ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

DEVELOPING ALGORITHMS TO UNDERSTAND THE ROLE OF MICROBIOME AND OTHER OMICS IN CANCER DISEASE MECHANISM AND THERAPY

TAYYİP KARAMAN
PH.D. THESIS

DEPARTMENT OF MEDICAL BIOTECHNOLOGY

SUPERVISOR
Prof. Osman Uğur Sezerman

ISTANBUL-2024

ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

DEVELOPING ALGORITHMS TO UNDERSTAND THE ROLE OF MICROBIOME AND OTHER OMICS IN CANCER DISEASE MECHANISM AND THERAPY

TAYYİP KARAMAN

PH.D. THESIS

DEPARTMENT OF MEDICAL BIOTECHNOLOGY

SUPERVISOR
Prof. Osman Uğur Sezerman

ISTANBUL-2024

Department:	Medical Biotechnology Department
Program:	Medical Biotechnology Doctorate Program
Thesis Title:	Developing Algorithms to Understand the Role of Microbiome and Other Omics in Cancer Disease Mechanism
Student's name and Surname:	Tayyip Karaman
Date of Defence:	27/06 /2024

This is to certify that I have examined this copy of Ph.D. thesis. I have found that the prepared after fulfilling the specified requirements in the associated legislation before the final examining committee, whose signatures are below.

Jury Member (Head of the Defense)	Prof. Günseli Bayram Akçapınar Acıbadem Mehmet Ali Aydınlar University	Signature
Jury Member (Thesis Supervisor)	Prof. Osman Uğur Sezerman Acıbadem Mehmet Ali Aydınlar University	Signature
Jury Member	Asst. Prof., Dr. Sinem Ökten Okullu Acıbadem Mehmet Ali Aydınlar University	Signature
Jury Member	Prof. Esra Erdal Dokuz Eylül University	Signature
Jury Member	Dr. Yavuz Oktay Dokuz Eylül University	Signature

DECLARATION

I declare that this thesis work is my own work, I had no unethical behavior at any stage from the planning to the writing of the thesis, I obtained all the information in this thesis in accordance with academic and ethical rules, and I cited all the information and comments that were not obtained with this thesis work, and I provided resources in the list of references. I also declare that there was no violation of any patents and copyrights during the study and writing of this thesis.

31 / 05 / 2024

Tayyip Karaman

PREFACE AND ACKNOWLEDGEMENT

Firstly, I would like to thank my advisor both in thesis and in life Prof. Ugur Sezerman, my beloved teachers Prof. Günseli Bayram Akçapınar, Prof. Emel Timuçin and Assoc. Prof. Sinem Öktem Okullu for sharing their valuable knowledge, science experiences, advice, and support with me, in my post-graduation career. I also would like to express my most profound thanks to Şükrü Karaman.

I would like to thank TUBITAK for supporting the projects (TUBITAK 2553, TUBITAK 2247-c and TUBITAK 2244) that I was involved with.

My family that was born from friendship deserve a huge thanks for making me experience the real meaning of friendship. I would like to thank Abdullah Kebapçı, Ahmet Sezgin, Burak Akmeşe, Enes Karakoyunlu, Furkan Haşlak, Hasan Cömert, Mahmut Şenbalkan, Yelkan Kandemir.

I would like to thank my Acıbadem family, Merve Çelik Yamacı, Ozan Savaşan, Esma Aybakan, and Merve Aydın, for making the time that I spent in Acıbadem University meaningful and for sharing lots of cheery memories.

I would like to thank Görkem, Selma, Sümeyye, Tuğçe and Sümeyye Akçelik for their friendship and any support they provided. Moreover, dear Sezerman Lab. members Ali Arslan, Berk, Eray, Tuğçe, Berkay, İbrahim, Özdeyiş, Özlem, Sena, Selin, Hüseyin deserve special thanks for sharing all good and bad together. I also would like to thank Arge-Team volleyball members Berkay, Feyzanur, Gülçin, Sevgi, Tuba, Selcen.

Lastly, I would like to give the biggest thanks to my family members who made me who I am. Each article I publish, each project I am involved, each sentence I write, each word I use for scientific purposes, the success belongs to my beautiful mother Ayşegül Karaman, my honorable father Şenol Karaman, my lovely sisters Sümeyra Karaman and Zeynep Rana Karaman.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND SYMBOL

NGS	Next Generation Sequencing
DNA	Deoxyribonucleic acid
ddNTP	Didioxynucleotide
HGP	Human Genome Project
SNP	Single Nucleotide Polymorphism
CVP	Copy Number Variation
RNA	Ribonucleic acid
mRNA	Messenger ribonucleic acid
mi-RNA	Micro ribonucleic acid
sRNA	Small ribonucleic acid
nc-RNA	Non-coding ribonucleic acid
PCR	Polymerase Chain Reaction
UC	Unclassified
Bp	Base pair
CRC	Colorectal Cancer
CMS1	Consensus Molecular Subtype 1
CMS2	Consensus Molecular Subtype 2
CMS3	Consensus Molecular Subtype 3
CMS4	Consensus Molecular Subtype 4
NCBI	National Center for Biotechnology Information
OTT	Open Tree of Life Taxonomy
RDP	Ribosomal Database Projects
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
HMDB	Human Metabolome Database
r	Pearson Correlation coefficient

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Developing Algorithms to Understand the Role of Microbiome and Other Omics in Cancer Disease Mechanism and Therapy

Number of omics studies significantly increased after development of sequencing technology. Data obtained from microbiome, transcriptome proteome, metabolome and other omics studies play a crucial role in understanding the disease mechanism and developing new diagnosis and treatment techniques. Each omics study provides very important outcomes, however combining multiple omics data can give even more detailed and specific result. There are so many parameters that directly or indirectly affect the disease mechanism of occurrence, especially complex diseases like cancer. This is the reason why multi-omics analysis is quite significant for evaluating those parameters to find the problem or to find a treatment methodology specifically. In this thesis, we obtained microbiome and transcriptome data from colorectal cancer patients. In addition to that, we made a pathway prediction analysis and then these three omics data (microbiome, transcriptome, and pathway abundance data) were performed integration analysis.

Keywords: Microbiome, Transcriptome, Pathway analysis, Multi-Omis data integration, Colorectal cancer.

ÖZET

Mikrobiyom ve Diğer Omik Çalışmalarının Kanser Mekanizmasındaki ve Kanser Terapisindeki Rolünü Anlamak için Kullanılan Algoritmaların Geliştirilmesi

Sekanslama teknolojilerinin gelişmesi ile omik çalışmaların sayısında çok ciddi bir artış görülmüştür. Mikrobiyom, Transkriptomik, Proteomik, Metabolomik ve diğer omik çalışmalar sonucunda elde edilen büyük veriler hem hastalıkların mekanizmalarının çözülmesinde hem de tanı ve tedavilerin geliştirilmesinde oldukça önemli rol oynamaktadır. Her bir omik çalışmadan çok değerli bulgular elde edilse de bu omik çalışmaları bütün olarak ele almak hastalık mekanizmalarının anlaşılmasında daha belirgin sonuçlara ulaşabilmeyi sağlayacaktır. Özellikle kanser gibi kompleks hastalıklarda, hastalığın oluşumuna etki eden birden çok parametre bulunmaktadır. Omik çalışma sonuçlarının makine öğrenmesi yardımı ile entegre edilmesi hangi etkenlerin hastalıkta rol aldığını tespit edebilmek adına oldukça özgül sonuçlara ulaşabilmeyi sağlayacaktır. Bu tez çalışmasında, kolorektal kanser hastalarından elde edilen mikrobiyom transkriptom ve yolak verilerinin entegrasyonu güdümsüz ve güdümlü makine öğrenme teknikleri kullanılarak bu verilerin birbirleri ile olan ilişkisi analiz edilmiştir.

Anahtar Sözcükler: Mikrobiyom, Transkriptom, Yolak analizi, Çoklu omik veri entegrasyonu, Kolorektal kanser

1 INTRODUCTION AND AIM

The next generation sequencing (NGS) has revolutionized the many biological research, especially in omics studies such as genomics, metagenomics, transcriptomics, proteomics etc. Rapid and accurate sequencing of genomes facilitated the identification of complex genetic traits, microbial communities, gene expression levels, disease pathways and disease mechanism of action. Data obtained from these omics' studies provide valuable insights into biological mechanisms. However, pivotal approach which is the integration of multi-omics data allows researchers to gain a more comprehensive understanding of complex biological systems.

Omics data sources, when analyzed individually, provide valuable insights into biological mechanisms. However, integrating multi-omics data has emerged as a pivotal approach, allowing researchers to gain a more comprehensive understanding of complex biological systems. Omics data integration facilitates the reconstruction of intrinsic molecular networks and signaling pathways implicated in diseases. Scientists unravel the interaction between genes, proteins, metabolites, microbial communities, and regulatory elements, by overlaying the information obtained from different omics. This integration helps researchers to reach more robust biomarkers to develop a more specific approaches in diagnosis, treatment and understanding the diseases mechanism of action.

In recent years, many studies that indicates the importance of multi-omics data integration were published. Despite its promises, omics data integration present challenges. These are data heterogeneity, integration methodologies, computational complexities, and validation. Advancements in bioinformatics, machine learning and statistical methodologies are essential for refining integration techniques and establishing robust workflows for reliable and reproducible analyses.

Integration of omics data is especially important for complex diseases like cancer. Because there is more than one parameter that affects the cancer occurrence

and progress. Evaluating cancer from a single point of view may give valuable outcomes however the result obtained from single omics always have missing information. Lacking information in understanding the disease mechanism may lead less accurate diagnosis and treatment approaches. Combining omics data can overcome this problem and it can provide a development in personalized medicine. With this thesis, best standard machine learning approaches were found for the integration of microbiome, transcriptome and pathway abundance data obtained from colorectal cancer patients.

2 BACKGROUND

2.1 Bioinformatics in Biotechnology

Biotechnology is the usage and enhancement of technology to solve biological problems in life sciences. From agriculture to medicine, it has a broad range of applications. It enables us to understand environmental changes, microbial life, organism's genes, functions of proteins, disease mechanism of occurrence, diagnosing and treatment approaches, drug discovery and development, waste cleanup etc. (1). Bioinformatics is one of the most important branches of biotechnology to convert biological information, biological data into meaningful outcomes by applying statistical and computational algorithms (2). It became popular especially after human genome project. However, the concept of bioinformatics has started in early 1960s with the analysis of protein sequences. Insulin protein structure was determined in late 1950s via crystallography and it encouraged the development of Edman degradation method which allowed scientists to sequence proteins. More than 10 protein families were sequenced in following years. This is the way bioinformatics was born. Discover of deoxyribonucleic acid (DNA) and increasing number of microbial studies indicated the significance of bioinformatics (3). These studies were followed by a milestone discovery of DNA Sanger Sequencing developed by Frederick Sanger and his colleagues. Sequencing technology enabled scientist to evaluate the nucleic acids and we have started to learn our DNA codes from these genomic data. Enhancement in sequencing technology led us to the Next Generation Sequencing (NGS) and this application became much faster and cheaper. It provided scientist to do more research on genomes and this gave rise to an exponential increase in the amount of data obtain from this research (4). After the huge development in computing technology, bioinformatics took a crucial role in analysis of all biological research to deal with the complex sequences.

2.2 Sequencing Technology and Omics Studies

Deciphering the genetic code of organisms has started a new age in science. Sequencing technology provided to read nucleic acids of the organisms and it pioneered the discovery of many diseases, development of diagnosis and treatment approaches, understanding the biological mechanisms and genetics. It has started in 1977 by the discovery of Sanger Sequencing method introduced by Frederick Sanger who became a Nobel prize winner in chemistry. The method he developed was based on DNA chain termination with dideoxy nucleotides (ddNTP), allowed for the determination of precise sequence of nucleotides in DNA strands (4, 5). Then, Human Genome Project (HGP) was launched in 1990 to sequence entire human genome. Before this breathtaking research finished, Next Generation Sequencing technology (NGS) was developed. NGS both lowered the cost of sequencing and reduced the time that is consumed during sequencing. This technology was spread all around the world and research studies about genomics, proteomics, metagenomics, transcriptomics, and metabolomics accelerated rapidly (6).

Advancement in NGS has led to generation of high-throughput data to be used in pharmaceutics, therapeutics, diagnostics, developmental biology, and other biotechnological applications (7). Genomics is the studying of entire genome of an organism. Analyzing entire genome is being used for identifying single nucleotide polymorphism (SNP), loss of variants, copy number variations (CVP), rare variants and DNA methylation level. This information obtained from genomic studies help scientist to understand the various types of diseases' mechanisms and provide novel therapeutics, diagnosis. Transcriptomics is the study of ribonucleic acids (RNAs) such as messenger RNA (mRNA), micro-RNA (miRNA), small RNA (sRNA), non-coding RNA (ncRNA) etc. It is widely used for understanding the gene expression levels by sequencing the mRNAs (6, 8). Reduction of time and cost caused that transcriptomics to take the place of microarray studies by the development of sequencing technology and enhanced the gene expression and disease relationship studies (8). Metagenomics is the study of genetic materials that are obtained from environmental sources. Analyzing the genetic materials of microbes (such as

bacteria, viruses, parasites, fungus, yeasts etc.) that belongs to a certain habitat is called as microbiome analysis. It is difficult to grow the organisms by using laboratory culturing and to detect Polymerase Chain Reactions (PCR) dependent methods. These methods provide only limited number of microbes to be identified. That is the reason why finding another approach to overcome this limitation was necessary. Millions of bacteria can be detected in a short period of time at each taxonomic level (phylum, order, class, family, genus, species) thanks to the sequencing methods. One way to obtain microbial population is the shotgun sequencing approach. This method gives sequence result of entire genome of microbes and this cause high price and consuming of time. It is still being used but only for certain projects to identify exact biomarkers. Accurate, high-throughput, fast and cost effective 16SrRNA sequencing method is commonly preferred one for the determination of microbial abundance. 16SrRNA gene which is presence in all bacteria and differentiate within species provide this easy approach. This gene has nine variable region and sequencing only this gene after PCR amplification is sufficient to detect bacteria communities (9, 10). Human gut microbiome analysis indicated the significance of the presence of microorganism that are living inside our body. Secreted proteins and metabolites from millions of microorganisms cause a huge impact on humans' life. Besides microbes' main role in digestion mechanism of humans, they also take an essential part in defense mechanism, metabolic regulation, and drug metabolism (11). Therefore, dysbiosis of microbiome cause severe effects and found to be related with many diseases, like cancer, neurodegenerative diseases, immune disorders etc. Developing platforms like PICRUSt2 (12), Piphillin, (13) PUMAA (14), iVikodak (15) and BURRITO (16) enables microbiome researchers to get pathway information obtained from microbiome results. This unravels the mystery behind interaction of microbes and host through the pathways.

2.3 Cancer and Omics Studies

Cancer is the most complex and heterogeneous disease that cause almost 10 million of humans' death all around the world each year. Its structure, mechanism of occurrence, mechanism of action, response to the drugs, genomic variation,

metabolic and microbial effects are highly changeable from one patient to another. (17). Colorectal cancer is taking the third place in cancer mortality causation worldwide. Tumors that are detected in intestines are formed because of host genetic mutations and/or microbial effects. One of the most studied gut bacteria specie is *Fusobacterium nucleatum*. It initiates or progress tumor formation in intestinal tissues via metabolism, virulence factors, immune modulation, and miRNAs. FadA, Fap2 and OMVs(proteases) are the main virulence that are directly involved in B-catenin signaling activation pathway and natural killer cells inactivation and provide bacteria localization and invasion. On the other hand, overexpression of CLDN1 genes in epithelial tissue of intestines causes invasiveness of cancer cells. It is one of the most prognostic biomarkers in CRC patients (18).

Development of sequencing technology provided detailed understanding of complex diseases. Increasing number of omics studies enlightened very significant pathways that are associated with the cancer mechanism. Colorectal cancer has four subgroups which are consensus molecular subtype 1 (CMS1), consensus molecular subtype 2 (CMS2), consensus molecular subtype 3 (CMS3) and consensus molecular subtype 4 (CMS4). These subgroups are determined based on patient's cancer cells pathway activation and molecular function. These parameters are determined based on transcriptome and/or microbiome data obtained from tumor tissue. Immune infiltration and activation found to be associated with CMS1 subgroup, Wnt and Myc activation and some cell cycle signatures are associated with CMS2 subgroup, metabolic deregulation belongs to the CMS3 subgroup and Stromal infiltration, TGFβ activation and angiogenesis are seen in CMS4 subgroup (19). This subgrouping provides a better way of diagnosis and treatment. However, individual evaluation of each parameter that takes a role in progression of cancer is not enough to understand whole mechanism. Therefore, integration of omics datasets is essential to understand the details of pathways and association between gene products of host, metabolites (produced by both host and microbial community), gene products of microbial community and pathways. Integration of meta datasets provides to gather all significant information related with the disease of interest and enables researchers to comprehensive perspective with remarkable outcomes. Despite the fact that this

approach is quite valuable for each disease, it is essential for cancer because of its complexity. Pathways disrupted by cancer differentiate from patient to patient and this cause a trouble for diagnosing and choosing a treatment approach. Same approach may not work on different patients because of heterozygosity and that's the reason why specific information for each patient is necessary. Integration of multi-omics data will contribute personalized medicine in recent future.

2.4 Multi-Omics Data Integration

Integration of omics data has broad range of studying areas, including food science and industry, disease biology, genomics and metagenomics, pathway and network interactions, systems microbiology, drug, and industrial product discovery. [20] Implementation of different types of data supplies a comprehensive look to a research question and detailed results thanks to help of machine learning techniques. Usage of this approach in disease biology is thought as the beginning steps of personalized medicine. The reason for this saying depends on improvement in computational algorithms and development in omics databases that provide a huge source (21, 22). Important databases that are used in omics studies are NCBI, Open Tree of life Taxonomy (OTT), SILVA, Ribosomal Database Project (RDP), Greengenes, Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database, Gene Ontology (GO), Human Metabolome Database (HMDB), UniProt, EMBL (23, 24, 25, 26, 27, 28, 29, 30, 31).

The data obtained from omics studies is quite large in size which makes it difficult to manipulate for further analysis. There are many approaches to be used in integration of omics data such as dimension reduction approaches(Principle Component Analysis (PCA), Principle Co-ordinate Analysis (PCoA)), correlation based approaches(Spearman correlation, Pearson correlation, Canonical correlation analysis (CCA), Co-inertia Analysis (CIA)), regression based approaches (Generalized Linear Mixed Model(GLMM), Partial Least Square (PLS), Least Absolute shrinkage and Selection Operator (LASSO), Kernal Association Test) and network based approaches (Transkingdom network, Weighted Gene Co-expression

Network Analysis(WGCNA), Genome Scale Metabolic Models(GEM)) (32). All these unsupervised and supervised machine learning methods are recently being used for various types of omics data for discovery of biomarkers, finding an association between datasets, comparison of samples with healthy controls or with subgroups, better visualization to interpreting the data and discovery of pathway interactions.

Nevertheless, there is no golden standard for the usage of these techniques for integration. This results from types of the datasets, high dimensionality of datasets, handling missing values and reduction of noisy variables in data. Partial Least Square regression has been widely used for handling multi-covariate datasets in different kind of scientific areas. Its computational power for high dimensional data provides a remarkable variable selection (33, 34). This technique relies on cross validation based methods that are efficient as tuning parameters. Although, its performance on high dimensional data is pretty acceptable, implementation of datasets is still an issue while integrating multi-dimensional omics data.

3 MATERIALS AND METHODS

The data is obtained from “Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer” study (19). They used transcriptome and microbiome data for tumor tissue obtained from colorectal cancer patients. Unsupervised and supervised machine learning models were used for the integration analysis.

3.1 Data Source and Patient Cohort

The transcriptome and microbiome datasets were obtained from "Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer" paper, upon author's request. Metadata file that includes patient's gender, CMS subtypes, cancer stage, age, tumor side and tumor site information were sent by author to be used for classification analysis. Microbial pathway abundance dataset was generated by using the PICRUST2 (ver. 2.4.1.) plugin of QIIME2 (ver. 2023.5) (12)(35). Transcriptome, microbiome and pathway abundance datasets were collected at the end of downloading and generation steps for the integration.

Data obtained from “Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer" study contains 34 colorectal cancer patients. Untreated tumors were collected from each patient with approval from University of Otago Human Ethics Committee. One of the patient's samples were excluded from the study because of degradation in RNA sample, transcriptome data couldn't be obtained. Integration analyses were carried on by using 33 samples transcriptome, microbiome, and pathway abundance datasets. Age of the patients are in range of 44-88 years, fourteen of them were male and twenty of them were female. Six of them belonged to the CMS1 group, thirteen of them belonged to CMS2, nine of them belonged to CMS3 group and remaining five were recorded as unclassified (UC). Six of the patient’s tumor were at stage1, thirteen of them were at stage2, thirteen of them were at stage3 and one of them were at stage4.

3.2 Microbiome Analysis

Illumina Miseq platform was used for amplicon sequencing of V3-V4 16SrRNA target region amplified by PCR to obtain 250bp long reads. FLASh (ver. 1.2.11) was used to merge forward and reverse reads. QIIME2 (ver. 2023.5) was used to remove chimeric sequences. Rest of the analysis were continued with QIIME2 platform. DADA2 (ver. 3.19) was used for quality control and removing chimeric sequences. Taxonomic analysis was done by q2-feature-classifier plugin of QIIME2 by taking Greengenes database as a reference and OTU tables were obtained. Differential abundance testing was applied by ANCOM-BC (ver. 3.19). Biom files obtained from QIIME2 analysis were used for microbial pathway prediction analysis.

3.3 RNA-Seq Analysis

Illumina HiSeq platform was used for sequencing of the tumor samples, to obtain 125bp long paired end reads. Fastq-mcf (ver. 1.1.2.537) was used to remove adopter sequences. SolexaQA++ (ver. 3.1.6) was used for trimming of long reads, reads shorter than 50bp and low-quality reads. STAR (ver. 2.5.2b) mapping tool was used for the mapping of the filtered reads; the reads were mapped to human genome reference GRCh38. Samtools (ver. 1.3.1.) were used for merging the reads and htseq-count. Raw counts were transformed to TPM values by using R programming language and Bioconducter package Deseq2 (ver. 1.10.1) (36) The RNA-Seq analysis was concluded with pathway enrichment analysis by using pathfindR package (ver. 2.1.0.) in R (37). Pathway IDs were retrieved from KEGG database with false discovery rate (FDR) <0.05, pathways with FDR-adjusted p-value < 0.05 were used for further analysis.

3.4 Data Manipulation and Standardization

Transcriptome, microbiome data (at each taxonomic level) and pathway abundance data were manipulated according to the different types of analysis approaches. One of the patient's microbiome data were excluded from the study because patient's transcriptome data was already absent. Remaining 33 samples were used to carry out the data integration analysis. All the results were recorded. Centered Log Transformation (CLR) method was used for the normalization of datasets by using the R programming language (ver. 4.4.0.) (38)(39)(40).

3.5 Data Pre-Processing

The samples that belong to the UC CMS subgroup were excluded from datasets to observe a difference in machine learning model accuracy and correlation coefficient of omics datasets. All the results that are obtained after exclusion of UC samples were also recorded. Manipulation of data was performed by dplyr (ver. 1.1.4.) and tidyverse (ver. 2.0.0.) packages in R (41). Shapiro-Wilk test was performed to observe the data distribution. It is followed by Kruskal-Wallis's test and the non-significant variables were filtered out from the dataset. N-Integration was repeated with pre-processed samples.

3.6 Data Integration Analysis

Among various types of multi-omics data integration Canonical Correlation Analysis and Principal Component Analysis were used as an unsupervised method. Also, Sparse Partial Least Square (sPLS), Block Sparse Partial Least Square Discriminant Analysis (Multi-block sPLS-DA) which is also called as N-Integration method were used as a supervised machine learning approach. Performing of supervised methods were done by using mixOmics (ver. 6.10.) packages in R (42). Visualization of the results were performed by ggplot2 (ver. 3.5.1.) package in R and QIIME2view plugin of QIIME2 analyzing platform.

3.7 Statistical Analysis

R programming language (ver. 3.5.1) was used for all statistical analysis. Distribution of datasets was checked by Shapiro-Wilk test. Kruskal-Wallis and pairwise Wilcoxon tests which are non-parametric alternative of ANOVA was used for determining pairwise comparison of subgroups with Benjamini Hochberg correction based on FDR corrected p value threshold of 0.05. The p-value lower than 0.05 was considered as statistically significant. MixOmics (ver. 6.10.) and caret (ver. 6.0.94) packages were used for machine learning methods (42, 43).

4 RESULTS

Transcriptome, microbiome, and microbial pathway abundance results are explained and association between these omics datasets obtained from data integration analysis is discovered.

4.1 Microbiome and Microbial Pathway Analysis

As a result of microbiome analysis, CMS1 patients were found to have enriched level of *Fusobacteria* and *Bacteroidetes* and decreased level of *Firmicutes* and *Proteobacteria* at phylum level showed in Figure 1, Figure 2, Figure 3 and Figure 4 respectively (p-value < 0.05).

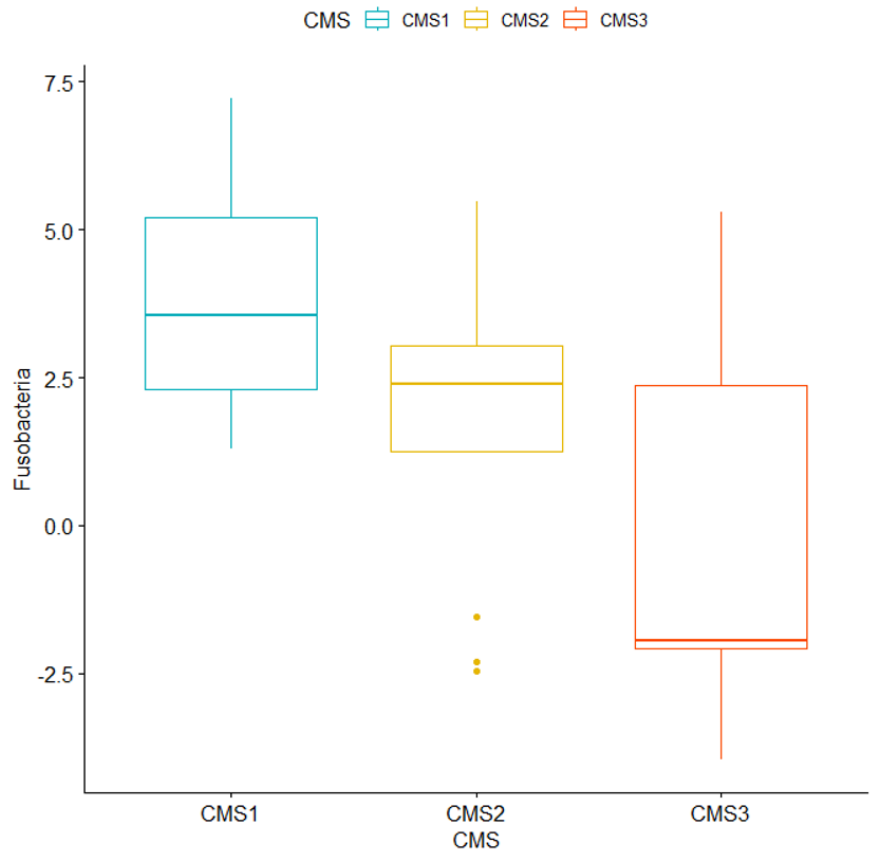

Figure 1. Boxplot shows the relative abundance difference of Fusobacteria between CMS1 (blue), CMS2 (yellow), and CMS3(red) subgroups (p value < 0.05).

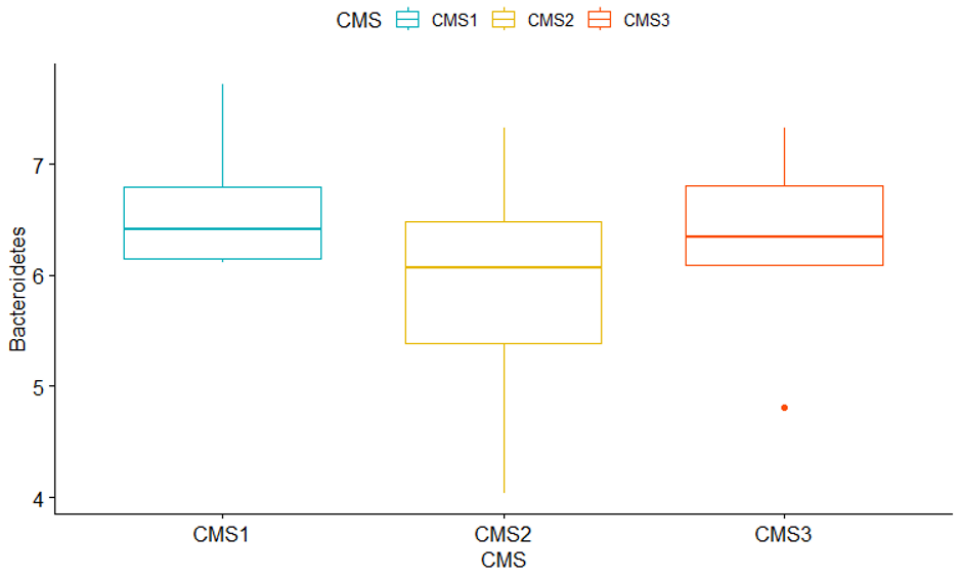

Figure 2. Boxplot shows the relative abundance difference of Bacteroidetes between CMS1 (blue), CMS2 (yellow), and CMS3(red) subgroups (p-value > 0.05).

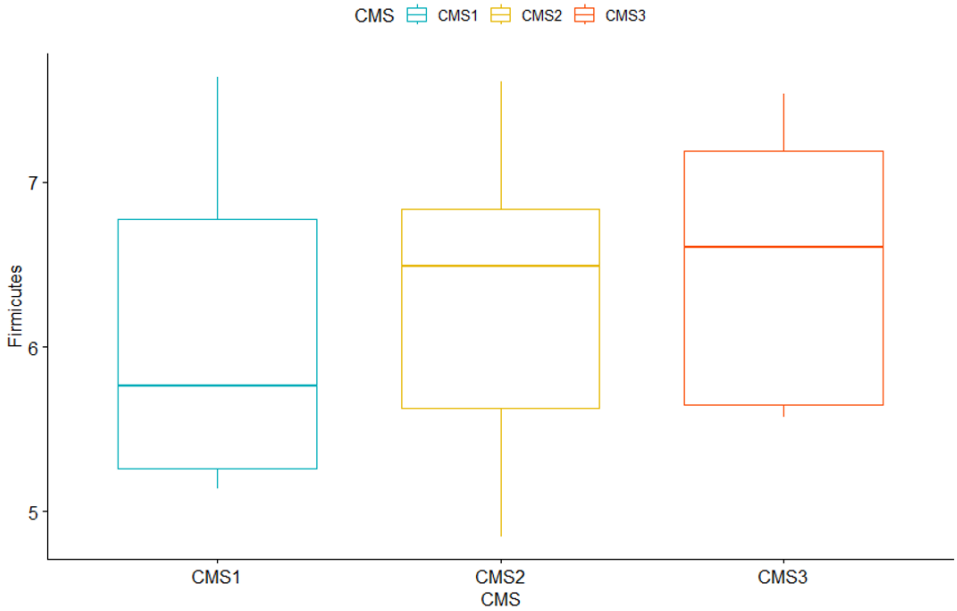

Figure 3. Boxplot shows the relative abundance difference of Firmicutes between CMS1 (blue), CMS2 (yellow), and CMS3(red) subgroups (p -value > 0.05).

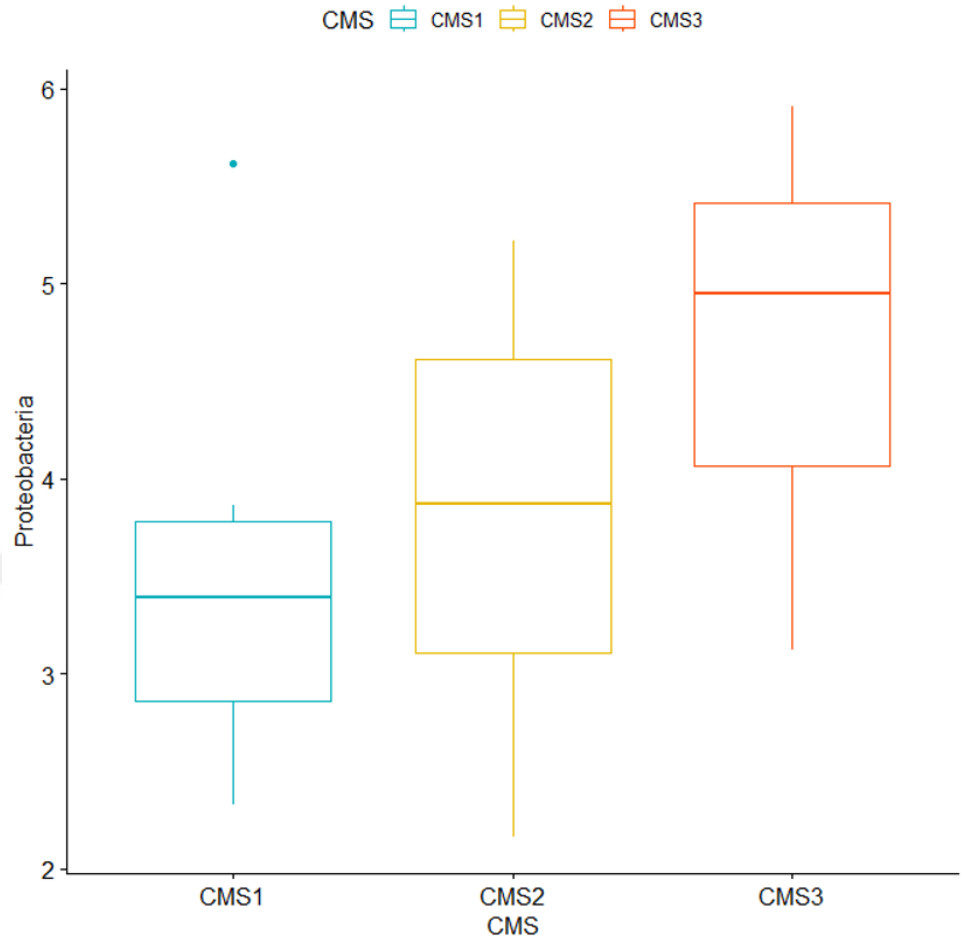

Figure 4. Boxplot shows the relative abundance difference of Proteobacteria between CMS1 (blue), CMS2 (yellow), and CMS3(red) subgroups (p-value < 0.05).

General overview of relative abundance of phylum is indicated in Figure 5. Statistical significance was calculated by Kruskal Wallis test is shown in Figure 6. There was no significant difference between CMS subgroups, the statistical results obtained from pairwise Wilcoxon test were indicated in table 1. Relative abundance of *Bacteroidetes* changes between CMS subgroups. At genus level, *Prevotella* was found significantly increased in CMS1 group (p-value < 0.05). Another significant enrichment was observed in population of *Bacteroides* in CMS2 and CMS3 groups comparing to other CMS clusters at genus level (p-value < 0.05). Average percentage of *Bacteroides* abundance in CMS1 is 48%, 65% in CMS2 and 29% in CMS3. *Fusobacterium* has second highest relative abundance in both CMS1 and CMS2 with 15% and 4% percentage respectively. *Faecalibacterium* and *Clostridium* found to be second highest percentage in relative abundance of CMS3 patients with 5.5%. 478

pathways were found, and non-significant pathways were excluded from microbiome data because of Picrust2 analysis. KEGG pathway database was used in pathway prediction analysis. Microbiome abundance according to their significant univariate scores. The most significant 30 pathways were kept for further analysis.

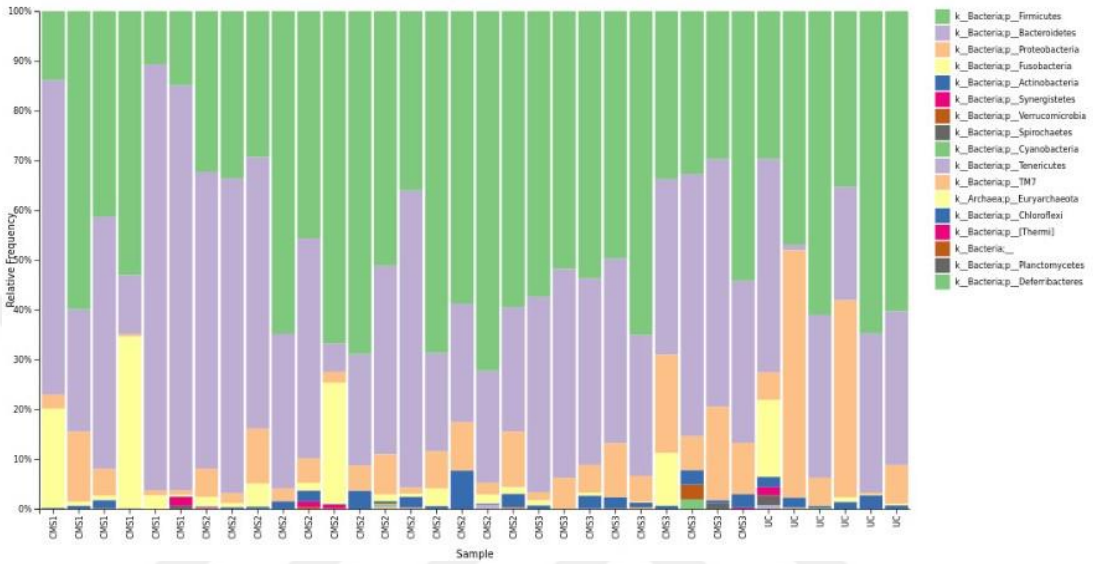

Figure 5. Heatmap indicates the relative abundance of phyla that belong to each CMS subgroup of colorectal cancer patients.

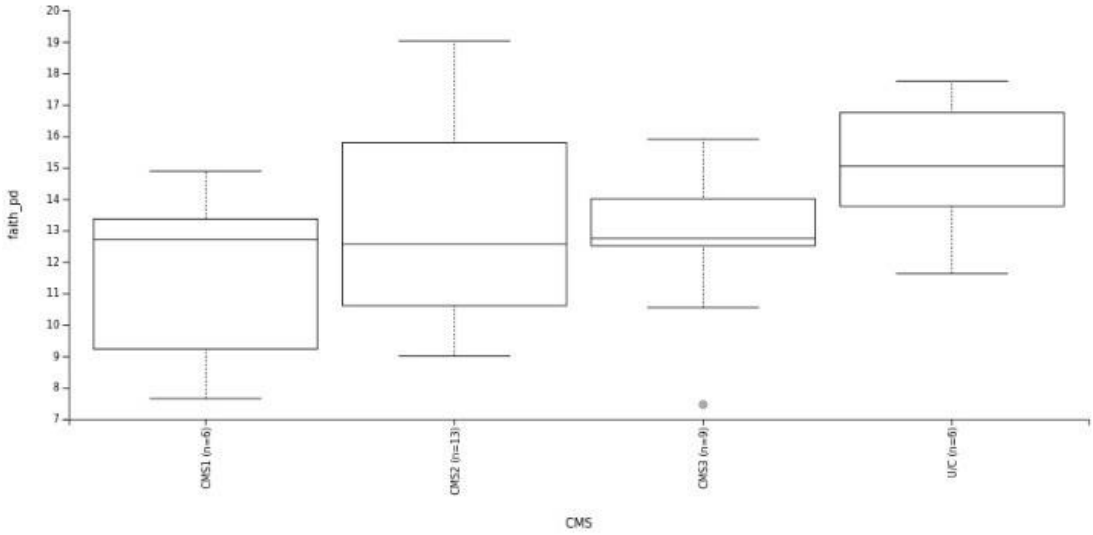

Figure 6. Faith pd interpretation results that indicates the significant differences between CMS subgroups. Kruskal Wallis' test resulted as no significancy was found between groups (p-value > 0.05).

Table 1. Pairwise Wilcoxon test was performed to observe the potential significant differences between subgroups.

Group 1	**Group 2**	**H**	**p-value**	**q-value**
CMS1	CMS2	0.769231	0.380455	0.570683
	CMS3	0.222222	0.637352	0.764822
	UC	3.692308	0.054664	0.231300
CMS2	CMS3	0.054627	0.8152	0.8152
	UC	1.3	0.254213	0.508426
CMS3	UC	3.125	0.0771	0.231300

4.2 RNA-Seq and Pathway Enrichment Analysis

A volcano plot was constructed to indicate up-regulated and down-regulated genes obtained from RNA-Seq analysis. RN7SK, RPPH1, RN7SL1, RN7SL2, ZNF460, SNORD17, SNORA73B, H1-4, H4C12, H3C7 genes were highly upregulated in cancer patients. In contrast SRPX were found to be downregulated in all patients. Volcano plot figure is shown in Figure 7. Principle Component Analysis was also performed for RNA-Seq results which is indicated in Figure 8. Dimension reduction of the dataset was used to obtain the cluster of CMS subgroups in colorectal cancer patients.

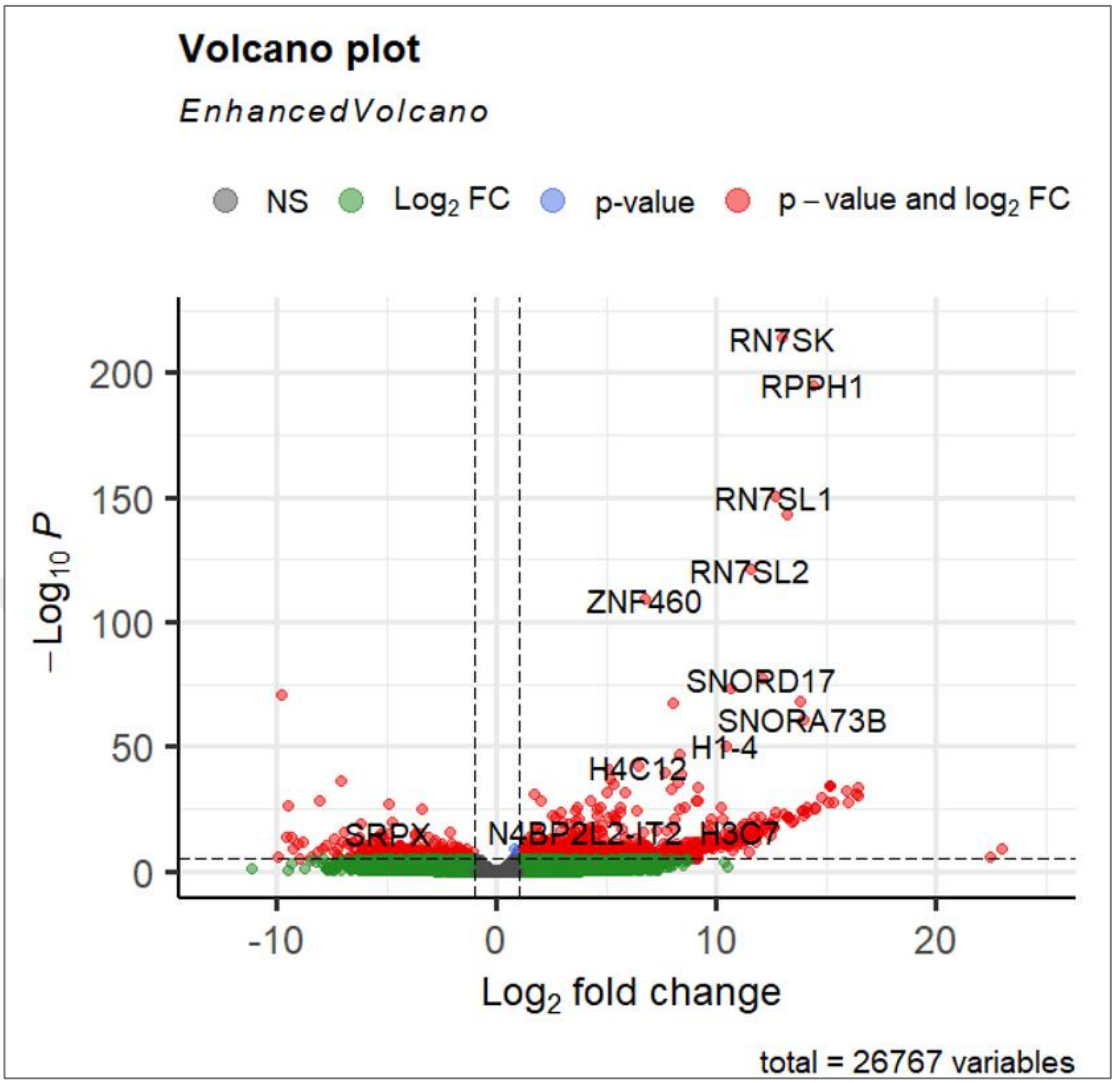

Figure 7. Differentially expressed genes between CMS subgroups are illustrated in volcano plot (p-value < 0.05).

An appropriate clustering wasn't observed in PCA obtained from microbiome data. However, PCA obtained from RNA-Seq data gave a better clustering for each CMS groups comparing to microbiome data. PCA graph is indicated in Figure 8. However, it still does not cluster the subgroups properly so that other dimension reduction approaches were used. The pathway enrichment results were shown in appendix section, Table 2.

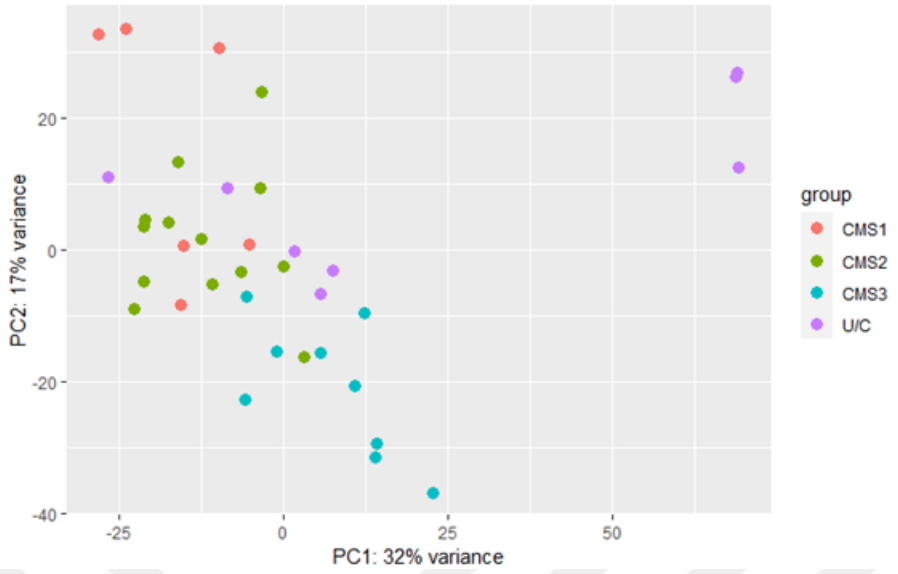

Figure 8. Principal component visualization graph indicates the clustering of CMS subgroups based on RNA-Seq analysis. CMS1 group is labeled with red color, CMS2 group is labeled with green color, CMS3 group is labeled with blue color, and CMS4 group is labeled with purple color.

4.3 Canonical Correlation Analysis

Microbiome and pathway results were firstly used to achieve multi-omics data integration by using Canonical Correlation Analysis. Phylum level abundance table and significant pathway feature tables were used to find an association between them. We obtained several significant correlations that are shown in Figure 8. *Proteobacteria* is positively correlated with Cathecol-ortho-cleavage pathway, Crn-forcat pathway, Glucose1Pmetab pathway, Kdo.naglipasyn pathway, L-lysine pathway, L-threonine pathway, and L-methionine biosynthesis pathway, Protocathechuate.ortho.cleavage pathway, Ubiquinol-7 biosynthesis pathway, Vanilin and Vanillate degradation pathway, Vanilin and Vanillate Degradation II pathway, Heparin degradation pathway, Ubisyn pathway and Biphenyl degradation pathway. In addition to that, Actinobacteria was found to be significantly correlated with Threocat pathway, Meth-acetate pathway and CMP-legionaminate biosynthesis I pathway.

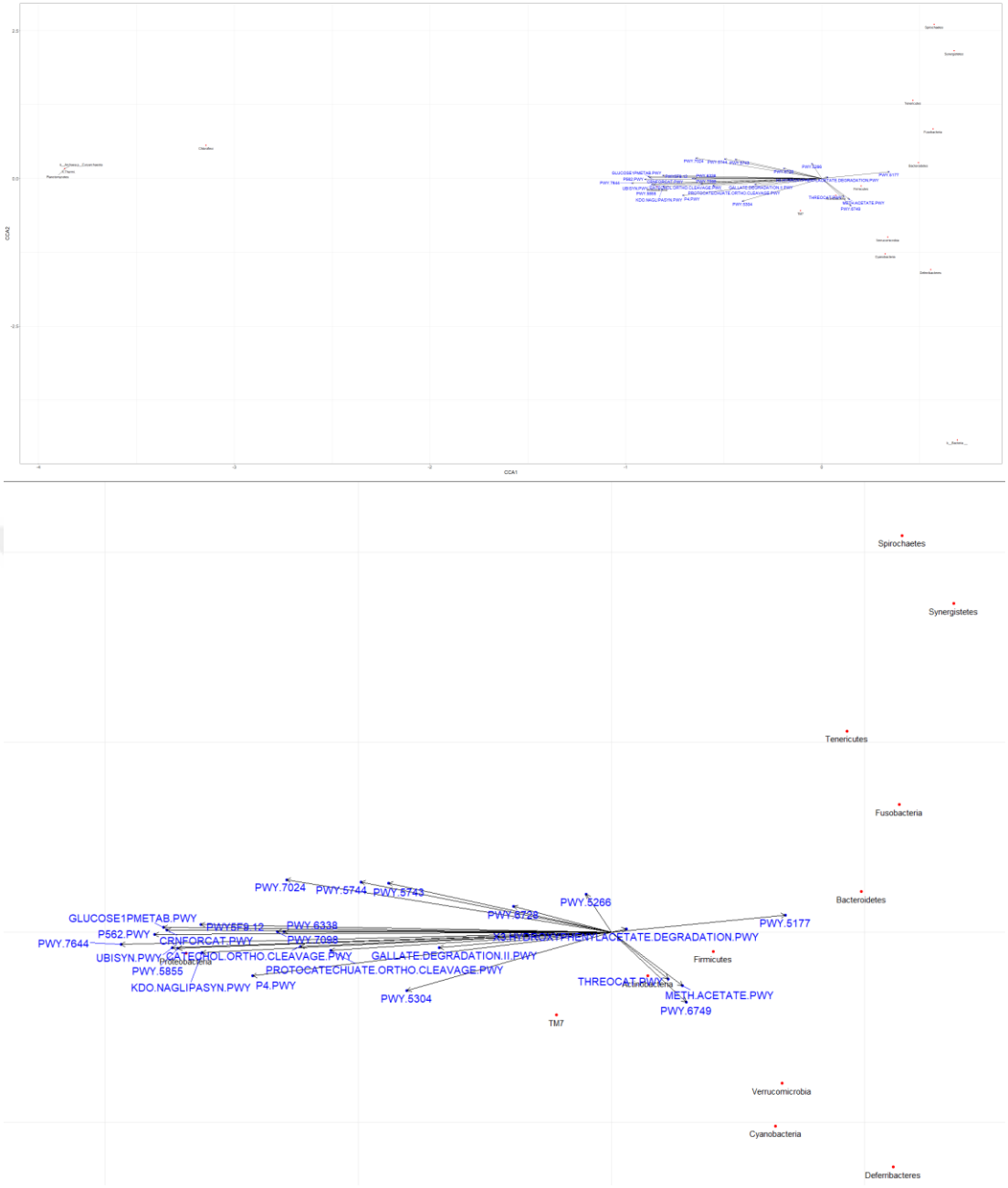

Figure 9. Canonical correlation between phyla and microbial pathway is indicated in figure.

4.4 Multi-Omics Data Integration

Transcriptome, microbiome, and microbial pathway abundance datasets were Integrated by using sPLS and multi-block sPLS methods based on partial least square regression model.

4.4.1 SPLS for transcriptome and microbial pathway datasets

Pathways obtained from microbial data and transcriptome of human tissue were subjected into sPLS to find an association between. Least square regression eliminated the non-significant genes and pathways. Explanatory power of components was lower than 0.8 so that results did not give exact outputs. However, it still gave quite important correlations. Network visualization was applied to sPLS results obtained from sPLS analysis. EIF3B gene found to be positively associated with both Superpathway of Sulfur Oxidation pathway and METH-ACETATE pathway with Pearson correlation coefficient (r) value of 0.60. Furthermore, TPX2, MCM7, AHCY, SSRP1 and ANAPC1 genes are positively correlated with METH.ACETATE pathway (r = 0.61, r = 0.61, r = 0.63, r = 0.6, r = 0.61 respectively). REPIN1 is positively associated with Superpathway of Sulfur Oxidation pathway. In contrast, negative association was found between DDX60, B2M, AFF4, HLALE genes and Superpathway of Sulfur Oxidation pathway. This network visualization is shown in Figure 10.

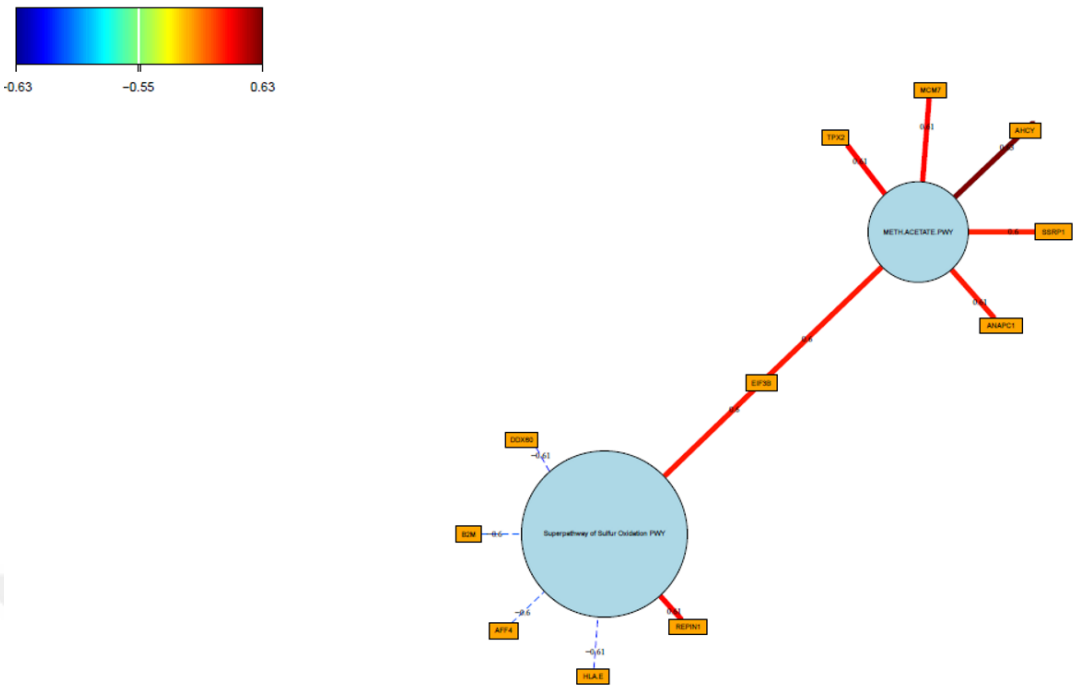

Figure 10. Network analysis that indicates the correlation between transcriptome and microbial pathway datasets (r = ± 0.63).

4.4.2 SPLS for microbiome and transcriptome datasets

As a result of sPLS regression method which is applied for microbiome (at phylum level) and transcriptome datasets, quite valuable associations were observed. Excluding the *Spirochaetes* and *Verrucomicrobia*, heatmap directly discriminate patients into two groups. Differentiated first group includes KLF3, CLDN7, PLEKHG3, MALL, RNA5-8N2, SNORD10, SYTL2, DNM2, CXADR, TFF3, SERPINA1, FAM120A, RTN4, MIB1, OIP5-AS1, S100A6, H1-5 gene sets and *Actinobacteria*, *Cyanobacteria*, *Tenericutes*, *TM7*, *Bacteroidetes*, *Firmicutes* and *Proteobacteria* phylum. Second group contains ADAR, CSNK2A1, RPL37, AHCTF1, ODC1, NUP98, CCDC14, ZNF292, CCDN2, IGKC, PRDX5, GPSM2, GPX2, PAN3, TKT, RPL22, AHCY, NME2, RBM12, CCT6A, RPLP2 and SLC7A1 genes and *Synergistetes*, *Fusobacteria* phyla. Heatmap results show the association of genus level microorganisms and genes obtained from transcriptome data. Spirochaetes genus type is found to be positively related with the OIP5-AS1, S100A6, H1-5 and ADAR genes (r = 0.71).

Moreover, *Fusobacteria* is discovered to be positively correlated with the KLF3, CLDN7, SYTL2, DNM2 and TFF3 genes (r = 0.70). Correlation coefficient of variables (both genus and microbiome) lower than 0.5 were not taken into consideration. *Spirochaetes* found to be negatively correlated with TFF3, SERPINA1 and PROX5 (r = -0.71). ODC1, NUP98, CCDC14, AHCY, CCT6A genes are negatively correlated with the *Fusobacteria* (r = -0.69). RPL22, AHCY, GPSM2 genes are found negatively associated with *Synergistetes*. All results were indicated in Figure 11.

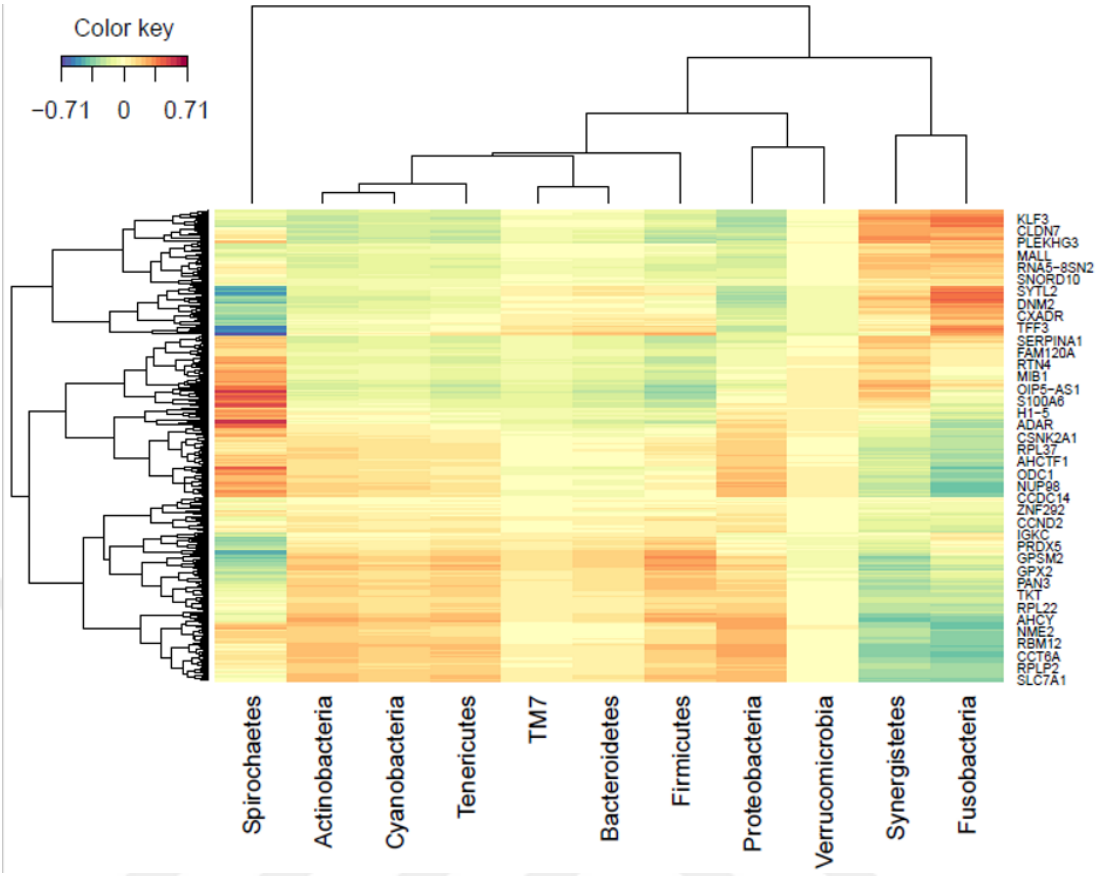

Figure 11. Heatmap illustration indicates the correlation between microbiome (phylum) and transcriptome datasets (r = ±0.71).

4.4.3 Multiblock sPLS-DA for transcriptome, microbiome and microbial pathway datasets

As a conclusion of multiblock sPLS-DA model, statistically significant loadings that are used for explaining the data based on consensus molecular subtypes are determined for each omics dataset. TPX2, CBX3, HSDP1, TOP2A, LBR and CSE1L are found to be most explanatory variables of transcriptome data indicated in Figure 12.

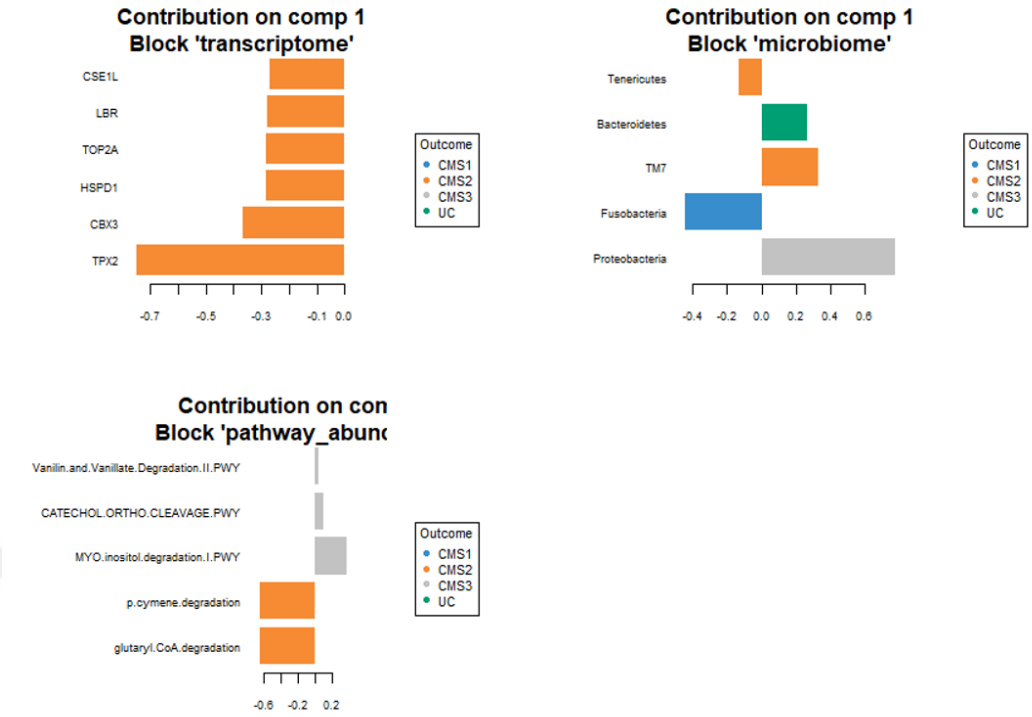

Figure 12. Transcriptome, microbiome, and microbial pathway significant features that construct the component 1.

Each dataset is subjected to PLS regression and components are constructed based on the significant variables. Component design weights for transcriptome, microbiome and pathway are 0.72, 0.62 and 0.58 respectively. The model worked with the 88% accuracy shown in Figure 13. It predicted two CMS1 and two CMS2 patients as unclassified. Model made no mistake for prediction of CMS3 and UC patients.

	predicted.as.CMS1	predicted.as.CMS2	predicted.as.CMS3	predicted.as.UC
CMS1	4	0	0	2
CMS2	0	11	0	2
CMS3	0	0	9	0
UC	0	0	0	5

Figure 13. Accuracy of the model that tries to predict CMS subgroups based on features obtained from transcriptome, microbiome, and microbial pathway datasets.

PCA analysis for each block resulted in slightly better clustering but because of the noisy variables in datasets, it still did not indicate a significant discrimination of CMS subgroups. PCA analysis result was shown in Figure 14.

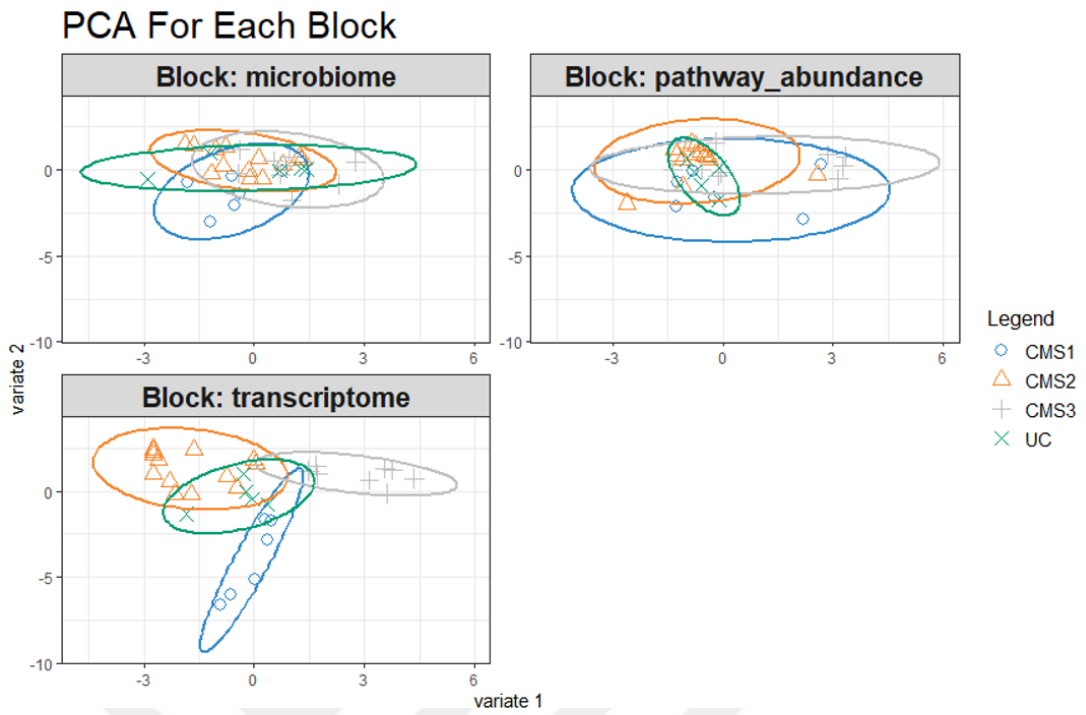

Figure 14. Clustering of CMS subgroups by using the principal component analysis.

4.4.4 Multiblock sPLS-DA for pre-processed transcriptome, microbiome and microbial pathway datasets

Unclassified samples removed from the data and feature selection based on statistical significance was applied for each dataset. Non-significant variables were removed, and data integration was performed with filtered features. Component design weights for transcriptome, microbiome and pathway are 0.80, 0.73 and 0.60 respectively. Model constructed based on CMS subgroup information worked perfectly by predicting each subgroup correctly as shown in Figure 15.

	predicted.as.CMS1	predicted.as.CMS2	predicted.as.CMS3
CMS1	6	0	0
CMS2	0	13	0
CMS3	0	0	9

Figure 15. Accuracy of the model that tries to predict CMS subgroups based on features obtained from pre-processed transcriptome, microbiome, and microbial pathway datasets.

The principal component analysis provided the most distinct clustering, especially for transcriptome block shown in Figure 16. Most discriminative variables were obtained from transcriptome data. PCA result obtained from all these features indicated a perfect clustering of CMS groups in transcriptome block of model.

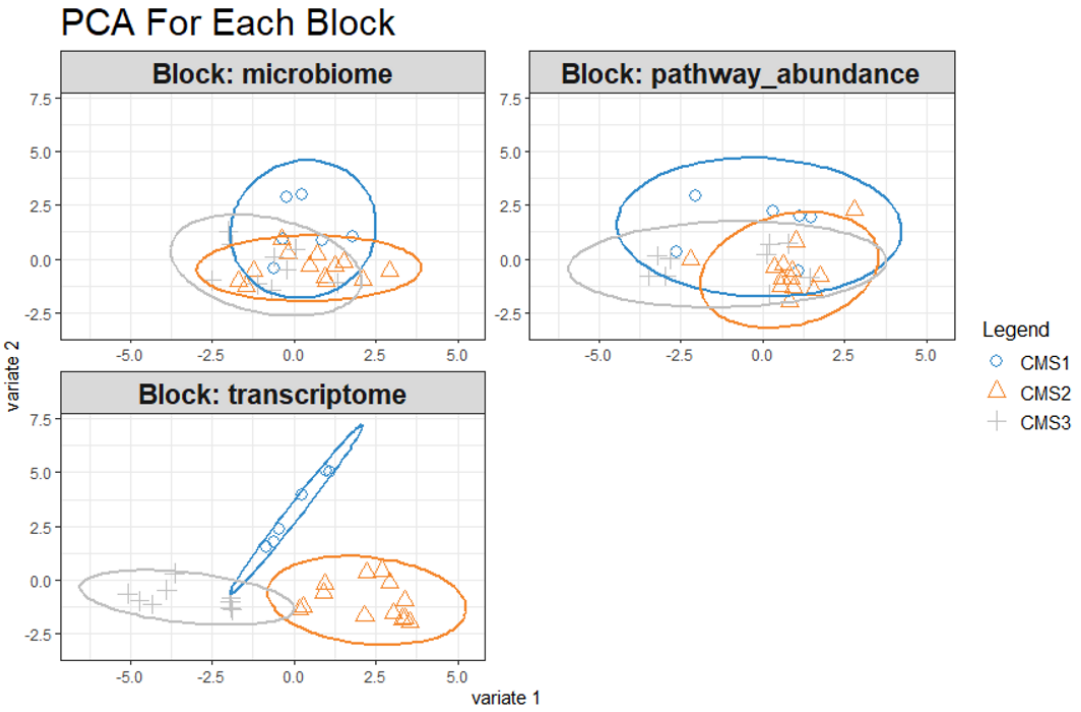

Figure 16. Clustering of CMS subgroups by using the principal component analysis.

Features contribute to model from transcriptome data are HSPD1, CBX3, CSE1L, LBR genes TPX2, from microbiome data *Actinobacteria*, *Spirochaetes*, *Fusobacteria*, *Bacteroidetes* and *Proteobacteria* phyla, from pathway data Cholorthocleavage pathway, crn-forcat pathway, myo-inositol degradation pathway, p-cymene degradation pathway and glutaryl-CoA degradation pathway. All these features were shown in Figure 17.

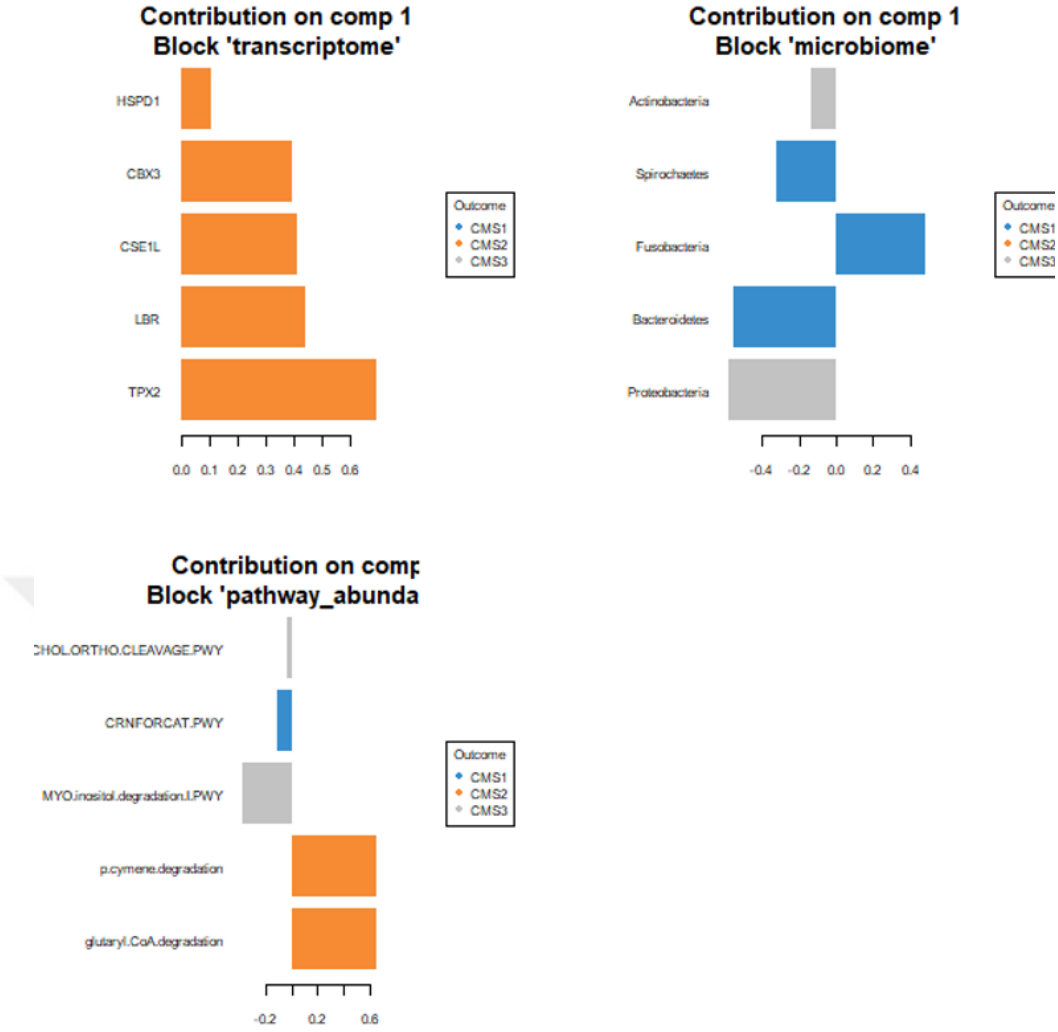

Figure 17. Pre-processed transcriptome, microbiome, and microbial pathway datasets significant features that construct the component 1.

Correlation results between transcriptome and pathway datasets indicated remarkable associations. HSPD1, CBX3, CSE1L, LBR and TPX2 genes were discovered to be positively correlated with p-cymene degradation and glutaryl CoA degradation pathways indicated in Figure 18 (r = 0.95).

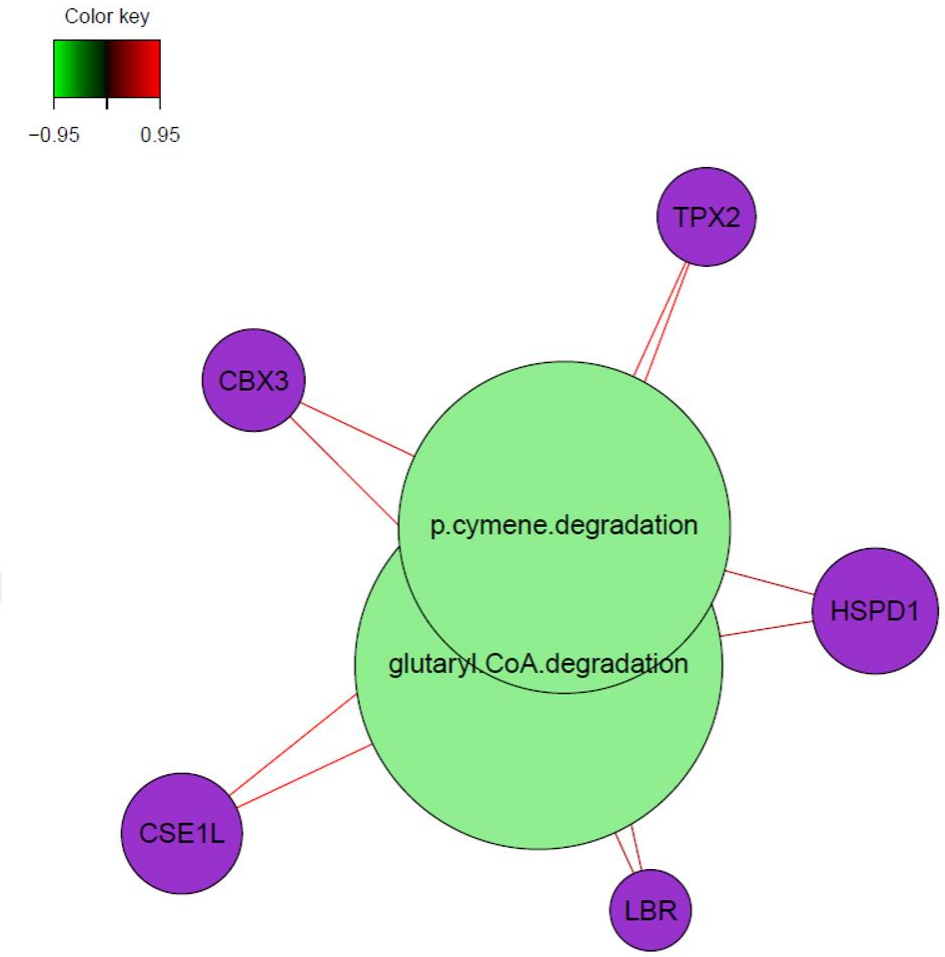

Figure 18. Network analysis applied to pre-processed dataset that indicates the correlation between transcriptome and microbial pathway datasets (r = ± 0.95).

Moreover, *Proteobacteria* showed negative correlation with HSPD1, CBX3, CSE1L and TPX2 shown in Figure 19 (r = -0.73). The first target in multiomics data integration is to concentrate on the positive correlation to observe if the features that are obtained from different datasets work together or not. However, negative correlation should not be neglected, especially in presence of microbial datasets. Because the bacteria living inside our gut are in a competition to occupy a place. The factors that inversely affect the bacteria surviving are remarkable output in microbiome research are.

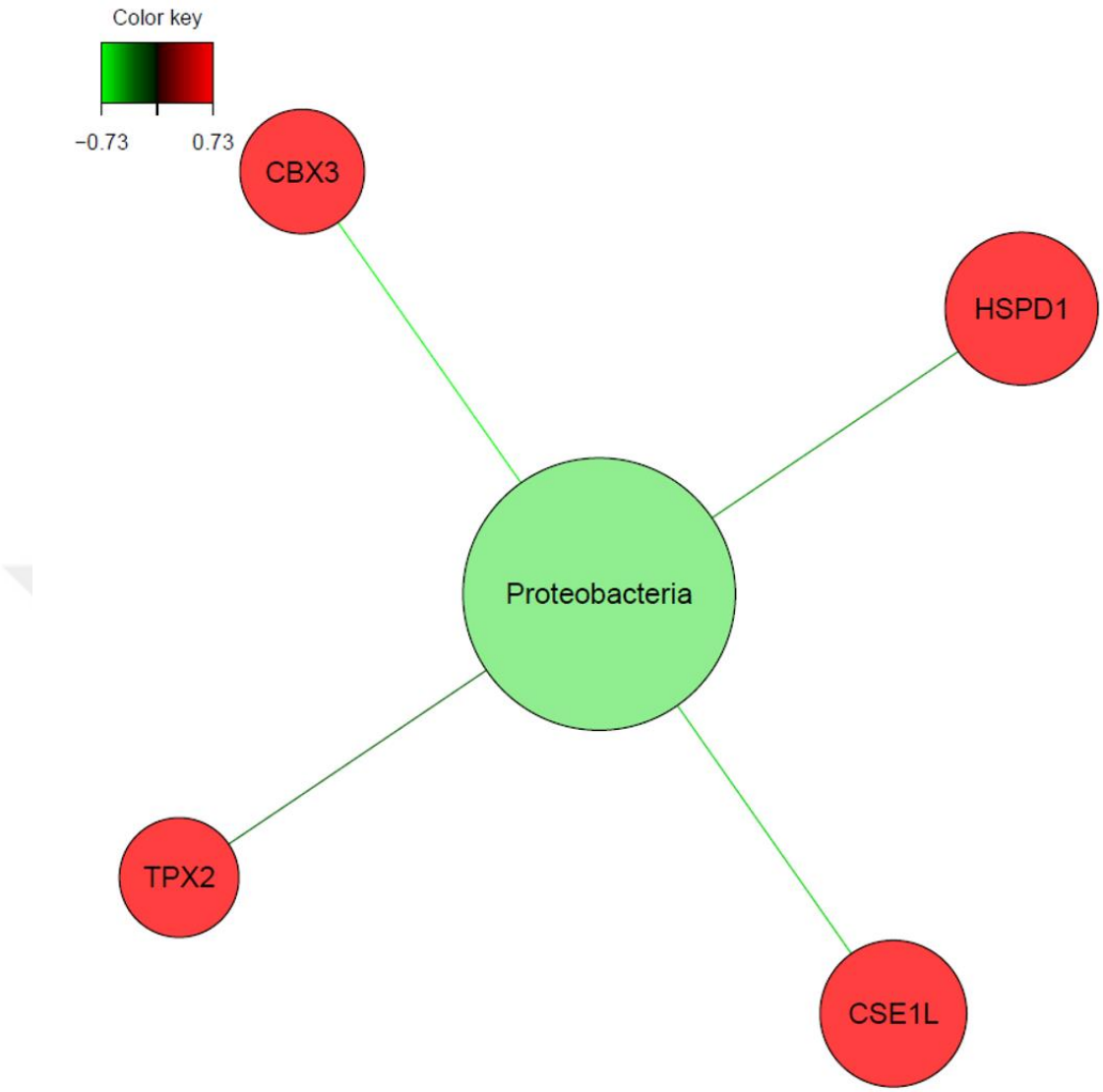

Figure 19. Network analysis applied to pre-processed dataset that indicates the correlation between transcriptome and microbiome (phylum) datasets (r = ± 0.73).

The comprehensive neural network illustration that includes other significant features provided broad perspective outputs (correlation cut off = 0.7). *Proteobacteria* found to be negatively correlated with p-cymene degradation, glutaryl CoA degradation and meth-acetate pathways. It also negatively correlates with HSPD1, CBX3, CSE1L, TPX2 and OAS3. In contrast, SEMA5A and CDH17 genes are found in positive association with meth-acetate pathway indicated in Figure 20. Overall correlation between each feature that are obtained from transcriptome, microbiome and microbial pathway datasets helps us to understand the potential associations clearly. Despite the fact that correlation between the

variables is quite high and the model that select the variables with high accuracy, the results still need further evaluation. In our case, we checked the literature and databases to interpret the results, but experimental validation is essential.

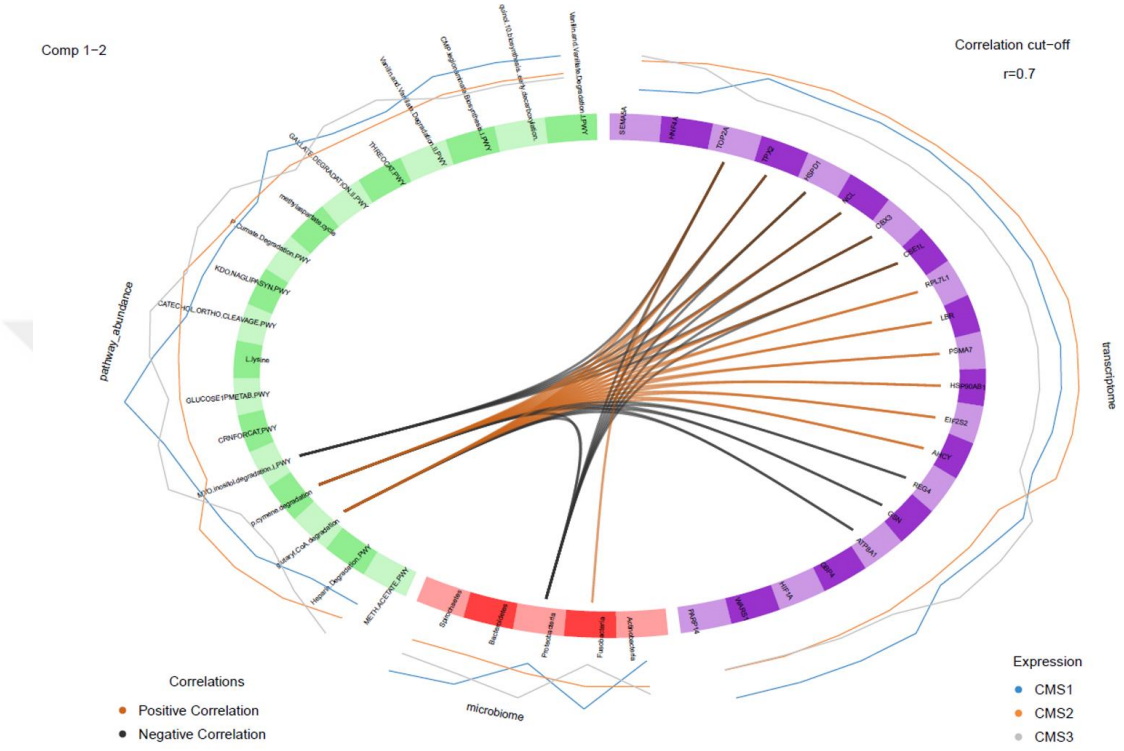

Figure 20. Network analysis applied to pre-processed datasets that indicates the correlation between transcriptome, microbiome, and microbial pathway (phylum) data (r = ± 0.7). Orange lines indicate the positive correlation, black lines indicate the negative correlation.

5 DISCUSSION

All in all, plenty of positive and negative correlation were obtained as a result of multiple omics datasets integration. The results were interpreted by comparing them with previous research outputs and with information obtained from Metacyc, and Genecards databases (44, 45).

Integration of large datasets have always been a trouble for data scientists who are working for finding the golden standard to achieve integration. Nevertheless, different types, contents and structure of datasets arouse difficulty in choosing the exact method. In this thesis we aimed to overcome this problem by using the transcriptome, microbiome, and pathway datasets. In microbiome analysis, same results were obtained with the previous study because I used the same analyzing platform QIIME2. Pathway abundance was predicted by using Picrust2 which is a plugin tool of QIIME2. Transcriptome analyses were performed by using pipeline generated by Epigenetiks company.

Pathway enrichment analysis was done, and human pathways associated with up-regulated and down-regulated genes were recorded. For the integration of datasets, microbiome OTU data (at each level), RNA-seq count data and microbial pathway data was preferred. Association of genes with phyla, genus and microbial pathways were checked with human pathway enrichment results to see if they work in a collaboration or work against each other. Metacyc and Genecard databases were also used to get functional information about correlated genes and pathways. Literature was checked for further information when significant functional hint is obtained.

5.1 Unsupervised Learning Approach for Multi-Omics Data Integration

Unsupervised methods are widely used in omics studies, especially for comparison targeted research. PCA, PCoA and CCA are the most preferred dimension reduction methods that are working based on the distances between or

within sample groups. PCA and CCA was used in our study for two reasons. First, PCA provided how the CMS subgroups differentiated from each other based on microbiome and transcriptome analysis. The variables obtained from this analysis were not informative enough for the perfect clustering. This information indicates the noisiness of the data which needs to be filtered out properly. After data pre-processing step, we obtained almost perfect clustering for transcriptome data. We continued the analysis by using those variables and obtained 100% accuracy in our supervised model at the end. Second one is to observe if CCA that maximize the covariates between datasets work properly or not. The results obtained from CCA and multi-block sPLS method are quite different. Maximizing power of CCA method might cause false positive correlation that's why more reliable methods should be preferred for discovery of associations between large datasets.

5.2 SPLS for Multi-Omics Data Integration

Partial Least Square was chosen out of plenty of approaches because it works with several omics' integration studies, and it is sustainable enough for development. The results obtained from sPLS gave a comprehensive perspective. The regression method tries to keep as many variables as possible to maximize the correlation between datasets. That is the reason why explanatory values of components obtained from this method might be low. Nonetheless, the outputs came by correlation analysis which is applied after regression are pretty acceptable. CLDN7 gene found to be associated with *Fusobacteria*. CLDN7 is a member of claudin family which is directly involved in structure of tight junction. Low expression of this gene is strongly associated with cancer (46). *Fusobacteria* species are commonly observed in colorectal cancer and one of its functions is disrupting the structure of intestinal barrier. In a study, researchers proposed that *Fusobacterium nucleus* damage oral mucosal epithelial barrier by inhibiting the CLDN4 which is also a member of claudin (47, 48).

Our findings suggest that this strong correlation between Fusobacteria and CLDN7 may have similar role that results in tight junction collapse. Moreover, KLF3 was also found to be positively correlated with Fusobacteria. Expression of

KLF3 is playing crucial role in activation of WNT1 and WNT/β-catein pathways which involve in colorectal cancer progression (49). In a study it was discovered that *Fusobacterium nucleatum* directly take part in Cdk5 activated Wnt/β-catein pathway and contribute progression of Colorectal cancer (50). In RNA-Seq enrichment analysis we obtained downregulation of WNT2, WNT3, NOTUM, LRP5, LRP6, CSNK2A2, GSK3B, CHD8, CREBBP, EP300, TBL1XR1, INVS, PLCB3 genes and upregulation of WNT2B, WNT5B, SERPINF1, DKK1, SFRP1, SFRP2, SFRP4, RSPO2, RSPO3, FZD1, FZD4, FRAT1, TCF7L1, CBY1, SOX17, CTNND2, CCND3, PRKACA, SKP1, RBX1, ROR1, PRICKLE1, DAAM2, RHOA, RAC2, MAPK10, PPP3CA, PPP3CB genes that are found to be involved in wnt signaling pathway (Appendix 1). It means that if we did not use the integration method, we were going to miss the information of KLF3 gene and *Fusobacteria* that work in collaboration. Same can be made for CLDN7 and *Fusobacteria* association. It doesn't give the pathway information to find out how they interact together in disruption of tight junction, even so it gave a strong clue for further analysis and experiments. These kinds of conclusions provide a specific pathway dependent diagnosis, and it is necessary to understand the complex diseases like cancer.

These results indicates that sPLS method can be used for getting an extensive point of view even though low explanatory variables that constructs components. It is especially more useful if we have small sample size because most models fail to differentiate variables in datasets that includes low number of samples. Partial Least Square provides a better potential association comparing to unsupervised learning methods (such as PCA, PCoA, CCA etc.) that are used for clustering.

5.3 Impact of the Pre-Processing Datasets

It is an obvious fact that, even potential outcomes are received by sPLS analysis, it can only be used for comparison, not favorable for discovery of a biomarker and for understanding the story of pathways that interact each other. So that, N-Integration method (from mixOmics) was checked. In this method, regression model eliminates most non-significant variables to discriminate the variable as much as

possible. In consequence of results of the model, explanation power of components was still low. The model predicted the outcomes with 87% accuracy, by using the component that has the highest value of explanation. The datasets were not subjected to this analysis. Main concentration was given to the model improvement.

Dimensionality of huge datasets causes a confusion in machine learning models because of the noisiness. Therefore, it is very important to find optimum pre-process approaches for each type of data. In our study, we firstly excluded the unclassified samples from the data. In the study which datasets were taken, researchers couldn't manage to classify those 5 samples. It is most likely because of the complexity of variables that are not distinct enough for classification so that it would be better to exclude them from the study for finding biomarkers with increasing accuracy level of model. Other approach was usage of feature selection algorithm to dispose of non-significant variables from the data. Thanks to the filtering out of unclassified patients, sample size decreased to 28 and because of the feature selection microbiome phyla variables were down to 5, transcriptome data variables were down to 851 and pathway variables were down to 26. The N-Integration process was performed by using these data only.

As stated above, N-Integration was used with clean data and filtering out non-significant variables and samples significantly increased the explanatory value of components. The component explanatory value close to 0.8 and higher than 0.8 are indicating the model work very well, however it is not achieved in many papers because of dimensionality of datasets. Datasets that are used in mixOmics tutorial includes miRNA, mRNA and proteome data that includes 150 samples and noisiness of the data they have does not affect the results. Integration of omics studies that have a correlation value of $r = \pm 0.73$, $r = \pm 0.64$, $r = \pm 0.5$ etc. are considered as significant outputs (51)(52)(53). In our study, we increased this value up to $r = \pm 0.95$ and by developing the model with pre-processing.

5.4 Multi-block sPLS for Multi-Omics Data Integration for Pre-Processed Datasets

The last integration model was constructed by HSPD1, CBX3, CSEL1, LBR and TPX2 genes from transcriptome data, *Actinobacteria*, *Spirochaetes*, *Fusobacteria*, *Bacteroidetes* and *Proteobacteria* phyla from microbiome data and cholorthocleavage pathway, crn-forcat pathway, MYO inositol degradation pathway, p-cymene degradation pathway and glutaryl.CoA degradation pathway from microbial pathway abundance data. Component design weights for transcriptome, microbiome and pathway are increased to 0.80, 0.73 and 0.60 respectively. The model predicted the three CMS groups with 100% accuracy. According to the correlation results, HSPD1, CBX3, CSEL1, LBR and TPX2 genes were discovered to be highly associated with p-cymene degradation and glutaryl-CoA degradation pathways. HSPD1 gene is a heat shock protein, related with transcription and peroxisomal lipid metabolism. It was recently discovered that activity of HSPD1 gene is necessary for metabolic fitness and loss of function in this gene cause energetic breakdown that directly affect the cancer cells dividing and expanding ability in a study published in 2021. [54] CBX3 is taking role in transcription, RNA polymerase and promoter opening pathways. Also, studies indicate that this gene is overexpressed in variety of cancers but still a poor prognosis biomarker in cancer patients (55)(56). CSEL1 gene is chromosome segregation 1 like gene that play role in p53 effector and Golgi to endoplasmic reticulum traffic pathways. Researchers discovered that CSEL1 directly promotes the development of tumor in various malignancies (57). LBR gene is Lamin B receptor that takes role in super pathway of cholesterol biosynthesis and transcription pathways. Its role in cancer was revealed as LBR protects the genome from tumorigenesis and the chromosomal instability. [58] TPX2 gene is a microtubule nucleation factor that relates with microtubule organization pathways. It is commonly found as upregulated in CRC patients (59).

P-cymene degradation pathway has an antitumor effect on colorectal cancer patients. Especially in cancer related to high fat diet. The mechanism behind this is thought as reduction of inflammatory factors like IL-1 and LEP, whereas

upregulation of IL-6 that promotes growth of *Bifidobacteria* and *Clostridium* IV (60). Glutaryl-CoA degradation pathway is essential for the metabolism of lysine, tryptophan and hydroxylysine. Research proved that inhibition of this pathway induces NRF2 glutarylation and increases the stability. This is followed by the upregulation of ATF4-ATF3 signaling and cancer cell death (61).

Interpretation of results will be made from pathway point of view. We discovered the p-cymene degradation pathway associated with four oncogenes. P-cymene increase in tumor tissue indicates that this might be one of the human immune mechanism of microbes that live in our body commensally. In immune mechanism the bacteria that are responsible for producing p-cymene might take role to fight against cancer cells. Further research is needed to specifically understand the role of p-cymene.

HSPD1, CBX3, CSEL1, NCL, TOP2A and TPX2 genes and *Proteobacteria* are negatively correlated with each other. *Proteobacteria* found to be negatively correlated with p-cymene degradation and glutaryl CoA degradation. P-cymene promotes the growth of genus *Bifidobacterium* and genus *Clostridium* (they belong to *Actinomycetota* and *Firmicutes* respectively) (60). Negative correlation between p-cymene degradation pathway and *Proteobacteria* is pretty meaningful. However, it is not easy to command on this so that this association still needs further analysis for the identification of positively correlated genus and pathways.

6 CONCLUSION

Development in technology provides many opportunistic advantages to reach necessary information. Evolution of sequencing devices caused a remarkable increase in number of studies about drug industry, medical biotechnology, diagnosing, treatment approaches, molecular biology etc. This increase revealed the complex disease mechanism, gene functions, pathway interactions, drug responses, microbial associations. Golden standard in analysis of these complex systems has not been reached. The statistical and machine learning approaches continuously change from one study to another for understanding all these mechanisms. Integration of multi-omics datasets is one of the most popular approaches to overcome this problem.

Multi-omics data integration presents a remarkable way to investigate the datasets from different perspective and provides astonishing associations between omics data. This enables researchers to understand the pathways associations, microbial interactions, and gene's role in diseases. It is thought to be the best option to understand the disease exact mechanism of action especially complex diseases like cancer. Personalized medicine will take an important part in our life with the improvement of this multi-omics data integration that will ensure detailed diagnosis and treatment approaches.

Enhancement of biotechnology, machine learning techniques and databases will lead the way for unraveling the mystery behind complex mechanism. Although unsupervised and supervised learning methods have limited power of explaining the complex datasets, deep learning approaches implementation in biotechnology with upgraded databases will overcome these limitations.

7 REFERENCES

1. Gupta, V., Sengupta, M., Prakash, J., Tripathy, B.C. (2017). An Introduction to Biotechnology. In: Basic and Applied Aspects of Biotechnology. Springer, Singapore.
2. Hogeweg P. The roots of bioinformatics in theoretical biology. PLoS Comput Biol. 2011;7(3):1–5.
3. Ouzounis CA, Valencia A. Early bioinformatics: The birth of a discipline - A personal view. Bioinformatics. 2003;19(17):2176–90.
4. Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed. 2013;98(6):236–8.
5. Asrat Z. Review on Mechanism of DNA Sequencing Technologies in Crop Improvement. Int J Res Stud Agric Sci. 2021;7(8):27–36.
6. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol [Internet]. 2021;82(11):801–11.
7. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. Biology (Basel). 2023;12(7)
8. Qin D. Next-generation sequencing and its clinical application. Cancer Biol Med. 2019;16(1):4–10.
9. Watts GS, Hurwitz BL. Metagenomic Next-Generation Sequencing in Clinical Microbiology. Clin Microbiol Newsl [Internet]. 2020;42(7):53–9.
10. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet [Internet]. 2019;20(6):341–55.
11. Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform. 2021;22(1):178–93.
12. Caicedo HH, Hashimoto DA, Caicedo JC, Pentland A, Pisano GP. Overcoming barriers to early disease intervention. Nat Biotechnol. 2020;38(6):669–73.
13. Narayan NR, Weinmaier T, Laserna-Mendieta EJ, Claesson MJ, Shanahan F, Dabbagh K, et al. Erratum: Piphillin predicts metagenomic composition and dynamics from DADA2- corrected 16S rDNA sequences (BMC Genomics (2020) 21 (56) DOI: 10.1186/s12864-019-6427-1). BMC Genomics. 2020;21(1):1–12.
14. Mitchell K, Ronas J, Dao C, Freise AC, Mangul S, Shapiro C, et al. PUMAA: A Platform for Accessible Microbiome Analysis in the Undergraduate Classroom. Front Microbiol. 2020;11(October):1–15.
15. Nagpal S, Haque MM, Singh R, Mande SS. IVikodak-A platform and standard workflow for inferring, analyzing, comparing, and visualizing the functional potential of microbial communities. Front Microbiol. 2019;10(JAN):1–15.
16. McNally CP, Eng A, Noecker C, Gagne-Maynard WC, Borenstein E. BURRITO: An interactive multi-omic tool for visualizing taxa-function relationships in microbiome data. Front Microbiol. 2018;9(MAR):1–11.

17. Ternes D, Karta J, Tsenkova M, Wilmes P, Haan S, Letellier E. Microbiome in Colorectal Cancer: How to Get from Meta-omics to Mechanism? Trends Microbiol. 2020;28(5):401–23.
18. Singh RP, Bashir H, Kumar R. Emerging role of microbiota in immunomodulation and cancer immunotherapy. Semin Cancer Biol. 2021;70:37–52
19. Purcell R V., Visnovska M, Biggs PJ, Schmeier S, Frizelle FA. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. Sci Rep [Internet]. 2017;7(1):1–12.
20. De Steenhuijsen Piters WAA, Heinonen S, Hasrat R, Bunsow E, Smith B, Suarez-Arrabal MC, et al. Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. Am J Respir Crit Care Med. 2016;194(9):1104–15.
21. Cao KAL, Costello ME, Lakis VA, Bartolo F, Chua XY, Brazeilles R, et al. MixMC: A multivariate statistical framework to gain insight into microbial communities. PLoS One. 2016;11(8).
22. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. Biotechnol Adv [Internet]. 2021;49(March):107739.
23. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50(D1):D20–6.
24. OpenTreeofLife, Karen A. Cranston, Benjamin Redelings, Luna Luisa Sanchez Reyes, Jim Allman, Emily Jane McTavish, & Mark T. Holder. (2019). Open Tree of Life Taxonomy (3.2).
25. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013;41(D1):590–6.
26. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42(D1):633–42.
27. Yi Y, Fang Y, Wu K, Liu Y, Zhang W. Comprehensive gene and pathway analysis of cervical cancer progression. Oncol Lett. 2020;19(4):3316–32.
28. Carolina N, Hill C. Eulogy for Junk DNA. Genomics. 2012;9(37):1159–61.
29. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: The human metabolome database. Nucleic Acids Res. 2007;35(SUPPL. 1):521–6
30. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, et al. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158–69.
31. Kanz C, Aldebert P, Althorpe N, Baker W, Baldwin A, Bates K, et al. The EMBL nucleotide sequence database. Nucleic Acids Res. 2005;33(DATABASE ISS.):21–6.
32. Misra BB, Langefeld C, Olivier M, Cox LA. Integrated omics: Tools, advances and future approaches. J Mol Endocrinol. 2019;62(1):R21–45.
33. Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. Neuroimage [Internet]. 2011;56(2):455–75.
34. Chun H, Kele¸s SK. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. J R Stat Soc B [Internet]. 2009;(72):3–25.

35. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.
36. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139–40.
37. Ulgen E, Ozisik O, Sezerman OU. PathfindR: An R package for comprehensive identification of enriched pathways in omics data through active subnetworks. Front Genet. 2019;10(SEP):1–33.
38. Bates, D., Maechler, M., Bolker, B. and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.
39. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
40. Wickham, H. (2009) ggplot2: elegant graphics for data analysis. Springer New York.
41. Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4,
42. F R, B G, A S, K LC. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol [Internet]. 2017;13(11):1–14.
43. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5):1–26.
44. Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, et al. MetaCyc: A multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. 2004;32(DATABASE ISS.):438–42.
45. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support. Bioinformatics. 1998;14(8):656–64.
46. Wang K, Xu C, Li W, Ding L. Emerging clinical significance of claudin-7 in colorectal cancer: A review. Cancer Manag Res. 2018;10:3741–52.
47. Karahashi Y, Cueno ME, Kamio N, Takahashi Y, Takeshita I, Soda K, et al. Fusobacterium nucleatum putatively affects the alveoli by disrupting the alveolar epithelial cell tight junction, enlarging the alveolar space, and increasing paracellular permeability. Biochem Biophys Res Commun [Internet]. 2023;682(October):216–22.
48. Yang CL, Yang M, Wang YF, Song CC, DU GH, Tang GY. [Effects of Fusobacterium nucleus derived outer membrane vesicles on claudin-4 expression in oral epithelial cells]. Shanghai Kou Qiang Yi Xue. 2023 Apr;32(2):126-131. Chinese. PMID: 37153991.
49. Shen W, Yuan L, Hao B, Xiang J, Cheng F, Wu Z, et al. KLF3 promotes colorectal cancer growth by activating WNT1. Aging (Albany NY). 2024;16(3):2475–93.
50. Li X, Huang J, Yu T, Fang X, Lou L, Xin S, et al. Fusobacterium nucleatum Promotes the Progression of Colorectal Cancer Through Cdk5-Activated Wnt/β-Catenin Signaling. Front Microbiol. 2021;11(January):1–13.

51. Xie Z, Huang J, Sun G, He S, Luo Z, Zhang L, et al. Integrated multi-omics analysis reveals gut microbiota dysbiosis and systemic disturbance in major depressive disorder. Psychiatry Res [Internet]. 2024;334(February):115804.
52. Jude MS, Yang CX, Leitao Filho FSS, Yang JSW, Shaipanich T, Kobor M, et al. Multi 'Omics Profiling of the HIV Airway Epithelium: Integration of the Microbiome, Methylome, and Transcriptome. 2020;(October):A2940–A2940.
53. Kazakova P, Abasolo N, de Cripan SM, Marquès E, Cereto-Massagué A, Garcia L, et al. Gut Microbiome and Small RNA Integrative-Omic Perspective of Meconium and Milk-FED Infant Stool Samples. Int J Mol Sci. 2023;24(9).
54. Parma B, Ramesh V, Gollavilli PN, Siddiqui A, Pinna L, Schwab A, et al. Metabolic impairment of non-small cell lung cancers by mitochondrial HSPD1 targeting. J Exp Clin Cancer Res. 2021;40(1):1–20.
55. Lin H, Zhao X, Xia L, Lian J, You J. Clinicopathological and Prognostic Significance of CBX3 Expression in Human Cancer: A Systematic Review and Meta-analysis. Dis Markers. 2020;2020:1–11.
56. Niu H, Chen P, Fan L, Sun B. Comprehensive pan-cancer analysis on CBX3 as a prognostic and immunological biomarker. BMC Med Genomics. 2022;15(1):1–21.
57. Li, H., Wang, L., Ruan, Z., Li, X., Yang, Y., Fang, J., & Wang, R. (2024). CSE1L as a prognostic biomarker associated with pan cancer immune infiltration and drug sensitivity. Expert Review of Clinical Immunology, 1–13.
58. Patil S, Deshpande S, Sengupta K. Nuclear envelope protein lamin B receptor protects the genome from chromosomal instability and tumorigenesis. Hum Mol Genet. 2023;32(5):745–63.
59. Shaath H, Vishnubalaji R, Elango R, Velayutham D, Jithesh PV, Alajez NM. Therapeutic targeting of the TPX2/TTK network in colorectal cancer. Cell Commun Signal. 2023;21(1):1–17.
60. Jin H, Leng Q, Zhang C, Zhu Y, Wang J. P-cymene prevent high-fat diet-associated colorectal cancer by improving the structure of intestinal flora. J Cancer. 2021;12(14):4355–61.
61. Minogue E, Cunha PP, Wadsworth BJ, Grice GL, Sah-Teli SK, Hughes R, et al. Glutarate regulates T cell metabolism and anti-tumour immunity. Nat Metab. 2023;5(October).

APPENDIX

APPENDIX 1: Pathway enrichment results.

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Term Description	**Highest p value**	**Upregulated**	**Downregulated**
Cell cycle	3.51E-01	RBL1, E2F1, E2F2, GSK3B, SKP2, CDC45, DBF4, CCNB1, CCNB2, CDC25B, CDC25C, PLK1, WEE1, PKMYT1, ANAPC1, ANAPC4, ANAPC5, CDC20, PTTG1, PTTG2, ESPL1, STAG2, TTK, BUB1, BUB1B, MAD1L1, FZR1, ATR, CREBBP, EP300, PRKDC, ORC1, ORC2, ORC6, MCM2, MCM4, MCM7	CCND3, CDKN1A, SKP1, RBX1, YWHAE, YWHAH, CDC26, ANAPC13, GADD45A

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Ubiquitin mediated proteolysis	2.13E-08	UBA6, UBE2C, UBE2O, UBE2S, BIRC6, UBE3C, SMURF1, ITCH, TRIP12, HUWE1, UBR5, HERC1, HERC2, HERC4, UBE4A, CBLB, TRAF6, COP1, BRCA1, FANCL, SKP2, VHL, ERCC8, CUL7, CDC20, FZR1, ANAPC1, ANAPC4, ANAPC5	UBB, UBE2D2, UBE2E3, UBE2E2, UBE2F, UBE2J1, UBE2L3, UBE2QL1, RBX1, SKP1, DDB2, RNF7, CDC26, ANAPC13

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Parkinson disease	3.30E-07	ITPR3, KLC3, KLC1, TUBA8	GNAL, GNAI1, GNAI2, ADCY5, PRKACA, UBB, UBE2L3, UBE2J1, SNCA, PSMC1, PSMC5, UCHL1, CALM2, CALM3, CALM1, RYR3, XBP1, SLC18A2, SLC39A7, NDUFV2, NDUFA1, NDUFA2, NDUFA4, NDUFA5, NDUFA6, NDUFA13, NDUFB1, NDUFB3, NDUFB4, NDUFB6, NDUFB8, NDUFS5

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Prion disease	1.87E-06	GRIN2B, GRIN2D, CACNA1D, ITPR3, HSPA1B, NOTCH1, IL1A, GSK3B, CSNK2A2, KLC3, KLC1, TUBA8	PRNP, NCAM1, LAMC1, PRKACA, BAX, RYR3, PPP3CA, PPP3CB, VDAC3, SLC25A4, SLC25A5, NDUFV2, NDUFA1, NDUFA2, NDUFA4, NDUFA5, NDUFA6, NDUFA13, NDUFB1, NDUFB3, NDUFB4, NDUFB6, NDUFB8, NDUFS5, NDUFS6, NDUFC1, NDUFC2, SDHC, SDHD, UQCRH, UQCRHL,

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Fanconi anemia pathway	1.17E+09	ATR, FANCM, FANCA, FANCB, FANCC, FANCE, FANCG, FANCL, FANCI, FANCD2, BRCA2, PALB2, BRCA1, BRIP1, FAN1, REV1, POLK, TOP3B, BLM, RPA4, SLX4	RPA1
Non-alcoholic fatty liver disease	1.38E-01	INSR, IRS2, GSK3B, ITCH, TRAF2, IL1A, CXCL8, CASP8	PIK3R1, AKT3, GSK3A, LEP, LEPR, ADIPOQ, PRKAG1, MAPK10, XBP1, BAX, NDUFV2, NDUFA1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Oxidative phosphorylation	4.77E+01	ATP6V0A2	NDUFS5, NDUFS6, NDUFV2, NDUFA1, NDUFA2, NDUFA4, NDUFA5, NDUFA6, NDUFA13, NDUFB1, NDUFB3, NDUFB4, NDUFB6, NDUFB8, NDUFC1, NDUFC2, SDHC, SDHD, UQCRH, UQCRHL, UQCRB, COX2, COX5A, COX6A1, COX6C, COX7A2, COX7C, ATP6V1D, ATP6V1E1, ATP6V1F,

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Diabetic cardiomyopathy	1.04E+02	SLC2A1, TNNI3, ATP2A1, SP1, CPT1B, PDK3, INSR, MTOR, GSK3B, PLCB3	PLN, CMA1, CYBB, RAC2, MAPK10, MMP2, CD36, PDK4, PIK3R1, AKT3, MPC1, PDHB, NDUFV2, NDUFA1, NDUFA2, NDUFA4, NDUFA5, NDUFA6, NDUFA13, NDUFB1, NDUFB3, NDUFB4, NDUFB6, NDUFB8, NDUFS5, NDUFS6, NDUFC1, NDUFC2, SDHC, SDHD, UQCRH, UQCRHL, UQCRB, COX2, COX5A

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Lysosome	5.55E+04	ATP6V0A2, IGF2R, AP1G2	ATP6V0D2, ATP6AP1, ATP6V0C, CTSB, CTSC, CTSG, CTSK, CTSL, CTSO, CTSS, CTSZ, LGMN, TPP1, GAA, NAGA, NAGLU, FUCA1, FUCA2, HEXB, MAN2B1, ARSB, IDS, LIPA, ACP5, ASAH1, AGA, PSAP, GM2A, PPT1, LAMP2, LAMP3, CD68, CD63, SCARB2, NPC2, LAPTM5, LAPTM4A, GNPTAB, GNPTG, AP1S2, AP3M1, AP3S1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Chemical carcinogenesis - reactive oxygen species	5.07E+03	EPHX4, MET, SOS1, BRAF, PTK2, VEGFA	GSTA4, MGST1, GSTM5, MGST2, GSTM2, NDUFV2, NDUFA1, NDUFA2, NDUFA4, NDUFA5, NDUFA6, NDUFA13, NDUFB1, NDUFB3, NDUFB4, NDUFB6, NDUFB8, NDUFS5, NDUFS6, NDUFC1, NDUFC2, SDHC, SDHD, UQCRH, UQCRHL, UQCRB, COX2, COX5A, COX6A1, COX6C

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Nucleocytoplasmic transport	2.21E+01	NUP98, NUP214, RANBP2, SUMO4, NUP85, NUP107, NUP43, AHCTF1, NUP205, NUP188, NUP155, NUP210, TPR, NUP153, KPNA7, XPO1, XPO4, XPO5, RANBP17, NCBP1, SRRM1, THOC1, THOC2	SUMO3, SUMO2, SUMO1, EEF1A1, EEF1A2, MAGOH

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Spliceosome	0,000207322902668227	RNVU1-18, RNU1-2, RNU1-4, RNU1-3, RNU1-1, RNVU1-7, RNU4-2, RNU4-1, RNU5A-1, DHX8, DHX15, PRPF40A, TCERG1, SF3B3, U2SURP, RP9, PRPF3, SNRNP200, HSPA1B, THOC1, THOC2, NCBP1, HNRNPU, SRSF10	SNRPF, LSM3, LSM6, SNU13, SNRNP27, ZMAT2, BCAS2, PLRG1, HSPA2, SYF2, MAGOH, HNRNPA1, SRSF8

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Thermogenesis	1.99E+05	KDM3A, KDM3B, SMARCA4, SMARCC1, SMARCC2, ARID1B, SOS1, TSC1, TSC2, RPTOR, MTOR, ACSL6, NDUFAF5, NDUFAF7, COX19, COA1, CPT1B	ADCY5, PRKACA, CREB3, SMARCD3, PLIN1, PRKAG1, NPR1, FGFR1, GRB2, RPS6KA2, NDUFS5, NDUFS6, NDUFV2, NDUFA1, NDUFA2, NDUFA4, NDUFA5, NDUFA6, NDUFA13, NDUFB1, NDUFB3, NDUFB4, NDUFB6, NDUFB8, NDUFC1, NDUFC2, NDUFAF1, SDHC, SDHD, UQCRH, UQCRHL

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

SNARE interactions in vesicular transport	4.33E+06	STX1A, VTI1A	VTI1B, BET1L, BET1, SNAP23, SNAP29, VAMP2, VAMP3, VAMP8, VAMP4, VAMP5, VAMP7, SEC22B
Endocytosis	7.27E+04	IGF1R, CBLB, TRAF6, ITCH, SMURF1, LDLR, ACTR3B, ACTR3C, HSPA1B, HGS, IGF2R, RAB11FIP1, PARD3, PARD6B, ASAP2, ARAP2, AGAP4, AGAP6, AGAP9	AP2M1, AP2S1, EHD2, TGFBR2, PDGFRA, SH3GL2, AMPH, DAB2, CCR5, ARRB1, ACTR2, ACTR3, ARPC2, ARPC3, ARPC4, ARPC5, WIPF1, HSPA2, CAPZB, SNX3, RAB7A, SNF8, CHMP4A

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Prostate cancer	2.22E+08	E2F1, E2F2, MMP3, IGF1R, GSK3B, CREBBP, EP300, MTOR, SOS1, BRAF	ERG, PLAT, PDGFC, PDGFD, IGF1, PDGFRA, FGFR1, PIK3R1, AKT3, CDKN1A, CREB3, TCF7L1, GRB2
Human cytomegalovirus infection	9.76E+09	TSC1, TSC2, MTOR, SOS1, SP1, TRAF5, TRAF2, CXCL8, PLCB3, ITPR3, GSK3B, VEGFA, PTK2, TAPBP, E2F1, E2F2, CASP8	PDGFRA, PIK3R1, AKT3, GRB2, ITGB3, RAC2, RHOA, IL1R1, CCL2, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, CALM2, CALM3, CALM1, PPP3CA, PPP3CB, GNAI1, GNAI2,GNAO1, ADCY5, PRKACA

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Proteoglycans in cancer	2.32E+05	BRAF, FLNB, ITPR3, NANOG, DROSHA, IGF1R, MTOR, ERBB3, TFAP4, VEGFA, MET, PLAUR, ITGA2, SOS1, PTK2, WNT2, WNT3	GRB2, RRAS, MRAS, ESR1, FLNC, RHOA, ANK2, PIK3R1, AKT3, PPP1R12B, TWIST2, DCN, IGF1, CAV1, CAV2, CD63, CDKN1A, TIMP3, MMP2, LUM, HPSE2, ITGB3, HGF, FZD1, FZD4, SDC2, RDX, MSN, FGF2, FGFR1, NUDT16L1, WNT2B, WNT5B, CTSL, PRKACA, HSPB2
DNA replication	0,000128169524388138	DNA2, POLA1, PRIM2, POLD1, POLE, POLE2, MCM2,	RPA1, POLE3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Regulation of actin cytoskeleton	2.70E+07	ITGA2, ITGB4, ITGB8, PTK2, DOCK1, SOS1, BRAF, PAK6, MYLK2, DIAPH2, PIKFYVE, DIAPH3, ACTR3B, ACTR3C, SSH2, IQGAP3, SPATA13	LPAR1, LPAR4, CXCL12, F2R, FGF1, FGF2, FGF7, FGF10, PDGFC, PDGFD, FGFR1, PDGFRA, ITGA7, ITGA8, ITGB3, ITGB7, RRAS, MRAS, ARHGEF6, PIK3R1, RHOA, RAC2, PAK3, MYLK, PPP1R12B, MYL9, MYL12A, MYH11, PIP4K2A, NCKAP1L, BRK1, ACTR2, ACTR3, ARPC2, ARPC3, ARPC4, ARPC5, PFN1, RDX, MSN, TMSB4X

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Breast cancer	7.32E+06	SP1, IGF1R, SOS1, BRAF, MTOR, NOTCH1, WNT2, WNT3, LRP5, LRP6, GSK3B, POLK, E2F1, E2F2, BRCA1, BRCA2	ESR1, PGR, FGF1, FGF2, FGF7, FGF10, FGFR1, IGF1, KIT, SHC4, GRB2, PIK3R1, AKT3, HEY1, CDKN1A, WNT2B, WNT5B, FZD1, FZD4, FRAT1, TCF7L1, GADD45A, BAX, DDB2
mTOR signaling pathway	1.40E+07	SLC7A5, FNIP1, DEPDC5, SKP2, RPTOR, MTOR, TTI1, LPIN3, TSC1, TSC2, WNT2, WNT3, LRP5, LRP6, GSK3B, INSR, IGF1R, SOS1	ATP6V1D, ATP6V1E1, ATP6V1F, ATP6V1G2, LAMTOR1, LAMTOR3, LAMTOR5, RRAGA, RRAGC, SESN2, EIF4E2, WNT2B, WNT5B, FZD1, FZD4, IGF1, GRB2, RPS6KA2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Thyroid hormone signaling pathway	2.55E+09	NCOR1, NCOA2, CREBBP, EP300, MED12, MED13L, MED13, MED14, ATP2A1, NOTCH1, PLCB3, GSK3B, TSC2, MTOR, SLC2A1	PRKACA, ITGB3, ESR1, RXRG, KAT2B, PLN, PLCD4, ATP1A2, ATP1B2, PIK3R1, AKT3, RCAN2, SLCO1C1
Progesterone-mediated oocyte maturation	9.79E+07	AURKA, SPDYA, PKMYT1, CCNB1, CCNB2, PLK1, CDC25B, CDC25C, IGF1R, BRAF, BUB1, MAD1L1, FZR1, ANAPC1	PGR, PIK3R1, MAPK10, GNAI1, GNAI2, ADCY5, PRKACA, CPEB1, RPS6KA2, IGF1, AKT3, CDC26, ANAPC13

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Protein export	3.94E+08	IMMP1L	OXA1L, SEC61B, SEC61G, SEC62, SEC63, SRP9, SRP14, SRP68, SPCS1, SPCS2, SEC11C, SEC11A
Retrograde endocannabinoid signaling	4.99E+05	CACNA1D, PLCB3, ITPR3, DAGLB, GABRE	GNAI1, GNAI2, GNAO1, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, MAPK10, ADCY5, PRKACA, NDUFV2, NDUFA1, NDUFA2, NDUFA4, NDUFA5, NDUFA6, NDUFA13, NDUFB1, NDUFB3, NDUFB4

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Tight junction	1.66E+07	CLDN1, CLDN9, OCLN, PARD6B, NF2, LLGL2, DLG1, PPP2R2D, PARD3, DLG3, SYMPK, RUNX1, CACNA1D, ACTR3B, ACTR3C, RAPGEF6, TJP2, TUBA8	CLDN11, CLDN5, JAM2, JAM3, BVES, PARD6G, MPDZ, AMOTL1, PPP2CB, RHOA, MAPK10, CD1A, CD1B, CD1D, CD1E, YBX3, RDX, MSN, ACTR2, ACTR3, ARPC2, ARPC3, ARPC4, ARPC5, PRKACA, RAB13, PRKAG1, MYH11, MYL6, MYL9, MYL12A, RAB8B, RAP1A, TUBA1A

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

TGF-beta signaling pathway	1.18E+05	AMH, INHBA, ACVR2B, SMURF1,	CHRD, GREM1, GREM2, DCN, FMOD, FST, BMP6, BMP5, GDF6, RGMA, TGFBR2, SMAD1, ID4, RBX1, SKP1, RHOA, PPP2CB
Transcriptional misregulation in cancer	1.46E+08	RUNX1, NCOR1, PER2, ETV6, KMT2A, CCNT2, DOT1L, HOXA9, JMJD1C, HMGA2, KDM6A, IGF1R, TLX1, BCL11B, REL, MMP3, ETV4, ELK4, PTK2, SP1, CXCL8, ASPSCR1, MET, POLK	CSF1R, SPI1, CD14, ZBTB16, LMO2, PBX3, SMAD1, MEF2C, FLT3, LYL1, HHEX, CD86, IGH, MAF, ITGB7, RXRG, TFE3, CDKN1A, ERG, PLAT, NGFR, FLI1, IGF1, TGFBR2, MITF, NUPR1, GADD45A, BAX, DDB2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Insulin signaling pathway	1.55E+07	INSR, IRS2, GSK3B, CBLB, ACACA, FASN, HKDC1, MTOR, RPTOR, TSC1, TSC2, SOS1, BRAF	PIK3R1, AKT3, PPP1R3C, CALM2, CALM3, CALM1, PYGM, PRKACA, PRKAR2B, SORBS1, RHOQ, PRKAG1, EIF4E2, SHC4, GRB2, MAPK10

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Salmonella infection	2.37E+08	DYNC1H1, TUBA8, KLC3, KLC1, GCC2, FLNB, EXOC4, ACTR3B, ACTR3C, MYO6, PIK3C2A, CYTH2, TRAF6, NOD1, CXCL8, TRAF2, CASP8	RAB5B, RAB5C, RAB7A, RAB7B, DYNLL1, DYNLL2, DYNLT1, DCTN3, DCTN6, ACTR1A, ACTR10, TUBA1A, TUBB6, TUBB4A, ARL8B, KIF5A, KIF5C, RHOA, RAB9A, RAB9B, PFN1, FLNC, S100A10, RHOG, MYL9, MYL12A, NCKAP1L, ABI1, BRK1, ACTR2, ACTR3, ARPC2, ARPC3, ARPC4

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Kaposi sarcoma-associated herpesvirus infection	0,000180868877013494	CREBBP, EP300, EIF2AK2, CASP8, TRAF2, E2F1, E2F2, MTOR, ITPR3, CXCL8, CXCL2, VEGFA, GSK3B	C3, IFNAR1, CCR5, BAX, CDKN1A, CD86, UBB, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, PIK3R1, AKT3, MAPK10, CD200R1, HCK, CALM2, CALM3, CALM1, PPP3CA, PPP3CB, FGF2, IL6ST, TCF7L1
Oocyte meiosis	4.21E+08	IGF1R, AURKA, PKMYT1, SPDYA, PLK1, CDC25C, CCNB2, ANAPC1, ANAPC4, ANAPC5	IGF1, PGR, ADCY5, PRKACA, CPEB1, RPS6KA2, YWHAE, YWHAH, RBX1, SKP1, CDC26, ANAPC13

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Autophagy - animal	7.05E+05	IGF1R, IRS2, TSC2, TSC1, RPTOR, MTOR, ATG2A, ATG2B, ATG9B, C9orf72, AMBRA1, TRAF6, VMP1, DAPK2, ATG16L1	PIK3R1, AKT3, MRAS, RRAS, PPP2CB, WIPI1, RAB39B, RAB1A, RRAGA, RRAGC, PRKACA, MAPK10, NRBF2, CAMKK2, GABARAP, GABARAPL1, GABARAPL2, ATG4A, ATG4C, LAMP2, RAB7A, RAB7B, VAMP8, SNAP29, CTSL, CTSB

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Ribosome biogenesis in eukaryotes	0,00094029369786839	CSNK2A2, HEATR1, PWP2, TCOF1, UTP14C, DROSHA, RCL1, POP1, RPP25, XRN1, GNL3L, NVL, MDN1, RBM28, XPO1, SPATA5, RNA5S9, SNORD3A, SNORD3C, SNORD3B-2, SNORD3B-1, RMRP	IMP3, SNU13, GAR1, FCF1, RPP30, SBDS, AK6
HIF-1 signaling pathway	0,000183359677901198	INSR, IGF1R, MTOR, VHL, CREBBP, EP300, VEGFA, NOS2, SLC2A1,	IGF1, PIK3R1, AKT3, EIF4E2, RBX1, CYBB, ANGPT1, TEK, CDKN1A, PDHB

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Viral carcinogenesis	2.72E+07	REL, EIF2AK2, CREBBP, EP300, SKP2, TRAF2, TRAF5, DLG1, RBL1, UBR4, CHD4, HDAC4, HDAC8, RASA2, CDC20, MAD1L1, VAC14, CASP8	CREB3, YWHAE, YWHAH, PIK3R1, VDAC3, CDKN1A, PRKACA, BAX, GRB2, CCND3, ATP6V0D2, PSMC1, KAT2B, GTF2A2, RHOA, GSN, IL6ST, CCR5, C3
Mismatch repair	0,026867125017708	MSH3, MLH3, RFC5, EXO1, RPA4, POLD1	RPA1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Hepatocellular carcinoma	4.49E+09	IGF1R, SOS1, BRAF, MTOR, E2F1, E2F2, TERT, TERC, POLK, WNT2, WNT3, LRP5, LRP6, GSK3B, MET, ARID1B, SMARCC1, SMARCC2, SMARCA4, ARID2, PBRM1	SHC4, GRB2, PIK3R1, AKT3, CDKN1A, GADD45A, BAX, DDB2, TGFBR2, WNT2B, WNT5B, FZD1, FZD4, FRAT1, TCF7L1, NFE2L2, GSTA4, MGST1, GSTM5, MGST2, GSTM2, HGF, SMARCD3
Glioma	3.32E+08	IGF1R, SOS1, BRAF, MTOR, E2F1, E2F2, POLK	PDGFRA, IGF1, CALM2, CALM3, CALM1, CAMK1, SHC4, GRB2, PIK3R1, AKT3, CDKN1A, GADD45A, BAX, DDB2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Ras signaling pathway	4.10E+08	VEGFA, INSR, IGF1R, MET, SOS1, GRIN2B, NF1, RASA2, RASAL2, STK4, PAK6, JMJD7-PLA2G4B, RAPGEF5, REL	FGF1, FGF2, FGF7, FGF10, IGF1, PDGFC, PDGFD, CSF1, KITLG, FLT3LG, VEGFC, HGF, ANGPT1, FGFR1, NGFR, PDGFRA, CSF1R, KIT, FLT3, TEK, GRB2, SHC4, RASGRP4, HTR7, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, PRKACA, CALM2, CALM3, CALM1, MRAS, RRAS, RAC2, PAK3, RHOA, PIK3R1, AKT3, PLA1A, RAP1A, RGL1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Homologous recombination	0,000233111215263407	BRCA1, BARD1, BRIP1, TOPBP1, PALB2, BRCA2, RPA4, RAD52, RAD51B, XRCC2, XRCC3, RAD54L, RAD54B, POLD1, BLM, TOP3B	RPA1
Nucleotide excision repair	9.38E+06	ERCC8, ERCC6, RPA4, POLD1, POLE, POLE2, RFC5	RBX1, DDB2, XPC, GTF2H5, RPA1, POLE3
RNA polymerase	6.00E+09	POLR2J3, POLR2J2, POLR3A, POLR3E, POLR1B, POLR1A	POLR2F, POLR2L, POLR3GL

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

ErbB signaling pathway	3.66E+09	CBLB, PTK2, PAK6, ERBB3, SOS1, BRAF, MTOR, GSK3B	NCK1, PAK3, MAPK10, SHC4, GRB2, PIK3R1, AKT3, CDKN1A
Focal adhesion	8.09E+07	LAMB3, LAMC2, IBSP, ITGA2, ITGB4, ITGB8, VEGFA, IGF1R, MET, ARHGAP5, MYLK2, FLNB, GSK3B, PTK2, PAK6, DOCK1, BRAF, SOS1	COL4A6, COL4A5, COL6A1, COL6A2, LAMA2, LAMA4, LAMC1, RELN, TNXB, ITGA7, ITGA8, ITGB3, ITGB7, PDGFC, PDGFD, IGF1, VEGFC, HGF, PDGFRA, RHOA, MYL9, MYL12A, PPP1R12B, MYLK, FLNC, PARVA, AKT3, PIK3R1, RAC2, PAK3, RAP1A, MAPK10, CAV1, CAV2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

MAPK signaling pathway	3.85E+07	CACNA1D, CACNG8, NF1, RASA2, VEGFA, ERBB3, INSR, IGF1R, MET, SOS1, BRAF, ELK4, JMJD7-PLA2G4B, IL1A, TRAF2, TRAF6, MAP4K3, STK4, STK3, MAP3K4, TAOK1, MAPK8IP3, FLNB, MAPKAPK5, CDC25B, HSPA1B	CACNA2D1, PRKACA, PPP3CA, PPP3CB, RASGRP4, RAP1A, FGF1, FGF2, FGF7, FGF10, IGF1, PDGFC, PDGFD, CSF1, KITLG, FLT3LG, VEGFC, HGF, ANGPT1, FGFR1, NGFR, PDGFRA, CSF1R, KIT, FLT3, TEK, GRB2, RRAS, MRAS, LAMTOR3, RPS6KA2, IL1R1, TGFBR2, CD14, RAC2, GADD45A, TAB2, MAP3K3, MAP3K12

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Renal cell carcinoma	1.29E+08	VHL, CREBBP, EP300, SLC2A1, VEGFA, MET, SOS1, BRAF, PAK6	RBX1, HGF, PIK3R1, AKT3, RAP1A, GRB2, PAK3, TFE3, CDKN1A
Platelet activation	0,0316348523083802	ITPR3, MYLK2, PLCB3, ITGA2, JMJD7-PLA2G4B	TBXA2R, F2R, RHOA, MYL12A, P2RX1, STIM1, MYLK, RAP1A, ITGB3, P2RY12, GNAI1, GNAI2, ADCY5, PTGIR, PRKACA, FCER1G, PIK3R1, BTK, GP1BA, AKT3, PTGS1, SNAP23, VAMP8

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

FoxO signaling pathway	0,000499613259961337	SKP2, CREBBP, EP300, STK4, IGF1R, INSR, IRS2, SOS1, BRAF, CCNB1, CCNB2, PLK1, PLK4	TGFBR2, PRKAG1, MAPK10, IGF1, PIK3R1, AKT3, SGK3, GRB2, CDKN1A, GADD45A, GABARAP, GABARAPL1, GABARAPL2, IL7R, S1PR1
Estrogen signaling pathway	1.17E+08	HSPA1B, NCOA2, KRT16, KRT23, SP1, SOS1, PLCB3, ITPR3	ESR1, FKBP5, HSPA2, KRT10, KRT24, PGR, MMP2, ADCY5, PRKACA, CREB3, SHC4, GRB2, PIK3R1, AKT3, GNAI1, GNAI2, GNAO1, CALM2, CALM3, CALM1
Adherens junction	0,0002194443772575	PARD3, FARP2, LMO7, CSNK2A2	RAC2, SORBS1, RHOA, TCF7L1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Melanoma	0,000328589803729029	MET, IGF1R, BRAF, E2F1, E2F2, POLK	FGF1, FGF2, FGF7, FGF10, HGF, IGF1, PDGFC, PDGFD, FGFR1, PDGFRA, PIK3R1, AKT3, CDKN1A, GADD45A, BAX, DDB2, MITF
Human T-cell leukemia virus 1 infection	2.33E+08	SLC2A1, TRRAP, XPO1, VAC14, MAD1L1, BUB1B, ANAPC1, ANAPC4, ANAPC5, CDC20, PTTG1, PTTG2, ESPL1, CCNB2, E2F1, E2F2, ATR, DLG1, ELK4	NRP1, CD4, PPP3CA, PPP3CB, IL2RG, B2M, SLC25A4, SLC25A5, VDAC3, CDC26, CCND3, CDKN1A, PIK3R1, AKT3, MAPK10, IL1R1, MAP3K3, GPS2, TGFBR2, ADCY5, PRKACA

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Cellular senescence	8.71E+08	RBL1, E2F1, E2F2, TSC1, TSC2, MTOR, MYBL2, LIN9, FOXM1, CCNB1, CCNB2, HUS1, ATR, IL1A, CXCL8, CACNA1D, TRPM7, ITPR3	TGFBR2, CCND3, PIK3R1, CDKN1A, RRAS, MRAS, AKT3, GADD45A, ZFP36L2, CALM2, CALM3, CALM1, PPP3CA, PPP3CB, SLC25A4, SLC25A5, VDAC3
Calcium signaling pathway	0,000577355785457147	ATP2A1, CACNA1D, GRIN2D, NTSR1, OXTR, LTB4R2, VEGFA, ERBB3, MET, MST1R, PLCB3, ITPR3, MYLK2	SLC8A2, SLC8A3, ATP2B4, ADRB2, HTR7, GNAL, PRKACA, PLN, STIM1, P2RX1, RYR3, CASQ2, CYSLTR1, EDNRB, F2R, HTR2A, PTGER1, PTGER3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Central carbon metabolism in cancer	5.16E+09	SLC2A1, MET, GLS, HKDC1, MTOR, SLC7A5	PDHB, KIT, RET, NTRK3, PDGFRA, FGFR1, FLT3, PIK3R1, AKT3
Fc gamma R-mediated phagocytosis	2.05E+09	JMJD7-PLA2G4B, ACTR3B, ACTR3C, DOCK1, ASAP2, MYO10	IGH, HCK, PIK3R1, AKT3, GSN, ACTR2, ACTR3, ARPC2, ARPC3, ARPC4, ARPC5, RAC2, CFL1, CFL2, FCGR2B, AMPH
Neurotrophin signaling pathway	9.35E+09	SOS1, BRAF, GSK3B, TP73, TRAF6, IRAK2	NTRK3, SH2B3, GRB2, RPS6KA2, RAP1A, MAP3K3, SHC4, PIK3R1, AKT3, CALM2, CALM3, CALM1, MATK, NGFR, ARHGDIB, RHOA, MAPK10, BAX

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Lysine degradation	0,00296348544666673	CAMKMT, KMT2A, KMT2D, KMT2C, KMT2B, KMT2E, SETD1B, ASH1L, SMYD3, EHMT1, SETDB1, EZH2, NSD1, SETD2, DOT1L, PLOD3	DLD, HADHA, ACAT1, ALDH9A1
Growth hormone synthesis, secretion and action	0,00034155747090643	CREBBP, EP300, PLCB3, ITPR3, CACNA1D, SOS1, IRS2, MTOR, GSK3B, PTK2	ADCY5, SST, SSTR3, GNAI1, GNAI2, PRKACA, CREB3, GHR, SHC4, GRB2, PIK3R1, AKT3, IGF1, MAPK10
Longevity regulating pathway	0,000336781609195542	INSR, IGF1R, IRS2, EHMT1, TSC1, TSC2	IGF1, PIK3R1, AKT3, ADCY5, PRKACA, CREB3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Valine, leucine and isoleucine degradation	3.01E+09	ACSF3	DLD, ACADM, ACADSB, HADHA, HADHB, ALDH9A1, AOX1, HMGCLL1, ACAT1
Protein processing in endoplasmic reticulum	1.94E+08	UGGT2, HSPA1B, YOD1, EIF2AK2, TRAF2	SEC61B, SEC61G, SEC62, SEC63, DAD1, TUSC3, SIL1, MAN1C1, ERP29, TXNDC5, OS9, ERLEC1, SSR4, TRAM1, TRAM1L1, DERL3, UBXN8, HSPA2, DNAJB2, CRYAB, NFE2L2, XBP1, MAPK10, BAX, UBE2J1, RNF5, RNF185, SEL1L, HERPUD1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Rap1 signaling pathway	3.97E+09	RAPGEF5, GRIN2B, PLCB3, VEGFA, INSR, IGF1R, MET, RAPGEF6, SIPA1L1, SIPA1L3, FARP2, PARD3, PARD6B, KRIT1, BRAF	RAP1A, MRAS, CALM2, CALM3, CALM1, F2R, LPAR1, LPAR4, ADCY5, FGF1, FGF2, FGF7, FGF10, IGF1, PDGFC, PDGFD, CSF1, KITLG, VEGFC, HGF, ANGPT1, FGFR1, NGFR, PDGFRA, CSF1R, KIT, TEK, GNAI1, GNAI2, GNAO1, RALB, RAC2, ITGB3, PFN1, RHOA, PARD6G, PIK3R1, AKT3, RRAS
Non-small cell lung cancer	0,0013108441612467	FHIT, E2F1, E2F2, STK4, MET, SOS1, BRAF, POLK	RXRG, PIK3R1, AKT3, HGF, GRB2, CDKN1A, GADD45A

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

mRNA surveillance pathway	0,000179262068030626	NCBP1, SRRM1, RNGTT, CPSF6, PCF11, WDR33, CSTF2, SYMPK, PABPC1L, SMG1, SMG5, PPP2R2D, PPP2R3B	MAGOH, TARDBP, WDR82, SSU72, CSTF2T, PABPC5, PPP2CB, PPP2R3C
Choline metabolism in cancer	0,0331591099695767	SOS1, JMJD7-PLA2G4B, TSC1, TSC2, MTOR, SP1, SLC22A1, SLC22A3, DGKE, DGKH, GPCPD1	PDGFC, PDGFD, PDGFRA, GRB2, MAPK10, PIK3R1, AKT3, RAC2
Mitophagy - animal	5.22E+08	AMBRA1, ATG9B, PGAM5, CSNK2A2, E2F1, SP1	MAPK10, UBB, GABARAP, GABARAPL1, GABARAPL2, MITF, TFE3, FUNDC1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Pathogenic Escherichia coli infection	0,000128183524872689	CYTH2, ACTR3B, ACTR3C, MYO5B, MYO5C, MYO6, MYO10, TUBA8, BAIAP2L1, TRAF6, TRAF2, CXCL8, CASP8, OCLN, CLDN1, CLDN9	IGH, NCKAP1L, ABI1, BRK1, ACTR2, ACTR3, ARPC2, ARPC3, ARPC4, ARPC5, RHOA, MYH11, LPAR1, LPAR4, F2R, TUBA1A, TUBB6, TUBB4A, NCK1, WIPF1, IL1R1, TAB2, MAPK10, BAX, PAK3, RAB1A, CLDN11, CLDN5
Spinocerebellar ataxia	0,0117100857880519	PLCB3, ITPR3, ATXN2L, ATXN2, GRIN2B, GRIN2D, SPTBN2, TRAF2, ATXN1	VDAC3, SLC25A4, SLC25A5, FGF14, PSMC1, PSMC5, MAPK10, XBP1, NRBF2, WIPI1, RELN, PIK3R1, AKT3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

p53 signaling pathway	0,00287977313076924	ATR, MDM4, CCNB1, CCNB2, GTSE1, CASP8, RRM2, TSC2, COP1, TP73	CDKN1A, CCND3, GADD45A, BAX, EI24, ZMAT3, IGF1, DDB2, RRM2B, SESN1, SESN2, CCNG1
Fluid shear stress and atherosclerosis	0,00259457091969194	PTK2, IL1A, VEGFA, ACVR2B, SUMO4	CAV1, CAV2, CALM2, CALM3, CALM1, SDC2, NFE2L2, GSTA4, MGST1, GSTM5, MGST2, GSTM2, TXN2, THBD, PLAT, CDH5, PECAM1, PIK3R1, AKT3, MEF2C, ITGB3, RHOA, DUSP1, IL1R1, MAPK10, MMP2, CTSL, CCL2, VCAM1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Bacterial invasion of epithelial cells	4.63E+09	MET, DOCK1, ACTR3B, ACTR3C, PTK2, ELMO3	PIK3R1, ACTR2, ACTR3, ARPC2, ARPC3, ARPC4, ARPC5, SHC4, CAV1, CAV2, RHOG, RHOA
AMPK signaling pathway	0,00831086931555911	FASN, ACACA, MLYCD, CPT1B, PPP2R2D, PPP2R3B, IGF1R, INSR, IRS2, TSC2, TSC1, MTOR, RPTOR	LEP, LEPR, CAMKK2, ADIPOQ, PRKAG1, CREB3, PPP2CB, PPP2R3C, RAB14, CD36, IGF1, PIK3R1, AKT3
Hepatitis C	0,0245884888590123	LDLR, CLDN1, CLDN9, OCLN, TRAF6, EIF2AK2, PPP2R2D, SOS1, BRAF	CD81, CLDN11, CLDN5, IFNAR1, PPP2CB, GRB2, YWHAE, YWHAH, PIK3R1, AKT3, CDKN1A, BAX

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

JAK-STAT signaling pathway	0,0234746238620112	IL11, IL24, IFNE, OSM, CNTF, CSF3, IL20RA, CREBBP, EP300, SOS1, MTOR	TSLP, LEP, IL2RG, IL5RA, IL7R, IL6ST, IFNAR1, OSMR, LIFR, CSF2RB, GHR, LEPR, PDGFRA, CCND3, CDKN1A, AOX1, FHL1, GRB2, PIK3R1, AKT3
Shigellosis	0,000354549320901773	PTK2, DOCK1, ACTR3B, ACTR3C, CYTH2, GSK3B, PLCB3, ITPR3, HKDC1, TRAF2, TRAF5, CXCL8, TRAF6, NOD1, RBCK1, MTOR	ACTR2, ACTR3, ARPC2, ARPC3, ARPC4, ARPC5, RHOA, MYL9, MYL12A, PFN1, PIK3R1, AKT3, GSK3A, PLCD4, CAPNS1, BAX, TAB2, IL1R1, CD14, MAPK10, UBE2D2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Human immunodeficiency virus 1 infection	0,000199495418559803	PTK2, MTOR, ITPR3, PAK6, TRAF6, TRAF5, TRAF2, CASP8, TAPBP, AP1G2, ATR, CDC25C, WEE1, CCNB1, CCNB2	CD4, SKP1, RBX1, CCR5, GNAI1, GNAI2, GNAO1, PIK3R1, AKT3, CALM2, CALM3, CALM1, PPP3CA, PPP3CB, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, RAC2, PAK3, CFL1, CFL2, TAB2, MAPK10, APOBEC3C, SAMHD1, BAX, AP1S2, B2M, RNF7
Longevity regulating pathway - multiple species	0,000182260748099597	INSR, IGF1R, IRS2, MTOR, RPTOR, HSPA1B	IGF1, PIK3R1, AKT3, ADCY5, PRKACA, PRKAG1, HSPA2, CRYAB

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Colorectal cancer	0,00301664522271728	GSK3B, BRAF, MSH3, POLK, SOS1, MTOR	TCF7L1, PIK3R1, AKT3, RALB, RAC2, RHOA, MAPK10, TGFBR2, BAX, CDKN1A, GADD45A, DDB2, GRB2
Hepatitis B	0,00178792835871102	CREBBP, EP300, CASP8, E2F1, E2F2, SOS1, CXCL8, BRAF, TRAF6	TGFBR2, CDKN1A, MAPK10, DDB2, PIK3R1, AKT3, BAX, ATP6AP1, VDAC3, GRB2, CREB3, TAB2, TICAM2, IFNAR1
Dopaminergic synapse	0,0206728021362605	PLCB3, ITPR3, CACNA1D, DRD4, PPP2R2D, PPP2R3B, GSK3B, GRIN2B	SLC18A2, CALM2, CALM3, CALM1, PPP3CA, PPP3CB, GNAL, ADCY5, PRKACA, CREB3, MAPK10

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Endometrial cancer	0,00476119248213705	SOS1, BRAF, GSK3B, POLK	PIK3R1, AKT3, GRB2, TCF7L1, CDKN1A, GADD45A, BAX, DDB2
Phagosome	0,000117268952416092	HGS, ATP6V0A2, DYNC1H1, TUBA8, PIKFYVE, ITGA2	VAMP3, STX12, SEC22B, HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, RAB5B, RAB5C, ATP6V1D, ATP6V1E1, ATP6V1F, ATP6V1G2, ATP6V0E1, ATP6V0D2, ATP6V0C, ATP6AP1, RAB7A, RAB7B, TUBA1A, TUBB6

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Glucagon signaling pathway	5.19E+09	CREBBP, EP300, PLCB3, ITPR3, CPT1B, ACACA, SLC2A1	AKT3, PRKACA, CREB3, PPP3CA, PPP3CB, CALM2, CALM3, CALM1, PYGM, PRKAG1, PDHB
Gap junction	0,022639068940854	SOS1, TUBA8, PLCB3, ITPR3	GJA1, LPAR1, GNAI1, GNAI2, PDGFC, PDGFD, PDGFRA, GRB2, TUBA1A, TUBB6, TUBB4A, ADCY5, PRKACA, HTR2A
Signaling pathways regulating pluripotency of stem cells	0,014702412381721	NANOG, INHBA, ACVR2B, WNT2, WNT3, GSK3B	LIFR, IL6ST, GRB2, PIK3R1, AKT3, SMAD1, ID4, WNT2B, WNT5B, FZD1, FZD4, FGF2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Long-term potentiation	0,0152485373505669	GRIN2B, GRIN2D, CREBBP, EP300, BRAF, ITPR3, PLCB3	PRKACA, PPP1R1A, RAP1A, PPP3CA, PPP3CB, CALM2, CALM3, CALM1, RPS6KA2
Small cell lung cancer	0,000355165452314091	FHIT, POLK, SKP2, E2F1, E2F2, LAMB3, LAMC2, ITGA2, PTK2, TRAF2, TRAF5, TRAF6, NOS2	RXRG, CDKN1A, GADD45A, BAX, DDB2, COL4A6, COL4A5, LAMA2, LAMA4, LAMC1, PIK3R1, AKT3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Chemokine signaling pathway	0,00100415136338604	CXCL2, CXCL8, SOS1, BRAF, GSK3B, PTK2, PLCB3, PARD3	CXCL12, CXCL14, CCL2, CCL11, CCL13, CCL15, CCL23, CCL19, CCL21, CCL22, CX3CR1, CCR7, CCR5, CCR10, GNAI1, GNAI2, ADCY5, PRKACA, HCK, SHC4, GRB2, PIK3R1, AKT3, GSK3A, RAC2, RHOA, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, RAP1A, ARRB1
Chronic myeloid leukemia	0,00103550457871261	SOS1, BRAF, E2F1, E2F2, POLK, RUNX1	PIK3R1, AKT3, GRB2, SHC4, CDKN1A, GADD45A, BAX, DDB2, TGFBR2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Circadian rhythm	0,00341487259895933	PER2, NPAS2	PER1, RORB, RBX1, SKP1, PRKAG1
Autophagy - other	8.12E+09	RPTOR, MTOR, ATG9B, ATG2A, ATG2B, ATG16L1	PPP2CB, WIPI1, GABARAP, GABARAPL1, GABARAPL2, ATG4A, ATG4C
Gastric cancer	0,00192788314526829	POLK, WNT2, WNT3, LRP5, LRP6, GSK3B, TERT, TERC, SOS1, BRAF, MTOR, E2F1, E2F2, MET	CDKN1A, GADD45A, BAX, DDB2, RXRG, WNT2B, WNT5B, FZD1, FZD4, FRAT1, TCF7L1, SHC4, GRB2, PIK3R1, AKT3, TGFBR2, HGF, FGF1, FGF2, FGF7, FGF10
Basal transcription factors	7.78E+09	TAF1, TAF2, TAF4, GTF2IRD1	GTF2A2, TAF4B, TAF12, TAF9B, TAF13, GTF2H5

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Vascular smooth muscle contraction	0,0217145348516359	JMJD7-PLA2G4B, CACNA1D, MYLK2, PLCB3, ITPR3, BRAF	KCNMA1, KCNMB1, CALM2, CALM3, CALM1, MYLK, MYL6, MYL9, MYH11, CALD1, ACTA2, ACTG2, PPP1R14A, PPP1R12B, RHOA, PTGIR, CALCRL, RAMP2, RAMP3, ADCY5, PRKACA, NPR1
Endocrine and other factor-regulated calcium reabsorption	0,00340840816680367	VDR, PLCB3	PTH1R, PRKACA, ESR1, KL, AP2M1, AP2S1, RAB11A, ATP1A2, ATP1B2, SLC8A2, SLC8A3, ATP2B4

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

B cell receptor signaling pathway	0,0324089838446819	SOS1, GSK3B	IGH, CD79A, BTK, RAC2, PPP3CA, PPP3CB, GRB2, CD81, PIK3R1, AKT3, FCGR2B, LILRB1, LILRB5, LILRA1
Lipid and atherosclerosis	0,0302423240048858	LDLR, POU2F1, CYP2B6, CXCL8, MMP3, HSPA1B, TRAF6, PTK2, CXCL2, TRAF2, GSK3B, PLCB3, CASP8	RHOA, PIK3R1, AKT3, BAX, MAPK10, CYBB, CCL2, SELP, VCAM1, HSPA2, CD14, LY96, TAB2, TICAM2, CD36, RXRG, RAP1A, XBP1, NFE2L2, CALM2, CALM3, CALM1, PPP3CA, PPP3CB

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Yersinia infection	0,0222904318242193	PTK2, DOCK1, ACTR3B, ACTR3C, TRAF6, TRAF2, CXCL8, GSK3B	RHOG, IGH, RAC2, ACTR2, ACTR3, ARPC2, ARPC3, ARPC4, ARPC5, WIPF1, RHOA, RPS6KA2, TAB2, MAPK10, PIK3R1, AKT3, CCL2, CD4
NOD-like receptor signaling pathway	0,041564009053569	NOD1, CXCL8, CXCL2, ATG16L1, RBCK1, TRAF2, TRAF5, TRAF6, TRPM7, PLCB3, ITPR3, NAMPT, CASP8, TP53BP1	CCL2, TAB2, MAPK10, GABARAP, GABARAPL1, GABARAPL2, CTSB, VDAC3, CYBB, TXNIP, TXN2, RHOA, YWHAE, IFNAR1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Acute myeloid leukemia	0,0131172727612405	MTOR, SOS1, BRAF, RUNX1, PER2	KIT, FLT3, PIK3R1, AKT3, GRB2, PIM2, CSF1R, ZBTB16, SPI1, CD14, TCF7L1
Thyroid cancer	0,0138230548118833	TPR, BRAF, POLK	RET, NCOA4, RXRG, CDKN1A, GADD45A, BAX, DDB2, TCF7L1
Axon guidance	0,0114864014815638	PTK2, ABLIM2, PAK6, LRIG2, EPHA1, RGS3, PLXNA3, PLXNA1, GSK3B, PLXNB1, MET, PARD3, PARD6B, SSH2	LRRC4C, NTN1, PPP3CA, PPP3CB, RAC2, NCK1, PAK3, RHOA, UNC5C, UNC5D, RGMA, MYL9, MYL12A, EFNB3, EPHA3, EPHA7, CXCL12, GNAI1, GNAI2, SLIT2, SLIT3, SEMA3D, SEMA3E, SEMA3G

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Leukocyte transendothelial migration	0,00057478674434365	CLDN1, CLDN9, OCLN, PTK2, ARHGAP5	JAM3, JAM2, PECAM1, CD99, CD99L2, CDH5, CLDN11, CLDN5, VCAM1, MSN, PIK3R1, CYBB, MMP2, THY1, RHOA, MYL9, MYL12A, RAP1A, CXCL12, GNAI1, GNAI2, RAC2
Pancreatic cancer	0,00788006026487999	BRAF, MTOR, VEGFA, E2F1, E2F2, POLK, BRCA2	PIK3R1, ARHGEF6, RAC2, AKT3, MAPK10, RALB, RALBP1, CDKN1A, GADD45A, BAX, DDB2, TGFBR2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Hippo signaling pathway	0,0191954165222646	PARD3, PARD6B, LIMD1, AJUBA, NF2, STK3, PPP2R2D, LATS1, LLGL2, DLG1, DLG3, DLG5, TP73, AMH, WNT2, WNT3, GSK3B	PARD6G, FRMD6, PPP2CB, FGF1, TGFBR2, BMP5, BMP6, GDF6, SMAD1, WNT2B, WNT5B, FZD1, FZD4, YWHAE, YWHAH, TCF7L1, CCND3, SNAI2
Fatty acid degradation	0,0084548763292134	ACSL6, CPT1B	ACAT1, HADHB, HADHA, ACADM, ACADSB, CYP2U1, ADH1B, ADH5, ALDH9A1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Coronavirus disease - COVID-19	0,0258483366063783	MMP3, CXCL8, ADAM17, CSF3, TRAF6, EIF2AK2	NRP1, CYBB, CCL2, IL6ST, RPS9, RPS16, RPS25, RPS27L, RPS29, RPL3, RPL5, RPL17-C18orf32, RPL35A, RPL41, RPL36AL, RPL36A-HNRNPH2, IGH, PIK3R1, TLR7, TAB2, MAPK10, IFNAR1, C3AR1, C3, CFD, C7, SELP, C1QA, C1QB, C1QC, C1R, C1S, C4A, C4B, MASP1, F13A1
Alcoholism	0,0358987438844975	GRIN2B, GRIN2D, HDAC4, HDAC8, SLC29A2, SOS1, BRAF	SLC18A2, GNAI1, GNAI2, GNAO1, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

PD-L1 expression and PD-1 checkpoint pathway in cancer	0,00490864656082138	EML4, MTOR, TRAF6, CSNK2A2	PIK3R1, AKT3, TICAM2, CD4, MAP3K3, PPP3CA, PPP3CB
Arrhythmogenic right ventricular cardiomyopathy	0,000278508564655881	ITGA2, ITGB4, ITGB8, CACNA1D, CACNG8, ATP2A1, DSP	ITGA7, ITGA8, ITGB3, ITGB7, SGCD, SGCA, SGCB, LAMA2, DES, TCF7L1, CACNA2D1, SLC8A2, SLC8A3, CDH2, GJA1
C-type lectin receptor signaling pathway	0,0100974286332395	ITPR3, CASP8, CBLB	LSP1, MRAS, RRAS, CALM2, CALM3, CALM1, PPP3CA, PPP3CB, MAPK10, FCER1G, CD209, RHOA, CCL22, PIK3R1, AKT3
Relaxin signaling pathway	0,00349196974031765	VEGFA, NOS2, MMP13, PLCB3	GNAO1, GNAI1, GNAI2, GNB1, GNB4, GNG2, GNG3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

VEGF signaling pathway	0,00852064036969162	VEGFA, JMJD7-PLA2G4B, PTK2	PPP3CA, PPP3CB, PIK3R1, RAC2, AKT3
Legionellosis	0,0086279496666866	CASP8, CLK1, HSF1, HSPA1B, CXCL8, CXCL2	C3, RAB1A, RAB1B, SEC22B, EEF1A1, EEF1A2, HSPA2, CD14
Tuberculosis	0,00456867862887744	CASP8, CREBBP, EP300, IRAK2, TRAF6, NOS2, IL1A, VDR, ATP6V0A2	BAX, AKT3, HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, CTSS, FCER1G, CD14, MAPK10, CD209, RHOA, LSP1, C3, MRC1, CALM2, CALM3, CALM1, RAB5B, RAB5C

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Ribosome	0,0119778858556761	RNA5S9	MRPS7, MRPS14, MRPS18C, RPS9, RPS16, RPS25, RPS27L, RPS29, MRPL20, MRPL27, MRPL33, MRPL35, RPL3, RPL5, RPL17-C18orf32, RPL35A, RPL41, RPL36AL, RPL36A-HNRNPH2
Pertussis	0,0240796502543234	C4BPA, TRAF6, CXCL8, NOD1, IL1A, NOS2	CALM2, CALM3, CALM1, RHOA, CFL1, CFL2, C1QA, C1QB, C1QC, C1R, C1S, C4A, C4B, C3, SERPING1, LY96, CD14, TICAM2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Aminoacyl-tRNA biosynthesis	0,0187739664866101	AARS2, NARS2, SARS2, CARS2, RARS2, FARS2	
GnRH secretion	0,0126704314531341	PLCB3, ITPR3, HCN3, CACNA1D, GNRH1	KCNN3, ARRB1, PIK3R1, AKT3
Phospholipase D signaling pathway	0,00801860199853201	INSR, SOS1, TSC1, TSC2, MTOR, JMJD7-PLA2G4B, DGKE, DGKH, CXCL8, PLCB3, CYTH2	PDGFC, PDGFD, KITLG, PDGFRA, KIT, GRB2, SHC4, MRAS, RRAS, RALB, PIK3R1, AKT3, RHOA, IGH, FCER1A, MS4A2, FCER1G, AGPAT1, F2R, LPAR1, LPAR4, LPAR6, PTGFR, ADCY5

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Vibrio cholerae infection	0,0465326708790209	ATP6V0A2, KDELR2, TJP2	ATP6V1D, ATP6V1E1, ATP6V1F, ATP6V1G2, ATP6V0E1, ATP6V0D2, ATP6AP1, ATP6V0C, SEC61B, SEC61G, PRKACA
Toxoplasmosis	0,00904311422661184	NOS2, TRAF6, HSPA1B, LAMB3, LAMC2, LDLR, CASP8	HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, TAB2, MAPK10, LY96, HSPA2, LAMA2, LAMA4, LAMC1, GNAI1, GNAI2, GNAO1, AKT3, CCR5

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Measles	0,00844389936465401	CD46, TRAF6, CSNK2A2, IL1A, EIF2AK2, HSPA1B, TP73, CASP8, CBLB, GSK3B	MSN, CD209, TLR7, TAB2, MAPK10, IFNAR1, RAB9A, RAB9B, HSPA2, BAX, CCND3, IL2RG, PIK3R1, AKT3, FCGR2B
Cardiac muscle contraction	0,00161686777069584	CACNA1D, CACNG8, ATP2A1, TNNI3, TNNT2	CACNA2D1, CASQ2, TPM2, UQCRH, UQCRHL, UQCRB, COX2, COX5A, COX6A1, COX6C, COX7A2, COX7C, SLC8A2, SLC8A3, SLC9A6, ATP1A2, ATP1B2
Prolactin signaling pathway	0,0195451068542435	SOS1, GSK3B	SHC4, PIK3R1, AKT3, GRB2, MAPK10, ESR1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Renin secretion	0,0209343765411945	CACNA1D, PLCB3, ITPR3	ADRB2, ADCYAP1, ADCYAP1R1, AQP1, PTGER2, ADCY5, PRKACA, KCNMA1, PDE1A, PDE1B, GNAI1, GNAI2, CALM2, CALM3, CALM1, PPP3CA, PPP3CB, NPR1, CTSB
Chemical carcinogenesis - receptor activation	0,0173411293066301	E2F1, CACNA1D, MTOR, VEGFA, VDR, SOS1, KPNA7, EPHX4, UGT1A6, UGT1A4, UGT1A7, CYP2B6	ARRB1, CHRNA3, PIK3R1, AKT3, FGF2, ADRB2, ADCY5, PRKACA, CREB3, GNAI1, GNAI2, RPS6KA2, ESR1, CCND3, GRB2, PGR, FGF7, FGF10

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Parathyroid hormone synthesis, secretion and action	0,0455089273339817	PLCB3, ITPR3, SP1, VDR, BRAF, LRP5, LRP6, MMP13	GNAI1, GNAI2, ADCY5, MAFB, RXRG, FGFR1, KL, EGR1, PTH1R, ARRB1, PRKACA, MEF2C, NACA, CREB3, CDKN1A, RHOA
Vasopressin-regulated water reabsorption	0,00631154388945661	DYNC1H1	PRKACA, ARHGDIB, CREB3, RAB11A, DYNLL1, DYNLL2, DCTN6, VAMP2, NSF, RAB5B, RAB5C
Insulin resistance	0,0243781798882403	INSR, MTOR, GSK3B, CPT1B, SLC27A5, IRS2, TRIB3, OGT, SLC2A1	RPS6KA2, PPP1R3C, PIK3R1, AKT3, MAPK10, PRKAG1, SLC27A3, SLC27A6, CD36, PYGM

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Hematopoietic cell lineage	0,0242898706184178	CSF3, IL11, IL1A, ITGA2	KITLG, FLT3LG, CSF1, CD34, FLT3, HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, KIT, IL7R, CD1A, CD1B, CD1D, CD1E, CD4, IGH, CD33, CSF1R, CD14, IL1R1, IL5RA, CD36, ITGB3, GP1BA
T cell receptor signaling pathway	0,0153035246588778	PAK6, DLG1, SOS1, GSK3B, CBLB	CD4, NCK1, GRB2, PAK3, RHOA, PPP3CA, PPP3CB, MAPK10, PIK3R1, AKT3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Osteoclast differentiation	0,0073067996496474	IL1A, TRAF2, TRAF6	CSF1, CSF1R, GRB2, PIK3R1, AKT3, IL1R1, TGFBR2, TAB2, MAPK10, BTK, LILRB1, LILRB5, LILRA1, FCGR2B, SIRPA, TYROBP, PPP3CA, PPP3CB, SPI1, MITF, CTSK, ACP5, ITGB3, IFNAR1
Oxytocin signaling pathway	0,0184048058604991	OXTR, JMJD7-PLA2G4B, PLCB3, CACNA1D, CACNG8, ITPR3, MYLK2	MEF2C, RYR3, CACNA2D1, CALM2, CALM3, CALM1, PPP3CA, PPP3CB, RGS2, CAMKK2, PRKAG1, CAMK1, NPR1, MYLK, MYL6, MYL9, RHOA

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Glutathione metabolism	0,0450068961652094	CHAC1, OPLAH, RRM2	GGT5, GSTA4, MGST1, GSTM5, MGST2, GSTM2, HPGDS, TXNDC12, GPX3, GPX1, GPX8, PRDX6, RRM2B
Necroptosis	0,0344981664722988	TRAF2, TRAF5, RBCK1, SPATA2L, CASP8, JMJD7-PLA2G4B, PGAM5, TRPM7, IL1A, FAF1, EIF2AK2	CYBB, SLC25A4, SLC25A5, VDAC3, PYGM, MAPK10, CHMP3, CHMP4A, IFNAR1, TICAM2, BAX

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

IL-17 signaling pathway	0,00425989927541625	CASP8, TRAF6, ANAPC5, TRAF5, TRAF2, GSK3B, CXCL2, CXCL8, CSF3, MMP3, MMP13	CCL11, TAB2, MAPK10, MAPK4, CCL2
Notch signaling pathway	0,00333002383075312	NOTCH1, DTX2, DTX3L, ADAM17, CREBBP, EP300, ATXN1	HEY1, KAT2B, CIR1
Other glycan degradation	0,00352638250567185		HEXB, MAN2B1, FUCA1, FUCA2, AGA
Antigen processing and presentation	0,0415930283149403	HSPA1B, TAPBP	HSPA2, B2M, LGMN, CTSB, HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

RNA degradation	0,038863244606365	EXOSC2, EXOSC3, EXOSC6, DIS3, ZCCHC7, CNOT2, CNOT4, PABPC1L, PAN3, DCP2, EDC4, XRN1	C1D, WDR61, CNOT8, BTG1, BTG2, PABPC5, NUDT16, LSM3, LSM6
Epstein-Barr virus infection	0,00895992682578154	TRAF6, EIF2AK2, CASP8, TAPBP, POLK, TRAF5, TRAF2, SKP2, E2F1, E2F2	HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, TAB2, IFNAR1, BAX, PSMC1, PSMC5, B2M, CIR1, CDKN1A, GADD45A, DDB2, PIK3R1, AKT3, MAPK10, VIM, IGH, BTK

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Apelin signaling pathway	0,042752161810187	PLCB3, ITPR3, NOS2, MYLK2, MTOR, HDAC4	ADCY5, PRKACA, APLNR, GNAI1, GNAI2, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, RYR3, CALM2, CALM3, CALM1, SLC8A2, SLC8A3, MYLK, RRAS, MRAS, AKT3, EGR1, PRKAG1, PLIN1, MEF2C, ACTA2, PLAT
Leishmaniasis	0,0438315511267751	TRAF6, IL1A, NOS2	TAB2, C3, IGH, CYBB, HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Base excision repair	0,00962469672336539	NEIL3, POLD1, POLE, POLE2, LIG3, PARP4	APEX1, POLE3
Propanoate metabolism	0,00524501866878486	ACACA, MLYCD	ACSS3, DLD, HADHA, ACAT1
Neutrophil extracellular trap formation	0,0480486192891434	PLCB3, MTOR, HDAC4, HDAC8	CYBB, RAC2, TLR7, VDAC3, SLC25A4, SLC25A5, IGH, PIK3R1, AKT3, C3, ITGB3, GP1BA, SELP, CTSG
Fatty acid elongation	0,0103838216669162		HADHB, HADHA, PPT1, ELOVL2, ELOVL4, ACOT2
Pyruvate metabolism	0,0209198909788256	ACACA	PDHB, DLD, ADH1B, ADH5, AKR1A1, ALDH9A1, MDH1, ACAT1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Bladder cancer	0,0126121403993962	BRAF, DAPK2, E2F1, E2F2, VEGFA, CXCL8	CDKN1A, MMP2
cGMP-PKG signaling pathway	0,0167116920556921	CACNA1D, GTF2IRD1, PLCB3, ITPR3, ATP2A1, MYLK2, INSR, IRS2	EDNRB, PPP3CA, PPP3CB, MEF2C, NPR1, KCNMA1, KCNMB1, SLC8A2, SLC8A3, ATP1A2, ATP1B2, ATP2B4, RGS2, PLN, RHOA, CALM2, CALM3, CALM1, MYLK, MYL9, CNGB1, ADCY5, PDE2A, PDE5A, ADORA3, GNAI1, GNAI2, AKT3, ADRB2, VDAC3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Tryptophan metabolism	0,0188932993911865		KMO, DLD, HADHA, ACAT1, TPH1, ALDH9A1, AOX1
GABAergic synapse	0,0143304859350119	GLS, GAD1, GABRE, CACNA1D	PRKACA, GABARAP, GABARAPL1, GABARAPL2, NSF, TRAK2, PLCL1, GNAI1, GNAI2, GNAO1, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, ADCY5
Complement and coagulation cascades	0,0112691500371479	PLAUR, CD46, C4BPA	F8, THBD, F2R, F13A1, PLG, SERPINA5, PROS1, PLAT, A2M, CFD, C3, C7, C1QA, C1QB, C1QC, C1R, C1S, MASP1, C4A, C4B, C3AR1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

cAMP signaling pathway	0,0187283747148662	HTR1D, OXTR, BRAF, CREBBP, EP300, AMH, TNNI3, ATP2A1, GRIN2B, GRIN2D, CACNA1D	ADCYAP1, VIP, ADRB2, PTGER2, ADCYAP1R1, VIPR2, NPR1, SST, NPY, PTGER3, GNAI1, GNAI2, ADCY5, CNGA3, CNGB1, CALM2, CALM3, CALM1, RRAS, MAPK10, RAP1A, RAC2, RHOA, PIK3R1, AKT3, PRKACA, CREB3, HHIP, F2R, MYL9, PLN, ATP1A2, ATP1B2, FXYD1, ATP2B4

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Insulin secretion	0,0139991292219702	SLC2A1, CACNA1D, PDX1, PLCB3, ITPR3, STX1A	ATP1A2, ATP1B2, KCNMA1, KCNMB1, KCNN3, ADCYAP1, ADCYAP1R1, ADCY5, PRKACA, CREB3, VAMP2
Cholinergic synapse	0,0151723942206175	PLCB3, ITPR3, CACNA1D	GNAI1, GNAI2, GNAO1, GNB1, GNB4, GNG2, GNG3, GNG5, GNG7, GNG10, GNG11, GNGT2, CHRNA3, ADCY5, PRKACA, CREB3, PIK3R1, AKT3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Wnt signaling pathway	0,0417658009690299	WNT2, WNT3, NOTUM, LRP5, LRP6, CSNK2A2, GSK3B, CHD8, CREBBP, EP300, TBL1XR1, INVS, PLCB3	WNT2B, WNT5B, SERPINF1, DKK1, SFRP1, SFRP2, SFRP4, RSPO2, RSPO3, FZD1, FZD4, FRAT1, TCF7L1, CBY1, SOX17, CTNND2, CCND3, PRKACA, SKP1, RBX1, ROR1, PRICKLE1, DAAM2, RHOA, RAC2, MAPK10, PPP3CA, PPP3CB
Long-term depression	0,046279027782611	BRAF, JMJD7-PLA2G4B, PLCB3, ITPR3, IGF1R	PPP2CB, GNAI1, GNAI2, GNAO1, IGF1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Regulation of lipolysis in adipocytes	0,0420495150444817	INSR, IRS2	ADRB2, ADCY5, PRKACA, NPR1, PLIN1, FABP4, PIK3R1, AKT3, PTGS1, PTGER3, NPY, GNAI1, GNAI2
Inflammatory mediator regulation of TRP channels	0,0280340160156978	JMJD7-PLA2G4B, PLCB3, ITPR3	HTR2A, CALM2, CALM3, CALM1, IL1R1, MAPK10, PIK3R1, PTGER2, ADCY5, PRKACA, IGF1
Salivary secretion	0,0220873563875305	PLCB3, ITPR3, PRH1, CST1	ADRB2, ADCY5, PRKACA, VAMP2, RYR3, CALM2, CALM3, CALM1, BST1, ATP1A2, ATP1B2, ATP2B4, KCNMA1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Influenza A	0,0226304468562464	EIF2AK2, CREBBP, EP300, IL1A, CXCL8, CASP8, NUP98, XPO1, KPNA7	TPSAB1, TPSB2, PLG, TLR7, CCL2, IFNAR1, HLA-DOA, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, PIK3R1, AKT3, CCND3, BAX, SLC25A4, SLC25A5, RAB11A
Hedgehog signaling pathway	0,0349437219295605	GSK3B, SMURF1, EFCAB7	PRKACA, SUFU, HHIP, BOC, ARRB1, SPOP
Chagas disease	0,0293757784557326	TRAF6, CXCL8, PPP2R2D, NOS2, PLCB3, CASP8	MAPK10, CCL2, PPP2CB, C3, C1QA, C1QB, C1QC, GNAI1, GNAI2, GNAO1, GNAL, PIK3R1, AKT3, TGFBR2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Epithelial cell signaling in Helicobacter pylori infection	0,0499444069685081	MET, ADAM17, CXCL8, NOD1, CXCL2, ATP6V0A2	JAM2, JAM3, MAPK10, PTPRZ1, ATP6V1D, ATP6V1E1, ATP6V1F, ATP6V1G2, ATP6V0E1, ATP6V0D2, ATP6AP1, ATP6V0C
GnRH signaling pathway	0,0291249214664448	GNRH1, PLCB3, ITPR3, CACNA1D, MAP3K4, JMJD7-PLA2G4B, SOS1	ADCY5, PRKACA, CALM2, CALM3, CALM1, MAP3K3, MAPK10, MMP2, GRB2, EGR1
Natural killer cell mediated cytotoxicity	0,0482150399305539	SH3BP2, SOS1, BRAF, ULBP1, ULBP2, ULBP3	RAC2, TYROBP, IGH, FCER1G, PIK3R1, SHC4, GRB2, CD48, PPP3CA, PPP3CB, IFNAR1

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

NF-kappa B signaling pathway	0,0489537057512915	CARD14, TRAF6, TRAF2, TRAF5, CXCL8, CXCL2, CSNK2A2	IGH, BTK, IL1R1, EDA2R, CD14, LY96, TICAM2, TAB2, TNFSF13B, GADD45A, VCAM1, CCL13, CCL19, CCL21, CXCL12
Pancreatic secretion	0,0354922286293095	PLCB3, ITPR3, ATP2A1	RAB11A, RAP1A, RHOA, CPA3, PNLIPRP2, BST1, ATP1A2, ATP1B2, KCNMA1, ATP2B4, ADCY5, SLC26A3
Gastric acid secretion	0,0384186498492776	PLCB3, ITPR3, MYLK2	CALM2, CALM3, CALM1, MYLK, ADCY5, SST, GNAI1, GNAI2, PRKACA, KCNK2

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Cholesterol metabolism	0,0488003725109282	LDLR	LPL, PLTP, CD36, SOAT1, APOC1, APOE, LIPA, VDAC3, NPC2
TNF signaling pathway	0,0481514263103719	BAG4, TRAF2, TRAF5, ITCH, PGAM5, CASP8, CXCL2, MMP3	TAB2, MAPK10, CREB3, CCL2, CSF1, VEGFC, VCAM1, PIK3R1, AKT3
Apoptosis	0,042752161810187	CASP8, TUBA8, SPTAN1, PARP4, DFFB, TRAF2, ITPR3, CASP2	BAX, TUBA1A, CTSB, CTSC, CTSK, CTSL, CTSO, CTSS, CTSZ, MAPK10, GADD45A, CSF2RB, PIK3R1, AKT3

Table 2. Pathway enrichment results that contain information of upregulated and downregulated genes found to be related to pathways (continued).

Sphingolipid signaling pathway	0,0483664203656586	TRAF2, PPP2R2D, PPP2R3B, PLCB3, ABCC1	ASAH1, BAX, MAPK10, AKT3, PPP2CB, PPP2R3C, FCER1A, MS4A2, FCER1G, PIK3R1, ADORA3, GNAI1, GNAI2, S1PR1, RAC2, S1PR3, RHOA
Cytosolic DNA-sensing pathway	0,0498339603519855	POLR3A, POLR3E	POLR3GL, POLR2F, POLR2L

9 CURRICULUM VITAE