Ayşegül Koçak

M.Sc. Thesis

AGU 2024

DEVELOPMENT OF A MACHINE LEARNING-BASED SYSTEM TO IDENTIFY DISEASE BIOMARKERS FROM HUMAN GUT MICROBIOTA

M.Sc. THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Ayşegül Koçak
April 2024

Updated: May 2023

DEVELOPMENT OF A MACHINE LEARNING-BASED SYSTEM TO IDENTIFY DISEASE BIOMARKERS FROM HUMAN GUT MICROBIOTA

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Ayşegül Koçak
April 2024

Updated: May 2023

SCIENTIFIC ETHICS COMPLIANCE

I hereby declare that all information in this document has been obtained in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name-Surname: Ayşegül Koçak

Signature :

REGULATORY COMPLIANCE

M.Sc. thesis titled **"DEVELOPMENT OF A MACHINE LEARNING-BASED SYSTEM TO IDENTIFY DISEASE BIOMARKERS FROM HUMAN GUT MICROBIOTA"** has been prepared in accordance with the Thesis Writing Guidelines of the Abdullah Gül University, Graduate School of Engineering & Science.

Prepared By	Advisor	Co Advisor
Ayşegül Koçak	Assoc. Prof. Burcu Bakır Güngör	Prof. Dr. Malik YOUSEF

Head of the Electrical and Computer Engineering Graduate Program
Assist. Prof. Samet GÜLER

Updated: May 2023

ACCEPTANCE AND APPROVAL

M.Sc. thesis titled **"DEVELOPMENT OF A MACHINE LEARNING-BASED SYSTEM TO IDENTIFY DISEASE BIOMARKERS FROM HUMAN GUT MICROBIOTA"** and prepared by Ayşegül Koçak has been accepted by the jury in the Electrical and Computer Engineering Graduate Program at Abdullah Gül University, Graduate School of Engineering & Science.

…….. /…….. / ……..

JURY:

Advisor : Assoc. Prof. Burcu Bakır Güngör

Member : Asst. Prof. Tayyip ÖZCAN

Member : Assoc. Prof. Rifat KURBAN

APPROVAL:

The acceptance of this M.Sc. thesis has been approved by the decision of the Abdullah Gül University, Graduate School of Engineering & Science, Executive Board dated ….. /….. / …….. and numbered ……..……. .

…….. /…….. / ……..

Graduate School Dean
Prof. İrfan ALAN

Updated: May 2023

ABSTRACT

DEVELOPMENT OF A MACHINE LEARNING- BASED SYSTEM TO IDENTIFY DISEASE BIOMARKERS FROM HUMAN GUT MICROBIOTA

Ayşegül KOÇAK
MSc. in Electrical and Computer Engineering
Advisor: Assoc. Prof. Burcu BAKIR GÜNGÖR
Co-advisor : Prof. Dr. Malik YOUSEF
April 2024

The human gut microbiota consists of a diverse ecosystem of organisms, encompasses billions of species. Recently developed next-generation sequencing methods have enabled researchers to examine the microbiota in greater detail, leading to new insights into its functions and dysfunctions.

This study aims to identify metagenomic biomarkers (Microorganism-Enzyme Pairs) for colorectal cancer (CRC). The tool that we used allows for the analysis of microorganisms and enzymes within the gut microbiota. It achieves this by initially clustering enzymes based on their correlations with species and subsequently utilizing these clustering results to evaluate the ability of groups to differentiate between patient and healthy cohorts.

By integrating species and enzymes, it is possible to identify pathogen microorganisms and enzyme clusters, that have the potential to distinguish cases (individuals with CRC) from controls (healthy individuals).

The identified enzyme clusters and associated species could potentially act as biomarkers for colorectal cancer (CRC), enabling early diagnosis and more effective treatment. This approach holds promise for further exploration of the gut microbiota and its importance in human health and illness.

Keywords: Bioinformatics, Machine Learning, Colorectal Cancer Diagnosis

ÖZET

İNSAN BAĞIRSAK MİKROBİYOTASINDAN HASTALIK BİYOBELİRTECLERİNİN TESPİTİ İCIN MAKİNE ÖGRENMESİ TEMELLİ SİSTEM GELİŞTİRİLMESİ

Ayşegül KOÇAK
Elektrik ve Bilgisayar Mühendisliği Bölümü Yüksek Lisans
Tez Danışmanı: Doç. Dr. Öğr. Üyesi Burcu BAKIR-GÜNGÖR
İkinci Tez Danışmanı: Prof. Dr. Malik YOUSEF
Nisan-2024

İnsan bağırsak mikrobiyotası, milyarlarca türü kapsayan karmaşık bir organizma topluluğudur. Son zamanlarda geliştirilen yeni nesil dizileme yöntemleri, araştırmacıların mikrobiyotayı daha detaylı bir şekilde incelemesine olanak tanımış ve bu sayede işlevleri ve disfonksiyonları hakkında yeni bilgiler elde edilmiştir.

Bu çalışma, kolorektal kanser (CRC) için metagenomik biyobelirteçler (Mikroorganizma-Enzim Çiftleri) tanımlamayı amaçlamaktadır. Kullandığımız araç, bağırsak mikrobiyotası içindeki mikroorganizmaların ve enzimlerin analizine olanak tanır. Bunu, öncelikle enzimleri türlerle olan korelasyonlarına göre kümeleyerek ve daha sonra bu kümeleme sonuçlarını hasta ve sağlıklı kohortlar arasında ayırt etme yeteneği gruplarını değerlendirecek şekilde gerçekleştirir.

Türler ve enzimler entegre edilerek, vakaları (CRC'li bireyler) kontrollerden (sağlıklı bireyler) ayırma potansiyeline sahip patojen mikroorganizmaları ve enzim kümeleri belirlenebilir.

Tanımlanan enzim kümeleri ve ilişkili türler, CRC için potansiyel biyobelirteçler olarak hizmet edebilir, erken tanı ve daha etkili tedavi sağlayabilir. Bu yaklaşım, bağırsak mikrobiyotasının ve insan sağlığı ile hastalıkları üzerindeki rolünün daha fazla keşfedilmesi için umut vaat etmektedir.

Anahtar kelimeler: Biyoinformatik, Makine Öğrenimi, Kolorektal Kanser Teşhisi

Acknowledgements

I would like to express my profound gratitude to my advisor, Dr. Burcu BAKIR-GÜNGÖR, for her invaluable support and guidance throughout my thesis. Additionally, my deepest appreciation goes to my co-advisor, Dr. Malik YOUSEF, for imparting his extensive knowledge, as well as for his patience and unwavering support.

I would also like to extend my cordial gratitude to Dr. Tayyip ÖZCAN and Dr. Rifat KURBAN for their participation on my dissertation defense committee.

I am also immensely grateful to my family. I extend heartfelt thanks to my mother, Aynur KOÇAK, and my father, Reşat KOÇAK, for their boundless patience, tolerance, and love. I owe a great deal of thanks to my sisters, Hilal KOÇAK and Arife KOÇAK, for always listening to my concerns and for their wise counsel.

Lastly, I would like to offer my warmest thanks to my dearest friends, Nurten BULUT, and to Hüseyin AKKAŞ for their constant support, motivation, and inspiration in every decision I have made in my life. They have consistently been there for me, especially during challenging times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CRC	Colorectal Cancer
EC	Enzyme Commission
NGS	Next-Generation Sequencing
RCE	Recursive Cluster Elimination
RFE	Recursive Feature Elimination

To my family

Chapter 1

Introduction

The burgeoning field of bioinformatics has heralded a paradigm shift in understanding complex diseases, notably colorectal cancer (CRC), through the lens of the human gut microbiota. Leveraging cutting-edge next-generation sequencing technologies, this study embarked on an ambitious quest to elucidate the intricate interactions between microbial species and enzymes within the gut microbiome, with a specific focus on identifying potential metagenomic biomarkers for CRC. By integrating taxonomic, functional, and strain-level profiling of diverse microbial communities, we identified distinctive microorganism-enzyme pairs that exhibit a significant association with CRC.

Utilizing a novel analytical framework that incorporates sophisticated grouping, ranking, and modeling-based approaches, our research has successfully pinpointed several microorganism-enzyme pairs. These pairs not only delineate a unique microbial signature of the CRC gut microbiome but also highlight the potential of these microbial constituents as non-invasive biomarkers for early CRC detection and prognosis. The identified biomarkers demonstrate a remarkable ability to distinguish between CRC patients and healthy controls, underscoring their diagnostic and therapeutic relevance.

The implications of our findings extend beyond the realm of diagnostics; they pave the way for exploring novel therapeutic targets within the gut microbiota, offering a beacon of hope for precision medicine in CRC. Moreover, this study sets the stage for future investigations aimed at validating these biomarkers in larger cohort studies and exploring their mechanistic roles in CRC pathogenesis. In essence, our research underscores the pivotal role of the gut microbiome in CRC and showcases the potential of bioinformatics approaches in unraveling the complexities of human diseases, thereby opening new avenues for personalized medicine.

Chapter 2

Literature Review

The microbiome of the human body is a diverse collection of bacteria, their enzymes, and the compounds they provide or modify [1]. The human gut microbiota, which consists of a diverse collection of microbes, is critical to human health and occupies an important role in medical conditions. In recent studies, the core contribution of commensal bacteria to both human physiological functions and diseases has been underscored. The microbiota of the human gastrointestinal tract is essential for maintaining gut homeostasis and general health. Dysbiosis, or modifications to the structure and function of the microbiota of the gastrointestinal tract, has been related to a number of illnesses and health conditions, emphasizing the importance of gut microbiological stability. The growing body of research on disease-related microbiota suggests that a reduction in microbial diversity is a common feature observed in dysbiosis [2]. An imbalance in the gut microbiota, known as dysbiosis, has been associated with inflammatory diseases, neurodevelopmental disorders, aging-related cognitive impairment, environmental pollutant exposure, and various systemic diseases [3].

Gut microbiota dysbiosis defined by an imbalance of microbial composition, has been implicated in driving the development of Colorectal cancer (CRC) [4]. Colorectal cancer (CRC) is a disease where dysbiosis plays a significant role in its development. Colorectal cancer (CRC) is a complicated disease marked by the uncontrolled proliferation of abnormal cells in the colon or rectum, influenced by genetic predisposition, lifestyle, and environmental factors. Colorectal cancer (CRC) is a major health issue, with a 4%–5% probability of occurrence, influenced by factors including age, history of chronic diseases, and lifestyle choices [5]. Colorectal cancer (CRC), causing nearly 500,000 deaths annually and ranking as the primary global causality of death, presents a substantial healthcare challenge [6]. While some research has highlighted the importance of the gut microbiota in the pathogenesis of colorectal cancer (CRC), this field of research is still relatively fresh. Despite the use of diverse classification standards and methodologies in past inquiries, the specific microbial strains

that significantly play a role in the development and advancement of CRC remain unidentified [4]. CRC poses a significant risk to human health and ranks as the third most prevalent type of cancer [4]. For individuals with colorectal cancer (CRC), it's essential to understand the composition of the human gut microbiome, encompassing the collective genetic material of the microorganisms in the gut and their potential functions [2]. Genetic research has been pivotal in understanding the hereditary components of CRC, with studies focusing on cancer susceptibility gene mutations, including Lynch syndrome (LS) and genetic colorectal cancer and polyposis susceptibility genes [9,10,11]. Molecular research has elucidated the intricate molecular pathways involved in CRC, including the roles of human condensins, YAP/TAZ signaling, and the sensitivities and dependencies of specific genetic mutations in CRC cell lines [12,13,14]. Clinical research has focused on different areas of CRC care, including the assessment of modifiable factors for cancer-specific survival, the characterization of serious intestinal congestion due to primary colorectal carcinoma, and the accuracy of algorithms for identifying cancer-related adverse events [15,16,17]. Research has also explored the impact of metabolic dysfunction, oxidative balance, and preoperative intentional weight loss on CRC, shedding light on the broader metabolic and lifestyle factors that affect the progression and prognosis of CRC. Furthermore, investigations into the feasibility of microwave ablation for colorectal liver metastases and the potential therapeutic role of gut metabolites have expanded the horizons of CRC treatment modalities [18,19]. Extensive research on CRC has offered a thorough insight into the disease, spanning genetics, molecular pathways, clinical management, and epidemiological aspects, significantly advancing our knowledge of CRC and informing future methods for preventing, diagnosing, and managing this intricate disease.

Advances in next-generation sequencing (NGS) technology now enable the production of enormous sums of reads in just one run. Metagenomic NGS technologies have been used to identify bacteria in the gastrointestinal system. This allows for a comprehensive examination of the genetic content in a sample, unveiling the characteristics of microbial communities in terms of biology and behavior. As a result, investigating the metagenomics of the gastrointestinal system provides vital insights into how the microbial community living in the human gut effects the functioning of the human body and the advancement of disorders [18]. The use of NGS technologies has significantly improved Our knowledge of the interaction between colorectal cancer (CRC) and the gut microbiota. NGS has permitted a complete analysis of the gut

microbiome, resulting in the identification of particular bacterial species associated with CRC. This has given us vital insights into the molecular pathways that underlie the emergence of colorectal cancer. NGS technologies have empowered a more thorough examination of the gut microbiota, offering a deeper understanding of its abundance, variety, and distinct microbial categories [19]. It has been confirmed that NGS is the preferred approach for examining the composition of gut microbiota, thanks to its superior resolution and capability to capture the diversity present in microbial communities [20]. These advanced sequencing methods have allowed for identifying and characterizing microbial enzymes engaged in diverse metabolic processes, illuminating their probable influence on human health [21].

The microbial community in the human gut forms a intricate ecosystem that harbors a diverse array of enzymes crucial for human health and disease. These microbial enzymes play a vital purpose in medication metabolism, neurotransmitter synthesis, and the reactivation of compounds like as estrogens. For instance, the gut microbiota has been found to contain decarboxylases that produce the neurotransmitter tryptamine. Additionally, the gut microbiota's β-glucuronidases have been shown to reactivate estrogens, contributing to the estrobolome that reactivates these compounds. Furthermore, the gut microbiota's involvement in drug metabolism, particularly in the metabolism of levodopa, has been identified, highlighting its impact on therapeutic interventions. Furthermore, the gut microbiota's role in cancer causation through "Phase IV Metabolism" of exogenous and internal chemicals has been defined, highlighting potential implications for disease development [22]. These references provide thorough understandings into the several purposes of microbial enzymes in the human microbiota of the gastrointestinal tract, encompassing drug metabolism, neurotransmitter production, and their impact on disease etiology.

Enzymes play diverse and critical roles in CRC, encompassing aspects such as biomarker potential, molecular pathways, metabolism, and drug response. Clarifying the intricate processes underlying the onset and progression of CRC requires an understanding of the diverse functions that enzymes play in the disease. Enzymes' involvement in colorectal cancer (CRC) is multifaceted and encompasses various aspects of cancer development and progression. There is evidence linking CRC to enzymes involved in signaling, restoration of DNA, and metabolism. For example, current research has examined the role of several enzymes and hematopoietic growth factors (FGFs) in both the detection and outlook of colorectal cancer [23]. Numerous research have

highlighted the complex relationship between microbial enzymes, colorectal cancer, and gut microbiota. For instance, colorectal cancer has been related to the gut microbiome, with distinct microbial fingerprints discovered in stool samples of colorectal cancer patients [24]. Furthermore, colorectal cancer has been related to the microbiota's manufacture of microbe-derived compounds and control over immune function [25]. Furthermore, it has been discovered that the colorectal cancer milieu has an enrichment of anticipated virulence-associated genes, which is particularly dependent on the genomes of particular bacteria. This finding raises the probability that microbial enzymes take part in the tumor microenvironment. [26].

Furthermore, it has been revealed that the enzyme cyclooxygenase-2 (COX-2) contributes to the reproduction and activity of colorectal tumor cells, highlighting its role in the progression of CRC [27]. Furthermore, genetic variants in enzymes that participate in the metabolism of folate, such as methylenetetrahydrofolate reductase, have drawn attention to epidemiological research on CRC, indicating a potential link between the development of CRC and polymorphisms in these enzymes [28]. Moreover, Uba2, an essential part of the SUMO-activating enzyme, has been shown to play significant roles in the advancement of several malignancies; its function in colorectal cancer, however, is still unknown [29]. Additionally, enzymes involved in amino acid metabolism of tumor-associated macrophages have been speculated to regulate CRC through cellular demise [30]. Furthermore, GSTM1 has been linked to the development and evolution of several deadly cancers, including colorectal cancer [31]. The role of cellular process-related enzymes, such as CD73, in encouraging the expansion of human colorectal tumor cells through both enzyme and non-enzyme activities has been studied [32]. Additionally, aberrant DNA methylation affecting the carcinogenesis of CRC has been linked to methyl-metabolizing enzyme and epigenetic regulator genetic variations [33]. Additionally, in both colorectal tumors and healthy colorectal tissue, the characteristics, variability, and activity of the enzymes were examined, as well as their capacity to inactivate paclitaxel (taxol) [34].

Colorectal cancer is influenced by a complex interplay of diverse microbial species. The gut microbiota has emerged as a significant component in the genesis and progression of colorectal cancer. Studies have identified specific microbial signatures associated with colorectal cancer, illuminating the complex connection between disease etiology and the microbiome [35]. Fusobacterium nucleatum, in particular, has garnered significant attention due to its prevalence in colorectal carcinoma and its possible contribution to the

development of cancer. The existence of F. nucleatum in the tissues of colorectal cancer has been linked to disease progression, highlighting the direct functional link between specific microbial components and cancer development [36]. Moreover, the stability and dysbiosis in the gut flora have been implicated in colorectal cancer, with alterations in microbial composition potentially triggering disease development [37]. The correlation between intestinal microbiota and carcinoma of the colorectal extends beyond its composition, with studies highlighting the role of microbial metabolites, immune modulation, and microbial transcriptome adaptations in the tumor microenvironment. Furthermore, the microbiome's stability, dysbiosis and response to antimicrobial interventions have been examined with consideration for carcinoma of the colorectal, shedding light on potential therapeutic avenues [50, 51]. The microbiome's influence on colorectal cancer is multifaceted, involving relationships with the immune system of the host, dietary factors, and the tumor microenvironment [35, 52, 53]. Furthermore, certain clinical outcomes have been linked to the gut flora, such as survival in late-stage colorectal cancer patients, emphasizing its potential as an anticipatory indicator [42]. The field of study on the relation between colorectal cancer and the microbiota of the intestinal tract is quickly developing, and an increasing amount of data is showing how intricately host physiology, microbial species, and disease progression are all related. Understanding the role of gut microbiota in colorectal cancer is essential for developing targeted therapies and individualized treatment plans.

Chapter 3

Materials and Methods

3.1 Data sets

Beghini and the research team compiled 1,262 metagenomic samples, including 600 control samples and 662 from colorectal cancer (CRC) cases, from nine separate datasets across eight countries [43]. For their study, they retrieved the original microbiome DNA data for each sample from the appropriate project websites. They then utilized MetaPhlAn 3 and HUMAnN 3 for conducting both taxonomic and functional analyses on all the subgroups within each dataset.

In their research methodology, following the approach described by Thomas and colleagues in their 2019 study [44], the team calculated the relative abundances for each of the samples within the datasets, categorizing them at the i) species level, ii) enzyme categories (according to the Enzyme Commission, EC), and iii) pathway categories. In this proposed project, which focuses on metagenomic data related to colorectal cancer (CRC) collected from the previously mentioned eight countries, the researchers plan to work with relative abundance counts for the 1,262 samples as follows: i) 983 unique species, ii) 2,894 unique enzymes, and iii) 550 unique pathways.

Our research utilized a metagenomic dataset compiled by Beghini et al., consisting of 1,262 metagenomic samples, including 600 from control individuals and 662 from patients with colorectal cancer (CRC), sourced from nine different datasets across eight countries. This dataset provided an extensive overview of the microbial composition linked to CRC. We processed the original microbiome DNA data for each sample using MetaPhlAn 3 for taxonomic identification and HUMAnN 3 for functional analysis, following the method outlined by Thomas et al. Through this process, we accurately identified the relative abundances at the species level and for enzyme categories, categorized by Enzyme Commission (EC) numbers.

In our analysis, we concentrated on the species and enzyme components of the metagenomic data to investigate the distinctive microbial and enzymatic profiles associated with CRC. The dataset revealed relative abundance values for 983 unique microbial species and 2,894 distinct enzymes, offering a detailed view into the microbial ecology and functional capabilities within the CRC microbiome. Our goal was to shed light on the roles of specific species and enzymes in the development of CRC, aiming to identify new biomarkers and therapeutic targets amidst the complex metagenomic landscape.

3.2 Preprocessing

In advancing the understanding of the microbiota's role in colorectal cancer (CRC), comprehensive datasets encapsulating microbial taxonomic compositions and their functional attributes are indispensable. The datasets, derived from the MetaPhlAn 3 taxonomic profiles and HUMAnN 3 functional profiles, have undergone a sequence of preprocessing steps, pivotal for the refinement of data quality and analytical readiness [45].

Table 3.2.1 Enzymes dataset after preprocessing

Row ID	class	1.1.1.122	1.1.1.125	1.1.1.127	1.1.1.130	1.1.1.131	1.1.1.132	
Row0	pos	6.38395	0.48866	24.1456	9.88266	14.1598	4.47724	
Row1	pos	0	3.50633	40.925	12.4527	10.8095	4.77108	
Row2	pos	1.31362	0	0	0	17.5954	2.55107	
Row3	pos	37.3938	0.216005	6.08372	2.6948	32.466	2.73441	
Row4	pos	3.0982	14.0057	72.7411	39.6646	1.30704	13.2858	…
Row5	pos	18.586	1.42644	14.4763	1.65687	23.3021	1.23378	
Row6	pos	14.6723	0	12.3031	5.84978	18.7827	3.11644	
Row7	pos	1.69378	2.32216	53.1471	10.9493	8.53768	8.3683	
Row8	pos	2.64107	0	4.84045	2.5696	10.8627	0.173282	
Row9	pos	6.89442	1.35666	49.9316	18.9629	9.31839	14.1257	
			…					

Table 3.2.2 Species dataset after preprocessing

Row ID	class		Collinsella aerofaciens	Alistipes putredinis	Streptococcus anginosus_group	Blautia_obeum	
Row0	pos		0.02906	0.00095	0.01205	0.04658	
Row1	pos		0.25828	1.92414	0	0.55916	
Row2	pos		0.01288	4.80856	0.00389	0	
Row3	pos		0.00086	9.35177	0	0.30297	
Row4	neg	…	0.94993	2.91432	0.00639	0.64185	
Row5	neg		0.00547	2.0925	0.00184	0.28482	…
Row6	pos		1.11336	0.43826	0	0.14736	
Row7	neg		0.14341	0.62694	0	0.12371	
Row8	neg		1.17221	2.09242	0.02602	1.29861	
Row9	pos		0.94884	1.24318	1.02093	0.49009	
			...				

3.3 Machine Learning

Machine Learning is a subset of artificial intelligence where algorithms and statistical models are employed to enable machines to improve at tasks with experience. Machine learning models learn from data, identifying patterns and making decisions. This method, which includes supervised learning (with labeled data), unsupervised learning (with unlabeled data), and reinforcement learning (with reward feedback), reduces the requirement for explicit programming to make judgments or predictions.

3.3.1 Feature Selection Methods

Feature selection in machine learning is the process of identifying and selecting a subset of relevant features (variables, predictors) for use in model construction. It's about choosing the most important features that contribute to the predictive power of the model while removing redundant, irrelevant, or noisy data. The goal is to improve the model's performance, reduce computational cost, and enhance interpretability. The importance of feature selection in machine learning lies in its ability to enhance model performance, efficiency, and interpretability. By identifying and retaining only the most relevant features, it reduces the risk of overfitting, as the model is less likely to learn from noise or irrelevant data. This process leads to simpler models that are faster to train, require less

computational resources, and are easier to understand and explain. In essence, feature selection helps in building robust, efficient, and interpretable predictive models that can perform better on unseen data, making it a crucial step in the machine learning pipeline.

3.3.1.1 Conditional Mutual Information Maximization (CMIM)

CMIM is used to evaluate the worth of a feature by measuring the mutual information between the feature and the class label, given the values of other features. It aims to find the set of features that are most informative about the outcome, helping to reduce dimensionality and improve model performance.

3.3.1.2 Fast Correlation-Based Filter (FCBF)

This algorithm identifies and eliminates features that are highly correlated with each other but have low correlation with the target variable. It aims to reduce redundancy and irrelevance in the feature set, leading to simpler and more efficient models.

3.3.1.3 Information Gain (IG)

Used in decision trees and feature selection, Information Gain measures how much information a feature gives about the class. It's calculated by the difference in entropy before and after the dataset is split on that feature, guiding the selection of the most informative features.

3.3.1.4 Minimum Redundancy Maximum Relevance (MRMR)

MRMR selects features that have the highest relevance with the target variable and the least redundancy with each other, balancing between feature relevance and redundancy, to enhance model performance with a compact feature set.

3.3.1.5 SelectKBest (SKB)

This method selects the top-k features based on a statistical test (like chi-squared, f-test, etc.). It's straightforward and effective for preliminary feature selection, where only the strongest features are retained.

3.3.1.6 XGBoost Feature Selection (XGB)

In XGBoost, feature selection is performed through the model itself by evaluating feature importances based on the contribution of each feature to the model's performance, typically using metrics like "gain" or "cover".

3.3.2 Classification Methods

Classification methods in machine learning are algorithms that are used to categorize or classify data points into predefined classes or categories. These methods analyze the input data and use it to determine the class label for each data point.

3.3.2.1 Adaboost

Adaboost is a boosting algorithm that builds a strong classifier by combining several weak learners, typically decision trees. It adjusts the weights of misclassified instances so they are more likely to be correctly classified in the next iteration, iteratively improving model accuracy.

3.3.2.2 Decision Tree

Decision Tree is a decision support tool that uses a tree-like model of decisions and their possible consequences. It splits the dataset into branches based on feature values, leading to decision nodes that represent the selected feature and leaf nodes that represent the outcome.

3.3.2.3 LogitBoost

LogitBoost is a boosting algorithm specifically designed for improving classification performance, particularly for binary classification. It modifies the weights of instances in the dataset, focusing on those that are hard to classify in previous rounds.

3.3.2.4 Random Forest

Random Forest is an ensemble method that creates a forest of decision trees, usually trained with variations in the data or features. It improves over a single decision tree by reducing overfitting and providing more accurate and stable predictions.

3.3.2.5 Optimized Support Vector Machine (SVM_opt)

Optimized Support Vector Machine refers to SVMs that have been optimized, often with hyperparameter tuning (like selecting the best kernel, regularization parameter, etc.) to achieve the best performance on a given dataset.

3.3.2.6 Stack_Logitboost_Kmenas

This model stacks LogitBoost and K-Means for better predictive performance. It leverages the strengths of both clustering (K-Means) and classification (LogitBoost) in a unified model, often improving accuracy and robustness.

3.3.2.7 Stack_SVM_Kmeans

Stack_SVM_Kmeans combines SVM and K-Means in a stacked model, utilizing SVM for classification and K-Means for clustering within the dataset. This approach aims to improve the model's ability to classify by leveraging the internal structure of the data.

3.3.2.8 eXtreme Gradient Boosting (XGBoost)

XGBoost is a highly efficient and scalable version of gradient boosting, XGBoost is known for its speed and performance. It uses a gradient boosting framework and is often used in data mining competitions for its ability to handle a wide variety of data types and tasks.

3.3.4 Feature Ranking and Selection Algorithms

3.3.4.1 Support Vector Machine - Recursive Feature Elimination (SVM-RFE)

Support Vector Machine Recursive Feature Elimination (SVM-RFE) is a feature selection technique used to identify the most significant features for building a predictive model. This method is particularly useful in cases where you have a large number of features and need to reduce the dimensionality of your data.

3.3.4.2 Support Vector Machine - Recursive Cluster Elimination (SVM-RCE)

Support Vector Machine Recursive Cluster Elimination (SVM-RCE) is a feature selection approach. This method integrates recursive feature elimination with support vector machines to iteratively filter out feature subsets with high accuracy rates, particularly suitable for large datasets. The authors demonstrated the effectiveness of SVM-RCE in feature selection tasks using Matlab. SVM-RCE aims to enhance machine learning model performance by systematically removing less relevant features, thereby enhancing prediction accuracy. Support vector machines and recursive cluster elimination are combined in SVM-RCE, which offers a reliable technique for choosing useful features for the creation of predictive models [59, 60]. This feature selection technique shows promise for various applications, especially in scenarios with high-dimensional datasets that necessitate efficient dimensionality reduction without compromising predictive performance. SVM-RCE's capability to identify and retain relevant features can significantly impact the accuracy and generalization abilities of machine learning models, making it a valuable tool in data analysis and predictive modeling.

3.3.5 Statistical Measures and Tests

Pearson Correlation Coefficient: It is a metric used in statistics to represent how much a linear relationship exists between two variables. It's a value between -1 and 1, where:

+1 indicates a perfect positive linear relationship (as one variable increases, the other variable increases at a consistent rate).

-1 indicates a perfect negative linear relationship (as one variable increases, the other decreases at a consistent rate).

0 indicates no linear relationship between the variables.

The formula for the Pearson Correlation Coefficient (denoted as r) for two variables X and Y is:

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}} \quad (3.3.5.1)$$

Where:

X and Y are the individual sample points.

$\bar{X}$ and $\bar{Y}$ are the mean values of the X and Y variables.

The coefficient's magnitude indicates the strength of the correlation, while its sign indicates the direction (positive or negative correlation). It's widely used in scientific research to examine the relationship between variables, particularly when figuring out whether a shift in one variable could indicate a shift in another. It only measures linear relationships and can be affected by outliers. It also doesn't imply causation, only association.

T – test: A statistical test called a t-test is used to compare two groups' means and determine whether there is a significant difference between them. It's a common tool in statistics and research for hypothesis testing, especially in situations where the sample sizes are small and the population variance is unknown**.** The basic idea behind the t-test is to assess whether the means of two groups are statistically different from each other. This involves calculating a t-value, which then is compared to a value from a t-distribution to determine the probability (p-value) of the observed difference occurring by chance. A low p-value (typically <0.05) indicates that the difference in means is statistically significant, suggesting it's unlikely to have happened due to random chance.

The formula and exact method of calculation can vary depending on the type of t-test being conducted, but generally, the t-test takes into account the means and standard deviations of the groups, as well as the number of subjects in each group.

P- Value : In hypothesis testing, the p-value is the probability of observing test results at least as extreme as the actual observed results, under the assumption that the null hypothesis is correct. A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, leading to its rejection. Conversely, a large p-value suggests weak evidence against the null hypothesis, so it's not rejected. The p-value is calculated depending on the statistical test being used (e.g., Z-test, T-test, chi-squared test). It's determined from the test statistic, which measures the degree of agreement between the sample data and the null hypothesis. P-values are used in many fields to determine the statistical significance of results from experiments and studies, influencing decision-making and conclusions.

3.3.6 Performance Evaluation Metrics

To assess the performance of each model, we determined various statistical metrics including sensitivity, specificity, and accuracy. These statistics were computed using the following definitions, where TP represents true positives, FP denotes false positives, TN signifies true negatives, and FN stands for false negatives:

$$\text{Sensitivity (SE, Recall)} = TP/(TP + FN) \quad \textbf{(3.3.6.1)}$$

$$\text{Specificity (SP)} = TN/(TN + FP) \quad \textbf{(3.3.6.2)}$$

$$\text{Accuracy (ACC)} = (TP + TN)/(TP + TN + FP + FN) \quad \textbf{(3.3.6.3)}$$

In addition, we computed the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC), evaluating the probability that, if selected at random, the classifier will score a positive instance higher than a negative one. The performance metrics presented in this research are based on the mean of 100 Monte Carlo Cross-Validation (MCCV) iterations. MCCV involves randomly partitioning the data into training and testing sets, typically without replacement, and this selection process is conducted multiple times to produce varied training and testing sets. In our case, we used 90% of the data for training and 10% for testing. We also addressed the issue of imbalanced data sets, which can skew the classifier's performance towards the more prevalent class. To mitigate this, we implemented an under-sampling strategy that equalizes the number of instances across classes, adjusting the majority class to match the minority class count, effectively reducing data subset size bias. Specifically, we employed an under-sampling ratio of 1:2.

3.4 Proposed Method

In this segment, miRcorrNet tool employed in this study, is to be introduced. Figure 3.4.1 illustrates the fundamental principles and main elements of these tools. For a contemporary summary and analysis of these methods, the 2021 review by Yousef, Kumar, and Bakir-Gungor [48] serves as a useful reference. The comprehensive approach consists of two principal components: the Grouping function, symbolized by G(), and the Ranking function, denoted by R(). The detailed functions and uses of these components are depicted in Figure 3.4.1.

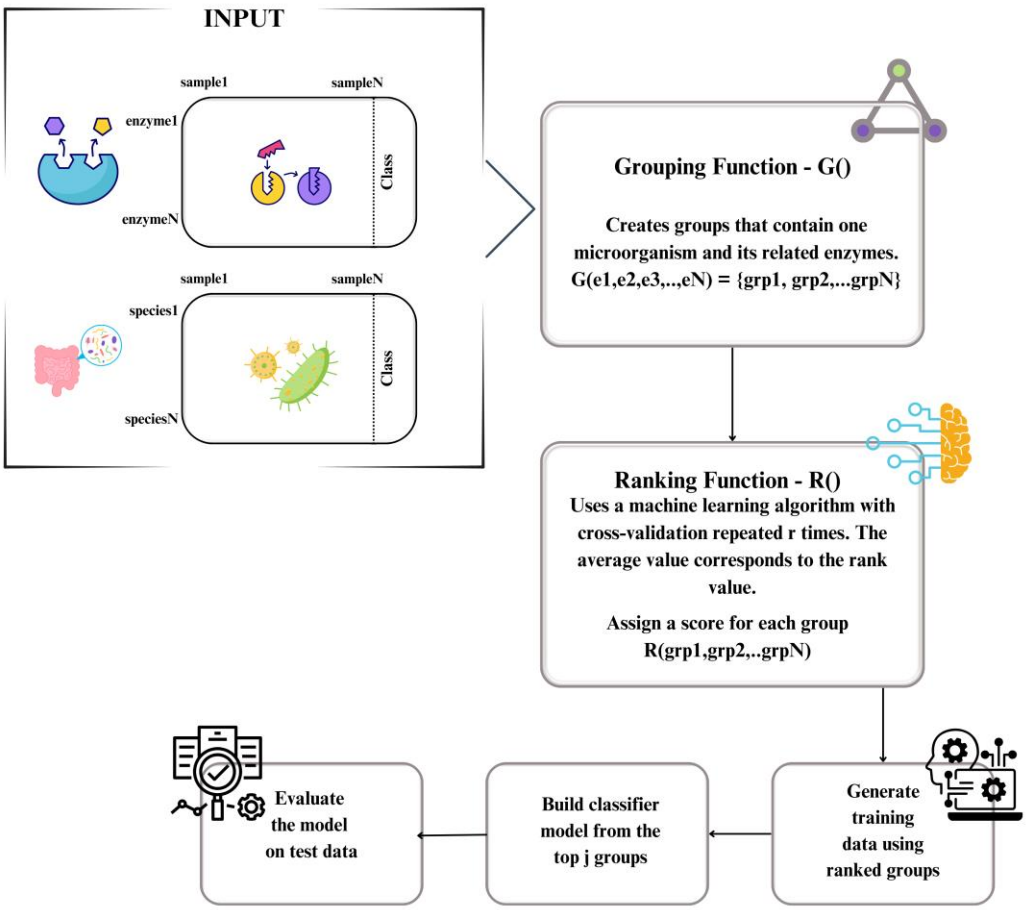

Figure 3.4.1 Basic workflow for categorizing data by utilizing the grouping function G() and employing the R() function to evaluate the groups.

In miRcorrNet, the G() function is designed to generate two distinct types of group outcomes: one includes clusters of enzymes that are associated with specific species and the other features clusters of species that are linked with particular enzymes. For instance, concerning Colorectal Cancer, the results produced by the G() function, as illustrated in Table 3.4.1 and Table 3.4.2, encompass both enzyme groups that are connected to specific species and groups of species categorized according to their association with enzymes.

Table 3.4.1 Microorganism groups based on their correlations with enzyme levels

Enzyme	Associated Microorganisms
1.1.1.136	Bacteroides_fragilis
1.1.1.154	Escherichia_coli
1.1.1.361	Eisenbergiella_tayi,Fusobacterium_necrophorum
1.1.1.40	Prevotella_copri
1.1.1.408	Prevotella_nigrescens,Fusobacterium_nucleatum
1.1.2.4	Slackia_exigua,Alloprevotella_tannerae,Prevotella_sp_oral_taxon_306
1.1.99.2	Prevotella_nigrescens,Fusobacterium_nucleatum
1.1.99.33	Methanobrevibacter_smithii
1.14.19.1	Prevotella_denticola,Citrobacter_freundii
1.17.1.2	Bifidobacterium_bifidum,Bifidobacterium_longum
1.18.1.6	Bifidobacterium_bifidum,Bifidobacterium_longum
1.5.1.29	Parvimonas_micra,Fusobacterium_nucleatum
1.2.1.59	Methanobrevibacter_smithii
1.2.99.7	Eubacterium_hallii,Blautia_wexlerae
1.21.4.2	Slackia_exigua,Prevotella_sp_oral_taxon_306
1.21.4.4	Slackia_exigua,Alloprevotella_tannerae,Prevotella_sp_oral_taxon_306
1.3.1.43	Eubacterium_hallii,Anaerostipes_hadrus
1.3.3.1	Bifidobacterium_longum,Anaerostipes_hadrus
1.3.99.3	Parvimonas_micra,Fusobacterium_necrophorum
1.3.99.22	Eubacterium_hallii,Anaerostipes_hadrus
…	…

Table 3.4.2 Enzymes groups based on their correlations with microorganism levels

Microorganisms	Related Enzymes
Fusobacterium nucleatum	1.1.1.408,1.1.99.2,1.5.1.29,1.6.99.5 ,2.4.1.80,2.7.1.68,3.1.3.29,3.2.1.65,3.4.19.1,3.4.22.37,4.3.1.14,4.4.1.17,5.4.3.3
Bacteroides fragilis	1.1.1.136,1.1.1.281,1.2.4.4,2.1.1.29,2.4.1.289,2.6.1.33,2.7.1.168,3.1.1.72,3.2.1.169,3.6.3.54,4.2.1.135
Prevotella nigrescens	1.1.1.408,1.1.99.2,1.6.99.5,2.3.1.247,3.1.3.29,3.2.1.65,3.4.19.1,3.4.21.26,3.4.22.37,4.4.1.17
Porphyromonas asaccharolytica	1.1.1.61,1.10.2.2,1.4.99.5,2.3.2.13,2.4.1.19,2.4.1.83,2.7.8.24,2.8.3.9,3.1.1.47,3.1.2.1,3.1.3.71,3.1.7.2,3.4.22.10,3.4.24.64,4.3.1.6,5.1.99.1,5.2.1.2,5.3.3.3
Slackia exigua	1.1.2.4,1.21.4.3,1.21.4.4,1.6.99.5,2.7.1.68,3.1.3.29,3.2.1.65,3.4.21.26,3.4.22.37,4.1.3.4
Cloacibacillus porcorum	2.6.1.36
Faecalibacterium prausnitzii	1.1.1.69,1.4.1.14,2.1.1.51,2.4.1.211,4.2.1.55,4.4.1.1,5.4.1.2
Prevotella copri	1.1.1.40,1.4.7.1,1.5.1.7,2.1.1.148,2.4.1.281,2.4.2.45,3.4.11.5,3.6.1.13,3.6.3.8,4.3.3.6,5.5.1.4,1.5.1.2
Flavonifractor plautii	1.3.1.31,1.4.1.11,3.5.99.3,4.2.1.119

Streptococcus salivarius	1.1.1.36,2.3.1.274,2.3.2.10,2.4.1.10,2.4.1.5,2.7.4.2,3.2.1.11,3.4.24.57,6.3.2.7
Escherichia coli	1.1.1.103,1.1.1.127,1.1.1.130,1.1.1.132,1.1.1.237,1.1.1.282,1.1.1.291,1.1.1.298,1.1.1.350,1.1.1.373,1.1.1.381,1.1.1.65,1.1.1.67v,1.1.1.8,1.1.1.83,1.1.5.3,1.1.5.4,1.1.99.1,1.11.1.
Eubacterium limosum	1.3.7.7
…	…

Table 3.4.1 fundamentally shows the groups of microorganisms generated by calculation of correlations between enzymes and microorganisms. Each entry consists of a list of microorganisms that are correlated with a particular enzyme above a threshold value (0.5). Table 3.4.2 outlines the clusters of enzymes identified through analyzing the correlations between enzymes and microorganisms. It details the enzymes that exhibited a correlation with a particular microorganism, surpassing the threshold of 0.5.

For instance, the enzyme group "1.3.1.31, 1.4.1.11, 3.5.99.3, 4.2.1.119" was found to correlate with "Flavonifractor plautii," as evidenced in Table 3.4.2. Consequently, it is suggested that there might be a relationship between these enzymes and the specific microorganism mentioned.

The dataset, composed solely of enzymes (e.g., 1.3.1.31, 1.4.1.11, 3.5.99.3, 4.2.1.119) with corresponding class labels—a subset of the original dataset—is crucial for effectively distinguishing between two classes: CRC patients and control samples. The role and importance of these enzymes in class differentiation are assessed by the Ranking function R(), which is the second component of our methodology. This function gives a score to each group created by the G() function using an internal cross-validation mechanism, evaluating their significance in the context of class differentiation. The Ranking algorithm R() is specifically designed to assess species-enzyme groups, quantitatively gauging each group's contribution to the overall classification accuracy.

The Ranking function R() is designed to evaluate and score groups based on their ability to classify or differentiate between two classes in datasets. It uses internal cross-validation, a technique in machine learning, to assess the predictive accuracy of each group. R(), the ranking technique, uses internal cross-validation to assign a score to each group. The ranking process is pivotal in determining the relative importance and effectiveness of each group in terms of their ability to classify the data into the respective classes. This function is crucial for assessing the relative importance or effectiveness of each group created by the G() function. This ranking allows you to select the top-

performing groups for further analysis or model building. It provides insights into which groups of enzymes (or species) are most significant in terms of their correlation and impact on the classification task. This process is integral in determining which groups are most effective in distinguishing between the classes in the data, thereby guiding the subsequent analysis or predictive modeling.

Sub-datasets are generated from the original dataset in the G() step, each containing enzymes linked to a particular species group while retaining the original class labels. The R() function assigns a score to each group based on their ability to differentiate between case and control classes, using cross-validation with a classification algorithm. When the R() function employs cross-validation, it systematically evaluates how well each group (created by the G() function) can differentiate between the case and control classes, thus ensuring that the ranking of groups is as accurate and generalizable as possible. Post-ranking, every group is scored, reflecting its discriminative power. The R() function's output is a ranked list of these groups. Subsequently, model testing can either focus on the highest-ranked group or extend to the top 'j' groups, with 'j' set at 10 for this study. This involves creating sub-datasets with enzymes from the top 10 ranked groups, keeping their original labels. A machine learning model is then developed using this tailored data and evaluated against the test set.

In essence, this process allows for a focused analysis of the most relevant groups (as determined by the R() function) and their effectiveness in distinguishing between different classes in the dataset. By limiting the analysis to the top-ranked groups, the aim is to optimize the model's predictive accuracy and relevance to the specific case-control scenarios being studied.

3.4.1 miRcorrNet

In this study, enzyme and species data were run in reverse as mentioned before but for the better understanding explanations and examples continued with enzymes groups based on their correlations with species levels. Suppose we have two datasets: one, Denzymes, containing enzyme data, and the other, Dspecies, featuring species data, both covering the same N samples. Table 3.2.1 and Table 3.2.2 exemplifies these two datasets. In this context, let L represent the count of enzymes in Denzymes, and K denote the number of species in Dspecies. The methodology employed in miRcorrNet system is depicted in Figure 3.4.1.1. This workflow includes multiple components, with the C(),

G(), and R() functions being the primary elements. The initial stage of the miRcorrNet workflow involves the input component, where two datasets, Denzymes and Dspecies, are introduced. Following this, the second and third components focus on the elimination of missing values and the normalization of the data, respectively. Subsequently, the fourth component involves identifying the abundance of microorganisms and enzymes in functional profiling through the application of a t-test. We restrict our consideration to only those enzymes and species that have p-values of 0.05 or lower. The fifth component of the workflow involves the C() function, which employs the Pearson Correlation Coefficient to identify associations between enzymes and species. This function utilizes the Pearson correlation coefficient to detect sets of enzymes that show negative correlations with specific species. A stringent threshold of -0.5 is set to ensure the detection of high-confidence associations. The outcome of the C() function is a "species-enzymes correlation" matrix, which then serves as the input for the subsequent component in the workflow. In the context of the C() function in miRcorrNet workflow, the Pearson Correlation Coefficient is used to detect the strength and direction of the association between specific enzymes and species, helping to identify which enzymes are significantly correlated (negatively) with particular species.

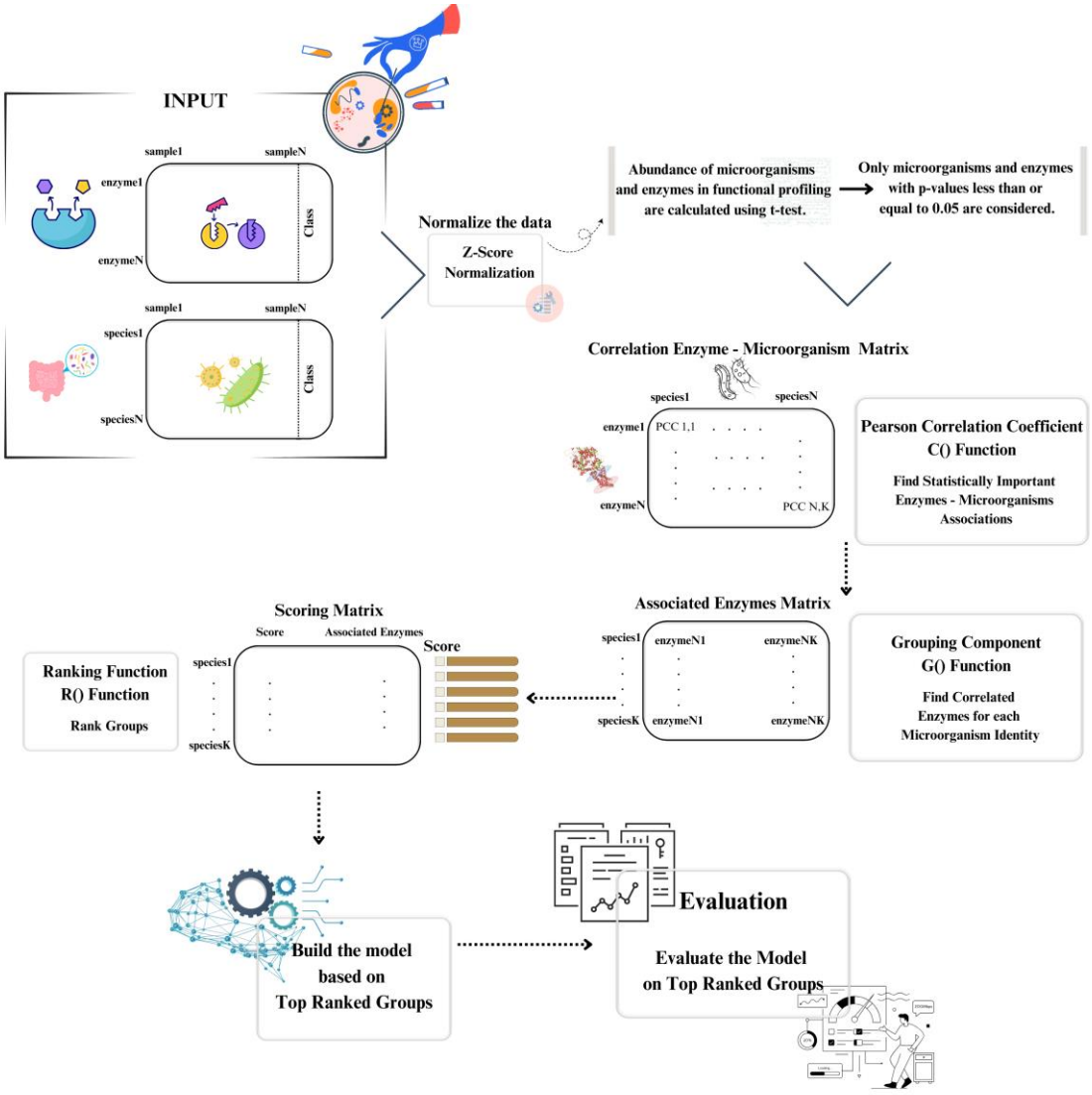

Figure 3.4.2.1 miRcorrNet Workflow

The next component in the workflow is the "Grouping Component," which plays a pivotal role in integrating species and enzymes. This process is essentially the G() function, responsible for creating groups that encompass associations between species and their corresponding enzymes. The objective of this function is to generate groups that will be utilized to form sub-datasets, each corresponding to a group while maintaining the original class labels. This component results in the creation of the "Associated Enzymes" matrix, which catalogs sets of correlated enzymes for each species. This matrix then becomes the input for the subsequent component, known as the "Ranking Component".

The "Ranking Component" assesses each enzyme group from the "Associated Enzyme" matrix, determining its importance in differentiating the two-class data. This involves constructing models based on the highest-ranked groups. In this process, ten models are developed, with the first model being based on the top-ranked group, the second model incorporating the first two ranked groups, and this pattern continues up to the top ten ranked groups. By "building on top groups," we refer to the utilization of enzymes associated with these groups. The final stage in the workflow is to evaluate the model that has been developed based on these top-ranked groups. This component processes the initial Denzymes data along with the "Associated Enzyme Matrix". When dealing with each species in the "Associated Enzymes Matrix", a smaller subset of the Denzymes data is generated, which includes only the enzymes specified in that particular species entry.This component generates L subsets of data, and each of these datasets is subsequently utilized in the following component for conducting scoring processes. The Score component processes each subset of data, implementing a cross-validation method and documenting its effectiveness. For model construction and prediction, a Random Forest classifier is employed as seen in Figure 3.4.1.2. The result of this component is a Scoring matrix that offers scores and lists the recognized enzymes for every species group. This matrix is instrumental in ranking the groups and will later be utilized to develop a model based on the top j groups. The model will be constructed focusing on the enzymes from the top-ranked groups.

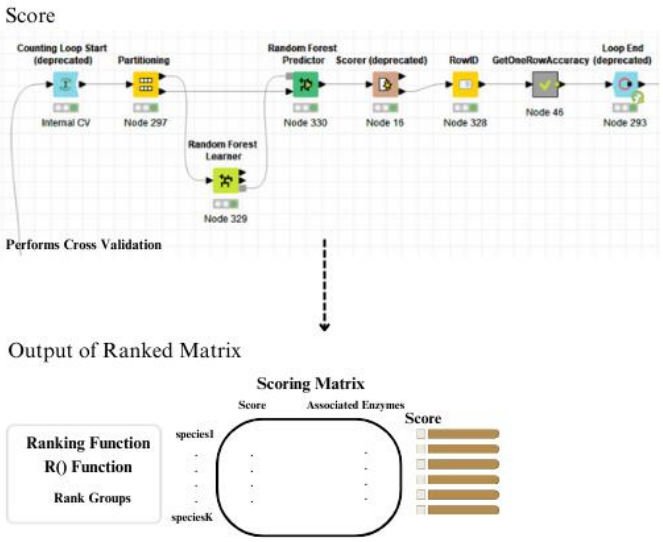

Figure 3.4.1.2 Details of the R() function

3.4.1.1 Determining the importance of enzymes and species in miRcorrNet

Figure 3.4.1.1 illustrates that miRcorrNet iterates the procedure N times, where in each iteration, 90% of the data is used for training and the remaining 10% for testing. Moreover, miRcorrNet employs a random selection process that under-samples samples at a 1:2 ratio. In every iteration, this method produces slightly varied lists of microorganisms and their corresponding enzymes, necessitating the use of a prioritization method for these lists. The RobustRankAggreg R package is integrated into miRcorrNet, assigning a P-value to each item in the combined list to indicate the ranking quality of each element/entity relative to the expected ranking. As depicted in Figure 3.4.1.1, during each of the N iterations, microorganisms are ordered based on the scores from the function R(). This process results in N distinct lists showing varying rankings of microorganisms. These varying lists are then used as input for RobustRankAggreg. Similarly, to rank enzymes, each enzyme is assigned the rank of its corresponding microorganism, creating N lists of ranked enzymes. These enzyme rankings are also fed into RobustRankAggreg for the final ranking process.

Chapter 4

Results

In this study, we investigated the association between the human gut microbiota and colorectal cancer (CRC) to identify potential microorganism-enzyme pairs as biomarkers for CRC. Our analysis involved comprehensive metagenomic sequencing data from CRC patients and healthy controls, examining both the diversity and functional capacity of the gut microbiota. Our analysis identified several significant microorganism-enzyme pairs with distinct abundance patterns between CRC patients and healthy controls. For example, [specific microorganisms and enzymes identified in the study], were notably associated with CRC cases. These findings suggest a potential mechanistic link between gut microbiota composition, enzymatic activity, and CRC pathogenesis.

4.1 Performance Evaluation of Enzyme and Species Clustering

The clustering of enzymes and species based on their correlations was evaluated to identify potential biomarkers for colorectal cancer (CRC). The assessment was performed using various machine learning models, and the results were evaluated using a range of performance metrics including accuracy, sensitivity, specificity, F-measure, and the area under the ROC curve (AUC).

4.1.1 Enzyme Clusters Based on Correlations with Species Levels

Enzyme clustering involved the evaluation of groups based on their ability to distinguish between CRC patients and control samples. The mean accuracy across different cluster group sizes remained relatively stable at approximately 0.67, indicating a consistent classification performance irrespective of the number of clusters. Notably,

sensitivity peaked in smaller cluster sizes, specifically with a single cluster, suggesting a potential focus on specific enzyme subsets for CRC detection. The specificity was relatively high across all cluster groups, with the best performance observed in single-cluster scenarios. However, F-measure and AUC values suggested a balanced performance when using between 6 to 10 clusters, peaking at an AUC of 0.74, which indicates a strong model's discriminative ability between positive and negative classes.

Table 4.2.1.1 Performance Evaluation of Enzyme groups based on their correlations with species levels

		Mean						
#Clusters	#Enzymes	Accuracy	Sensitivity	Specificity	F-measure	**Area Under Curve**	Iteration (Max*)	learner (First)
10	371	0.67	0.65	0.68	0.66	**0.73**	9	svm
9	365	0.66	0.64	0.69	0.65	**0.73**	8	svm
8	357	0.67	0.65	0.69	0.66	**0.73**	7	svm
7	347.8	0.67	0.66	0.69	0.66	**0.74**	6	svm
6	330.8	0.68	0.67	0.69	0.67	**0.74**	5	svm
5	260.2	0.67	0.64	0.7	0.66	**0.73**	4	svm
4	211.4	0.67	0.62	0.72	0.65	**0.74**	3	svm
3	158.1	0.66	0.61	0.7	0.64	**0.73**	2	svm
2	122.6	0.64	0.59	0.69	0.61	**0.71**	1	svm
1	41.1	0.63	0.54	0.72	0.58	**0.68**	0	svm
		Standard Deviation						
#Clusters	#Enzymes	Accuracy	Sensitivity	Specificity	F-measure	**Area Under Curve**	Iteration (Max*)	learner (First)
10	92	0.03	0.05	0.05	0.04	**0.03**	9	svm
9	93	0.04	0.05	0.04	0.04	**0.03**	8	svm
8	96	0.04	0.07	0.05	0.05	**0.03**	7	svm
7	98	0.03	0.05	0.03	0.04	**0.03**	6	svm
6	96	0.04	0.04	0.05	0.03	**0.03**	5	svm
5	156	0.04	0.05	0.06	0.04	**0.04**	4	svm
4	157	0.05	0.06	0.07	0.05	**0.04**	3	svm
3	154	0.03	0.05	0.08	0.03	**0.03**	2	svm
2	153	0.03	0.09	0.09	0.05	**0.04**	1	svm
1	80	0.02	0.13	0.1	0.07	**0.02**	0	svm
Enzyme groups based on their correlations with species levels								

4.1.2 Species Clusters Based on Correlations with Enzyme Levels

In species clustering, groups were again evaluated for their discriminatory power. Accuracy was consistently at 0.7 across all cluster sizes, while sensitivity showed lower values, indicating challenges in the true positive rate. The specificity was notably high, with a peak of 0.91, suggesting excellent performance in identifying true negatives. Similar to enzyme clustering, the F-measure and AUC indicated a better balance in larger cluster sizes, with an AUC of 0.74, reinforcing the model's effectiveness.

Table 4.3.2.1 Performance Evaluation of Species groups based on their correlations with enzymes levels

		Mean						
#Clusters	#Species	Accuracy	Sensitivity	Specificity	F-measure	**Area Under Curve**	Iteration (Max*)	learner (First)
10	27.5	0.7	0.49	0.91	0.62	**0.74**	9	svm
9	25.2	0.7	0.49	0.9	0.62	**0.74**	8	svm
8	22.3	0.7	0.49	0.9	0.62	**0.74**	7	svm
7	19.3	0.7	0.49	0.9	0.61	**0.73**	6	svm
6	16.1	0.69	0.48	0.9	0.61	**0.73**	5	svm
5	12.7	0.69	0.47	0.9	0.6	**0.72**	4	svm
4	9.9	0.68	0.47	0.89	0.59	**0.72**	3	svm
3	7.1	0.68	0.47	0.89	0.59	**0.72**	2	svm
2	4.7	0.68	0.46	0.89	0.59	**0.71**	1	svm
1	2.3	0.67	0.43	0.9	0.56	**0.68**	0	svm
		Standart Deviation						
#Clusters	#Species	Accuracy	Sensitivity	Specificity	F-measure	**Area Under Curve**	Iteration (Max*)	learner (First)
10	5	0.03	0.06	0.05	0.04	**0.02**	9	svm
9	5	0.03	0.07	0.05	0.05	**0.02**	8	svm
8	4	0.02	0.07	0.05	0.05	**0.02**	7	svm
7	4	0.03	0.07	0.06	0.05	**0.03**	6	svm
6	3	0.03	0.06	0.05	0.05	**0.02**	5	svm
5	2	0.03	0.06	0.05	0.05	**0.02**	4	svm
4	2	0.03	0.06	0.05	0.05	**0.02**	3	svm
3	1	0.02	0.06	0.05	0.04	**0.02**	2	svm
2	1	0.02	0.06	0.05	0.04	**0.02**	1	svm
1	0.4	0.03	0.09	0.04	0.07	**0.03**	0	svm
Species groups based on their correlations with enzymes levels								

4.2 Variation Analysis

The standard deviation of performance metrics across cluster sizes was analyzed to understand the stability of the models. It was observed that there was minimal variation in accuracy, sensitivity, specificity, and F-measure, suggesting a stable classification performance across different iterations of the clustering process. This stability is crucial for ensuring the reliability of enzyme and species clusters as potential biomarkers for CRC.

4.3 Comparative Performance Analysis

Comparative analysis using different feature selection methods and classifiers was conducted to validate the robustness of the identified biomarkers. The results demonstrated that specific machine learning models, such as Random Forest and XGBoost, consistently showed superior performance, particularly when combined with feature selection techniques like Information Gain and SelectKBest. These methods yielded high AUC values, indicating their potential in enhancing the predictive power of the biomarker identification process.

The feature selection process and the subsequent machine learning model application on the curated groups of enzymes and species delineated the potential of these biological entities in distinguishing between CRC patients and healthy controls. The AUC values derived from these models substantiate their candidacy as robust biomarkers.

4.4 Comparative Analysis of Feature Selection Techniques

To further assess the robustness of the biomarkers identified for colorectal cancer (CRC), a comparison was made using different feature selection (FS) methods combined with various classifiers. The Area Under the Curve (AUC) was used as a performance metric for these models, with 150 selected features for enzymes and 5 for species.

The Conditional Mutual Information Maximization (CMIM) method when used with an Adaboost classifier, provided an AUC of 0.569 for enzymes and 0.556 for species. On the other hand, the Random Forest (RF) classifier showed an AUC of 0.555 for enzymes and 0.557 for species using the same FS method. These values indicate moderate predictive power, suggesting a reasonable ability to differentiate between the CRC and control groups.

Feature selection using Information Gain (IG) yielded significantly higher AUC values when paired with the Adaboost classifier (0.684 for enzymes, 0.540 for species) and even higher with the RF classifier (0.732 for enzymes, 0.538 for species). These results suggest that IG is a powerful method for selecting relevant features, particularly when combined with RF, resulting in a model with good classification abilities.

The Maximum Relevance Minimum Redundancy (MRMR) FS method exhibited lower AUC values, with the RF classifier achieving an AUC of 0.526 for enzymes and 0.491 for species. This outcome indicates a need for careful consideration of feature redundancy when selecting biomarkers for CRC.

Interestingly, the SelectKBest (SKB) FS method achieved some of the highest AUC scores. When used with an Adaboost classifier, the AUC was 0.681 for enzymes and an impressive 0.731 for species. The RF classifier further improved these results, producing an AUC of 0.743 for enzymes and 0.689 for species, highlighting the strength of this combination in distinguishing between CRC patients and healthy controls.

The Extreme Gradient Boosting (XGBoost) method also yielded high AUC values, particularly with the XGBoost classifier, resulting in AUCs of 0.764 for both enzymes and species. This indicates a strong predictive model that effectively differentiates between the two classes.

The Recursive Feature Elimination (RFE) method with an SVM classifier resulted in AUCs of 0.730 for enzymes and 0.669 for species, suggesting the effectiveness of RFE in reducing feature dimensionality while maintaining predictive accuracy.

The miRcorrNet approach, which combines machine learning with network analysis, was used to identify species and enzyme groups based on their correlations. The Random Forest classifier provided an AUC of 0.714 for species groups and 0.734 for enzyme groups, underscoring the potential of miRcorrNet in biomarker discovery.

Overall, the comparison table revealed that the choice of FS method and classifier has a significant impact on the model's ability to predict CRC from metagenomic data.

The Random Forest classifier, particularly when used with SKB or XGBoost for feature selection, showed superior performance, suggesting their use in future studies aiming to develop robust diagnostic models for CRC.

Table 4.4.4 Comparison table with different FS Methods

	# selected features	**150**	**5**
FS Type	**Classifier**	**Enzymes**	**Species**
		AUC	**AUC**
CMIM	Adaboost	0.568986111	0.556388889
	DT	0.504166667	0.49
	LogitBoost	0.582458333	0.55
	RF	0.555111111	0.556902778
	SVM_opt	0.517805556	0.492555556
	Stack_Logitboost_Kmenas	0.526763889	0.503333333
	Stack_SVM_Kmeans	0.530430556	0.49375
	XGBoost	0.543625	0.541916667
FCBF	Adaboost	0.520958333	0.512819444
	DT	0.529166667	0.516916667
	LogitBoost	0.508125	0.508611111
	RF	0.52075	0.519097222
	SVM_opt	0.465472222	0.497083333
	Stack_Logitboost_Kmenas	0.482194444	0.488638889
	Stack_SVM_Kmeans	0.488527778	0.465847222
	XGBoost	0.539819444	0.511722222
IG	Adaboost	0.683680556	0.5395
	DT	0.5825	0.540347222
	LogitBoost	0.683694444	0.544138889
	RF	**0.732236111**	0.537555556
	SVM_opt	0.620472222	0.488305556
	Stack_Logitboost_Kmenas	0.560944444	0.521680556
	Stack_SVM_Kmeans	0.622472222	0.548819444
	XGBoost	**0.718583333**	0.532958333

MRMR	Adaboost	0.507472222	0.490097222
	DT	0.5025	0.488458333
	LogitBoost	0.507902778	0.492583333
	RF	0.526194444	0.490833333
	SVM_opt	0.511638889	0.495027778
	Stack_Logitboost_Kmenas	0.520319444	0.508041667
	Stack_SVM_Kmeans	0.485027778	0.496152778
	XGBoost	0.515263889	0.489194444
SKB	Adaboost	**0.681097222**	**0.730805556**
	DT	0.620833333	0.622930556
	LogitBoost	**0.711458333**	**0.739833333**
	RF	**0.742805556**	**0.688875**
	SVM_opt	**0.697888889**	**0.708888889**
	Stack_Logitboost_Kmenas	0.648138889	**0.685638889**
	Stack_SVM_Kmeans	0.640166667	**0.714597222**
	XGBoost	**0.735888889**	**0.723041667**
XGB	Adaboost	**0.720638889**	**0.716541667**
	DT	0.613333333	0.641847222
	LogitBoost	**0.723180556**	**0.726041667**
	RF	**0.764527778**	**0.695583333**
	SVM_opt	0.507847222	**0.711125**
	Stack_Logitboost_Kmenas	0.669236111	**0.679986111**
	Stack_SVM_Kmeans	0.658388889	**0.724361111**
	XGBoost	**0.764444444**	**0.719958333**
RCE	**SVM**	**0.737663934**	0.65709016
RFE	**SVM**	**0.729596528**	0.668625
	selected species :5		
miRcorrNet - Species groups based on their correlations with enzymes levels	**Random Forest**		**0.714016393**
	selected enzymes: 150		
miRcorrNet - Enzyme groups based on their correlations with species levels	**Random Forest**	**0.734398907**	

4.5 Analysis of Correlations Between Species and Enzyme Levels

The investigation into the microbiota's potential role in colorectal cancer (CRC) revealed intricate correlations between certain microbial species and enzyme levels. The analysis focused on clusters of enzymes that demonstrated significant correlations with particular species, as well as clusters of species that were notably associated with enzyme levels. The results offer insights into the interplay between microbial presence and enzymatic activity, potentially contributing to the pathogenesis of CRC or serving as biomarkers for the disease.

4.5.1 Species Groups Based on Their Correlations with Enzymes Levels

Noteworthy correlations between enzymes and species were observed. For instance, the enzyme 1.5.1.29 showed a significant correlation with species such as Parvimonas micra and Fusobacterium nucleatum. Similarly, enzyme 3.1.1.15 was correlated with Eisenbergiella tayi, Parvimonas micra, and Fusobacterium necrophorum. These associations indicate a potential link between the enzymatic functions and the microbial species that may play a role in the CRC microenvironment.

Further correlations were seen with enzyme 1.3.99.3 and species such as Parvimonas micra and Fusobacterium necrophorum, suggesting that the enzymatic activity might be influenced by the presence of these bacteria. The enzyme 4.2.1.43 was linked with multiple species, including Eisenbergiella tayi, Parvimonas micra, and Fusobacterium necrophorum, emphasizing the multi-faceted interactions within the gut microbiota.

Table 4.5.1.1 Species groups based on their correlations with enzymes levels

Enzyme	Related Species
1.5.1.29	Parvimonas_micra,Fusobacterium_nucleatum
3.1.1.15	Eisenbergiella_tayi,Parvimonas_micra,Fusobacterium_necrophorum
1.3.99.3	Parvimonas_micra,Fusobacterium_necrophorum
1.1.1.361	Eisenbergiella_tayi,Fusobacterium_necrophorum

4.2.1.43	Eisenbergiella_tayi,Parvimonas_micra,Fusobacterium_necrophorum
2.8.3.17	Peptostreptococcus_stomatis
3.2.1.65	Slackia_exigua,Prevotella_intermedia,Prevotella_nigrescens,Eubacterium_infirmum, Fusobacterium_nucleatum
2.3.1.247	Prevotella_nigrescens,Gemella_bergeri,Fusobacterium_nucleatum
3.4.19.1	Prevotella_nigrescens,Gemella_bergeri,Fusobacterium_nucleatum
1.6.99.5	Slackia_exigua,Prevotella_nigrescens,Fusobacterium_nucleatum

4.5.2 Enzyme Groups Based on Their Correlations with Species Levels

Conversely, certain species were found to be associated with clusters of enzymes. Fusobacterium nucleatum, a species often implicated in CRC, was correlated with a vast array of enzymes, including 1.1.1.408, 1.1.99.2, and 1.5.1.29. This extensive list of enzymes may shed light on the metabolic pathways influenced by or contributing to the presence of F. nucleatum in CRC.

Bacteroides fragilis, another species of interest, was associated with enzymes such as 1.1.1.136 and 1.1.1.281, reflecting its metabolic activities in the gut. Prevotella nigrescens showed correlations with enzymes that could be integral to its role in the gut ecosystem and potential involvement in CRC pathophysiology.

The species Porphyromonas asaccharolytica and Slackia exigua were each associated with distinct sets of enzymes, indicating their unique contributions to or alterations in metabolic processes in the gut microbiome of CRC patients.

Overall, the analysis underscores the potential of these species and enzyme correlations as biomarkers for CRC. Their significance lies in the possibility of distinguishing CRC-associated microbial and enzymatic profiles from those of healthy controls. These findings also pave the way for future explorations into therapeutic targets within the gut microbiota, offering potential strategies for precision medicine in the context of CRC.

Table 4.5.2.1 Enzyme groups based on their correlations with species levels

Species	**Related Enzymes**
Fusobacterium nucleatum	1.1.1.408,1.1.99.2,1.5.1.29,1.6.99.5 ,2.4.1.80,2.7.1.68,3.1.3.29,3.2.1.65,3.4.19.1,3.4.22.37,4.3.1.14,4.4.1.17,5.4.3.3
Bacteroides fragilis	1.1.1.136,1.1.1.281,1.2.4.4,2.1.1.29,2.4.1.289,2.6.1.33,2.7.1.168,3.1.1.72,3.2.1.169,3.6.3.54,4.2.1.135
Prevotella nigrescens	1.1.1.408,1.1.99.2,1.6.99.5,2.3.1.247,3.1.3.29,3.2.1.65,3.4.19.1,3.4.21.26,3.4.22.37,4.4.1.17
Porphyromonas asaccharolytica	1.1.1.61,1.10.2.2,1.4.99.5,2.3.2.13,2.4.1.19,2.4.1.83,2.7.8.24,2.8.3.9,3.1.1.47,3.1.2.1,3.1.3.71,3.1.7.2,3.4.22.10,3.4.24.64,4.3.1.6,5.1.99.1,5.2.1.2,5.3.3.3

Slackia exigua	1.1.2.4,1.21.4.3,1.21.4.4,1.6.99.5,2.7.1.68,3.1.3.29,3.2.1.65,3.4.21.26,3.4.22.37,4.1.3.4
Escherichia coli	1.1.1.103,1.1.1.127,1.1.1.130,1.1.1.132,1.1.1.237,1.1.1.282,1.1.1.291,1.1.1.298,1.1.1.350,1.1.1.373,1.1.1.381,1.1.1.65,1.1.1.67v,1.1.1.8,1.1.1.83,1.1.5.3,1.1.5.4,1.1.99.1,1.11.1.21,1.11.1.6,1.11.1.9,1.12.99.6,1.13.11.29,1.14.11.17,1.14.11.47,1.14.12.1,1.14.12.17,1.14.12.19,1.14.13.107,1.14.13.239,1.14.13.59,1.14.14.51,1.14.99.46,1.16.1.9,1.17.1.9,1.17.7.1,1.18.1.3,1.19.1.1,1.2.1.10,1.2.1.16,1.2.1.19,1.2.1.21,1.2.1.5,1.2.1.71,1.2.1.72,1.2.1.8,1.2.1.88,1.2.4.2,1.2.5.1,1.2.99.6,1.3.1.28,1.3.1.84,1.3.1.87,1.3.3.3,1.3.5.3,1.3.8.13,1.4.1.4,1.4.3.19,1.4.3.21,1.4.99.6,1.5.1.42,1.5.5.1,1.6.1.1,1.6.5.2,1.6.5.5,1.6.6.9,1.7.1.15,1.7.1.4,1.7.2.3,1.7.5.1,1.7.99.4,1.8.1.2,1.8.1.8,1.8.5.3,1.8.5.7,1.97.1.9,2.1.1.166,2.1.1.170,2.1.1.172,2.1.1.174,2.1.1.181,2.1.1.200,2.1.1.201,2.1.1.222,2.1.1.265,2.1.1.298,2.1.1.35,2.1.1.61,2.1.1.77,2.3.1.109,2.3.1.118,2.3.1.15,2.3.1.16,2.3.1.191,2.3.1.193,2.3.1.210,2.3.1.242,2.3.1.243,2.3.1.251,2.3.1.267,2.3.1.40,2.3.1.41,2.3.1.46,2.3.1.61,2.3.1.n5,2.3.2.2,2.3.3.16,2.3.3.9,2.4.1.12,2.4.1.15,2.4.1.180,2.4.1.58,2.4.2.15,2.4.2.4,2.4.2.43,2.4.99.13,2.5.1.141,2.5.1.18,2.5.1.64,2.5.1.94,2.6.1.57,2.6.1.66,2.6.1.82,2.6.1.88,2.7.1.165,2.7.1.184,2.7.1.19,2.7.1.192,2.7.1.194,2.7.1.197,2.7.1.201,2.7.1.202,2.7.1.51,2.7.1.53,2.7.1.58,2.7.1.60,2.7.1.73,2.7.1.83,2.7.1.89,2.7.11.5,2.7.2.15,2.7.4.23,2.7.4.292.7.7.1,2.7.7.19,2.7.7.58,2.7.7.61,2.7.7.73,2.7.7.76,2.7.7.n1,2.7.8.20,2.7.8.37,2.7.8.42,2.8.1.12,2.8.1.2,2.8.3.19,2.8.3.21,3.1.1.1,3.1.1.32,3.1.1.4,3.1.1.45,3.1.1.85,3.1.11.3,3.1.11.5,3.1.2.12,3.1.2.14,3.1.2.2,3.1.2.28,3.1.21.1,3.1.21.7,3.1.3.10,3.1.3.12,3.1.3.22,3.1.3.26,3.1.3.50,3.1.3.6,3.1.3.74,3.1.3.81,3.1.3.82,3.1.3.89,3.1.3.97,3.1.4.14,3.1.4.17,3.1.4.55,3.2.1.196,3.2.1.199,3.2.1.28,3.2.2.28,3.3.2.1,3.4.11.1,3.4.11.23,3.4.13.21,3.4.19.5,3.4.24.55,3.4.24.70,3.4.25.2,3.5.1.105,3.5.1.110,3.5.1.78,3.5.1.96,3.5.3.23,3.5.3.26,3.5.4.13,3.5.4.4,3.5.4.42,3.6.1.12,3.6.1.25,3.6.1.26,3.6.1.45,3.6.1.65,3.6.1.67,3.6.1.9,3.6.3.19,3.6.3.2,3.6.3.33,3.8.1.7,4.1.1.112,4.1.1.17,4.1.1.18,4.1.1.47,4.1.1.85,4.1.2.20,4.1.2.21,4.1.2.52,4.1.2.53,4.1.3.1,4.1.3.17,4.1.3.40,4.1.99.1,4.1.99.3,4.2.1.104,4.2.1.12,4.2.1.40,4.2.1.6,4.2.1.7,4.2.1.70,4.2.1.75,4.2.1.79,4.2.1.90,4.2.2.n2,4.3.1.18,4.3.2.7,4.4.1.13,4.4.1.15,5.1.3.20,5.1.3.29,5.3.1.22,5.3.1.28,5.3.1.31,5.3.2.2,5.3.3.14,5.4.99.19,5.4.99.20,5.4.99.21,5.4.99.22,5.4.99.27,5.4.99.28,5.4.99.29,6.2.1.1,6.2.1.17,6.2.1.26,6.2.1.33,6.2.1.48,6.3.1.11,6.5.1.4,7.1.1.3,7.1.1.7,7.4.2.1,7.5.2.1,7.6.2.2,7.6.2.5
Faecalibacterium prausnitzii	1.1.1.69,1.4.1.14,2.1.1.51,2.4.1.211,4.2.1.55,4.4.1.1,5.4.1.2
Prevotella copri	1.1.1.40,1.4.7.1,1.5.1.7,2.1.1.148,2.4.1.281,2.4.2.45,3.4.11.5,3.6.1.13,3.6.3.8,4.3.3.6,5.5.1.4,1.5.1.2
Flavonifractor plautii	1.3.1.31,1.4.1.11,3.5.99.3,4.2.1.119
Streptococcus salivarius	1.1.1.36,2.3.1.274,2.3.2.10,2.4.1.10,2.4.1.5,2.7.4.2,3.2.1.11,3.4.24.57,6.3.2.7

Chapter 5

Biological Validation of Result Obtained from mircorrnet

5.1 Species Groups Based on Their Correlations with Enzymes Levels

FMN reductase (EC 1.5.1.29) is a crucial enzyme involved in various biological processes, plays a significant role in transferring reducing equivalents in different cellular pathways [49]. Flavin mononucleotide (FMN) reductases are crucial enzymes involved in various biological processes, including colorectal cancer. Research has shown that NADPH-cytochrome P450 oxidoreductase and nitric oxide synthase (NOS) are examples of FMN reductases that are involved in electron transfer activities that are crucial for the activity of enzymes like cytochrome P450 [61, 62, 63]. These reductases contain domains for binding FMN, FAD, and NADPH, which facilitate the transfer of reducing equivalents necessary for catalytic reactions [49].

Parvimonas micra found in the human oral and gastrointestinal microbiota, has garnered increasing attention in research due to its possible link to a number of illnesses, including colorectal cancer (CRC). It has been discovered that patients with colorectal cancer who have this bacterium had a lower five-year cancer-specific survival rate [52]. The potential function of Parvimonas micra in colorectal cancer is supported by emerging research. For example, studied the significance of Parvimonas micra and how it relates to other microbial biomarkers as a possible non-invasive intestines biomarker for colorectal cancer [53]. Additionally, a study by revealed a connection between Parvimonas micra and colorectal cancer [54]. Furthermore, a study by revealed that Parvimonas micra tumor colonization is linked to a lower prognosis in patients with colorectal cancer [52].

Fusobacterium nucleatum are normally found in the human oral cavity, however they have been linked to colorectal cancer (CRC) more and more recently [67, 68]. Fusobacterium nucleatum has attracted considerable interest for its possible involvement in accelerating and intensifying the development and advancement of colorectal cancer via multiple pathways [55]. Additionally, It's emphasized the association of Fusobacterium nucleatum in the gastrointestinal tract with the emergence of CRC, indicating its role in the carcinogenesis of CRC [57]. Furthermore, provided a thorough evaluation summarizing the role of Fusobacterium nucleatum in CRC, including its detection methods, underlying pathogenic mechanisms, and potential prevention strategies [58]. Furthermore, the review by emphasized Fusobacterium nucleatum's procarcinogenic activity, which may be related to its modification of the tumor immunological environment and activation of carcinogenesis and inflammatory pathways [59]. This further underscores the significance of Fusobacterium nucleatum in the context of CRC development. The high frequency of Fusobacterium nucleatum in colorectal tumors and its link to a lower chance of human survival were also covered, confirming the bacteria's significance in the pathophysiology of colorectal cancer [60].

Our research has identified enzyme 1.5.1.29, known as FMN reductase, in association with the bacterial species Parvimonas micra and Fusobacterium nucleatum as a potential biological indicator for Colorectal Cancer. This finding delineates a significant link between the metabolic pathways involving FMN reductase and the microbial ecosystem within the colorectal cancer milieu, providing new insights into the disease's molecular landscape and highlighting potential targets for diagnostic and therapeutic strategies.

L-arabinonolactonase (EC 3.1.1.15) is a essential enzyme in the alternative pathway of L-arabinose metabolism [61]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of L-arabinonolactonase on colorectal cancer.

Eisenbergiella tayi is a species of bacteria initially identified from human blood [62]. According to metagenomic sequencing, people with particular medical problems, such as overweight conditions and gestational diabetes mellitus, have higher relative abundances of Eisenbergiella tayi [63], [64]. This suggests a potential association between the abundance of Eisenbergiella tayi and specific health conditions.

Fusobacterium necrophorum is a type of anaerobic, Gram-negative bacteria commonly found in the human digestive tract, yet it can become pathogenic under specific circumstances [65]. Fusobacterium nucleatum and Fusobacterium necrophorum are frequently detected in colorectal cancer (CRC) tissues. These bacteria are linked to the development and worsening of colorectal cancer (CRC) [78, 79]. Studies have consistently reported the enrichment of Fusobacterium species, particularly F. nucleatum and F. necrophorum, in CRC tissues compared to healthy controls [78, 80]. Furthermore, Fusobacterium species have been linked to the promotion of chemoresistance in CRC and have been found to influence disease progression through various mechanisms [69].

The findings from our study underscore the novel discovery that L-arabinonolactonase (EC 3.1.1.15), in conjunction with the bacterial species Eisenbergiella tayi, Parvimonas micra, and Fusobacterium necrophorum, acts as a possible colorectal cancer biomarker.

Acyl-CoA dehydrogenases (EC 1.3.99.3) represent a group of enzymes essential for the metabolism of fatty acids and the breakdown of branched-chain amino acids [70]. Deficiencies in these enzymes can lead to fatty acid oxidation disorders, which can result in life-threatening metabolic derangements [71]. The direct impact of acyl-CoA dehydrogenase activity on colorectal cancer (CRC) development or progression is not explicitly outlined. However, fatty acid metabolism is essential to the development of colorectal cancer (CRC). Cancer cells' metabolic reprogramming involves alterations in fatty acid oxidation (FAO) and synthesis, impacting cell proliferation and survival [72]. To the best of my knowledge, and after a thorough investigation of the available literature, there is no explicit delineation of the direct impact that acyl-CoA dehydrogenase activity has on the development of colorectal cancer (CRC).

In our research, we identified the enzyme acyl-CoA dehydrogenase (EC 1.3.99.3), along with the presence of Parvimonas micra and Fusobacterium necrophorum, as a potential biomarker for Colorectal Cancer.

Glucose-6-phosphate 3-dehydrogenase (G6PD) (EC - 1.1.1.361) is essential to the metabolism of carbohydrates. It catalyzes the conversion of glucose-6-phosphate to 6-phosphogluconolactone and the reduction of NADP to NADPH, which is the initial step in the pentose phosphate pathway. This process is vital for generating NADPH, a reducing agent that helps protect cells from oxidative damage and is also used in anabolic reactions,

such as fatty acid synthesis [73]. An enzyme engaged in the pentose phosphate cycle, glucose-6-phosphate 3-dehydrogenase (G6PD) is essential for preserving the redox balance of cells. Research has indicated a potential link between altered G6PD activity and various health conditions, including cancer. Furthermore, it has been suggested that second primary cancers are a commonly encountered medical problem, indicating the need for further investigation into the role of G6PD in cancer growth [74]. Additionally, the incidence of second primary cancers has been found to negatively influence the prognosis of cancer patients [43].

In our study, glucose-6-phosphate 3-dehydrogenase (EC 1.1.1.361), in conjunction with the presence of Eisenbergiella tayi and Fusobacterium necrophorum, was found to be a possible colorectal cancer biological marker.

The enzyme 2-dehydro-3-deoxy-L-arabinonate dehydratase (DAH) (Enzyme 4.2.1.43) has a vital part in the metabolic pathways related to carbohydrate processing. It is involved in the dehydration process, which is essential for the metabolism of certain sugars, thereby contributing significantly to various biological processes, including cellular metabolism and energy production [75]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of 2-dehydro-3-deoxy-L-arabinonate dehydratase (DAH) on colorectal cancer.

The enzyme 4.2.1.43, known as 2-dehydro-3-deoxy-L-arabinonate dehydratase, alongside the presence of Eisenbergiella tayi, Parvimonas micra, and Fusobacterium necrophorum, was identified as a possible colorectal cancer biomarker.

The enzyme (R)-3-(aryl)lactate CoA-transferase (Enzyme 2.8.3.17) belongs to a newly identified family of CoA-transferases [76]. The involvement of CoA (coenzyme A) in colorectal cancer is a subject of interest due to its role in cellular metabolism and its possible influence on the onset of cancer. Research has demonstrated that colorectal cancer cells overexpress HMG-CoA reductase, an enzyme involved in the mevalonate pathway that is controlled by statins. Statins have also been demonstrated to stimulate death in cancer cell lines [77]. Additionally, the expression of hydroxymethylglutaryl-Coenzyme A synthase 2 has been affiliated with the effects of chemoradiotherapy on human colorectal cancer cells [78]. Moreover, the proliferation of colorectal cancer has been linked to the NAD+ salvage pathway, suggesting a possible connection to CoA

metabolism [79]. Furthermore, it has been discovered that microRNA-26a controls the metabolizing process of glucose in colorectal cancer cells by specifically targeting PDHX, which prevents pyruvate from being converted to acetyl coenzyme A in the citric acid cycle [80]. These results imply a major role for CoA and associated pathways in the pathogenesis and management of colorectal cancer.

Peptostreptococcus stomatis is a species of bacteria that belongs to the genus Peptostreptococcus. This genus is part of the Firmicutes phylum and is known for being anaerobic, meaning it does not require oxygen for growth. This genus of bacteria is commonly found in the dental cavity, digestive tract, and other parts of the human body. They can play roles in both healthy microbiota balance and in disease states [81]. Peptostreptococcus stomatis has been recognized as a possible indicator for colorectal cancer (CRC). Studies have indicated an intensified abundance of Peptostreptococcus stomatis in CRC, pointing out that it may be important for detecting the first signs of colorectal cancer [82]. Furthermore, Peptostreptococcus stomatis has been associated with gastric carcinogenesis, with key points in the ecological network of gastric cancer [83]. Additionally, it has been discovered that Peptostreptococcus stomatis is a prevalent high-risk pathogen of colorectal cancer (CRC) worldwide, and it is regarded as a significant variable in CRC models that forecast risks across various areas [84]. Additionally, it has been identified as one of the species enriched in gastric tumors, suggesting its potential significance in the progression of gastric cancer [98, 99]. Additionally, there is evidence that Peptostreptococcus stomatis may be a useful biomarker for the noninvasive detection and prognosis of colorectal laterally spreading cancers. These two biomarkers have a strong association [87]. Moreover, Peptostreptococcus stomatis has been associated with consensus molecular subtype 1 (CMS1) tumors, which are closely linked to the serrated pathway in CRC carcinogenic pathways [88]. These findings collectively suggest that Peptostreptococcus stomatis may serve as a meaningful microbial marker for the early detection, diagnosis, and risk prediction of colorectal and gastric cancers.

In our study, the enzyme 2.8.3.17, known as 3-(aryl)acryloyl-CoA:(R)-3-(aryl)lactate CoA-transferase, alongside Peptostreptococcus stomatis, was pinpointed as a potential biomarker for Colorectal Cancer. This finding not only enriches our understanding of CoA-related metabolic pathways in cancer development but also emphasizes the significant role of specific bacterial species in

influencing colorectal cancer's landscape, thereby offering novel insights into the disease's diagnostic and therapeutic avenues.

Levanases (EC 3.2.1.65) are enzymes present in certain microorganisms, such as Bacillus and Pseudomonas, that facilitate the breakdown of the β-2,6-linked main chain of levan, a type of natural fructan [89]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of Levanases on colorectal cancer.

Slackia exigua is a bacterium that has been associated with various health conditions. It has been identified within the gastrointestinal microbiome and has been affiliated with gastric carcinogenesis [96, 103, 104]. It has been revealed that Slackia exigua is a human intestine bacterium that can convert genistein and daizenzin [92]. It has been discovered to be considerably more abundant in the microbiota of gastric carcinogenesis, indicating a possible function in the development of the illness and dysbiosis of the gut [93]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of Slackia exigua on colorectal cancer.

Prevotella intermedia is linked to a wide range of gastrointestinal (GI) malignancies, such as gastrointestinal and colorectal cancers (CRC). It has been discovered that the microbiome of patients with gastric cancer and colorectal cancer has increased levels in Prevotella intermedia, which forms a robust co-occurrence network with disease development [107, 108]. Furthermore, Prevotella intermedia, along with other oral pathogens, has been connected to a higher risk of colorectal cancer, indicating its potential role in colorectal carcinogenesis [109, 110].

Prevotella nigrescens is a prevalent species in the human oral cavity, commonly found in dental plaque and associated with periodontal diseases [97]. Research has indicated that differences in the gut microbiota of patients with colorectal cancer (CRC) may be related to Prevotella, notably Prevotella nigrescens [98]. Furthermore, it has been discovered that Prevotella nigrescens is more abundant in CRC patients, indicating a possible connection between this bacterium and CRC [99].

Eubacterium infirmum is a species of the genus Eubacterium that has been associated with human health [100]. It has been identified in tumors, peri-implantitis sites, and in the root canals of diabetic patients [101]. The development of colorectal cancer has been linked to the gut microbiome; individuals with colorectal cancer have lower

levels of some bacteria, such as Eubacterium rectale and Faecalibacterium prausnitzii, than healthy controls [35]. It has also been discovered that the development of colorectal cancer is associated with a rise in the Eubacterium coprostanoligenes and Bacteroides families [102].

The enzyme EC 3.2.1.65 known as levanase, in conjunction with Slackia exigua, Prevotella intermedia, Prevotella nigrescens, Eubacterium infirmum, and Fusobacterium nucleatum, was identified as a potential biomarker for Colorectal Cancer.

3-keto-5-aminohexanoate cleavage (Enzyme 2.3.1.247) is a crucial catalyst in the cleavage of 3-keto-5-aminohexanoate, a reaction integral to specific metabolic pathways encompassed in the degradation of certain amino acids or compounds [103]. This enzyme is essential to metabolic processes, such as the breakdown and synthesis of different biochemical molecules. Its precise mechanism, substrate specificity, and biological role depend on the organism in which it is located. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of 3-keto-5-aminohexanoate cleavage on colorectal cancer.

Gemella bergeri belongs to the Gemella genus, which includes other species like Gemella sanguinis, Gemella haemolysans, and Gemella morbillorum [45, 118]. The presence of Gemella in colorectal cancer patients' gut microbiota has been identified through 16S ribosomal RNA sequencing, along with other bacteria like Fusobacterium and Parvimonas [105].

The enzyme 2.3.1.247, known as 3-keto-5-aminohexanoate cleavage enzyme, alongside Prevotella nigrescens, Gemella bergeri, and Fusobacterium nucleatum, emerged as a potential biomarker for Colorectal Cancer.

Acylaminoacyl-peptidase (ACPH, EC 3.4.19.1) is part of the prolyl oligopeptidase (POP) family and catalyzes the cleavage of N-acylated peptides into acylamino acids and peptides with free N-termini [106]. This enzyme affects the DNA damage response and is linked to a number of diseases, including cancer, cataract development, and Alzheimer's disease [105].

The enzyme acylaminoacyl-peptidase (EC 3.4.19.1), in association with Prevotella nigrescens, Gemella bergeri, and Fusobacterium nucleatum, has been identified as a potential biomarker for Colorectal Cancer.

NADH dehydrogenase (quinone) (Enzyme 1.6.99.5) plays a crucial role in electron transfer processes in various organisms. It aids in the production of energy by catalyzing the transfer of electrons from NADH to quinones. Quinone, also known as NADH dehydrogenase, is essential for cellular metabolism, especially in cancer cells. Studies have shown that NADH dehydrogenase is involved in electron transfer from NADH to quinones, impacting energy generation [121, 122]. Furthermore, mutations in mitochondrial DNA affecting NADH dehydrogenase have been linked to tumorigenesis, altering reactive oxygen species generation and apoptosis [109]. In colorectal cancer, the NAD+/NADH ratio increases during cancer progression, impacting oxidative stress levels [110]. Furthermore, the finding that overexpression of NADH dehydrogenase complex I is present in metastatic colon cancer cells implies that the enzyme is involved in the spread of cancer [111]. Furthermore, altered alcohol dehydrogenase and aldehyde dehydrogenase activities in the sera of individuals with colorectal cancer may operate as disease diagnostic indicators [112]. Mitochondrial function, including NADH dehydrogenase activity, influences the intrusion of colon cancer cells, with variations in TFAM, ND6, COX-II, NDUFA9, SDHB, UQCRC1/2, and cytochrome C oxidase subunits impacting cancer cell behavior [113]. Furthermore, the mitochondrial electron transport chain, where NADH dehydrogenase is a key player, is crucial regarding the cancer cells' energy metabolism, affecting drug resistance and autophagy [114]. In general, the growth and spread of colorectal cancer as well as other cancers are significantly influenced by NADH dehydrogenase (quinone) and the pathways that are linked to it. Comprehending the complex pathways associated with NADH dehydrogenase is imperative in the advancement of focused therapeutic interventions and diagnostic methodologies for the treatment of cancer.

The Enzyme NADH dehydrogenase (quinone) (EC 1.6.99.5), along with the presence of Slackia exigua, Prevotella nigrescens, and Fusobacterium nucleatum, has been pinpointed as a potential biomarker for Colorectal Cancer.

5.2 Enzyme Groups Based on Their Correlations with Species Levels

Fusobacterium nucleatum is linked to the progression and worsening of colorectal cancer by promoting inflammation, aiding tumor growth, and influencing the tumor microenvironment as mentioned above.

4-phospho-D-threonate 3-dehydrogenase (EC 1.1.1.408) is involved in metabolic pathways, affecting the biosynthesis of compounds like 2,4-dihydroxybutyric acid [115]. An important part of cancer biology is the metabolism. Synthetic metabolic pathways for biosynthesis can offer insights into potential intervention targets. Additionally, the expression of enzymes such as D-3-phosphoglycerate dehydrogenase can impact serine synthesis, which is relevant in cancer progression [116].

L-2-hydroxyglutarate dehydrogenase (L2HGDH) (EC 1.1.99.2) is vital for the metabolism of L-2-hydroxyglutarate (L-2HG), an oncometabolite associated with various cancers, comprising colorectal cancer. Research has shown that L-2HG can inhibit α-ketoglutarate-dependent enzymes, affecting cellular processes [117]. Moreover, changes in enzymes such as isocitrate dehydrogenase can change the tricarboxylic acid cycle, which can result in an overproduction of 2-hydroxyglutarate and aid in the occurrence of cancer [118].

The influence of **FMN reductase (EC 1.5.1.29)** enzyme on colorectal cancer has been mentioned above.

The effect of the enzyme **NADH dehydrogenase (quinone)(EC 1.6.99.5)** on colorectal cancer is explained above.

Ceramide glucosyltransferase (EC 2.4.1.80) (CGT) plays a crucial role in cellular processes, particularly in cancer. Research has demonstrated a correlation between changes in ceramide and glucosylceramide levels and the advancement of cancer [119]. In colorectal cancer, the enzyme CGT has been implicated in multidrug resistance mechanisms [120]. Moreover, it has been demonstrated that glucosylceramide synthase inhibition increases the anti-proliferative effects of chemotherapy on cancer cells. Ceramide glucosyltransferase activity is dependent on this enzyme [121].

1-phosphatidylinositol-4-phosphate 5-kinase (EC 2.7.1.68) (PIP5K) is an essential enzyme involved in lipid metabolism and cellular signaling pathways. Studies have indicated that PIP5Ks are essential for controlling a number of cellular functions, including autophagy, cell morphogenesis, and pain signaling [136, 137, 138]. In the context of cancer, the inhibition of PIP5Ks has been investigated as a potential therapeutic approach. Research has emphasized PIP5Ks' significance in conditions like cancer, neurodegeneration, and immunological disorders [123]. Moreover, the phosphorylation

status of PIP5K1C has been associated with invasive ductal carcinoma of the breast, suggesting a possible predictive value in cancer progression [125].

N-acylneuraminate-9-phosphatase (Ezyme 3.1.3.29) is also referred to as PRL-3 phosphatase and has been linked to the advancement of colorectal cancer. Research has indicated that PRL-3 expression is elevated in metastases of colorectal cancer patients relative to primary tumors. In colorectal cancer, this phosphatase is essential for facilitating cell invasion, proliferation, and metastasis [140, 141, 142]. Moreover, it has been discovered that PRL-3 is overexpressed in colorectal cancer liver metastases, highlighting its importance in the process of metastatic progression. Furthermore, PRL-3 has been linked to boosting cancer evolution in colorectal cancer cells via a number of different pathways, such as Csk down-regulation that activates Src [128].

The effect of the **levanase (EC 3.2.1.65)** enzyme on colorectal cancer has been stated above.

The impact of the **acylaminoacyl-peptidase (EC 3.4.19.1)** enzyme on colorectal cancer is referenced above.

Gingipain R (EC 3.4.22.37) a cysteine proteinase produced by Porphyromonas gingivalis, has been connected to several factors that contribute to the development of cancer and progression. Studies have shown that gingipains, including gingipain R, can activate matrix metalloproteinases like MMP9, promoting cancer invasiveness [129]. Furthermore, through the MAPK/ERK signaling pathway, gingipains have been connected to the growth of colorectal cancer cells; gingipain-deficient mutants exhibit decreased cancer cell proliferation [130]. Moreover, esophageal cancer has been linked to P. gingivalis, specifically gingipain R, which influences clinicopathological features and patient survival [131]. Gingipains have been reported to play a role in the growth and metastasis of cancer by encouraging the invasion of oral squamous cell carcinoma through the induction of proMMP9 and its activation [132]. Additionally, gingipains have been implicated in creating a cancer-promoting microenvironment, influencing tumor progression in oral cancer [133]. Overall, the data point to the critical role that gingipain R and other gingipains play in the initiation and spread of cancer, impacting mechanisms including invasion, proliferation, and the tumor microenvironment. Gaining insight into

the mechanisms via which gingipains support various cancer-related activities may help develop new therapeutic approaches that specifically target these virulence factors.

3-Aminobutyryl-CoA ammonia-lyase (EC 4.3.1.14) is essential in lysine fermentation, converting 3-aminobutyryl-CoA into crotonyl-CoA. This enzyme is crucial for the production of acetate, butyrate, and ammonia during the fermentation process [68]. This enzyme's involvement in lysine fermentation may have implications for colorectal cancer metabolism.

Holocytochrome-c synthase (HCCS) (EC 4.4.1.17) is a crucial enzyme responsible for attaching heme to cytochromes c, a process essential for mitochondrial function. The final byproduct of HCCS activity, cytochrome c, is involved in both apoptosis and oxidative phosphorylation [148, 149]. Prostaglandin E synthase (PGES) isomer activity in the context of colorectal cancer has been investigated to learn more about its relationship to clinicopathological variables and prognosis for patients [136]. Furthermore, by controlling signaling pathways, chondroitin synthase-1 (CHSY1) has been connected to encouraging proliferation and preventing apoptosis in colorectal cancer [137]. Additionally, fatty acid synthase (FASN) is frequently upregulated in colorectal cancer, suggesting its potential as a diagnostic tool for this particular kind of cancer [138]. Thymidylate synthase has been connected to drug resistance and has been linked to a poor prognosis in colorectal cancer patients [139]. In conclusion, heme's attachment to cytochromes c, which is necessary for mitochondrial activity, is facilitated by the important enzyme HCCS. Comprehending the expression and modulation of enzymes such as thymidylate synthase, PGES, CHSY1, FASN, and FASN offers important insights into the pathogenesis and possible targets for treatment in colorectal cancer.

Lysine 5,6-aminomutase (5,6-LAM) (EC 5.4.3.3) is an enzyme that is essential to the metabolism of lysine. In the context of colorectal cancer, lysine methyltransferase 9 (KMT9) has been found to be a key regulator of colorectal cancer [140]. Lysine methyltransferases, which mediate histone methylation, have been identified as a critical mechanism in controlling pathological processes in digestive malignancies, such as colorectal cancer [141]. In summary, lysine 5,6-aminomutase is a key enzyme in lysine metabolism, with intricate mechanisms involving radical intermediates and cofactors. Understanding its function and regulation can provide insights into both normal metabolic processes and potential implications in diseases like colorectal cancer.

Fusobacterium nucleatum has been identified as a possible biomarker for colorectal cancer along with the previously described enzymes.

Bacteroides fragilis (ETBF) is one of the bacteria linked to the development of colorectal cancer, has received considerable focus. The significance of ETBF in colorectal carcer has been shown by studies. Because ETBF can alter epithelial cells and regulate the immune system reaction, it has been associated with an increased risk of colorectal cancer [160, 161, 162]. The secretion of the B. fragilis toxin (BFT) by ETBF strains is a pivotal mechanism through which these bacteria contribute to cancer development [163, 164]. Moreover, dysbiosis resulting from an unbalanced microbiome in the intestinal tract, leading to elevated levels of bacteria like Bacteroides fragilis, has been related to a higher chance of colorectal cancer. This imbalance can disrupt the normal functions of the intestinal microbiota, potentially fostering carcinogenesis [147]. Additionally, the presence of ETBF has been linked to exacerbating tumorigenesis in colorectal cancer models [145].

UDP-N-acetylglucosamine 6-dehydrogenase (1.1.1.136), plays a vital role in producing UDP-N-acetylglucosamine, an essential metabolite in the hexosamine biosynthetic pathway (HBP). The transformation of UDP-N-acetylglucosamine into UDP-N-acetylglucosaminuronic acid is catalyzed by this enzyme [148]. Dysregulation of the hexosamine biosynthesis pathway enzymes, such as UDP-N-acetylglucosamine 6-dehydrogenase, can affect glycosylation processes that are essential for the advancement of colorectal cancer. Upregulation of enzymes like UDP-N-acetylglucosamine pyrophosphorylase may contribute to altered glycosylation patterns observed in colorectal cancer cells [149].

GDP-4-dehydro-6-deoxy-D-mannose reductase (EC 1.1.1.281) is a key enzyme in the production of GDP-L-fucose, a crucial nucleotide sugar required for numerous glycosylation processes. Key to the hexosamine biosynthesis pathway, this enzyme catalyzes the conversion of GDP-4-keto-6-deoxy-D-mannose to GDP-L-fucose [150]. Glycosylation pattern changes seen in a variety of biological settings, including cancer, have been linked to changes in enzymes involved in GDP-L-fucose production, such as GDP-4-keto-6-deoxy-D-mannose epimerase/reductase. Dysregulation of GDP-L-fucose biosynthesis-related enzymes, such as GDP-4-dehydro-6-deoxy-D-mannose reductase, may impact glycosylation processes that are essential for the advancement of colorectal

cancer. Studies have indicated that deficiencies in fucosylation, a process dependent on GDP-L-fucose, can lead to colitis and adenocarcinoma [151]. The dysregulation of enzymes responsible for GDP-L-fucose synthesis has been linked to changes in Notch signaling, a pathway implicated in colorectal cancer development [152].

3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring) (EC 1.2.4.4) is a key player in colorectal cancer due to its involvement in branched-chain amino acid metabolism. This enzyme is particularly significant in colorectal cancer as it interacts with lactate dehydrogenase [153]. Dysregulation of enzymes engaged in metabolizing 3-methyl-2-oxobutanoate, such as dehydrogenases has been linked to colorectal cancer's advancement [154].

tRNA (cytosine-5-)-methyltransferase (EC 2.1.1.29) performs a crucial function in the development of colorectal cancer by modifying RNA [155]. Enhanced gene copy numbers and elevated protein expressions of tRNA (cytosine-5-)-methyltransferase have been described in colorectal cancer, one of the human cancers where dysregulation of tRNA methyltransferases, such as NSUN2, has been seen. [156].

N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase (EC 2.4.1.289) is an enzyme , which is involved in glycosylation activities, has been linked to the advancement of colorectal cancer. Research has demonstrated that the enzyme b1,6-N-acetylglucosaminyl transferase V (GNT-V) is overexpressed in liver metastases and original colorectal tumors, suggesting a possible function for GNT-V in the proliferation of cancer [157]. This enzyme has been addressed for biomarker identification because it is known to be increased in malignant cancer cells [158].

dTDP-4-amino-4,6-dideoxy-D-glucose transaminase (EC 2.6.1.33) is a crucial component in the synthesis of various bioactive molecules, and is of interest in the context of colorectal cancer, which involves several biochemical pathways [159].

D-glycero-alpha-D-manno-heptose-7-phosphate kinase (EC 2.7.1.168) plays a vital role in producing heptose derivatives, key components of lipopolysaccharides (LPS) in various bacteria. This enzyme phosphorylates D-glycero-D-manno-heptose-7-phosphate to create intermediates such as D-glycero-D-manno-heptose-1,7-bisphosphate. Research indicates that in bacteria like E. coli, it functions as part of a bifunctional protein,

serving both as a kinase and an adenylyltransferase [181, 182, 183]. The impact of E. coli on colorectal cancer has been elucidated above.

Acetylxylan esterase Akt2 (EC 3.1.1.72) overexpression has been associated with advanced colorectal cancer stages and the formation of metastases [163].

Protein O-GlcNAcase (EC 3.2.1.169) is an enzyme that removes O-linked β-N-acetylglucosamine (O-GlcNAc) from modified proteins, has attracted considerable interest in cancer research, especially in studies of colorectal cancer. Research has demonstrated that O-GlcNAcase plays a critical role in cancer malignancy and chemotherapy resistance. The dysregulation of O-GlcNAcase has been associated with altered protein O-GlcNAcylation levels, impacting various cellular processes in cancer cells. Moreover, O-GlcNAcase's potential as a therapeutic target has been illuminated by its association with radioresistance and DNA repair processes in colorectal cancer [185, 186, 187].

Cu+-exporting ATPase (EC 3.6.3.54) are implicated in the intricate mechanisms of colorectal cancer, encompassing the dysregulation of copper transport proteins such as these ATPases [167]. In the context of colorectal cancer, there is evidence of altered ATPase activity. Research has shown a notable reduction in Na^+/K^+ ATPase functioning in samples of colorectal tumors influenced by NAADP [168]. Furthermore, the up-regulation of the Na^+,K^+-ATPase α3-isoform has been observed in human colorectal cancer cells, potentially compensating for the decreased expression of the α1-isoform [169]. Moreover, the use of copper nanoparticles as radiosensitizers for colorectal carcinoma treatment has shown promise [170]. These nanoparticles induce autophagy in cancer cells, providing insights into their mechanism of action for cancer treatment. Additionally, the knockdown and overexpression of clusterin have been found to modulate Cu-ATPase-mediated copper export capacity in cells [171].

UDP-N-acetylglucosamine 4,6-dehydratase (configuration-retaining) (EC 4.2.1.135) is an essential enzyme for the production of several sugar compounds. Studies have emphasized the importance of enzymes like this in bacterial proteins [172]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of UDP-N-acetylglucosamine 4,6-dehydratase on colorectal cancer.

Bacteroides fragilis has been identified as a possible biomarker for colorectal cancer along with the previously described enzymes.

The effect of the **Prevotella nigrescens** strain on colorectal cancer is described above.

The impact of the **4-phospho-D-threonate 3-dehydrogenase (Enzyme 1.1.1.408), L-2-hydroxyglutarate dehydrogenase (Enzyme 1.1.99.2), NADH dehydrogenase (quinone) (Enzyme 1.6.99.5), 3-keto-5-aminohexanoate cleavage (Enzyme 2.3.1.247), N-acylneuraminate-9-phosphatase (Ezyme 3.1.3.29), levanase (Enzyme 3.2.1.65), acylaminoacyl-peptidase (Enzyme 3.4.19.1), gingipain R (Enzyme 3.4.22.37)** and **Holocytochrome-c synthase (Enzyme 4.4.1.17)** on colorectal cancer is referenced above.

Prolyl oligopeptidase (POP) (Enzyme 3.4.21.26) is a serine endopeptidase that has been implicated to a number of illnesses, such as colorectal cancer. Fibroblast activation protein (FAP) and POP are frequently overexpressed in cancers arising from epithelial cells. The specificity with which cancer stromal fibroblasts produce FAP points to FAP as a possible target for therapeutic intervention in the treatment of cancer [173]. Studies have shown that POP is involved in inflammatory diseases, including atherosclerosis, indicating its role in immune system modulation [174].

Prevotella_nigrescens has been identified as a possible biomarker for colorectal cancer along with the previously described enzymes.

Porphyromonas asaccharolytica is a species within the Porphyromonas genus has been linked to the occurrence of colorectal carcinoma. Research has indicated that patients with colorectal cancer have considerably higher concentrations of Porphyromonas species, such as P. asaccharolytica, in their mucosal and specimens from stool compared to control patients [175]. Additionally, a possible correlation between the Porphyromonas genus and the Porphyromonadaceae family and colorectal carcinoma has been further suggested by metagenomic studies that indicate an overrepresentation of these organisms in colorectal cancer cases relative to normal specimens [197, 198]. Porphyromonas species have also been discovered to be viable diagnostic indicators for colorectal cancer (CRC), with P. asaccharolytica being one of the bacteria that are detected in notably high abundance in tumoral tissues and fecal samples of individual

with CRC [199, 200]. Furthermore, pancreatic, colorectal, and orodigestive cancers have all been connected to Porphyromonas species [180].

4-hydroxybutyrate dehydrogenase (Enzyme 1.1.1.61) is an enzyme involved in the catabolism of 4-hydroxybutyrate, and has also been investigated in different biological pathways [181]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of 4-hydroxybutyrate dehydrogenase on colorectal cancer.

Quinol-cytochrome-c reductase's (Enzyme 1.10.2.2) role in the electron transport chain is a critical factor in the apparition of cancer [182]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of Quinol-cytochrome-c reductase on colorectal cancer.

Glycine dehydrogenase (cyanide-forming)- HCN synthase (Enzyme 1.4.99.5) is a crucial enzyme that facilitates the replacement of glycine to produce cyanide, a process relevant to cancer metabolism [183]. Studies have shown that glycine metabolism, particularly in conjunction with serine, is essential for cancer cells, providing precursors for macromolecules and antioxidant defense mechanisms [184].

Protein-glutamine gamma-glutamyltransferase (Enzyme 2.3.2.13) is significant for the advancement of cancer of the gastrointestinal tract. Studies have revealed that enzymes such as GGT1 and GGT5, which belong to the gamma-glutamyltransferase family, are crucial in promoting cancer development, enhancing drug resistance, and participating in redox processes [208, 209]. In colorectal carcinoma, glutamate metabolism is critical to cell survival and proliferation. Research has demonstrated that blocking the metabolism of glutamine can impede the combination and invasion of tumor cells [187]. The importance of glutamine in cancer metabolism has been highlighted by the discovery that it is a necessary substrate for the TCA cycle in colorectal tumor cells [188]. Furthermore, it has been discovered that SIRT4's suppression of glutamine metabolism increases colorectal cancer cells' susceptibility to chemotherapy [189]. Overall, the dysregulation of protein-glutamine gamma-glutamyltransferase enzymes and glutamine metabolism in colorectal cancer underscores their importance as potential therapeutic targets.

Cyclomaltodextrin glucanotransferase (Enzyme 2.4.1.19) contributes indirectly to colorectal cancer treatment by producing cyclodextrins that enhance drug delivery and efficacy through improved solubility, stability, and targeted therapy [190].

Dolichyl-phosphate beta-D-mannosyltransferase (DPMT) (Enzyme 2.4.1.83) is a key enzyme involved in protein glycosylation processes, in particular when mannose residues are transferred to proteins [191]. Genes related to DPMT like DPM2 have been associated with colorectal cancer survival [192] and identified as unfavorable prognostic markers in other types of cancer [193].

Phosphatidylcholine synthase (Enzyme 2.7.8.24) is a key player in colorectal cancer, contributing to cancer cell resistance to chemotherapy drugs like oxaliplatin. Abnormal phosphatidylcholine biosynthesis, facilitated by enzymes such as lysophosphatidylcholine acyltransferase 2 (LPCAT2), has been linked to chemoresistance in colorectal cancer [194]. Furthermore, lysophosphatidylcholine acyltransferase 1 (LPCAT1) is significantly overexpressed in colorectal adenocarcinomas, suggesting its potential involvement in colorectal cancer pathogenesis [195].

Butyrate-acetoacetate CoA-transferase (Enzyme 2.8.3.9) is a crucial enzyme in butyrate synthesis, particularly in the final step of the process. This enzyme helps convert butyryl-CoA to butyrate in the human the gastrointestinal tract, where it is mostly used by bacteria that produce butyrate [219, 220, 221]. Numerous investigations have found the gene encoding this enzyme, underscoring its importance in the gut microbiota [199]. Additionally, the butyrate-acetoacetate CoA-transferase pathway, utilizing acetate as a cosubstrate, has been proposed as a major pathway by which the gut ecology produces butyrate [200]. Studies have demonstrated a possible connection between dysbiosis and disease states by linking changes in butyryl-CoA transferase gene expression to diseases such as colorectal cancer [201]. Understanding the regulation of butyrate oxidation, which can be influenced by histone deacetylases, is crucial in comprehending the protective effects of butyrate in colorectal cancer [202].

The enzymes 1-alkyl-2-acetylglycerophosphocholine esterase (EC 3.1.1.47), acetyl-CoA hydrolase (EC 3.1.2.1), 2-phosphosufolactate phosphatase (EC 3.1.3.71) and guanosine-3',5'-bis(diphosphate) 3'-diphosphatase (EC 3.1.7.2) are enzymes that act

specifically on various ester bonds under EC 3.1 within the class of hydrolases. These enzymes stimulate the hydrolysis of ester bonds to break them. Ester bond-acting enzymes are essential to many biological processes, involving the initiation and dissemination of malignancies like colorectal cancer (CRC) [203].

Streptopain (Enzyme 3.4.22.10) also known as SpeB, is a cysteine protease produced by Streptococcus pyogenes, and it has been implicated in various pathogenic processes. Streptopain, along with other bacterial proteases, can obstruct the host's intrinsic defense mechanism by affecting processes like kinin release [204]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of Streptopain on colorectal cancer.

Mitochondrial processing peptidase (MPP) (Enzyme 3.4.24.64) is a crucial enzyme responsible for cleaving mitochondrial targeting sequences from proteins destined for the mitochondria [205]. In the context of colorectal cancer, mitochondrial function, including histone acetylation, has been implicated in the disease. Mitochondria have been reported to elevate H3K27ac marks in colorectal cancer, suggesting a potential link between mitochondrial processes and cancer progression [206].

Beta-alanyl-CoA ammonia-lyase (EC 4.3.1.6) also referred to as Acl. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of beta-alanyl-CoA ammonia-lyase on colorectal cancer.

Methylmalonyl-CoA epimerase (EC 5.1.99.1) is a key enzyme involved in various metabolic pathways, like the process that transforms propionyl-CoA into succinyl-CoA [207]. In reference to colorectal cancer, understanding the importance of methylmalonyl-CoA epimerase is significant due to its involvement in intermediary metabolism [208]. Alterations in metabolic pathways, including those involving methylmalonyl-CoA epimerase, can have implications for tumorigenesis. Even though the references do not specifically mention a direct association between methylmalonyl-CoA epimerase and colon cancer, it is important to comprehend the enzyme's function in metabolic processes and the possibility of its dysregulation in disease states in order to guide future research in the field of cancer research.

Maleylacetoacetate isomerase (EC 5.2.1.2) also known as GSTZ1, is essential to the breakdown of tyrosine. In the context of colorectal cancer (CRC), changes to metabolic processes, including those involving amino acids like tyrosine, can have significant impacts on cancer cell growth and proliferation [209].

Vinylacetyl-CoA DELTA-isomerase (EC 5.3.3.3) is essential in fatty acid metabolism [210]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of vinylacetyl-CoA DELTA-isomerase on colorectal cancer.

Porphyromonas asaccharolytica has been identified as a possible biomarker for colorectal cancer along with the previously described enzymes.

The effect of the **Slackia exigua** on colorectal cancer has been stated above.

D-lactate dehydrogenase (cytochrome) (EC 1.1.2.4) is significant in colorectal cancer, as research has shown that its levels, along with other forms of lactate dehydrogenase, correlate with the prognosis of patients with this disease [211].

Sarcosine reductase (EC 1.21.4.3), a derivative of glycine, has been identified as a differential metabolite significantly increased in colorectal cancer. Metabolomic studies have revealed elevated levels of sarcosine in colorectal tumors, suggesting its involvement in tumorigenesis [212]. Studies have shown that sarcosine and related molecules could potentially serve as biomarkers for early colorectal cancer [213].

Betaine reductase (EC 1.21.4.4) has been a subject of interest in the context of colorectal cancer due to its potential implications in cancer risk. Studies have explored the association between plasma betaine levels and colorectal carcinogenesis, with limited research available on this specific relationship [214].

The effect of the enzyme **NADH dehydrogenase (quinone)(EC 1.6.99.5)** on colorectal cancer is explained above.

The effect of the enzyme **1-phosphatidylinositol-4-phosphate 5-kinase (EC 2.7.1.68)** on colorectal cancer is explained above.

The effect of the enzyme **N-acylneuraminate-9-phosphatase (Enzyme 3.1.3.29)** on colorectal cancer is explained above.

The effect of the **levanase enzyme (EC 3.2.1.65)** on colorectal cancer has been stated above.

The effect of the **prolyl oligopeptidase (POP) enzyme (Enzyme 3.4.21.26)** on colorectal cancer has been stated above.

The effect of the **gingipain R enzyme (EC 3.4.22.37)** on colorectal cancer has been stated above.

Hydroxymethylglutaryl-CoA lyase (HMGCL) (EC 4.1.3.4) plays a significant role in cancer metabolism, particularly in colorectal cancer [215]. Additionally, HMGCL has been implicated in hepatocellular carcinoma (HCC) suppression through β-hydroxybutyrate production, indicating its potential role in inhibiting cancer progression [216].

Along with the previously mentioned enzymes, Slackia_exigua has been found as a potential diagnostic biological marker for colorectal cancer.

Escherichia coli is a prevalent bacterium in the human digestive system, is linked with the emergence of colon tumor, as research has shown a significant connection between E. coli adhering to the mucosa and the disease. [217]. A considerable proportion of individuals with colorectal cancer contain E. coli strains that produce colibitin, indicating a possible role in oncogenesis [218]. The association between precancerous and cancerous colorectal tissues and colibactin-producing E. coli suggests a potential role for this bacteria in fostering colorectal carcinogenesis [219]. E. Coli strains that produce a lot of colibactin have also been linked to colorectal tumor formation, underscoring their potential impact on cancer development [142]. Furthermore, enteropathogenic E. coli (EPEC), a kind of mucosa-associated E. coli, has been linked to the emergence of colon tumorigenesis, suggesting involvement in the molecular pathways of colorectal cancer [220]. Escherichia coli has been identified as a significant microbial player in colorectal cancer oncogenesis due to its prevalence in colorectal cancer tissues and its role in tumorigenesis in animal models [221]. The course of the tumor's development linked to colitis is significantly afflicted by chronic inflammation in the colon, which is made

possible by E. coli colonization [222]. In conclusion, it seems that Escherichia coli, especially the types that produce colibactin, is a fundamental feature in the emergence of colorectal carcinoma. Understanding the mechanisms by which E. coli contributes to colorectal oncogenesis is essential for developing targeted interventions to prevent and treat this prevalent form of cancer.

L-threonine 3-dehydrogenase (EC 1.1.1.103), 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (EC 1.1.1.127), 3-dehydro-L-gulonate 2-dehydrogenase (EC 1.1.1.130), GDP-mannose 6-dehydrogenase (EC 1.1.1.132), hydroxyphenylpyruvate reductase (EC 1.1.1.237), quinate/shikimate dehydrogenase [NAD(P)+] (EC1.1.1.282), 2-hydroxymethylglutarate dehydrogenase (EC 1.1.1.291), 3-hydroxypropionate dehydrogenase (NADP+) (EC 1.1.1.298), ureidoglycolate dehydrogenase (NAD+) (EC 1.1.1.350), sulfolactaldehyde 3-reductase (EC 1.1.1.373), 3-hydroxy acid dehydrogenase (EC 1.1.1.381), pyridoxine 4-dehydrogenase (EC 1.1.1.65), mannitol 2-dehydrogenase (EC 1.1.1.67), glycerol-3-phosphate dehydrogenase (NAD+) (EC 1.1.1.8) and D-malate dehydrogenase (decarboxylation) (EC 1.1.1.83) enzymes are a subgroup of NAD+ or NADP+ as acceptors. NAD+ and NADP+ play crucial roles in colorectal cancer. Research has demonstrated that the NAD biosynthesis pathway facilitated by Nampt is operational in gastrointestinal tumor cells, underscoring the significance of NAD in the advancement of malignancy [223].

Glycerol-3-phosphate dehydrogenase (GPDH) (EC 1.1.5.3) is a key enzyme involved in various biological processes, including cancer. Research has demonstrated its significance in cancer biology, particularly in colorectal cancer. In colorectal cancer, GPDH has been found to be upregulated, indicating its role in supplying glycerol 3-phosphate for lipid biosynthesis, essential for cancer cell growth [224]. The observed metabolic alterations in cancer cells could perhaps be attributed to this deregulation of GPDH activity.

Malate dehydrogenase (quinone) (EC 1.1.5.4) plays a vital involvement in cancer metabolism and is affiliated with multiple cancer types. In bacteria such as Escherichia coli and Corynebacterium glutamicum, malate dehydrogenase has been studied, emphasizing its importance in cellular metabolism [225]. MDH1 has been connected to promoting glycolysis in actively proliferating cells, consisting of tumor cells in the context of cancer [226].

Choline dehydrogenase (EC 1.1.99.1) vital in the choline-glycine betaine pathway, has connections to multiple types of cancers, including gastrointestinal cancer. Research findings have indicated that choline dehydrogenase is essential in conferring osmotic tolerance in Escherichia coli [227]. Colorectal cancer stem cells, characterized by specific markers like CD44 and aldehyde dehydrogenase, have been determined to represent targets for efficient therapy of colorectal carcinoma [228]. To comprehend the intricacy of the disease, researchers have looked into the biological traits of tumor-generating subpopulations in colon cancer, such as cellular plasticity and aldehyde dehydrogenase enzyme activity [229].

Catalase-peroxidase (EC 1.11.1.21), catalase (EC 1.11.1.6), glutathione peroxidase (EC 1.11.1.9) enzymes are a subgroup of **Peroxidases (EC 1.11.1). Peroxidases**, including **glutathione peroxidase (GPx),** have been a subject of relevance for the study of carcinoma of the colon due to their role in combating oxidative stress, a major contributor to cancer development [230]. Studies have shown that selenium, a key component of selenoenzymes like GPx, may have a chemopreventive effect on colorectal cancer through its antioxidative properties [231]. Furthermore, the upregulation of GPx has been linked to protecting colorectal cells from oxidative damage induced by reactive oxygen species [232].

Hydrogenase (acceptor) (EC 1.12.99.6) is an enzyme responsible for converting molecular hydrogen into protons and electrons reversibly, plays roles in multiple biological processes, including cancer. Its involvement is especially significant in colorectal cancer, where research indicates that the activity of hydrogenase may contribute to the prevention of cancerous transformations [233].

Stizolobate synthase (EC 1.13.11.29) is an enzyme that breaks specific carbon-carbon bonds in dihydroxyphenylalanine, leading to the creation of certain acids [234]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of stizolobate synthase on colorectal cancer.

Taurine dioxygenase (EC 1.14.11.17), [50S ribosomal protein L16]-arginine 3-hydroxylase (EC 1.14.11.47), anthranilate 1,2-dioxygenase (deaminating, decarboxylating) (EC 1.14.12.1), nitric oxide dioxygenase(EC 1.14.12.17), 3-phenylpropanoate dioxygenase (EC 1.14.12.19), limonene 1,2-monooxygenase (EC 1.14.13.107), carnitine monooxygenase (EC 1.14.13.239), L-lysine N6-monooxygenase (NADPH) (EC 1.14.13.59), (S)-limonene 6-monooxygenase (EC

1.14.14.51) and pyrimidine oxygenase (EC 1.14.99.46) enzymes classified under EC 1.14 are known as "Acting on paired donors, with incorporation or reduction of molecular oxygen (oxygenases)". Heme oxygenase-1 (HO-1) in particular is an oxygenase that is important in colorectal cancer. Research has indicated that HO-1's involvement of p53 confers antitumoral effects in colorectal cancer [235]. Furthermore, in colorectal cancer cells, overexpression of HO-1 can cause ferroptosis and ultimately result in cancer cell death [236].

Ferric-chelate reductase (NADPH) (EC 1.16.1.9) is an enzyme that helps to maintain iron balance and aids in iron absorption [237]. Posttranscriptional regulation of ferric chelate reductase highlights its significance in adapting to iron deficiency conditions [238]. The emergence of colorectal carcinoma is significantly influenced by iron, and the effects of different types of iron vary. Compared to other types of iron, heme iron in particular has been proposed to have a more substantial part in the progression of colon carcinoma [239].

Formate dehydrogenase (EC 1.17.1.9) has been pertained to colon cancer research, particularly in the context of identifying potential markers for cancer stemness. Furthermore, the growth of colorectal cancer has been connected to the activity of two important enzymes involved in alcohol metabolism: aldehyde dehydrogenase and alcohol dehydrogenase [240]. Moreover, lactate dehydrogenase (LDH) levels have been investigated for their prognostic value in colorectal cancer patients. Studies have shown that elevated LDH levels may have implications for patient outcomes and treatment responses [211]. In conclusion, formate dehydrogenase and related enzymes play significant roles in the molecular pathways and potential therapeutic targets in colorectal cancer.

(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin) (EC 1.17.7.1) is also known as GcpE or IspG. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of stizolobate synthase on colorectal cancer.

Ferredoxin-NAD+ reductase (EC 1.18.1.3), flavodoxin-NADP+ reductase (EC 1.19.1.1), acetaldehyde dehydrogenase (acetylating) (EC 1.2.1.10), succinate-semialdehyde dehydrogenase [NAD(P)+] (EC 1.2.1.16), aminobutyraldehyde dehydrogenase (EC 1.2.1.19), glycolaldehyde dehydrogenase (EC 1.2.1.21), aldehyde dehydrogenase [NAD(P)+] (EC 1.2.1.5), succinylglutamate-semialdehyde dehydrogenase (EC 1.2.1.71), erythrose-4-phosphate dehydrogenase (EC 1.2.1.72),

betaine-aldehyde dehydrogenase (EC 1.2.1.8), L-glutamate gamma-semialdehyde dehydrogenase (EC 1.2.1.88), oxoglutarate dehydrogenase (succinyl-transferring) (EC 1.2.4.2), pyruvate dehydrogenase (quinone) (EC 1.2.5.1), carboxylate reductase (EC 1.2.99.6), 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EC 1.3.1.28), acrylyl-CoA reductase (NADPH) (EC 1.3.1.84), 3-(cis-5,6-dihydroxycyclohexa-1,3-dien-1-yl)propanoate dehydrogenase (EC 1.3.1.87), coproporphyrinogen oxidase (EC 1.3.3.3), protoporphyrinogen IX dehydrogenase (quinone) (EC 1.3.5.3), crotonobetainyl-CoA reductase (EC 1.3.8.13), glutamate dehydrogenase (NADP+) (EC 1.4.1.4), glycine oxidase (EC 1.4.3.19), primary-amine oxidase (EC 1.4.3.21), D-arginine dehydrogenase (EC 1.4.99.6), FMN reductase (NADH) (EC 1.5.1.42), electron-transferring-flavoprotein dehydrogenase (EC 1.5.5.1), NAD(P)+ transhydrogenase (Si-specific) (EC 1.6.1.1), NAD(P)H dehydrogenase (quinone) (EC 1.6.5.2), NADPH:quinone reductase (EC 1.6.5.5), trimethylamine-N-oxide reductase (EC 1.6.6.9), nitrite reductase (NADH) (EC 1.7.1.15), nitrite reductase [NAD(P)H] (EC 1.7.1.4), trimethylamine-N-oxide reductase (EC 1.7.2.3), nitrate reductase (quinone) (EC 1.7.5.1), nitrate reductase (EC 1.7.99.4), assimilatory sulfite reductase (NADPH) (EC 1.8.1.2), protein-disulfide reductase (EC 1.8.1.8), respiratory dimethylsulfoxide reductase (EC 1.8.5.3), glutathionyl-hydroquinone reductase (EC 1.8.5.7), selenate reductase (EC 1.97.1.9) enzymes classified under EC 1 are known as Oxidoreductases. The oxidoreductases that play a major part in colorectal cancer include those that are overexpressed in a variety of malignancies. Furthermore, in colorectal cancer, xanthine oxidoreductase levels have been related to malignancy growth and metastasis [241]. Overall, the pathophysiology and advancement of colorectal cancer may be aided by the dysregulation of a number of oxidoreductases, including xanthine oxidoreductase, microsomal epoxide hydrolase, and NQO1. Knowing how these oxidoreductases function in colorectal cancer may help identify new treatment targets and indicators of the illness.

23S rRNA (uridine2552-2'-O)-methyltransferase (EC 2.1.1.166), 16S rRNA (guanine527-N7)-methyltransferase (EC 2.1.1.170), 16S rRNA (guanine1207-N2)-methyltransferase (EC 2.1.1.172), 23S rRNA (guanine1835-N2)-methyltransferase (EC 2.1.1.174), 23S rRNA (adenine1618-N6)-methyltransferase (EC 2.1.1.181), tRNA (cytidine32/uridine32-2'-O)-methyltransferase (EC 2.1.1.200), 2-methoxy-6-polyprenyl-1,4-benzoquinol methylase (EC 2.1.1.201), 2-polyprenyl-6-hydroxyphenol methylase (EC 2.1.1.222), tellurite methyltransferase (EC 2.1.1.265),

ribosomal protein L3 N5-glutamine methyltransferase (EC 2.1.1.298), tRNA (uracil54-C5)-methyltransferase (EC 2.1.1.35), tRNA 5-(aminomethyl)-2-thiouridylate-methyltransferase (EC 2.1.1.61), protein-L-isoaspartate(D-aspartate) O-methyltransferase (EC 2.1.1.77) enzymes classified under EC 2.1.1 are known as Methyltransferases. Methyltransferases regulate a number of activities, including DNA repair, gene expression, and epigenetic changes, which is important in colorectal cancer. Genes such as p16, THBS1, and hMLH1 are methylated in colorectal cancers that exhibit the CpG island methylator phenotype (CIMP) [242]. Histone methyltransferases like EZH2 are known to repress gene expression in digestive cancers, including colorectal cancer [141]. Additionally, the m6A methyltransferase METTL3 has been implicated in promoting cancer progression [243]. METTL3 has been found to mediate chemotherapy resistance in colorectal cancer by enhancing m6A modification levels [244]. Furthermore, it has been determined that m6A methyltransferase KIAA1429 is a carcinogenic contributor to colorectal cancer [245].

Arginine N-succinyltransferase (EC 2.3.1.109), N-hydroxyarylamine O-acetyltransferase (EC 2.3.1.118), glycerol-3-phosphate 1-O-acyltransferase (EC 2.3.1.15), acetyl-CoA C-acyltransferase (EC 2.3.1.16), UDP-3-O-(3-hydroxyacyl)glucosamine N-acyltransferase (EC 2.3.1.191), tRNAMet cytidine acetyltransferase (EC 2.3.1.193), dTDP-4-amino-4,6-dideoxy-D-galactose acyltransferase (EC 2.3.1.210), Kdo2-lipid IVA palmitoleoyltransferase (EC 2.3.1.242), acyl-Kdo2-lipid IVA acyltransferase (EC 2.3.1.243), lipid IVA palmitoyltransferase (EC 2.3.1.251), [ribosomal protein S5]-alanine N-acetyltransferase (EC 2.3.1.267), acyl-[acyl-carrier-protein]-phospholipid O-acyltransferase (EC 2.3.1.40), beta-ketoacyl-[acyl-carrier-protein] synthase I (EC 2.3.1.41), homoserine O-succinyltransferase (EC 2.3.1.46), dihydrolipoyllysine-residue succinyltransferase (EC 2.3.1.61), gamma-glutamyltransferase (EC 2.3.2.2), citrate synthase (unknown stereospecificity) (EC 2.3.3.16), malate synthase (EC 2.3.3.9) enzymes classified under EC 2.3 are known as Acyltransferases. Acyltransferases play a crucial role in colorectal cancer by influencing various pathways connected to chemoresistance and metabolism of lipids. Researches have highlighted the significance of specific acyltransferases in colorectal carcinoma. For example, in colorectal cancer, there is an upregulation of Agpat4, an acyltransferase that converts lysophosphatidic acid (LPA) into phosphatidic acid, indicating an inadequate prognosis [246]. Furthermore, lysophosphatidylcholine acyltransferase 2 LPCAT2-induced lipid droplet biogenesis

impedes cell death pathways, contributing to chemoresistance in colorectal cancer [247]. Moreover, lysophosphatidylcholine acyltransferase 1 (LPCAT1) and lysophosphatidylcholine acyltransferase 4 (LPCAT4) have been associated with colorectal cancer progression and lipid accumulation in cancer cells [248]. These acyltransferases are involved in the conversion of lysophosphatidylcholine into phosphatidylcholine, impacting lipid metabolism in cancer cells. Furthermore, in a number of cancer forms, including colorectal cancer, acylglycerol-3-phosphate acyltransferase 2 (AGPAT2) has been connected to lipid droplet buildup and the longevity of cells [249]. Additionally, it has been demonstrated that the dysregulation of acyltransferases, such as LPCAT2, increases the formation of lipid droplets, which aids in the chemotherapy resistance of colorectal cancer cells [194]. This dysregulation affects pathways related to caspase activation, ER stress, and immunogenic cell death, ultimately influencing the response to chemotherapy in colorectal cancer. In conclusion, acyltransferases, particularly LPCAT1, LPCAT2, LPCAT4, and AGPAT2, play critical roles in lipid metabolism, chemoresistance, and cancer progression in colorectal cancer. Understanding the impact of these acyltransferases on cellular pathways can provide insights into potential therapeutic targets and strategies to overcome chemoresistance in colorectal cancer.

Cellulose synthase (UDP-forming) (EC 2.4.1.12), alpha,alpha-trehalose-phosphate synthase (UDP-forming) (EC 2.4.1.15), lipopolysaccharide N-acetylmannosaminouronosyltransferase (EC 2.4.1.180), lipopolysaccharide glucosyltransferase I (EC 2.4.1.58), guanosine phosphorylase(EC 2.4.2.15), thymidine phosphorylase (EC 2.4.2.4), lipid IVA 4-amino-4-deoxy-L-arabinosyltransferase (EC 2.4.2.43), (Kdo)-lipid IVA 3-deoxy-D-manno-octulosonic acid transferase (EC 2.4.99.13) enzymes classified under EC 2.4 are known as Glycosyltransferases. Glycosyltransferases play a vital role in colorectal cancer (CRC) by influencing various aspects of tumor progression and prognosis. Studies have identified specific glycosyltransferase enzymes like GCNT3, FUT4, FUT9, and β3GnT8 that are associated with CRC development and outcomes. It has been suggested that these enzymes are indicators for CRC chemoresistance and prognosis [296, 297]. Additionally, alterations in glycosyltransferase gene expression have been linked to poor prognostic CRC subtypes, particularly those connected to a lack of mismatch repair and insufficient glycan biosynthesis [252]. In conclusion, dysregulated glycosyltransferases and aberrant

glycosylation patterns are key factors in CRC pathogenesis, influencing tumor progression, prognosis, and chemoresistance.

Heme o synthase (EC 2.5.1.141), glutathione transferase (EC 2.5.1.18), 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase (EC 2.5.1.64), adenosyl-chloride synthase (EC 2.5.1.94), aromatic-amino-acid transaminase (EC 2.6.1.57), valine-pyruvate transaminase (EC 2.6.1.66), putrescine-2-oxoglutarate transaminase (EC 2.6.1.82), methionine transaminase (EC 2.6.1.88), glycerate 2-kinase (EC 2.7.1.165), sulfofructose kinase (EC 2.7.1.184), phosphoribulokinase (EC 2.7.1.19), protein-Npi-phosphohistidine-N-acetylmuramate phosphotransferase (EC 2.7.1.192), protein-Npi-phosphohistidine-L-ascorbate phosphotransferase (EC 2.7.1.194), protein-Npi-phosphohistidine-D-mannitol phosphotransferase (EC 2.7.1.197), protein-Npi-phosphohistidine-trehalose phosphotransferase (EC 2.7.1.201), protein-Npi-phosphohistidine-D-fructose phosphotransferase (EC 2.7.1.202), L-fuculokinase (EC 2.7.1.51), L-xylulokinase (EC 2.7.1.53), 2-dehydro-3-deoxygalactonokinase (EC 2.7.1.58), N-acylmannosamine kinase (EC 2.7.1.60), inosine kinase (EC 2.7.1.73), pseudouridine kinase (EC 2.7.1.83), thiamine kinase (EC 2.7.1.89), [Isocitrate dehydrogenase (NADP+)] kinase (EC 2.7.11.5), propionate kinase (EC 2.7.2.15), ribose 1,5-bisphosphate phosphokinase (EC 2.7.4.23), Kdo2-lipid A phosphotransferase (EC 2.7.4.29), nicotinamide-nucleotide adenylyltransferase (EC 2.7.7.1), polynucleotide adenylyltransferase (EC 2.7.7.19), (2,3-dihydroxybenzoyl)adenylate synthase (EC 2.7.7.58), citrate lyase holo-[acyl-carrier protein] synthase (EC 2.7.7.61), sulfur carrier protein ThiS adenylyltransferase (EC 2.7.7.73), molybdenum cofactor cytidylyltransferase (EC 2.7.7.76), phosphatidylglycerol-membrane-oligosaccharide glycerophosphotransferase (EC 2.7.8.20), alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase (EC 2.7.8.37), Kdo2-lipid A phosphoethanolamine 7''-transferase (EC 2.7.8.42), molybdopterin synthase (EC 2.8.1.12), 3-mercaptopyruvate sulfurtransferase (EC 2.8.1.2), CoA:oxalate CoA-transferase (EC 2.8.3.19), L-carnitine CoA-transferase (EC 2.8.3.21) enzymes classified under EC 2 are known as Transferases. Polypeptide N-acetylgalactosaminyltransferase isozymes T1 and T2, which are enzymes like UDP-GalNAc, are important in the development of colorectal cancer. Studies have demonstrated that the expression of these enzymes varies between healthy tissues and cancerous tumors of the colorectal [253]. Moreover,

alterations in mitochondrial lipid peroxides and enzymes that are antioxidants, such as glutathione transferase, have been reported in colorectal adenocarcinoma tissues [254].

Carboxylesterase(EC 3.1.1.1), phospholipase A1(EC 3.1.1.32), phospholipase A2 (EC 3.1.1.4), carboxymethylenebutenolidase(EC 3.1.1.45), pimelyl-[acyl-carrier protein] methyl ester esterase(EC 3.1.1.85), exodeoxyribonuclease (lambda-induced)(EC 3.1.11.3), exodeoxyribonuclease V (EC 3.1.11.5), S-formylglutathione hydrolase(EC 3.1.2.12), oleoyl-[acyl-carrier-protein] hydrolase(EC 3.1.2.14), palmitoyl-CoA hydrolase (EC 3.1.2.2), 1,4-dihydroxy-2-naphthoyl-CoA hydrolase(EC 3.1.2.28), deoxyribonuclease I(EC 3.1.21.1), deoxyribonuclease V(EC 3.1.21.7), glucose-1-phosphatase (EC 3.1.3.10), trehalose-phosphatase(EC 3.1.3.12), mannitol-1-phosphatase (EC 3.1.3.22), 4-phytase(EC 3.1.3.26), sorbitol-6-phosphatase (EC 3.1.3.50), 3'-nucleotidase(EC 3.1.3.6), pyridoxal phosphatase(EC 3.1.3.74), diacylglycerol diphosphate phosphatase(EC 3.1.3.81), D-glycero-beta-D-manno-heptose 1,7-bisphosphate 7-phosphatase (EC 3.1.3.82), 5'-deoxynucleotidase(EC 3.1.3.89), 3',5'-nucleoside bisphosphate phosphatase(EC 3.1.3.97), [acyl-carrier-protein] phosphodiesterase(EC 3.1.4.14), 3',5'-cyclic-nucleotide phosphodiesterase(EC 3.1.4.17), phosphoribosyl 1,2-cyclic phosphate phosphodiesterase(EC 3.1.4.55), limit dextrin alpha-1,6-maltotetraose-hydrolase(EC 3.2.1.196), sulfoquinovosidase(EC 3.2.1.199), alpha,alpha-trehalase(EC 3.2.1.28), double-stranded uracil-DNA glycosylase(EC 3.2.2.28), isochorismatase (EC 3.3.2.1), leucyl aminopeptidase(EC 3.4.11.1),PepB aminopeptidase (EC 3.4.11.23), dipeptidase E(EC 3.4.13.21), beta-aspartyl-peptidase(EC 3.4.19.5), pitrilysin (EC 3.4.24.55), oligopeptidase A (EC 3.4.24.70), HslU-HslV peptidase(EC 3.4.25.2), chitin disaccharide deacetylase (EC 3.5.1.105), ureidoacrylate amidohydrolase (EC 3.5.1.110), Glutathionylspermidine amidase (EC 3.5.1.78), succinylglutamate desuccinylase (EC 3.5.1.96), N-succinylarginine dihydrolase (EC 3.5.3.23), (S)-ureidoglycine aminohydrolase (EC 3.5.3.26), dCTP deaminase (EC 3.5.4.13), adenosine deaminase (EC 3.5.4.4), N-isopropylammelide isopropylaminohydrolase (EC 3.5.4.42), dCTP diphosphatase (EC 3.6.1.12), triphosphatase (EC 3.6.1.25), CDP-diacylglycerol diphosphatase (EC 3.6.1.26), UDP-sugar diphosphatase (EC 3.6.1.45), (d)CTP diphosphatase (EC 3.6.1.65), dihydroneopterin triphosphate diphosphatase (EC 3.6.1.67), nucleotide diphosphatase (EC 3.6.1.9), maltose-transporting ATPase (EC 3.6.3.19), phosphoribosyl 1,2-cyclic phosphate phosphodiesterase (EC 3.1.4.55), limit dextrin

alpha-1,6-maltotetraose-hydrolase (EC 3.2.1.196), sulfoquinovosidase (EC 3.2.1.199), alpha,alpha-trehalase (EC 3.2.1.28), double-stranded uracil-DNA glycosylase (EC 3.2.2.28), isochorismatase (EC 3.3.2.1), leucyl aminopeptidase (EC 3.4.11.1), PepB aminopeptidase (EC 3.4.11.23), dipeptidase E (EC 3.4.13.21), beta-aspartyl-peptidase (EC 3.4.19.5), pitrilysin (EC 3.4.24.55), oligopeptidase A (EC 3.4.24.70), HslU-HslV peptidase (EC 3.4.25.2), chitin disaccharide deacetylase (EC 3.5.1.105), ureidoacrylate amidohydrolase (EC 3.5.1.110), Glutathionylspermidine amidase (EC 3.5.1.78), succinylglutamate desuccinylase (EC 3.5.1.96), N-succinylarginine dihydrolase (EC 3.5.3.23), (S)-ureidoglycine aminohydrolase (EC 3.5.3.26), dCTP deaminase (EC 3.5.4.13), adenosine deaminase (EC 3.5.4.4), N-isopropylammelide isopropylaminohydrolase (EC 3.5.4.42), dCTP diphosphatase (EC 3.6.1.12), triphosphatase (EC 3.6.1.25), CDP-diacylglycerol diphosphatase (EC 3.6.1.26), UDP-sugar diphosphatase (EC 3.6.1.45), (d)CTP diphosphatase (EC 3.6.1.65), dihydroneopterin triphosphate diphosphatase (EC 3.6.1.67), nucleotide diphosphatase (EC 3.6.1.9), maltose-transporting ATPase (EC 3.6.3.19), Mg2+-importing ATPase (EC 3.6.3.2), vitamin B12-transporting ATPase (EC 3.6.3.33), 4-chlorobenzoyl-CoA dehalogenase (EC 3.8.1.7) enzymes classified under EC 3 are known as Hydrolases. The pathophysiology of carcinoma of the colon is significantly afflicted by hydrolases. Epoxide hydrolases, like microsomal epoxide hydrolase (mEH), have been linked to a higher chance of cancer of the gastrointestinal tract [255]. Moreover, the role of hydrolases extends to the molecular level, where ubiquitin carboxyl-terminal hydrolase L1 (UCHL1) has been investigated in nasopharyngeal carcinoma [256]. In colorectal cancer, mono-ADP-ribosylhydrolase locus deletions have been found to impair PARP1 transferase activity [257]. Additionally, significant differential expression of epoxide hydrolase 4 (EPHX4) has been observed in colorectal tumors [258]. Furthermore, the study of endocannabinoids and ceramides in colorectal cancer tissues has highlighted alterations in these molecules and their metabolic enzymes [259]. In conclusion, hydrolases, including epoxide hydrolases and other related enzymes, contribute key functions in the appearance and growth of carcinoma of the colon.

Oxaloacetate decarboxylase (EC 4.1.1.112), ornithine decarboxylase (EC 4.1.1.17), lysine decarboxylase (EC 4.1.1.18), tartronate-semialdehyde synthase (EC 4.1.1.47), 3-dehydro-L-gulonate-6-phosphate decarboxylase (EC 4.1.1.85), 2-dehydro-3-deoxyglucarate aldolase (EC 4.1.2.20), 2-dehydro-3-deoxy-6-

phosphogalactonate aldolase (EC 4.1.2.21), 4-hydroxy-2-oxoheptanedioate aldolase (EC 4.1.2.52), 2-keto-3-deoxy-L-rhamnonate aldolase (EC 4.1.2.53), isocitrate lyase (EC 4.1.3.1), 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17), chorismate lyase (EC 4.1.3.40), tryptophanase (EC 4.1.99.1), deoxyribodipyrimidine photo-lyase (EC 4.1.99.3), cyanase (EC 4.2.1.104), phosphogluconate dehydratase (EC 4.2.1.12), glucarate dehydratase (EC 4.2.1.40), galactonate dehydratase (EC 4.2.1.6), altronate dehydratase (EC 4.2.1.7), pseudouridylate synthase (EC 4.2.1.70), uroporphyrinogen-III synthase (EC 4.2.1.75), 2-methylcitrate dehydratase (EC 4.2.1.79), L-rhamnonate dehydratase (EC 4.2.1.90), D-Serine ammonia-lyase (EC 4.3.1.18), glutathione-specific gamma-glutamylcyclotransferase (EC 4.3.2.7), cysteine-S-conjugate beta-lyase (EC 4.4.1.13) and D-cysteine desulfhydrase (EC 4.4.1.15) enzymes classified under EC 4 are known as Lyases. Lyases crucial in cancer biology, especially regarding the use of enzyme inhibitors targeting key lyases involved in cancer progression. For instance, S1P lyase mRNA expression reduction has been observed in intestinal cancer and metastatic tumors, indicating its potential involvement in cancer development [260]. Colorectal cancer is a multifaceted disease driven by various molecular mechanisms. One crucial enzyme involved in cancer progression is S1P lyase, which influences cancer cell proliferation and invasion [261]. Furthermore, it has been discovered that ATP citrate lyase (ACLY) is a crucial metabolic enzyme that is increased in tumor cells, especially carcinoma of the colorectal, where it is essential to the cells' proliferation [262].

ADP-glyceromanno-heptose 6-epimerase (EC 5.1.3.20), L-fucose mutarotase (EC 5.1.3.29), Hydroxypyruvate isomerase (EC 5.3.1.22), D-sedoheptulose 7-phosphate isomerase (EC 5.3.1.28), sulfoquinovose isomerase (EC 5.3.1.31), oxaloacetate tautomerase (EC 5.3.2.2), trans-2-decenoyl-[acyl-carrier protein] isomerase (EC 5.3.3.14), 16S rRNA pseudouridine516 synthase (EC 5.4.99.19), 23S rRNA pseudouridine2457 synthase (EC 5.4.99.20), 23S rRNA pseudouridine2604 synthase (EC 5.4.99.21), 23S rRNA pseudouridine2605 synthase (EC 5.4.99.22), tRNA pseudouridine13 synthase (EC 5.4.99.27), tRNA pseudouridine32 synthase (EC 5.4.99.28), 23S rRNA pseudouridine746 synthase (EC 5.4.99.29) enzymes classified under EC 5 are known as Isomerases. Since they are in charge of several paths and procedures linked to carcinogenesis and growth of cancer, isomerases are important players in colorectal cancer. For instance, ribose-5-phosphate isomerase A (RPIA) has been shown to regulate cancer growth and tumorigenesis [263]. Modifications in

complementary polyadenylation (APA) patterns have been reported in the development of colorectal cancer, impacting genes such as peptidylpropyl isomerase E (PPIE) [264] and cyclophilin A [265]. These changes in APA patterns are indicative of the dynamic gene regulation processes occurring during different stages of cancer progression [266]. Overall, by affecting several cellular processes and pathways involved in carcinogenesis and cancer progression, isomerases including ribose-5-phosphate isomerase A, peptidyl-prolyl isomerases, and protein disulfide isomerases perform vital roles in cancer of the gastrointestinal tract. Targeting these isomerases could be one of the therapeutic strategies used to treat carcinoma of the colorectal.

Acetate-CoA ligase (EC 6.2.1.1), propionate-CoA ligase (EC 6.2.1.17), O-succinylbenzoate-CoA ligase (EC 6.2.1.26), 4-chlorobenzoate-CoA ligase (EC 6.2.1.33), carnitine-CoA ligase (EC 6.2.1.48), glutamate-putrescine ligase (EC 6.3.1.11), RNA 3'-terminal-phosphate cyclase (ATP) (EC 6.5.1.4) enzymes classified under EC 6 are known as Ligases. Ligases have been discovered to have a substantial impact on the occurrence and dissemination of colorectal cancer especially E3 ubiquitin ligases. The pathogenesis of CRC involves a complex interplay of genetic alterations, including somatic mutations, abnormal gene expression, and epigenetic changes, which can lead to dysregulation of E3 ligases [267]. Studies have highlighted the role of specific E3 ligases in CRC.

Ubiquinol oxidase (H+-transporting) (EC 7.1.1.3), quinol oxidase (electrogenic, proton-motive force generating) (EC 7.1.1.7), ABC-type polar-amino-acid transporter (EC 7.4.2.1), ABC-type maltose transporter (EC 7.5.2.1), ABC-type xenobiotic transporter (EC 7.6.2.2), ABC-type heme transporter (EC 7.6.2.5) enzymes classified under EC 7 are known as Translocases. Translocases play a crucial role in the emergence and dissemination of colorectal cancer. According to studies, the overexpression of the fatty acid translocase CD36 increases MMP28 and decreases E-Cadherin expression, hence promoting colorectal cancer metastasis. Additionally, increased CD36 overexpression has been linked to a higher potential for metastasis in colorectal cancer cells [268]. Moreover, mitochondrial translocases have also been implicated in colorectal cancer [269]. In conclusion, translocases, including fatty acid translocases and mitochondrial translocases, play crucial roles in colorectal cancer progression, metastasis, and maintenance of mitochondrial function.

Together with the previously described enzymes, Escherichia coli has been identified as a possible indicator for colorectal cancer.

Faecalibacterium prausnitzii is an important component of the gut microbiota, has been extensively researched for its potential role in various health conditions, including colorectal cancer. Research has indicated that an unbalanced gut microbiome, characterized by a low abundance of Faecalibacterium prausnitzii, may be involved in intestinal inflammatory reactions that result in intestinal illnesses [270].

Gluconate 5-dehydrogenase (EC 1.1.1.69) enzyme is a subgroup of NAD+ or NADP+ as acceptors. The effect of NAD+ and NADP+ in colorectal cancer is mentioned above.

Glutamate synthase (NADH) (EC 1.4.1.14) serves a vital role in the metabolism of amino acids, particularly in the assimilation of nitrogen. Research has shown that glutamate synthase exhibits tissue-specific distribution patterns and controlled by mechanisms that detect light and metabolites [271]. Within the realm of colorectal cancer, the production of glutamate synthase and related enzymes has been linked to cancer progression. For instance, elevated levels of a relevant enzyme called mPGES-1 have been linked to a poorer outcome in later stages of colorectal cancer, suggesting a major involvement in the progression of the disease [272]. In summary, the intricate interplay between enzymes like glutamate synthase, fatty acid synthase, and mPGES-1, along with the modulation of ATP synthase activity and other molecular factors, contributes to the complex landscape of colorectal cancer progression.

rRNA (guanine-N1-)-methyltransferase (EC 2.1.1.51) enzyme is categorized as methyltransferases fall under EC 2.1.1. The effect of Methyltransferases in colorectal cancer is mentioned above.

1,3-beta-galactosyl-N-acetylhexosamine phosphorylase (EC 2.4.1.211) enzyme classified under EC 2.4 are known as Glycosyltransferases. The effect of Glycosyltransferases in colorectal cancer is mentioned above.

3-hydroxybutyryl-CoA dehydratase (EC 4.2.1.55) plays a key role in converting 3-hydroxybutyryl-CoA to crotonyl-CoA, which is a crucial step in specific autotrophic carbon fixation pathways [273]. Studies have indicated that 4-hydroxybutyryl-CoA dehydratases, which are closely related to 3-hydroxybutyryl-CoA dehydratase, are sensitive to oxygen [274]. This oxygen sensitivity is significant in understanding the enzyme's function in aerobic environments such as cancer tissues. Involvement of 3-hydroxybutyryl-CoA dehydratase in colorectal cancer may be linked to its oxygen sensitivity, stereochemical specificity, and unique structural features.

Cystathionine gamma-lyase (EC 4.4.1.1) is a substantial enzyme responsible for the production of hydrogen sulfide (H2S) in various tissues [275]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of Cystathionine gamma-lyase on colorectal cancer.

Precorrin-8X methylmutase (EC 5.4.1.2) also known as CobH, is essential in the biosynthesis of vitamin B12 [377]. Biochemical pathways, especially those connected to cancer of the gastrointestinal tract, have been related to vitamin B12.

Together with the previously mentioned enzymes, Faecalibacterium prausnitzii has been found as an intriguing biomarker for colorectal cancer.

Prevotella copri is a bacterium in the intestinal tract of humans has been connected to a range of health issues. Studies have identified Prevotella copri as a potential biomarker for colorectal cancer [276].

Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) (EC 1.1.1.40) enzyme is a subgroup of NAD+ or NADP+ as acceptors. The effect of NAD+ and NADP+ in colorectal cancer is mentioned above.

Glutamate synthase (ferredoxin) (EC 1.4.7.1) is a key player in colorectal cancer, influencing treatment response, prognosis, and cell fate decisions through its involvement in p53-dependent pathways and apoptosis. Research has demonstrated that ferredoxin reductase is linked to p53-dependent apoptosis in colorectal cancer, especially in those receiving therapy with 5-fluorouracil. [277].

Saccharopine dehydrogenase (NAD+, L-lysine-forming) (EC 1.5.1.7) enzyme is a subgroup of NAD+ or NADP+ as acceptors. The effect of NAD+ and NADP+ in colorectal cancer is mentioned above.

Thymidylate synthase (FAD) (EC 2.1.1.148) enzyme classified under EC 2.1.1 are known as Methyltransferases. The effect of Methyltransferases in colorectal cancer is mentioned above.

4-O-beta-D-mannosyl-D-glucose phosphorylase (EC 2.4.1.281) and decaprenyl-phosphate phosphoribosyltransferase (EC 2.4.2.45) enzymes classified under EC 2.4 are known as Glycosyltransferases. The effect of Glycosyltransferases in colorectal cancer is mentioned above.

Prolyl aminopeptidase (EC 3.4.11.5) is an enzyme significantly involved in numerous cancers, encompassing colorectal carcinoma. Research has demonstrated a

connection between aminopeptidases, such as PAP, and the construction and dissemination of tumors [278].

ADP-ribose diphosphatase (EC 3.6.1.13), Poly(ADP-ribose) polymerase (PARP) enzymes, specifically PARP1, have emerged as significant targets in cancer therapy, including colorectal cancer. PARP inhibitors have shown promise in preclinical trials, indicating their potential in cancer treatment [279].

Ca2+-transporting ATPase (EC 3.6.3.8) and their activity alterations have been connected to gastrointestinal cancer growth. Research has revealed that in cancer tissues, the activity of Ca2+ ATPase of the plasma membrane tends to be higher compared to controls [168].

Pyridoxal 5'-phosphate synthase (glutamine hydrolysing) (EC 4.3.3.6) has a major effect on the occurrence of colorectal carcinoma [280]. Furthermore, glutamine has been found to induce remodeling of tight junctions in CRC cells, impacting epithelial integrity [281].

Inositol-3-phosphate synthase (EC 5.5.1.4) enzyme classified under EC 5 are known as Isomerases. The effect of Isomerases in colorectal cancer is mentioned above.

Pyrroline-5-carboxylate reductase (EC 1.5.1.2) enzyme is a subgroup of NAD+ or NADP+ as acceptors. The effect of NAD+ and NADP+ in colorectal cancer is mentioned above.

In addition to the previously mentioned enzymes, prevotella copri has been found as a promising biomarker for colorectal cancer.

Flavonifractor plautii has been highlighted in multiple studies as an important bacterium linked to colorectal cancer. [50, 340]. Patients with colorectal cancer have reportedly shown it to be enhanced suggesting a possible role in the onset of the illness [282].

2-enoate reductase (EC 1.3.1.31) and L-erythro-3,5-diaminohexanoate dehydrogenase (EC 1.4.1.11) enzymes is a subgroup of NAD+ or NADP+ as acceptors. The effect of NAD+ and NADP+ in colorectal cancer is mentioned above.

Hydroxydechloroatrazine ethylaminohydrolase (EC 3.5.99.3) enzyme is classified under EC 3 are known as Hydrolases. The effect of Hydrolases in colorectal cancer is mentioned above.

Enoyl-CoA hydratase 2 (EC 4.2.1.119) enzyme is classified under EC 4 are known as Lyases. The effect of Lyases in colorectal cancer is mentioned above.

Flavonifractor plautii, along with the presence of 2-enoate reductase (EC 1.3.1.31) and L-erythro-3,5-diaminohexanoate dehydrogenase (EC 1.4.1.11), hydroxydechloroatrazine ethylaminohydrolase (EC 3.5.99.3) and enoyl-CoA hydratase 2 enzymes (EC 4.2.1.119) has been identified as a prospective colorectal cancer biological indicator.

Streptococcus salivarius, a member of the salivarius group of viridans streptococci, has been extensively studied for its beneficial properties. Research has shown that S. salivarius exhibits anti-inflammatory effects in various contexts [283]. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of Streptococcus salivarius on colorectal cancer.

Acetoacetyl-CoA reductase (EC 1.1.1.36) enzyme is a subgroup of NAD+ or NADP+ as acceptors. The effect of NAD+ and NADP+ in colorectal cancer is mentioned above.

Phosphate acyltransferase (EC 2.3.1.274), UDP-N-acetylmuramoylpentapeptide-lysine N6-alanyltransferase (EC 2.3.2.10) and **phosphomevalonate kinase (EC 2.7.4.2)** enzymes classified under EC 2 are known as Transferases. The effect of Transferases in colorectal cancer is mentioned above.

Levansucrase (EC 2.4.1.10) and dextransucrase (EC 2.4.1.5) enzymes classified under EC 2.4 are known as Glycosyltransferases. The effect of Glycosyltransferases in colorectal cancer is mentioned above.

Dextranase (EC 3.2.1.11) and O-sialoglycoprotein endopeptidase (EC 3.4.24.57) enzymes classified under EC 3 are known as Hydrolases. The effect of Hydrolases in colorectal cancer is mentioned above.

UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase (EC 6.3.2.7) is also referred to MurD. To the best of my knowledge, and after a thorough investigation of the available literature, no studies have directly addressed the direct impact of this enzyme on colorectal cancer.

Streptococcus salivarius has been identified as a possible biomarker for colorectal cancer along with the previously described enzymes.

Chapter 6

Conclusions and Future Prospects

6.1 Conclusions

In conclusion, this thesis has successfully employed advanced bioinformatics and machine learning techniques to identify specific microorganism-enzyme pairs in the human gut microbiota that are associated with colorectal cancer (CRC). Through meticulous data analysis and the integration of taxonomic and functional profiling, we have unveiled distinctive microbial signatures and enzyme clusters, pinpointing their potential as non-invasive biomarkers for early detection and prognosis of CRC. Our findings not only shed light on the complex interplay between the gut microbiota and colorectal cancer but also highlight the potential of these identified biomarkers in revolutionizing the diagnostic and therapeutic landscapes of CRC. By providing a novel perspective on the microbial and enzymatic underpinnings of CRC, this research contributes significantly to the burgeoning field of personalized medicine, paving the way for targeted interventions and improved clinical outcomes. Future research should focus on validating these biomarkers in larger cohort studies and exploring their mechanistic roles in the pathogenesis of CRC. Additionally, the exploration of the gut microbiota's dynamic nature and its interaction with host physiology could further unravel the multifaceted contributions to colorectal carcinogenesis. This thesis underscores the power of integrating bioinformatics and machine learning in biomedical research, offering new avenues for understanding and combating complex diseases like colorectal cancer.

6.2 Societal Impact and Contribution to Global Sustainability

The societal impact of this thesis on identifying microorganism-enzyme pairs associated with colorectal cancer (CRC) extends beyond the scientific community, touching various facets of public health, healthcare economics, and global sustainability. The early detection and accurate diagnosis of CRC facilitated by the biomarkers identified in this study could lead to significant improvements in patient outcomes and healthcare cost reductions.

Firstly, the global burden of CRC, one of the leading causes of cancer-related mortality worldwide, necessitates innovative approaches to its diagnosis and treatment. By pinpointing specific biomarkers within the human gut microbiota, this research contributes to a deeper understanding of CRC pathogenesis, enabling the development of non-invasive, early detection methods. Such advancements can significantly reduce the mortality rate of CRC through timely and effective treatment, enhancing patient survival and quality of life.

Secondly, from a healthcare economics perspective, the early detection of CRC through non-invasive means can lead to considerable cost savings. It reduces the need for extensive and expensive late-stage treatments, such as surgery and chemotherapy, which are often required when the disease is diagnosed at advanced stages. These savings can be redirected to other critical healthcare needs, contributing to a more sustainable healthcare system.

Furthermore, this research aligns with global sustainability goals by promoting health and well-being (Sustainable Development Goal 3). By advancing our understanding of the gut microbiome's role in cancer and leveraging this knowledge for early diagnosis, this study contributes to the broader goal of ensuring healthy lives and promoting well-being for all ages.

Moreover, the interdisciplinary approach of this thesis, integrating bioinformatics and machine learning, sets a precedent for future research in precision medicine. It

emphasizes the importance of harnessing technological advancements in addressing global health challenges, fostering innovation, and contributing to sustainable healthcare solutions.

In conclusion, the societal impact and contribution to global sustainability of this thesis are profound. It not only paves the way for improved clinical practices and healthcare outcomes in CRC but also exemplifies the critical role of scientific research in advancing global health and sustainability agendas. Through innovative methodologies and cross-disciplinary collaborations, this research underscores the potential of bioinformatics and machine learning to contribute to a more sustainable and healthier future.

6.3 Future Prospects

Building on the findings of this thesis, the clinical validation of identified microorganism-enzyme pairs as biomarkers for colorectal cancer (CRC) represents an immediate and critical next step. Large-scale studies are necessary to verify these biomarkers' diagnostic and prognostic significance, potentially leading to non-invasive screening tests that can transform early detection and personalized treatment strategies for CRC. Moreover, understanding the mechanisms underlying these microorganism-enzyme interactions in CRC progression is crucial. Future research should delve into the biological pathways and interactions, using in vitro and in vivo models to elucidate their functional roles and impact on disease dynamics, which may uncover novel therapeutic targets.

The methodologies and insights from this study also hold promise for broader applications beyond CRC, including investigations into the microbiome’s role in other diseases. This could lead to a more comprehensive understanding of the microbiome’s influence on human health across various conditions, including other forms of cancer, inflammatory diseases, and metabolic disorders. Integrating metagenomic data with other omics disciplines—such as genomics, proteomics, and metabolomics—can provide a holistic view of disease mechanisms, enhancing our understanding of the complex interplay between the microbiome and human health.

In the realm of personalized medicine, the identified biomarkers have the potential to tailor therapeutic strategies to individual microbiome profiles, leading to personalized treatment plans. This could eventually extend to therapeutic interventions, such as modifying the gut microbiota through probiotics, prebiotics, or microbiota transplants, offering innovative treatment modalities for CRC. Additionally, the global prevalence of CRC necessitates research that accommodates population diversity in gut microbiota composition, aiming for diagnostic and therapeutic strategies that are applicable universally, thereby addressing a significant public health concern worldwide.

The advancement in bioinformatics tools and machine learning models is pivotal for the future of microbiome research, enabling the analysis of complex data sets and enhancing the predictive power of computational models. This technological progress will undoubtedly catalyze further discoveries and innovations in understanding and treating CRC and other microbiome-related conditions, marking a significant stride towards the era of data-driven, personalized healthcare.

BIBLIOGRAPHY

[1] D. Yan, L. Cao, M. Zhou, and H. Mohimani, "TransDiscovery: Discovering Biotransformation from Human Microbiota by Integrating Metagenomic and Metabolomic Data," *Metabolites*, vol. 12, no. 2, Art. no. 2, Feb. 2022, doi: 10.3390/metabo12020119.

[2] J. Halfvarson *et al.*, "Dynamics of the human gut microbiome in inflammatory bowel disease," *Nat. Microbiol.*, vol. 2, p. 17004, Feb. 2017, doi: 10.1038/nmicrobiol.2017.4.

[3] R. Wisaksana, G. Darmawan, N. Agustanti, and D. Girawan, "Gut Microbiota in Human Immunodeficiency Virus Infection," *Indones. J. Gastroenterol. Hepatol. Dig. Endosc.*, vol. 20, no. 1, Art. no. 1, Jul. 2020, doi: 10.24871/201201938-43.

[4] X. Fan, Y. Jin, G. Chen, X. Ma, and L. Zhang, "Gut Microbiota Dysbiosis Drives the Development of Colorectal Cancer," *Digestion*, vol. 102, no. 4, pp. 508–515, Sep. 2020, doi: 10.1159/000508328.

[5] I. Mármol, C. Sánchez-de-Diego, A. Pradilla Dieste, E. Cerrada, and M. J. Rodriguez Yoldi, "Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer," *Int. J. Mol. Sci.*, vol. 18, no. 1, p. 197, Jan. 2017, doi: 10.3390/ijms18010197.

[6] K. D. Miller *et al.*, "Cancer treatment and survivorship statistics, 2016," *CA. Cancer J. Clin.*, vol. 66, no. 4, pp. 271–289, Jul. 2016, doi: 10.3322/caac.21349.

[7] F. Duraturo, R. Liccardo, M. De Rosa, and P. Izzo, "Genetics, diagnosis and treatment of Lynch syndrome: Old lessons and current challenges," *Oncol. Lett.*, vol. 17, no. 3, pp. 3048–3054, Mar. 2019, doi: 10.3892/ol.2019.9945.

[8] M. B. Yurgelun *et al.*, "Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer," *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.*, vol. 35, no. 10, pp. 1086–1095, Apr. 2017, doi: 10.1200/JCO.2016.71.0012.

[9] B. A. Seifert *et al.*, "Determining the clinical validity of hereditary colorectal cancer and polyposis susceptibility genes using the Clinical Genome Resource Clinical Validity Framework," *Genet. Med. Off. J. Am. Coll. Med. Genet.*, vol. 21, no. 7, pp. 1507–1516, Jul. 2019, doi: 10.1038/s41436-018-0373-1.

[10] H. Wang, Y. Chen, D. Yang, and L. Ma, "Perspective of Human Condensins Involved in Colorectal Cancer," *Front. Pharmacol.*, vol. 12, 2021, Accessed: Feb. 02, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphar.2021.664982

[11] S. Mouillet-Richard and P. Laurent-Puig, "YAP/TAZ Signalling in Colorectal Cancer: Lessons from Consensus Molecular Subtypes," *Cancers*, vol. 12, no. 11, p. 3160, Oct. 2020, doi: 10.3390/cancers12113160.

[12] I. A. Voutsadakis, "Sensitivities and Dependencies of BRAF Mutant Colorectal Cancer Cell Lines with or without PIK3CA Mutations for Discovery of Vulnerabilities with Therapeutic Potential," *Med. Kaunas Lith.*, vol. 58, no. 10, p. 1498, Oct. 2022, doi: 10.3390/medicina58101498.

[13] Z.-H. Chen *et al.*, "Assessment of Modifiable Factors for the Association of Marital Status With Cancer-Specific Survival," *JAMA Netw. Open*, vol. 4, no. 5, p. e2111813, May 2021, doi: 10.1001/jamanetworkopen.2021.11813.

[14] J. V. Veld *et al.*, "Definition of large bowel obstruction by primary colorectal cancer: A systematic review," *Colorectal Dis. Off. J. Assoc. Coloproctology G. B. Irel.*, vol. 23, no. 4, pp. 787–804, Apr. 2021, doi: 10.1111/codi.15479.

[15] T. Fujiwara *et al.*, “Accuracy of algorithms to identify patients with a diagnosis of major cancers and cancer-related adverse events in an administrative database: a validation study in an acute care hospital in Japan,” *BMJ Open*, vol. 12, no. 7, p. e055459, Jul. 2022, doi: 10.1136/bmjopen-2021-055459.
[16] P. Karra *et al.*, “Metabolic dysfunction and obesity-related cancer: Beyond obesity and metabolic syndrome,” *Obes. Silver Spring Md*, vol. 30, no. 7, pp. 1323–1334, Jul. 2022, doi: 10.1002/oby.23444.
[17] Y. Zhang *et al.*, “Gut metabolite Urolithin A mitigates ionizing radiation-induced intestinal damage,” *J. Cell. Mol. Med.*, vol. 25, no. 21, pp. 10306–10312, 2021, doi: 10.1111/jcmm.16951.
[18] “The gut microbiota in IBD | Nature Reviews Gastroenterology & Hepatology.” Accessed: Jun. 16, 2022. [Online]. Available: https://www.nature.com/articles/nrgastro.2012.152
[19] E.-J. Yun *et al.*, “Diet Is a Stronger Covariate than Exercise in Determining Gut Microbial Richness and Diversity,” *Nutrients*, vol. 14, no. 12, Art. no. 12, Jan. 2022, doi: 10.3390/nu14122507.
[20] A. Rintala *et al.*, “Gut Microbiota Analysis Results Are Highly Dependent on the 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor,” *J. Biomol. Tech. JBT*, vol. 28, no. 1, pp. 19–30, Apr. 2017, doi: 10.7171/jbt.17-2801-003.
[21] K. Wang *et al.*, “Microbial-host-isozyme analyses reveal microbial DPP4 as a potential antidiabetic target,” *Science*, vol. 381, no. 6657, p. eadd5787, Aug. 2023, doi: 10.1126/science.add5787.
[22] B. B. Williams *et al.*, “Discovery and Characterization of Gut Microbiota Decarboxylases that Can Produce the Neurotransmitter Tryptamine,” *Cell Host Amp Microbe*, vol. 16, no. 4, 2014, doi: 10.1016/j.chom.2014.09.001.
[23] W. Jelski and B. Mroczko, “Biochemical Markers of Colorectal Cancer - Present and Future,” *Cancer Manag. Res.*, vol. 12, pp. 4789–4797, 2020, doi: 10.2147/CMAR.S253369.
[24] K. L. Krigul, O. Aasmets, K. Lüll, T. Org, and E. Org, “Using fecal immunochemical tubes for the analysis of the gut microbiome has the potential to improve colorectal cancer screening,” scite.ai. Accessed: Jan. 26, 2024. [Online]. Available: https://doi.org/10.1101/2021.03.15.435399
[25] S. Oke and A. Martín, “Insights into the role of the intestinal microbiota in colon cancer,” *Ther. Adv. Gastroenterol.*, vol. 10, no. 5, 2017, doi: 10.1177/1756283x17694832.
[26] M. B. Burns, J. Lynch, T. K. Starr, D. Knights, and R. Blekhman, “Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment,” *Genome Med.*, vol. 7, no. 1, 2015, doi: 10.1186/s13073-015-0177-8.
[27] H. Sheng, J. Shao, M. K. Washington, and R. N. DuBois, “Prostaglandin E2 increases growth and motility of colorectal carcinoma cells,” *J. Biol. Chem.*, vol. 276, no. 21, pp. 18075–18081, May 2001, doi: 10.1074/jbc.M009689200.
[28] A. Koushik *et al.*, “Nonsynonymous Polymorphisms in Genes in the One-Carbon Metabolism Pathway and Associations with Colorectal Cancer,” *Cancer Epidemiol. Biomarkers Prev.*, vol. 15, no. 12, pp. 2408–2417, Dec. 2006, doi: 10.1158/1055-9965.EPI-06-0624.
[29] H. Cheng, X. Sun, J. Li, P. He, W. Liu, and X. Meng, “Knockdown of Uba2 inhibits colorectal cancer cell invasion and migration through downregulation of the Wnt/β-catenin signaling pathway,” *J. Cell. Biochem.*, vol. 119, no. 8, pp. 6914–6925, 2018, doi: 10.1002/jcb.26890.

[30] M. Jiang *et al.*, “The Role of Amino Acid Metabolism of Tumor Associated Macrophages in the Development of Colorectal Cancer,” *Cells*, vol. 11, no. 24, p. 4106, Dec. 2022, doi: 10.3390/cells11244106.
[31] J. Li, W. Xu, F. Liu, S. Huang, and M. He, “GSTM1 polymorphism contribute to colorectal cancer in Asian populations: a prospective meta-analysis,” *Sci. Rep.*, vol. 5, no. 1, Art. no. 1, Jul. 2015, doi: 10.1038/srep12514.
[32] R. Wu *et al.*, “Effects of CD73 on human colorectal cancer cell growth in vivo and in vitro,” *Oncol. Rep.*, vol. 35, no. 3, pp. 1750–1756, Mar. 2016, doi: 10.3892/or.2015.4512.
[33] S. de Vogel *et al.*, “Genetic variants of methyl metabolizing enzymes and epigenetic regulators: associations with promoter CpG island hypermethylation in colorectal cancer,” *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.*, vol. 18, no. 11, pp. 3086–3096, Nov. 2009, doi: 10.1158/1055-9965.EPI-09-0289.
[34] C. Martínez, E. García-Martín, R. M. Pizarro, F. J. García-Gamito, and J. a. G. Agúndez, “Expression of paclitaxel-inactivating CYP3A activity in human colorectal cancer: implications for drug therapy,” *Br. J. Cancer*, vol. 87, no. 6, pp. 681–686, Sep. 2002, doi: 10.1038/sj.bjc.6600494.
[35] Y. Chen, Y. Yang, and J. Gu, “ Clinical Implications of the Associations Between Intestinal Microbiome and Colorectal Cancer Progression ,” *Cancer Manag. Res.*, vol. 12, pp. 4117–4128, Jun. 2020, doi: 10.2147/CMAR.S240108.
[36] M. Castellarin *et al.*, “Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma,” *Genome Res.*, vol. 22, no. 2, pp. 299–306, Feb. 2012, doi: 10.1101/gr.126516.111.
[37] L. Lauka *et al.*, “Role of the intestinal microbiome in colorectal cancer surgery outcomes,” *World J. Surg. Oncol.*, vol. 17, no. 1, 2019, doi: 10.1186/s12957-019-1754-x.
[38] Y. Yang *et al.*, “Dysbiosis of human gut microbiome in young-onset colorectal cancer,” *Nat. Commun.*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-27112-y.
[39] R. Siddiqui, A. Boghossian, A. M. Alharbi, H. Alfahemi, and N. A. Khan, “The Pivotal Role of the Gut Microbiome in Colorectal Cancer,” *Biology*, vol. 11, no. 11, 2022, doi: 10.3390/biology11111642.
[40] R. Wu, Q. Shen, P. Li, and N. Shang, “Sturgeon Chondroitin Sulfate Restores the Balance of Gut Microbiota in Colorectal Cancer Bearing Mice,” *Int. J. Mol. Sci.*, vol. 23, no. 7, 2022, doi: 10.3390/ijms23073723.
[41] M. Shoji *et al.*, “Characteristics of the gut microbiome profile in obese patients with colorectal cancer,” *JGH Open*, vol. 5, no. 4, 2021, doi: 10.1002/jgh3.12529.
[42] J. W. Debelius *et al.*, “The local tumor microbiome is associated with survival in late-stage colorectal cancer patients,” scite.ai. Accessed: Feb. 02, 2024. [Online]. Available: https://doi.org/10.1101/2022.09.16.22279353
[43] F. Beghini *et al.*, “Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3,” eLife. Accessed: Nov. 23, 2023. [Online]. Available: https://elifesciences.org/articles/65088/figures
[44] A. M. Thomas *et al.*, “Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation,” *Nat. Med.*, vol. 25, no. 4, Art. no. 4, Apr. 2019, doi: 10.1038/s41591-019-0405-7.
[45] F. Beghini *et al.*, “Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3,” *eLife*, vol. 10, p. e65088, May 2021, doi: 10.7554/eLife.65088.

[46] M. Yousef, S. Jung, L. C. Showe, and M. K. Showe, “Recursive Cluster Elimination (RCE) for classification and feature selection from gene expression data,” *BMC Bioinformatics*, vol. 8, no. 1, p. 144, May 2007, doi: 10.1186/1471-2105-8-144.
[47] M. Yousef, B. Bakir-Gungor, A. Jabeer, G. Goy, R. Qureshi, and L. C Showe, “Recursive Cluster Elimination based Rank Function (SVM-RCE-R) implemented in KNIME,” *F1000Research*, vol. 9, p. 1255, 2020, doi: 10.12688/f1000research.26880.2.
[48] M. Yousef, A. Kumar, and B. Bakir-Gungor, “Application of Biological Domain Knowledge Based Feature Selection on Gene Expression Data,” *Entropy*, vol. 23, no. 1, p. 2, Dec. 2020, doi: 10.3390/e23010002.
[49] D. W. Konas, K. Zhu, M. Sharma, K. S. Aulak, G. W. Brudvig, and D. J. Stuehr, “The FAD-shielding Residue Phe1395 Regulates Neuronal Nitric-oxide Synthase Catalysis by Controlling NADP+ Affinity and a Conformational Equilibrium within the Flavoprotein Domain,” *J. Biol. Chem.*, vol. 279, no. 34, 2004, doi: 10.1074/jbc.m400872200.
[50] D. Hamdane, C. Xia, S. C. Im, H. Zhang, J. J. P. Kim, and L. Waskell, “Structure and Function of an NADPH-Cytochrome P450 Oxidoreductase in an Open Conformation Capable of Reducing Cytochrome P450,” *J. Biol. Chem.*, vol. 284, no. 17, 2009, doi: 10.1074/jbc.m807868200.
[51] N. Kimmich, A. Das, I. F. Sevrioukova, Y. T. Meharenna, S. G. Sligar, and T. L. Poulos, “Electron Transfer between Cytochrome P450cin and Its FMN-containing Redox Partner, Cindoxin,” *J. Biol. Chem.*, vol. 282, no. 37, 2007, doi: 10.1074/jbc.m703790200.
[52] T. Löwenmark *et al.*, “Tumour Colonisation of Parvimonas micra Is Associated with Decreased Survival in Colorectal Cancer Patients,” *Cancers*, vol. 14, no. 23, Art. no. 23, Jan. 2022, doi: 10.3390/cancers14235937.
[53] T. Löwenmark *et al.*, “Parvimonas micra as a putative non-invasive faecal biomarker for colorectal cancer,” *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Sep. 2020, doi: 10.1038/s41598-020-72132-1.
[54] J. Xu *et al.*, “Alteration of the abundance of Parvimonas micra in the gut along the adenoma-carcinoma sequence,” *Oncol. Lett.*, vol. 20, no. 4, pp. 1–1, Oct. 2020, doi: 10.3892/ol.2020.11967.
[55] C. Gethings-Behncke *et al.*, “Fusobacterium nucleatum in the Colorectum and Its Association with Cancer Risk and Survival: A Systematic Review and Meta-analysis,” *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.*, vol. 29, no. 3, pp. 539–548, Mar. 2020, doi: 10.1158/1055-9965.EPI-18-1295.
[56] S. Zundler *et al.*, “Rectal Cancer Presenting with Absceding Infection Due to Fusobacterium nucleatum,” *Pathogens*, vol. 11, no. 10, Art. no. 10, Oct. 2022, doi: 10.3390/pathogens11101113.
[57] J. Wu, Q. Li, and X. Fu, “Fusobacterium nucleatum Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity,” *Transl. Oncol.*, vol. 12, no. 6, pp. 846–851, Jun. 2019, doi: 10.1016/j.tranon.2019.03.003.
[58] F.-M. Shang and H.-L. Liu, “Fusobacterium nucleatum and colorectal cancer: A review,” *World J. Gastrointest. Oncol.*, vol. 10, no. 3, pp. 71–81, Mar. 2018, doi: 10.4251/wjgo.v10.i3.71.
[59] H. Hussan, S. K. Clinton, K. Roberts, and M. T. Bailey, “Fusobacterium’s link to colorectal neoplasia sequenced: A systematic review and future insights,” *World J.*

Gastroenterol., vol. 23, no. 48, pp. 8626–8650, Dec. 2017, doi: 10.3748/wjg.v23.i48.8626.

[60] K. Ganesan, S. Guo, S. Fayyaz, G. Zhang, and B. Xu, "Targeting Programmed Fusobacterium nucleatum Fap2 for Colorectal Cancer Therapy," *Cancers*, vol. 11, no. 10, Art. no. 10, Oct. 2019, doi: 10.3390/cancers11101592.

[61] S. Watanabe, N. Shimada, K. Tajima, T. Kodaki, and K. Makino, "Identification and Characterization of l-Arabonate Dehydratase, l-2-Keto-3-deoxyarabonate Dehydratase, and l-Arabinolactonase Involved in an Alternative Pathway of l-Arabinose Metabolism," *J. Biol. Chem.*, vol. 281, no. 44, 2006, doi: 10.1074/jbc.m606727200.

[62] I. Amir, P. Bouvet, C. Legeay, U. Gophna, and A. Weinberger, "Eisenbergiella tayi gen. nov., sp. nov., isolated from human blood," *Int. J. Syst. Evol. Microbiol.*, vol. 64, Nov. 2013, doi: 10.1099/ijs.0.057331-0.

[63] Y. A. Millet *et al.*, "Modulation of the Gut Microbiome by Novel Synthetic Glycans for the Production of Propionate and the Reduction of Cardiometabolic Risk Factors," scite.ai. Accessed: Feb. 13, 2024. [Online]. Available: https://doi.org/10.1101/2022.04.04.487010

[64] M. Crusell *et al.*, "Gestational diabetes is associated with change in the gut microbiota composition in third trimester of pregnancy and postpartum," *Microbiome*, vol. 6, no. 1, 2018, doi: 10.1186/s40168-018-0472-x.

[65] J. S. Brazier, "Human infections with Fusobacterium necrophorum," *Anaerobe*, vol. 12, no. 4, pp. 165–172, Aug. 2006, doi: 10.1016/j.anaerobe.2005.11.003.

[66] C. Cueva, M. Silva, I. Pinillos, B. Bartolomé, and M. V. Moreno-Arribas, "Interplay between Dietary Polyphenols and Oral and Gut Microbiota in the Development of Colorectal Cancer," *Nutrients*, vol. 12, no. 3, 2020, doi: 10.3390/nu12030625.

[67] G. Longhi, D. van Sinderen, M. Ventura, and F. Turroni, "Microbiota and Cancer: The Emerging Beneficial Role of Bifidobacteria in Cancer Immunotherapy," *Front. Microbiol.*, vol. 11, 2020, doi: 10.3389/fmicb.2020.575072.

[68] A. Heine, G. Herrmann, T. Selmer, F. Terwesten, and K. Reuter, "High resolution crystal structure of*Clostridium propionicum*β-alanyl-CoA:ammonia lyase, a new member of the 'hot dog fold' protein superfamily," *Proteins Struct. Funct. Bioinforma.*, vol. 82, no. 9, 2014, doi: 10.1002/prot.24557.

[69] Z. Zhou, J. Chen, H. Yao, and H. Hu, "Fusobacterium and Colorectal Cancer," *Front. Oncol.*, vol. 8, 2018, doi: 10.3389/fonc.2018.00371.

[70] J.-J. P. Kim and R. Miura, "Acyl-CoA dehydrogenases and acyl-CoA oxidases," *Eur. J. Biochem.*, vol. 271, no. 3, 2004, doi: 10.1046/j.1432-1033.2003.03948.x.

[71] C. Thorpe and J.-J. P. Kim, "Structure and mechanism of action of the Acyl-CoA dehydrogenases [1]," *FASEB J.*, vol. 9, no. 9, 1995, doi: 10.1096/fasebj.9.9.7601336.

[72] M. P. Oliveira and M. Liesa, "The Role of Mitochondrial Fat Oxidation in Cancer Cell Proliferation and Survival," *Cells*, vol. 9, no. 12, 2020, doi: 10.3390/cells9122600.

[73] H. Li *et al.*, "DHA promotes healthy adipose tissue remodeling and carbohydrate utilization in grass carp (*Ctenopharyngodon idellus*) and largemouth bass (*Micropterus salmoides*)," *Aquac. Rep.*, vol. 33, p. 101856, Dec. 2023, doi: 10.1016/j.aqrep.2023.101856.

[74] T. Tabuchi, Y. Ito, A. Ioka, I. Miyashiro, and H. Tsukuma, "Incidence of metachronous second primary cancers in Osaka, Japan: Update of analyses using population-based cancer registry data," *Cancer Sci.*, vol. 103, no. 6, 2012, doi: 10.1111/j.1349-7006.2012.02254.x.

[75] D. Schomburg and M. Salzmann, “2-Dehydro-3-deoxy-L-arabinonate dehydratase,” *Enzyme Handb. 1*, 1990, doi: 10.1007/978-3-642-86605-0_167.
[76] J. Heider, “A new family of CoA-transferases,” *FEBS Lett.*, vol. 509, no. 3, 2001, doi: 10.1016/s0014-5793(01)03178-7.
[77] J. N. Poynter *et al.*, “Statins and the Risk of Colorectal Cancer,” *N. Engl. J. Med.*, vol. 352, no. 21, 2005, doi: 10.1056/nejmoa043792.
[78] S.-G. Yeo *et al.*, “Hydroxymethylglutaryl-Coenzyme A Synthase 2 Expression Is Associated With Chemoradiotherapy Responses in Colorectal Cancer,” *Dis. Colon Amp Rectum*, vol. 55, no. 6, 2012, doi: 10.1097/dcr.0b013e3182505080.
[79] C. Ye *et al.*, “Targeting the NAD+ salvage pathway suppresses APC mutation-driven colorectal cancer growth and Wnt/β-catenin signaling via increasing Axin level,” *Cell Commun. Signal.*, vol. 18, no. 1, 2020, doi: 10.1186/s12964-020-0513-5.
[80] B. Chen *et al.*, “MicroRNA-26a regulates glucose metabolism by direct targeting PDHX in colorectal cancer cells,” *BMC Cancer*, vol. 14, no. 1, 2014, doi: 10.1186/1471-2407-14-443.
[81] F. Mekhalif *et al.*, “Peptostreptococcus faecalis sp. nov., new bacterial species isolated from healthy indigenous congolese volunteer,” *Heliyon*, vol. 8, no. 3, p. e09102, Mar. 2022, doi: 10.1016/j.heliyon.2022.e09102.
[82] G. Zeller *et al.*, “Potential of fecal microbiota for early-stage detection of colorectal cancer,” *Mol. Syst. Biol.*, vol. 10, no. 11, p. 766, Nov. 2014, doi: 10.15252/msb.20145645.
[83] C. Peng, N. Li, Y. Hu, and N. Lü, “Impact factors that modulate gastric cancer risk in *Helicobacter pylori*-infected rodent models,” *Helicobacter*, vol. 24, no. 4, 2019, doi: 10.1111/hel.12580.
[84] S. Han *et al.*, “Using whole-genome sequencing (WGS) to plot colorectal cancer-related gut microbiota in a population with varied geography,” *Gut Pathog.*, vol. 14, no. 1, 2022, doi: 10.1186/s13099-022-00524-x.
[85] Y. Xie *et al.*, “Microbiota in Tumors: From Understanding to Application,” *Adv. Sci.*, vol. 9, no. 21, 2022, doi: 10.1002/advs.202200470.
[86] X. Liu *et al.*, “Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer,” *EBioMedicine*, vol. 40, 2019, doi: 10.1016/j.ebiom.2018.12.034.
[87] X. Shen *et al.*, “Fecal Enterotoxigenic Bacteroides fragilis–Peptostreptococcus stomatis–Parvimonas micra Biomarker for Noninvasive Diagnosis and Prognosis of Colorectal Laterally Spreading Tumor,” *Front. Oncol.*, vol. 11, 2021, Accessed: Jun. 17, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fonc.2021.661048
[88] L. C. DeDecker, B. Coppedge, J. Avelar-Barragan, W. E. Karnes, and K. Whiteson, “Microbiome distinctions between the CRC carcinogenic pathways,” *Gut Microbes*, vol. 13, no. 1, 2021, doi: 10.1080/19490976.2020.1854641.
[89] D. P. P. Pires, H. Oliveira, L. D. R. Melo, S. Sillankorva, and J. Azeredo, “Bacteriophage-encoded depolymerases: their diversity and biotechnological applications,” *Appl. Microbiol. Biotechnol.*, vol. 100, no. 5, 2016, doi: 10.1007/s00253-015-7247-0.
[90] O. O. Coker *et al.*, “Mucosal microbiome dysbiosis in gastric carcinogenesis,” *Gut*, vol. 67, no. 6, 2017, doi: 10.1136/gutjnl-2017-314281.
[91] T. Zhang *et al.*, “Research trends on the relationship between Microbiota and Gastric Cancer: A Bibliometric Analysis from 2000 to 2019,” *J. Cancer*, vol. 11, no. 16, 2020, doi: 10.7150/jca.44126.

[92] A. Matthies, M. Blaut, and A. Braune, “Isolation of a Human Intestinal Bacterium Capable of Daidzein and Genistein Conversion,” *Appl. Environ. Microbiol.*, vol. 75, no. 6, 2009, doi: 10.1128/aem.01795-08.
[93] X. Zhang *et al.*, “Caecal microbiota could effectively increase chicken growth performance by regulating fat metabolism,” *Microb. Biotechnol.*, vol. 15, no. 3, 2021, doi: 10.1111/1751-7915.13841.
[94] M. Contaldo *et al.*, “Oral Microbiota and Salivary Levels of Oral Pathogens in Gastro-Intestinal Diseases: Current Knowledge and Exploratory Study,” *Microorganisms*, vol. 9, no. 5, 2021, doi: 10.3390/microorganisms9051064.
[95] J. Dong, Y. Li, H. Xiao, M. Cui, and S. Fan, “Commensal microbiota in the digestive tract: a review of its roles in carcinogenesis and radiotherapy,” *Cancer Biol. Med.*, vol. 18, no., 2021, doi: 10.20892/j.issn.2095-3941.2020.0476.
[96] Y. Yang *et al.*, “Prospective study of oral microbiome and colorectal cancer risk in low-income and African American populations,” *Int. J. Cancer*, vol. 144, no. 10, 2018, doi: 10.1002/ijc.31941.
[97] M. Kuboniwa *et al.*, “Quantitative detection of periodontal pathogens using real-time polymerase chain reaction with TaqMan probes,” *Oral Microbiol. Immunol.*, vol. 19, no. 3, 2004, doi: 10.1111/j.0902-0055.2004.00135.x.
[98] Q. Sheng *et al.*, “Characteristics of fecal gut microbiota in patients with colorectal cancer at different stages and different sites,” *Oncol. Lett.*, 2019, doi: 10.3892/ol.2019.10841.
[99] S. Obuya *et al.*, “A signature of Prevotella copri and Faecalibacterium prausnitzii depletion, and a link with bacterial glutamate degradation in the Kenyan colorectal cancer patients,” *J. Gastrointest. Oncol.*, vol. 13, no. 5, 2022, doi: 10.21037/jgo-22-116.
[100] S. W. Ryu *et al.*, “Gut Microbiota Eubacterium callanderi Exerts Anti-Colorectal Cancer Activity,” *Microbiol. Spectr.*, vol. 10, no. 6, 2022, doi: 10.1128/spectrum.02531-22.
[101] A. Kensara, E. Hefni, M. Williams, H. Saito, E. F. Mongodin, and R. Masri, “Microbiological Profile and Human Immune Response Associated with Peri-Implantitis: A Systematic Review,” *J. Prosthodont.*, vol. 30, no. 3, 2020, doi: 10.1111/jopr.13270.
[102] W. Phannasorn, A. Pharapirom, P. Thiennimitr, H. Guo, S. Ketnawa, and R. Wongpoomchai, “Enriched Riceberry Bran Oil Exerts Chemopreventive Properties through Anti-Inflammation and Alteration of Gut Microbiota in Carcinogen-Induced Liver and Colon Carcinogenesis in Rats,” *Cancers*, vol. 14, no. 18, 2022, doi: 10.3390/cancers14184358.
[103] M. Bellinzoni *et al.*, “3-Keto-5-aminohexanoate Cleavage Enzyme,” *J. Biol. Chem.*, vol. 286, no. 31, 2011, doi: 10.1074/jbc.m111.253260.
[104] M. Sakamoto, I. N. Rôças, J. F. Siqueira, and Y. Benno, “Molecular analysis of bacteria in asymptomatic and symptomatic endodontic infections,” *Oral Microbiol. Immunol.*, vol. 21, no. 2, 2006, doi: 10.1111/j.1399-302x.2006.00270.x.
[105] P. Yadav *et al.*, “Carboxypeptidase in prolyl oligopeptidase family: Unique enzyme activation and substrate-screening mechanisms,” *J. Biol. Chem.*, vol. 294, no. 1, 2019, doi: 10.1074/jbc.ra118.004254.
[106] H. Wright, A. Kiss, Z. Szeltner, L. Polgár, and V. Fülöp, “Crystallization and preliminary crystallographic analysis of porcine acylaminoacyl peptidase,” *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.*, vol. 61, no. 10, 2005, doi: 10.1107/s1744309105029222.

[107] A. M. Melo, T. M. Bandeiras, and M. Teixeira, “New Insights into Type II NAD(P)H:Quinone Oxidoreductases,” *Microbiol. Mol. Biol. Rev.*, vol. 68, no. 4, 2004, doi: 10.1128/mmbr.68.4.603-616.2004.
[108] A. Heikal *et al.*, “Structure of the bacterial type <scp>II NADH</scp> dehydrogenase: a monotopic membrane protein with an essential role in energy generation,” *Mol. Microbiol.*, vol. 91, no. 5, 2014, doi: 10.1111/mmi.12507.
[109] J. S. Park *et al.*, “A heteroplasmic, not homoplasmic, mitochondrial DNA mutation promotes tumorigenesis via alteration in reactive oxygen species generation and apoptosis,” *Hum. Mol. Genet.*, vol. 18, no. 9, 2009, doi: 10.1093/hmg/ddp069.
[110] H. Sun *et al.*, “Increased nicotinamide adenine dinucleotide pool promotes colon cancer progression by suppressing reactive oxygen species level,” *Cancer Sci.*, vol. 110, no. 2, 2018, doi: 10.1111/cas.13886.
[111] J. Márquez *et al.*, “NADH dehydrogenase complex�I is overexpressed in incipient metastatic murine colon cancer cells,” *Oncol. Rep.*, 2018, doi: 10.3892/or.2018.6892.
[112] W. Jelski, B. Zalewski, L. Chrostek, and M. Szmitkowski, “Alcohol dehydrogenase (ADH) isoenzymes and aldehyde dehydrogenase (ALDH) activity in the sera of patients with colorectal cancer,” *Clin. Exp. Med.*, vol. 7, no. 4, 2007, doi: 10.1007/s10238-007-0140-0.
[113] C.-C. Lin, L.-T. Liu, L.-H. Ou, S.-C. Pan, C.-I. Lin, and Y.-H. Wei, “Role of mitochondrial function in the invasiveness of human colon cancer cells,” *Oncol. Rep.*, 2017, doi: 10.3892/or.2017.6087.
[114] J. S. Lee *et al.*, “Targeting Oxidative Phosphorylation Reverses Drug Resistance in Cancer Cells by Blocking Autophagy Recycling,” *Cells*, vol. 9, no. 9, 2020, doi: 10.3390/cells9092013.
[115] C. J. R. Frazão, N. Wagner, K. Rabe, and T. Walther, “Construction of a synthetic metabolic pathway for biosynthesis of 2,4-dihydroxybutyric acid from ethylene glycol,” *Nat. Commun.*, vol. 14, no. 1, 2023, doi: 10.1038/s41467-023-37558-x.
[116] C. Larsson and E. Albertsson, “Enzymes Related to Serine Synthesis in Spinach Chloroplasts,” *Physiol. Plant.*, vol. 45, no. 1, 1979, doi: 10.1111/j.1399-3054.1979.tb01654.x.
[117] A. M. Intlekofer *et al.*, “Hypoxia Induces Production of L-2-Hydroxyglutarate,” *Cell Metab.*, vol. 22, no. 2, 2015, doi: 10.1016/j.cmet.2015.06.023.
[118] M. Godel *et al.*, “Targeting Mitochondrial Oncometabolites: A New Approach to Overcome Drug Resistance in Cancer,” *Pharmaceutics*, vol. 13, no. 5, 2021, doi: 10.3390/pharmaceutics13050762.
[119] L. Chen, H.-X. Chen, Y. Li, L. Li, Y. Qiu, and J. Ren, “Endocannabinoid and ceramide levels are altered in patients with colorectal cancer,” *Oncol. Rep.*, vol. 34, no. 1, 2015, doi: 10.3892/or.2015.3973.
[120] Y.-Y. Liu, T. Han, A. E. Giuliano, and M. C. Cabot, “Ceramide glycosylation potentiates cellular multidrug resistance,” *FASEB J.*, vol. 15, no. 3, 2001, doi: 10.1096/fj.00-0223com.
[121] M. K. Yandım, E. Apohan, and Y. Baran, “Therapeutic potential of targeting ceramide/glucosylceramide pathway in cancer,” *Cancer Chemother. Pharmacol.*, vol. 71, no. 1, 2012, doi: 10.1007/s00280-012-1984-x.
[122] M. R. Lundquist *et al.*, “Phosphatidylinositol-5-Phosphate 4-Kinases Regulate Cellular Lipid Metabolism By Facilitating Autophagy,” *Mol. Cell*, vol. 70, no. 3, 2018, doi: 10.1016/j.molcel.2018.03.037.
[123] G. G. Aldred *et al.*, “The rational design of ARUK2007145, a dual inhibitor of the α and γ isoforms of the lipid kinase phosphatidylinositol 5-phosphate 4-kinase (PI5P4K),” *RSC Med. Chem.*, vol. 14, no. 10, 2023, doi: 10.1039/d3md00355h.

[124] B. D. Wright *et al.*, "Development of a High-Throughput Screening Assay to Identify Inhibitors of the Lipid Kinase PIP5K1C," *SLAS Discov.*, vol. 20, no. 5, 2015, doi: 10.1177/1087057114564057.
[125] N. Durand *et al.*, "The phosphorylation status of PIP5K1C at serine 448 can be predictive for invasive ductal carcinoma of the breast," *Oncotarget*, vol. 9, no. 91, 2018, doi: 10.18632/oncotarget.26357.
[126] B. J. Stephens, H. Han, V. Gokhale, and D. D. V. Hoff, "PRL phosphatases as potential molecular targets in cancer," *Mol. Cancer Ther.*, vol. 4, no. 11, 2005, doi: 10.1158/1535-7163.mct-05-0248.
[127] H. Xu *et al.*, "PRL-3 improves colorectal cancer cell proliferation and invasion through IL-8 mediated glycolysis metabolism," *Int. J. Oncol.*, vol. 51, no. 4, 2017, doi: 10.3892/ijo.2017.4090.
[128] Y. Zhao *et al.*, "Identification and functional characterization of paxillin as a target of protein tyrosine phosphatase receptor T," *Proc. Natl. Acad. Sci.*, vol. 107, no. 6, 2010, doi: 10.1073/pnas.0914884107.
[129] I. L.-I. de Mendoza, X. M. Mendia, A. M. G. de la Fuente, G. Quindós, and J. M. Aguirre-Urízar, "Role of *Porphyromonas gingivalis* in oral squamous cell carcinoma development: A systematic review," *J. Periodontal Res.*, vol. 55, no. 1, 2019, doi: 10.1111/jre.12691.
[130] W. Mu, Y. Jia, X. Chen, H. Li, Z. Wang, and B. Cheng, "Intracellular Porphyromonas gingivalis Promotes the Proliferation of Colorectal Cancer Cells via the MAPK/ERK Signaling Pathway," *Front. Cell. Infect. Microbiol.*, vol. 10, 2020, doi: 10.3389/fcimb.2020.584798.
[131] S. Gao *et al.*, "Presence of Porphyromonas gingivalis in esophagus and its association with the clinicopathological characteristics and survival in patients with esophageal cancer," *Infect. Agent. Cancer*, vol. 11, no. 1, 2016, doi: 10.1186/s13027-016-0049-x.
[132] H. Inaba *et al.*, "*Porphyromonas gingivalis*promotes invasion of oral squamous cell carcinoma through induction of proMMP9 and its activation," *Cell. Microbiol.*, vol. 16, no. 1, 2013, doi: 10.1111/cmi.12211.
[133] L. Wen *et al.*, "*Porphyromonas gingivalis* Promotes Oral Squamous Cell Carcinoma Progression in an Immune Microenvironment," *J. Dent. Res.*, vol. 99, no. 6, 2020, doi: 10.1177/0022034520909312.
[134] I. Wimplinger *et al.*, "Mutations of the Mitochondrial Holocytochrome c–Type Synthase in X-Linked Dominant Microphthalmia with Linear Skin Defects Syndrome," *Am. J. Hum. Genet.*, vol. 79, no. 5, 2006, doi: 10.1086/508474.
[135] J. M. Stevens, "Cytochrome c as an experimental model protein," *Metallomics*, vol. 3, no. 4, 2011, doi: 10.1039/c0mt00089b.
[136] T. Seo *et al.*, "Microsomal prostaglandin E synthase protein levels correlate with prognosis in colorectal cancer patients," *Virchows Arch.*, vol. 454, no. 6, 2009, doi: 10.1007/s00428-009-0777-z.
[137] L. Zeng, J. Qian, X. Luo, A. Zhou, Z. Zhang, and Q. Fang, "CHSY1 promoted proliferation and suppressed apoptosis in colorectal cancer through regulation of the NFκB and/or caspase-3/7 signaling pathway," *Oncol. Lett.*, 2018, doi: 10.3892/ol.2018.9385.
[138] S. Ogino *et al.*, "Cohort Study of Fatty Acid Synthase Expression and Patient Survival in Colon Cancer," *J. Clin. Oncol.*, vol. 26, no. 35, 2008, doi: 10.1200/jco.2008.18.2675.

[139] J. L. Eiseman *et al.*, “Distribution of 1-(2-Deoxy-2-fluoro-β-<scp>d</scp>-arabinofuranosyl) Uracil in Mice Bearing Colorectal Cancer Xenografts,” *Clin. Cancer Res.*, vol. 10, no. 19, 2004, doi: 10.1158/1078-0432.ccr-03-0686.
[140] C. Berlin *et al.*, “KMT9 Controls Stemness and Growth of Colorectal Cancer,” *Cancer Res.*, vol. 82, no. 2, 2022, doi: 10.1158/0008-5472.can-21-1261.
[141] Y. Chen *et al.*, “The role of histone methylation in the development of digestive cancers: a potential direction for cancer management,” *Signal Transduct. Target. Ther.*, vol. 5, no. 1, 2020, doi: 10.1038/s41392-020-00252-1.
[142] A. Biondi, F. Burzotta, and M. Vacante, “Familial adenomatous polyposis and changes in the gut microbiota: New insights into colorectal cancer carcinogenesis,” *World J. Gastrointest. Oncol.*, vol. 13, no. 6, 2021, doi: 10.4251/wjgo.v13.i6.495.
[143] F. Haghi, E. Goli, B. Mirzaei, and H. Zeighami, “The association between fecal enterotoxigenic B. fragilis with colorectal cancer,” *BMC Cancer*, vol. 19, no. 1, 2019, doi: 10.1186/s12885-019-6115-1.
[144] M. Oliero *et al.*, “Prevalence of pks + bacteria and enterotoxigenic Bacteroides fragilis in patients with colorectal cancer,” *Gut Pathog.*, vol. 14, no. 1, 2022, doi: 10.1186/s13099-022-00523-y.
[145] S. Hwang *et al.*, “Enterotoxigenic *Bacteroides fragilis* infection exacerbates tumorigenesis in AOM/DSS mouse model,” *Int. J. Med. Sci.*, vol. 17, no. 2, 2020, doi: 10.7150/ijms.38371.
[146] S. H. Ko *et al.*, “Bacteroides fragilis Enterotoxin Upregulates Heme Oxygenase-1 in Intestinal Epithelial Cells via a Mitogen-Activated Protein Kinase- and NF-κB-Dependent Pathway, Leading to Modulation of Apoptosis,” *Infect. Immun.*, vol. 84, no. 9, 2016, doi: 10.1128/iai.00191-16.
[147] D. Toumazi and C. Constantinou, “A Fragile Balance: The Important Role of the Intestinal Microbiota in the Prevention and Management of Colorectal Cancer,” *Oncology*, vol. 98, no. 9, 2020, doi: 10.1159/000507959.
[148] Q. Xu *et al.*, “Dual Role of *gnaA* in Antibiotic Resistance and Virulence in Acinetobacter baumannii,” *Antimicrob. Agents Chemother.*, vol. 63, no. 10, 2019, doi: 10.1128/aac.00694-19.
[149] H. M. Itkonen *et al.*, “UAP1 is overexpressed in prostate cancer and is protective against inhibitors of N-linked glycosylation,” *Oncogene*, vol. 34, no. 28, 2014, doi: 10.1038/onc.2014.307.
[150] M. Rizzi *et al.*, “GDP-4-keto-6-deoxy-D-mannose epimerase/reductase from Escherichia coli, a key enzyme in the biosynthesis of GDP-L-fucose, displays the structural characteristics of the RED protein homology superfamily,” *Structure*, vol. 6, no. 11, 1998, doi: 10.1016/s0969-2126(98)00144-0.
[151] Y. Wang *et al.*, “Fucosylation Deficiency in Mice Leads to Colitis and Adenocarcinoma,” *Gastroenterology*, vol. 152, no. 1, 2017, doi: 10.1053/j.gastro.2016.09.004.
[152] T. Ayukawa *et al.*, “Rescue of Notch signaling in cells incapable of GDP-<scp>l</scp> -fucose synthesis by gap junction transfer of GDP- <scp>l</scp> -fucose in *Drosophila*,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 38, 2012, doi: 10.1073/pnas.1202369109.
[153] M. M. Heemskerk, V. van Harmelen, K. W. van Dijk, and J. B. van Klinken, “Reanalysis of mGWAS results and in vitro validation show that lactate dehydrogenase interacts with branched-chain amino acid metabolism,” *Eur. J. Hum. Genet.*, vol. 24, no. 1, 2015, doi: 10.1038/ejhg.2015.106.
[154] Y. Peng, X. Zhao, C.-F. Zeng, J. Xu, L.-N. Qi, and L.-Q. Li, “Integrated omics analysis: the relationship between significantly increased Klebsiella post-

hepatectomy and decreased hub-metabolite 3-methyl-2-oxobutanoic acid is associated with induced liver failure," *J. Gastrointest. Oncol.*, vol. 13, no. 1, 2022, doi: 10.21037/jgo-21-906.

[155] M. Wnuk, P. Slipek, M. Dziedzic, and A. Lewińska, "The Roles of Host 5-Methylcytosine RNA Methyltransferases during Viral Infections," *Int. J. Mol. Sci.*, vol. 21, no. 21, 2020, doi: 10.3390/ijms21218176.

[156] M. Okamoto *et al.*, "Frequent Increased Gene Copy Number and High Protein Expression of tRNA (Cytosine-5-)-Methyltransferase (NSUN2) in Human Cancers," *DNA Cell Biol.*, vol. 31, no. 5, 2012, doi: 10.1089/dna.2011.1446.

[157] N. C. Bird, D. Mangnall, and A. W. Majeed, "Biology of colorectal liver metastases: A review," *J. Surg. Oncol.*, vol. 94, no. 1, 2006, doi: 10.1002/jso.20558.

[158] Y.-S. Kim *et al.*, "Lectin precipitation using phytohemagglutinin-L_4 coupled to avidin–agarose for serological biomarker discovery in colorectal cancer," *Proteomics*, vol. 8, no. 16, 2008, doi: 10.1002/pmic.200800034.

[159] M. W. Ruszczycky, S. H. Choi, and H. W. Liu, "Stoichiometry of the Redox Neutral Deamination and Oxidative Dehydrogenation Reactions Catalyzed by the Radical SAM Enzyme DesII," *J. Am. Chem. Soc.*, vol. 132, no. 7, 2010, doi: 10.1021/ja909451a.

[160] P. L. Taylor *et al.*, "Structure and Function of Sedoheptulose-7-phosphate Isomerase, a Critical Enzyme for Lipopolysaccharide Biosynthesis and a Target for Antibiotic Adjuvants," *J. Biol. Chem.*, vol. 283, no. 5, 2008, doi: 10.1074/jbc.m706163200.

[161] B. Kneidinger *et al.*, "Biosynthesis Pathway of ADP- <scp>l</scp> - *glycero* -β- <scp>d</scp> - *manno* -Heptose in *Escherichia coli*," *J. Bacteriol.*, vol. 184, no. 2, 2002, doi: 10.1128/jb.184.2.363-369.2002.

[162] D. F. Xiang, M. Xu, M. K. Ghosh, and F. M. Raushel, "Metabolic Pathways for the Biosynthesis of Heptoses Used in the Construction of Capsular Polysaccharides in the Human Pathogen *Campylobacter jejuni*," *Biochemistry*, vol. 62, no. 21, 2023, doi: 10.1021/acs.biochem.3c00390.

[163] P. G. Rychahou *et al.*, "Akt2 overexpression plays a critical role in the establishment of colorectal cancer metastasis," *Proc. Natl. Acad. Sci.*, vol. 105, no. 51, 2008, doi: 10.1073/pnas.0810715105.

[164] N. Very *et al.*, "Thymidylate synthase O-GlcNAcylation: a molecular mechanism of 5-FU sensitization in colorectal cancer," *Oncogene*, vol. 41, no. 5, 2021, doi: 10.1038/s41388-021-02121-9.

[165] Y. Zhou *et al.*, "And-1 O-GlcNAcylation regulates homologous recombination repair and radioresistance in colorectal cancer," *Clin. Transl. Med.*, vol. 12, no. 4, 2022, doi: 10.1002/ctm2.785.

[166] P. Chaiyawat, D. Chokchaichamnankit, K. Lirdprapamongkol, C. Srisomsap, J. Svasti, and V. Champattanachai, "Alteration of O-GlcNAcylation affects serine phosphorylation and regulates gene expression and activity of pyruvate kinase M2 in colorectal cancer cells," *Oncol. Rep.*, vol. 34, no. 4, 2015, doi: 10.3892/or.2015.4178.

[167] S. Lutsenko, E. S. LeShane, and U. Shinde, "Biochemical basis of regulation of human copper-transporting ATPases," *Arch. Biochem. Biophys.*, vol. 463, no. 2, 2007, doi: 10.1016/j.abb.2007.04.013.

[168] I. Kushkevych, M. A. Bychkov, S. Bychkova, M. Gajdács, R. O. Merza, and M. Vítězová, "ATPase Activity of the Subcellular Fractions of Colorectal Cancer Samples under the Action of Nicotinic Acid Adenine Dinucleotide Phosphate," *Biomedicines*, vol. 9, no. 12, 2021, doi: 10.3390/biomedicines9121805.

[169] H. Sakai *et al.*, "Up-regulation of Na^+,K^+-ATPase α3-isoform and down-regulation of the α1-isoform in human colorectal cancer," *FEBS Lett.*, vol. 563, no. 1–3, 2004, doi: 10.1016/s0014-5793(04)00292-3.
[170] Z. Liu *et al.*, "Investigation of Copper Cysteamine Nanoparticles as a New Type of Radiosensitiers for Colorectal Carcinoma Treatment," *Sci. Rep.*, vol. 7, no. 1, 2017, doi: 10.1038/s41598-017-09375-y.
[171] S. Materia, M. A. Cater, L. W. J. Klomp, J. F. B. Mercer, and S. L. Fontaine, "Clusterin (Apolipoprotein J), a Molecular Chaperone That Facilitates Degradation of the Copper-ATPases ATP7A and ATP7B," *J. Biol. Chem.*, vol. 286, no. 12, 2011, doi: 10.1074/jbc.m110.190546.
[172] P. Morgan, R. F. Sala, and M. E. Tanner, "Eliminations in the Reactions Catalyzed by UDP-*N*-Acetylglucosamine 2-Epimerase," *J. Am. Chem. Soc.*, vol. 119, no. 43, 1997, doi: 10.1021/ja971718q.
[173] K. W. Jackson *et al.*, "Suppression of Tumor Growth in Mice by Rationally Designed Pseudopeptide Inhibitors of Fibroblast Activation Protein and Prolyl Oligopeptidase," *Neoplasia*, vol. 17, no. 1, 2015, doi: 10.1016/j.neo.2014.11.002.
[174] Y. Waumans, L. Baerts, K. Kehoe, A.-M. Lambeir, and I. D. Meester, "The Dipeptidyl Peptidase Family, Prolyl Oligopeptidase, and Prolyl Carboxypeptidase in the Immune System and Inflammatory Disease, Including Atherosclerosis," *Front. Immunol.*, vol. 6, 2015, doi: 10.3389/fimmu.2015.00387.
[175] I. Huybrechts *et al.*, "The Human Microbiome in Relation to Cancer Risk: A Systematic Review of Epidemiologic Studies," *Cancer Epidemiol. Biomark. Amp Prev.*, vol. 29, no. 10, 2020, doi: 10.1158/1055-9965.epi-20-0288.
[176] J. Sun and I. Kato, "Gut microbiota, inflammation and colorectal cancer," *Genes Amp Dis.*, vol. 3, no. 2, 2016, doi: 10.1016/j.gendis.2016.03.004.
[177] E. Vogtmann *et al.*, "Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing," *Plos One*, vol. 11, no. 5, 2016, doi: 10.1371/journal.pone.0155362.
[178] E. Reitano *et al.*, "Oral Bacterial Microbiota in Digestive Cancer Patients: A Systematic Review," *Microorganisms*, vol. 9, no. 12, 2021, doi: 10.3390/microorganisms9122585.
[179] J. Yang *et al.*, "Development of a colorectal cancer diagnostic model and dietary risk assessment through gut microbiome analysis," *Exp. Amp Mol. Med.*, vol. 51, no. 10, 2019, doi: 10.1038/s12276-019-0313-4.
[180] I. Allali *et al.*, "Gut microbiome of Moroccan colorectal cancer patients," *Med. Microbiol. Immunol. (Berl.)*, vol. 207, no. 3–4, 2018, doi: 10.1007/s00430-018-0542-5.
[181] T. Kardon, G. Noël, D. Vertommen, and E. V. Schaftingen, "Identification of the gene encoding hydroxyacid-oxoacid transhydrogenase, an enzyme that metabolizes 4-hydroxybutyrate," *FEBS Lett.*, vol. 580, no. 9, 2006, doi: 10.1016/j.febslet.2006.02.082.
[182] F. A. Dick, S. Karamanou, and B. L. Trumpower, "QSR1, an Essential Yeast Gene with a Genetic Relationship to a Subunit of the Mitochondrial Cytochromebc 1 Complex, Codes for a 60 S Ribosomal Subunit Protein," *J. Biol. Chem.*, vol. 272, no. 20, 1997, doi: 10.1074/jbc.272.20.13372.
[183] J. W. Locasale, "Serine, glycine and one-carbon units: cancer metabolism in full circle," *Nat. Rev. Cancer*, vol. 13, no. 8, 2013, doi: 10.1038/nrc3557.
[184] I. Amelio, F. Cutruzzolà, A. Антонов, M. Agostini, and G. Melino, "Serine and glycine metabolism in cancer," *Trends Biochem. Sci.*, vol. 39, no. 4, 2014, doi: 10.1016/j.tibs.2014.02.004.

[185] A. Pompella, A. Corti, and A. Visvikis, “Redox Mechanisms in Cisplatin Resistance of Cancer Cells: The Twofold Role of Gamma-Glutamyltransferase 1 (GGT1),” *Front. Oncol.*, vol. 12, 2022, doi: 10.3389/fonc.2022.920316.
[186] J.-R. Wei, J. Dong, and L. Li, “Cancer-associated fibroblasts-derived gamma-glutamyltransferase 5 promotes tumor growth and drug resistance in lung adenocarcinoma,” *Aging*, vol. 12, no. 13, 2020, doi: 10.18632/aging.103429.
[187] X. Xu *et al.*, “Tumor suppressor NDRG2 inhibits glycolysis and glutaminolysis in colorectal cancer cells by repressing c-Myc expression,” *Oncotarget*, vol. 6, no. 28, 2015, doi: 10.18632/oncotarget.4544.
[188] Y. Zhao *et al.*, “Colorectal cancers utilize glutamine as an anaplerotic substrate of the TCA cycle in vivo,” *Sci. Rep.*, vol. 9, no. 1, 2019, doi: 10.1038/s41598-019-55718-2.
[189] B. Xu, L. Cai, H. Xiao, G. Cai, P. Sun, and G. Huang, “SIRT4 enhances the sensitivity of colorectal cancer cells to 5-FU by inhibiting ammonia-mediated autophagy.” Jan. 23, 2023. doi: 10.21203/rs.3.rs-1695035/v2.
[190] S. V. Reddy, S. S. More, and G. S. Annappa, “Purification and properties of beta-cyclomaltodextrin glucanotransferase from *Bacillus flexus* SV 1,” *J. Basic Microbiol.*, vol. 57, no. 11, 2017, doi: 10.1002/jobm.201700270.
[191] X. Wang *et al.*, “Deletion of AS87_03730 gene changed the bacterial virulence and gene expression of Riemerella anatipestifer,” *Sci. Rep.*, vol. 6, no. 1, 2016, doi: 10.1038/srep22438.
[192] Y. Huang, J. Zhou, H. Zhong, N. Xie, F. Zhang, and Z. Zhang, “Identification of a novel lipid metabolism-related gene signature for predicting colorectal cancer survival,” *Front. Genet.*, vol. 13, 2022, doi: 10.3389/fgene.2022.989327.
[193] J. Lu *et al.*, “<scp>DPM2</scp> serve as novel oncogene and prognostic marker transactivated by <scp>ESR1</scp> in breast cancer,” *Environ. Toxicol.*, vol. 39, no. 3, 2023, doi: 10.1002/tox.24059.
[194] A. K. Cotte *et al.*, “Lysophosphatidylcholine acyltransferase 2-mediated lipid droplet production supports colorectal cancer chemoresistance,” *Nat. Commun.*, vol. 9, no. 1, 2018, doi: 10.1038/s41467-017-02732-5.
[195] S. Han, Y. Xi, and Y. Pan, “Can Mitochondria DNA Provide a Novel Biomarker for Evaluating the Risk and Prognosis of Colorectal Cancer?,” *Dis. Markers*, vol. 2017, 2017, doi: 10.1155/2017/5189803.
[196] P. Louis and H. J. Flint, “Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine,” *FEMS Microbiol. Lett.*, vol. 294, no. 1, 2009, doi: 10.1111/j.1574-6968.2009.01514.x.
[197] A. Rivière, M. Gagnon, S. Weckx, D. Roy, and L. D. Vuyst, “Mutual Cross-Feeding Interactions between Bifidobacterium longum subsp. longum NCC2705 and Eubacterium rectale ATCC 33656 Explain the Bifidogenic and Butyrogenic Effects of Arabinoxylan Oligosaccharides,” *Appl. Environ. Microbiol.*, vol. 81, no. 22, 2015, doi: 10.1128/aem.02089-15.
[198] P. Louis and H. J. Flint, “Development of a Semiquantitative Degenerate Real-Time PCR-Based Assay for Estimation of Numbers of Butyryl-Coenzyme A (CoA) CoA Transferase Genes in Complex Bacterial Samples,” *Appl. Environ. Microbiol.*, vol. 73, no. 6, 2007, doi: 10.1128/aem.02561-06.
[199] C. Charrier *et al.*, “A novel class of CoA-transferase involved in short-chain fatty acid metabolism in butyrate-producing human colonic bacteria,” *Microbiology*, vol. 152, no. 1, 2006, doi: 10.1099/mic.0.28412-0.
[200] B. Hippe, J. Zwielehner, K. Liszt, C. Laßl, F. M. Unger, and A. G. Haslberger, “Quantification of butyryl CoA:acetate CoA-transferase genes reveals different

butyrate production capacity in individuals according to diet and age," *FEMS Microbiol. Lett.*, vol. 316, no. 2, 2011, doi: 10.1111/j.1574-6968.2010.02197.x.
[201] X. B. Lin *et al.*, "The Role of Intestinal Microbiota in Development of Irinotecan Toxicity and in Toxicity Reduction through Dietary Fibres in Rats," *PLoS ONE*, vol. 9, no. 1, 2014, doi: 10.1371/journal.pone.0083644.
[202] A. Han, N. E. Bennett, A. Bettaieb, J. Whelan, and D. R. Donohoe, "Butyrate decreases its own oxidation in colorectal cancer cells through inhibition of histone deacetylases," *Oncotarget*, vol. 9, no. 43, 2018, doi: 10.18632/oncotarget.25546.
[203] D. Wang, L.-W. Zou, Q. Jin, J. Hou, G. Ge, and L. Yang, "Human carboxylesterases: a comprehensive review," *Acta Pharm. Sin. B*, vol. 8, no. 5, 2018, doi: 10.1016/j.apsb.2018.05.005.
[204] J. Potempa and R. N. Pike, "Corruption of Innate Immunity by Bacterial Proteases," *J. Innate Immun.*, vol. 1, no. 2, 2008, doi: 10.1159/000181144.
[205] N. Kunová *et al.*, "Mitochondrial Processing Peptidases—Structure, Function and the Role in Human Diseases," *Int. J. Mol. Sci.*, vol. 23, no. 3, 2022, doi: 10.3390/ijms23031297.
[206] K. Ohshima, R. Oi, S. Nojima, and E. Morii, "Mitochondria govern histone acetylation in colorectal cancer," *J. Pathol.*, vol. 256, no. 2, 2021, doi: 10.1002/path.5818.
[207] J. Kühnl *et al.*, "Functional analysis of the methylmalonyl-CoA epimerase from *Caenorhabditis elegans*," *FEBS J.*, vol. 272, no. 6, 2005, doi: 10.1111/j.1742-4658.2005.04579.x.
[208] R. J. Chandler *et al.*, "Propionyl-CoA and adenosylcobalamin metabolism in Caenorhabditis elegans: Evidence for a role of methylmalonyl-CoA epimerase in intermediary metabolism," *Mol. Genet. Metab.*, vol. 89, no. 1–2, 2006, doi: 10.1016/j.ymgme.2006.06.001.
[209] R. L. Siegel *et al.*, "Colorectal cancer statistics, 2020," *CA. Cancer J. Clin.*, vol. 70, no. 3, 2020, doi: 10.3322/caac.21601.
[210] S. Bornemann, "Flavoenzymes that catalyse reactions with no net redox change," *Nat. Prod. Rep.*, vol. 19, no. 6, 2002, doi: 10.1039/b108916c.
[211] G. Li, Z. Wang, J. Xu, H. Wu, and S. Cai, "The prognostic value of lactate dehydrogenase levels in colorectal cancer: a meta-analysis," *BMC Cancer*, vol. 16, no. 1, 2016, doi: 10.1186/s12885-016-2276-3.
[212] C. M. Metallo, "Expanding the Reach of Cancer Metabolomics," *Cancer Prev. Res. (Phila. Pa.)*, vol. 5, no. 12, 2012, doi: 10.1158/1940-6207.capr-12-0433.
[213] T. Yoshie *et al.*, "Regulation of the metabolite profile by an <scp>*APC*</scp> gene mutation in colorectal cancer," *Cancer Sci.*, vol. 103, no. 6, 2012, doi: 10.1111/j.1349-7006.2012.02262.x.
[214] S. Bae *et al.*, "Plasma Choline Metabolites and Colorectal Cancer Risk in the Women's Health Initiative Observational Study," *Cancer Res.*, vol. 74, no. 24, 2014, doi: 10.1158/0008-5472.can-14-1835.
[215] R. Makuku *et al.*, "The Role of Ketogenic Diet in the Treatment of Neuroblastoma," *Integr. Cancer Ther.*, vol. 22, 2023, doi: 10.1177/15347354221150787.
[216] X. Chen *et al.*, "HMGCL-induced β-Hydroxybutyrate Production Attenuates Hepatocellular Carcinoma via DPP4-mediated Ferroptosis Susceptibility," scite.ai. Accessed: Mar. 18, 2024. [Online]. Available: https://doi.org/10.21203/rs.3.rs-2055839/v1
[217] M. Bonnet *et al.*, "Colonization of the Human Gut by *E. coli* and Colorectal Cancer Risk," *Clin. Cancer Res.*, vol. 20, no. 4, 2014, doi: 10.1158/1078-0432.ccr-13-1343.

[218] T. Faïs, J. Delmas, N. Barnich, R. Bonnet, and G. Dalmasso, "Colibactin: More Than a New Bacterial Toxin," *Toxins*, vol. 10, no. 4, 2018, doi: 10.3390/toxins10040151.
[219] E. A. Alhinai, G. Walton, and D. M. Commane, "The Role of the Gut Microbiota in Colorectal Cancer Causation," *Int. J. Mol. Sci.*, vol. 20, no. 21, 2019, doi: 10.3390/ijms20215295.
[220] C. P. Kaur, J. Vadivelu, and C. Samudi, "Impact of <scp>*Klebsiella pneumoniae*</scp> in lower gastrointestinal tract diseases," *J. Dig. Dis.*, vol. 19, no. 5, 2018, doi: 10.1111/1751-2980.12595.
[221] M. A. Hernández-Luna, S. López-Briones, and R. Luría-Pérez, "The Four Horsemen in Colon Cancer," *J. Oncol.*, vol. 2019, 2019, doi: 10.1155/2019/5636272.
[222] M. Prorok-Hamon *et al.*, "Colonic mucosa-associated diffusely adherent*afaC*+ *Escherichia coli*expressing*lpfA*and*pks*are increased in inflammatory bowel disease and colon cancer," *Gut*, vol. 63, no. 5, 2013, doi: 10.1136/gutjnl-2013-304739.
[223] C. Zhang, J. Tong, and G. Huang, "Nicotinamide Phosphoribosyl Transferase (Nampt) Is a Target of MicroRNA-26b in Colorectal Cancer Cells," *PLoS ONE*, vol. 8, no. 7, 2013, doi: 10.1371/journal.pone.0069963.
[224] J. Turyn *et al.*, "Increased Activity of Glycerol 3-phosphate Dehydrogenase and Other Lipogenic Enzymes in Human Bladder Cancer," *Horm. Metab. Res.*, vol. 35, no. 10, 2003, doi: 10.1055/s-2003-43500.
[225] M. E. van der Rest, C. Frank, and D. Molenaar, "Functions of the Membrane-Associated and Cytoplasmic Malate Dehydrogenases in the Citric Acid Cycle of *Escherichia coli*," *J. Bacteriol.*, vol. 182, no. 24, 2000, doi: 10.1128/jb.182.24.6892-6899.2000.
[226] E. A. Hanse, C. Ruan, M. Kachman, D. Wang, X. H. Lowman, and A. Kelekar, "Cytosolic malate dehydrogenase activity helps support glycolysis in actively proliferating cells and cancer," *Oncogene*, vol. 36, no. 27, 2017, doi: 10.1038/onc.2017.36.
[227] B. Landfald and A. R. Strøm, "Choline-glycine betaine pathway confers a high level of osmotic tolerance in Escherichia coli," *J. Bacteriol.*, vol. 165, no. 3, 1986, doi: 10.1128/jb.165.3.849-855.1986.
[228] Y.-S. Ma, W. Li, Y. Liu, Y. Shi, Q. Lin, and D. Fu, "Targeting Colorectal Cancer Stem Cells as an Effective Treatment for Colorectal Cancer," *Technol. Cancer Res. Amp Treat.*, vol. 19, 2020, doi: 10.1177/1533033819892261.
[229] H. Mirzaei *et al.*, "Deciphering biological characteristics of tumorigenic subpopulations in human colorectal cancer reveals cellular plasticity," *J. Res. Med. Sci.*, vol. 21, no. 1, 2016, doi: 10.4103/1735-1995.187355.
[230] Y. Murawaki *et al.*, "Aberrant expression of selenoproteins in the progression of colorectal cancer," *Cancer Lett.*, vol. 259, no. 2, 2008, doi: 10.1016/j.canlet.2007.10.019.
[231] U. Peters *et al.*, "Variation in the Selenoenzyme Genes and Risk of Advanced Distal Colorectal Adenoma," *Cancer Epidemiol. Biomark. Amp Prev.*, vol. 17, no. 5, 2008, doi: 10.1158/1055-9965.epi-07-2947.
[232] Y. Huang, W. Fang, Y. Wang, W. Yang, and B. Xiong, "Transforming growth factor-β1 induces glutathione peroxidase-1 and protects from H2O2-induced cell death in colon cancer cells via the Smad2/ERK1/2/HIF-1α pathway," *Int. J. Mol. Med.*, 2012, doi: 10.3892/ijmm.2012.901.
[233] F. Lu, "Hypothetical Hydrogenase Activity of Human Mitochondrial Complex I and Its Role in Preventing Cancer Transformation," *Explor. Res. Hypothesis Med.*, vol. 000, no. 000, 2022, doi: 10.14218/erhm.2022.00083.

[234] K. Saito and A. Komamine, “Biosynthesis of Stizolobinic Acid and Stizolobic Acid in Higher Plants,” *Eur. J. Biochem.*, vol. 82, no. 2, 1978, doi: 10.1111/j.1432-1033.1978.tb12033.x.
[235] N. C. Andrés *et al.*, “Heme oxygenase-1 has antitumoral effects in colorectal cancer: Involvement of p53,” *Exp. Mol. Pathol.*, vol. 97, no. 3, 2014, doi: 10.1016/j.yexmp.2014.09.012.
[236] S. Tan *et al.*, “The Mechanisms of Ferroptosis and the Applications in Tumor Treatment: Enemies or Friends?,” *Front. Mol. Biosci.*, vol. 9, 2022, doi: 10.3389/fmolb.2022.938677.
[237] E. L. Connolly, J. P. Fett, and M. L. Guerinot, “Expression of the IRT1 Metal Transporter Is Controlled by Metals at the Levels of Transcript and Protein Accumulation,” *Plant Cell*, vol. 14, no. 6, 2002, doi: 10.1105/tpc.001263.
[238] E. H. Verbon, P. Trapet, I. A. Stringlis, S. Kruijs, P. A. H. M. Bakker, and C. M. J. Pieterse, “Iron and Immunity,” *Annu. Rev. Phytopathol.*, vol. 55, no. 1, 2017, doi: 10.1146/annurev-phyto-080516-035537.
[239] A. Fonseca-Nunes, P. Jakszyn, and A. Agudo, “Iron and Cancer Risk—A Systematic Review and Meta-analysis of the Epidemiological Evidence,” *Cancer Epidemiol. Biomark. Amp Prev.*, vol. 23, no. 1, 2014, doi: 10.1158/1055-9965.epi-13-0733.
[240] G. Yin *et al.*, “*Alcohol dehydrogenase* and *aldehyde dehydrogenase* polymorphisms and colorectal cancer: The Fukuoka Colorectal Cancer Study,” *Cancer Sci.*, vol. 98, no. 8, 2007, doi: 10.1111/j.1349-7006.2007.00519.x.
[241] B. C. Yoo *et al.*, “Identification of hypoxanthine as a urine marker for non-Hodgkin lymphoma by low-mass-ion profiling,” *BMC Cancer*, vol. 10, no. 1, 2010, doi: 10.1186/1471-2407-10-55.
[242] M. Toyota, N. Ahuja, M. Ohe-Toyota, J. G. Herman, S. B. Baylin, and J.-P. J. Issa, “CpG island methylator phenotype in colorectal cancer,” *Proc. Natl. Acad. Sci.*, vol. 96, no. 15, 1999, doi: 10.1073/pnas.96.15.8681.
[243] Z. Liu, L.-M. Li, H.-L. Sun, and S.-M. Liu, “Link Between m6A Modification and Cancers,” *Front. Bioeng. Biotechnol.*, vol. 6, 2018, doi: 10.3389/fbioe.2018.00089.
[244] H. Li, Z. Liu, and H. Wang, “Expression and clinical significance of METTL3 in colorectal cancer,” *Medicine (Baltimore)*, vol. 102, no. 37, 2023, doi: 10.1097/md.0000000000034658.
[245] Y. Zhou *et al.*, “m6A methyltransferase KIAA1429 acts as an oncogenic factor in colorectal cancer by regulating SIRT1 in an m6A-dependent manner,” *Cell Death Discov.*, vol. 8, no. 1, 2022, doi: 10.1038/s41420-022-00878-w.
[246] P. Balijepalli, C. C. Sitton, and K. E. Meier, “Lysophosphatidic Acid Signaling in Cancer Cells: What Makes LPA So Special?,” *Cells*, vol. 10, no. 8, 2021, doi: 10.3390/cells10082059.
[247] C. Zhang *et al.*, “New dawn for cancer cell death: Emerging role of lipid metabolism,” *Mol. Metab.*, vol. 63, p. 101529, Sep. 2022, doi: 10.1016/j.molmet.2022.101529.
[248] F. Mansilla *et al.*, “Lysophosphatidylcholine acyltransferase 1 (LPCAT1) overexpression in human colorectal cancer,” *J. Mol. Med.*, vol. 87, no. 1, 2008, doi: 10.1007/s00109-008-0409-0.
[249] Z. Yu, Y. Zhou, Y. Li, and Z. Dong, “Integration of clinical and spatial data to explore lipid metabolism-related gene for predicting prognosis and immune microenvironment in gliomas,” scite.ai. Accessed: Mar. 22, 2024. [Online]. Available: https://doi.org/10.21203/rs.3.rs-2541686/v1

[250] L. Fernández *et al.*, "The role of glycosyltransferase enzyme GCNT3 in colon and ovarian cancer prognosis and chemoresistance," *Sci. Rep.*, vol. 8, no. 1, 2018, doi: 10.1038/s41598-018-26468-4.
[251] A. Blanas *et al.*, "Transcriptional activation of fucosyltransferase (FUT) genes using the CRISPR-dCas9-VPR technology reveals potent N-glycome alterations in colorectal cancer cells," *Glycobiology*, vol. 29, no. 2, 2018, doi: 10.1093/glycob/cwy096.
[252] M. Noda *et al.*, "Glycosyltransferase Gene Expression Identifies a Poor Prognostic Colorectal Cancer Subtype Associated with Mismatch Repair Deficiency and Incomplete Glycan Synthesis," *Clin. Cancer Res.*, vol. 24, no. 18, 2018, doi: 10.1158/1078-0432.ccr-17-3533.
[253] "Expression of UDP-GalNAc: polypeptide N -acetylgalactosaminyltransferase isozymes T1 and T2 in human colorectal cancer - [scite report]." Accessed: Mar. 23, 2024. [Online]. Available: https://scite.ai/reports/expression-of-udp-galnac-polypeptide-n-gjbj4z
[254] Ö. Kanbagli, G. Özdemirler, T. Bulut, S. Yamaner, G. Aykaç-Toker, and M. Uysal, "Mitochondrial Lipid Peroxides and Antioxidant Enzymes in Colorectal Adenocarcinoma Tissues," *Jpn. J. Cancer Res.*, vol. 91, no. 12, 2000, doi: 10.1111/j.1349-7006.2000.tb00912.x.
[255] C. Sachse *et al.*, "A pharmacogenetic study to investigate the role of dietary carcinogens in the etiology of colorectal cancer," *Carcinogenesis*, vol. 23, no. 11, 2002, doi: 10.1093/carcin/23.11.1839.
[256] L. L. Li *et al.*, "The Tumor Suppressor UCHL1 Forms a Complex with p53/MDM2/ARF to Promote p53 Signaling and Is Frequently Silenced in Nasopharyngeal Carcinoma," *Clin. Cancer Res.*, vol. 16, no. 11, 2010, doi: 10.1158/1078-0432.ccr-09-3178.
[257] A. Sakthianandeswaren *et al.*, "*MACROD2* Haploinsufficiency Impairs Catalytic Activity of PARP1 and Promotes Chromosome Instability and Growth of Intestinal Tumors," *Cancer Discov.*, vol. 8, no. 8, 2018, doi: 10.1158/2159-8290.cd-17-0909.
[258] S. Mamoor, "Differential expression of EPHX4 in colorectal cancer.," scite.ai. Accessed: Mar. 23, 2024. [Online]. Available: https://doi.org/10.31219/osf.io/dnr8f
[259] L. Chen, H.-X. Chen, Y. Li, L. Li, Y. Qiu, and J. Ren, "Endocannabinoid and ceramide levels are altered in patients with colorectal cancer," *Oncol. Rep.*, vol. 34, no. 1, 2015, doi: 10.3892/or.2015.3973.
[260] N. J. Pyne, F. Tonelli, K. G. Lim, J. Long, J. Edwards, and S. Pyne, "Sphingosine 1-phosphate signalling in cancer," *Biochem. Soc. Trans.*, vol. 40, no. 1, 2012, doi: 10.1042/bst20110602.
[261] Y. Yatomi, M. Kurano, H. Ikeda, K. Igarashi, K. Kano, and J. Aoki, "Lysophospholipids in laboratory medicine," *Proc. Jpn. Acad. Ser. B*, vol. 94, no. 10, 2018, doi: 10.2183/pjab.94.025.
[262] J. Basappa *et al.*, "ACLY is the novel signaling target of PIP2/PIP3 and Lyn in acute myeloid leukemia," *Heliyon*, vol. 6, no. 5, 2020, doi: 10.1016/j.heliyon.2020.e03910.
[263] L. Jin and Y. Zhou, "Crucial role of the pentose phosphate pathway in malignant tumors (Review)," *Oncol. Lett.*, 2019, doi: 10.3892/ol.2019.10112.
[264] L. Coppin, J. Leclerc, A. Vincent, N. Porchet, and P. Pigny, "Messenger RNA Life-Cycle in Cancer Cells: Emerging Role of Conventional and Non-Conventional RNA-Binding Proteins?," *Int. J. Mol. Sci.*, vol. 19, no. 3, 2018, doi: 10.3390/ijms19030650.

[265] K. Y. C. Fung *et al.*, "A combined free-flow electrophoresis and DIGE approach to identify proteins regulated by butyrate in HT29 cells," *Proteomics*, vol. 11, no. 5, 2011, doi: 10.1002/pmic.201000429.
[266] A. Curinha, S. Bráz, I. Pereira-Castro, A. Cruz, and A. Moreira, "Implications of polyadenylation in health and disease," *Nucleus*, vol. 5, no. 6, 2014, doi: 10.4161/nucl.36360.
[267] L. Liu, C. C. Wong, B. Gong, and J. Yu, "Functional significance and therapeutic implication of ring-type E3 ligases in colorectal cancer," *Oncogene*, vol. 37, no. 2, 2017, doi: 10.1038/onc.2017.313.
[268] J. Drury *et al.*, "Upregulation of CD36, a Fatty Acid Translocase, Promotes Colorectal Cancer Metastasis by Increasing MMP28 and Decreasing E-Cadherin Expression," *Cancers*, vol. 14, no. 1, 2022, doi: 10.3390/cancers14010252.
[269] P. Müller *et al.*, "Tomm34 is commonly expressed in epithelial ovarian cancer and associates with tumour type and high FIGO stage," *J. Ovarian Res.*, vol. 12, no. 1, 2019, doi: 10.1186/s13048-019-0498-0.
[270] S. Verhoog *et al.*, "Dietary Factors and Modulation of Bacteria Strains of Akkermansia muciniphila and Faecalibacterium prausnitzii: A Systematic Review," *Nutrients*, vol. 11, no. 7, 2019, doi: 10.3390/nu11071565.
[271] A. Suzuki and D. B. Knaff, "Glutamate synthase: structural, mechanistic and regulatory properties, and role in the amino acid metabolism," *Photosynth. Res.*, vol. 83, no. 2, 2005, doi: 10.1007/s11120-004-3478-0.
[272] M. Nakanishi, V. Gokhale, E. J. Meuillet, and D. W. Rosenberg, "mPGES-1 as a target for cancer suppression," *Biochimie*, vol. 92, no. 6, 2010, doi: 10.1016/j.biochi.2010.02.006.
[273] I. A. Berg, D. Kockelkorn, W. Buckel, and G. Fuchs, "A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea," *Science*, vol. 318, no. 5857, 2007, doi: 10.1126/science.1149976.
[274] M. Könneke *et al.*, "Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO_2 fixation," *Proc. Natl. Acad. Sci.*, vol. 111, no. 22, 2014, doi: 10.1073/pnas.1402028111.
[275] G. Yang *et al.*, "H_2S as a Physiologic Vasorelaxant: Hypertension in Mice with Deletion of Cystathionine γ-Lyase," *Science*, vol. 322, no. 5901, 2008, doi: 10.1126/science.1162667.
[276] Y. Yao *et al.*, "A New Biomarker of Fecal Bacteria for Non-Invasive Diagnosis of Colorectal Cancer," *Front. Cell. Infect. Microbiol.*, vol. 11, 2021, doi: 10.3389/fcimb.2021.744049.
[277] I. Romero-Canelón and P. J. Sadler, "Next-Generation Metal Anticancer Complexes: Multitargeting via Redox Modulation," *Inorg. Chem.*, vol. 52, no. 21, 2013, doi: 10.1021/ic400835n.
[278] J. J. Miettinen *et al.*, "Aminopeptidase Expression in Multiple Myeloma Associates with Disease Progression and Sensitivity to Melflufen," *Cancers*, vol. 13, no. 7, 2021, doi: 10.3390/cancers13071527.
[279] S. Vyas and P. Chang, "New PARP targets for cancer therapy," *Nat. Rev. Cancer*, vol. 14, no. 7, 2014, doi: 10.1038/nrc3748.
[280] Y. Hao *et al.*, "Oncogenic PIK3CA mutations reprogram glutamine metabolism in colorectal cancer," *Nat. Commun.*, vol. 7, no. 1, 2016, doi: 10.1038/ncomms11971.
[281] C.-Y. Huang, J. Chen, and W.-T. Kuo, "Glutamine induces remodeling of tight junctions in Caco-2 colorectal cancer cell," *Med. Oncol.*, vol. 40, no. 1, 2022, doi: 10.1007/s12032-022-01896-5.

[282] X. Liao *et al.*, “Metagenomics-based systematic analysis reveals that gut microbiota Gd-IgA1-associated enzymes may play a key role in IgA nephropathy,” *Front. Mol. Biosci.*, vol. 9, 2022, doi: 10.3389/fmolb.2022.970723.
[283] G. Kaci *et al.*, “Inhibition of the NF-κB Pathway in Human Intestinal Epithelial Cells by Commensal Streptococcus salivarius,” *Appl. Environ. Microbiol.*, vol. 77, no. 13, 2011, doi: 10.1128/aem.03021-10.

CURRICULUM VITAE

2014 – 2019	B.Sc., Computer Engineering, Erciyes University, Kayseri, TURKEY
2021 – Present	M.Sc., Electrical and Computer Engineering, Abdullah Gül University, Kayseri, TURKEY
2020 – 2021	Research Assistant, Computer Engineering, Gazi University, Ankara, TURKEY
2021 – Present	Research Assistant, Computer Engineering, Abdullah Gül University, Kayseri, TURKEY