HEALTHCARE: GENETIC SEQUENCING

Members :→

1. SHUBHAM KUMAR — 23122033
2. PARUL SHARMA — 23122026
3. [redacted]LERAN ANTHONY CARVALHO — 23122021

HEALTHCARE: GENETIC SEQUENCING REPORT

1. INTRODUCTION

1.1. BACKGROUND

Genomic data analysis plays a crucial role in modern biological research and healthcare. Understanding genetic information helps scientists and clinicians make informed decision about diseases, treatments and more.

This project aims to perform basic genomic data analysis using python on a simple dataset to demonstrate fundamental techniques and workflow in this field.

1.2. PROJECT SCOPE

This project aims to perform comprehensive data analysis on a randomly generated dataset.

The scope of the project is:

1. Data analysis
2. Basic statistical analysis
3. Interpretation of result

2. CODE EXPLANATION

This is a menu driven program in which we can choose the following options:

1. Generate random genomics data
2. Perform analysis on genomic data
   a) Compare reference bases and variation bases.
   b) Calculate reference percentage of homozygous allele and heterozygous allele
   c) Calculate genetic density
3. Generate random dna sequencing data
4. Perform analysis on the sequencing data
   a) determine the gc-content

This code includes various functions which specifically makes use of different looping statements and conditional statements to generate specific results.

The code includes the following functions:

1) compare_ref_var:

This function compares the values of reference base and variation base. This function makes use of the conditional statements to compare both the values.

The function take the data as an argument which is generated previously.

If both the values are same, the function prints "The gene is a Homozygous allle." else the function prints "The gene is a Heterozygous allle".

This function returns two lists: homo_list and hetero_list.

2) calc_percentage:

This function calculates the percentage of homozygous alleles and heterozygous alleles for the individuals present in the randomly generated data.

This function take homo_list and hetero_list as its arguments and returns the percentage for both.

3) genetic_density:

This function calculates the genetic density based on the generated data. The function specifics the chromosome length and variants count and based on that calculates returns the value for genetic density.

4) generate_dna_sequence:

This function generates a random dna sequence based on the length of the chromosomes that can be inputed by the user.

5) gc_content: calculates the gc_ content.

3. INSIGHTS

This project mainly focuses on the following insights:

(1) Comparing Reference Base and Variation Base

The "Variant Base" column specifies the observed genetic variance or variation at the specified genomic position for an individual. It represents a deviation from the reference base.

(2) Calculating percentage of each nucleotide present in the Chromosome

It can be used to determine which chromosome have the most or least genetic variations, which can be indicative of genomic regions with higher or lower mutation rates.

(3) Calculating Genetic Density

Genetic Density is a valuable metric in genomics and genetics research because it helps researchers understand how closely genetic variations are distributed along a chromosome or genomic region.

High genetic density in a specific region of a chromosome suggests that this region contains a higher concentration of genetic variations.

(4) Variant Counts

To understand the overall genetic diversity within a group of individuals.

REMARK:

All the important terms are explained in the comments of the code provided for the ease of understanding

4. CONCLUSIONS

In Summary, genomics data analysis has the potential to advance our understanding of genetics, biology and medicine, with practical applications in healthcare, research, and the development of novel therapeutics.

It can contribute to the identification of genetic variations associated with specific diseases or traits, aiding in the development of diagnostics and targeted therapies.

Other uses of analysing genomics data:

- Tailoring drug treatments based on an individual's genetic profile to enhance treatment efficacy and minimize side effects.
- Understanding genetic diversity within and between populations.