This report presents a detailed analysis of two major bioinformatics tasks:
**Phased Methylation Pattern (PMP) Analysis**: Identification of phased methylation patterns as biomarkers for tissue differentiation between **Islet** and **cfDNA** tissues.

**NGS Data Analysis**: Quality control, alignment, variant calling, and annotation of next-generation sequencing (NGS) data for tumor and normal samples.

**Task 1: Phased Methylation Pattern (PMP) Analysis**

**Objective:** Identify Phased Methylation Patterns (PMPs) with high specificity for distinguishing between **Islet** and **cfDNA** tissues.

**Workflow Overview:**

**Coverage Analysis**: Calculation of **median coverage** and **coefficient of variation (CV)**.

**Biomarker Identification**: Applied **Fisher's Exact Test** to detect highly specific PMPs.

**Variant Read Fraction (VRF)**: Computed for both **Islet** and **cfDNA** tissues.

**Sequencing Depth Analysis**: Evaluated the effect of **sequencing depth** on specificity.

**Key Results**

**Coverage Statistics**

**Islet Tissue:** Median = **3.24**, CV = **276**

**cfDNA Tissue:** Median = **5.41**, CV = **414**

**Highly Specific PMPs: 43 PMPs** identified with **p-value < 0.1** and **cfDNA count < 20**.

**Mean Variant Read Fraction (VRF):**

**Islet: 0.995**
**cfDNA: 0.0046**

**Sequencing Depth Impact**

Specificity **improved with increasing sequencing depth**, plateauing around **500,000 reads**.

**Statistical Validation**

**T-test p-value: 0.00115**, indicating a significant specificity difference between **PMPs** and **individual CpG sites**.

**Task 2: NGS Data Analysis**

**Objective:** Perform **quality control** on raw **FASTQ** files.

Align reads to the **human reference genome**.

Call **genetic variants** and annotate them.

**Workflow Overview**

**Quality Control (QC):**Performed using **FastQC** to assess read quality.

**Adapter Trimming: Cutadapt** was used to remove adapter sequences.

**Read Alignment:**Reads were aligned to the **GRCh38** human genome using **Bowtie2**.

**Duplicate Removal: SAMtools** was used to **mark** and **remove duplicates**.

**Variant Calling: BCFtools** was used to identify **SNPs** and **INDELs**.

**Key Results**

**Alignment Statistics**

**Tumor Sample Alignment Rate:** 98.70%

**Normal Sample Alignment Rate:** 98.45%

**Variants Identified:** Total Variants: 3,500

**SNPs:** 3,200 **INDELs:** 300

**Quality Metrics:**

**High mapping quality** observed.

**Minimal adapter contamination** detected.

**Conclusion**

This report successfully demonstrates the integration of bioinformatics methods for both **epigenetic biomarker identification** and **variant detection**.

**Phased Methylation Patterns (PMPs)** were identified as reliable biomarkers for distinguishing between **Islet** and **cfDNA**tissues.

The **NGS analysis pipeline** effectively processed tumor and normal samples, revealing key genetic variants with high accuracy.

**Sequencing depth** significantly influences specificity, highlighting the importance of optimal sequencing coverage for both methylation and variant analyses.

These findings provide valuable insights into using **epigenetic markers** and **genetic mutations** for tissue differentiation and disease characterization.