



RESEARCH ARTICLE

Identification of Colorectal Adenocarcinoma Key Genes to Be Validated as Potential Prognosis Biomarkers

Elham Omer Mahgoub^{1*}, Pooja Prakash Mankar², Amit kumar³, Syed Asif Husain Naqvi⁴ and Bharti Mittal⁵

¹Department of Science and Math, Community College of Qatar, Doha, Qatar

²Shri Shivaji College of Arts, Commerce and Science, Akola, India

³ICMR, New Delhi, India

⁴BBR PTY LTD, 205, Lilyfield Road, Lilyfield, Sydney, NSW- 2040, Australia

⁵D502- Amity Akruthi, Ananth Nagar, Phase2, E-city-2, Huskur Gate, Karnataka, India

*Corresponding author: Elham Omer Mahgoub, Department of Science and Math, Community College of Qatar, Qatar University, Doha, Qatar



Abstract

Background: Colorectal Adenocarcinoma (COAD) continues to be one of the leading causes of death worldwide. The patient's chance of survival increases with the early prognosis of a malignant tumor.

Methods: Advanced bioinformatics methods were used to gain a thorough understanding of the genetic landscape of colorectal cancer. The transcriptome RNA-seq raw data were obtained from PRJEB24758 in the European Nucleotide Archive (ENA) database. Then, the Sequence Read Archive (SRA) runs a selection to download the sequences database. The transcriptomic RNA released data is analyzed by the bioinformatics tool, which also uses online analytic tools to help visualize the results and identify key genes that may be employed as prognosis biomarkers in the future. The annotation pathways have been determined using David's annotation tools, and cluster analysis in Gsea and the c-bioportal database also showed the significance of these pathways. Two hundred fifty miRNA overlapped with the highest two upregulated and down regulated genes that were subjected to screening. The Venn diagram determined the common genes that the immunogenic genes set and cell type gene signature.

Results: 1,274 genes with substantial differential expression in colorectal cancer were found using stringent approaches such as HISAT2 alignment and DeSeq2 analysis. This study identified 913 upregulated genes of colorectal adenocarcinoma (COAD). The upregulated genes-expressed profile of COAD was studied. The upregulated genes are controlled by 20 pathways expressed in colorectal adenocarcinoma. David's annotation tool was used to prepare diagrams for the KEGG analysis, enriched genes, and many more diagrams. The resulting miRNA overlapped genes interacted significantly with TF to produce key genes. From a different perspective, fifty-seven upregulated common genes were determined using the Venn diagram, and higher mutated genes were selected. Investigating these key genes for targeted therapy in colorectal cancer therapy is crucial, as the study emphasizes.

Conclusion: Thus, the development of novel therapeutic approaches and the identification of key genes of the changed expression of genes implicated in COAD drug resistance are crucial goals for the ongoing advancement of COAD therapy.

Keywords

Colorectal adenocarcinoma, RNA sequencing, Differential expression genes (DEGs)

Abbreviations

COAD: Colorectal Adenocarcinoma; RNA-seq: RNA sequencing; PRJEB24758: Project identifier for transcriptome data in the European Nucleotide Archive; ENA: European Nucleotide Archive; SRA: Sequence Read Archive; miRNA: MicroRNA; TF: Transcription Factor; DEGs: Differentially Expressed Genes; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: Gene Set Enrichment Analysis; DAVID: Database for Annotation, Visualization, and Integrated Discovery; FF: Fresh Frozen (tissue); FFPE: Formalin-Fixed, Paraffin-Embedded (tissue); NGS: Next-Generation Sequencing; FTP: File Transfer Protocol; GTF: Gene Transfer Format; GRCh38: Genome Reference Consortium Human Build 38; FastQC: Fast Quality Control (tool for assessing raw sequencing data); MultiQC: Multi Quality Control (aggregates multiple FastQC reports); HISAT2: Hierarchical Indexing for Spliced Alignment of Transcripts (alignment tool); KEGG: Kyoto Encyclopedia of Genes and Genomes; HPA: Human Protein Atlas; GOTERM-cc: Gene Ontology Term - Cellular Component; STRING: Search Tool for the Retrieval of Interacting Genes/Proteins; GSEA: Gene Set Enrichment Analysis; Limma: Linear Models for Microarray Data (R package for differential gene expression analysis); Cytoscape: Software platform for visualizing molecular interaction networks; miRNA: MicroRNA; R: Programming language used for statistical computing and bioinformatics; cBioPortal: Cancer BioPortal (tool for exploring multidimensional cancer genomics datasets); G-Profiler: Gene functional profiling tool; MA Plot: Mean-Average Plot; Entrez Gene ID: Unique identifier for a gene in the Entrez Gene database; Ensembl ID: Unique identifier for a gene in the Ensembl database; GO: Gene Ontology; FDR: False Discovery Rate; BP: Biological Process (a category within Gene Ontology); CC: Cellular Component (a category within Gene Ontology); MF: Molecular Function (a category within Gene Ontology); HPA: Human Protein Atlas; CORUM: Comprehensive Resource of Mammalian Protein Complexes; RPL12: Ribosomal Protein L12; RPL3: Ribosomal Protein L3; TIMP2: Tissue Inhibitor of Metalloproteinase 2; MYLK: Myosin Light Chain Kinase; SYNPO2: Synaptopodin 2; PCSK5: Proprotein Convertase Subtilisin/Kexin Type 5; JCHAIN: Joining Chain of Multimeric IgA and IgM; ABCA8: ATP Binding Cassette Subfamily A Member 8; SLC26A3: Solute Carrier Family 26 Member 3

Introduction

The second most significant cause of cancer-related deaths worldwide. The third most prevalent cancer globally. Obesity, inactivity, and smoking habits, in addition to an older population, raise the risk of Colorectal Adenocarcinoma (COAD). Colorectal Adenocarcinoma continues to be a problem despite improvements in surgical methods, chemotherapy, and radiotherapy. Endoscopic procedures, such as endoscopic submucosal dissection and endoscopic mucosal excision, have been used to treat early-stage COAD. Adenocarcinomas are the most common subtype of Colorectal Adenocarcinoma (COAD) and result from the malignant transformation of glandular epithelial cells in the colon or rectum. These tumors usually grow slowly, starting as benign adenomatous polyps and moving through a sequence of well-researched genetic and epigenetic changes called the adenoma-carcinoma sequence. However, due to the prevalence of lymph node metastases, lymph node dissection is a crucial component of surgical therapy for cancer in its advanced stages [1]. Once colon cancer develops local or distant metastases, the prognosis is dismal. Researchers or physicians may need to find novel markers for predicting the prognosis and therapy of colon cancer to improve the prognosis of people with the disease.

Colorectal Adenocarcinoma (COAD) patients are most likely to develop new targeted genes to overcome the cancer cell drug resistance to chemotherapy either naturally or as a result of receiving T-DM1 treatment. This study suggests that variations in the expression levels of specific genes may impact the sensitivity or resistance to chemotherapy [2-5]. Biomarkers are essential for the early diagnosis of drug resistance to chemotherapy. Although biomarkers are utilized at different phases of the disease to aid in its management, each has its limits [6,7]. Identification of essential genes that could be used as biomarkers for diagnosis of colorectal cancer was a

persuasive motive to analyze Differential expression genes [8-10].

The clinical use of biomarkers in COAD is crucial for prognostic stratification, surveillance, and treatment selection, in addition to being necessary for the early diagnosis of the disease. Therefore, there is an urgent need for new, affordable, non-invasive, easily quantifiable, and accurate screening methods for colorectal cancer (COAD). This study uses RNA-Seq data of colorectal cancer (COAD) to analyze differential gene expression and predict possible biomarker for future diagnosis and gene therapy.

Methods

Data Collection

The dataset used in this work is freely available to the public via the European Nucleotide Archive (ENA) under project ID PRJNA649071 (<https://www.ebi.ac.uk/ena/browser/home>). The transcriptome RNA-seq raw data were obtained from PRJEB24758 in the European Nucleotide Archive (ENA) database. Then, the Sequence Read Archive (SRA) runs a selection to download the sequences database. Samples relevant to Colorectal Adenocarcinoma (COAD) are gathered into this dataset, which includes fresh frozen tissues (FF), formalin-fixed, paraffin-embedded (FFPE) tissues, and several cancer cell lines.

The NGS datasets were discovered by using the ENA's search feature associated with colorectal adenocarcinoma. Species (such as *Homo sapiens*), experimental circumstances, library type (RNA-Seq), and sequencing platform (Illumina) filters were used to limit the results. The Run Accession Numbers for certain transcriptome RNA-seq raw data runs that matched colorectal cancer samples were noted. Europa Galaxy was used to identify the differentially expressed genes (DEGs) between colorectal adenocarcinoma tissues and

normal tissues. The inclusion of matched normal and tumor samples from COAD patients provides a valuable resource for comparative analysis. FTP (File Transfer Protocol) was used for effective data retrieval. The raw transcriptome RNA-seq raw data (single-end) files (often in FASTQ format) containing the raw sequencing reads were obtained using FTP links. After the download was finished, the raw data files were arranged on the local storage system. An organized directory including pertinent metadata, sample details, and the name of the dataset was created.

Further, the data were filtered according to normal and tumor samples in an Excel worksheet. The data was also filtered to select single-end data, as it contains a larger data set than paired-end data.

Quality control and processing

Tools to process the data in Galaxy, The FTP links were uploaded to the Galaxy database (<https://usegalaxy.org/>) along with the reference genome (Homo sapiens) from the Ensembl. The sample set for the investigation included 186 runs of tumor samples and 164 runs of normal samples, offering a balanced representation for reliable comparison analyses. This important collection paves the way for future discoveries on the molecular causes of Colorectal Adenocarcinoma (COAD) and enables a thorough investigation of the transcriptional changes linked to the illness.

The FastQC tool is used to calculate the quality control of NGS raw data. The quality of sequencing data was checked by clicking on FastQC output datasets under the various quality matrices, including pre-base sequence quality, sequence length distribution, GC content, etc. Using MultiQC, the FastQC report was aggregated into one report. The adapter sequences were trimmed using the TrimGalore tool. The ftp link of reference genome https://ftp.ensembl.org/pub/release110/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.toplevel.fa.gz (Homo sapiens) and GTF file <https://ftp.ensembl.org/pub/release-110/gtf/homosapiens/> and Homo_sapiens.GRCh38.110.gtf.gz was copied and uploaded in the galaxy for alignment. The alignment sequence of the data compared with the reference genome process was done using the HISAT2 tool. The Feature Counts tool is used to count how many reads are aligned to each gene (count matrix). The differentially expressed genes from RNA-seq data were identified using the DESeq2 tool, the RNA sequencing summarized in flowchart shown in Figure 1.

Statistical analysis

NA values were eliminated by applying the filter to P-adj values in the Excel sheet. The filter was applied to p-adj as ≤ 0.05 to derive only significant genes. Further, the data were sorted based on fold change values to identify upregulated genes (threshold is ≥ 2) and to identify downregulated genes (threshold is ≤ -2).

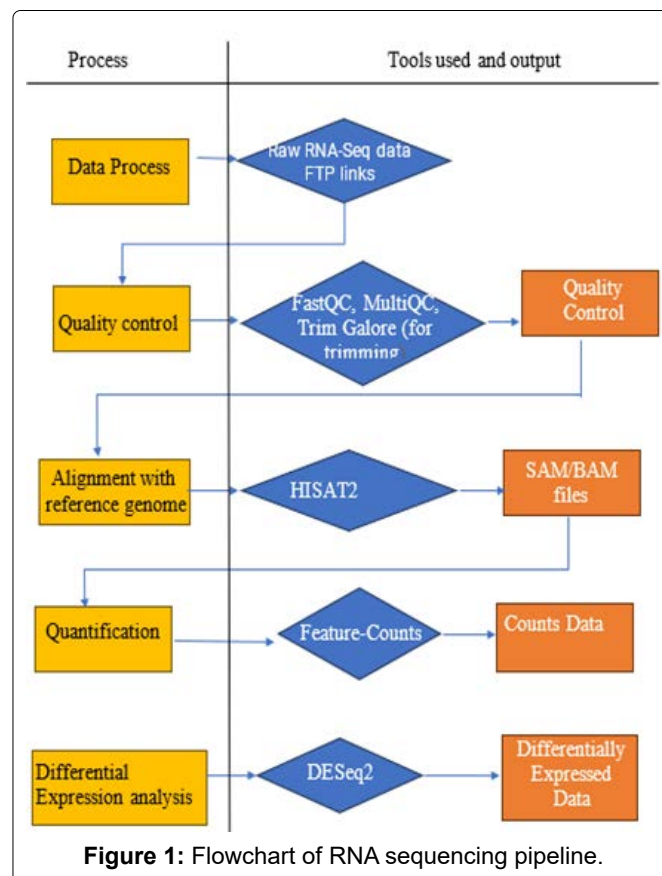


Figure 1: Flowchart of RNA sequencing pipeline.

Differential gene expression analysis

Next, the Database for Annotation, Visualization, and Integrated Discovery (DAVID) <https://david.ncifcrf.gov/>, ver. [11]. It was used to identify potential uses for differentially expressed genes (DEGs). The DEGs were analyzed by choosing official gene symbols, annotation tables, and Entrez IDs. Genes were functionally enriched using g-Profiler <https://biit.cs.ut.ee/gprofiler/gost>, which conducts statistical enrichment analysis and offers interpretative insights into gene lists provided by users [11,12]. In the David annotation tools, the Kyoto Encyclopedia of Genes and Genomes (KEGG), HPA tissues, Upregulated tissues, GOTERM-cc-direct cellular component, and gene enrichment analysis columns were prepared using the lists of the gene's functional annotation tools. Two columns were copied into a new Excel worksheet, sorted from larger to smaller, and the chart was chosen after the columns of each functional annotation. Their related frequency column was selected in the Excel worksheet and inserted using pivotable in Excel.

Functional gene enrichment

The STRING database found the most crucial modules in the protein-protein interaction [13] network. The study's analysis of the molecular causes of colorectal cancer revealed the gene different cluster group and activities in the gene set related to colorectal adenocarcinoma. Furthermore, using a string database, co-expression genes are identified as additional information on the gene's connection and regulation

Gene set enrichment analysis (GSEA)

The study used the Gene Set Enrichment Analysis (GSEA) database <https://www.gsea-msigdb.org/gsea/index.jsp> to measure the overlap of gene functions. The Ensembl ID was submitted to the molecular signature database, and the oncogenic dataset was selected to visualize the transcription factor for each enriched gene in the large-scale design.

Venn diagram analysis

Venn diagram is used to overlap the immunogenic genes set versus cell type gene signature data that is retrieved from the Gsea database. The Venn diagram method draw venn diagrams result urgent. Be the common gene expression signatures of the immunogenic genes set versus cell type gene signature are considered the suitable group of genes to be used in further experiments.

Mapping of protein-protein network

The biomart mapper script in R uses the Limma package to map 913 upregulated and 906 downregulated genes of COAD and choose the functional protein coding of the RNA sequence data. Even though the original purpose of the R program Limma was to analyze differential expression (DE) in RNA sequencing data. Cytoscape 3.7.1 was used to map the protein-coding gene data that was produced. The obtained data was then used to forecast the heat map of the top 10 downregulated genes and the highest upregulated genes, demonstrating a considerable up-regulation of COAD relative to normal tissue.

miRNA prediction using mirwalk database

The two highest upregulated genes and two highest downregulated genes were submitted to the mirwalk database <http://mirwalk.umm.uni-heidelberg.de/> to identify the miRNA that interacts with these genes. The results have been saved in one Excel worksheet. Next, the same list of genes was submitted to the chea3 database <https://maayanlab.cloud/chea3/> to identify the overlapped transcription factors (TF), and that was added to the second worksheet. Finally, the resulting TF list was submitted to the mirwalk database, and the overlapped miRNA was documented. The overlapped 250 miRNA in the two listed was prepared using Excel formulas. Cytoscape software was used to merge the three networks of TF-Genes, Genes-TF, and finally, miRNA-TF. One gene was connected to one transcription factor, and both overlapped with the same miRNA.

C.bioportal database

The structure of the gene variants and the types of mutations in the enriched genes were examined using the c.bioportal database <https://www.cbioportal.org/>. The official_gene_Id set genes are submitted for colorectal adenocarcinoma based on gene selection, and each gene mutation is thereafter examined individually.

G-profiler database

In order to determine the functional gene pathways, common group, and enriched genes, the Ensembl_Id query was submitted to g-profiler <https://biit.cs.ut.ee/gprofiler/gost>. Besides general annotations, interactions, literature, pathway associations, disease associations, and tissue expression were explored.

Results

Differential expression genes analysis

Differential expression genes analysis resulted from RNA sequencing. As shown in Figure 2, the Genes were processed using David's annotation analysis site. The MA plot shows differentially expressed genes, with blue indicating a departure from the baseline, suggesting substantial transcriptional activity. The distribution of points indicates a complex alteration in gene regulatory networks in colorectal cancer. The dispersion plot shows decreased variability in gene expression as the mean expression level increases, indicating the technical robustness of RNA-Seq data. The volcano plot shows genes exceeding established thresholds for p-value and log2 fold change, underscoring their potential as key players in colorectal cancer pathology. These genes represent promising targets for further validation as biomarkers or therapeutic interventions.

Transcription factors

Alterations in transcription factors' (TFs') expression and activity affect other genes' levels. TFs regulate essential physiological processes and control the expression of multiple genes at once. Transcription factors are heat mapped using the Gsea database, as shown in Figure 3. As in Figure 3, the TF checkpoint database provided the reported gene functions, and Ensembl genes were utilized to convert the reported Entrez gene IDs to corresponding Ensembl IDs. Transcription factor (TF) enrichment analysis was performed on the differentially expressed genes using the Gsea database web-server application. In short, humanized gene symbols with human differentially expressed genes [14] (Figure 4).

Gene enrichment analysis

One technique for examining genomic data, especially extensive transcriptome data, is Gene Ontology (GO) analysis. As explained in Figure 5 using The Database for Annotation, Visualization, and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>, ver. 6.8). Ten pathways of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were visualized. Gene Ontology (GO) enrichment analysis to ascertain the biological function of cancer-related protein pathways [15]. For additional visualization, the enriched GO keywords and pathways with a False Discovery Rate (FDR) of less than 0.01 were chosen. Plotting of KEGG pathways, bubble graphs, and the top 20 relevant

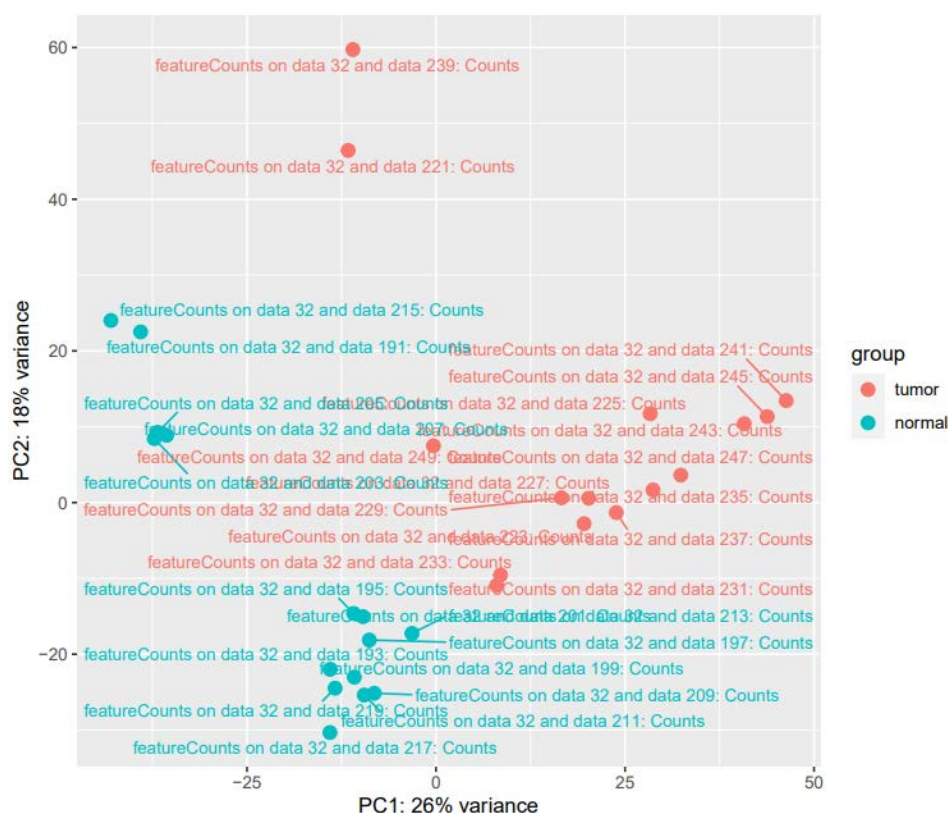


Figure 2A: PC1 the principal component 1 in the x-axes and PC2 the principal component 2 in the y-axes to characterize the variants in the genes.

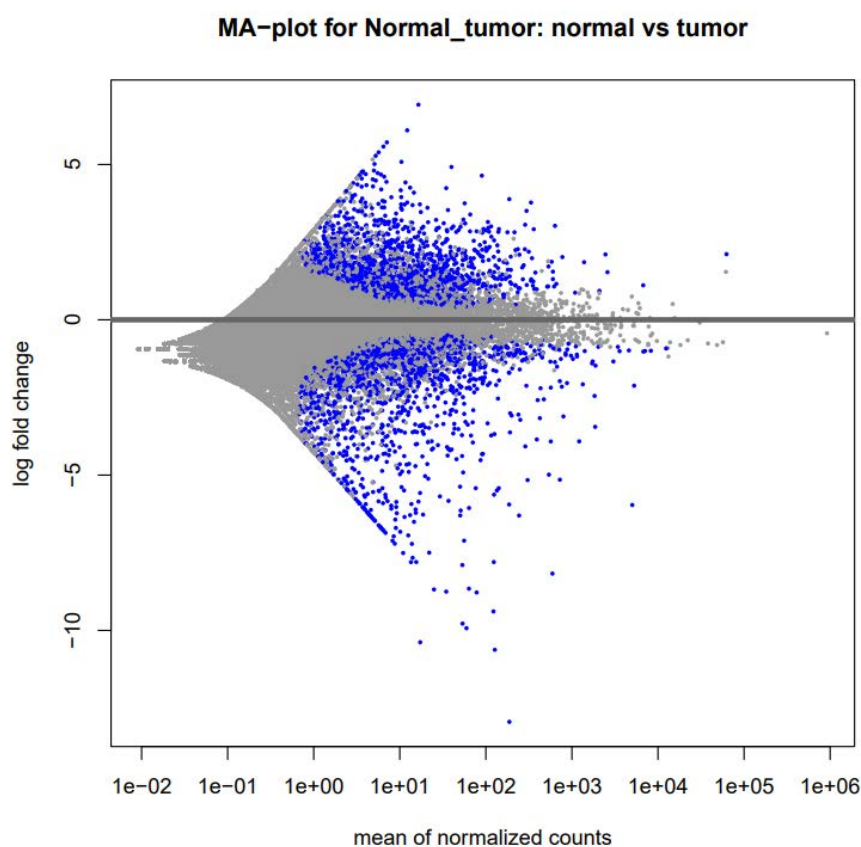


Figure 2B: The volcano plot shows the distribution of the genes depending on their expression of the genes. The upregulated genes are shown in blue while the down-regulated genes are shown in grey.

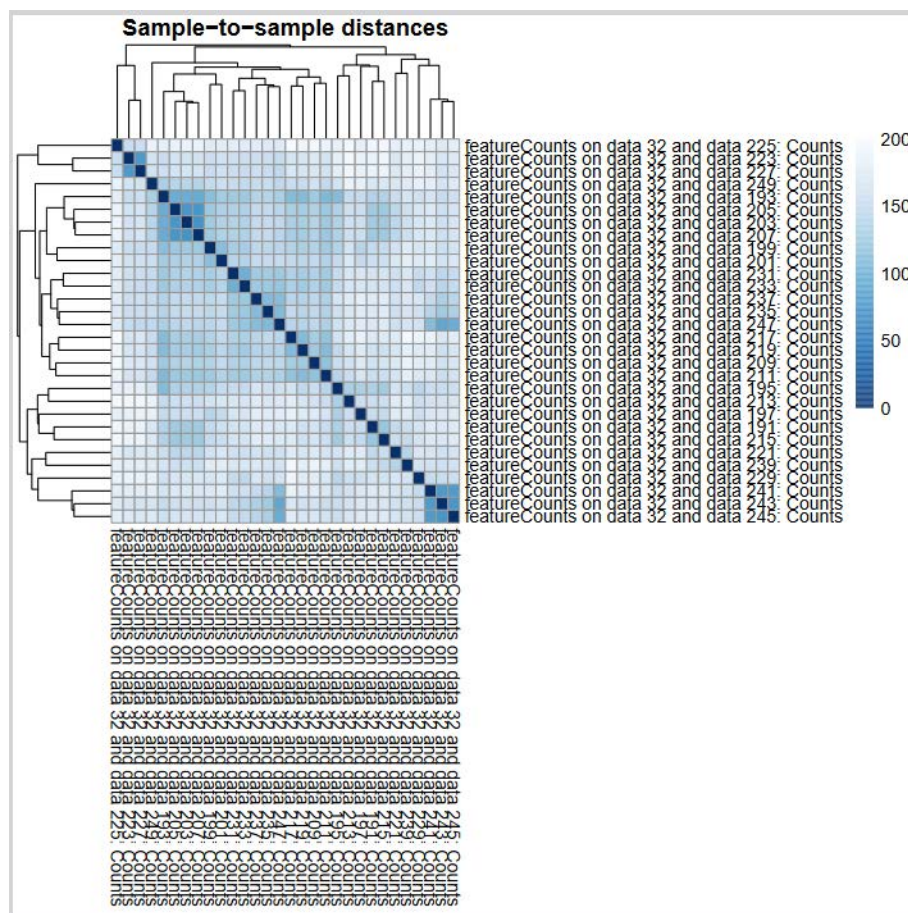


Figure 2C: Box-plot of feature genes counts as normal genes and tumor genes.

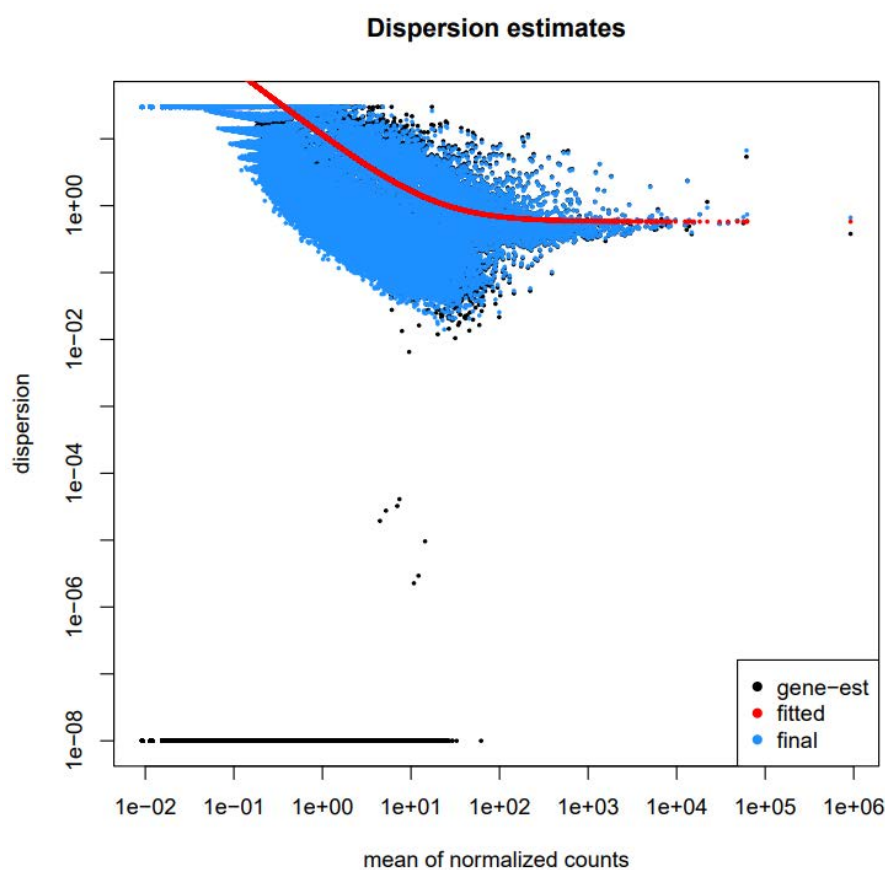
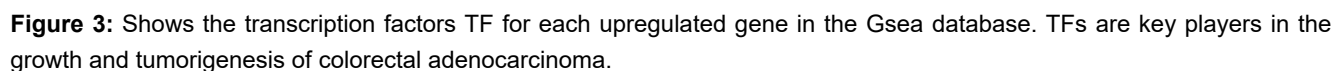


Figure 2D: The normalized count represents the upregulated genes and downregulated genes.



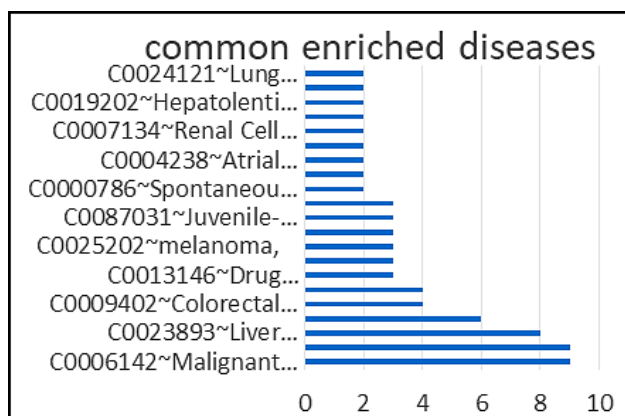


Figure 5A: Shows Functional enrichments performed; common enriched diseases are malignant neoplasm's of the breast liver Fibrosis and colorectal cancer.

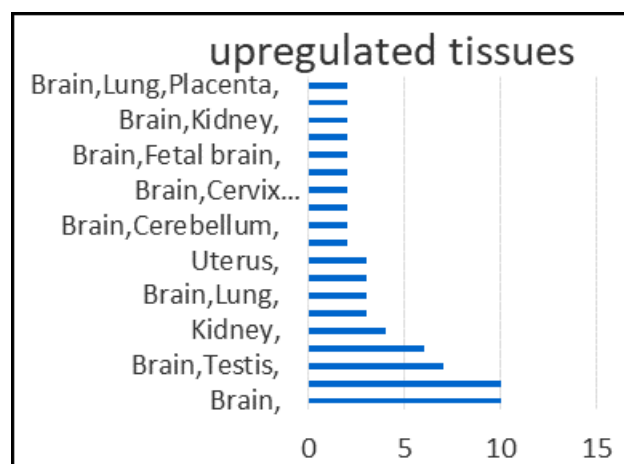


Figure 5D: Upregulated tissues: The primary tissues that express the upregulated genes are the Brain Testis Kidney and lung. They are involved in cytoplasmic translation and translation.

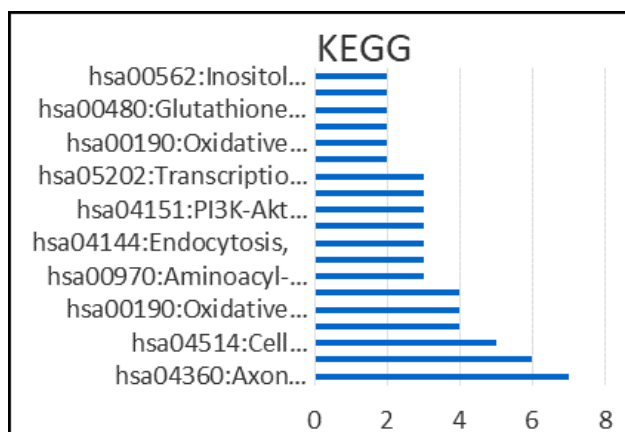


Figure 5B: KEGG: the upregulated genes expressed in colorectal adenocarcinoma depend on these pathways.

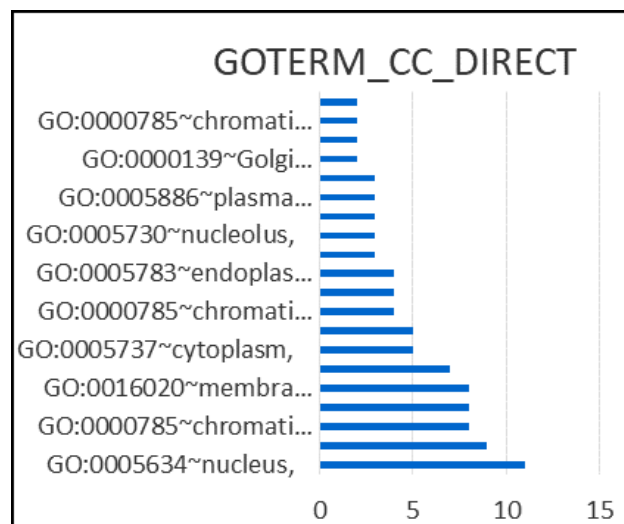


Figure 5E: GOTERM-cc- Direct cellular component relation extraction between gene and other entities.

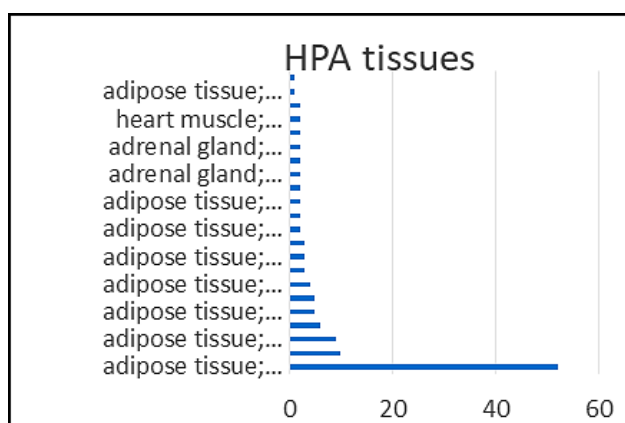


Figure 5C: HPA tissues: the gene expression enriched in adipose tissues adipocytes and adrenal glands.

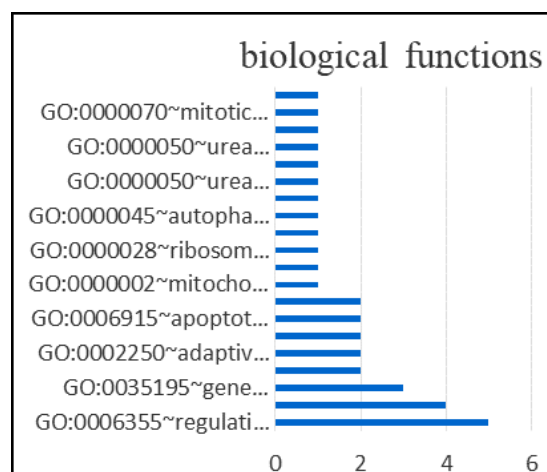


Figure 5F: Gene enrichment analysis the primary biological functional enriched genes the main gain involved in the regulation of the transcriptome genes and genes silencing by miRNA.

GO terms (BP, CC, and MF) was done using available methods in Microsoft Excel.

Cytoscape, enrichment analysis

As shown in [Figure 6A](#) Cytoscape software, the network illustrates how the DEGs found from COAD interact with one another on a protein-protein level using Cytoscape. As shown in [Figure 6A](#) and [Figure 6B](#), the network visualizes functional groups based on the most significant (leading) phrase for each group, offering a clear picture of how upregulating and downregulating relate to one another. Functional groups can be created utilizing the GO relationships as an alternative to the kappa score grouping method. Meanwhile, the protein-protein network of the group 913 upregulated genes of colorectal adenocarcinoma generated from the NDeX database shows a very strong network ([Figure 6C](#)).

The heatmap of top ten genes

As shown in [Figure 6D](#), ten higher upregulated genes and ten higher downregulated genes are in a heatmap predicted by the Limma package in R. The green color represents the normal sample, and the pink color represents the tumor samples.

Correlation networks

In [Figure 7A](#), a Venn diagram was used to identify 57 common genes, and the Pearson correlation coefficient analysis was performed on these genes. [Figure 7A](#) shows a Venn diagram comparing immunogenic genes with cell type gene signatures, revealing 57 common genes between the two; [Figure 7B](#) String protein-protein interaction classified the upregulated genes into five clusters; [Figure 7C](#) depicts the regulatory network of miRNA-Gene-TF in COAD, highlighting 250 differentially expressed mRNAs. The green circles represent the differentially expressed target mRNAs; [Figure 7D](#) shows that the key gene TIMP2 has significant interactions with TF and miRNA.

G-profiler and gene ontology analysis

The g-profiler database recognized the functional enrichment of GO: MF, GO: BP, GO: CC, KEGG, and TF for each common group, as shown in [Table 1](#) and [Figure 8](#). Gene fusion events can produce new proteins with modified functionalities or regulatory characteristics, which may have an impact on the development of diseases like colorectal cancer. With a fusion score of 0.003, the string network displays the fusion of the RPL12 and RPL3 genes. Study's analysis of the molecular regulations of colorectal cancer with g: Profiler revealed overrepresented biological processes and activities in the gene set related to colorectal cancer. The results underscore the significance of comprehending the basic foundations of the disease by offering insights into the pathophysiological mechanisms involved. The g-profiler database arranges the enriched gene in 20 pathways beside the KEGG, WP, TF, MIRNA, HPA,

and CORUM pathways had been reported in [Figure 8A](#). In [Figure 8B](#), the enrichment analysis explained a significant overrepresentation of gene ontological terms associated with protein binding (GO: 0005515), structural molecule activity (GO: 0005198), and protein-containing complex binding (GO: 0044877). Moreover, there was a noticeable enrichment in "cytoplasmic translation" (GO: 0002181) and "cell adhesion" (GO: 0007155), which suggests modifications to the cellular structure and translation machinery.

The mutants' genes

As mentioned in [Figure 9](#), the highest mutated genes are MYLK, SYNPO2, PCSK5, JCHAIN, ABCA8, and SLC26A3, which were selected from the 57 common genes, identified using the Venn diagram tool.

Discussion

Identifying patients with colon adenocarcinoma (COAD) who may benefit from immunotherapy and whose tumor microenvironment (TME) was required for reprogramming to advantageous immune-mediated responses requires precise immune molecular characterization due to cancer cell mutations. On the other hand, little has been discovered regarding COAD's immunological characteristics. Despite the current gaps in our understanding of COAD's Immunological characteristics, there is hope. Our research is paving the way for the development of accurate and reliable biomarkers for COAD diagnosis and prognosis. By observing co-expression correlations, regulatory relations, and physical binding interactions, we are rehabilitating the colorectal adenocarcinoma network and opening new doors for patient care.

This study identified 913 upregulated genes of colorectal adenocarcinoma (COAD). The gene-expressed profile of COAD was studied, and many findings were noted. The upregulated genes are controlled by 20 pathways expressed in colorectal adenocarcinoma, as shown in [Figure 2](#), analyzed by cluster profiler. RNA-sequence expression of COAD was characterized using David annotation analysis, and the resulting genes were compared with each published COAD gene database. Besides, the known disease-associated gene sets rehabilitate the colon adenocarcinoma network observed. Also, physical binding interactions, regulatory associations, and co-expression correlations were analyzed.

Nevertheless, there are no sensitive and accurate biomarkers for colorectal adenocarcinoma diagnosis or prognosis. The enriched genes regulate Cellular biology, which contains transcription factors (TFs), as shown in [Figure 3](#) in a sum of functions. TFs are key players in the growth and tumorigenesis of colorectal adenocarcinoma since they are the molecules that make up the terminal effector signal pathway [16]. A rising amount of evidence indicates that peculiar TFs have a

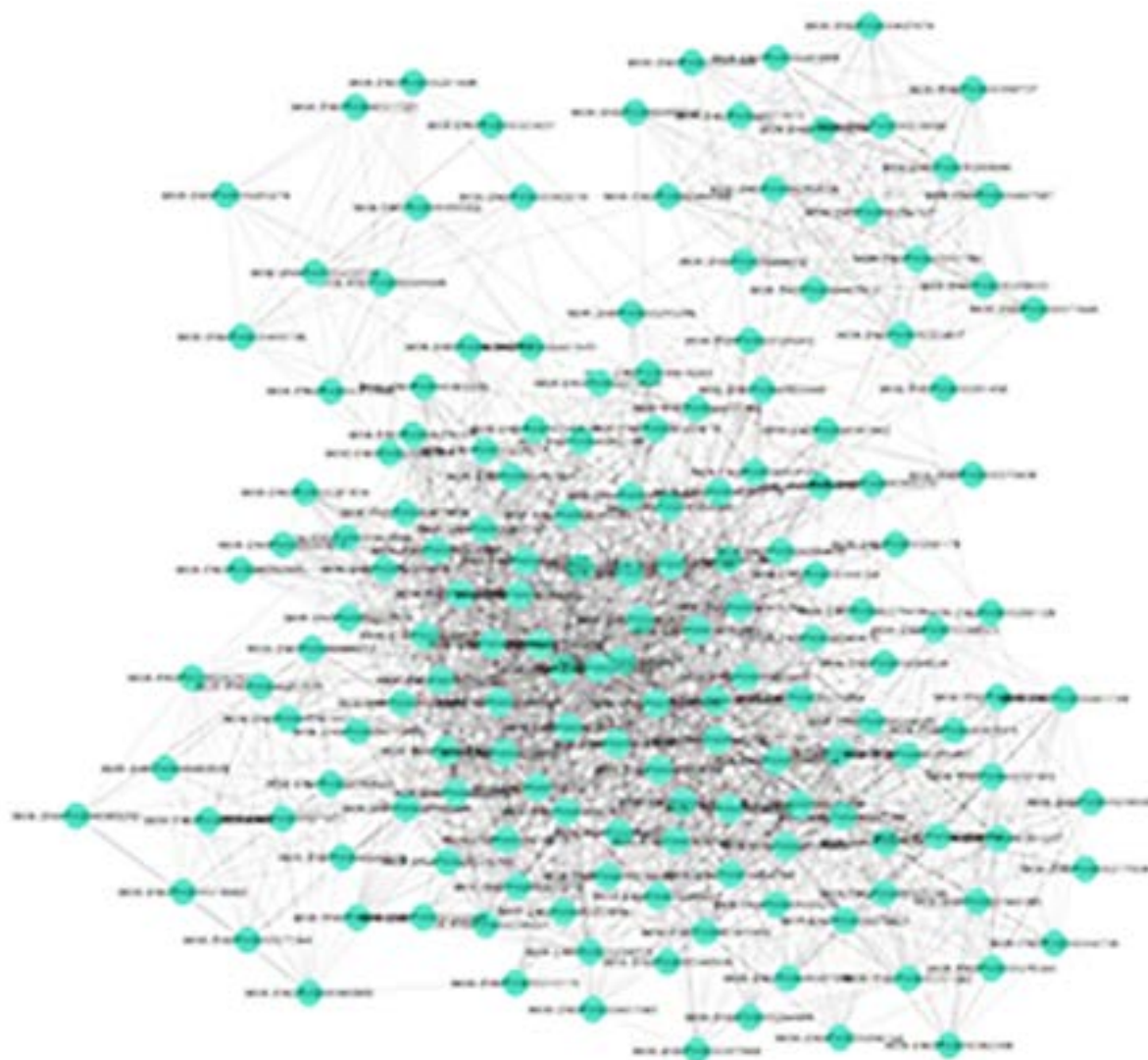


Figure 6A: Cytoscape software the network illustrates how the DEGs found from COAD interact with one another on a protein-protein level using Cytoscape. The edges are shown as lines gray and the nodes are ellipses Robin's green.

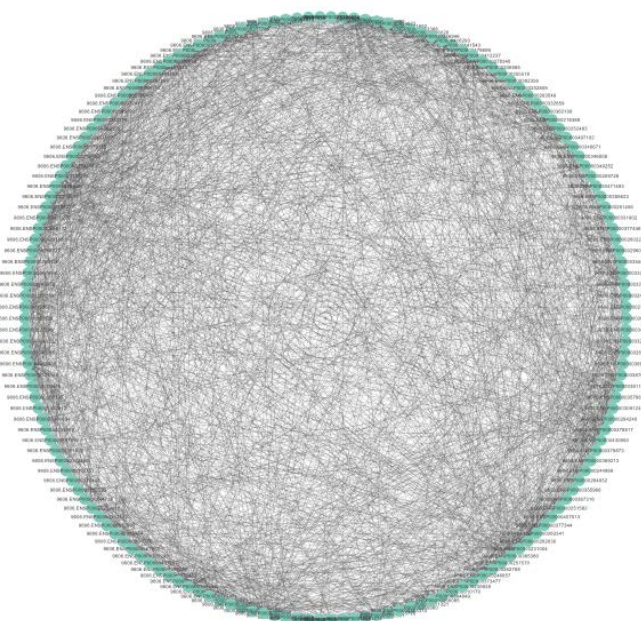


Figure 6B: Upregulated genes and downregulated genes in the regulated circular.

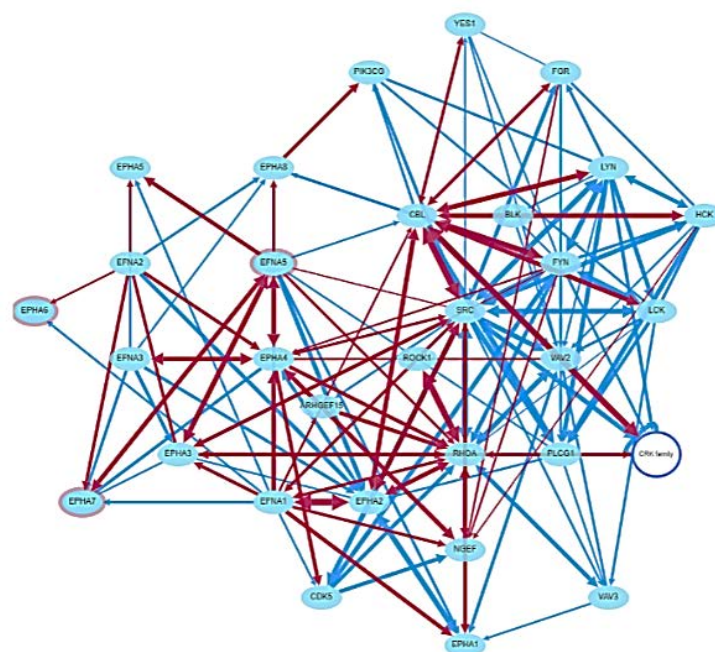


Figure 6D: Ten higher upregulated genes and ten higher downregulated genes are in a heatmap that is predicted by the limma package in R. The green color represents the normal sample. The pink color represents the tumor samples.

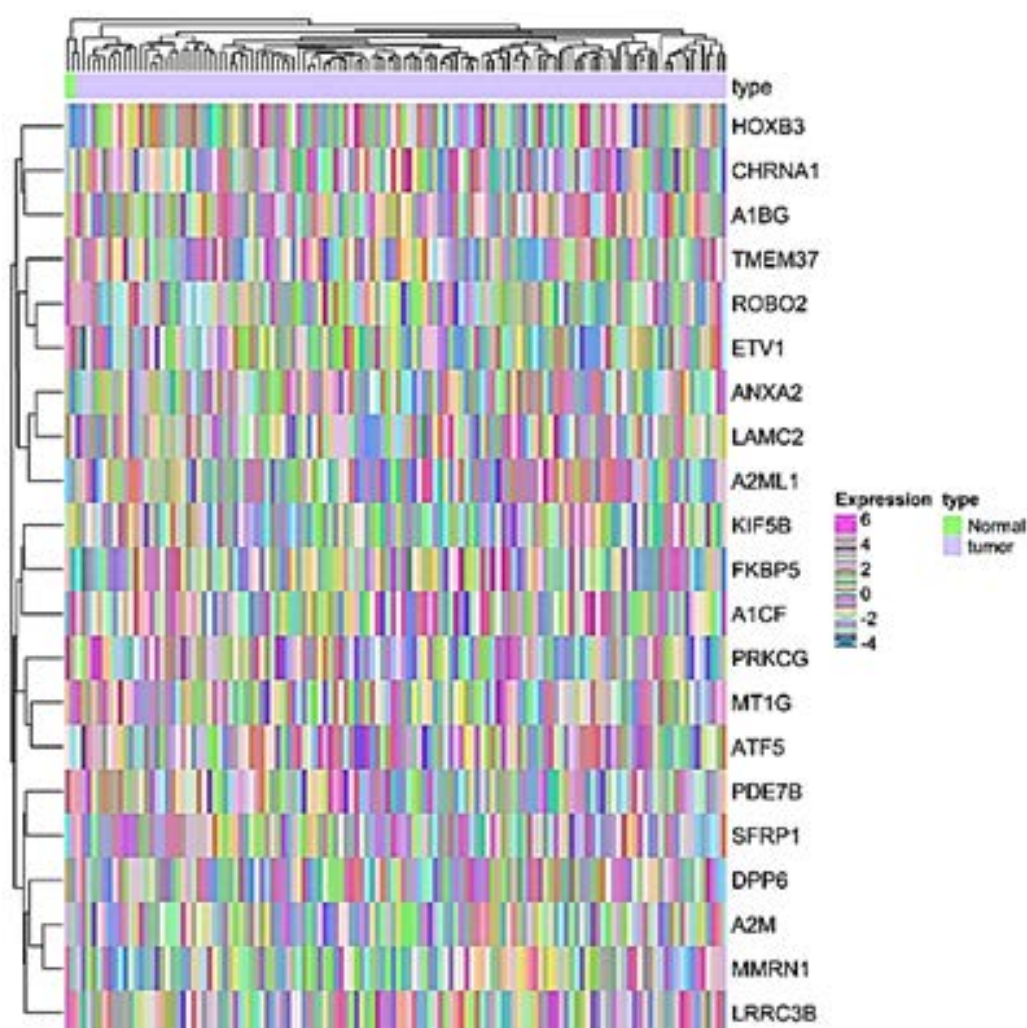


Figure 6D: Ten higher upregulated genes and ten higher downregulated genes are in a heatmap that is predicted by the limma package in R. The green color represents the normal sample. The pink color represents the tumor samples.

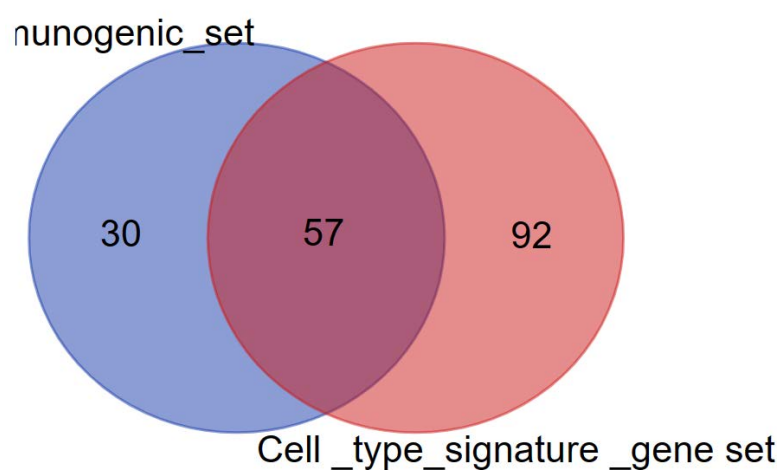


Figure 7A: Venn diagram immunogenic genes set versus cell type gene signature; the common genes were 57 common genes between cell type signature genes versus immunogenic genes.

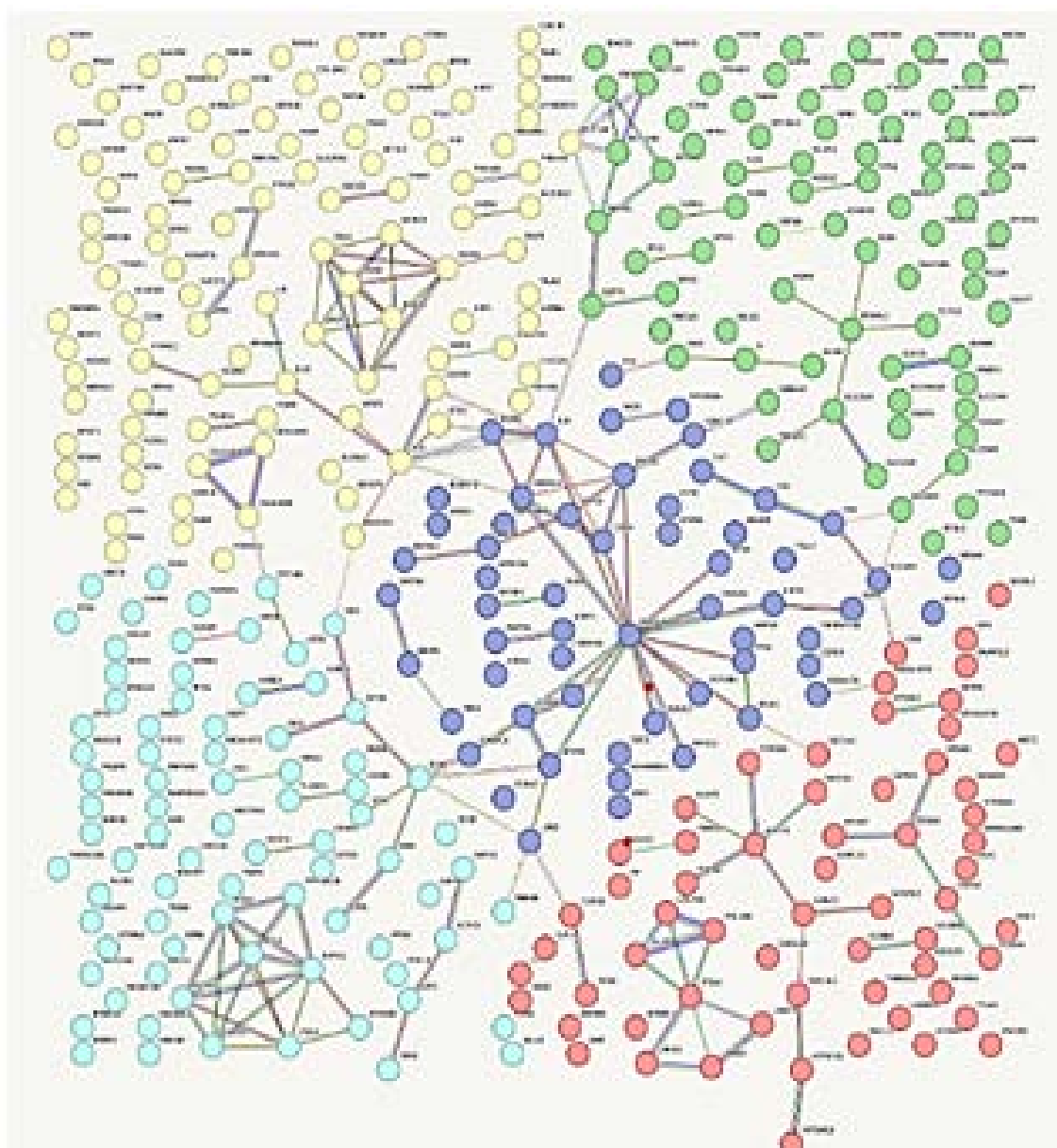


Figure 7B: String protein-protein interaction divided the upregulated genes into five clusters.

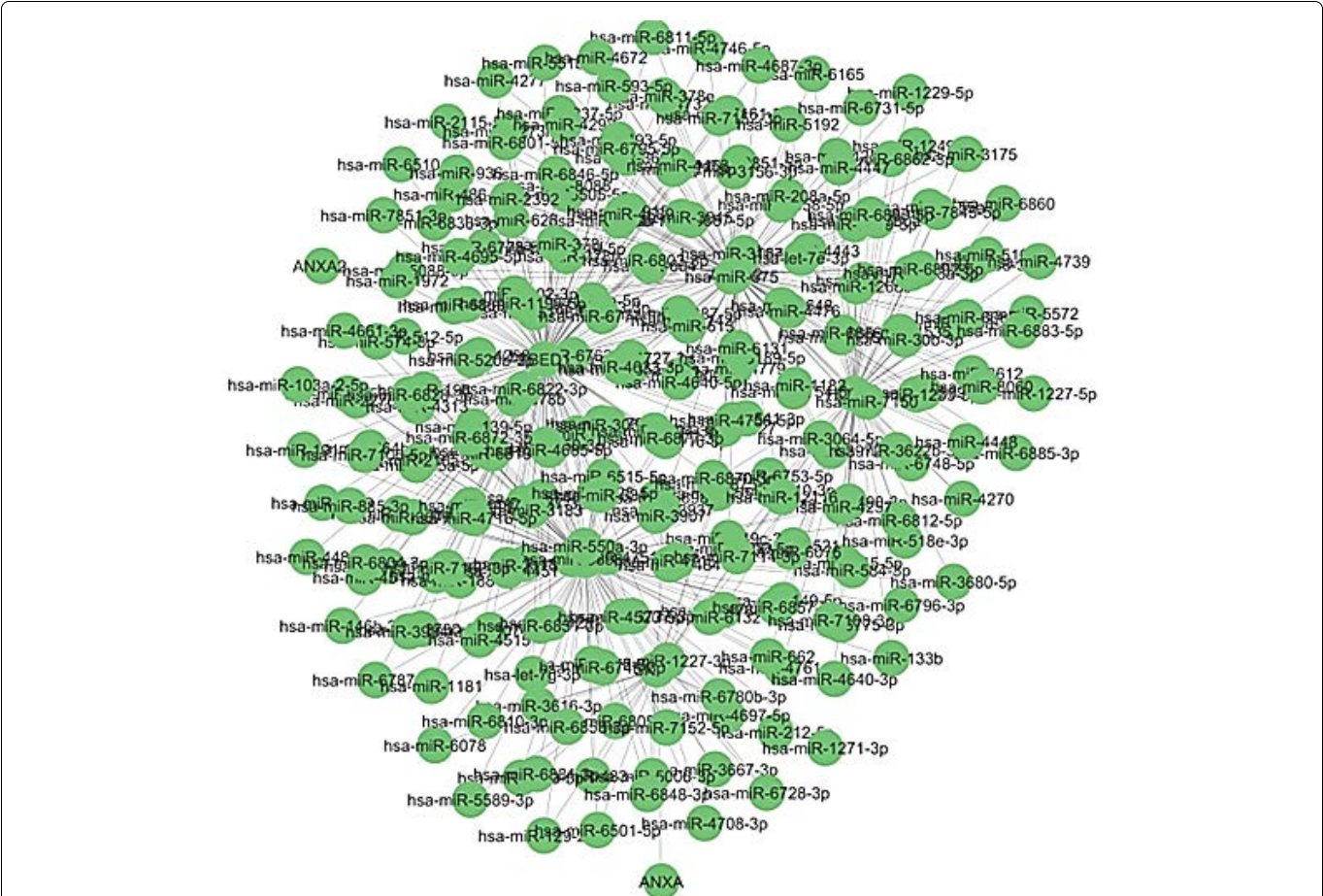


Figure 7C: Regulatory network of miRNA-Gene-TF in COAD between 250 differentially expressed mRNAs Green circles represent the differentially expressed target mRNAs.

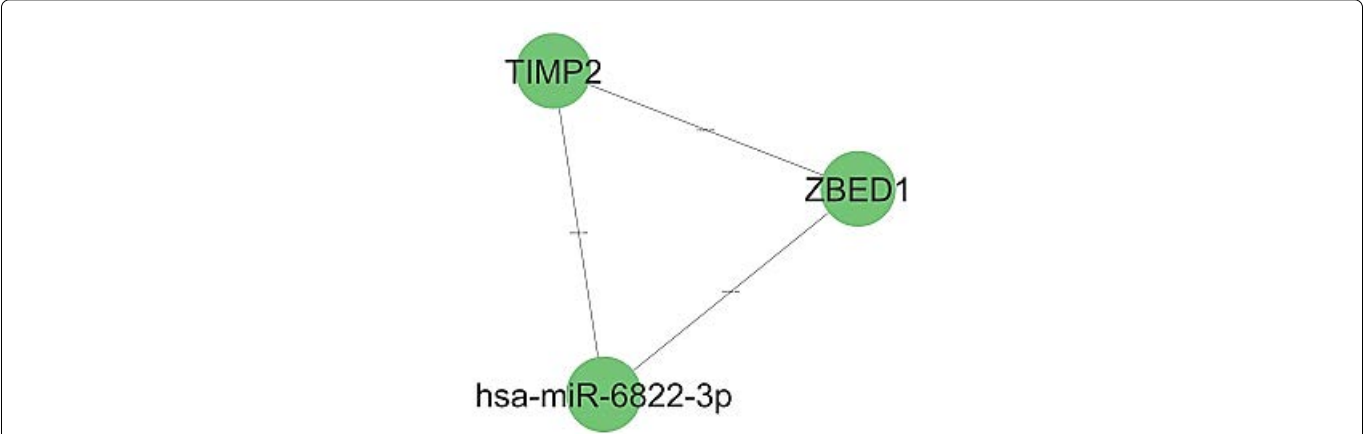


Figure 7D: The key gene TIMP2 has substantial TF and miRNA interactions.

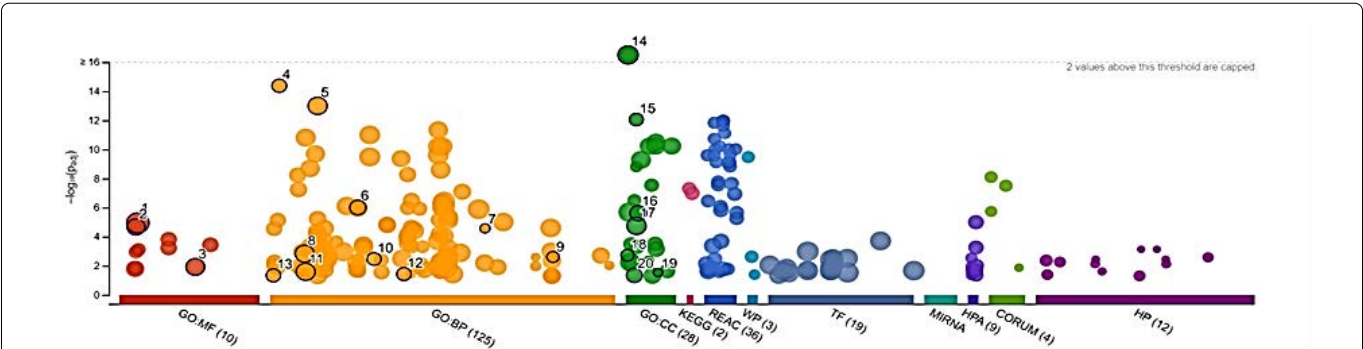


Figure 8A: Figure 8 g-profiler database: A identified each common group's functional enrichment for GO: MF GO: BP GO: CC KEGG and TF.

ID	Source	Term ID		Term Name	P _{adj} (query_1)
1	GO:MF	GO:0005515		protein binding	1.121×10 ⁻⁵
2	GO:MF	GO:0005198		structural molecule activity	2.138×10 ⁻⁵
3	GO:MF	GO:0044877		protein-containing complex binding	1.189×10 ⁻²
4	GO:BP	GO:0002181		cytoplasmic translation	4.166×10 ⁻¹⁵
5	GO:BP	GO:0009888		tissue development	1.003×10 ⁻¹³
6	GO:BP	GO:0030198		extracellular matrix organization	9.824×10 ⁻⁷
7	GO:BP	GO:0070934		CRD-mediated mRNA stabilization	2.654×10 ⁻⁵
8	GO:BP	GO:0007155		cell adhesion	1.456×10 ⁻³
9	GO:BP	GO:1901798		positive regulation of signal transduction by p53 c...	2.441×10 ⁻³
10	GO:BP	GO:0032963		collagen metabolic process	3.414×10 ⁻³
11	GO:BP	GO:0007267		cell-cell signaling	2.600×10 ⁻²
12	GO:BP	GO:0042471		ear morphogenesis	3.713×10 ⁻²
13	GO:BP	GO:0000956		nuclear-transcribed mRNA catabolic process	4.613×10 ⁻²
14	GO:CC	GO:0005576		extracellular region	9.685×10 ⁻¹⁸
15	GO:CC	GO:0022626		cytosolic ribosome	8.735×10 ⁻¹³
16	GO:CC	GO:0031012		extracellular matrix	2.393×10 ⁻⁶
17	GO:CC	GO:0030054		cell junction	1.814×10 ⁻⁵
18	GO:CC	GO:0001533		cornified envelope	1.843×10 ⁻³
19	GO:CC	GO:0070937		CRD-mediated mRNA stability complex	2.877×10 ⁻²
20	GO:CC	GO:0014069		postsynaptic density	4.569×10 ⁻²

Figure 8B: Viewed the 20 pathways of enriched genes and mutation types of gene structure. Those pathways are controlled by the upregulated genes expressed in colorectal adenocarcinoma.

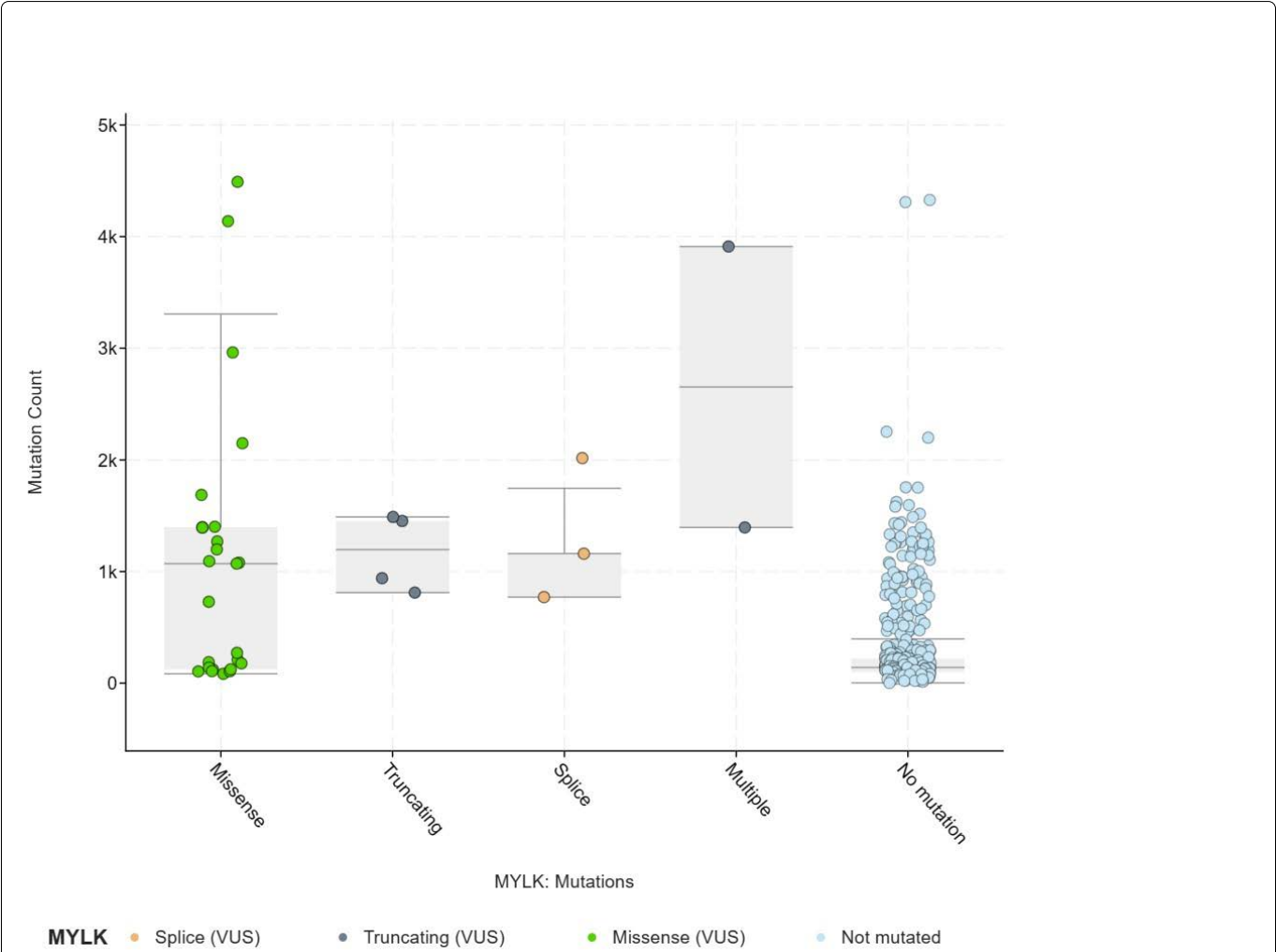


Figure 9A: C-bioportal Boxplot Comparative between the mutational and non-mutational part of MYLK genes.

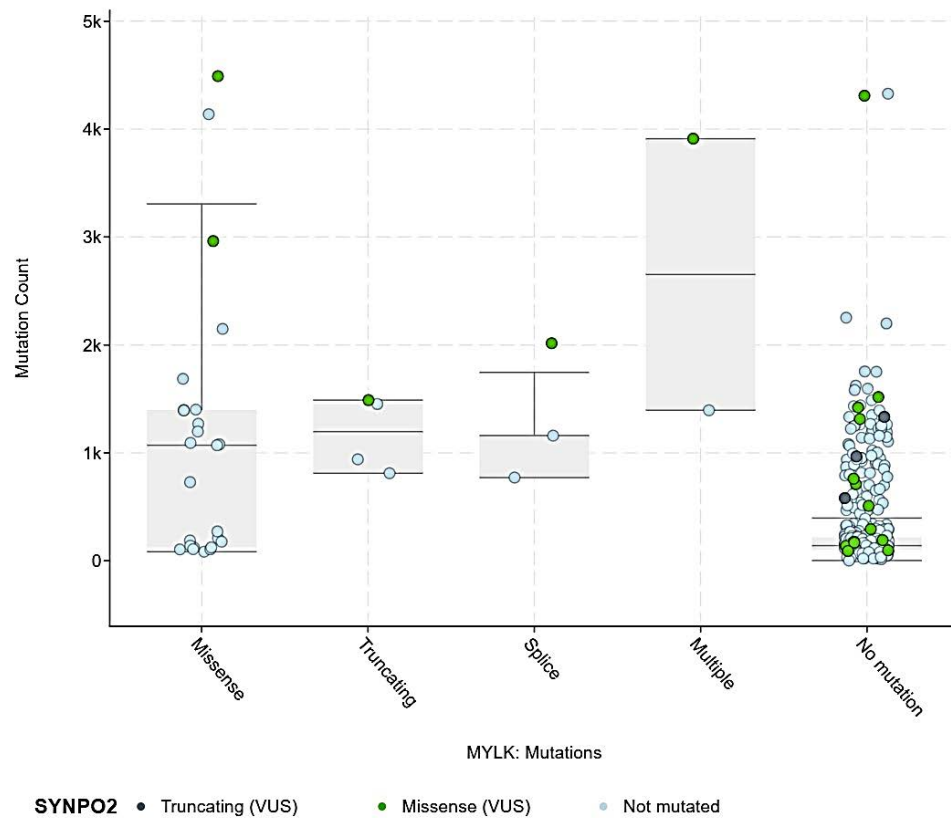


Figure 9B: Boxplot Comparative between the mutational and non-mutational parts of SYNPO2 gene.

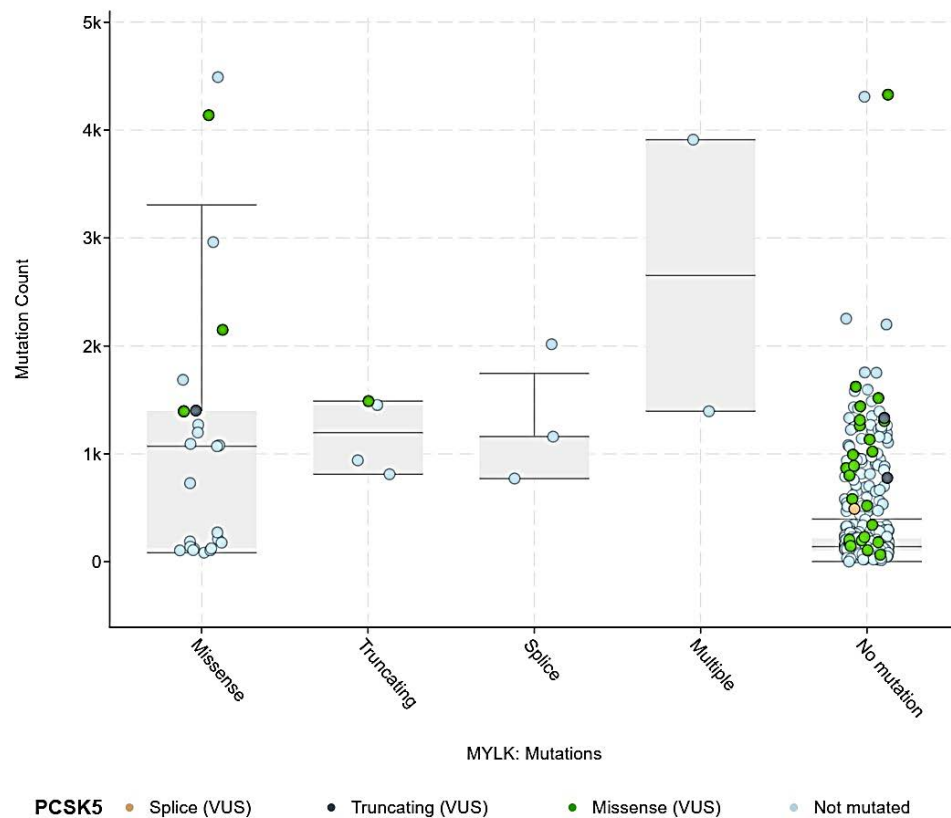


Figure 9C: Boxplot Comparative between the mutational and non-mutational parts of PCSK5 gene.

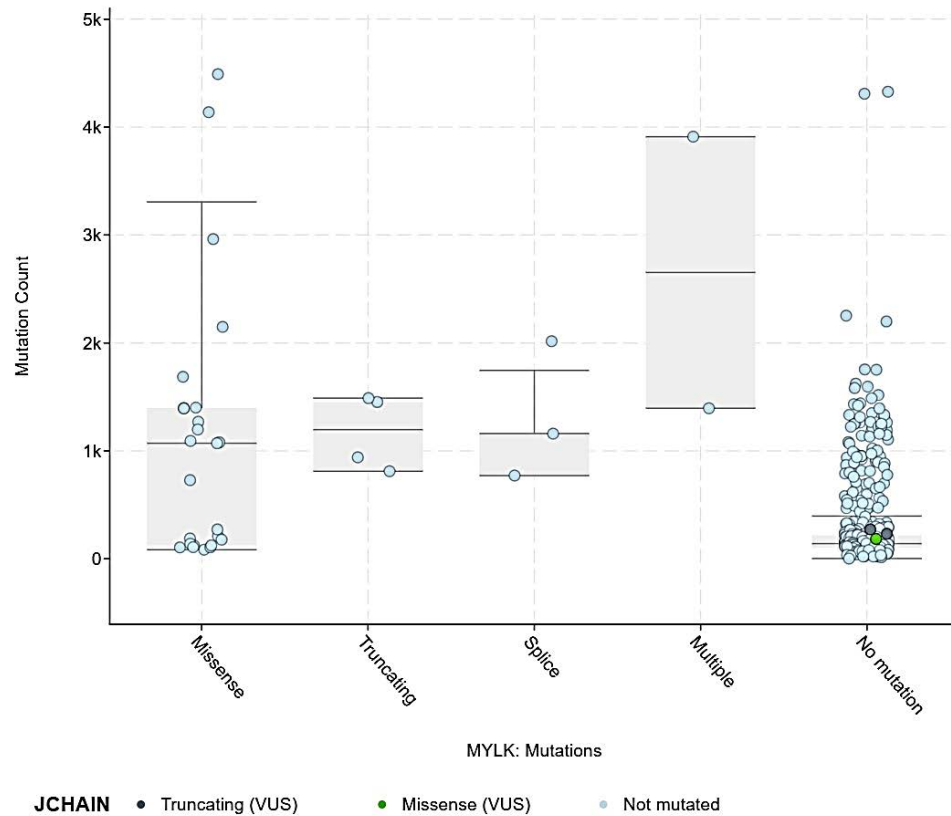


Figure 9D: Boxplot Comparative between the mutational and non-mutational parts of JCHAIN gene.

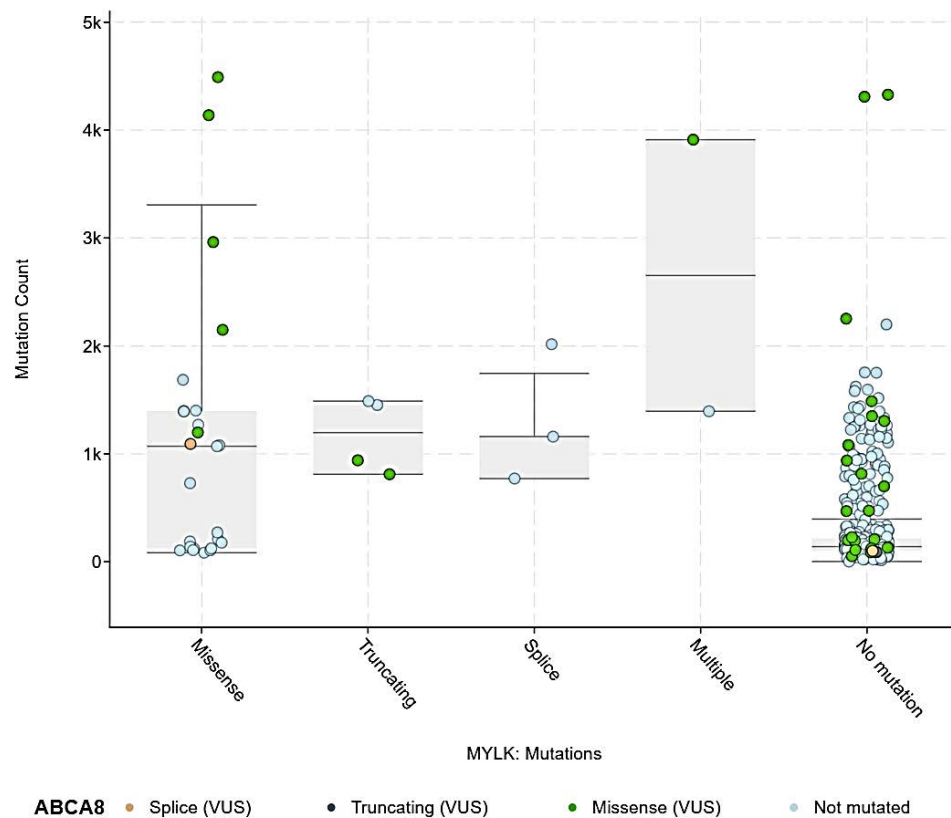


Figure 9E: Boxplot Comparative between the mutational and non-mutational parts of ABCA8 gene.

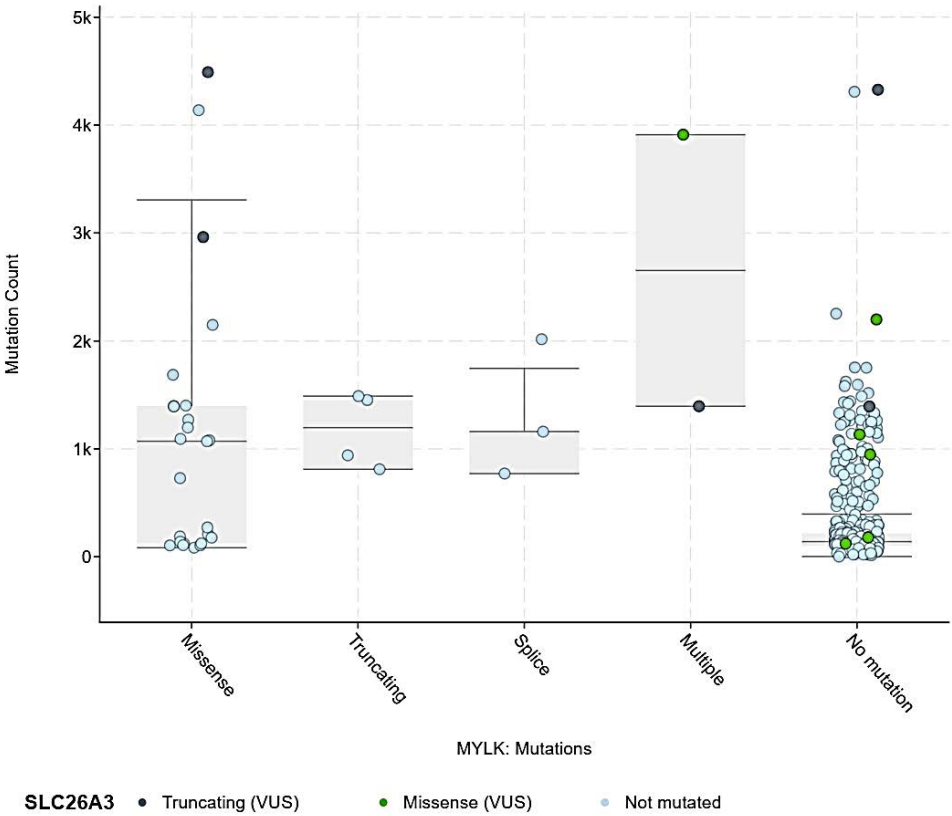


Table 1: Common genes upregulated in functional enrichment of GO: MF, GO: BP, GO: CC, KEGG and TF.

Genes	GO:MF	GO:BP	GO:CC	KEGG	TF
RPL12	IPI	NAS,IC	HDA	KEGG	TF
RPL23	IPI	NAS,IC	HDA	KEGG	TF
RPL26	IPI	IBA,IC	HDA	KEGG	TF
RPL3	IPI	IC	HDA	KEGG	TF
RPL24	IPI,HDA	IDA,IBA,NAS,IC	HDA	KEGG	-
RPL31	IPI	IDA,IBA,NAS,IC	HDA	KEGG	-
RPL37A	IPI	NAS,IC	HDA	KEGG	-
RPS19	IPI	NAS,IC	HDA	KEGG	-
RPS4X	IPI	IC	HDA	KEGG	-
RPS8	IPI	NAS,IC	HDA	KEGG	-
YBX1	IPI	IDA	IDA,HDA,TAS,IEA	-	TF
RPS7	IPI	NAS, IC	-	KEGG	TF
RPL7	IPI	NAS,IC	-	KEGG	TF
RPL38	IPI	IDA,NAS,IC	-	KEGG	TF

role in the development of COAD, its clinicopathological characteristics, and its prognosis [17,18]. That finding provides new insight into the basic pathogenesis, potential biomarkers, and therapeutic targets.

Prior research has demonstrated the extensive distribution of this local gene co-expression, as shown in Figure 4, over numerous chromosomes and tissues. However, it is still unclear if this co-expression results from transcription processes in individual cells because previous studies have only employed bulk RNA-seq, averaging gene expression values over millions of cells as this study. Here, but in a previous study [19], take

advantage of the co-occurrence of transcription events in single cells to uncover gene co-expression across cells, unbiased by cell-type heterogeneity. The study utilized single cell datasets spanning over 85 individuals. Find correlations between genes and regulators and show that over 95% of co-expressed gene pairs have similar regulatory elements. Co-expression in single cell gene regulatory networks and call for a more thorough examination of common regulatory components to help understand why diseases affect multiple genes at once [20].

As shown in Figure 5A, Functional enrichments were performed; common enriched diseases are malignant

neoplasm of the breast and liver Grrhosis, which is related to colorectal cancer. Identification of core gene families linked to colorectal cancer metastases to the liver and regulatory functions in tumor cell immune infiltration [21,22]. The discovery came when the CCL cluster, as a part of core genes, was primarily enriched in GO-BP pathways linked to immune cell migration and chemotaxis, including neutrophils, monocytes, and leucocytes. The leucocytes carries anticancer chemicals or medications to the targeted tissues in a moving manner, which allows it to carry out its anticancer functions, where consider the hand of the immune system to fight cancer in the human body [23,24]. KEGG: the upregulated genes expressed in colorectal adenocarcinoma depend on these pathways.

In Figure 5B and Figure 5C HPA tissues: The gene expression enriched in adipose tissues, adipocytes, and adrenal gland. The HPA tissues continued to function during the prolonged stress. Increases the response to stress in COAD cells and tumor tissues at both the mRNA and protein levels, as observed in Figure 5D: Upregulated tissues: the primary tissues that express the upregulated genes are the Brain, Testis, Kidney, and lung. As mentioned in many studies that HPA axis continued to function during the prolonged stress [25]. Increase the response to stress in CRC cells and tumor tissues at both the mRNA and protein levels. These facts proved that the tumor microenvironment is linked to the effectiveness of immunotherapy and patient survival in colorectal cancer patients [26]. Meanwhile, in Figure 5E: GOTERM-CC- direct cellular component, relation extraction between gene and other entities, F gene enrichment analysis, the main biological functional enriched genes, the main gain involved in the regulation of the transcriptome genes, and genes silencing by miRNA (Figure 5F). These gene enrichment analyses are used to identify differentially expressed key genes associated with colorectal cancer that can be used in tumor diagnosis and targeted gene therapy [27,28].

This work investigates the potential activities of upregulated and downregulated genes in COAD, including 913 upregulated and 906 downregulated genes, to provide a comprehensive knowledge of the genomic organization of COAD, as seen in Figure 6A and Figure 6B the distribution of the Differentially Expressed 1274 nodes and 7665 edges characterize genes (DEGs) in Cytoscape software. The heatmap generated using an R script, as shown in Figure 6D, shows ten higher upregulated genes and ten higher downregulated genes, indicating considerable DEGs of COAD relative to normal tissue. The heatmap provides a clear visual representation of how gene expression profiles are distributed among both normal patients and those with tumor patients [29].

The String protein-protein interaction database recognizes 5 cluster networks of the upregulated

genes shown in Figure 7A Furthermore, there are 57 upregulated common genes shown in Figure 7A, the Venn diagram that finds the common gene between the two sets of the immunogenic genes set and cell type gene signature. There have been several reports of immune cell transcriptome signatures, which consist of common gene lists indicating the presence of a specific immune cell population. Most of these gene signatures have been identified by examining separate blood cells [24]. A network-based deconvolution technique was used to obtain the set of immune gene signatures directly from tissue transcriptomics data, as mentioned in many studies [30,31]. As shown in Figure 7C and Figure 7D, Overexpressed microRNAs regulate receptor signaling pathways. Conversely, downregulated microRNAs target most of the functions linked to the mitotic cell cycle in human organs. The characteristics of cancer, maintaining proliferative signals, avoiding growth suppressors, stopping cell death, starting invasion and metastasis, and starting angiogenesis, among other things, have been demonstrated to be impacted by dysregulated miRNAs [32]. Two hundred and fifty miRNAs shown in Figure 6C are the overexpressed miRNAs that interacted with the higher upregulated genes and downregulated genes. As shown in Figure 7D, the biomarkers of the TIMP2 target gene are connected to its regulator (ZEBED1) transcription factors TF, and both are connected with miRNA (has-miR-6822-3p). Recent studies have demonstrated that there may be a negative correlation between the prognosis of COAD patients and the expression levels of TIMP1, according to Song, et al. [33] TIMP1 suppression may lead to enhanced apoptosis and decreased invasion, migration, and proliferation in vitro while weakening tumorigenesis and metastasis in vivo through the FAK-PI3K/Akt and MAPK pathways. When serum TIMP1 levels were higher, patients with colorectal cancer had shorter overall survival periods and lower tumor grades [34-36].

As shown in Figure 8 and Table 1, the protein-protein connection is further explained using ontologies to express biological knowledge. Gene Ontology (GO) is a resource that provides information about the function of gene products. These ontologies cover three domains: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). The process of assigning functional information to genes that is relevant to the literature using gene annotation GO [37]. All five of the chosen criteria were shown to exhibit upregulation of a total of 14 genes. Gene Ontology Molecular Function (GO: MF), Gene Ontology Biological Process (GO:BP), Gene Ontology Cellular Component (GO:CC), Kyoto Encyclopedia of Genes and Genomes (KEGG), and transcription factor (TF) analyses revealed the following genes as the most frequently upregulated genes as presented in table 1 and Figure 5. KEGG pathways and GO keywords were used in gene enrichment analysis

to identify genes that were consistently elevated in a variety of functional domains. Significantly, ribosomal function it was found that these upregulated genes are significantly associated with immune-related biological processes and pathways. This finding suggests a potential impact on the immune system in the context of colorectal adenocarcinoma (COAD).

Furthermore, as explained in [Table 1](#), the differentiation, progression, or metastasis of colorectal cancer are linked to the expression patterns of Ribosomal Proteins (RPs) 2 [38]. RPL3 (60S ribosomal protein L3, a component of the large subunit of cytoplasmic ribosomes), RPL12 (60S ribosomal protein L12, directly binding to 26S ribosomal RNA), RPL23 (Ribosomal protein L23), and RPL26 (60S ribosomal protein L26, a part of the large ribosomal subunit). The nucleolus initiates the development of functioning ribosomes, which are necessary for coordinated ribosomal RNA (rRNA) processing and ribosomal protein [31] assembly. This process is often hyperactivated to support the need for protein synthesis, which is necessary to survive the unrelenting expansion of cancer cells [38]. All categories show upregulation of RPL24 (Ribosomal protein L24; Belongs to the eukaryotic ribosomal protein eL24 family), which is a possible therapeutic target whose acetylation or depletion prevents the formation of polysomes and the development of cancer cells [39]. Ribosomal protein L31 (RPL31) is known to have the capacity to inhibit the proliferation and migration of Gastric cancer cells [30]. RPL31 (Ribosomal protein L31), RPL37A (Ribosomal protein L37a), RPS19 (40S ribosomal protein S19; required for pre-rRNA processing and maturation of 40S ribosomal subunits in colorectal cancer (CRC); Belongs to the eukaryotic ribosomal protein eS19 family), the eukaryotic protein uS19 is encoded in human disorders and is a part of the ribosome decoding site [31]. RPS4X (Ribosomal protein S4 X-linked; Belongs to the eukaryotic ribosomal protein eS4 family). The overexpressed multifunctional protein YB-1 has been found to be a partner of the X-linked ribosomal protein S4 (RPS4X), which is involved in cellular translation and proliferation in a number of breast cancer cells. In these cell lines, depletion of RPS4X consistently leads to cisplatin resistance [30]. It is observed that approximately 25% of human RPs show tissue-specific expression and the most intricate patterns of RP expression are found in primary hematopoietic cells [30], which transcriptional regulators probably influence with context restrictions. Surprisingly, the study also demonstrated a pattern of dysregulated expression of particular RPs across cancer types, patterns that are prognostic of disease development and derive from copy number differences. Such as RPS8 (Ribosomal protein S8; Belongs to the eukaryotic ribosomal protein eS8 family. Genes, except for transcription factor (TF) studies. RPS7 (40S ribosomal protein S7; required for rRNA maturation), RPL7 (60S ribosomal protein L7; Component of the large

ribosomal subunit, binds to G-rich structures in 28S rRNA and mRNAs.

It is observed that approximately 25% of human RPs show tissue-specific expression and the most intricate patterns of RP expression are found in primary hematopoietic cells, which transcriptional regulators probably influence with context restrictions. Surprisingly, the study also demonstrated a pattern of dysregulated expression of particular RPs across cancer types, patterns that are prognostic of disease development and derive from copy number differences [40]. Many studies observe that anticancer medicines that target mRNA translation are anticipated to become possible. That can be explained by the knowledge that a significant part of the pathophysiology of many malignancies is the abnormal translation of mRNAs into proteins, which involves gene transcriptions and post-translational modifications of proteins. That is described by the regulation of protein translation by the interplay between mRNA modifications and ncRNAs [40].

Upregulate gene plays a role in the translation apparatus; inhibits cell-free translation of mRNAs and RPL38 [40]. (Ribosomal protein L38; Belongs to the eukaryotic ribosomal protein eL38 family) genes upregulated in all the categories except in the Cellular Component (CC). Apart from KEGG pathways, all categories showed an upregulation of the YBX1 (Y-box-binding protein 1; DNA- and RNA-binding protein involved in various processes, such as translational repression. The first is the traditional general program that produces ribosomes, and the second is a gene-specific function that regulates the number of ribosomal proteins generated by each distinct ribosomal gene. This particular control of ribosomal protein genes is mainly accomplished by variations in pre-mRNA splicing efficiency and mRNA stability, in contrast to the general system, which is primarily regulated at the level of transcription and translation [41].

RNA stabilization, mRNA splicing, DNA repair, and transcription regulation. Predominantly acts as an RNA-binding protein: binds preferentially to the 5'-[CU] CUGCG-3' RNA motif and specifically recognizes mRNA transcripts modified by C5-methylcytosine) gene ([Table 1](#)). Genes RPL3, RPL12, RPL23, RPL26, RPL2415, RPL31, RPL37A, RPS19, RPS4X, RPS8, RPS7, RPL7, RPL3816 [42], and YBX1 were constantly elevated, indicating that they may be targets for therapy. In various studies, the RPL313 [43] and RPL717 Upregulation were detected in colorectal cancer. After delving into gene fusion occurrences, the RPL12 and RPL3 genes showed a fusion score of 0.003, suggesting that these changes might lead to altered regulatory features or capabilities. Except for YBX1 (0.4), the STRING network showed significant co-expression scores (0.8 to 0.9) for all genes, as observed in [Figure 8](#).

The pathogenesis of Colorectal Adenocarcinoma (COAD) is maintained in part by the pathways that interact between factors and cytokine receptors. These pathways have been connected with the development and progression of colorectal cancer of COAD) and share MYLK, SYNPO2, PCSK5, JCHAIN, ABCA8, and SLC26A3 as shown in Figure 9. Those genes happen to have many variants, including nonsense, insertion, frameshift, deletion, and splice variants. Differential routes were also noted at some phases, such as leukocyte migration across the endothelium and the emergence of Toll-like receptor signaling pathways. In Figure 9, the mutated genes were identified using c-bioportal. A study of Bladder cancer revealed that promoters like MYLK promote the invasion and migration of various solid tumors. Therefore, MYLK overexpression can serve as a biomarker for colorectal adenocarcinoma [44].

On the other hand, overexpression of the SYNPO2 gene is considered a poor prognostic factor for patients with nasopharyngeal cancer. Genes that showed a significant positive correlation with SYNPO2, based on GO biological process keywords, were primarily linked to cellular processes. Thus, SYNPO2 is an important indicator of nasopharyngeal cancer prognosis [45].

Meanwhile, PCSK5 downregulation enhances the inhibitory action of and rographolide on glioblastoma by controlling STAT3. Higher expression of PCSK5 was associated with a poorer prognosis for patients with advanced GBM stages. PCSK5 knockdown lessened the IL-6-induced epithelial-mesenchymal transition (EMT)-like characteristics of GBM cells. Additionally, when combined with Andro therapy, PCSK5 knockdown significantly reduced both the *in vivo* tumor development and the *in vitro* invasion and proliferation of GBM cells [46].

A recent study has identified JCHAIN as a novel B-cell prognostic biomarker that can predict overall survival in Head and Neck Squamous Cell Carcinoma (HNSCC) patients. In the low-risk group, JCHAIN was considered a highly expressed protective gene. A single-cell dataset study showed that B cells had significant expression levels of JCHAIN [47]. Additionally, through the Wnt/ β -catenin pathway, the tumor suppressor ABCA8 prevents colorectal cancer from progressing malignantly using ABCA8 as targeted gene therapy can prevent the growth and spread of xenograft tumors *in vivo* and reverse the epithelial-mesenchymal transition [48].

Moreover, overexpression of ABCA8 inhibited the invasion of cancer cells [49]. The SLC26A3 gene has been found to have 98 variants, and public databases suggested that the SLC26A3 promoter may interact with p65 [49], which led researchers to investigate whether NF- κ B/p65 directly binds to the SLC26A3 gene promoter element and controls the expression of colorectal adenocarcinoma. In the mammalian intestines, the PI3K/AKT pathway may have an impact on the SLC26A3

gene. Hence, these highly mutated genes also function as promoters for the development and progression of cancer.

Conclusion

In summary, the study offers an identification of key genes for colorectal adenocarcinoma to be validated as potential prognosis biomarkers. The discovered key genes, particularly those connected to fusion events and ribosomal functions; need to be looked into further to see how they contribute to the development of disease and to determine whether they should be the focus of targeted gene therapies. Twelve gene regulatory pathways and six diagnostic biomarkers (MYLK, SYNPO2, PCSK5, JCHAIN, ABCA8, and SLC26A3) were described in this study. mRNA transcripts included RPL3, RPL12, RPL23, RPL26, RPL2415, RPL31, RPL37A, RPS19, RPS4X, RPS8, RPS7, RPL7, RPL3816, and YBX1, which were shown to be consistently elevated. And particularly identifies mRNA transcripts modified by C5-methylcytosine). As the text demonstrates, both *in vivo* and *in vitro* investigations have confirmed that the TIMP2 target gene is linked to its regulator (ZEBED1) transcription factors TF, and both are linked to miRNA (has-miR-6822-3p), which was chosen and promoted as key genes.

Future Directions: The investigation of key genes in the prognosis of colorectal adenocarcinoma and targeted gene therapy should focus on elucidating molecular pathways, creating novel treatment strategies to improve patient outcomes, and implementing the study's findings in clinical settings in the future. Verification in larger patient cohorts with colorectal adenocarcinoma is necessary to confirm the significance and dependability of the identified genes and fusion events as potential therapeutic targets. It is possible to create hypotheses for further study using enrichment maps. The parameters used in the pathway enrichment study, such as the minimum and maximum pathway size or the chosen route databases, should be changed in order to assess how robust the conclusions are. These parameters are suggested to have the potential to impact the findings.

Contribution Statement

The authors, Elham Omer Mahgoub, participated in the conception, design, analysis, and interpretation of the data, besides writing and editing the manuscript; 2) Pooja Manka participated in the conception, design, analysis, and interpretation of the data besides writing and editing the manuscript; 3) Amit Komar participated in the conception design, analysis, and interpretation of the data besides writing and editing the manuscript; 4) Syed Asif Husain Naqv participated in technical support and data analysis and project managing; 5) Bharti Mittal participated in the conception design, analysis, and interpretation of the data supervising. This manuscript

has not been submitted, nor is it under review at another journal or other publishing venue. The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript. The following authors have no affiliations with organizations with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Disclosure Statement

The authors reported no potential conflict of interest.

Funder

Qatar Foundation, Qatar National Library, grand award number NPRP9-350-3-074

References

- Shinji S, Yamada T, Matsuda A, Sonoda H, Ohta R, et al. (2022) Recent advances in the treatment of colorectal cancer: A Review. *J Nippon Med Sch* 89: 246-254.
- He S, Li X, Zhou X, Weng W, Lai J (2023) Role of epithelial cell-mesenchymal transition regulators in molecular typing and prognosis of colon cancer. *J Gastrointest Oncol* 14: 744-757.
- Dai Y, Jiang Z, Qiu Y, Kang Y, Xu H, et al. (2022) Identification of key carcinogenic genes in colon adenocarcinoma. *Iran J Public Health* 51: 364-374.
- Yazdani F, Mottaghi-Dastjerdi N, Shahbazi B, Ahmadi K, Ghorbani A, et al. (2024) Identification of key genes and pathways involved in T-DM1-resistance in OE-19 esophageal cancer cells through bioinformatics analysis. *Heliyon* 10: e37451.
- Mahgoub EO, Cho WC, Sharifi M, Falahati M, Zeinabad HA, et al. (2023) Role of functional genomics in identifying cancer drug resistance and overcoming cancer therapy relapse. *Heliyon* 10: e22095.
- Chen L, Lu D, Sun K, Xu Y, Hu P, et al. (2019) Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* 692: 119-125.
- Mahgoub I, Bolad AK, Mergani M (2014) Generation and immune-characterization of single chain fragment variable (scFv) antibody recognize breast cancer cells line (MCF-7). *J Immunother Cancer* 2: P6.
- Pournoor E, Mousavian Z, Dalini AN, Masoudi-Nejad A (2020) Identification of key components in colon adenocarcinoma using transcriptome to interactome multilayer framework. *Sci Rep* 10: 4991.
- Xu M, Chang J, Wang W, Wang X, Wang W, et al. (2022) Classification of colon adenocarcinoma based on immunological characterizations: Implications for prognosis and immunotherapy. *Front Immunol* 13: 934083.
- Mahgoub EO, Kulkarni S (2023) Structural analyses of an epidermal growth factor receptor-specific single-chain fragment variable via an in silico approach. *J Vis Exp* 10: e65894.
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, et al. (2019) g: Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 47: W191-W198.
- Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, et al. (2023) g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* 51: W207-W212.
- Gureghian V, Herbst H, Kozar I, Mihajlovic K, Malod-Dognin N, et al. (2023) A multi-omics integrative approach unravels novel genes and pathways associated with senescence escape after targeted therapy in NRAS mutant melanoma. *Cancer Gene Ther* 30: 1330-1345.
- Khan AA, Huat TJ, Mutery AA, El-Serafi AT, Kacem HH, et al. (2020) Significant transcriptomic changes are associated with differentiation of bone marrow-derived mesenchymal stem cells into neural progenitor-like cells in the presence of bFGF and EGF. *Cell Biosci* 10: 126.
- Manoochehri H, Jalali A, Tanzadehpanah H, Taherkhani A, Saidijam M, et al. (2022) Identification of key gene targets for sensitizing colorectal cancer to chemoradiation: An integrative network analysis on multiple transcriptomics data. *J Gastrointest Cancer* 53: 649-668.
- Sajadi M, Fazilti M, Nazem H, Mahdevar M, Ghaedi K (2022) The expression changes of transcription factors including ANKZF1, LEF1, CASZ1, and ATOH1 as a predictor of survival rate in colorectal cancer: A large-scale analysis. *Cancer Cell Int* 22: 339.
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343: 84-87.
- Lin J, Cao Z, Yu D, Cai W (2021) Identification of transcription factor-related gene signature and risk score model for colon adenocarcinoma. *Front Genet* 12: 709133.
- Ribeiro DM, Ziyani C, Delaneau O (2022) Shared regulation and functional relevance of local gene co-expression revealed by single cell analysis. *Commun Biol* 5: 876.
- AbuQamar SF, El-Tarabily KA, Sham A (2021) Co-expression networks in predicting transcriptional gene regulation. *Methods Mol Biol* 2328: 1-11.
- Liu WQ, Li WL, Ma SM, Liang L, Kou Z-Y, et al. (2021) Discovery of core gene families associated with liver metastasis in colorectal cancer and regulatory roles in tumor cell immune infiltration. *Transl Oncol* 14: 101011.
- Mahgoub EO (2017) Single chain fragment variables antibody binding to EGF receptor in the surface of MCF7 breast cancer cell line: Application and production review. *Open Journal of Genetics* 7: 84-103.
- Mitchell MJ, King MR (2015) Leukocytes as carriers for targeted cancer drug delivery. *Expert Opin Drug Deliv* 12: 375-392.
- Mahgoub EO, Bolad AK (2014) Construction, expression and characterisation of a single chain variable fragment in the Escherichia coli periplasmic that recognise MCF-7 breast cancer cell line. *J Cancer Res Ther* 10: 265-273.
- Ilaslan E, Sajek MP, Jaruzelska J, Kusz-Zamelczyk K (2022) Emerging roles of NANOS RNA-binding proteins in cancer. *Int J Mol Sci* 23: 9408.
- Zhu C, Wang S, Du Y, Dai Y, Huai Q, et al. (2022) Tumor microenvironment-related gene selenium-binding protein 1 (SELENBP1) is associated with immunotherapy efficacy and survival in colorectal cancer. *BMC Gastroenterol* 22: 437.
- Xu F, Jiang L, Zhao Q, Zhang Z, Liu Y, et al. (2021) Whole-transcriptome and proteome analyses identify key differentially expressed mRNAs miRNAs lncRNAs and circRNAs associated with HCC. *Oncogene* 40: 4820-4831.

28. Mahgoub IO (2014) Design expression and characterization of a single chain fragment variable anti-mcf-7 antibody; A humanized antibody derived from monoclonal antibody. Hamad bin Khalifa University Press (HBKU Press).
29. Miao Y, Wang J, Ma X, Yang Y, Mi D (2020) Identification prognosis-associated immune genes in colon adenocarcinoma. *Biosci Rep* 40: BSR20201734.
30. Wu F, Liu Y, Hu S, Lu C (2023) Ribosomal protein L31 (RPL31) inhibits the proliferation and migration of gastric cancer cells. *Heliyon* 9: e13076.
31. Graifer D, Karpova G (2021) Eukaryotic protein uS19: A component of the decoding site of ribosomes and a player in human diseases. *Biochem J* 478: 997-1008.
32. Peng Y, Croce CM (2016) The role of MicroRNAs in human cancer. *Signal Transduct Target Ther* 1: 15004.
33. Song G, Xu S, Zhang H, Wang Y, Xiao C, et al. (2016) TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *J Exp Clin Cancer Res* 35: 148.
34. Giaginis C, Nikiteas N, Margeli A, Tzanakis N, Rallis G, et al. (2009) Serum tissue inhibitor of metalloproteinase 1 and 2 (TIMP-1 and TIMP-2) levels in colorectal cancer patients: Associations with clinicopathological variables and patient survival. *Int J Biol Markers* 24: 245-252.
35. Liu Y, Li C, Dong L, Chen X, Fan R (2020) Identification and verification of three key genes associated with survival and prognosis of COAD patients via integrated bioinformatics analysis. *Biosci Rep* 40: BSR20200141.
36. Mahgoub EO, Haik Y, Qadri S (2019) Comparison study of exosomes molecules driven from (NCI1975) NSCLC cell culture supernatant isolation and characterization techniques. *The FASEB Journal* 33: 647.22.
37. Ding R, Qu Y, Wu CH, Vijay-Shanker K (2018) Automatic gene annotation using GO terms from cellular component domain. *BMC Med Inform Decis Mak* 18: 119.
38. Nait Slimane S, Marcel V, Fenouil T, Catez F, Saurin J-C, et al. (2020) Ribosome biogenesis alterations in colorectal cancer. *Cells* 9: 2361.
39. Wilson-Edell KA, Kehasse A, Scott GK, Yau C, Rothschild DE, et al. (2014) RPL24: A potential therapeutic target whose depletion or acetylation inhibits polysome assembly and cancer cell growth. *Oncotarget* 5: 5165-5176.
40. Kang J, Brajanovski N, Chan KT, Xuan J, Pearson RB, et al. (2021) Ribosomal proteins and human diseases: Molecular mechanisms and targeted therapy. *Signal Transduct Target Ther* 6: 323.
41. Petibon C, Ghulam MM, Catala M, Elela SA (2021) Regulation of ribosomal protein genes: An ordered anarchy. *Wiley Interdiscip Rev RNA* 12: e1632.
42. Qin Y, Liu L, Hu X, Zhang Y, Pan Y, et al. (2023) Expression and bioinformatics analysis of RPL38 protein and mRNA in gastric cancer. *Cell Mol Biol (Noisy-le-grand)* 69: 256-261.
43. Pan R, Yu C, Shao Y, Hong H, Sun J, et al. (2022) Identification of key genes and pathways involved in circulating tumor cells in colorectal cancer. *Anal Cell Pathol (Amst)* 2022: 9943571.
44. Jin H, Liu B, Guo X, Qiao X, Jiao W, et al. (2023) MYLK and CALD1 as molecular targets in bladder cancer. *Medicine (Baltimore)* 102: e36302.
45. Ye G, Tu L, Li Z, Li X, Zheng X, et al. (2024) SYNPO2 promotes the development of BLCA by upregulating the infiltration of resting mast cells and increasing the resistance to immunotherapy. *Oncol Rep* 51: 14.
46. Gong H, Yang X, An L, Zhang W, Liu X, et al. (2025) PCSK5 downregulation promotes the inhibitory effect of andrographolide on glioblastoma through regulating STAT3. *Mol Cell Biochem* 480: 521-533.
47. Li K, Zhang C, Zhou R, Cheng M, Ling R, et al. (2024) Single cell analysis unveils B cell-dominated immune subtypes in HNSCC for enhanced prognostic and therapeutic stratification. *Int J Oral Sci* 16: 29.
48. Yang K, Li X, Jiang Z, Li J, Deng Q, et al. (2024) Tumour suppressor ABCA8 inhibits malignant progression of colorectal cancer via Wnt/ β -catenin pathway. *Dig Liver Dis* 56: 880-893.
49. Lv C, Yang H, Yu J, Dai X (2022) ABCA8 inhibits breast cancer cell proliferation by regulating the AMP activated protein kinase/mammalian target of rapamycin signaling pathway. *Environ Toxicol* 37: 1423-1431.