intergenic sequences (70 %)***
    - ✓ Unique or low copy number (10 %)
    - ✓ Moderate to highly repetitive (55 %)
      - ***Tandem Repeats (10%)***
        - *Satellites (Macro-satellite)* in the centromere and telomere
        - *Mini-satellites (VNTR)*
        - *Micro-satellites (STRs)*
      - ***Interspersed Repeats (45 %)***
        - *DNA transposons (3 %)*
        - *Retrotransposons (42 %)*
    - ✓ Others (5 %)



Tandem repeats



- ***Satellite (macro-satellite) DNA:*** Regions of *5-300 bp* repeated *106 – 107 times*
  - Mainly in ***centromeres*** and ***telomeres***
  - *α-satellite: Centromeric A/T-rich* repeats (171 bp) *unique* to each chromosome
    - ✓ Fluorescence in situ hybridization (**FISH**) can use chromosome specific probes to detect them
  - *Telomeric repeats: TTAGGG* repeated hundreds to thousands of times at the *termini of human* chromosomes with a *3′ overhang* of single-stranded DNA
    - ✓ The repeated sequences form loops that bind to shelterin
    - ✓ *Shelterin:* A protein complex binds and *protects the chromosome termini* from degradation

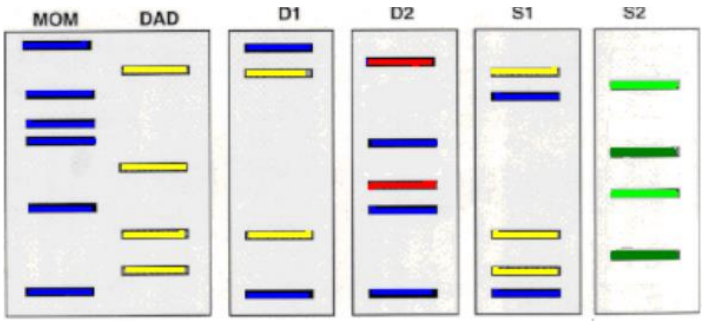


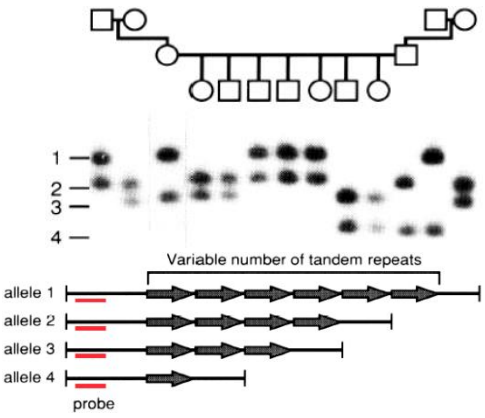
  - *Telomeric repeat-containing RNA (TERRA):* a *long non-coding RNA* transcribed from telomeres:
    - ✓ Maintaining the ***integrity*** of chromosome termini
    - ✓ Regulating ***telomerase activity***
    - ✓ Maintaining the ***heterochromatic*** state of telomeres
    - ✓ Protecting DNA from ***deterioration or fusion*** with neighboring chromosomes

- ***Mini-Satellite Sequences***
  - Variable number of tandem repeats (***VNTRs***) of *20 – 100 bp* repeated *20-50 times*
  - They differ mainly in the ***number of repeats*** between individuals

- ***Micro-Satellite Sequences***
  - Also called ***STRs*** (Short Tandem Repeats), and SSR (Simple Sequence Repeats)
  - They are sequences of *2 – 10 bp* repeated *10-100 times*

- ***Polymorphism:*** it is the **variation** in the sequence of DNA between individuals, and it is caused by:
  - *Polymorphisms of VNTR and STR*
    - ✓ The number of STRs and VNTRs are highly variable among individuals (polymorphic)
    - ✓ They are useful in DNA profiling for ***forensic testing*** and paternity tests





- *Single nucleotide polymorphism (SNPs):* Single-nucleotide substitutions of one base for another
  - ✓ 2 or more versions of a sequence must be present in at least one percent of the general population
  - ✓ SNPs occur throughout the human genome - about one in every 300 nucleotide base pairs which means around 10 million SNPs within the 3 billion nucleotide human genome
    - Only 500,000 SNPs are thought to be relevant
  - ✓ Types of SNPs:
    1. ***Linked SNPs:*** located *outside* the gene, and do not affect protein amount or function
    2. ***Cuasative SNPs:*** located *inside* the gene
       - *Non-coding SNP:* Located in the *regulatory region* (affecting gene **expression**), and changes the **amount** of protein produced
       - *Coding SNP:* Located in the *coding region*, and changes the **amino acid sequence** (**protein type**) affecting the **function** of the protein

Interspersed repeats

- ***Transposons (jumping genes):*** They are segments of DNA that can **move** from their original position in the genome to a new location
- Classes of transposons:
  - *DNA transposons* (*3%* of human genome)
  - *RNA transposons or retrotransposons* (*42%* of human genome)
    - ✓ Long interspersed elements (*LINEs*, 21%)
      - Example is *L1*
    - ✓ Short interspersed elements (*SINEs*, 13%)
      - Example is *Alu* (300 bp) which can integrate into a coding region
    - ✓ *Retrovirus-like elements* (8%)

- A retrotransposon present at one site in chromosomal ***DNA*** is transcribed into ***RNA***, then RNA is converted back into DNA by reverse transcriptase where the ***retrotransposon DNA*** can then integrate into a ***new chromosomal site***
- LINEs contain ***reverse transcriptase*** genes and the integrase gene that is necessary for integration into cellular DNA

- Over 99% of the transposons in the human genome ***lost their ability to move***, but we still have some active transposable elements that can sometimes cause disease
  - Examples of these diseases: Hemophilia A and B, severe combined immunodeficiency, porphyria, predisposition to cancer, and Duchenne muscular dystrophy

- Transposons are still active in some organisms such as pigs

Past Papers

1. What is the function of reverse transcriptase?
   A. Synthesizes RNA from DNA
   B. Transcribes DNA from RNA
   C. Synthesizes DNA from RNA
   D. Transcribes RNA from protein

2. VNTR alleles are hypervariable regions of human DNA that differ from each other in:
   A. location of internal sites recognized by restriction enzymes
   B. variable number of point mutations
   C. number of copies of an internally repeated DNA sequence
   D. variable location on different chromosomes

3. Which of the following is incorrect?
   A. Alu is an example of a SINE
   B. L1 is an example of a LINE
   C. Tandem repeats are found more than interspersed elements in the human genome
   D. 2% of the human genome is protein coding gene exons
   E. All of the above are correct

4. Which of the following are examples of Satellites?
   A. Centromeres
   B. Telomeres
   C. Alu
   D. A+B

5. Which of the following statements about genomes isn't correct?
   A. Almost 20% of the human genome isn't relevant
   B. Greater number of nucleotides per genome indicates higher complexity
   C. It's not a necessity to have variations of mini-satellite sequences between different alleles of the same individual
   D. B&C

6. Which of the following statements about transposons is incorrect?
   A. RNA transposons amounts are fewer than DNA transposons within human genome
   B. 99% of transposons within human genome lost their ability of movement
   C. All the possible movements of transposons cause diseases
   D. A & C

7. Retrotransposons represent this percentage of our genome:
   A. 30%
   B. 10%
   C. 5%
   D. 45%
   E. 21%

8. If you knew someone was genetically predisposed to have a certain disease, the type of SNP you expect to be responsible for this disease is:

   A. Linked SNP
   B. Causative SNP
   C. Cannot be determined
   D. Neither of the above
   E. Both of the above

9. *One of the following is TRUE in regards to the protein shelterin:*

   A. It is synthesized from centromeres
   B. It creates Barr bodies of one of the X-chromosomes
   C. It binds to telomeres protecting them from degradation
   D. It converts euchromatin to heterochromatin
   E. It increases the stability of centromeres

ARKAN

ACADEMY

علم في كل مكان

Arkan academy

www.arkan-academy.com

Arkanacademy

+962 790408805