GenomeBeans

Sequencing Analysis Platform

Your goto bioinformatics service for multi-omic studies.

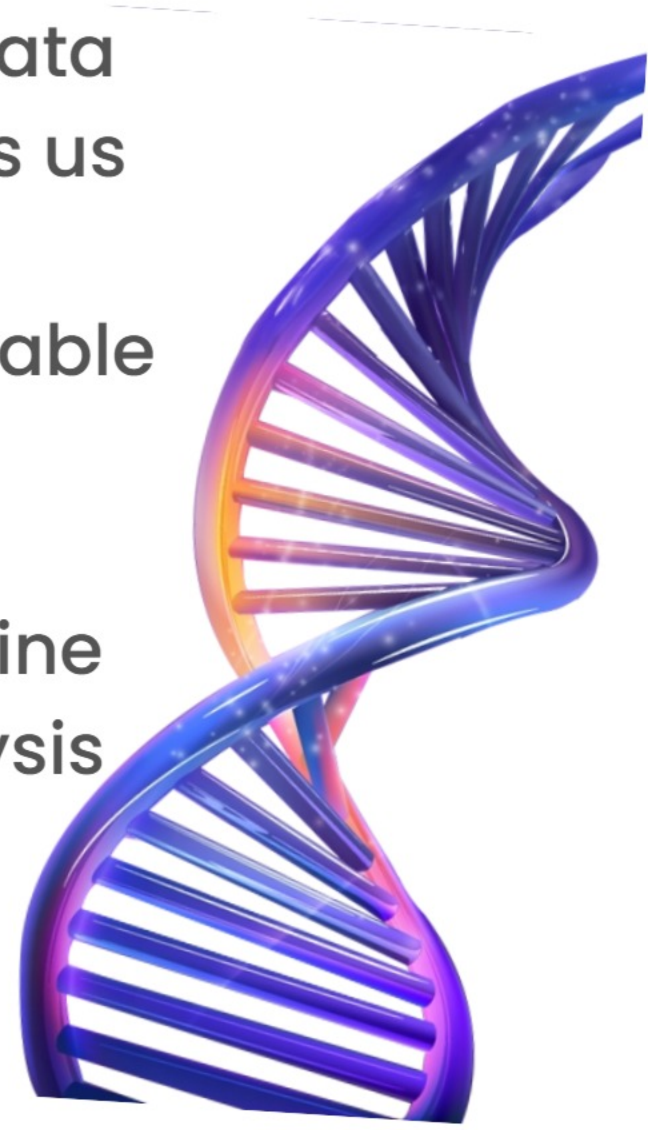
Introduction

Who Are We

- Determined to provide best solutions, making easier interpretation and decision making for drug development and clinical research.
- Passionate to generate coherent understanding of biological systems
- Commitment to uphold the highest standards of integrity, transparency, and ethics
- Unwavering core values that shape our culture and guide our every endeavor

What We Do (Vision)

- We offer customized high-quality solutions to the scientific community across the globe for their sequencing analysis requirements, without any expertise in programming.
- We believe in high data standards and understand data privacy being top priority among our clients to makes us their trusted collaborator for biological projects right from the raw complex data to useful and understandable data.
- Addressing the current advancement in AI and Machine Learning by incorporating the algorithms in our analysis for faster and more accurate results.



Solutions

Bulk RNASeq Analysis

- PreProcessing
- Generating Normalized Read Counts
- Visualization
  - Sample distribution PCA
  - Heatmap Differentially expressed genes
  - Network plot Gene Enrichment

Variant Analysis

- Somatic (tumor normal pair)
- Germline
- Visualization
  - Differentially Expressed Mutations
  - IGV/VCF viewer

Metagenomics Analysis

- 16s Analysis
- Shotgun data
  - Illumina data
  - ONT data
- Taxonomy Classification
- Binning
- Phenotype Annotation
- Visualization

Single Cell Analysis

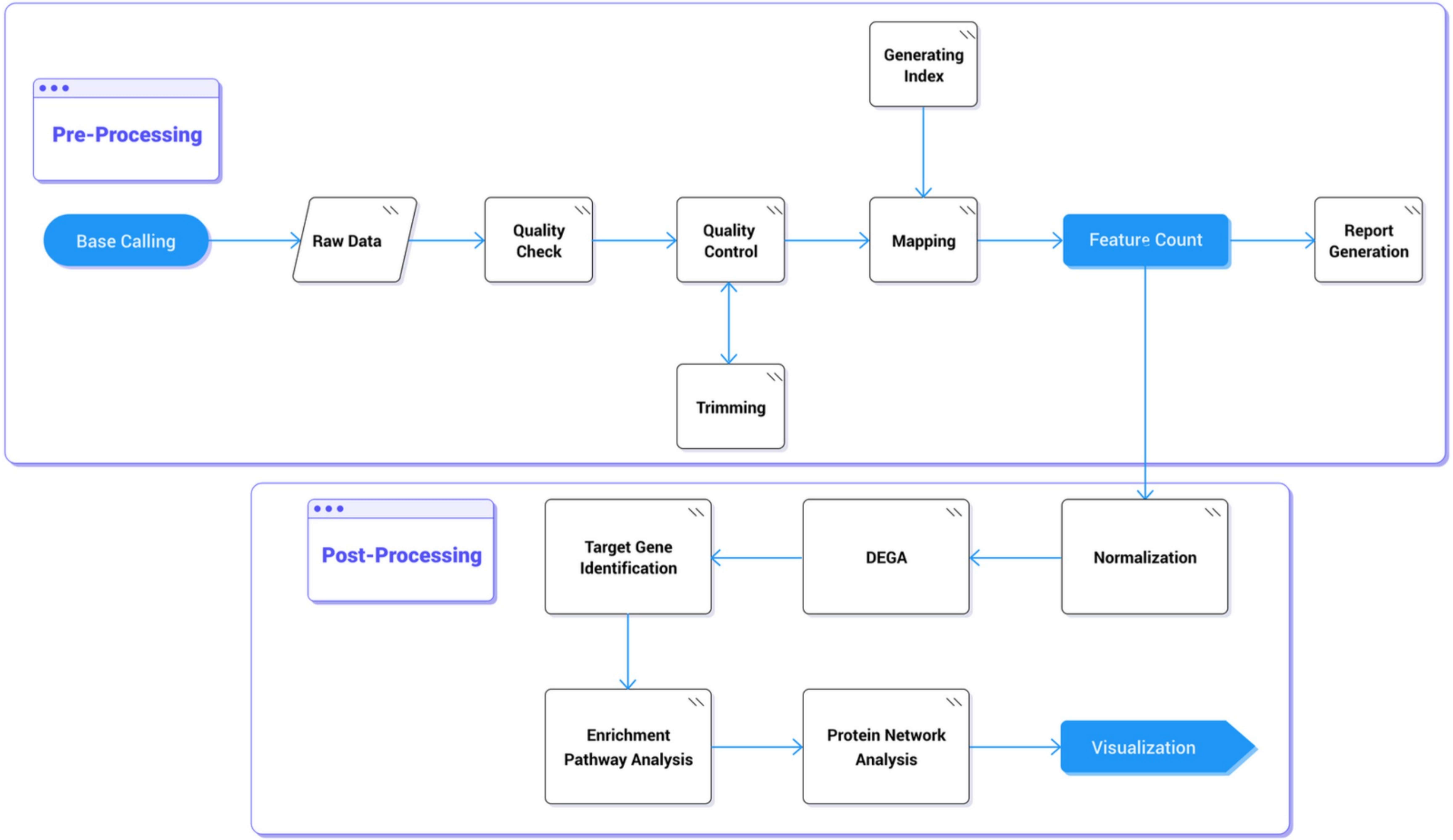
- RNASeq
- MethylSeq
- ATACSeq
- Visualization
  - Cell Cluster
  - Annotation
  - Dimensionality reduction

Immunomics Analysis

- PreProcessing
- AIRRtsv file
- Visualization
  - CDRH3 length distribution
  - Clonal lineage plot
  - Diversity Analysis
  - Repertoire overlap

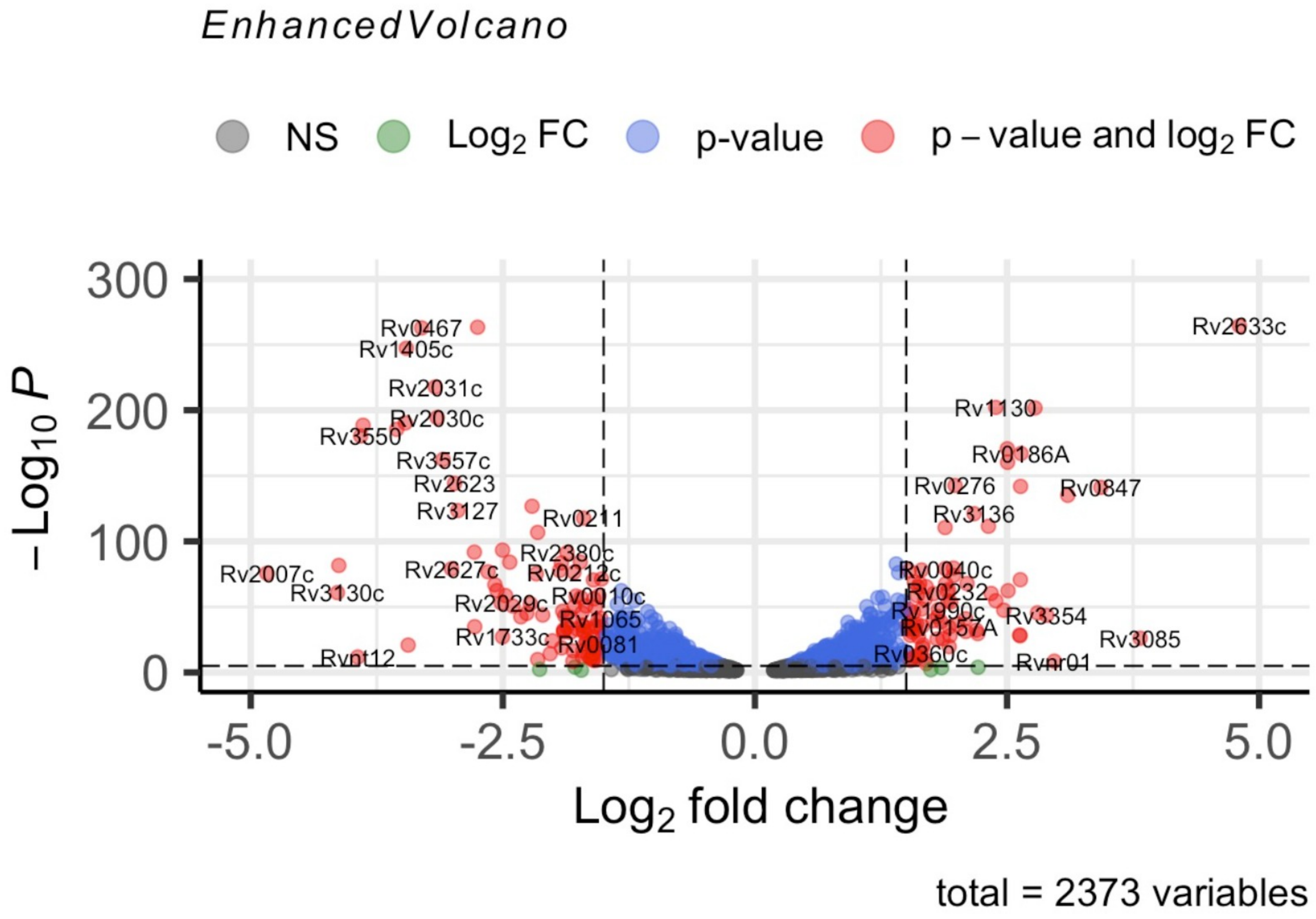
Bulk RNASeq

Workflow



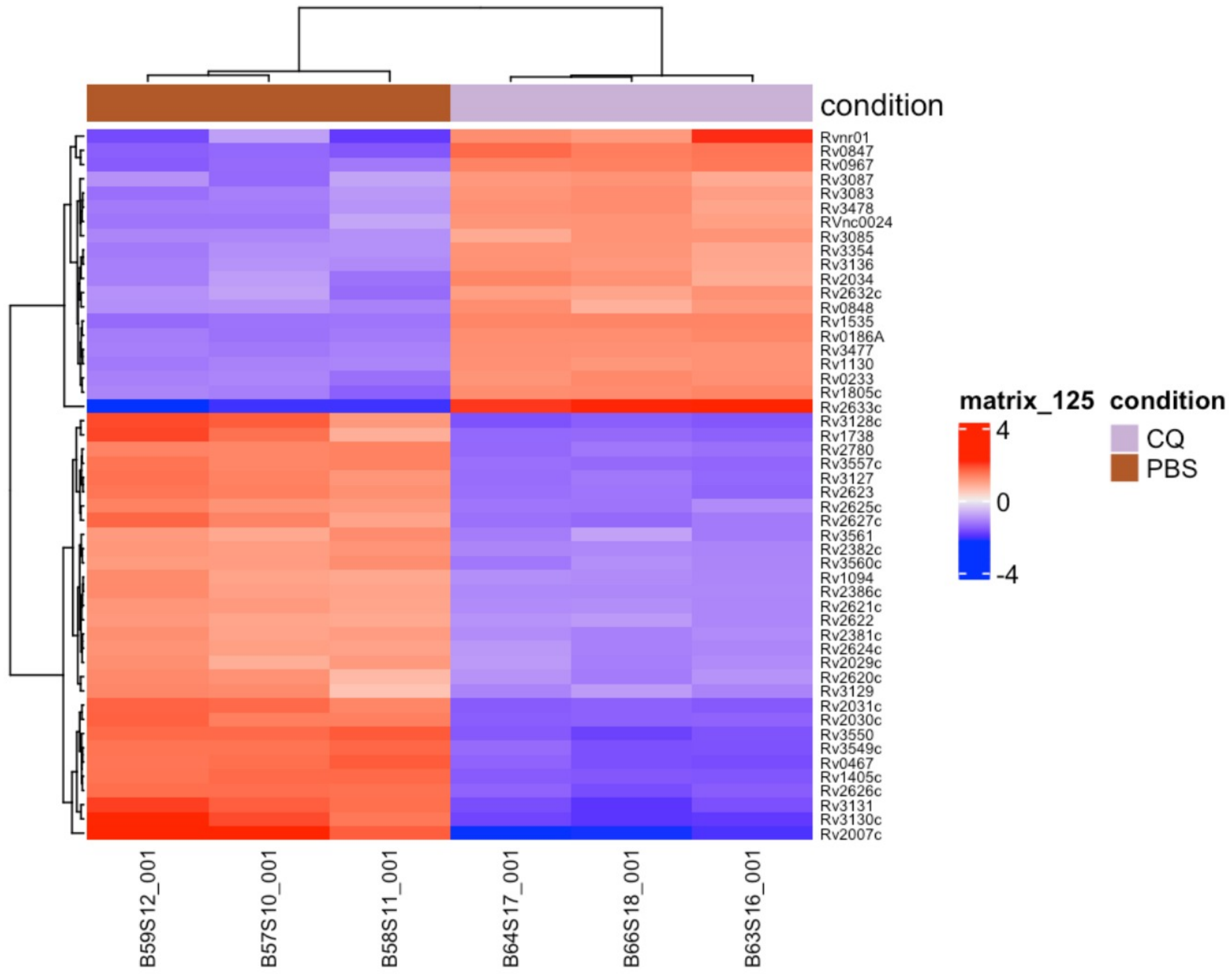
Bulk RNASeq

Mycobacter Case study



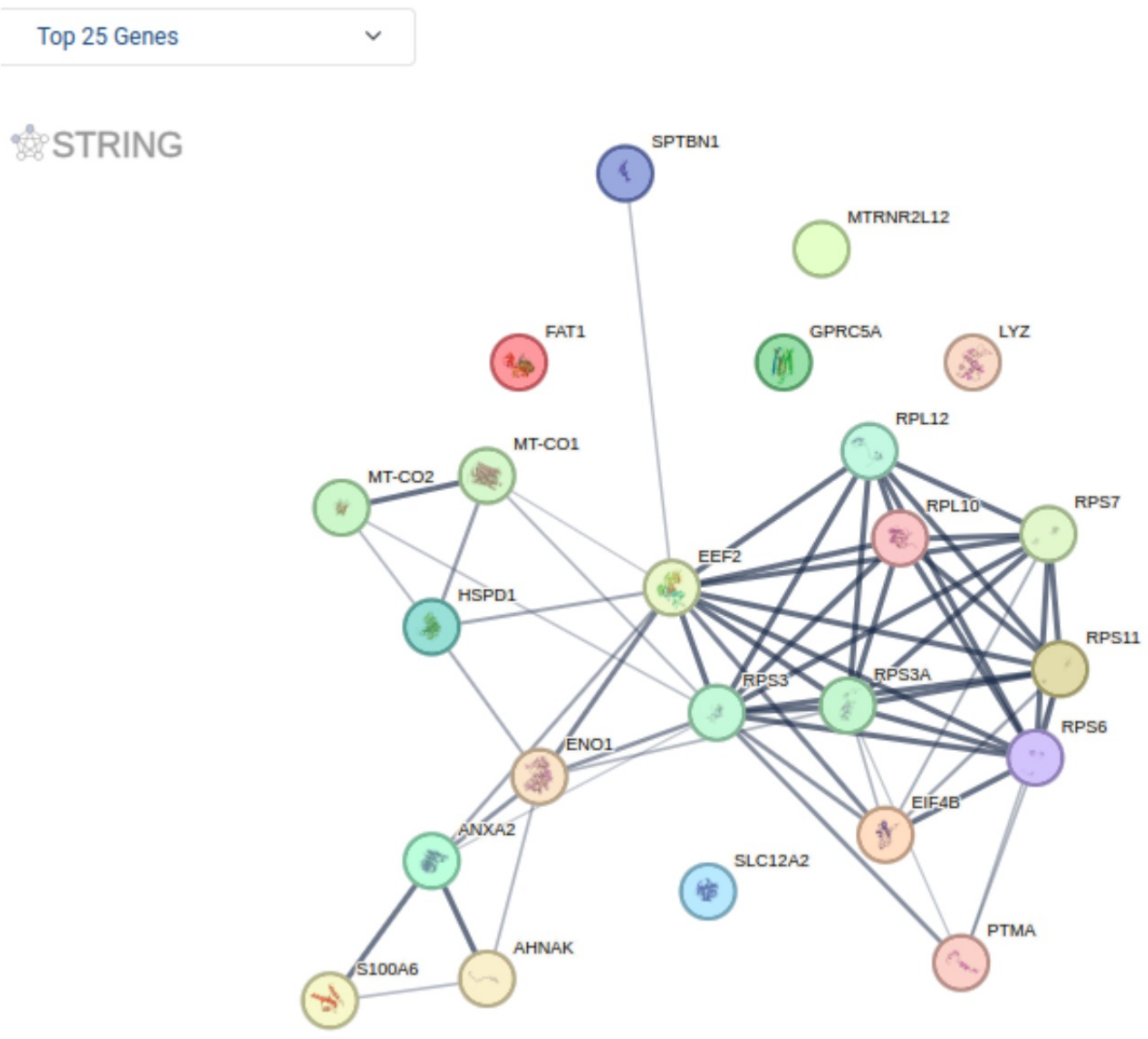
Volcano Plot

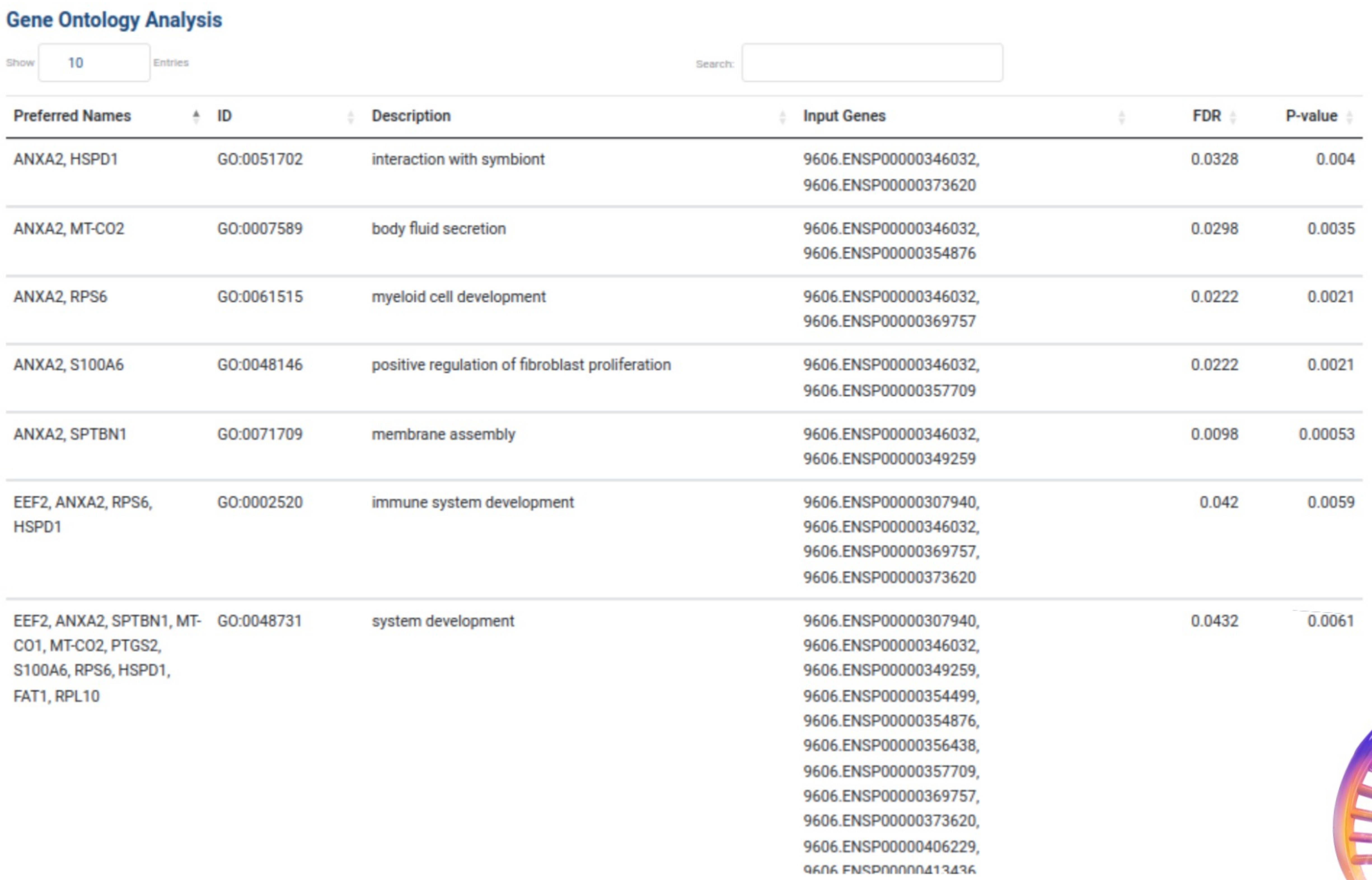
https://www.ncbi.nlm.nih.gov/sra/SRX276898



Heatmap for differentially expressed genes

Bulk RNASeq





Gene Ontology Analysis

Show 10 Entries

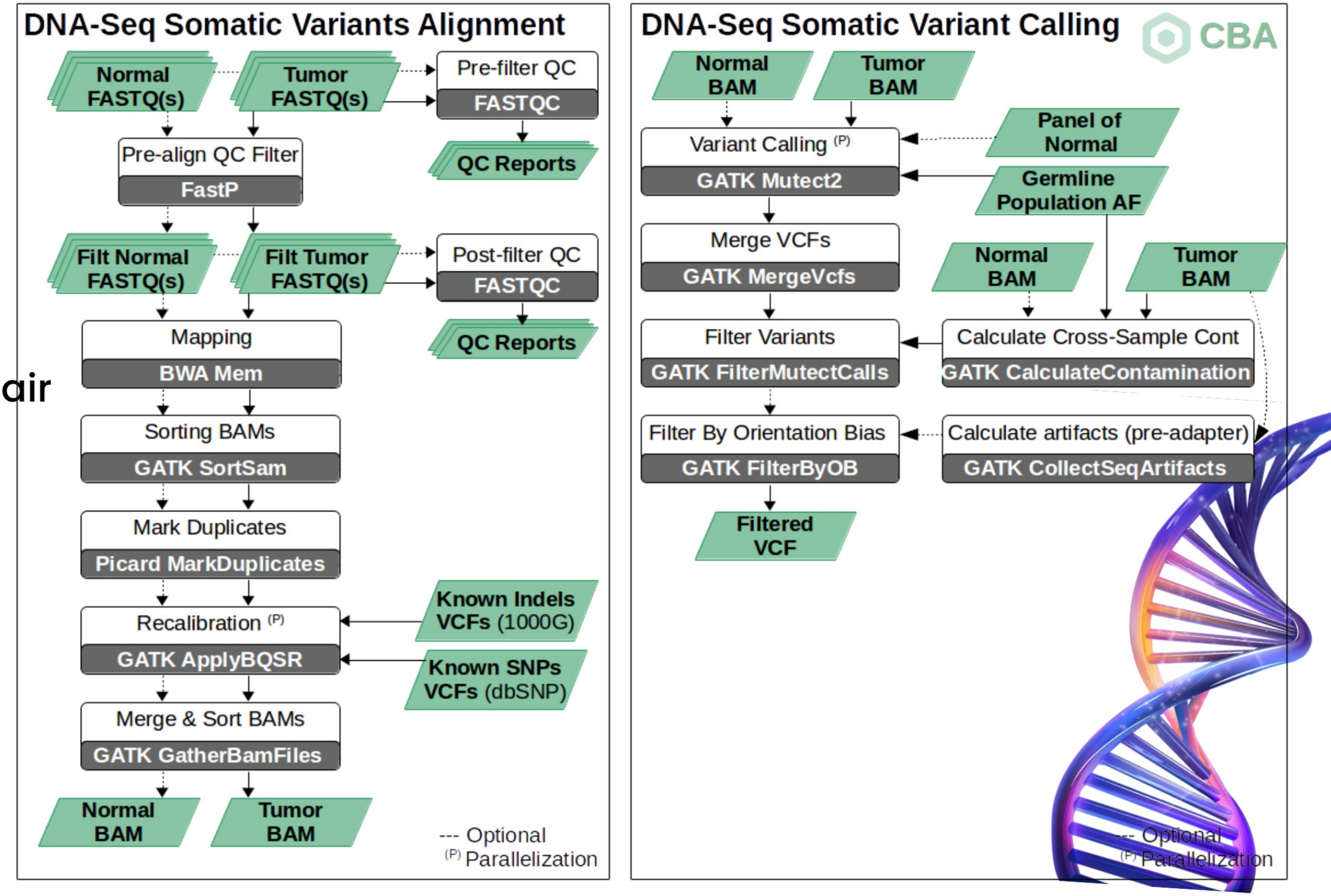
Search:

Preferred Names	ID	Description	Input Genes	FDR	P-value
ANXA2, HSPD1	GO:0051702	interaction with symbiont	9606.ENSP00000346032, 9606.ENSP00000373620	0.0328	0.004
ANXA2, MT-CO2	GO:0007589	body fluid secretion	9606.ENSP00000346032, 9606.ENSP00000354876	0.0298	0.0035
ANXA2, RPS6	GO:0061515	myeloid cell development	9606.ENSP00000346032, 9606.ENSP00000369757	0.0222	0.0021
ANXA2, S100A6	GO:0048146	positive regulation of fibroblast proliferation	9606.ENSP00000346032, 9606.ENSP00000357709	0.0222	0.0021
ANXA2, SPTBN1	GO:0071709	membrane assembly	9606.ENSP00000346032, 9606.ENSP00000349259	0.0098	0.00053
EEF2, ANXA2, RPS6, HSPD1	GO:0002520	immune system development	9606.ENSP00000307940, 9606.ENSP00000346032, 9606.ENSP00000369757, 9606.ENSP00000373620	0.042	0.0059
EEF2, ANXA2, SPTBN1, MT-CO1, MT-CO2, PTGS2, S100A6, RPS6, HSPD1, FAT1, RPL10	GO:0048731	system development	9606.ENSP00000307940, 9606.ENSP00000346032, 9606.ENSP00000349259, 9606.ENSP00000354499, 9606.ENSP00000354876, 9606.ENSP00000356438, 9606.ENSP00000357709, 9606.ENSP00000369757, 9606.ENSP00000373620, 9606.ENSP00000406229, 9606.ENSP00000413436	0.0432	0.0061

Gene Enrichment via Pathway and gene ontology

Variant Analysis

- GATK Best Practice workflow is followed for germline and somatic variant analysis
- Haplotype caller for germline variants
- Mutect2 for somatic variants (tumor normal pair or tumor only mode)
- Annotations using funcotator by GATK
- Filtered VCF files in tabular format for easier interpretation & Visualization on any given platform



DNA-Seq Somatic Variants Alignment

Normal FASTQ(s)
Tumor FASTQ(s)
Pre-filter QC
FASTQC
QC Reports
Pre-align QC Filter
FastP
Filt Normal FASTQ(s)
Filt Tumor FASTQ(s)
Post-filter QC
FASTQC
QC Reports
Mapping
BWA Mem
Sorting BAMs
GATK SortSam
Mark Duplicates
Picard MarkDuplicates
Recalibration (P)
GATK ApplyBQSR
Known Indels VCFs (1000G)
Known SNPs VCFs (dbSNP)
Merge & Sort BAMs
GATK GatherBamFiles
Normal BAM
Tumor BAM

--- Optional
(P) Parallelization

DNA-Seq Somatic Variant Calling

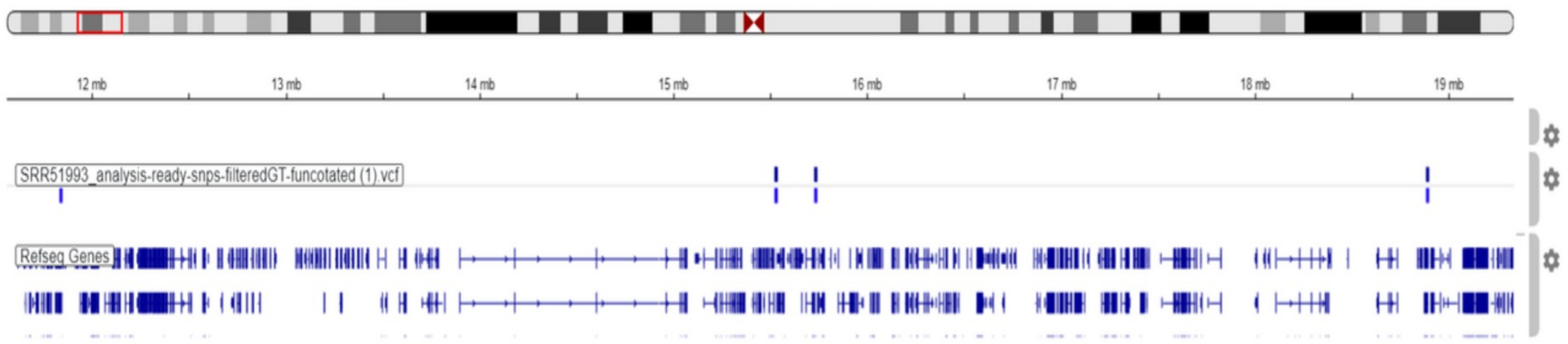
CBA

Normal BAM
Tumor BAM
Variant Calling (P)
GATK Mutect2
Panel of Normal
Germline Population AF
Merge VCFs
GATK MergeVcfs
Normal BAM
Tumor BAM
Filter Variants
GATK FilterMutectCalls
Calculate Cross-Sample Cont
GATK CalculateContamination
Filter By Orientation Bias
GATK FilterByOB
Calculate artifacts (pre-adapter)
GATK CollectSeqArtifacts
Filtered VCF

--- Optional
(P) Parallelization

Variant Analysis

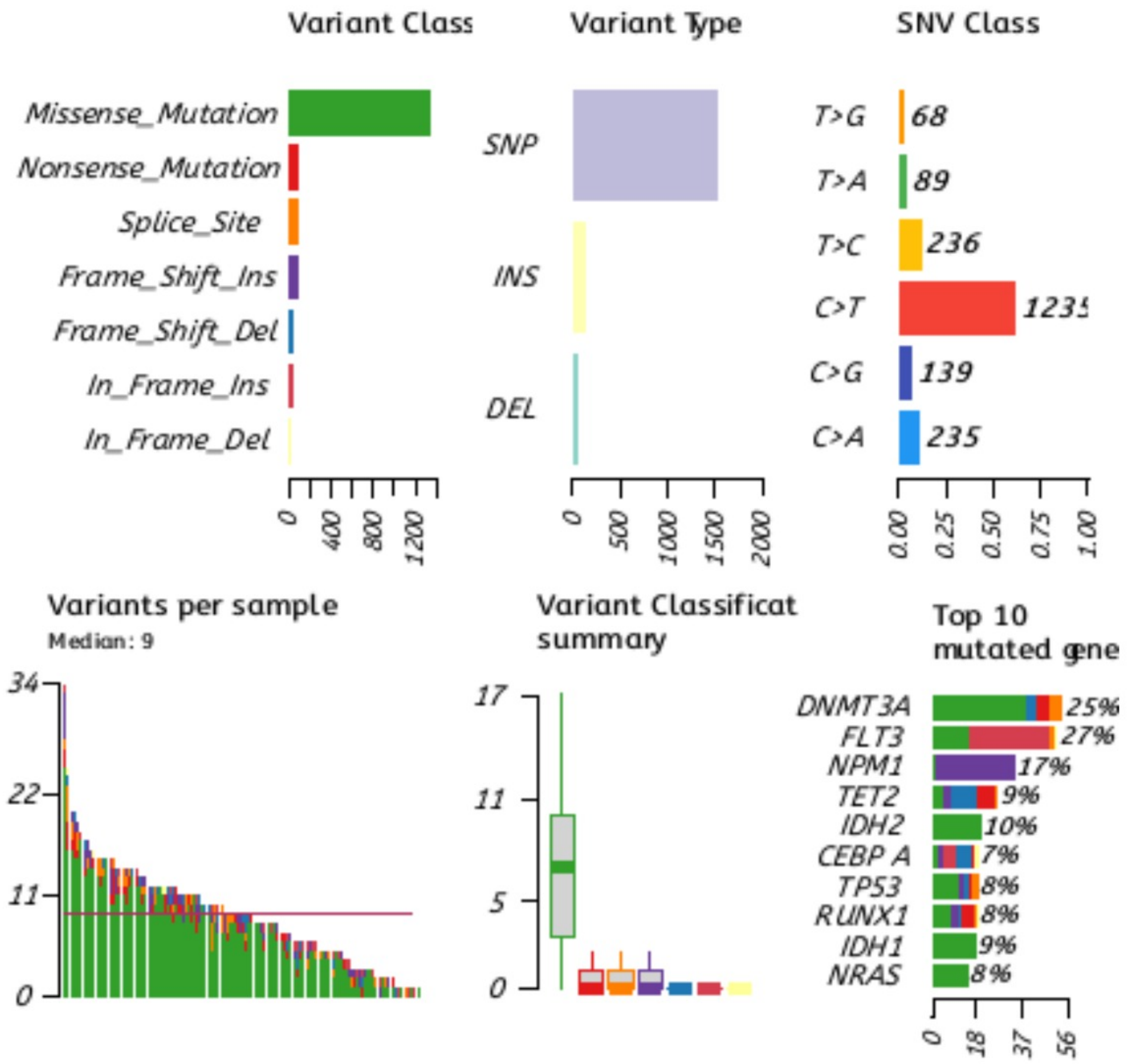
- VCF/bam files along with the index file used for visualization
- Maftools used for further visualization of mutated genes and classification of variants



Chr chr1
Pos 15,525,025
Loc 15,525,025-15,525,025
ID .
Ref A
Alt C
Qual .
Filter haplotype

Cravat A->C

AS_FilterStatus SITE
AS_SB_TABLE 0,0|0,0
DP 1
ECNT 2
FUNCOTATION [CASP9|hg38|chr1|15525025|15525025|FIVE_PRIME_FLANK||SNP|A|A|C|g.chr1:15525025A>C|ENST00003338...



Metagenomic Analysis

- Explore the diversity, structure and function of microbial populations in various ecosystems like soil, human gut, etc.
- Applications:
  - Environmental monitoring
  - Human microbiome studies
  - Biotechnology (biofuel production)

Types of Analysis:

- Shotgun Metagenomics: DNASeq, using all genetic material.
- Amplicon Sequencing: Use specific marker gene like 16S rRNA to study diversity.

Metagenomic Analysis

Workflow for Illumina Reads

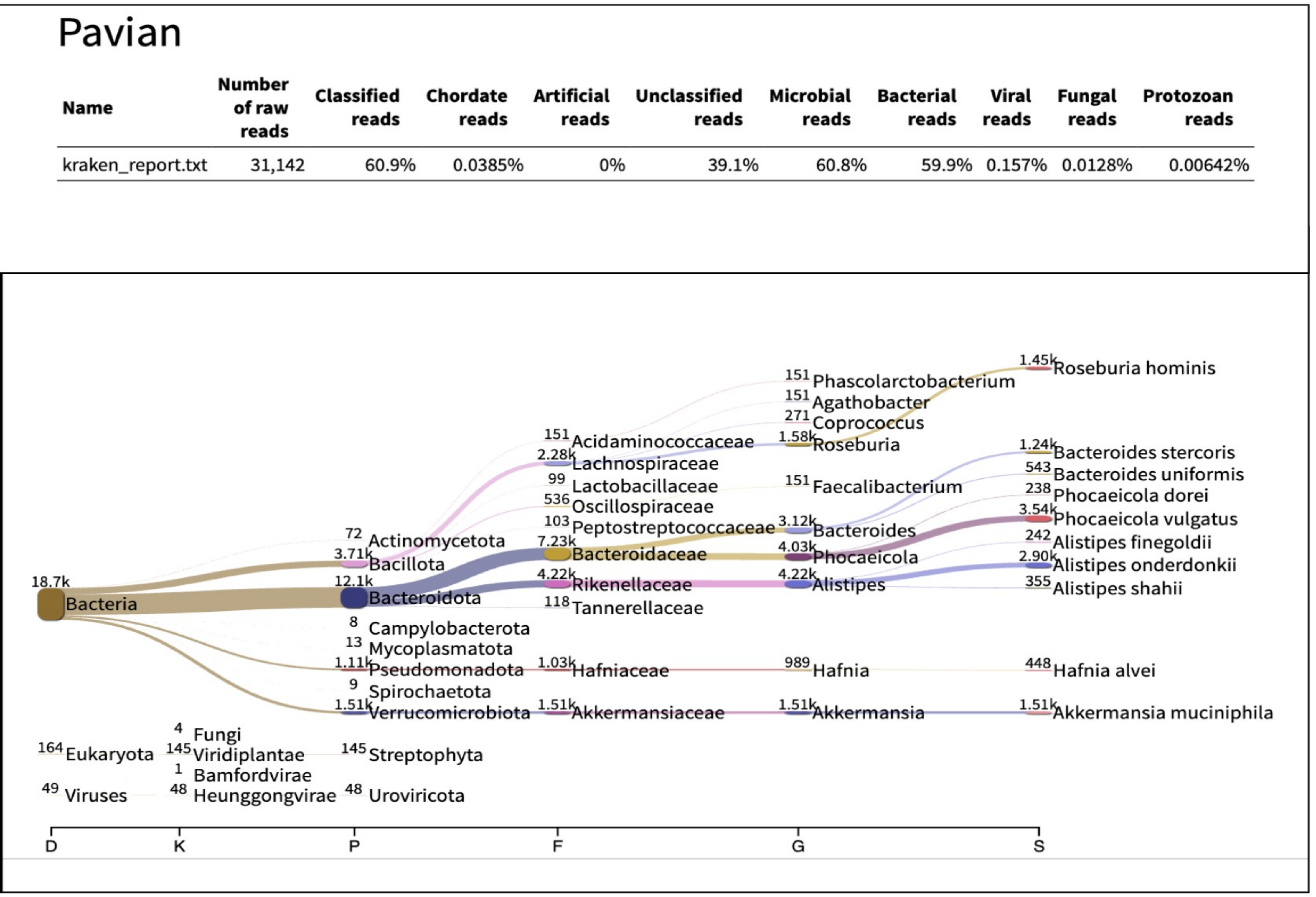
QC → Trimming → Host Removal → Taxonomic Classification → Assembly → Binning → Functional Annotation

Workflow for ONT Reads

QC → Trimming → Post Trimming QC → Extract Assigned Reads → Identification of AMR gene

Metagenomic Analysis

Results



Pavian

Name	Number of raw reads	Classified reads	Chordate reads	Artificial reads	Unclassified reads	Microbial reads	Bacterial reads	Viral reads	Fungal reads	Protozoan reads
kraken_report.txt	31,142	60.9%	0.0385%	0%	39.1%	60.8%	59.9%	0.157%	0.0128%	0.00642%

Upcoming

Single Cell Sequencing Analysis

- RNASeq: Tumor Heterogeneity
- ATACSeq: Transposase-accessible chromatin sequencing
- MethylSeq sequencing: DNA Methylation (Epigenetic analysis)

Immune Repertoire Analysis

- Preprocessing: AIRRSeq-tsv
- Postprocessing:
  - Repertoire overlap between samples
  - CDRH3 length distribution
  - Diversity Analysis
  - Clonal Diversity
  - Identify top clonotypes

Future Work & Vision

- Integration with machine learning for better predictions and classifications.
- Use of multi-omics for comprehensive analysis

Get In Touch

Email Adress

contact@genomebeans.com

Website

www.genomebeans.com

Whatever your goal, we can help you achieve it.