MicroRNAs as predictive biomarkers for diagnosis and prognosis of colorectal cancer using *in silico* approaches

UNIVERSITY *of the*
WESTERN CAPE

A Thesis Submitted in Partial Fulfillment of the Requirement of the Degree of Doctor of Philosophy (PhD)

Adewale Oluwaseun Fadaka

Faculty of Natural Science, Department of Biotechnology, University of the Western Cape

Supervisor: Dr. Ashwil Klein

Co-supervisor: Dr. Ashley Pretorius

December 2019

ABSTRACT

Colorectal cancer (CRC) is referred to as cancers that arise in the colon or rectum. Rectal cancer is most often defined as cancers originating within 15 cm from the anal verge. The crude incidence of CRC in sub-Sahara African populations has been found to be 4.04/100,000 (4.38 for men and 3.69 for women). CRC stage correlates well with survival/cure rates with the majority of patients diagnosed with CRC presenting with advanced disease and a low survival/cure rate. Although most CRC-related deaths are preventable through colonoscopy screening, it is estimated that less than 50% of eligible patients are screened for CRC using the aforementioned procedure. The majority (70%) of CRCs arises sporadically, while the remaining 30% represent patients with a familial or inherited form of the disease. Mutations within tumor suppressor genes (APC, TP53), oncogenes (KRAS, BRAF, Bcl2, PI3K) and other genes, such as DNA mismatch repair (MMR) genes, accompany the stepwise transition from single crypt lesions to benign adenomatous polyps and finally development of malignant carcinomas, known as the adenoma-carcinoma sequence. Screening can detect colorectal polyps that may be removed before they become cancerous, as well as to detect cancer at an early stage when treatment is usually less extensive and more successful. Accepted screening methods include the guaiac-based Fecal Occult Blood Test (gFOBT), flexible sigmoidoscopy, stool DNA test, computed tomography (CT) colonography, double-contrast barium enema, colonoscopy and the use of biomarkers. Of these screening options, more sensitive and specific diagnostic, prognostic/survival markers are still under development. Preventive measures for CRC include maintaining healthy body weight, being physically active, minimizing consumption of red and processed meat and alcohol, and avoidance of smoking. Early diagnosis and the search for the non-invasive biomarker is currently one of the most rapidly growing areas in cancer research, and an effective way to prolong the life of patients with CRC. MicroRNAs have proven to be a fit in this context. These microRNAs function as important regulators at the post-transcriptional level of a wide range of cellular processes by modulating gene expression levels. Basically, they regulate the expression of over 30% of human genes. Present estimates suggest that nearly a third of all cellular transcripts may be regulated by the few hundred human microRNAs currently known to exist. The abnormal expression level of microRNAs has been noted as an important issue in cancer development. Therefore, microRNAs therapy is becoming an increasingly valuable tool in the management of several cancers be it diagnosis and/or prognosis. Recently, they have gained substantial attention as therapeutic targets. Nevertheless, the complexity of gene networks that a single microRNA may control and the potential adverse effects of the microRNA and/or anti-microRNA *in vivo*

deliveries remain to be further explored. Given the ever-expanding number of microRNAs, understanding their functional aspects represents a promising research area. Also, the discovery of these microRNAs predominant at different stages of CRC will allow a more comprehensive assessment and understanding of microRNA effects and provide exciting opportunities for new pathogenesis, diagnosis and treatment insights into CRC management. Despite a better understanding of the molecular mechanisms involved in colorectal carcinogenesis, little or no progress has been made in the early detection, management and prevention of CRC at each stage by microRNA with *in silico* predictions and molecular approached. This study is aimed at investigating the potential of microRNA in the early detection and staging of CRC as well as their mechanism of gene regulation using argonaute protein.

Experimentally validated and mature microRNAs implicated in CRC were extracted from various CRC microRNA databases (miRCancer, miR2disease, dbDEMC, and HMDD) with each microRNA's reference paper showing experimental evidence for their association with CRC. Total mature microRNAs sequences deposited in miRBase were also extracted and duplicates were subsequently removed with CD-HIT software program generating two datasets (the query and reference dataset respectively). These datasets (query and reference) were used in the BLAST program for sequence similarity search. CH-HIT-EST-2D and BLASTn algorithms were used for this purpose. The final list of potential candidate microRNAs was used alongside their validated microRNAs in miRDB, TargetScan, and mirDip target prediction databases to predict the genes associated with these microRNAs after duplicate removal and text mining. The two gene lists were further subjected to intersection analysis after duplicate removal and the result was saved for further target prediction analysis. To determine the involvement of these microRNA in CRC, the gene list obtained were used alongside with CRC expressed genes extracted from gbCRC and CoReCG databases for gene intersection analysis (the predicted targets and the list of genes generated by CRC databases) with the aid of the Venn diagram. The involvement of these genes in the cancer subtype was further strengthened with the DAVID database. This database was used to examine the gene functions by annotation of the microRNA target genes. KEGG was used for the pathway analysis of these genes while STRING was employed for the interactions of proteins and further visualized by Cytoscape.

To prioritize the gene list, an in silico approach was carried out with the cBioPortal database for complex genomic exploration analysis in CRC clinical data. With this approach, the genetic

alteration, distribution of mutation across protein domains, network visualization and the survival analysis for the microRNA target genes were performed.

After gene prioritization, prognostic and expression analyses were performed on both the candidate microRNAs and its target genes. The candidate microRNAs were subjected to PROGmiRV2 and SurvMicro databases for their prognostic analysis while SurvExpress and PrognoScan were used for their target genes. dbDEMC 2.0 and FIREBROWSE databases were used respectively to carry out the expression of both the microRNAs and the target genes.

Functional analysis was finally performed on only target genes that are statistically significant to infer biological functions. The secondary structure of the microRNAs was revealed by MFold, Gene ontology, co-expression analysis, were performed alongside, and finally, the promoter sequences of these genes were extracted from the eukaryotic promoter database and verified with UCSC, NCBI, and Ensembl to prevent discrepancies. The sequence manipulator suite was also used to confirm the CpG island of these sequences and their triplex-helix structures were revealed by the Trident software.

Furthermore, protein selection and preparation were carried out using PDB and Schrödinger suits. The molecular docking analysis was performed using PATCHDOCK webserver and visualized by discovery studio visualizer. The results of the study reveal that the candidate microRNAs have strong binding affinity towards their targets suggesting a crucial factor in the silencing mechanism. Furthermore, the molecular docking of the receptor to both the microRNA and microRNA-mRNA duplex were analyzed computationally to understand their interaction at the molecular level.

After duplicate removal, 125 mature microRNAs linked to CRC were obtained after combination from miRCancer, miR2disease, dbDEMC, and HMDD databases and these microRNAs were labeled query dataset. A total of 2226 mature microRNA unique sequences were also downloaded from miRBase as the reference dataset. The result of the sequence similarity search produced 26 and 43 unique microRNAs respectively for both BLASTN and CD-HIT-EST-2D. The parameters used include the expected value of 1e-3, word size of 7 and a similarity between 90-99% and threshold of 0.90 and a word size of 7 for BLASTN and CD-HIT-EST-2D respectively. After further screening by intersection analysis for unique microRNA and literature mining, six microRNAs were identified as candidates for CRC diagnosis. Since one out of the 6 candidate microRNAs did not generate any target gene only 5 candidate microRNAs alongside their validated microRNAs were used to generate a total of

82 genes as enriched in DAVID after the target prediction analysis and 2 ‘hub’ genes. Using the cBioPortal, the 82 enriched genes were prioritized on the bases of genetic alteration/ frequency of alteration in CRC to 22 genes. Prognostic and expression analysis of the candidate microRNAs confirmed that there is no link to CRC and could serve as potential novel microRNAs while the prognostic and expression for their target genes concluded that seven genes namely APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8 and GNAS at $p < 0.05$ are statistically significant and showed good prognostic values with clear implications in CRC. Finally, the biological function of these microRNAs and their target genes was confirmed by GO term, MFold, GeneMANIA, SEECancer, and Trident software.

Using *in silico* approach, this study identified 5 candidate microRNAs, two hub genes alongside seven significant target genes. The patterns of expression obtained in these genes relative to their microRNAs and considering the survival analysis result could be inferred that patients with alterations in the microRNA prioritized target genes have significantly better overall survival than patients without these alterations. These could be further exploited and could potentially serve as a resource for explicitly selecting targets for diagnosis, drug development, and management of CRC. Although validation studies are required to ascertain the biological fitness of these findings.

KEYWORDS

Colorectal cancer; Prognostics; Diagnostics; Early detection; In silico analysis; MicroRNA; Differential expression; Target prediction; Promoter analysis; Biomarkers; Carcinogenesis; Bioinformatics; Database; Genetic alteration; BLAST; CD-HIT; BLAT; DAVID; KEGG; STRING, Cytoscape, MFold, GO term; molecular docking; Argonaute protein.

DECLARATION

I, Adewale Oluwaseun Fadaka declare that this thesis is my research work and has never been submitted for any evaluation, degree or examination in any other Institution. Also, all the sources used or quoted (books, published research and review articles, manuscripts, web pages, reports, online data, government documents, computer programs, dataset/databases, and electronic books/articles) have been indicated and duly acknowledged by means of complete references.

Adewale Oluwaseun Fadaka **December 2019**

Signed: 

ACKNOWLEDGMENTS

"Those who are truly grateful are deeply moved by the privilege of living."

First and foremost, I would like to thank **God Almighty** for giving me the strength, knowledge, ability and opportunity to undertake this research study and to persevere and complete it satisfactorily. Without his blessings, this achievement would be impossible.

Furthermore, I would like to express my sincere gratitude to my Supervisors **Dr. Ashley Pretorius and Dr. Ashwil Klein** for their continuous support of my Ph.D. study and related research experiences, for their patience, motivation, and immense knowledge. Their guidance helped me during the period of this research and writing of this thesis. I could not have imagined having a better advisor and mentor for my study.

Besides my PIs, I would like to thank the **Bioinformatics Research Group** (and specifically Nasr Eshibona and Chipampe) for their insightful comments, encouragement, positive contributions, as well as criticisms. Also for the constructive questions during my presentations which helped me to widen my research from various perspectives.

I thank my fellow lab mates in **Biotechnology** Department (Life science) for their stimulating discussions, for their relentless support when working together before deadlines, and for all the fun we had during the research years. Also, I thank my colleagues (Senior and mate) at Afe Babalola University for their constant check-ups. In particular, I am grateful to **Dr. Adewale** and his family for enlightening me in diverse areas of life and their constant prayers.

My acknowledgment would be incomplete without thanking the biggest source of my strength. My parents, my four sisters, the Adelegans, the Falades, and the friends turn family for supporting me spiritually throughout the writing of this thesis.

To **Dr. Akinsoji Taiwo,** a lot to say to you but conclusively, thanks for showing a profit whenever I count all my assets. Thank you for the undying love you have shown.

More so, **Prof M.O Ajisafe** is highly appreciated. He gave me all the freedom to pursue my research, while silently and non-obtrusively ensuring that I stay on course and do not deviate from the core of my research. He taught me how to own a table and bring enough to occupy the table. Without his able guidance, prayers and financial support, this thesis would not have been possible. I am eternally grateful to him for everything to say the least.

Dr. Bakare is a great friend and a selfless being. He made living in South Africa worthwhile. Thanks a lot.

A big thank you to **Professor Olubunmi Magbagbeola** for her support, encouragement and credible ideas have been great contributors to the completion of the thesis. I see you beyond my Lecture and my Master's Supervisor. She is a mentor and a mother. Her ways of life have shaped my world and have made me see life in a different dimension. Also, I will not forget the parting gift when I was leaving for South Africa for this program. Your advice and chats will ever be greener in my heart. I love you and appreciate you.

I also take pride in acknowledging the insightful guidance of **Dr. Ayorinde James**, for sparing his valuable time whenever I approach him. Thanks for showing me the way ahead. He did not only push me to realize my potentials, but he made me think out of the box and most times like there is no box. Also, thanks to his beautiful wife **(Dr. Olorunwa James (MBBS))** for the care and support. I cannot thank you enough.

I would like to intensely appreciate **Dr. Gbemisola Saibu** whose dreams for me to undertake this degree in South Africa have resulted in massive achievement and without her assistance and nurturing, I would not have own a Ph.D. Biotechnology (UWC) today (Maybe Ph.D. Biochemistry in UNILAG "laughs"). It is true that if God ever existed, he would be in the form of a mother because only a mother can love, assist, and give without expecting anything in return.

I sincerely appreciate the effort of **Mr. Musharraf Foorooskar** for his assistance with respect to software applications. Thanks a million.

I also thank the anonymous references used to achieve the overall success of the project. Your contributions are massively significant. 'In the end, we will remember not the words of our enemies, but the silence of our friends'. Therefore, to everybody who I did not mention/spell out their names but made a significant contribution to my progress and to the success of this thesis, just know that silence is better than unmeaning words and it is the only friend who never betrays. Thanks a lot. I double sight you all.

DEDICATION

To God Almighty and my loving parents Mr. and Mrs. Fadaka. The most overwhelming key to my success is the positive involvement of these people in my life.

LIST OF PUBLICATIONS

Published Journal Papers

1. Fadaka, A. O., Ojo, B. A., Adewale, O. B., Esho, T., and Pretorius, A. (2018). Effect of dietary components on miRNA and colorectal carcinogenesis. *Cancer cell international, 18*(1), 130.

2. Fadaka, A. O., Pretorius, A., and Klein, A. (2019). Biomarkers for Stratification in Colorectal Cancer: MicroRNAs. *Cancer Control, 26*(1), 1073274819862784.

3. Fadaka, A. O., Pretorius, A., and Klein, A. (2019). In silico identification of microRNAs as candidate colorectal cancer biomarkers. *Tumor Biology*, 1-15.

4. Fadaka, A. O., Pretorius, A., and Klein, A. (2019). Functional Prediction of Candidate MicroRNAs for CRC Management Using in Silico Approach. *International Journal of Molecular Science*. *2019, 20, 5190*

5. Fadaka, A. O., Pretorius, A., and Klein, A. (2019). MicroRNA Assisted Gene Regulation in Colorectal Cancer. *International Journal of Molecular Science*. *20*(19), *4899.*

LIST OF CONFERENCE PAPERS

1. Adewale Oluwaseun Fadaka, Ashley Pretorius, Ashwil Klein. Identification of MicroRNAs as Biomarkers in Colorectal Cancer Diagnostics: An *in silico* approach. In: Proceedings of the Experimental Biology, 2020 (ASBMB). San Diego California, USA. (Accepted)

2. Adewale O. Fadaka, Ashley Pretorius, Ashwil Klein. MicroRNA assisted gene regulation in colorectal cancer. Keystone Symposia. Noncoding RNAs: Mechanism, Function, and Therapies (A1). Fairmont Chateau Whistler - Whistler, British Columbia, Canada. (Accepted)

LIST OF ABBREVIATIONS

3’-UTR 3’ untranslated region; 5’-UTR 5’ untranslated region; BRCA1 Breast Cancer Type 1 Susceptibility Protein; CoReCG Colorectal Cancer Gene Database; CT Computed tomography; CD-HIT Cluster Database at High Identity with Tolerance; DAVID Database for Annotation, Visualization, and Integrated Discovery; GbCRC Gene Browser for Colorectal Cancer; GO Gene Ontology; OS Overall survival; RISC RNA-induced silencing complex; SNP Single nucleotide polymorphism; STRING Search Tool for the Interacting Gene/Proteins; TF Transcription Factor; TGFβ Transcription Growth Factor beta; TNM Tumour Node Metastasis; Tp53 Tumour Protein P53; MMR: Mismatch repair; LOH: Loss of heterozygosity; EMT: Epithelial-mesenchymal transition; TSA: Traditional serrated adenoma; SSA/P: Sessile serrated adenoma/polyps; MSI-H: Microsatellite instability-high; HNPCC: Hereditary non-polyposis colorectal cancer; FAP: Familial adenomatous polyposis; MSI: Microsatellite instability; CIN: Chromosomal instability; DGCR8: DiGeorge syndrome Critical Region 8 protein CRC: Colorectal cancer; microRNAs: microRNAs; RISC: RNA-induced silencing complex; UTR: The 3′untranslated region; DGCR-8: DiGeorge syndrome Critical Region 8 protein; AGO-2: Argonaute protein; ERβ: estrogen receptor beta; SCFAs: Short-chain fatty acids; CASP3: cysteine-aspartic acid protease 3; APC: adenomatous polyposis coli; MMPs: matrix metalloproteinases; DCC: deleted in colorectal carcinoma; EGFR: epidermal growth factor receptor; ICAM: intercellular adhesive molecules; PDCD4: programmed cell death 4; PTEN: phosphatase and tensin homolog; CDK4,6: cyclin-dependent kinase 4,6; ECM: extracellular matrix; EMT: epithelial-to-mesenchymal transition; RECK: reversion-inducing cysteine-rich protein with kazal motifs; TIMP3: tissue inhibitor of metalloproteinase 3; uPAR: plasminogen activator, urokinase. receptor; TGFβRI/II: transforming growth factor βreceptor I/II; ZEB1/2: zinc-finger E-box binding homeobox-1; CTGF: connective tissue growth factor; TSP1: thrombospondin-1; AGO: Argonaute protein.

TABLE OF CONTENTS

CHAPTER 1.0

INTRODUCTION

1.1 Background to Research Project

Cancer was reported to be the leading cause of death accounting for about 13% of all mortality globally (Ferlay *et al.*, 2010). The occurrence of cancer is greater in America and Europe when compared to low and middle-income countries but they still suffer an appreciable burden of this disease. Almost 24 million people have been predicted to have cancer by 2050 and about 70% will be located in the developing countries (Saluja *et al.*, 2014). Diagnoses of new cancer cases are expected to exceed 20 million per year by 2030 as the population demographics change with two-thirds occurring in the developing countries (Alwan, 2011). Colorectal cancer is referred to as cancers that arise in the colon (large intestine) or rectum. Rectal cancer is most often defined as cancers originating within 15 cm from the anal verge (F.-y. Li *et al.*, 2009).

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in males and the second in females, and the fourth greatest cause of cancer-related deaths worldwide (Xu *et al.*, 2016). Carcinogenic processes involve the stepwise accumulation of mutations and/or epigenetic alterations, leading to the transformation of normal colonic epithelia. this may develop and progress over a period of 10 to 20 years (Winawer *et al.*, 2002). The lifetime incidence of CRC in an average-risk population is 5% and progressively increases after age 50. Genetic and environmental factors play a role in colon cancer formation (Rao *et al.*, 2013). Cronjé *et al.* (2009) reported the troubling trend for a disproportionately large number of patients present with CRC in South-Africa. However, the age at which these patients present with CRC may be a marker for the involvement of hereditary factors that often have specific pathological features. Over the past decade, the overall incidence of CRC in South-Africa has increased markedly. CRC has recorded the 10th most common cancer diagnosed in males and females in South Africa in 1989 but was further ranked among the foremost five cancers (5th among males and 3rd among females) (Mqoqi *et al.*, 2004).

The epidemiology of CRC in white South Africans appears to follow the classic Western trend, although the molecular pathology has not been comprehensively investigated. CRC among black South Africans is far less common, but evidence showed the marked increase in some centers (Angelo *et al.*, 2001). In Africa, this disease was considered to be rare but this is no longer the case (Adesanya *et al.*, 2000). The crude incidence of CRC in sub-Saharan Africa has been found to be 4.04/100,000 population (4.38 for men and 3.69 for women) (Graham *et al.*,

2012). The cancer stage correlates well with survival/cure rates as outlined by Fredericks *et al.* (2015). Unfortunately, the majority of patients diagnosed with CRC will develop metastatic disease. Although most CRC-related deaths are preventable through screening colonoscopy, it is estimated that less than 50% of eligible patients are screened for CRC (Baron *et al.*, 2013). The major challenges in effective and appropriate CRC treatment are the late presentation of the patients, significant involvement of younger patients, aggressive tumor type, lack of tailored/targeted therapy, intra-hospital obstacles, and the patients' aversion to unfavorable surgical treatment (Irabor *et al.*, 2014). Mutations in tumor suppressor genes (APC, TP53), oncogenes (KRAS, BRAF, Bcl2, PI3K) and other genes, such as DNA mismatch repair (MMR) genes, accompany the stepwise transition from single crypt lesions to benign adenomatous polyps and finally development of malignant carcinomas (Migliore *et al.*, 2011), known as the adenoma-carcinoma sequence (Fearon *et al.*, 1990; Muto *et al.*, 1975; Vogelstein *et al.*, 1988). The majority (70%) of CRCs arises sporadically, while the remaining 30% represent patients with a familial or inherited form of the disease (Grady, 2003). Accepted screening methods include the guaiac-based fecal occult blood test (FOBT), flexible sigmoidoscopy, stool DNA test, computed tomography (CT) colonography, double-contrast barium enema, and colonoscopy. Of these screening options, prognostic survival markers of patients are still under development. Preventive measures for CRC include maintaining a healthy body weight, being physically active, minimizing consumption of red and processed meat and alcohol, and avoidance of smoking (Botteri *et al.*, 2008). Screening can detect colorectal polyps that can be removed before they become cancerous, as well as to detect cancer at an early stage when treatment is usually less extensive and more successful. Early diagnosis and the search for the non-invasive biomarker is currently one of the most rapidly growing areas in cancer research, and an effective way to prolong the life of patients with CRC. Interestingly, novel non-invasive markers with potential clinical value are discovered to detect early CRC and then improve the prognosis for CRC patients. It has become increasingly clearer over the past decade that a large class of small non-coding RNAs, known as microRNAs (microRNAs), function as important regulators at the post-transcriptional level of a wide range of cellular processes by modulating gene expression levels (Catalanotto *et al.*, 2016). Current estimates suggest that nearly a third of all cellular transcripts may be regulated by the few hundred existing human microRNAs (D. P. Bartel, 2004). microRNAs are short endogenous mediators of about 18–22 base pairs nucleotide, non-coding RNAs which play key roles in biological processes involved in an organism's development, cell specialization and homeostasis (Diederichs *et al.*, 2016; Tomaru *et al.*, 2006). The abnormal expression level of microRNAs is realized as an important issue in

cancer development. Therefore, microRNA therapy is becoming a bright target (Osaki *et al.*, 2015). Many microRNAs exhibit tissue-specific patterns of expression and are deregulated in various cancers, where they can either be oncogenic (oncomirs) or tumor-suppressive. These small molecules regulate gene expression through binding to the target mRNA, which influences mRNA stability or suppress translation (Fabian *et al.*, 2010). Basically, microRNAs are predicted to regulate or influence 30%-80% of human genes (Lu *et al.*, 2012). Despite a better understanding of the molecular mechanisms involved in colorectal carcinogenesis, little or no progress has been made in the early detection, management and prevention of CRC at each stage using microRNA in South-Africa and Africa as a whole.

1.2 Overview of microRNAs

MicroRNAs are a class of approximately 22 oligonucleotides, evolutionarily conserved non-coding RNA molecules naturally occurring in the genomes of plants and animals. They regulate post-transcriptional protein expression, typically by binding to the 3' untranslated region (3'-UTR) of the complementary mRNA sequence, resulting in translational repression and gene silencing (Cannell *et al.*, 2008). Studies have shown that thousands of human protein-coding genes are regulated by microRNAs, indicating that microRNAs are "chief regulators" of many important biological processes. Their roles have been demonstrated in both normal and pathological cellular processes. Also, due to their ability to target multiple genes, they can regulate the expression of several proteins. Studies demonstrated that these noncoding RNAs can act on several key cellular processes, including cell differentiation, cell cycle progression, and apoptosis. In tumors, some microRNAs function as oncogenes, others as tumor suppressors; upregulation of oncogenic miRNAs (oncomiRs) was demonstrated in cancer cells (Tanase *et al.*, 2011).

MicroRNA was initially discovered in Caenorhabditis elegans by Victor Ambros' laboratory in 1993 while studying the gene lin-14. At the same time, Gary Ravkun identified the first microRNA target gene. These two discoveries identified a novel mechanism of posttranscriptional gene regulation (Lee *et al.*, 1993).

However, they were not recognized as a distinct class of biological regulators until the early 2000s (Reinhart *et al.*, 2000). After then, an increasing number of microRNAs have been recognized in mammals. In humans, over 700 microRNAs have been identified and fully sequenced, and the estimated number of microRNA genes in a human genome is over one

thousand. Based on *in silico* models, microRNAs in humans have a direct effect on at least 30% of the genes in the whole genome. Research revealed their different expression pattern in different tissues (Pace *et al.*, 2000; Wienholds *et al.*, 2005) and multiple roles for microRNAs in plant and animal development and in many other biological processes (J.-R. M. B. D. Bartel, 1953; Harfe *et al.*, 2005; Poy *et al.*, 2004; Wilfred *et al.*, 2007; Wu *et al.*, 1998). Aberrant expressions of microRNAs are implicated in disease states. MicroRNA-based therapies are currently under investigation (Fasanaro *et al.*, 2010; Lally *et al.*, 2013; Thakur *et al.*, 2014; Trang *et al.*, 2008). Identification of more microRNAs and proper understanding of their mechanism of tissue expression regulation represent a key issue in cancer research.

The biogenesis of microRNAs is a multi-steps and requires specific cellular machinery (Hastings *et al.*, 2001). microRNAs are encoded as short inverted repeats composed of a double-stranded RNA stem-loop of about 70 bp long and are found in both introns and intergenic clusters in the genome (Ardekani *et al.*, 2010). RNA polymerase II is responsible for the synthesis of the introns and exons of both protein-coding and non-coding transcripts from where miRNAs are derived(Morey *et al.*, 2004). In the nucleus, microRNAs are transcribed as primary pri-miRNA transcripts and then are processed to form the precursor pre-miRNA stem-loop structure before transportation into the cytoplasm where they are cleaved by the Dicer RNAase III endonuclease and produce mature miRNA (21–23 nucleotides) (Bilen *et al.*, 2006).

Based on function, microRNAs have been demonstrated to play a crucial role in a wide range of developmental processes such as metabolism, cell proliferation, apoptosis, developmental timing, and neuronal cell fate (Berezikov *et al.*, 2005; Croce *et al.*, 2005; Mattick *et al.*, 2005). Other regulatory roles include neuronal gene expression (Klein *et al.*, 2005), brain morphogenesis (Giraldez *et al.*, 2005), muscle differentiation (Naguibneva *et al.*, 2006), and stem cell division (Hatfield *et al.*, 2005). The role of microRNAs as an important source in carcinogenesis is still very much unappreciated. But altered patterns of microRNAs in cells have been shown to be responsible for changes that cause mutation.

1.3 Research Aims

The main aim of this research was to investigate the potential of candidate microRNA for early detection and staging of CRC as well as their mechanism in gene expression. This involved an *in silico* and molecular docking to identify microRNAs and target genes, including their general

functions, triplex binding formation and mode of gene expression in CRC. In summary, the primary aims of this research project were:

1. Identify candidate microRNAs, which can be used as diagnostic and prognostic biomarkers for CRC, using an *in silico* approach.
2. Identify the target genes for these candidate microRNAs and implicate them in the onset and progression of CRC using an *in silico* approach.
3. Prioritize the target genes implicated in CRC, through the pathway and co-expression analysis and relate these genes back to their regulating microRNAs thus creating a shortlist of microRNAs and their target genes for further analysis.
4. Perform prognostic and predictive analysis of the prioritized microRNAs and their target genes using various databases as well as *in silico* expression analysis of these microRNAs and their target genes.
5. To investigate the underlying mechanism(s) by which the candidate microRNAs interact with the prioritized microRNA target genes promoter regions for gene transcriptional modification and to distinctly infer function(s).
6. To determine the possible mode of gene regulation in CRC using argonaute protein.

1.4 Statement of Problem

Early CRC detection strategies are faced with the major challenge of developing a standardized biochemical diagnostic and prognostic approach that is non-invasive, more sensitive and specific for CRC stages. To date, the most effective treatment option for CRC remains surgery, which is costly and has associated side effects, such as damage to nearby organs, adhesion, among others. Chemotherapeutic agents that often follow surgery lack tissue selectivity. Even the blood test, carcinoembryonic antigen exhibits low sensitivity and specificity, especially in the context of early disease. CRC has been such a serious health problem because it is largely asymptomatic until the latter stages often times when cancer has already metastasized (Kawamura *et al.*, 2014). If CRC is detected early, it is a largely treatable disease that can benefit from curative surgery. In most African countries including South Africa, the major challenge for CRC treatment is early detection. Current treatment options are often too late, typically after metastasis has occurred. The most reliable detection method is colonoscopy, but it is a specialized and expensive procedure, which is not readily available, and presents a small significant risk for perforations. MicroRNAs (microRNAs) are small regulatory RNAs that are

involved in the maintenance of cell homeostasis but are often altered in tumor cells. Therefore, the identification of these microRNAs that are expressed and predict each stage of CRC would be a focal point in this research.

1.5 Rationale of the Research

The involvement of non-coding RNAs in carcinogenesis and tumor progression has been confirmed by numerous researchers in the past decades (Seton-Rogers, 2013). In general, non-coding RNAs especially microRNAs are attracting considerable interest and there is increasing evidence that the expression of these microRNAs plays an important role in cancer development and progression. Recently, microRNAs have gained substantial attention as therapeutic targets. Nevertheless, the complexity of gene networks that a single microRNA may control and the potential adverse effects of the microRNA and/or anti-microRNA *in vivo* deliveries remained underexplored. In summary, given the ever-expanding number of microRNAs, understanding their functional aspects remains a promising mission for exploration. Large sample data analysis, the discovery of microRNAs predominant at different stages of CRC will allow for more comprehensive assessments and understanding of its effects will provide exciting opportunities for new pathogenetic and treatment insights into CRC management.

1.6 Significance of the Study

Considering CRC as one of the public health problems with increase mortality rate coupled with low overall survival due to late detection, the findings of this study will rebound to benefit the society and cancer research community as stated below.

1. To understand the roles of microRNAs in carcinogenesis. Given the shortcomings of chemotherapeutics, this study further proposes certain dietary components with cancer-preventive capabilities, including therapeutic and chemopreventive properties.
2. CRC remains the third leading cause of cancer-related mortality in the world with the major challenge of early detection. Current treatment options also present various side effects. This study would set the tone and used as a baseline for the use of microRNAs as diagnostic tools in the early detection and diagnosis of CRC.

3. This study would identify five candidate microRNAs and their target genes that can potentially be used in the detection and diagnosis of CRC. If CRC is detected earlier, it is a largely treatable disease.
4. The identification and functional inference of microRNAs would aid their specific usage in CRC. This study will further attach functions to these candidate microRNAs through their targets as well as their expression pattern with clinical prognosis.
5. The mechanism by which these microRNAs regulates their targets would also be discovered through the RNA induced silencing complex (RISC).

Contributions to knowledge

Several treatment options are being used in the detection of CRC such as colposcopy, and biomarkers (CEA). Disease detection is often too late, typically after metastasis and these methods are expensive and require a specialized procedure, which is not readily available, invasive, and presents a small significant risk for perforations. MicroRNAs have the ability to override these disadvantages presented by other methods/procedures of detection. These non-coding RNAs can be found in all parts of the human body including the blood, saliva, as well as tissue samples. The discovery of novel non-invasive markers with significant clinical value may potentially be used to detect early CRC and improve the overall management of patients with CRC. Therefore, the study potentially identified five candidate microRNAs that are differentially expressed (using the tumor node metastasis (TNM)) and are specific in the early detection and diagnosis of CRC. Also, the biological roles together with the mechanism of actions may also be a lead for therapeutic target and CRC management.

1.7 Research Methodology

The implementation of a suitable methodology that adequately encompasses the specific research aim and objectives set out for this project was obligatory. The process of achieving the research objectives is based on both *in silico* approach and molecular docking/interaction.

For objective 1: Identify candidate microRNAs, which can be used as diagnostic and prognostic biomarkers for CRC, using an *in silico* approach.

Total microRNAs for CRC were retrieved from available microRNA databases (Griffiths-Jones, 2004; Jones, 2004) (Cui *et al.*, 2014; Xie *et al.*, 2013). Thereafter, curation was

performed to verify that all the retrieved microRNAs are experimentally validated (dataset 1). The total human microRNAs available to date were also downloaded from the microRNA database (mirBase) (http://www.mirbase.org/, released 20) as the reference sequence (Dataset 2). Local similarity tools for microRNA identification was further performed on both datasets using in house" created scripts (Section 1.8: Data mining technique). Finally, a list of prioritized microRNA was generated by duplication removal and text mining.

For objective 2: Identify the target genes for these candidate microRNAs and implicate them in the onset and progression of CRC using an *in silico* approach.

To infer functions to microRNAs, their gene target functions were elucidated. Target prediction analysis was performed for each of the microRNAs prioritized. The predicted genes were then checked for specificity in CRC. The gene list was finally prioritized using an "in house" set criteria. In other to understand the functions of the microRNAs, their gene functions were further examined.

For objective 3: Prioritize the target genes implicated in CRC, through the pathway and co-expression analysis and relate these genes back to their regulating microRNAs thus creating a shortlist of microRNAs and their target genes for further analysis.
In order to establish any possible roles of the target genes in the CRC pathway, first, a co-expression analysis was performed on the microRNA target gene list using the STRING, KEGG, DAVID and cBioPortal (prioritization).

For objective 4: Perform prognostic and predictive analysis of the prioritized microRNAs and their target genes using various databases as well as *in silico* expression analysis of these microRNAs and their target genes.
Prognosis analysis of biomarkers is used to detect or confirm the presence of a disease or condition of interest or to identify individuals with a subtype of the disease while predictive analysis is used to identify the likelihood of a clinical event, disease recurrence or progression in patients who have the disease or medical condition of interest. In light of this, available *in silico* tools were employed to perform the predictive as well as the prognostic analysis of both the prioritized microRNAs and their target genes. Kaplan-Meier plotter, SurvExpress, and ProgGene were used for the gene list while SurvmicroRNA, ProgmiRVs and miRPower were

used for the corresponding regulating microRNA/s for the prognostic and predictive analysis. Firebrowser was used to carry out *in silico* expression analysis of the prioritized microRNAs and their target genes.

For objective 5: To investigate the underlying mechanism(s) by which the candidate microRNAs interact with the prioritized microRNA target genes promoter regions for gene transcriptional modification and to distinctly infer function(s).

The secondary structure and the thermodynamic energies of the microRNAs were revealed by the Mfold algorithm. The triplex binding ability of the oligonucleotide with the target promoters were analyzed by Trident. Finally, evolutionary stage-specific somatic events and co-expression analysis of the target genes in CRC were analyzed by SEECancer and GeneMANIA plugin in Cytoscape.

For objective 6: To determine the possible mode of gene regulation in CRC using argonaute protein.

Protein selection (Argonaute protein) and preparation were carried out using PDB, PROCHECK, and Schrödinger suits. The molecular docking analysis was performed using PATCHDOCK webserver and visualized by discovery studio visualizer.

1.8 Data Mining Techniques

BLAST for similarity searches

The BLAST program for nucleotide-nucleotide blast was used to return the most similar microRNA sequences from the microRNA database specified. The series of step followed on this program were (1) finding the appropriate blast program used (Blastn), (2) entry of query sequences, (3) selection of database to search, (4) running of the BLAST search, and (5) interpreting the E-values. This option was used to first build a dataset for all the Homo sapiens mature microRNAs extracted from miRBase after removal of duplication to use as the reference set in this study using the command line below to make BLAST database:

```
c: programfile NCBI blast-2.7.1 bin> makeblastdb.exe –in c: blast
allmicroRNAs.text.txt –parse_seqids –dbtype nucl –out c: blast result.out
```

Where -in is the input file, -Parse_seqids is sequence id parsing that is; to resolve into its component parts and describe their syntactic roles. –dbtype nucl is specifying the input, with the molecule type (nucleotide).

Validated microRNAs for CRC were then scanned after creating the dataset containing all the mature microRNAs to identify the number of mismatches between the query dataset and the reference sequences. The number of mismatches allowed was three sequences with less than three mismatches having more than 90% similarity. The command line used to run the program was:

```
c: programfile NCBI blast-2.7.1 bin> blastn.exe –db c: blast\result.out –evalue 1e-3 –word_size 7 –query CRCmicroRNA.text.txt –out c:\blast\result.out
```

The e-value is a parameter that describes the numbers of hits one can expect to see by chance when searching a database of a known size specified as 1e-3, -db which is the database generated by previous command line, -word_size is the similarity region, -query is the query file, and -out is the result file name.

Although CD-HIT-EST-2D is limited to certain clustering thresholds, it compares two nucleotide datasets (db1 and db2) for similarity check. Db1 contained the non-redundant mature microRNA sequences, which are associated with CRC, which were used as the query dataset and db2 contained all mature microRNAs, which were used as the reference dataset. The input was in FASTA format datasets (db1, db2) and the output was two files (a FASTA file of sequences in db2 that is not similar to db1 and a text file that lists similar sequences between db1 and db2. The text file showed the similarity in clusters generated and percentage of similarity between sequences. This was done with the command line below:

```
cd-hit-est-2d -i CRCmicroRNA.text –i2 allmicroRNAs.text –o result –c 0.90 –n 5
```

Where - i is db1 which includes the validated CRC microRNAs, - i2 is db2 which include all microRNAs, -O is the output, -c sequence identity threshold and -n 5 is the word-length.

MiRDB for target prediction

This database was accessed at http://www.mirdb.org/index.html for the identification of microRNA target genes for each of the validated and candidate microRNAs. With the default settings, each of the candidate microRNAs and the validated microRNAs was submitted differently for gene prediction. The two gene lists were subjected to intersection analysis after duplicate removal and the result was saved in an excel file for further target prediction analysis.

TargetScanHuman for prediction of microRNA targets

This tool was accessed at http://www.targetscan.org/vert_72/ (version 7.2) for the prediction of both the candidate and the validated microRNAs. This generated two lists of genes for the candidate and validated microRNAs. After the removal of duplicates, the two lists were

subjected to duplicate removal and the unique gene list created after intersection analysis was saved in an excel file for further analysis.

MirDIP microRNAs data integration portal

This prediction tool was accessed at http://ophid.utoronto.ca/mirDIP/index.jsp for the identification of target microRNAs in this study. This prediction tool can be used to Find overlapping interactions among selected microRNAs and genes in Homo sapiens in its bidirectional search or to find microRNAs that target a gene, or genes targeted by a microRNA in Homo sapiens (Unidirectional search). This unidirectional search of this tool was used for both the candidate and the validated microRNAs in other to follow a uniform process as followed by previous prediction tools. The two gene lists generated at high score class was subjected to duplicate removal and gene intersection analysis. The unique gene list was then saved in an excel file for further analysis.

Extraction of CRC expressed genes

Gene browser for colorectal cancer (gbCRC) and CRC for the gene database (CoReCG) were used for the extraction of the expressed genes associated with CRC. The gbCRC was accessed at http://gbcrc.bioinfo-minzhao.org/download.cgi. The human CRC genes with high confidence (2+ data sources and 5+ PubMed abstracts) were downloaded in a flat file and saved. The genes available at CoReCG was accessed at http://lms.snu.edu.in/corecg/gene_browse.php from the browse icon and gene information. Two gene lists were generated and saved.

Gene intersection analysis

The Venn diagram software was used to determine the intersection genes between the list of genes obtained from the candidate microRNAs and the validated microRNAs. This tool generated a pictorial Venn diagram of the intersect between the list input and a textual output indicating which genes were in each intersection or were unique to a certain list. This tool was also used to generate intersections between the predicted targets and the list of genes generated by CRC databases used in this study.

DAVID database

Database for annotation, visualization and integrated discovery (DAVID) was used to examine the gene functions by annotation of the microRNA target genes. This database was accessed at

https://david.ncifcrf.gov/summary.jsp. From the DAVID homepage in the list section, four steps were followed. The steps include: (1) Enter gene list, (2) select identifier, (3) list type, and (4) submit a list. Each of the five gene lists was submitted individually, the official gene symbol was selected as the identifier followed by selecting the gene list and submitted for result output. Homo sapiens was selected as the background for the search. The database displays the annotation summary results with several options. For the link gene-disease associations, the GAD_DISEASE was selected. The chart output places the genes in various cancer processes with special emphasis on CRC for this study. The analysis generated lists of genes involved in CRC and was saved in an Excel file.

Gene visualization using KEGG Pathway analysis

Under the several options available for the annotation summary results in DAVID, the KEGG_pathway was selected from the pathway section after the submitted gene list associated with genetic disease was submitted to DAVID. KEGG pathway analyses were carried out in DAVID, Bioinformatics Resources at http://david.abcc.ncifcrf.gov. Results were given as pathway graphs for the genes that are involved in known CRC processes.

Analysis of Gene/protein interaction network

Gene IDs for a list of genes identified by DAVID implicated in CRC were used as input for the generation of gene networks using the STRING database Version 10.5. First, the genes were used as input to generate an expression network among each other. The basic steps employed in this study for STRING interaction are as follows: List of differentially expressed genes from DAVID (82) → Paste in STRING (Multiple proteins) → Specific organism (Homo sapiens) → Go → Continue→ actual result page. To produce this expression network, parameters were chosen as follows; (i) a confidence level of 0.9 and (ii) a network depth of 5. Secondly, an extended network was produced to determine which of the genes showed links to known genes involved in cancer and specifically in CRC. The results were given as a ‘stick/line and ball’ graphics with the ‘balls’ (proteins/genes) connected by different color lines with each color indicating the type of evidence for the connection between each protein/gene.

Cytoscape for visualization

STRING network was downloaded as a simple tabular text output and was edited and uploaded into Cytoscape as a setup installation file (SIF file). The SIF included four columns with the interaction name in between the protein symbol columns and the intersection weight in the last

column. The following steps were used to input the file into Cytoscape: From file→ import→ network→ file (choose the appropriate file) → open→ ok). The visual display was then adjusted (nodes, edges, and network) using appropriate layout and plugins to achieve a less dense or clustered image. A complication with Cytoscape software is that it does not perform a complete workflow starting from the protein or gene list up to protein interaction mapping, network analysis, and visualization. It does not perform the first step, the interaction mapping. Therefore, a protein list was uploaded into STRING website first to map out the interaction between the protein/gene list. This interaction list was then uploaded into Cytoscape as a SIF file.

1.9 Linkage of scientific papers

Cancer is the second leading cause of death worldwide and is responsible for estimated deaths of 9.6 million in 2018. Globally, about 1 in 6 deaths is due to cancer. As a subtype of this disease, colorectal cancer (CRC) is the third most common cancer after lung and breast cancer.

The process of colorectal carcinogenesis, which involves the stepwise accumulation of mutations and/or epigenetic alterations, leading to the transformation of normal colonic epithelia may develop and progress over a period of time. This disease subtype is the third most commonly diagnosed cancer in males and the second in females, and the fourth greatest cause of cancer-related deaths worldwide. The lifetime incidence of CRC in an average-risk population is 5% and progressively increases after age 50.

CRC stage correlates well with survival/cure rates with the majority of patients diagnosed with CRC presenting with advanced disease and a low survival/cure rate. Although most CRC-related deaths are preventable through colonoscopy screening, it is estimated that less than 50% of eligible patients are screened for CRC using the aforementioned procedure. The majority (70%) of CRCs arises sporadically, while the remaining 30% represent patients with a familial or inherited form of the disease. Mutations within tumor suppressor genes (APC, TP53), oncogenes (KRAS, BRAF, Bcl2, PI3K) and other genes, such as DNA mismatch repair (MMR) genes, accompany the stepwise transition from single crypt lesions to benign adenomatous polyps and finally development of malignant carcinomas, known as the adenoma-carcinoma sequence. Screening can detect colorectal polyps that may be removed before they become cancerous, as well as to detect cancer at an early stage when treatment is usually less extensive and more successful.

Accepted screening methods include the guaiac-based Fecal Occult Blood Test (gFOBT), flexible sigmoidoscopy, stool DNA test, computed tomography (CT) colonography, double-contrast barium enema, colonoscopy and the use of biomarkers. Of these screening options, more sensitive and specific diagnostic, prognostic/survival markers are still under development. Preventive measures for CRC include maintaining healthy body weight, being physically active, minimizing consumption of red and processed meat and alcohol, and avoidance of smoking.

Early diagnosis and the search for non-invasive biomarker is currently one of the most rapidly growing areas in cancer research, and an effective way to prolong the life of patients with CRC. The discovery of novel non-invasive markers with significant clinical value may potentially be used to detect early CRC and improve the overall management of CRC patients.

It has become increasingly clear over the past decade that a large class of small non-coding RNAs, known as microRNAs, function as important regulators at the post-transcriptional level of a wide range of cellular processes by modulating gene expression levels. Present estimates suggest that nearly one-third of all cellular transcripts may be regulated by the few hundred human microRNAs currently known to exist. Many microRNAs exhibit tissue-specific patterns of expression and are deregulated in various cancers, where they can be either oncogenic (oncomirs) or tumor-suppressive. Therefore, microRNAs therapy is becoming an increasingly valuable tool in the management of several cancers.

Despite a better understanding of the molecular mechanisms involved in colorectal carcinogenesis, little or no progress has been made in the early detection, management and prevention of CRC at each stage using microRNA. Given the ever-expanding number of microRNAs, understanding their functional aspects represents a promising research area. Large sample data analysis, the discovery of microRNAs predominant at different stages of CRC will allow a more comprehensive assessment and understanding of microRNA effects and provide exciting opportunities for new pathogenesis, diagnosis and treatment insights into CRC management.

Early CRC detection strategies are faced with the major challenge of developing a standardized biochemical diagnostic approach that is non-invasive, more sensitive and specific for CRC stages. The most effective treatment option for CRC up to date remains surgery, which is costly and has associated side effects, such as damage to nearby organs, adhesion, amongst others. In addition, chemotherapeutic agents that often follow surgery, lack tissue selectivity.

Furthermore, blood test which employs carcinoembryonic antigen (CEA) exhibits low sensitivity and specificity, especially in the context of early disease. CRC has been such a serious health problem since it is largely asymptomatic up until the latter stages of the disease often times when the cancer has already metastasized and untreatable. If CRC is detected early, it is a treatable disease that can benefit from curative surgery. Current treatment options are often too late, typically after metastasis has occurred at which point prognosis is poor. The most reliable detection method is colonoscopy, which is a specialized and expensive procedure and not readily available, presenting a marginal risk for perforations of the colon, an associated side effect of this procedure. The lack of early diagnostic biomarkers and the use of invasive approaches for the diagnosis of CRC lessen the chances of a good prognostic outcome for those affected that prompted the need for a more sensitive, accurate and non-invasive biomarker. MicroRNAs are small regulatory RNAs that are involved in the maintenance of cell homeostasis but are often altered in tumor cells. Therefore, the identification of microRNAs, that are involved in the onset and progression of CRC can be useful in staging of the disease, prognosis as well as treatment outcomes, will be the focal points of this research study. Identification of such a biomarker will negate the need for invasive diagnostic methods as well as being used routinely to screen for the presence of CRC.

This thesis, therefore, sorts to identify microRNAs that are specific for CRC and can detect this disease at each stage of tumor, node and metastasis (TNM).

Paper 1

The probability that a person will develop CRC is about 4%–5% worldwide. Furthermore, many personal habits are regarded to be risk factors of this disease as they increase the chances of development. Risk factors related to lifestyle, can be reduced by implementing modest lifestyle changes in terms of dietary and physical activities such as exercises. For instance, it is thought that a sedentary lifestyle can increase the risk of developing CRC, although this relationship between CRC and inactivity is not completely defined.

Early CRC detection tools are faced with several challenges, thereby limiting the development of a standardized biochemical diagnostic approaches which are non-invasive, more sensitive and specific for CRC stages. Also research on the discovery of drugs for the treatment of CRC is still ongoing, with several shortcomings due to the complex genetic and epigenetic events involved in its pathogenesis. However, strong evidence continues to show that certain dietary components possess cancer-protective capabilities, including therapeutic and chemopreventive

properties. These dietary factors may play a role in several stages of carcinogenesis, such as cell-cycle modulation, inflammation, apoptosis, DNA repair, and angiogenesis. MicroRNAs are intrinsically involved in similar stages of carcinogenesis, which widens the understanding between microRNAs and certain dietary components. Several factors have been linked to the disease risk factors but adopting a healthy lifestyle could be a preventable means. As a result of these, diet has been implicated in a crucial role in preventing CRC. Therefore, Paper 1 (*Effect of dietary components on miRNA and colorectal carcinogenesis*) extensively discussed the diet-microRNA interplay and identification specific microRNAs that are expressed in CRC.

Paper 2

The formation of CRC involves multistep genomic changes, including the activation of oncogenes and inactivation of tumor suppressor genes. Numerous microRNAs have been reported to play a role in cancer development, such as carcinogenesis, progression, and metastasis. Only a couple of studies have explored circulating microRNAs in patients with CRC. Furthermore, there is extremely limited research on the identification of commonly and differentially expressed microRNA for CRC staging. Efforts to depict clinical, pathological, and molecular features in patients have reached disputable ends with respect to tumor grade and disease stage at diagnosis. It is also generally acknowledged that diagnosis in patients is always difficult because of the vulnerability of both patient and the specialist to the presenting symptoms, leading to a frequent unfavorable outcome of the disease. If specific microRNAs are expressed in a certain stage of CRC, then early detection of this disease will be largely treatable. Paper 2 (*Biomarkers for Stratification in Colorectal Cancer: MicroRNAs*) further discussed the concept of this research using microRNA as a biomarker for CRC stratification for better management from paper 1. The aim of this paper was to discuss the staging of CRC with respect to specific microRNAs for early detection, treatment, efficacy, and effective management of the disease.

Paper 3

The involvement of microRNA in cancers plays a significant role in their pathogenesis. Specific expressions of these noncoding RNAs also serve as biomarkers for early CRC diagnosis, but their laboratory/molecular identification is challenging and expensive. Paper 3 (*In silico identification of microRNAs as candidate colorectal cancer biomarkers)* therefore identifies candidate microRNAs and their target genes that are specific for CRC for prognosis and diagnosis. Molecular approaches to identify these microRNAs for diagnostic and

predictive biomarkers for CRC have suffered major restraints, such as the appropriate procedure to follow, time consumption, laborious, and expensive. Since the development of a diagnostic microRNA biomarker to distinctly monitor colorectal carcinogenesis is important to enhance the diagnosis rate at latter stages, the paper aimed to use *in silico* approaches to the identify microRNAs for diagnosis in CRC. After understanding the synthesis and molecular mechanism of colorectal carcinogenesis and their diet interplay in paper 1, paper 2 succinctly identifies previous microRNAs that are specific to CRC stages using TNM and Duke's staging. Paper 3 through sequence similarity search, identified some microRNAs that are specific to CRC and further identifies their target genes. Gene ontology, pathway enrichment, prognosis and expression analysis where also carried out.

Using *in silico* approaches, paper 3 identified five candidate microRNAs with seven target genes. The patterns of expression on these genes could be inferred that patients with alterations in the microRNA prioritized target genes have significantly better overall survival than patients without these alterations. These microRNAs could potentially serve as a resource for explicitly selecting targets for diagnosis, drug development, and management of CRC. The paper also identified two hub genes namely CTNNB1 and EGFR). There is a need to associate function(s) these microRNAs through their targets.

Paper 4

In silico approaches have been developed to connect sequences of microRNAs and their targets to infer function in Cancer studies. Biologists are mainly particular about the structural and functional properties of any newly derived sequence (protein or nucleotide). *In silico* predictions are therefore important for this discovery based on successful knowledge-based principles. These principles rely on the fact that the best way to predict the structure and/ or function is to find similar sequences in existing databases using the information about them to infer conclusions about properties of the new sequence. Algorithms have been developed and implemented as computer programs (local or web-based tools) to perform this function.

Paper 4 (*Functional Prediction of Candidate MicroRNAs for CRC Management Using in Silico Approach*) determined the functional roles of the identified microRNAs in the previous paper (paper 3) using an *in silico* pipeline. Understanding microRNA's secondary structures, thermodynamic parameters, and targets may deliver greater promise towards their diagnostic potentials and mechanisms in the management of CRC. The paper predicted the secondary structure as well as the promoter regions involved in the triplex formation with targets. Also,

co-expression and both evolutionary stage-specific and variant events of the microRNA target genes were identified. This paper suggests a conformational role to modulate target interactions and therefore can be used to explain the different degree of genetic regulation in CRC.

Paper 5

Molecular simulation has emerged as an efficient and cost-effective tool in binding analysis from lead identification to optimization and beyond. The process of molecular interaction through a non-covalent bond with high affinity and specificity to form a specific complex is crucial to all processes in living organisms. Protein functions are majorly determined based on their binding interaction with other molecules or ligands.

The main focus of paper 5 (*MicroRNA Assisted Gene Regulation in Colorectal Cancer*) was to predict the molecular mechanism of gene regulation based microRNA-mRNA duplex as a lead in the silencing mechanism. This paper further discusses the mechanism by which the microRNAs obtained in paper 3 could assist in gene regulation in CRC after which the functions have been identified in paper 4. Molecular docking approaches of microRNAs and targets obtained previously were used for protein-ligand binding interaction with argonaute protein in other to determine the mechanism of gene silencing as assisted by these microRNAs. The result could be used for intermolecular recognition mechanism.

CHAPTER 2.0

EFFECT OF DIETARY COMPONENTS ON MICRORNA AND COLORECTAL CARCINOGENESIS

ADEWALE OLUWASEUN FADAKA,[1,3]* BABAJIDE A. OJO[2], OLUSOLA BOLAJI ADEWALE[3], TEMITOPE ESHO[4], AND ASHLEY PRETORIUS[5]

[1]Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa. [2]Department of Nutritional Science, Oklahoma State University, 301, Human Sciences, Stillwater, OK 74075 USA. [3]Department of Biochemistry, Afe Babalola University, P.M.B. 5454, Ado-Ekiti, Ekiti State Nigeria. [4]Institute of Biochemistry II, Medical Faculty, University of Cologne, Joseph-Stelzmann Str. 52, 50931 Cologne, Germany. [5]Biotechnology Innovation Division, Aminotek PTY LTD, Suite 2C, Oude Westhof Village Square Bellville, 7530 South Africa.

Accepted/ Published: *Cancer Cell International*, (2018) **18**(**1**): 130.
DOI.org/10.1186/s12935-018-0631-y **(Impact factor: 3.439)**

Statement of contributions of joint authorship

Adewale Fadaka (PhD candidate)

Conceived the concept of the manuscript as well as the write up and compilation.

Babajide Ojo (Collaborator)

Constitution of the Dietary analysis part of the review as well as co-author of the manuscript.

Olusola Adewale and Esho Temitope (Collaborators)

General review and editing as well as co-author of the manuscript.

Ashley Pretorius (Co-Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

This Chapter is an exact copy of the journal paper below

Fadaka *et al. Cancer Cell Int (2018) 18:130*
https://doi.org/10.1186/s12935-018-0631-y

Cancer Cell International

REVIEW

Open Access

Effect of dietary components on miRNA and colorectal carcinogenesis

Adewale Oluwaseun Fadaka[1,3*], Babajide A. Ojo[2], Olusola Bolaji Adewale[3], Temitope Esho[4] and Ashley Pretorius[5]

Abstract

Background: Colorectal cancer (CRC) is one of the most common cancers diagnosed and among the commonest causes of cancer-related mortality globally. Despite the various available treatment options, millions of people still suffer from this illness and most of these treatment options have several limitations. Therefore, a less expensive, non-invasive or a treatment that requires the use of dietary products remains a focal point in this review.

Main body: Aberrant microRNA expression has been revealed to have a functional role in the initiation and progression of CRC. These has shown significant promise in the diagnosis and prognosis of CRC, owing to their unique expression profile associated with cancer types and malignancies. Moreover, microRNA therapeutics show a great promise in preclinical studies, and these encourage further development of their clinical use in CRC patients. Additionally, emerging studies show the chemo-preventive potential of dietary components in microRNA modulation using several CRC models. This review examines the dietary interplay between microRNAs and CRC incidence. Improving the understanding of the interactions between microRNAs and dietary components in the carcinogenesis of CRC will assist the study of CRC progression and finally, in developing personalized approaches for cancer prevention and therapy.

Conclusion: Although miRNA research is still at its infancy, it could serve as a promising predictive biomarkers and therapeutic targets for CRC. Given the ever-expanding number of miRNAs, understanding their functional aspects represents a promising option for further research.

Keywords: Colorectal cancer, microRNA, Biomarkers, Diet, Chemoprevention

Background

Colorectal cancer is the third most commonly diagnosed cancer and the fourth leading cause of cancer-related deaths in the world [1]. It is the fourth most common cancer in South Africa in both male and female and the 6th most lethal of all known cancers [2]. Bray et al. [3] has predicted that there will be an increase in all CRC incidence cases from over 12 million in 2008 to about 22.2 million by 2030 [4]. About 24 million new cases of CRC was expected to be diagnosed by 2050, out of which 70% of these cases would be found in the developing countries [5]. Occurrence of this disease is greater in America and Europe when compared to low and middle-income countries. However, there is still high burden of this disease due to lack of early diagnosis of CRC as a result of limited resources in these low and medium-income countries, such as most African countries. Although, this is avoidable because CRC is one of the cancers that is almost 100% preventable [6], but most of the world's population still lack information of this disease as well as its relationship with diet.

Treatment options for CRC is largely dependent on the stage of the tumor, that is, how far it has metastasized. A common non-invasive screening test currently employed is the Fecal Occult Blood Test (FOBT), but the test presents poor sensitivity and specificity [7]. Other screening tests such as the Fecal Immunochemical Test (FIT), the fecal DNA test and the plasma SEPT-9 gene

*Correspondence: silvernonferous@gmail.com; 3782900@myuwc.ac.za
[1] Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa
Full list of author information is available at the end of the article

methylation test [8], as well as colonoscopy [9] are in use. Some clinicians use the FOBT and colonoscopy together or at different times [10]. Serum biomarker test such as carcinoembryonic antigen (CEA), carbohydrate antigens (CA) 19-9, and CA 125 used for CRC diagnosis are also non-specific [11]. To date, colonoscopy is the most commonly used test in the detection of CRC, which has been found to reduce the risk of CRC by 30–75%, but the limitation to this technique is its high cost and invasiveness [12], which makes it ineffective in resource-limited settings. Chemotherapeutic agents that often used in post-surgery lack tissue selectivity. At early stage, CRC may not show obvious signs or symptoms such as colon and/or rectal bleeding, belly pain, change in bowel habit (diarrhea), constipation, stool narrowing, and sudden weight loss. This disease can be asymptomatic until latter stages when the cancer has metastasized [13]. Globally, the major challenge to CRC treatment is early detection, which makes the current treatment options to be administered so late, typically after the cancer has metastasized. If the cancer is detected early, and polyps are removed by surgery, this will reduce both the incidence and mortality cases of CRC. To achieve this, more non-invasive, selective and specific diagnostic tools which can detect the tumour at an early need to be reviewed.

Non-coding RNAs, most especially miRNAs, are attracting considerable interest, with increasing evidences on the role of miRNAs' expression in CRC development and progression [14]. This has led to the use of miRNAs as therapeutic targets. Nevertheless, the mechanism through which a single miRNA controls gene networks by and the possible in vivo adverse effects of the miRNA and/or anti-miRNA are yet to be fully explored. As earlier mentioned, early CRC detection tools are faced with several challenges, thereby limiting the development of a standardized biochemical diagnostic approaches which are non-invasive, more sensitive and specific for CRC stages. Several factors have been linked to the disease risk factors but adopting a healthy lifestyle could be a preventable means. As a result of these, diet has been implicated in a crucial role in preventing CRC [15]. Therefore, diet-miRNA interplay and identification of the miRNAs that are expressed in CRC would be a focal point in this review.

Colorectal cancer

Colorectal cancer is the occurrence of abnormal growth in the colon or rectum. It is the fourth most common cause of cancer-related deaths and one of the most ranked type of cancer worldwide [16]. It is the second and third most common cancer in women and men, respectively, and also accounted for about 10% of the total cancer cases worldwide [16]. The cancer begins with an abnormal growth of the cells lining the colon and rectum. These cells divide uncontrollably and rapidly thereby leading to the formation of a non-cancerous growth or benign tumour known as a polyp. The polyp grows gradually and over a period of 10–20 years (Fig. 1) [17]. An adenomatous polyp or adenoma is the most common type, and about one-third to one-half of all individuals will eventually develop one or more adenomas [18]. Although not all polyps give rise to CRC, but CRC is almost always developed from a polyp and all adenomas have the possibility to be cancerous [18]. The possibility that an adenoma will become cancerous increases as it becomes bigger [19]. Cancer arising from the inner lining of the colorectum is called adenocarcinoma, and accounts for approximately 96% of all CRCs [20]. Series of DNA changes in a polyp's cell result in its the development into malignant tumor over a period of time (Fig. 1). Initially, these cancer cells are confined to the surface of a polyp, but can grow into the wall of the colon or rectum, which eventually spreads to lymph nodes and other organs, such as the liver or lungs [21].

CRC pathogenesis and miRNA involvement

The molecular mechanism of colorectal carcinogenesis is a multistep process involving genetic, epigenetic and aberrant immunologic pathway as a major contributor of colorectal carcinogenesis [22–24]. miRNAs are often dysregulated in tumors either by genetic or epigenetic factors, and are currently being investigated for their potential as biomarkers in cancer diagnostics (Fig. 2) [25]. This non-coding RNA has been implicated in the mechanism by which gene expression of various cancer-associated genes are controlled and their expression may be altered in the process. A series of studies have highlighted the role of miRNAs in the development of this disease. CRC-related miRNAs have been demonstrated to regulate the genes by various mechanisms, including epigenetic modifications, long non-coding RNA–miRNA, and long non-coding RNA–protein interactions, and by their actions as miRNA precursors. Since miRNAs can be detected in human body fluid and have good specificity and accessibility, they have been suggested to be used as novel potential biomarkers for CRC diagnosis and prognosis as well as in the prediction of the response to therapy [26]. miRNAs have been implicated in a number of events, such as epigenetic, transcriptional, and post-transcriptional regulation [27]. These non-coding RNAs exhibit unique profiles in various human cancers such as colorectal cancer, reflecting disease progression [28]. Studies have previously reported the involvement of miRNAs in cancer initiation and progression but recently, their roles as drivers of tumor suppressor and oncogenic function have been evaluated in several cancer types [29].

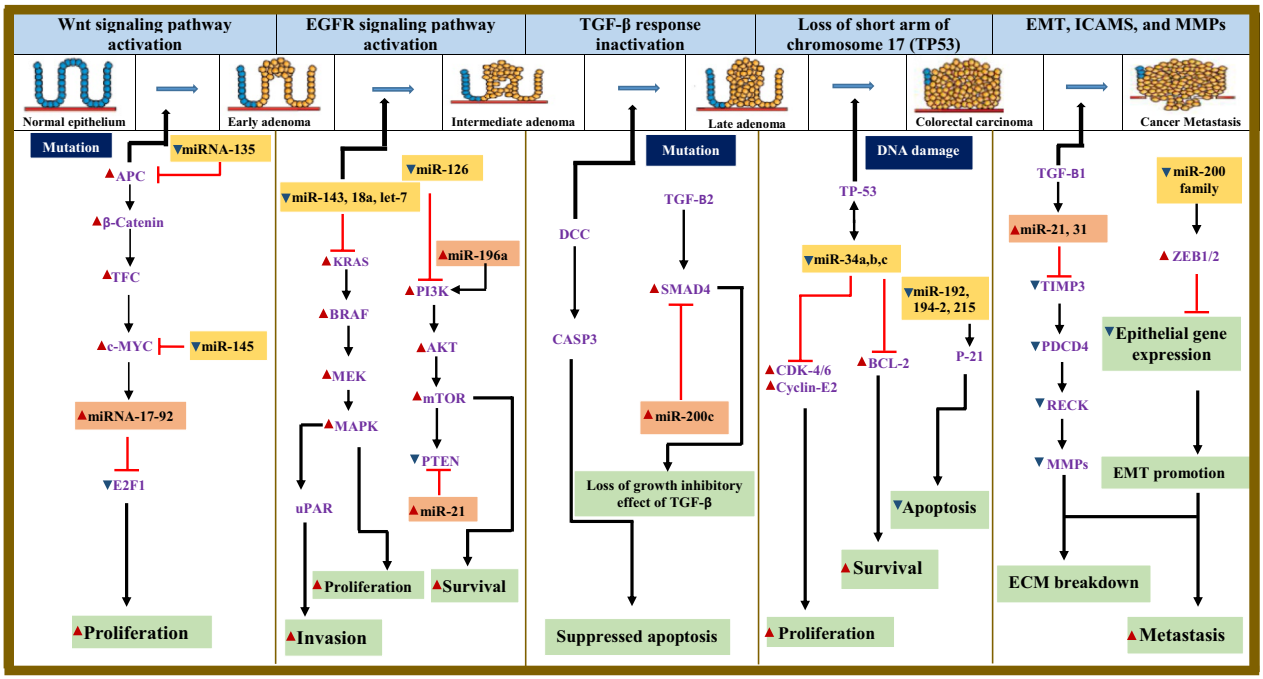


Fig. 1 microRNAs and genetic/epigenetic alterations involved in adenoma-carcinoma sequence. Red arrows—up-regulation; blue arrows—downregulation. Experimentally validated miRNAs are shown alongside with their target genes in altered expression in CRC

Several studies have also shown the association of non-coding RNAs in colorectal carcinogenesis through the stimulation or inhibition of apoptosis, cell proliferation, differentiation, invasion and metastasis [30–35].

Genetic regulation of colorectal carcinogenesis

Genetic instability has been considered fundamental to the multistep process of tumor growth and metastatic progression for decades [36]. A wide range of genetic mutations is found in most cancer subtypes (Fig. 2). The most common gene mutation in CRC is the APC gene from the normal epithelium followed by the K-ras, DCC and p53 genes leading to sporadic carcinoma [37]. DNA hypermethylation of tumor suppressor gene promoter regions has been intensively studied to demonstrate its critical role in gene silencing (Fig. 2). Histone modification includes histone methylation and deacetylation, both of which have been shown to be associated with DNA methylation [22].

Epigenetic modifications and miRNA in CRC

Epigenetic alterations have the ability to deregulate the expression of any type of transcript. However, the exact mechanisms of epigenetic regulation of non-coding RNAs are still unclear although, these RNAs are subject to the same epigenetic regulatory mechanisms as protein-coding genes. Several studies have reported the regulatory mechanism of miRNA to clarify the network that underlie the aberrant expression in tumor metastasis. Furthermore, aberrant epigenetic regulation affects abnormal miRNA expression in cancers. miR-21, miR-106, and miR-144 were reportedly upregulated in patients samples with CRC compared with normal individuals [38]. miR 143 and miR-145 were significantly downregulated in colorectal adenoma compared to normal colon sample [39]. These miRNAs were further confirmed to be significantly reduced in colorectal neoplasia and act as tumor suppressor miRNAs in the colorectum [40–44]. Zhang et al. [45] also revealed the induction of apoptosis through BCL-2 inhibition by miR-148a upregulation in CRC while the downregulation was linked to increased tumor size [46]. Attenuated miR-34a and miR-200c expression are associated with metastasis in CRC [47, 48]. Lujambio et al. [49] identified cancer-specific CpG island hypermethylation of the promoter lesion with the transcription of miR-148a, miR-34b/c, and miR-9. miR-34a also have effect on colorectal cancer invasion and metastasis in conjunction with IL-6R, ZNF281, MET, snail family zinc finger 1 and 2 (SNAI1, SNAI2) and β-catenin (CTNNB1) [47, 50–52].

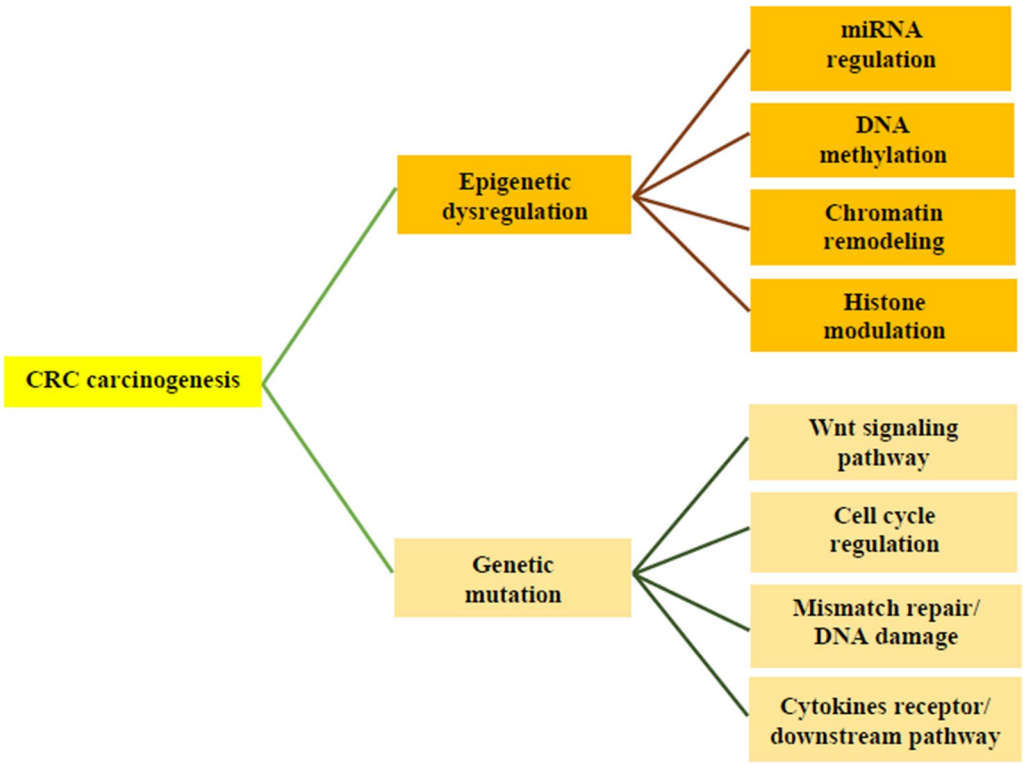


Fig. 2 Genetic and epigenetic mechanisms of colorectal carcinogenesis. Mutation of genes involved in the Wnt signaling pathway plays a superior role in colorectal carcinogenesis. Genes that are related to cell cycle progression. DNA repair, and cytokine signaling have also been shown to be pivotal in colorectal carcinogen. DNA hypermethylation of tumor suppressor gene promoter regions has been intensively studied to demonstrate its critical role in gene silencing. Histone modification includes histone methylation and deacetylation, both of which have been shown to be associated with DNA methylation

Overview of miRNA

Micro RNAs are short single stranded non-coding RNAs, consisting of about 19–25 nucleotides. They are responsible for the regulation of translation of genes by binding to the 3′-untranslated region of target mRNAs through sequence-specific manner. These miRNAs reportedly play vital roles in inflammation and carcinogenesis, which can be linked to their oncogenic or tumor suppressive properties [53]. Alterations in miRNA expression are implicated in different human cancers, which include breast cancer, CRC, liver cancer and lung cancer [54]. For gene silencing, cells can use miRNA, which binds and represses messenger RNA (mRNA), thereby turning off genes that are not required in translating genetic information into proteins. This miRNA participates in the regulatory mechanisms of cell's development through death, and the dysregulation can be implicated in several diseases such as cancer and heart diseases [55].

miRNAs have been recognized as potential biomarkers for early detection, as well as prognostic and therapeutic approach for CRC because of their high level of specificity and selectivity.

Synthesis of miRNA

As earlier mentioned, miRNA is an important class of post-transcriptional regulators of about 22 nucleotides in length [56], and it carries out its biological functions by binding to the 3′ untranslated regions (UTRs) of its target messenger RNA/s (mRNA/s), thereby repressing its expression [57]. A single miRNA may regulate multiple targets and thus act as a chief controller of gene expression. Human genes (about 30%) can be regulated by miRNAs as suggested by bioinformatic analysis, despite the constitution of 1–3% miRNA of the human genome [58]. Several miRNA-coding genes operate as independent transcription units, which contain their own promoters and regulatory elements. However, about a quarter of miRNA genes are intronic and transcribed alongside their host genes [59].

Like proteins, genes coding for miRNAs are contained in the nucleus. miRNA can be synthesized from the introns of a functional gene coding for a specific mRNA or from its own gene (Fig. 3). The same enzyme that produces mRNA (RNA polymerase II) transcribes each gene of coded miRNA resulting in a primary miRNA

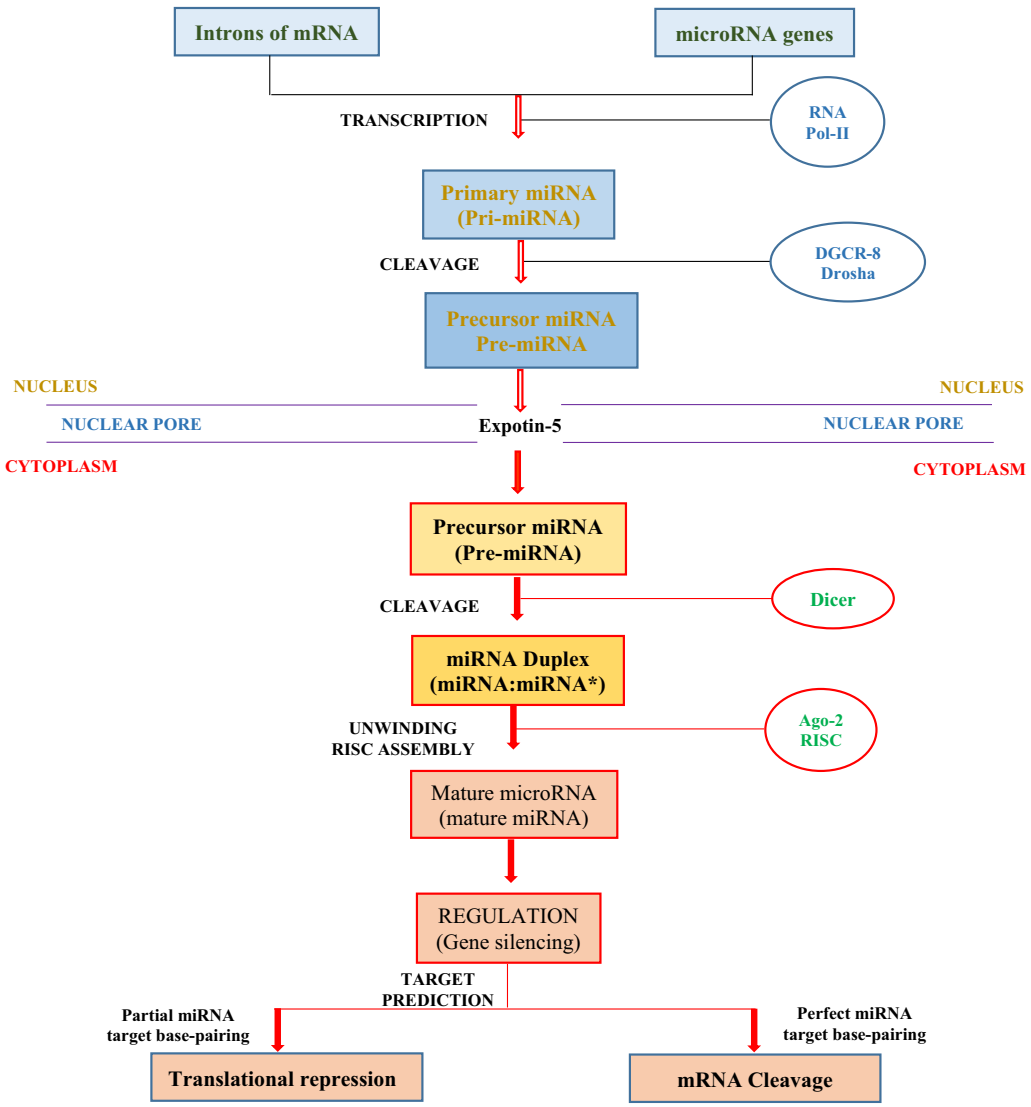


Fig. 3 The schematic diagram shows the synthesis and mode of regulation of miRNA from the nucleus to the cytoplasm. The spherical cycles indicate the proteins/enzyme and co-factors responsible for the synthesis of the molecule. Also, the big spherical cycle at the beginning of the reaction indicated that the enzyme is responsible for both mRNA and miRNA synthesis. *DGCR-8* DiGeorge syndrome Critical Region 8 protein, *AGO-2* argonaute protein

(pri-miRNA), which consists of a 5′ G-cap, at least an approximately 60–70-nucleotide hairpin structure and a 3′ poly (A) tail [60]. Polycistronic pri-mRNA may contain up to seven hairpin structures that produce different mature miRNAs. This pri-mRNA is the final microRNA with regulatory function after several steps. After transcription, the double-stranded stem is recognized by the cofactor DiGeorge syndrome Critical Region 8 protein (DGCR8). An enzyme (Drosha) associates with DGCR8 to form a microprocessing complex capable of cutting the pri-miRNA into a smaller precursor miRNA (pre-miRNA) by the removal of 5′ cap, the 3′ poly (A) tail and sequences flanking the hairpin structure. Precursor-miRNA is then moved from the nucleus through the nuclear pore to the cytoplasm with the aid of Exportin 5, moves where it inactivates mRNA of one or multiple genes [61]. In the cytoplasm, the stem-loop of the pre-miRNA is further cleaved by a large microRNA protein called dicer (an endoribonuclease) to form a short double-stranded microRNA molecule (about 20–25 nucleotides long) consisting of mature miRNA strand and its complementary strand [62].

Consequently, argonaute protein (AGO-2) interacts with dicer to bind the mature miRNA due to the

asymmetric thermostability. The miRNA is unwound, and one strand is released (the passenger strand). The guide strand interacts with AGO-2 (responsible for the endonuclease activity that induces mRNA cleavage) and some additional proteins to form the RNA Induced Silencing Complex (RISC). This is guided to the mRNA target to activate one or multiple genes [63]. The mRNA of a target gene is complementary to the sequence of the miRNA that enables base pairing. Complete and partial complementarity between the seed region (nucleotide positions 2–8) of miRNA and its target mRNAs results in mRNA degradation and translational inhibition or repression, respectively [64]. In the case of translational inhibition or repression, the RISC complex prevents the ribosome subunit from binding. The mechanism by which RISC induces translational repression is more complex and may include cap-dependent inhibition of translation initiation, eukaryotic translation initiation factor-6 recruitment to RISC, nascent protein degradation, ribosomal drop-off and prevention of the interaction between poly (A)-binding proteins and eukaryotic translation initiation factor-4G subsequent to mRNA deadenylation [64, 65]. In both cases (mRNA degradation and translational inhibition or repression), the mRNA will not be translated into a protein and the gene is silenced. Since their discovery in the 1900s, major parts of the miRNA's pathways still remain unclear. However, with their essential role in many biological processes (metabolism, stem cell division, development, apoptosis, cell proliferation, cell cycle control and cell stem differentiation), mRNA offers great potential in medicine and might lead to key treatment of various diseases in the future.

General functions of miRNA

Several biological functions of miRNA have been reported to be related to various disease mechanisms, regulation of cellular activities and cancer progression [66–69].

Seed region of about 2–8 nucleotides base pair allows miRNAs to bind at different degrees of complementarity. These therefore enables the recognition and binding of a variety of mRNAs which potentially regulate translation and expression of its protein products. Any change in the levels of a specific miRNA expression affects several biological pathways. Partial base pairing inhibits translation without interfering with the integrity of mRNA [70]. The observed discrepancies between mRNA and protein expression levels may be explained by the miRNA action, and the information on miRNA expression and function suggests the regulation of protein expression.

miRNAs are involved in various biological activities including cell differentiation, proliferation, apoptosis, and migration, which are key regulators in various pathogenesis and progression of different diseases, especially cancers [71–73]. miR-15 and miR-16, the first two miRNAs associated with cancer, play a significant role in the regulation of apoptosis by targeting the anti-apoptotic bcl-2 mRNA [74]. Also, the expression of human Ras, regulated by let-7 in cell culture, was also reported as the first miRNA-target interaction with relevance to cancer [75]. Subsequently, numerous publications have reported the role of miRNAs in tumors [76–80].

miRNA tumor-specific metabolic reprogramming

Cancer cells are shown to experience characteristic changes in their metabolic programs suggesting that metabolic shifts supports tumor cells growth and survival [81]. Report have it that the miRNA expression patterns in human cancers are not the same and that different cancer types have distinct expression patter [82]. This is so because the processing of primary miRNA transcripts to mature RNA is transcribed by RNA polymerase II (Fig. 3). This RNA polymerase II is also responsible for the transcription of mRNAs. Several alterations in miRNA levels have been revealed between colorectal cancer and normal colonic mucosa [83–85]. Gao et al. [86] reported that the c-Myc oncogenic transcription factor, which is known to regulate microRNAs and stimulate cell proliferation, transcriptionally represses miR-23a and miR-23b, resulting in greater expression of their target protein. Interestingly, c-Myc directly binds to the transcription subunit of microRNA (miR)-23a/b and subsequently contributes to the up-regulation of mitochondrial glutaminase 1 via the induction of ASCT2/SLC1A5 transporter. Moreover, the association of c-Myc with miR-17-92 cluster has been shown to inhibit the activity of phosphatase and tensin homologue deleted on chromosome 10 (PTEN) and activates PI3K-Akt-mTOR axis leading to cell survival in early stage adenoma in CRC [87]. The complex crosstalk between miRNA and Myc is considered to be partially responsible for metabolic reprogramming. In addition, metformin induces miR-27b-mediated suppression of ENPP1, which reduces chemoresistance and tumor seeding potential [88].

Expression of specific miRNAs in cancer

Understanding the deregulation of miRNA expression observed in cancer cells is crucial. Studies have confirmed that when a miRNA is down-regulated in cancer and targets an oncogene, it may act as a tumor suppressor, or may act as an oncogene when up-regulated and targets a tumor suppressor or a gene important for differentiation [89–91].

Carden et al. [92] reported that increased miR-663 expression in breast tumors consistently correlates with

increased patient survival, which demonstrates its epigenetic regulation and role in breast tumorigenesis. Also, miR-663a down-regulation was observed in human non-small cell lung cancer progression by targeting AP-1 component JunD in the cancer cells [93]. miR-34a, a chief regulator of tumor suppression, maintains its own expression levels through upstream signaling and activate tumor suppressor pathways, which are regulated by p53 [94]. Wiggins et al. [95] reported that this miRNA inhibits its cancer cells lacking endogenous p53.

miRNA has also been implicated in the repression of over 700 transcripts associated with cellular proliferation, survival, and plasticity [96]. High expression of miR-21 predicts poor survival in CRC patient [97–101]. In a contradicting report of Lee et al. [102], the expression of miR-21 in the periphery of primary tumors demonstrated the significance of miRNA as a better prognosis in patients with advanced stage CRC. Molecular validation result of miR-22 expression revealed a significant increase in gastric cancer tissues when compared to adjacent non-cancerous tissues, and that low expression of miR-22 is associated with aggressive gastric cancer phenotype and its poor survival [103]. As suggested in previous studies, miR-22 is associated with several cellular processes, and their deregulation is a hallmark of several human cancers such as ovarian, prostate, colon and liver cancers [104–106]. James et al. [107] also reported the clinical utility of miR-21 and let-7g in prostate cancer. Li et al. [108] investigated the level and role of miR-106a expression in pancreatic cancer and reported that pancreatic cancer cell invasion was dependent on miR-106a regulation [109, 110].

Diet interaction with microRNAs in colorectal cancer

Research on the discovery of drugs for the treatment of cancer is still ongoing, with several shortcomings due to the complex genetic and epigenetic events involved in its pathogenesis. However, strong evidence continues to show that certain dietary components possess cancer-protective capabilities, including therapeutic and chemopreventive properties. These dietary factors may play a role in several stages of carcinogenesis, such as cell-cycle modulation, inflammation, apoptosis, DNA repair and angiogenesis [111]. miRNAs are intrinsically involved in similar stages of carcinogenesis, which widens the understanding between miRNAs and certain dietary components (Fig. 4). Certain dietary components of plant origin may be less bio-available and thus, escape digestion into the large intestine. Therefore, these bioactive components may then play a role in modulating CRC.

Resveratrol

Resveratrol, polyphenols found largely in the skin of grapes, raspberries, mulberries, and blueberries, are generally considered to have several health benefits due to its antioxidative properties. Few studies have shown the potential role of resveratrol against colon cancer. Treatment of SW480 colon cancer cells with 50 μM of resveratrol for 14 h prevented the expression of several oncogenic miRNAs, such as miR-21 which is induced in chronic inflammation [112]. Furthermore, the expression of a tumor-suppressor miRNA, miR-663, was significantly higher in cancer cells when its expression was compared to untreated cells. The use of resveratrol in the treatment of colon cancer cells has led to a reduction in TGFβ1 and its downstream effector SMAD3, this could be explained by the target of miR-663 on TGFβ1 transcripts [112]. This finding on miR-663 is of significance since activation of the TGFβ1 pathway increases angiogenesis and metastasis in later stages of cancer [112, 113]. In addition, resveratrol was also shown to inhibit the up-regulation of miR-122, an oncogenic and inflammation-induced miRNA, which is dependent manner on miR-663 [114]. Another study showed that resveratrol inhibited the growth of human colon cancer cells by up-regulating miR-34a, which in turn down-regulated the E2F3 and Sirt1 genes [115]. Therefore, resveratrol possesses anti-inflammation and anti-cancer capabilities, which might be linked to its antioxidative properties.

Fatty acids

Short-chain fatty acids (SCFAs) and omega-3 fatty acids have been shown to have cancer-protective properties. Short-chain fatty acids are important end products of the gut microbial fermentation of non-digestible carbohydrates in the diet. Butyrate most importantly is vital for the proliferation and sustenance of colonic epithelial cells. As such, several studies have reported the involvement of microbial-derived butyrate from dietary fiber fermentation as a histone deacetylase (HDAC) inhibitor and thus plays a protective role against colorectal tumorigenesis [116–118]. However, the chemoprotective effect of butyrate on specific miRNAs remains to be fully elucidated. In human colon carcinoma cells, microarray analysis revealed that treatment with 2 mM butyrate changed the levels of various aberrantly expressed miRNAs [119]. Notably, butyrate treatment decreased the expression of miR-106a and miR-106b, which was accompanied by a reduction in cell proliferation [119]. Furthermore, in other human colon cancer models, treatment with 1–25 mM butyrate attenuated the expression of an oncogenic miR-17-92 cluster of miRNAs, while inhibiting cell differentiation and promoting apoptosis [120, 121].

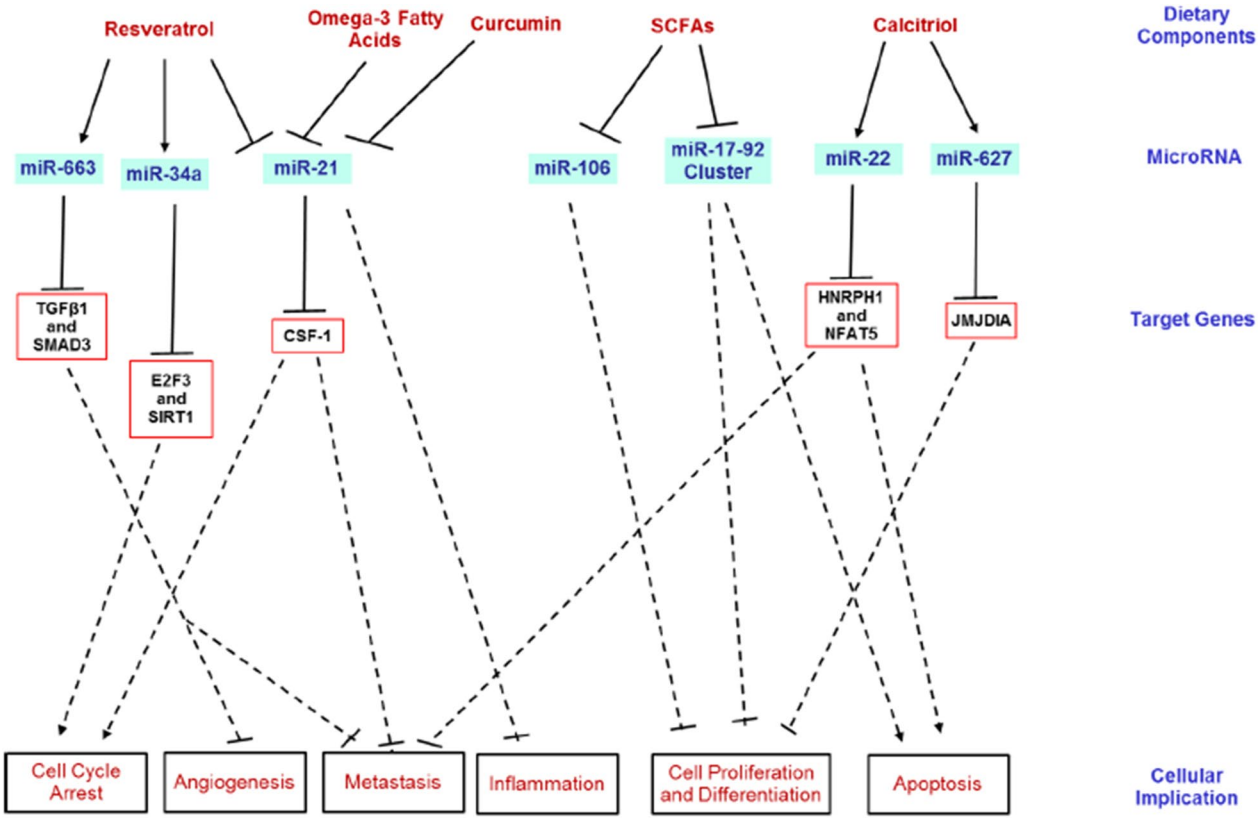


Fig. 4 Dietary modulation of microRNAs in colorectal cancer. Several dietary components show chemo-preventive and therapeutic potential in CRC pathogenesis through the modulation of miRNAs in difference signaling pathways. Line arrows indicate up-regulation; blunted lines with flat heads indicate inhibition, while dotted lines indicate multiple steps. *CSF-1* Colony stimulating factor 1, *E2F3* E2F transcription factor 3, *HNRPH1* heterogeneous nuclear ribonucleoprotein H1, *JMJDIA* Jumonji domain-containing protein 1A, *NFAT5* nuclear factor of activated T-cells 5, *SIRT1* Sirtuin (silent mating type information regulation 2 homolog) 1, *SMAD3* mother against decapentaplegic homolog 3, *TGFβ1* transforming growth factor beta 1

Omega-3 fatty acids may have protective effect against inflammatory diseases, including cancer [122, 123]. Specifically, it was reported that fish oil prevented the down-regulation of several miRNAs in the colon of rats 34-weeks post-injection with azoxymethane. Such miRNAs include miR-15b, miR-107, let-7d, miR-191, and miR-324-5p. This effect corresponds to a significant reduction in colon tumorigenesis [124]. Similarly, the expression of miR-21 was significantly diminished in breast cancer cell lines treated with fish oil, thus repressing CSF-1 levels which have a significant role in breast tumorigenesis and metastasis [125]. Put together, these findings strongly suggest the chemo-preventive potential of SCFAs and omega-3 fatty acids (that could be obtained through the diet).

Curcumin

Curcumin, a phytochemical found in turmeric, has been widely studied for its several health benefits, including antioxidant, anti-inflammatory and anti-cancer properties. Studies have also reported miRNA modulation in various cancer models. Recently, curcumin was reported to reduce the expression of miR-21, which is over-expressed in many tumors leading to cancer progression and metastasis [126]. Treatment of human colon carcinoma cells (HCT-116) with curcumin reduced miR-21 activity in a dose-dependent manner, thereby leading to cell cycle arrest at the G_2/M phase, thus reducing cell proliferation and tumor growth [126]. A similar effect of curcumin on miR-21 was also reported in a pancreatic cancer cell model [127]. Other studies have also reported the beneficial effect of curcumin in miRNA modulation in various cancer models, including pancreatic cancer [127–129] and lung cancer [130].

Vitamin D

Vitamins A, D, and E have been reported to play an anti-cancer role involving the modulation of miRNAs [131, 132], amongst which vitamin D have an active chemo-preventive role in CRC development. Early

epidemiological evidence suggested an inverse relationship between vitamin D levels and CRC [133]. Further studies in human colon cancer cells revealed that treatment with 10 μM calcitriol (an active form of vitamin D) induced miR-22 expression which further inhibits cell proliferation and migration. These effects were time- and dose-dependent, and also dependent on the activation of vitamin D receptor [134]. Moreover, up-regulation of miR-22 by vitamin D in the colon cancer cells is necessary for the repression of several vitamin D target genes, such as HNRPH1 and NFAT5, which mediate apoptosis inhibition and cancer invasion, respectively [134]. In addition, the expression of miR-627 was up-regulated following incubation of human CRC cells (HT-29) with calcitriol, which down-regulates JMJD1A (a gene involved in histone methylation), and prevent cell proliferation and differentiation [135]. Thus, current knowledge posits that vitamin D has cancer-suppressive potentials, which may be mediated via microRNA activation.

Selenium

Selenium is an essential trace mineral with an antioxidant activity, which was shown to be beneficial in promoting cardiac health and preventing cancer development [136]. Although its role in cancer prevention has been widely reported [136], little is known about its effect on miRNA activity in cancer models. Of note, incubation of human prostate cancer cells (LNCaP) with sodium selenite (2.5 μM) up-regulated members of the miR-34 family, resulting in a selenium-induced expression and activation of the tumor-suppressor p53, and its downstream targets [137]. Other metabolites of selenium, including methylselenocysteine and selenomethionine, have been found to possess HDAC-inhibiting activity in human colon cancer cells [138], but the knowledge of possible miRNAs involved is still vague.

Soy isoflavones

Diadzein, genistein, and glycitein are soy isoflavones that were reported to have anti-tumor properties via the modulation of the estrogen receptor [131]. Their chemopreventive and anti-metastasis potential via the modulation of miRNAs was reported in pancreatic cancer [139], prostate cancer [140], and ovarian cancer [141] models. It is interesting to investigate the potential role of the soy isoflavones in colon cancer, since these isoflavones act via the modulation of estrogen receptor. It was suggested that an up-regulation of the estrogen receptor beta (ERβ) signaling in SW480 colon cancer cells showed antiproliferative effects by silencing the effect of oncogenic miRNAs [142].

Ellagitannin

Ellagitannins are hydrolyzable polymeric polyphenols found in many fruits and nuts. Initial characterization of ellagitannins showed their potent antioxidant, anti-inflammatory, anti-proliferation and pro-apoptotic capabilities [143]. More recently, ellagitannin was shown to possess anti-neoplastic properties in a human liver cancer cell line HepG2, while modulating the expression of 25 miRNAs [144]. However, the specific mechanisms of the ellagitannin-miRNA interplay in cancer is still unknown.

Caloric restriction

Caloric restriction (CR) generally refers to a $\leq$60% dietary energy deficit without malnutrition [145]. The beneficial effects of CR have been reported in various conditions, including aging and cancer. CR has long been known to play a vital role in colon cancer prevention, but specific mechanisms and miRNAs involved still requires further evaluation [146, 147]. The anti-cancer effects of CR may be due to its influence on cellular senescence [148]. The Hayflick limit, which described cellular senescence as a stable cell cycle arrest regardless of growth conditions, was thought to protect against the heightened proliferation of cancer cells [149, 150]. In paradox, senescent cells may also contribute to tumorigenesis in various tissues, through the production of an array of cytokines, chemokines, proteases and growth factors, collectively referred to as the senescence-associated secretory phenotype (SASP) [151, 152]. Unsurprisingly, NFκβ is known to play a role in regulating various inflammatory pathways involved in producing the senescence secretome, that drives the chronic low-grade inflammation capable of driving tumor initiation and progression [153, 154].

Some of the consequences of overnutrition-induced obesity are hyperinsulinemia and hyperleptinemia, resulting in insulin and leptin resistance respectively. These may serve as growth factors leading to the activation of NFκβ, thus leading to chronic inflammation characteristic of many tumors [154–156]. On the other hand, CR may impact the obesity-cancer pathway, by reducing serum insulin, leptin, and associated inflammation by limiting NFκβ—related gene expression [157, 158]. Specifically, injection of mice on a 30% CR diet with MC38 colon tumor cells, led to a reduction in tumor size, serum growth factors and a downregulation of inflammatory genes induced by NFκβ [157]. Similarly, 5-week feeding of a 30% CR diet in mice showed inhibitory effects on pancreatic tumor growth, IGF-1 and NFκβ-related inflammatory gene expression [158]. Still, possible miRNAs involved in the anti-tumor effects of CR in relation to the NFκβ-SASP pathway are still largely

unclear. Few breast cancer models have shown that CR may impact miRNAs, by showing inhibitory effects on miR17/20a and miR200a, leading to a reduction in extracellular matrix proteins, tumor progression and metastasis [159, 160]. Put together, it may be hypothesized that CR possesses anti-cancer effects by decreasing chronic inflammation through the limitation of NFκβ activity in senescent cells. However, this concept, potential mechanisms, and miRNAs involved are interesting subjects for future studies. Understanding this effect of CR may be important in preventing colorectal cancer and other cancers in our obese and older adult populations where low-grade inflammation and cellular senescence are more observed, respectively.

Conclusion

It is now a known fact that CRC is a major depravity that affects the world based on lifestyle changes and sometimes based on age or hereditary factors. Regular screening for CRC is essential and should be done to detect tumor early before it metastasizes. Several screening and treatment methods have been employed for CRC, which have been of help to date but present several limitations. Recently, the involvement of 18–22 nucleotide to the foreknown miRNA, and its relation to dietary factors and tumorigenesis. This microRNA can be differentially and commonly expressed depending on its stage and location of the tumor. The ability of microRNA to differentiate between CRC patients and healthy patients in a non-invasive approach for CRC detection makes it a good diagnostic biomarker. Currently, little is known on the impact of diet on miRNAs in CRC, as most studies were only centered on in vitro models. Studies providing information on the use of miRNA-specific knockout should be considered in various in vivo models. Apart from the few described in this review, other dietary components such as folate and methyl-deficient diets, indoles and isothiocyanates (from cruciferous vegetables) and tea catechins have been widely shown to possess chemo-preventive properties but their effect via the modulation of microRNAs in the colon and rectal cancer is still unclear. Collectively, bioactive components from the diet modulate several miRNAs which are involved in cancer development and growth via several mechanisms. Due to their potent chemo-preventive properties, it is therefore pertinent for public health specialists and health organizations to consider incorporating these dietary components into the nutrition sensitization program to prevent or reduce the menace of CRC and other malignancies.

Abbreviations

CRC: colorectal cancer; miRNAs: microRNAs; RISC: RNA-induced silencing complex; UTR: the 3′untranslated region; DGCR-8: DiGeorge syndrome Critical Region 8 protein; AGO-2: argonaute protein; ERβ: estrogen receptor beta; SCFAs: short-chain fatty acids; CASP3: cysteine-aspartic acid protease 3; APC: adenomatous polyposis coli; MMPs: matrix metalloproteinases; DCC: deleted in colorectal carcinoma; EGFR: epidermal growth factor receptor; ICAM: intercellular adhesive molecules; PDCD4: programmed cell death 4; PTEN: phosphatase and tensin homolog; CDK4,6: cyclin-dependent kinase 4,6; ECM: extracellular matrix; EMT: epithelial-to-mesenchymal transition; RECK: reversion-inducing cysteine-rich protein with kazal motifs; TIMP3: tissue inhibitor of metalloproteinase 3; uPAR: plasminogen activator, urokinase receptor; TGFβRI/II: transforming growth factor βreceptor I/II; ZEB1/2: zinc-finger E-box binding homeobox-1; CTGF: connective tissue growth factor; TSP1: thrombospondin-1.

Authors' contributions

AF and AP conceived the concept, design, and first draft of the study. BO helped with the constitution of the Dietary analysis part of the review, ET and OA contributed to the revision and critical intellectual contents. All authors contributed significantly. All authors read and approved the final manuscript

Author details

[1] Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa. [2] Department of Nutritional Science, Oklahoma State University, 301, Human Sciences, Stillwater, OK 74075, USA. [3] Department of Biochemistry, Afe Babalola University, P.M.B. 5454, Ado-Ekiti, Ekiti State, Nigeria. [4] Institute of Biochemistry II, Medical Faculty, University of Cologne, Joseph-Stelzmann Str. 52, 50931 Cologne, Germany. [5] Biotechnology Innovation Division, Aminotek PTY LTD, Suite 2C, Oude Westhof Village Square, Bellville 7530, South Africa.

Acknowledgements

Authors will like to appreciate the contributions of the following for their support both with scientific materials, emotionally and physically. Bioinformatics Research Group of the University of the Western Cape, Dr. Taiwo D. Akinsoji (MBBS), Dr. Ashwil Klein, and The School of Post Graduate Studies, University of the Western Cape for their 1-week publication retreat.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Not applicable.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

Not applicable.

Publisher's Note

Received: 22 June 2018 Accepted: 31 August 2018
Published online: 06 September 2018

References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2016. https://doi.org/10.1136/gutjnl-2015-310912.
2. Brand M, Gaylard P, Ramos J. Colorectal cancer in South Africa: an assessment of disease presentation, treatment pathways and 5-year survival. SAMJ S Afr Med J. 2018;108:118–22.
3. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. Lancet Oncol. 2012;13:790–801.

4. Alwan A. Global status report on noncommunicable diseases 2010. Geneva: World Health Organization; 2011.
5. Saluja S, Alatise OI, Adewale A, Misholy J, Chou J, Gonen M, Weiser M, Kingham TP. A comparison of colorectal cancer in Nigerian and North American patients: is the cancer biology different? Surgery. 2014;156:305–10.
6. CECP-Nigeria. Colon cancer prevention in Nigeria—an achievable goal. Guardiance Newspaper, © 2015 Committee Encouraging Corporate Philanthropy (CECP-Nigeria). 2015.
7. Ahmed FE, Jeffries CD, Vos PW, Flake G, Nuovo GJ, Sinar DR, Naziri W, Marcuard SP. Diagnostic microRNA markers for screening sporadic human colon cancer and active ulcerative colitis in stool and tissue. Cancer Genom Proteom. 2009;6:281–95.
8. Song L-L, Li Y-M. Current noninvasive tests for colorectal cancer screening: an overview of colorectal cancer screening tests. World J Gastrointest Oncol. 2016;8:793.
9. Lieberman D. Progress and challenges in colorectal cancer screening and surveillance. Gastroenterology. 2010;138:2115–26.
10. Heo J, Jeon SW, Jung MK, Kim SK, Kim J, Kim S. Endoscopic resection as the first-line treatment for early colorectal cancer: comparison with surgery. Surg Endosc. 2014;28:3435–42.
11. Vukobrat-Bijedic Z, Husic-Selimovic A, Sofic A, Bijedic N, Bjelogrlic I, Gogov B, Mehmedovic A. Cancer antigens (CEA and CA 19-9) as markers of advanced stage of colorectal carcinoma. Med Arch. 2013;67:397.
12. Sovich JL, Sartor Z, Misra S. Developments in screening tests and strategies for colorectal cancer. BioMed Res Int. 2015. https://doi.org/10.1155/2015/326728.
13. Kawamura M, Toiyama Y, Tanaka K, Inoue Y, Kusunoki M. Can circulating microRNAs become the test of choice for colorectal cancer? Curr Colorectal Cancer Rep. 2014;10:403–10.
14. Reddy KB. MicroRNA (miRNA) in cancer. Cancer Cell Int. 2015;15:38.
15. Hou N, Huo D, Dignam JJ. Prevention of colorectal cancer and dietary management. Chin Clin Oncol. 2013;2:13.
16. Stewart B, Wild CP. World cancer report 2014. Geneva: World Health Organisation; 2017.
17. Winawer SJ, Zauber AG. The advanced adenoma as the primary target of screening. Gastrointest Endosc Clin. 2002;12:1–9.
18. Levine JS, Ahnen DJ. Adenomatous polyps of the colon. N Engl J Med. 2006;355:2551–7.
19. Pickhardt PJ, Kim DH, Pooler BD, Hinshaw JL, Barlow D, Jensen D, Reichelderfer M, Cash BD. Assessment of volumetric growth rates of small colorectal polyps with CT colonography: a longitudinal study of natural history. Lancet Oncol. 2013;14:711–20.
20. Mundade R, Imperiale TF, Prabhu L, Loehrer PJ, Lu T. Genetic pathways, prevention, and treatment of sporadic colorectal cancer. Oncoscience. 2014;1:400.
21. Rothwell PM, Wilson M, Elwin C-E, Norrving B, Algra A, Warlow CP, Meade TW. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet. 2010;376:1741–50.
22. Yang T, Owen JL, Lightfoot YL, Kladde MP, Mohamadzadeh M. Microbiota impact on the epigenetic regulation of colorectal cancer. Trends Mol Med. 2013;19:714–25.
23. Arends MJ. Pathways of colorectal carcinogenesis. Appl Immunohistochem Mol Morphol. 2013;21:97–102.
24. Cheetham S, Gruhl F, Mattick J, Dinger M. Long noncoding RNAs and the genetics of cancer. Br J Cancer. 2013;108:2419.
25. Mazeh H, Mizrahi I, Ilyayev N, Halle D, Brücher BL, Bilchik A, Protic M, Daumer M, Stojadinovic A, Avital I. The diagnostic and prognostic role of microRNA in colorectal cancer—a comprehensive review. J Cancer. 2013;4:281.
26. Xie X, Tang B, Xiao Y-F, Xie R, Li B-S, Dong H, Zhou J-Y, Yang S-M. Long non-coding RNAs in colorectal cancer. Oncotarget. 2016;7:5226.
27. Zhang J, Zhang A, Wang Y, Liu N, You Y, Kang C, Pu P. New insights into the roles of ncRNA in the STAT3 pathway. Future Oncol. 2012;8:723–30.
28. Wapinski O, Chang HY. Long noncoding RNAs and human disease. Trends Cell Biol. 2011;21:354–61.
29. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. Cancer Discov. 2011;1:391–407.
30. Liu Q, Huang J, Zhou N, Zhang Z, Zhang A, Lu Z, Wu F, Mo Y-Y. LncRNA loc285194 is a p53-regulated tumor suppressor. Nucleic Acids Res. 2013;41:4976–87.
31. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 1071;2010:464.
32. Zhang L, Yang F, Yuan J-H, Yuan S-X, Zhou W-P, Huo X-S, Xu D, Bi H-S, Wang F, Sun S-H. Epigenetic activation of the MiR-200 family contributes to H19-mediated metastasis suppression in hepatocellular carcinoma. Carcinogenesis. 2012;34:577–86.
33. Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, Ansell PJ, Zhao J, Weng C, Klibanski A. Activation of p53 by MEG3 non-coding RNA. J Biol Chem. 2007;282:24731–42.
34. Ling H, Spizzo R, Atlasi Y, Nicoloso M, Shimizu M, Redis R, Nishida N, Gafà R, Song J, Guo Z. CCAT2, a novel non-coding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. Genome Res. 2013. https://doi.org/10.1101/gr.152942.112.
35. Zheng J, Li X-D, Wang P, Liu X-B, Xue Y-X, Hu Y, Li Z, Li Z-Q, Wang Z-H, Liu Y-H. CRNDE affects the malignant biological characteristics of human glioma stem cells by negatively regulating miR-186. Oncotarget. 2015;6:25339.
36. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.
37. Wu WK, Wang XJ, Cheng AS, Luo MX, Ng SS, To KF, Chan FK, Cho CH, Sung JJ, Yu J. Dysregulation and crosstalk of cellular signaling pathways in colon carcinogenesis. Crit Rev Oncol Hematol. 2013;86:251–77.
38. Kong YW, Ferland-McCollough D, Jackson TJ, Bushell M. microRNAs in cancer management. Lancet Oncol. 2012;13:e249–58.
39. Michael MZ, O'Connor SM, van Holst Pellekaan NG, Young GP, James RJ. Reduced accumulation of specific MicroRNAs in colorectal Neoplasia11Note: Susan M. O'Connor and Nicholas G. van Holst Pellekaan contributed equally to this work. Mol Cancer Res. 2003;1:882–91.
40. Cawley K, Logue SE, Gorman AM, Zeng Q, Patterson J, Gupta S, Samali A. Disruption of microRNA biogenesis confers resistance to ER stress-induced cell death upstream of the mitochondrion. PLoS ONE. 2013;8:e73870.
41. Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA, Sjoblom T, Barad O, Bentwich Z, Szafranska AE, Labourier E. The colorectal microRNAome. Proc Natl Acad Sci. 2006;103:3687–92.
42. Wang C-J, Zhou Z-G, Wang L, Yang L, Zhou B, Gu J, Chen H-Y, Sun X-F. Clinicopathological significance of microRNA-31,-143 and-145 expression in colorectal cancer. Dis Markers. 2009;26:27–34.
43. Schepeler T, Reinert JT, Ostenfeld MS, Christensen LL, Silahtaroglu AN, Dyrskjøt L, Wiuf C, Sørensen FJ, Kruhøffer M, Laurberg S. Diagnostic and prognostic microRNAs in stage II colon cancer. Cancer Res. 2008;68:6416–24.
44. Bandrés E, Cubedo E, Agirre X, Malumbres R, Zarate R, Ramirez N, Abajo A, Navarro A, Moreno I, Monzo M. Identification by real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. Mol Cancer. 2006;5:29.
45. Zhang H, Li Y, Huang Q, Ren X, Hu H, Sheng H, Lai M. MiR-148a promotes apoptosis by targeting Bcl-2 in colorectal cancer. Cell Death Differ. 2011;18:1702.
46. Chen Y, Song Y, Wang Z, Yue Z, Xu H, Xing C, Liu Z. Altered expression of MiR-148a and MiR-152 in gastrointestinal cancers and its clinical significance. J Gastrointest Surg. 2010;14:1170–9.
47. Siemens H, Neumann J, Jackstadt R, Mansmann U, Horst D, Kirchner T, Hermeking H. Detection of miR-34a promoter methylation in combination with elevated expression of c-Met and β-catenin predicts distant metastasis of colon cancer. Clin Cancer Res. 2012. https://doi.org/10.1158/1078-0432.CCR-12-1703.
48. Lu Y-X, Yuan L, Xue X-L, Zhou M, Liu Y, Zhang C, Li J-P, Zheng L, Hong M, Li X-N. Regulation of colorectal carcinoma stemness, growth, and metastasis by an miR-200c-Sox2–negative feedback loop mechanism. Clin Cancer Res. 2014. https://doi.org/10.1158/1078-0432.CCR-13-2348.
49. Lujambio A, Calin GA, Villanueva A, Ropero S, Sánchez-Céspedes M, Blanco D, Montuenga LM, Rossi S, Nicoloso MS, Faller WJ. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci. 2008;105:13556–61.

50. Siemens H, Jackstadt R, Hünten S, Kaller M, Menssen A, Götz U, Hermeking H. miR-34 and SNAIL form a double-negative feedback loop to regulate epithelial–mesenchymal transitions. Cell Cycle. 2011;10:4256–71.
51. Rokavec M, Öner MG, Li H, Jackstadt R, Jiang L, Lodygin D, Kaller M, Horst D, Ziegler PK, Schwitalla S. IL-6R/STAT3/miR-34a feedback loop promotes EMT-mediated colorectal cancer invasion and metastasis. J Clin Invest. 2014;124:1853–67.
52. Hahn S, Jackstadt R, Siemens H, Hünten S, Hermeking H. SNAIL and miR-34a feed-forward regulation of ZNF281/ZBP99 promotes epithelial–mesenchymal transition. EMBO J. 2013;32:3079–95.
53. Okayama H, Schetter AJ, Harris CC. MicroRNAs and inflammation in the pathogenesis and progression of colon cancer. Dig Dis. 2012;30(suppl 2):9–15.
54. Nagy ZB, Wichmann B, Kalmár A, Galamb O, Barták BK, Spisák S, Tulassay Z, Molnár B. Colorectal adenoma and carcinoma specific miRNA profiles in biopsy and their expression in plasma specimens. Clin Epigenetics. 2017;9:22 **(PMID: PMC5310023)**.
55. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12:861.
56. Catalanotto C, Cogoni C, Zardo G. MicroRNA in control of gene expression: an overview of nuclear functions. Int J Mol Sci. 2016;17:1712.
57. Valinezhad Orang A, Safaralizadeh R, Kazemzadeh-Bavili M. Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. Int J Genom. 2014. https://doi.org/10.1155/2014/970607.
58. Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. Cell. 2009;136:642–55.
59. Wu WK, Law PT, Lee CW, Cho CH, Fan D, Wu K, Yu J, Sung JJ. MicroRNA in colorectal cancer: from benchtop to bedside. Carcinogenesis. 2010;32:247–53.
60. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23:4051–60.
61. Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol. 2005;6:376.
62. Siomi H, Siomi MC. Posttranscriptional regulation of microRNA biogenesis in animals. Mol Cell. 2010;38:323–32.
63. Eamens AL, Smith NA, Curtin SJ, Wang M-B, Waterhouse PM. The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. RNA. 2009;15:2219–35.
64. van den Berg A, Mols J, Han J. RISC-target interaction: cleavage and translational suppression. Biochim Biophys Acta (BBA) Gene Regul Mech. 2008;1779:668–77.
65. Wakiyama M, Takimoto K, Ohara O, Yokoyama S. Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. Genes Dev. 2007;21:1857–62.
66. Lizarbe MA, Calle-Espinosa J, Fernández-Lizarbe E, Fernández-Lizarbe S, Robles MÁ, Olmo N, Turnay J. Colorectal cancer: from the genetic model to posttranscriptional regulation by noncoding RNAs. BioMed Res Int. 2017. https://doi.org/10.1155/2017/7354260.
67. Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM, Zhang GZ. Biological functions of microRNAs: a review. J Physiol Biochem. 2011;67:129–39.
68. Vidigal JA, Ventura A. The biological functions of miRNAs: lessons from in vivo studies. Trends Cell Biol. 2015;25:137–47.
69. MacFarlane L-A, Murphy PR. MicroRNA: biogenesis, function and role in cancer. Curr Genom. 2010;11:537–61.
70. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem. 2010;79:351–79.
71. Volinia S, Galasso M, Costinean S, Tagliavini L, Gamberoni G, Drusco A, Marchesini J, Mascellani N, Sana ME, Jarour RA. Reprogramming of miRNA networks in cancer and leukemia. Genome Res. 2010;20:589–99.
72. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136:215–33.
73. Di Leva G, Garofalo M, Croce CM. MicroRNAs in cancer. Annu Rev Pathol. 2014;9:287–314.
74. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci. 2006. https://doi.org/10.1073/pnas.0506654102.
75. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ. RAS is regulated by the let-7 microRNA family. Cell. 2005;120:635–47.
76. Weiss CN, Ito K. A macro view of microRNAs: the discovery of microRNAs and their role in hematopoiesis and hematologic disease. Int Rev Cell Mol Biol. 2017;334:99–175.
77. Paladini L, Fabris L, Bottai G, Raschioni C, Calin GA, Santarpia L. Targeting microRNAs as key modulators of tumor immune response. J Exp Clin Cancer Res. 2016;35:103.
78. Calvin Li S, T Vu L, Jianying Luo J, F Zhong J, Li Z, A Dethlefs B, G Loudon W, H Kabeer M. Tissue elasticity bridges cancer stem cells to the tumor microenvironment through micrornas: implications for a "watch-and-wait" approach to cancer. Curr Stem Cell Res Ther. 2017;12:455–70.
79. Lages E, Ipas H, Guttin A, Nesr H, Berger F, Issartel J-P. MicroRNAs: molecular features and role in cancer. Front Biosci. 2012;17:2508.
80. Gambari R, Brognara E, Spandidos DA, Fabbri E. Targeting oncomiRNAs and mimicking tumor suppressor miRNAs: new trends in the development of miRNA therapeutic strategies in oncology. Int J Oncol. 2016;49:5–32.
81. Fadaka A, Ajiboye B, Ojo O, Adewale O, Olayide I, Emuowhochere R. Biology of glucose metabolization in cancer cells. J Oncol Sci. 2017;3:45–51.
82. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA. MicroRNA expression profiles classify human cancers. Nature. 2005;435:834.
83. Motoyama K, Inoue H, Takatsuno Y, Tanaka F, Mimori K, Uetake H, Sugihara K, Mori M. Over-and under-expressed microRNAs in human colorectal cancer. Int J Oncol. 2009;34:1069–75.
84. Rossi S, Kopetz S, Davuluri R, Hamilton SR, Calin GA. MicroRNAs, ultraconserved genes and colorectal cancers. Int J Biochem Cell Biol. 2010;42:1291–7.
85. Slaby O, Svoboda M, Michalek J, Vyzula R. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. Mol Cancer. 2009;8:102.
86. Gao P, Tchernyshyov I, Chang T-C, Lee Y-S, Kita K, Ochi T, Zeller KI, De Marzo AM, Van Eyk JE, Mendell JT. c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature. 2009;458:762.
87. Yoshida GJ. Emerging roles of Myc in stem cell biology and novel tumor therapies. J Exp Clin Cancer Res. 2018;37:173.
88. Yoshida GJ. Metabolic reprogramming: the emerging concept and associated therapeutic strategies. J Exp Clin Cancer Res. 2015;34:1.
89. Calin GA, Croce CM. MicroRNA-cancer connection: the beginning of a new tale. Cancer Res. 2006;66:7390–4.
90. Xishan Z, Ziying L, Jing D, Gang L. MicroRNA-320a acts as a tumor suppressor by targeting BCR/ABL oncogene in chronic myeloid leukemia. Sci Rep. 2015;5:12460.
91. Peng Y, Croce CM. The role of MicroRNAs in human cancer. Signal Transduct Target Ther. 2016;1:15004.
92. Carden T, Singh B, Mooga V, Bajpai P, Singh KK. Epigenetic modification of miR-663 controls mitochondria-to-nucleus retrograde signaling and tumor progression. J Biol Chem. 2017;292:20694–706.
93. Zhang Y, Xu X, Zhang M, Wang X, Bai X, Li H, Kan L, Zhou Y, Niu H, He P. MicroRNA-663a is downregulated in non-small cell lung cancer and inhibits proliferation and invasion by targeting JunD. BMC Cancer. 2016;16:315.
94. Bader AG. miR-34—a microRNA replacement therapy is headed to the clinic. Front Genet. 2012;3:120.
95. Wiggins JF, Ruffino L, Kelnar K, Omotola M, Patrawala L, Brown D, Bader AG. Development of a lung cancer therapeutic based on the tumor suppressor microRNA-34. Cancer Res. 2010;70:5923–30.
96. Slabáková E, Culig Z, Remšík J, Souček K. Alternative mechanisms of miR-34a regulation in cancer. Cell Death Dis. 2017;8:e3100.
97. Xia X, Yang B, Zhai X, Liu X, Shen K, Wu Z, Cai J. Prognostic role of microRNA-21 in colorectal cancer: a meta-analysis. PLoS ONE. 2013;8:e80426.
98. Kjaer-Frifeldt S, Hansen T, Nielsen B, Joergensen S, Lindebjerg J, Soerensen F, dePont Christensen R, Jakobsen A. The prognostic importance of miR-21 in stage II colon cancer: a population-based study. Br J Cancer. 2012;107:1169.
99. Kang WK, Lee JK, Oh ST, Lee SH, Jung CK. Stromal expression of miR-21 in T3-4a colorectal cancer is an independent predictor of early tumor relapse. BMC Gastroenterol. 2015;15:2.
100. Zhang J-X, Song W, Chen Z-H, Wei J-H, Liao Y-J, Lei J, Hu M, Chen G-Z, Liao B, Lu J. Prognostic and predictive value of a microRNA signature

in stage II colon cancer: a microRNA expression analysis. Lancet Oncol. 2013;14:1295–306.
101. Nielsen BS, Jørgensen S, Fog JU, Søkilde R, Christensen IJ, Hansen U, Brünner N, Baker A, Møller S, Nielsen HJ. High levels of microRNA-21 in the stroma of colorectal cancers predict short disease-free survival in stage II colon cancer patients. Clin Exp Metastasis. 2011;28:27–38.
102. Lee KS, Nam SK, Koh J, Kim D-W, Kang S-B, Choe G, Kim WH, Lee HS. Stromal expression of microRNA-21 in advanced colorectal cancer patients with distant metastases. J Pathol Transl Med. 2016;50:270.
103. Zuo Q, Cao L, Yu T, Gong L, Wang L, Zhao Y, Xiao B, Zou Q. MicroRNA-22 inhibits tumor growth and metastasis in gastric cancer by directly targeting MMP14 and Snail. Cell Death Dis. 2015;6:e2000.
104. Song SJ, Pandolfi PP. miR-22 in tumorigenesis. Cell Cycle. 2014;13:11–2.
105. Yang F, Hu Y, Liu H-X, Wan Y-JY. MiR-22-silenced cyclin A expression in colon and liver cancer cells is regulated by bile acid receptor. J Biol Chem. 2015;290:6507–15.
106. Wan W-N, Zhang Y-Q, Wang X-M, Liu Y-J, Zhang Y-X, Que Y-H, Zhao W-J, Li P. Down-regulated miR-22 as predictive biomarkers for prognosis of epithelial ovarian cancer. Diagn Pathol. 2014;9:178.
107. James AB, Fadaka AO, Magbagbeola OA, Oturu A, Kolawole OO, Ogunjimi A, Oshodi T, Habeebu M, Onawoga FO, Ajogbeje EO. Haematological and miRNAs (let-7g, miR-21, miR-141) expression modulation profile in serum samples of Human prostate cancer. FASEB J. 2017;31:757.12.
108. Li P, Xu Q, Zhang D, Li X, Han L, Lei J, Duan W, Ma Q, Wu Z, Wang Z. Upregulated miR-106a plays an oncogenic role in pancreatic cancer. FEBS Lett. 2014;588:705–12.
109. Yang G, Zhang R, Chen X, Mu Y, Ai J, Shi C, Liu Y, Shi C, Sun L, Rainov NG. MiR-106a inhibits glioma cell growth by targeting E2F1 independent of p53 status. J Mol Med. 1037;2011:89.
110. Wang Z, Liu M, Zhu H, Zhang W, He S, Hu C, Quan L, Bai J, Xu N. miR-106a is frequently upregulated in gastric cancer and inhibits the extrinsic apoptotic pathway by targeting FAS. Mol Carcinog. 2013;52:634–46.
111. Syed DN, Suh Y, Afaq F, Mukhtar H. Dietary agents for chemoprevention of prostate cancer. Cancer Lett. 2008;265:167–76.
112. Tili E, Michaille J-J, Alder H, Volinia S, Delmas D, Latruffe N, Croce CM. Resveratrol modulates the levels of microRNAs targeting genes encoding tumor-suppressors and effectors of TGFβ signaling pathway in SW480 cells. Biochem Pharmacol. 2010;80:2057–65.
113. Yoo YA, Kang MH, Kim JS, Oh SC. Sonic hedgehog signaling promotes motility and invasiveness of gastric cancer cells through TGF-β-mediated activation of the ALK5-Smad 3 pathway. Carcinogenesis. 2008;29:480–90.
114. Tili E, Michaille J-J, Adair B, Alder H, Limagne E, Taccioli C, Ferracin M, Delmas D, Latruffe N, Croce CM. Resveratrol decreases the levels of miR-155 by upregulating miR-663, a microRNA targeting JunB and JunD. Carcinogenesis. 2010;31:1561–6.
115. Kumazaki M, Noguchi S, Yasui Y, Iwasaki J, Shinohara H, Yamada N, Akao Y. Anti-cancer effects of naturally occurring compounds through modulation of signal transduction and miRNA expression in human colon cancer cells. J Nutr Biochem. 2013;24:1849–58.
116. Gonçalves P, Martel F. Butyrate and colorectal cancer: the role of butyrate transport. Curr Drug Metab. 2013;14:994–1008.
117. Sengupta S, Muir JG, Gibson PR. Does butyrate protect from colorectal cancer? J Gastroenterol Hepatol. 2006;21:209–18.
118. Fung KY, Cosgrove L, Lockett T, Head R, Topping DL. A review of the potential mechanisms for the lowering of colorectal oncogenesis by butyrate. Br J Nutr. 2012;108:820–31.
119. Hu S, Dong TS, Dalal SR, Wu F, Bissonnette M, Kwon JH, Chang EB. The microbe-derived short chain fatty acid butyrate targets miRNA-dependent p21 gene expression in human colon cancer. PLoS ONE. 2011;6:e16221.
120. Hu S, Liu L, Chang EB, Wang J-Y, Raufman J-P. Butyrate inhibits pro-proliferative miR-92a by diminishing c-Myc-induced miR-17-92a cluster transcription in human colon cancer cells. Mol Cancer. 2015;14:180.
121. Humphreys KJ, Cobiac L, Le Leu RK, Van der Hoek MB, Michael MZ. Histone deacetylase inhibition in colorectal cancer cells reveals competing roles for members of the oncogenic miR-17-92 cluster. Mol Carcinog. 2013;52:459–74.
122. Williams CD, Whitley BM, Hoyo C, Grant DJ, Iraggi JD, Newman KA, Gerber L, Taylor LA, McKeever MG, Freedland SJ. A high ratio of dietary n-6/n-3 polyunsaturated fatty acids is associated with increased risk of prostate cancer. Nutr Res. 2011;31:1–8.
123. Gopinath B, Buyken AE, Flood VM, Empson M, Rochtchina E, Mitchell P. Consumption of polyunsaturated fatty acids, fish, and nuts and risk of inflammatory disease mortality. Am J Clin Nutr. 2011;93:1073–9.
124. Davidson LA, Wang N, Shah MS, Lupton JR, Ivanov I, Chapkin RS. n-3 Polyunsaturated fatty acids modulate carcinogen-directed non-coding microRNA signatures in rat colon. Carcinogenesis. 2009;30:2077–84.
125. Mandal CC, Ghosh-Choudhury T, Dey N, Choudhury GG, Ghosh-Choudhury N. miR-21 is targeted by omega-3 polyunsaturated fatty acid to regulate breast tumor CSF-1 expression. Carcinogenesis. 2012;33:1897–908.
126. Mudduluru G, George-William JN, Muppala S, Asangani IA, Kumarswamy R, Nelson LD, Allgayer H. Curcumin regulates miR-21 expression and inhibits invasion and metastasis in colorectal cancer. Biosci Rep. 2011;31:185–97.
127. Bao B, Ali S, Kong D, Sarkar SH, Wang Z, Banerjee S, Aboukameel A, Padhye S, Philip PA, Sarkar FH. Anti-tumor activity of a novel compound-CDF is mediated by regulating miR-21, miR-200, and PTEN in pancreatic cancer. PLoS ONE. 2011;6:e17850.
128. Sun M, Estrov Z, Ji Y, Coombes KR, Harris DH, Kurzrock R. Curcumin (diferuloylmethane) alters the expression profiles of microRNAs in human pancreatic cancer cells. Mol Cancer Ther. 2008;7:464–73.
129. Ali S, Ahmad A, Banerjee S, Padhye S, Dominiak K, Schaffert JM, Wang Z, Philip PA, Sarkar FH. Gemcitabine sensitivity can be induced in pancreatic cancer cells through modulation of miR-200 and miR-21 expression by curcumin or its analogue CDF. Cancer Res. 2010;70:3606–17.
130. Zhang J, Zhang T, Ti X, Shi J, Wu C, Ren X, Yin H. Curcumin promotes apoptosis in A549/DDP multidrug-resistant human lung adenocarcinoma cells through an miRNA signaling pathway. Biochem Biophys Res Commun. 2010;399:1–6.
131. Parasramka MA, Ho E, Williams DE, Dashwood RH. MicroRNAs, diet, and cancer: new mechanistic insights on the epigenetic actions of phytochemicals. Mol Carcinog. 2012;51:213–30.
132. Saini S, Majid S, Dahiya R. Diet, microRNAs and prostate cancer. Pharm Res. 2010;27:1014–26.
133. Gandini S, Boniol M, Haukka J, Byrnes G, Cox B, Sneyd MJ, Mullie P, Autier P. Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. Int J Cancer. 2011;128:1414–24.
134. Alvarez-Díaz S, Valle N, Ferrer-Mayorga G, Lombardía L, Herrera M, Domínguez O, Segura MF, Bonilla F, Hernando E, Muñoz A. MicroRNA-22 is induced by vitamin D and contributes to its antiproliferative, antimigratory and gene regulatory effects in colon cancer cells. Hum Mol Genet. 2012;21:2157–65.
135. Padi SK, Zhang Q, Rustum YM, Morrison C, Guo B. MicroRNA-627 mediates the epigenetic mechanisms of vitamin D to suppress proliferation of human colorectal cancer cells and growth of xenograft tumors in mice. Gastroenterology. 2013;145:437–46.
136. Tinggi U. Selenium: its role as antioxidant in human health. Environ Health Prev Med. 2008;13:102–8.
137. Sarveswaran S, Liroff J, Zhou Z, Nikitin AY, Ghosh J. Selenite triggers rapid transcriptional activation of p53, and p53-mediated apoptosis in prostate cancer cells: implication for the treatment of early-stage prostate cancer. Int J Oncol. 2010;36:1419–28.
138. Nian H, Bisson WH, Dashwood W-M, Pinto JT, Dashwood RH. α-Keto acid metabolites of organoselenium compounds inhibit histone deacetylase activity in human colon cancer cells. Carcinogenesis. 2009;30:1416–23.
139. Li Y, VandenBoom TG, Wang Z, Kong D, Ali S, Philip PA, Sarkar FH. miR-146a suppresses invasion of pancreatic cancer cells. Cancer Res. 2010;70:1486–95.
140. Majid S, Dar AA, Saini S, Chen Y, Shahryari V, Liu J, Zaman MS, Hirata H, Yamamura S, Ueno K. Regulation of minichromosome maintenance gene family by microRNA-1296 and genistein in prostate cancer. Cancer Res. 2010;70:2809–18.
141. Parker L, Taylor D, Kesterson J, Metzinger D, Gercel-Taylor C. Modulation of microRNA associated with ovarian cancer cells by genistein. Eur J Gynaecol Oncol. 2009;30:616–21.
142. Edvardsson K, Nguyen-Vu T, Kalasekar SM, Pontén F, Gustafsson J-Å, Williams C. Estrogen receptor β expression induces changes in

the microRNA pool in human colon cancer cells. Carcinogenesis. 2013;34:1431–41.
143. Okuda T, Yoshida T, Hatano T. Ellagitannins as active constituents of medicinal plants. Planta Med. 1989;55:117–22.
144. Wen XY, Wu SY, Li ZQ, Liu ZQ, Zhang JJ, Wang GF, Jiang ZH, Wu SG. Ellagitannin (BJA3121), an anti-proliferative natural polyphenol compound, can regulate the expression of MiRNAs in HepG2 cancer cells. Phytother Res. 2009;23:778–84.
145. Mercken EM, Carboneau BA, Krzysik-Walker SM, de Cabo R. Of mice and men: the benefits of caloric restriction, exercise, and mimetics. Ageing Res Rev. 2012;11:390–8.
146. Steinbach G, Heymsfield S, Olansen NE, Tighe A, Holt PR. Effect of caloric restriction on colonic proliferation in obese persons: implications for colon cancer prevention. Cancer Res. 1994;54:1194–7.
147. Kritchevsky D. Colorectal cancer: the role of dietary fat and caloric restriction. Mutat Res Fundam Mol Mech Mutagen. 1993;290:63–70.
148. Fontana L, Mitchell SE, Wang B, Tosti V, van Vliet T, Veronese N, Bertozzi B, Early DS, Maissan P, Speakman JR. The effects of graded caloric restriction: XII. Comparison of mouse to human impact on cellular senescence in the colon. Aging Cell. 2018;17:e12746.
149. Hayflick L. The limited in vitro lifetime of human diploid cell strains. Exp Cell Res. 1965;37:614–36.
150. Mooi W, Peeper D. Oncogene-induced cell senescence—halting on the road to cancer. N Engl J Med. 2006;355:1037–46.
151. Ghosh K, Capell BC. The senescence-associated secretory phenotype: critical effector in skin cancer and aging. J Invest Dermatol. 2016;136:2133–9.
152. Rao SG, Jackson JG. SASP: tumor suppressor or promoter? Yes! Trends Cancer. 2016;2:676–87.
153. Vaughan S, Jat PS. Deciphering the role of nuclear factor-κB in cellular senescence. Aging (Albany NY). 2011;3:913.
154. Karin M. Nuclear factor-κB in cancer development and progression. Nature. 2006;441:431.
155. Hursting SD, Dunlap SM, Ford NA, Hursting MJ, Lashinger LM. Calorie restriction and cancer prevention: a mechanistic perspective. Cancer Metab. 2013;1:10.
156. Stattin P, Palmqvist R, Söderberg S, Biessy C, Ardnor B, Hallmans G, Kaaks R, Olsson T. Plasma leptin and colorectal cancer risk: a prospective study in Northern Sweden. Oncol Rep. 2003;10:2015–21.
157. Harvey AE, Lashinger LM, Otto G, Nunez NP, Hursting SD. Decreased systemic IGF-1 in response to calorie restriction modulates murine tumor cell growth, nuclear factor-κB activation, and inflammation-related gene expression. Mol Carcinog. 2013;52:997–1006.
158. Harvey AE, Lashinger LM, Hays D, Harrison LM, Lewis K, Fischer SM, Hursting SD. Calorie restriction decreases murine and human pancreatic tumor cell growth, nuclear factor-κB activation, and inflammation-related gene expression in an insulin-like growth factor-1—dependent manner. PLoS ONE. 2014;9:e94151.
159. Jin L, Lim M, Zhao S, Sano Y, Simone BA, Savage JE, Wickstrom E, Camphausen K, Pestell RG, Simone NL. The metastatic potential of triple-negative breast cancer is decreased via caloric restriction-mediated reduction of the miR-17 ~ 92 cluster. Breast Cancer Res Treat. 2014;146:41–50.
160. Devlin KL, Sanford T, Harrison LM, LeBourgeois P, Lashinger LM, Mambo E, Hursting SD. Stage-specific MicroRNAs and their role in the anticancer effects of calorie restriction in a rat model of ER-positive luminal breast cancer. PLoS ONE. 2016;11:e0159686.

CHAPTER 3.0

BIOMARKERS FOR STRATIFICATION IN COLORECTAL CANCER: MICRORNAS

ADEWALE OLUWASEUN FADAKA*, ASHLEY PRETORIUS, AND ASHWIL KLEIN

Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa.

Accepted/ published: *Cancer Control* (2019) **26**(1)
DOI: 10.1177/1073274819862784 **(Impact factor: 1.990)**

Statement of Contributions of Joint Authorship

Adewale Fadaka (PhD candidate)

Conceived the concept of the manuscript as well as the write up and compilation. Preparation of figures and tables.

Ashley Pretorius (Co-Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

Ashwil Klein (Principal Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

This Chapter is an exact copy of the journal paper below.

Review

Biomarkers for Stratification in Colorectal Cancer: MicroRNAs

Cancer Control
Volume 26: 1-11

DOI: 10.1177/1073274819862784
journals.sagepub.com/home/ccx

Adewale Oluwaseun Fadaka, PhD[1], Ashley Pretorius, PhD, MBBS[1], and Ashwil Klein, PhD[1]

Abstract

Colorectal cancer (CRC) is one of the most widely recognized and deadly malignancies worldwide. In spite of the fact that the death rates have declined over the previous decade, particularly because of enhanced screening or potential treatment alternatives, CRC still remains the third leading cause of cancer-related mortality in the world, with an estimated incidence of over 1 million new cases and approximately 600 000 deaths estimated yearly. Unlike prostate and lung cancer, CRC is not easily detectable in its early stage, which may also account for its high mortality rate. MicroRNAs (miRNAs) are a class of noncoding RNAs. The roles of these noncoding RNAs have been implicated in cancer pathogenesis, most especially CRC, due to their ability to posttranscriptionally regulate the expression of oncogenes and tumor suppressor genes. Dysregulated expression of many miRNAs regulates the expression of hundreds of growth regulatory genes and pathways that are important in the multistep model of colorectal carcinogenesis. If CRC is detected early, it is a largely treatable disease. Early diagnosis, including the identification of premalignant adenomas, is regarded a major concept for improving patient survival in CRC treatment. Several lines of research suggest that miRNAs are closely implicated in the metastatic process in CRC and some of these miRNAs could be useful as promising clinical tools for identifying specific stages of CRC due to their differential expression. This review discusses the correlation between CRC staging relative to the specific expression of miRNA for early detection, treatment, and disease management.

Keywords

microRNA, prognosis, colorectal cancer, diagnosis, biomarker, staging

Received January 11, 2019. Received revised June 8, 2019. Accepted for publication June 13, 2019.

Introduction

Colorectal cancer (CRC) is the most common malignancy in the gastrointestinal tract/bowel or large intestine, the third most commonly diagnosed cancer, and also the third cause of cancer-related demise worldwide.[1] More often than not, CRC is thought of as a typical disease affecting old individuals, with most cases analyzed amid the fifth and sixth decades and a higher predominance among men.[2] It is a multifactorial disease process, with etiology encompassing genetic factors, environmental exposures (including diet), and inflammatory conditions of the digestive tract. Colorectal cancer develops through a gradual accumulation of genetic and epigenetic changes, resulting in the transformation of normal colonic mucosa into invasive cancer.[3] Over 90% of colorectal carcinomas are adenocarcinomas (adenoma–carcinoma sequence) arising from epithelial cells of the colorectal mucosa,[4] and the neoplastic transformation time is considered to be 10 to 15 years, which represents the available time to detect and remove these adenomas before their progression.[3] Based on the differentiation of colorectal adenocarcinoma specified by a group of gland forming cells, colorectal

[1] Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa

Corresponding Author:
Adewale Oluwaseun Fadaka, Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa.
Emails: silvernonferous@gmail.com; 3782900@myuwc.ac.za

carcinomas can be divided into well, moderately, and poorly differentiated adenomas with varying gland formation. Over 95%, 50% to 95%, and <70%, respectively, of these adenocarcinomas are gland forming and are the basis for CRC diagnosis through histological grading. Also, approximately 70% of the diagnosed CRC are moderately differentiated, while others such as poor and well-differentiated CRCs are reportedly 20% and 10%, respectively. Some of the CRCs may also be undifferentiated.[5] The epidemiology of CRC can be categorized into modifiable risk factors, which include age; family history of familial adenomatous polyposis (FAP), Lynch syndrome, and inflammatory bowel diseases; and nonmodifiable risk factors (red and processed meat consumption, obesity, alcohol, and smoking). The larger part of CRCs is sporadic (70%-80%), with age being the most critical risk factor. Other inherited forms of this disease are FAP (less than 1%), nonpolyposis hereditary CRC or Lynch syndrome (2%-5%), or *MYH* gene–associated polyposis (<1%), which constitute a small proportion of reported cases.[6] Moreover, cases associated with hereditary components have been estimated to be 20% to 25% and are termed familial CRC.[7] The ideal technique to precisely identify CRC and recurrence at the most punctual conceivable time is an exceedingly debatable concept in research. It is well known that most recurrences occur within 5 years.[8] Although researchers have provided improved CRC diagnosis, good treatment option, and a suitable way to predict recurrence and/or prognosis in CRC in recent time, the proper staging of CRC noninvasively for effective diagnosis can also be a good lead to its management, thereby increasing the overall survival of patients suffering from this cancer subtype.[9,10] The involvement of short oligonucleotide noncoding ribonucleic acid as biomarkers with specific attributes that are distinct for human processes are proven indicators for improved diagnosis and treatment intervention for CRC.[11-14]

MicroRNAs (miRNAs) are small, 18 to 25 noncoding nucleotide sequences of RNA. These sequences control the expression of several target genes at the same time either by translational repression or degradation of the messenger RNA (mRNA) transcript after targeting the 3′-Untranslaterd region (3′UTR).[15] Many major cellular functions such as development, differentiation, growth, and metabolism are regulated by these miRNAs.[16] Therefore, they play a central role in research and clinical settings as potential valuable biomarkers and novel therapeutics for cancer.[17-19] A single miRNA has been reported to regulate up to several hundred mRNAs simultaneously and affects a number of target transcripts. As of 2010, approximately 2200 miRNA genes were suggested to exist in the mammalian genome[16] and one-third of the human genome is estimated to be regulated by miRNAs.[20] Knowing the expression, distribution, and longevity of these noncoding class of RNA in tissues is essential for the understanding of both physiological and pathological mechanisms. In addition, determination of the tissues that express specific miRNAs and their stages will help to develop a miRNA in biological samples into a biomarker for a specific disease. Recently, miRNA expression in multiple human tissues has been provided in an atlas (https://ccb-web.cs.uni-saarland.de/tissueatlas) for the elucidation of the role of miRNAs in tissue development and tissue-specific diseases such as CRC and has reported that these miRNAs have the half-life of about 1 to 14 days at 4°C.[21] MicroRNAs have been found in an assortment of body liquids, where they are astoundingly stable.[22-24] Extracellular miRNAs could serve as diagnostic biomarkers relevant to both prevention and treatment of human cancer. Notwithstanding, broad research is fundamental for distinguishing the attributes of extracellular miRNAs to portray their roles in tumorigenesis and prevention.[22] Accordingly, there may be a poor correlation between cellular and extracellular miRNAs and between miRNAs detectable in various biological fluids.[25,26] The tumorigenesis of CRC involves multistep genomic changes, including the activation of oncogenes and inactivation of tumor suppressor genes. Numerous miRNAs have been reported to play a role in cancer development, such as carcinogenesis, progression, and metastasis.[27,28] Only a couple of studies have explored circulating miRNAs in patients with CRC.[15,29] Also, there is extremely limited research on the identification of commonly and differentially expressed miRNA for CRC staging if there is any. Efforts to depict clinical, pathological, and molecular features in patients have reached disputable ends with respect to tumor grade and disease stage at diagnosis.[30] Also, it is generally acknowledged that diagnosis in patients is always difficult because of the vulnerability of both patient and the specialist to the presenting symptoms, leading to a frequent unfavorable outcome of the disease. If specific miRNAs are expressed in a certain stage of CRC, then early detection of this disease will be largely treatable. The review aims to discuss the staging of CRC with respect to specific miRNAs for early detection, treatment, efficacy, and effective management of the disease.

Molecular Pathogenesis of CRC

Suppressor pathway or pathway of chromosomal instability (CIN) was first proposed as the mechanism of colorectal carcinogenesis.[31] The accumulation of mutations leads to oncogene activation such as Kirsten rat sarcoma (KRAS) and inactivation of tumor suppressor genes such as Deleted in Colorectal Cancer (*DCC*), Total Protein-53 (*TP-53*), SMAD family member 4, Mothers against decapentaplegic homolog 4 (*SMAD4*), and Adenomatous polyposis coli (*APC*).[32] Regardless of the order of this molecular alteration, their accumulation is responsible for neoplastic transformation.[33] Mutations in the genes *MSH2*, *MSH3*, *MSH6*, *Exo1*, *PMS1*, *PSM2*, *MLH1*, and *MLH3* responsible for DNA repair during replication are associated with the second mechanism of colorectal carcinogenesis. These mismatch repair (*MMR*) genes play a crucial role in the identification and repair of errors after replication in order to prepare them for cell division. Accumulation of errors in repetitive DNA fragments causes mutations in target genes.[34] Approximately 20% of sporadic CRC and Lynch syndrome are reportedly caused by mutations in mismatch DNA repair genes, that is, defective DNA MMR system (microsatellite instability).[32,35] The last pathway of aberrant hypermethylation was identified as a mechanism of gene

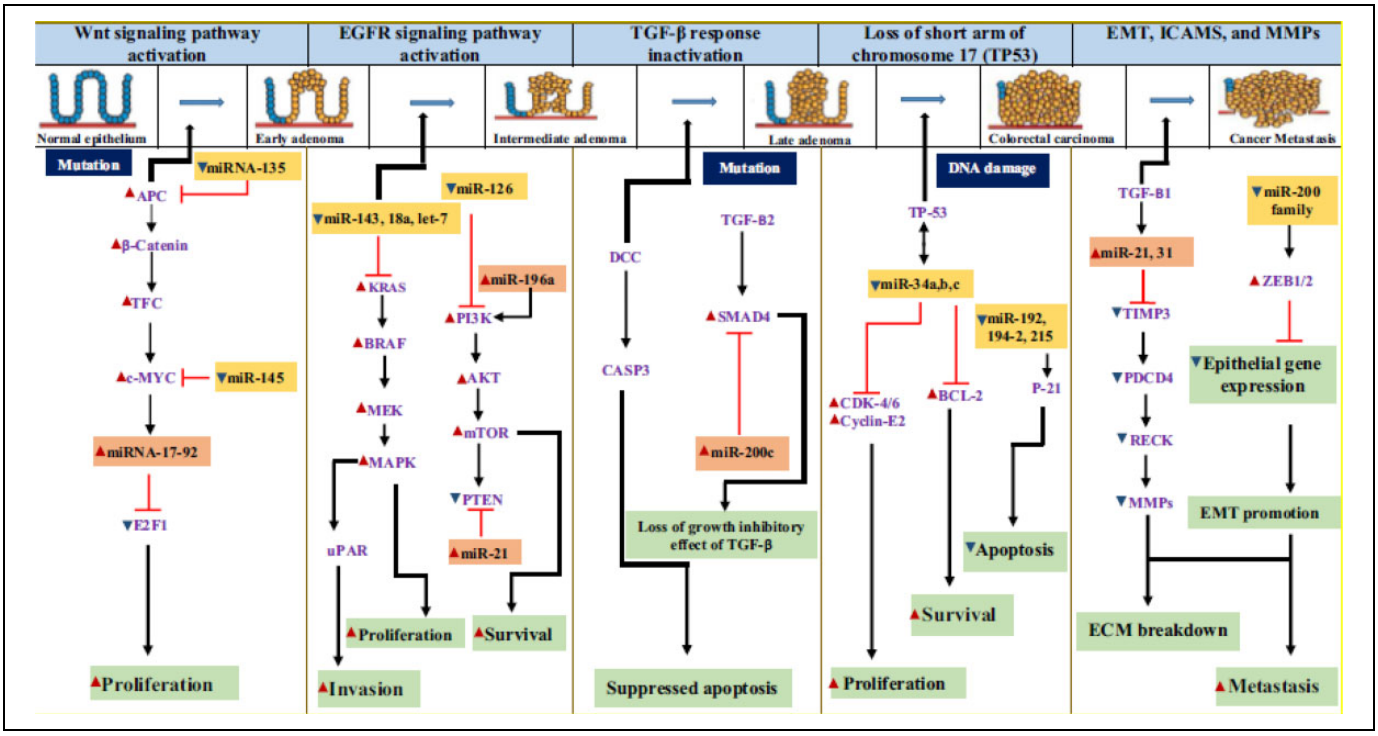


Figure 1. Involvement of microRNA (miRNA) in colorectal cancer.[41] Red arrows—upregulation; blue arrows—downregulation. Experimentally validated miRNAs are shown alongside with their target genes in altered expression in CRC. CRC indicates colorectal cancer.

function silencing in the field of epigenetics.[36] The CpG island methylator, also known as CIMP, is referred to as dinucleotide methylation, which occurs in the transcription start site upstream of many genes. They are attributed to 15% to 20% of sporadic CRC.[37] The hypermethylation of the promoter region of any gene is mainly carried on by the CpG island methylator phenotype. The positive tumor of this methylator methylates certain marker genes. Examples of these genes are the calcium voltage-gated channel subunit α1 G, the protein-coding gene, suppressor of cytokine signaling-1, Runt-related transcription factor-3, the induction of neuronal differentiation by the overexpression of NEUROG-1, and finally the insulin-like growth factor 2.[38] More than 2 of these genes are targeted and methylated by CIMP. Histologically, the differentiation of these tumors is poorly defined. They also exhibit microsatellite instability and are known to be B-RAF mutation carriers.[39] The precursor lesions of the methylator tumors are the sessile serrated adenomas.[40] A superior comprehension of carcinogenesis pathways has allowed the improvement of diagnostic and prognostic biomarkers and furthermore the examination of new remedial targets and predictors of CRC treatment response.

MicroRNA

In our previous review, the overview of miRNA together with the synthesis, general functions, metabolic reprogramming, and their specific expression were discussed extensively.[41] Furthermore, the mechanism of action underlying the initiation, progression, and metastasis of CRC with respect to miRNAs was also examined (Figure 1). Emerging evidence suggests the promising potential of these miRNAs as potential noninvasive biomarkers for CRC screening.[42-48]

Biomarkers

Recently, interest has focused on the search for biomarkers in CRC. Tremendous research on CRC has revealed the 3 major pathways for carcinogenesis: chromosomal abnormalities, microsatellite instability pathway, and methylation pathway described by the epigenetic methylation of a large number of genes. Of the molecules associated with prognosis implicated in CIN pathway, only the epidermal growth factor receptor (EGFR) pathway as a biomarker is used for diagnosis due to its clinical relevance. This is because of the complexity and redundancy of several pathways occurring in cellular processes, as well as the lack of therapies that can effectively target various biomarkers.[49] Epidermal growth factor receptor pathway has also been reported as the main target for the treatment of a specific type of CRC.[50] Also, mutations observed in the pathways of the RAS family and the abnormal activation of the EGFR occur in a number of CRC cases.

The microsatellite instability status was also confirmed as the primary biomarker for stratification of stage II CRC. The CIMP pathway as reported is associated with a group of clinical and histological features involved with approximately 15% to 20% of CRC with MMR gene *MLH*.[51] The precursor lesions in CIMP cancers are serrated polyps, not adenomatous lesions,

with the underlying genetic changes frequently occurring in the *BRAF* oncogene.[52] Mutation in the Raf family B-Raf (B-Raf proto-oncogene, serine/threonine kinase) has been observed in the transformation of normal tissue layer or membrane into abnormal cell multiplication, such as crypt foci or sessile serrated polyps. The frequency of mutation in BRAF varies among human cancer, ranging from about 80% in skin cancer to around 0% to 18% among other cancers.[53] About 1% to 3% and 5% were reported for lung and CRC, respectively. In nearly almost all the cases of BRAF mutation reported, thymine (T) is substituted with adenine (A) at the position 1799 nucleotide, thereby changing the amino acid valine (Val) to glutamic acid (Glu) at codon 600. This segment of activation has also been reported in various cancer types, including CRC.[54-62] The methylation of *BRAF* gene promoter region causes loss of p16, leading to the cell progression to advanced polyps.[63] Increase in activity also prompts the methylation of *MutL homolog 1* gene, silencing transcription. Loss of function of this gene results in MMR deficiency and subsequently the high microsatellite instability in CRC phenotype.[64,65]

Moreover, aggregating evidence confirmed that cancer cells release some miRNAs into systemic circulation.[66-68] This unique feature of miRNAs is one of the focal reasons behind the ongoing exploration and explosion of miRNA biomarker studies in cancer research. There are various types of biomarkers depending on their functions. Examples include diagnostic biomarkers (to identify/monitor or detect the type of tumor and/or reoccurrence, eg, carcinoembryonic antigen [CEA]), predictive biomarkers (to predict the efficacy or response to different treatments or therapeutic intervention), and prognostic biomarker (to indicate the progress of disease and to estimate the risk of disease recurrence, ie, estimation of survival outcome and treatment strategy).[69]

MicroRNAs have emanated as tumor-related biomarkers that reflect not only the existence of early-stage tumors but also the dynamics and status of advanced stage tumors, tumor recurrence, and drug sensitivities.[66] Cancer-associated miRNAs are present in blood in a very stable and detectable form that is protected from endogenous ribonuclease activities and other conditions. Previous studies have demonstrated the ease of quantification of these circulating miRNAs using various methods.[70-73] Several miRNA expressions have been implicated in various categories of a biomarker for the detection of tumor. High expression levels of miR-92a, miR-141, let-7a, miR-1229, miR-1246, miR-150, miR-21, miR-223, miR-23a, and miR-378 have been analyzed to be associated with diagnostic biomarkers, while high expression levels of miR-141, miR-320, miR-596, and miR-203 are majorly for prognosis, malignant potential, and tumor recurrence. High expression levels of miR-106a, miR-484, and miR-130b are associated with predictive biomarkers, and low expression levels of miR-106a, miR-484, and miR-130b are prognostic in nature.[74]

MiRNAs as CRC Diagnostic Tools

In CRC, abnormally expressed miRNAs disrupt cellular signal transduction and cell survival pathways, for example, Wnt signaling pathway, EGFR, and p53, connecting miRNA to known events in the pathway of cancer transformation.[75] Accumulating evidence suggests that miRNAs may also have intense clinical applications. MicroRNA expression profiles have the ability to discriminate tumors from different cancer subtypes.[76] Also, the expression of individual miRNAs may be used to predict patient survival, tumor stage, the presence of lymph node metastases, and the response to therapy in CRC.[75,77,78] Studies investigated the differential expression of a panel of 95 miRNAs and also demonstrated that miR-92 was significantly elevated in the plasma of patients with CRC and that it has potential as a noninvasive molecular biomarker for CRC screening with high sensitivity and specificity.[79] These researchers also showed that the discovery of miR-92a may differentiate CRC from other gastrointestinal cancers and inflammatory bowel diseases. Cheng et al[80] proposed that plasma miR-141 may represent a novel biomarker that complements CEA in detecting CRC with distant metastasis and that high levels of miR-141 in plasma were associated with poor prognosis. Furthermore, 7 miRNAs (let-7a, miR-1229, miR-1246, miR-150, miR-21, miR-223, and miR-23a) were validated using quantitative real-time polymerase chain reaction (RT-PCR). These miRNAs were confirmed to be suitable biomarkers to detect CRC. They also possess high sensitivity and specificity.[81] Another study also discovered miR-378 in biological fluid as a screening biomarker that can discriminate patients with CRC from a healthy individual.[82]

Table 1. Correlation of CRC TNM Stages With Prognosis (Modified From Cancer Therapy Advisor).[a]

Stages	TNM	5-Year Survival (%)
0, I	Tis, T_1, N_0, M_0	>90
I	T_2, N_0, M_0	80-85
II	T_{3-4}, N_0, M_0	70-75
III	T_2, N_{1-3}, M_0	70-75
III	T_3, N_{1-3}, M_0	50-65
III	T_4, N_{1-2}, M_0	25-45
IV	M_1	<3

Abbreviations: CRC, colorectal cancer; TNM, tumor–node–metastasis.
[a]T_{IS} indicates carcinoma in situ intraepithelia or invasion of lamina propria; T_1, tumor invasion of submucosa; T_2, tumor invasion of muscularis propria; T_3, tumor invasion through the muscularis propria into pericolorectal tissues; T_4, penetration of tumor through the surface of the visceral peritoneum and further directly invading other organs.

Colorectal Cancer Prognosis

Early detection of distant metastasis and selective criteria regarding which individuals would benefit most from invasive treatments is essential for improving long-term survival. The most important predictor of outcome is the stage of disease at diagnosis (Table 1). Therapeutic prognosis is an evaluative segment of medicine and research that includes the science of estimating the intricacy and recurrence of CRC and an anticipated survival of patients.[83] A substantial number of variables, including tumor grade, tumor size and staging, and lymph node status

including different viewpoints, may impact, influence, or correlate with prognosis for patients with CRC. Therefore, prognostication of CRC is an imperative element for providing compelling/effective treatment for patients with the colorectal tumor. However, survival studies have shown inconsistent results. Fu et al[84] found out that younger patients tend to have a poorer prognosis compared to their older counterparts. Others studies did not agree with these findings and suggested that older patients have a poorer prognosis.[30,85-88] Manjelievskaia et al[89] reported that there is no survival difference between both patients. The use of biological markers to help prognostication is important. A good tumor biomarker should be less invasive, have a long half-life, and be estimated accurately and precisely by a simple and inexpensive blood test. It is also crucial to put into consideration how specific and sensitive they are to change so that it can be followed over time by serial measurements.[90] A couple of biomarkers meet these criteria. MicroRNA is a flawless precedent. The only potentially curative modality employed in patients with stages I-III and selected stage IV patient with the oligometastatic disease is surgical resection.

Patients with prior phases of the disease, including stage I (5-year survival >90%) and stage II (5-year survival 70%-85%), do not require adjuvant treatment aside from those with high-risk stage II disease. Individuals with resectable CRC are at a higher risk of locoregional relapse and require chemoradiotherapy in addition to adjuvant chemotherapy for risk reduction. The majority of patients with metastatic CRC (stage IV) are treated with a palliative intent to prolong life while preserving the quality of life. With modern chemotherapy regimens, the median overall survival of patients with metastatic CRC is less than 3 years (5-year survival approximately 10%; Table 1).[91] These numbers indicate clear improvements in outcomes for patients with CRC, and many promising novel therapies remain under development.[92] Prognostic biomarkers have been described in CRC.[80] *BRAF* mutations occur in 7% to 10% of patients and are associated with poor outcomes, especially in the metastatic setting. A study suggests that patients with primary tumors originating from the ascending colon have significantly worse overall survival compared to those with descending colon, including rectum, irrespective of the type of therapy used.[93] MSI-H, found in 22% of stage II and 11% of stage III patients, is associated with better outcomes in the adjuvant setting.[94]

Staging and Grading of CRC

There were concerns regarding the stratification of patients with bowel cancer in order to establish an appropriate surgical treatment.[95] Stage refers to the extent of cancer, that is, how large the tumor is, and the degree of metastasis. Knowing the stage of cancer helps to understand the degree and the chances of survival, plan the best treatment, and identify clinical trials that may be treatment options. The first clinical staging system is followed by Dukes' monumental work, which creates in his first articles a purely pathological classification based on the extent of the primary tumor[96] and highlights the implications of the histologic grading as a prognostic factor.[97] Numerous

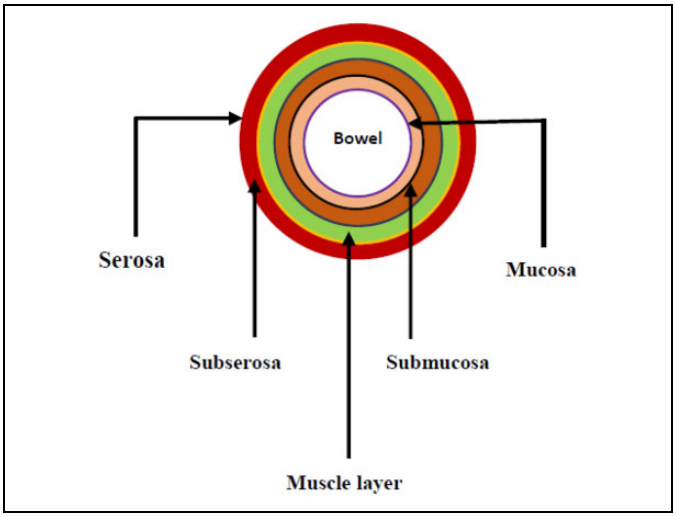


Figure 2. Cross section of the colon.

staging systems have been proposed and have been used for the classifications of various cancer subtypes, while others may be specific to a particular type of cancer (Table 1). Most staging systems include information about tumor location, cell type, tumor type, the degree of metastasis, and tumor grade (Figure 2). The most common types of staging system aside the tumor–node–metastasis (TNM) is the Dukes' staging system. As time progresses and new information unfolded, the Dukes' staging was constantly modified by Kirklin, Astler-Coller, the Australian, and clinicopathological (mucinous adenocarcinoma) classification.

The original description of cancer staging by Dukes is till date in use for the evaluation of prognosis and to a limited extent is used to determine the treatment for patients with CRC.[98] This classification was formally limited to 3 classes A, B, and C, and finally, letter D as a class for stratification was added to infer the presence of metastasis. Table 2 represents the TNM classification as a universal system for CRC stratification. This system corresponds to Dukes' mode of classification and is divided into 4 different categories.[99] The TNM classification of staging used to classify the magnitude of cancer is established on the tumor's anatomical information that is the size and degree (T), the node(s) involvement (N), and finally, whether or not the tumor has affected other organs through the blood stream (metastasis; M), grouping the cases with similar prognostic. The system is maintained collaboratively by the International Union for Cancer Control (IUCC) and the American Joint Committee for Cancer (AJCC), resulting in periodical and simultaneously publication of the *TNM Classification of Malignant Tumours* and the *AJCC Cancer Staging Manual*.[100] Currently, despite some critics, it is the most used clinically.

This system of classification was designed in such a way to prevent confusion and alleviate ambiguity by following physiopathological considerations after several repetitive revision of the Dukes' procedure. Obrocea et al[100] reported that research studies gave an improved understanding of cancer pathogenesis

Table 2. TNM Classification of Colorectal Cancer.

T: Primary tumor	N: Regional lymph node	M: Distant metastasis
Tx: Tumor cannot be assessed	Nx: Nodes cannot be assessed	Mx: Distant metastasis cannot be assessed
T_0: No evidence of primary tumor	N_0: No node metastasis	M_0: No distant metastasis
Tis: Carcinoma in situ	N_1: Metastasis in 1-3 nodes	M_1: Distant metastasis
T_2: Tumor invades muscularis propria	N_2: Metastasis in 4 or more nodes	
T_3: Tumor invades through into subserosa		
T_4: Tumor directly invades other organs		

Abbreviation: TNM, tumor–node–metastasis.

Table 3. Summary of CRC Classification System Based on TNM From AJCC, Modified Dukes' Staging, and Dukes' Staging System.[a,109]

Stage	T	N	M	Dukes	MAC
0	T_{IS}	N_0	M_0	-	-
I	T_1	N_0	M_0	A	A
	T_2	N_0	M_0	A	B_1
IIA	T_3	N_0	M_0	B	B_2
IIB	T_4	N_0	M_0	B	B_3
IIIA	T_{1-2}	N_1	M_0	C	C_1
IIIB	T_{3-4}	N_1	M_0	C	C_2/C_3
IIIC	Any T	N_2	M_0	C	$C_1/C_2/C_3$
IV	Any T	Any N	M_1	-	D

Abbreviations: AJCC, American Joint Committee for Cancer; CRC, colorectal cancer; TNM, tumor–node–metastasis.

[a]T_{IS} indicates carcinoma in situ intraepithelia or invasion of lamina propria; T_1, tumor invasion of submucosa; T_2, tumor invasion of muscularis propria; T_3, tumor invasion through the muscularis propria into pericolorectal tissues; T_4, penetration of tumor through the surface of the visceral peritoneum and further directly invading other organs.

and focused on the significant role of more nonanatomical biomarkers in order to create the prognosis and response to treatment for patients with CRC in such extent that a staging of disease made only on anatomical ground no longer responds to the recent advances in clinical evaluation and therapeutic decisions.

The TNM mode of cancer classification presents a great advantage over the Dukes' staging system (based on histopathology; Table 2). It allows the assessment of TNM categories by physical examination, imaging, endoscopy, and/or surgical exploration. Previously, data accumulated from pathologic staging were utilized essentially to determine prognosis. Current, the CRC staging has assumed additional roles, in particular, the determination of optimal therapy and assessment of response to treatment. The staging system is regularly used by cancer registries compared physicians. The staging system describes CRC as in situ (presence of abnormal cell without spread), localized (cancer is restricted to a particular location with no sign of spread), regional (spread of cancer to nearby lymph nodes, tissues, or organs), distant (spread of cancer to distance body parts), and unknown (limited information to discover the stage).

On a whole, for the first stage also known as Dukes A, the tumor growth captures the wall of the muscle, that is, the submucosa or muscular wall (T_1 and/or T_2). The second stage involved the lesions invasion stating from the propria (muscularis) through the subserosa and pericolic tissues (Dukes B, T_3). The lesions could also penetrate and target other organs through the visceral peritoneum (T_4). In the third stage, the tumors have metastasized, indicating their involvement in lymph nodes (N_1, 1-3 nodes involved and $N_2 \geq 4$ nodes). Lastly, stage IV (Dukes' D) lesions metastasize to other organs such as the liver, after perforation of a tumor into the peritoneal (Figure 2).

Grading

The most significant prognostic factor in CRC is the TNM staging established in accordance with the IUCC and AJCC, and therefore, it has crucial role in therapeutic decision-making in this cancer subtype[101,102]; regardless of its strong prognostic estimation of this staging system, it only indicates the anatomic degree of a tumor in some cases, without any correlation with patient survival.[103] Poor histological differentiation is currently considered to be a major adverse prognostic factor in CRC. Therefore, histological grading is incorporated in the histopathological report of CRC in routine practice.[104] Studies show that a 2-grade system can represent prognostic markers independent from TNM and with a better reproducibility.[100,105] According to this system, low-grade CRC includes well-differentiated and moderately well-differentiated adenocarcinoma and high-grade CRC weakly differentiated adenocarcinoma, mucinous adenocarcinoma, signet-ring carcinoma, and medullary and undifferentiated carcinoma, accordingly.[100,106-108] Tumor regression grade of the 4-grade system recommended are grade 0 (complete response)—no living cells; grade 1 (moderate response)—reduced number of cancer cells; grade 2 (minimal response)—insignificant cancer outgrown by fibrosis; and grade 3 (poor response)—minimal or no tumor kill, extensive residual cancer (Table 3).

Specific miRNA Expression in CRC Initiation and Progression

The most imperative predictor of outcome is the stage of disease at diagnosis. In general, surgical resection is the main potential curative modality and is utilized in individuals with stages I-III and select stage IV patients with oligometastatic disease. Most CRCs emerge from adenomatous polyps over a time of years to decades by aggregation of serial physical changes (serial somatic mutations) because of basic acquired or gained CIN.[110] As indicated by the adenoma-carcinoma sequence model, the initiating mutation is in the *APC* gene.[111] Consequent changes incorporate KRAS and BRAF, with implications for treatment and prognosis, respectively. Different occasions incorporate p53 alterations and loss of chromosome

18q.[112] Familial adenomatous polyposis is portrayed by germ line transformations in APC, leading to Wnt pathway activation.[113] Hereditary nonpolyposis colorectal cancer (HNPCC) and around 15% of sporadic cases are portrayed by germ line or somatic DNA repair deformities or methylation changes in the *MMR* genes, which may prompt genomic instability because of the disabled capacity to correct DNA replication errors.[114] This prompts mutations in malignancy-related genes subsequently driving carcinogenesis. Also, the contraction and expansion of microsatellites compared to the normal tissue are also implicated in carcinogenesis.[115] The hypermethylation phenotype (CIMP+) is characterized by DNA methylation of CpG islands of numerous genes such as those involved in MMR, resulting in silencing of gene expression typically causing serrated adenomas.[116]

MicroRNAs Implicated in Each Stage of CRC Using the TNM Staging Classification

This section attempts to discuss both the differential miRNA expressions across various cancers, across all the stages of CRC, and also those that are commonly expressed to bring about good treatment outcome and better survival for patients with CRC since their expression levels in cancers may assist therapeutic decisions and have advantage as a therapeutic target through miRNA inhibition or replacement strategies. Several studies have examined the expression patterns of miRNA through various techniques (deep sequencing, quantitative RT-PCR, and microarray) and affirmed their reliability and reproducibly altered in CRC.[76,78,117-119] All these studies revealed that miRNA is differentially expressed in CRC compared to normal tissues. This is in line with the hypothesis that aberrant miRNA expression is pivotal in colorectal carcinogenesis and development.[120] Studies have affirmed that specific miRNAs have imperative oncogenic capacities while others have critical tumor suppressive capacities and that these capacities should be assessed for each miRNA independently with regard to the particular tissue or cancer type.

Cheng et al[80] experimentally determined that miR-141 is differentially expressed in the late stage of CRC. This can, therefore, be used within colorectal tissue as a differential diagnostic biomarker for M_1. Furthermore, miR-143 and miR-145 were suggested to play a tumor suppressive function in CRC.[121] Wang et al[122] built a diagnostic model for CRC by experimentally validating miR-21, miR-31, miR-203, miR-92a, miR-181b, miR-145, miR-143, miR-30c, miR-17, and let-7g and then identified a profile that combined 6 miRNAs, which can serve as a novel noninvasive biomarker for CRC diagnosis. MicroR-193a-3p was also predicted as a tumor-suppressive miRNA involved in the development of CRC (early stage of colorectal carcinogenesis) and also have an effect on the sensitivity of anti-EGFR therapy.[123] Expression of miR-181c has been assessed to suggest the recurrence of stage II CRC.[124] Both miR-17-3p and miR-221 were found to be commonly expressed in all the stages (stages I, II, III, and IV) of CRC, with a sensitivity of 64% and 86% and a specificity of approximately 70% and 41%, respectively, in plasma.[79,125] In feces, miR-17 and 21 are as well been shown to be commonly expressed in all the stages of CRC.[126,127] Also, miR-91a, miR-106a, miR-135a, and miR-135b were implicated, but their stages are not reported.[128] Huang et al[44] surveyed the expressions of 12 miRNAs in plasma samples from patients with advanced CRC and healthy controls utilizing RT-PCR and discovered that miR-29a and miR-92a possess significant diagnostic value for advanced neoplasia and proposed that these miRNAs have solid potential as novel noninvasive biomarkers for early CRC detection. From our ongoing research, 5 novel miRNAs have been discovered using in silico approaches and have been found to be linked/implicated in CRC but await molecular validation for the stratification of this disease at each stage of TNM.

Conclusions

The enthusiasm for biomarkers relating to CRC is obviously expanding. They shape another part of clinical and laboratory research, which helps interpret these ideas to more significant applications in disease management. MicroRNA biomarkers are an emerging field that can potentially assist in guiding the diagnosis, prognosis, treatment, and management of CRC. The potential miRNA advantage for clinical translation in CRC is a focal point for better understanding of staging and specific treatment efficacy in surgery for CRC. Staging supplies information regarding the prognosis and may suggest the requirement for other therapy. Accurate assessment of CRC with specific miRNA TNM classes is vital for deciding the best stage-specific management to improve the predictive and prognosis of the disease.

Future Perspective

Research has confirmed the increasing rate of incidence and mortality of CRC subtype worldwide, and as such, it has become a public health issue globally. The future perspective of this review is aimed at the provision of the current findings in the diagnosis and management of this disease as well as latest discoveries and future viewpoint in the field of oncology as a means to assist in the insight of the cancer subtype.

Since the major causes of CRC are both environmental factors and genetic factors, their exploitation can bring about new diagnosis and treatment strategies.

For CRC treatment, a therapeutic model that distinguishes individuals into various categories with clinical decisions, practices, mediations, and additionally items being custom-made to the individual patient depending on their anticipated reaction or risk of disease such as personalized medicine is fast becoming a significant tool. Therefore, it is noteworthy to carry out comprehensive research of the tumor features of individual patients to tailor the best therapy.

Finally, the greater part of current research in this field is largely dependent on the development of a new treatment that is noninvasive, less expensive, sensitive, specific, and more

effective compared to the conventional therapies. MicroRNAs have proven to be widely distributed all over the body in terms of their abundance and their expression profiles have also been exploited to be different among cancer subtypes and are tissue-specific. The development of miRNAs as biomarkers will improve diagnosis as well as detection in the early stage of this disease since this disease is largely treatable when detected earlier. Discoveries in this area and their clinical significance will improve the overall survival and disease management of patients with CRC subtype in the nearest future.

Acknowledgments

The authors thank the Plant Omics Laboratory and the Bioinformatics Research Group of the University of the Western Cape for their usual support. Also, Dr Taiwo Akinsoji (MBBS, MPH) of the University of Illinois at Springfield for co-proof reading the final manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Adewale Oluwaseun Fadaka https://orcid.org/0000-0002-3952-2098

References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2016. doi:10.1136/gutjnl-2015-310912.
2. Brenner H, Altenhofen L, Hoffmeister M. Sex, age, and birth cohort effects in colorectal neoplasms: a cohort analysis. *Ann Intern Med*. 2010;152(11):697-703.
3. Binefa G, Rodríguez-Moranta F, Teule À, Medina-Hayas M. Colorectal cancer: from prevention to personalized medicine. *World J Gastroenterol*. 2014;20(22):6786-6808.
4. Bosman FT, Carneiro F, Hruban RH. Carcinoma of the colon and rectum. In: Theise ND ed. *WHO Classification of Tumours of the Digestive System*. Lyon, France: IARC Press; 2010:134-146.
5. Fleming M, Ravula S, Tatishchev SF, Wang HL. Colorectal carcinoma: pathologic aspects. *J Gastrointest Oncol*. 2012;3(3):153-173.
6. Farrington SM, Tenesa A, Barnetson R, et al. Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet*. 2005;77(1):112-119.
7. Pinol V, Castells A, Andreu M, et al. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA* 2005;293(16):1986-1994.
8. Walker AS, Johnson EK, Maykel JA, et al. Future directions for the early detection of colorectal cancer recurrence. *J Cancer*. 2014;5(4):272-280.
9. Sawyers CL. The cancer biomarker problem. *Nature*. 2008;452(7187):548-552.
10. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends—an update. *Cancer Epidemiol Biomarkers Prev*. 2015;25(1):16-27.
11. Biomarkers Definitions Working Group, Atkinson AJ Jr, Colburn WA, et al. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69(3):89-95.
12. Aronson JK. Biomarkers and surrogate endpoints. *Br J Clin Pharmacol*. 2005;59(5):491-494.
13. Newton K, Newman W, Hill J. Review of biomarkers in colorectal cancer. *Colorectal Dis*. 2012;14(1):3-17.
14. Rawson JB, Bapat B. Epigenetic biomarkers in colorectal cancer diagnostics. *Expert Rev Mol Diagn*. 2012;12(5):499-509.
15. Schepeler T, Reinert JT, Ostenfeld MS, et al. Diagnostic and prognostic microRNAs in stage II colon cancer. *Cancer Res*. 2008;68(15):6416-6424.
16. Ardekani AM, Naeini MM. The role of microRNAs in human diseases. *Avicenna J Med Biotechnol*. 2010;2(4):161-179.
17. Keller A, Meese E. Can circulating miRNAs live up to the promise of being minimal invasive biomarkers in clinical settings? *Wiley Interdiscip Rev RNA*. 2016;7(2):148-156.
18. Gambari R, Brognara E, Spandidos DA, Fabbri E. Targeting oncomiRNAs and mimicking tumor suppressor miRNAs: new trends in the development of miRNA therapeutic strategies in oncology. *Int J Oncol*. 2016;49(1):5-32.
19. Fehlmann T, Ludwig N, Backes C, Meese E, Keller A. Distribution of microRNA biomarker candidates in solid tissues and body fluids. *RNA Biol*. 2016;13(11):1084-1088.
20. Urbich C, Kuehbacher A, Dimmeler S. Role of microRNAs in vascular diseases, inflammation, and angiogenesis. *Cardiovasc Res*. 2008;79(4):581-588.
21. Ludwig N, Leidinger P, Becker K, et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Res*. 2016;44(8):3865-3877.
22. Cheng G. Circulating miRNAs: roles in cancer diagnosis, prognosis and therapy. *Adv Drug Deli Rev*. 2015;81:75-93.
23. Rehbein G, Schmidt B, Fleischhacker M. Extracellular microRNAs in bronchoalveolar lavage samples from patients with lung diseases as predictors for lung cancer. *Clin Chim Acta*. 2015;450:78-82.
24. Izzotti A, Carozzo S, Pulliero A, Zhabayeva D, Ravetti JL, Bersimbaev R. Extracellular microRNA in liquid biopsy: applicability in cancer diagnosis and prevention. *Am J Cancer Res*. 2016;6(7):1461-1493.
25. Molina-Pinelo S, Suárez R, et al. Association between the miRNA signatures in plasma and bronchoalveolar fluid in respiratory pathologies. *Dis Markers*. 2012;32(4):221-230.
26. Izzotti A, Balansky R, Ganchev G, et al. Blood and lung microRNAs as biomarkers of pulmonary tumorigenesis in cigarette smoke-exposed mice. *Oncotarget*. 2016;7(51):84758-84774.
27. Cho WC. MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. *Int J Biochem Biol*. 2010;42(8):1273-1281.

28. Aqeilan R, Calin G, Croce C. miR-15a and miR-16-1 in cancer: discovery, function and future perspectives. *Cell Death Differ*. 2010;17(2):215-220.
29. Wang L-G, Gu J. Serum microRNA-29a is a promising novel marker for early detection of colorectal liver metastasis. *Cancer Epidemiol*. 2012;36(1):e61-e67.
30. Campos FG, Figueiredo MN, Monteiro M, Nahas SC, Cecconello I. Incidence of colorectal cancer in young patients. *Rev Col Bras Cir*. 2017;44(2):208-215.
31. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61(5):759-767.
32. Worthley DL, Leggett BA. Colorectal cancer: molecular features and clinical opportunities. *Clin Biochem Rev*. 2010;31(2): 31-38.
33. Rustgi AK. Hereditary gastrointestinal polyposis and nonpolyposis syndromes. *N Engl J Med*. 1994;331(25):1694-1702.
34. Loeb LA. A mutator phenotype in cancer. *Cancer Res*. 2001; 61(8):3230-3239.
35. Nojadeh JN, Behrouz Sharif S, Sakhinia E. Microsatellite instability in colorectal cancer. *EXCLI J*. 2018;17:159-168.
36. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer—a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer*. 2006;6(2):107-116.
37. Toyota M, Ohe-Toyota M, Ahuja N, Issa J-PJ. Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype. *Proc Natl Acad Sci U S A*. 2000;97(2): 710-715.
38. Weisenberger DJ, Siegmund KD, Campan M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet*. 2006;38(7):787-793.
39. Ogino S, Kawasaki T, Ogawa A, Kirkner GJ, Loda M, Fuchs CS. TGFBR2 mutation is correlated with CpG island methylator phenotype in microsatellite instability-high colorectal cancer. *Hum Pathol*. 2007;38(4):614-620.
40. Nosho K, Irahara N, Shima K, et al. Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample. *PLoS One* 2008;3(11): e3698.
41. Fadaka AO, Ojo BA, Adewale OB, Esho T, Pretorius A. Effect of dietary components on miRNA and colorectal carcinogenesis. *Cancer Cell Int*. 2018;18:130.
42. Lawrie CH, Gal S, Dunlop HM, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol*. 2008;141(5): 672-675.
43. Heneghan HM, Miller N, Kerin MJ. MiRNAs as biomarkers and therapeutic targets in cancer. *Curr Opin Pharmacol*. 2010;10(5): 543-550.
44. Huang Z, Huang D, Ni S, Peng Z, Sheng W, Du X. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. *Int J Cancer*. 2010;127(1):118-126.
45. Cortez MA, Bueso-Ramos C, Ferdin J, Lopez-Berestein G, Sood AK, Calin GA. MicroRNAs in body fluids—the mix of hormones and biomarkers. *Nat Rev Clin Oncol*. 2011;8(8):467-477.
46. Yu D-C, Li Q-G, Ding X-W, Ding Y-T. Circulating microRNAs: potential biomarkers for cancer. *Int J Mol Sci*. 2011;12(3): 2055-2063.
47. Schwarzenbach H, Nishida N, Calin GA, Pantel K. Clinical relevance of circulating cell-free microRNAs in cancer. *Nat Rev Clin Oncol*. 2014;11(3):145-156.
48. Toiyama Y, Okugawa Y, Goel A. DNA methylation and microRNA biomarkers for noninvasive detection of gastric and colorectal cancer. *Biochem Biophys Res Commun*. 2014;455(1-2): 43-57.
49. Chand M, Keller DS, Mirnezami R, et al. Novel biomarkers for patient stratification in colorectal cancer: a review of definitions, emerging concepts, and data. *World J Gastrointest Oncol*. 2018; 10(7):145-158.
50. Plano D, Alcolea V, Sanmartín C, Sharma AK. Methods of selecting combination therapy for colorectal cancer patients: a patent evaluation of US20160025730A1. *Expert Opin Ther Pat*. 2017; 27(5):527-538.
51. Mojarad EN, Kuppen PJ, Aghdaei HA, Zali MR. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench*. 2013;6(3):120-128.
52. Kambara T, Simms L, Whitehall V, et al. BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut*. 2004;53(8):1137-1144.
53. Namba H, Nakashima M, Hayashi T, et al. Clinical implication of hot spot BRAF mutation, V599E, in papillary thyroid cancers. *J Clin Endocrinol Metab*. 2003;88(9):4393-4397.
54. Tan YH, Liu Y, Eu KW, et al. Detection of BRAF V600E mutation by pyrosequencing. *Pathology*. 2008;40(3):295-298.
55. Li WQ, Kawakami K, Ruszkiewicz A, Bennett G, Moore J, Iacopetta B. BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer independently of microsatellite instability status. *Mol Cancer*. 2006;5:2.
56. Elisei R, Ugolini C, Viola D, et al. BRAFV600E mutation and outcome of patients with papillary thyroid carcinoma: a 15-year median follow-up study. *J Clin Endocrinol Metab*. 2008;93(10): 3943-3949.
57. Puxeddu E, Moretti S, Elisei R, et al. BRAFV599E mutation is the leading genetic event in adult sporadic papillary thyroid carcinomas. *J Clin Endocrinol Metab*. 2004;89(5):2414-2420.
58. Maldonado JL, Fridlyand J, Patel H, et al. Determinants of BRAF mutations in primary melanomas. *JNCI*. 2003;95(24):1878-1890.
59. Gear H, Williams H, Kemp EG, Roberts F. BRAF mutations in conjunctival melanoma. *Invest Ophthalmol Vis Sci*. 2004;45(8): 2484-2488.
60. Deng G, Bell I, Crawley S, et al. *BRAF* mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin Cancer Res*. 2004;10(1 pt 1):191-195.
61. Benlloch S, Payá A, Alenda C, et al. Detection of BRAF V600E mutation in colorectal cancer: comparison of automatic sequencing and real-time chemistry methodology. *J Mol Diagn*. 2006; 8(5):540-543.
62. Qi Li W, Kawakami K, Ruszkiewicz A, Bennett G, Moore J, Iacopetta B. BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer

independently of microsatellite instability status. *Mol Cancer* 2006;5:2.
63. Bettington M, Walker N, Clouston A, Brown I, Leggett B, Whitehall V. The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology*. 2013;62(3):367-386.
64. Mrkonjic M, Roslin NM, Greenwood CM, et al. Specific variants in the MLH1 gene region may drive DNA methylation, loss of protein expression, and MSI-H colorectal cancer. *PLoS One*. 2010;5(10):e13314.
65. Vilar E, Gruber SB. Microsatellite instability in colorectal cancer—the stable evidence. *Nat rev Clin Oncol*. 2010;7(3):153-162.
66. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Nati Acad Sci U S A*. 2008;105(30):10513-10518.
67. Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res*. 2008;18(10):997-1006.
68. Turchinovich A, Weiz L, Langheinz A, Burwinkel B. Characterization of extracellular circulating microRNA. *Nucleic Acids Res*. 2011;39(16):7223-7233.
69. Jothimani G, Sriramulu S, Chabria Y, Sun X-F, Banerjee A, Pathak S. A review on theragnostic applications of microRNAs and long non-coding RNAs in colorectal cancer. *Curr Top Med Chem*. 2018;18(10):2614-2629.
70. Kosaka N, Iguchi H, Yoshioka Y, Takeshita F, Matsuki Y, Ochiya T. Secretory mechanisms and intercellular transfer of microRNAs in living cells. *J Biol Chem*. 2010;285(23):17442-17452. doi:10.1074/jbc. M110. 107821.
71. Arroyo JD, Chevillet JR, Kroh EM, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Nati Acad Sci U S A*. 2011;108(12):5003-5008.
72. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat Cell Biol*. 2011;13(4):423-433.
73. Ichikawa D, Komatsu S, Konishi H, Otsuji E. Circulating microRNA in digestive tract cancers. *Gastroenterology*. 2012;142(5):1074-1078.e1.
74. Kawaguchi T, Komatsu S, Ichikawa D, et al. Circulating microRNAs: a next-generation clinical biomarker for digestive system cancers. *Int J Mol Sci*. 2016;17(9):1459.
75. Slaby O, Svoboda M, Michalek J, Vyzula R. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. *Mol Cancer*. 2009;8:102.
76. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;435(7043):834-838.
77. Slaby O, Svoboda M, Fabian P, et al. Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology*. 2007;72(5-6):397-402.
78. Schetter AJ, Leung SY, Sohn JJ, et al. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA*. 2008;299(4):425-436.
79. Ng EK, Chong WW, Jin H, et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut*. 2009;58(10):1375-1381.
80. Cheng H, Zhang L, Cogdell DE, et al. Circulating plasma miR-141 is a novel biomarker for metastatic colon cancer and predicts poor prognosis. *PLoS One*. 2011;6:e17745.
81. Ogata-Kawata H, Izumiya M, Kurioka D, et al. Circulating exosomal microRNAs as biomarkers of colon cancer. *PLoS One*. 2014;9:e92921.
82. Zanutto S, Pizzamiglio S, Ghilotti M, et al. Circulating miR-378 in plasma: a reliable, haemolysis-independent biomarker for colorectal cancer. *Br J Cancer*. 2014;110(4):1001-1007.
83. Ohno-Machado L. Modeling medical prognosis: survival analysis techniques. *J Biomed Inform*. 2001;34(6):428-439.
84. Fu J-F, Huang Y-Q, Yang J, Yi C-H, Chen H-L, Zheng S. Clinical characteristics and prognosis of young patients with colorectal cancer in Eastern China. *World J Gastroenterol*. 2013;19(44):8078-8084.
85. McKay A, Donaleshen J, Helewa RM, et al. Does young age influence the prognosis of colorectal cancer: a population-based analysis. *World J Surg Oncol*. 2014;12:370.
86. Wang M-J, Ping J, Li Y, et al. The prognostic factors and multiple biomarkers in young patients with colorectal cancer. *Sci Rep*. 2015;5:10645.
87. Kneuertz PJ, Chang GJ, Hu CY, et al. Overtreatment of young adults with colon cancer: more intense treatments with unmatched survival gains. *JAMA Surg*. 2015;150(5):402-409.
88. Jiang Z, Wang X, Tan X, Fan Z. Effect of age on survival outcome in operated and non-operated patients with colon cancer: a population-based study. *PLoS One*. 2016;11(1):e0147383.
89. Manjelievskaia J, Brown D, McGlynn KA, Anderson W, Shriver CD, Zhu K. Chemotherapy use and survival among young and middle-aged patients with colon cancer. *JAMA Surg*. 2017;152(5):452-459.
90. McShane LM, Altman DG, Sauerbrei W. *Identification of Clinically Useful Cancer Prognostic Factors: What are We Missing?* Oxford, England: Oxford University Press; 2005.
91. Adam R, De Gramont A, Figueras J, et al. The oncosurgery approach to managing liver metastases from colorectal cancer: a multidisciplinary international consensus. *Oncologist*. 2012;17(10):1225-1239. theoncologist. 2012-0121.
92. Labianca R, Beretta GD, Kildani B, et al. Colon cancer. Critical reviews in oncology/hematology. *Crit Rev Oncol Hematol*. 2010;74:106-133.
93. Tamas K, Walenkamp A, De Vries E, et al. Rectal and colon cancer: not just a different anatomic site. *Cancer Treat Rev*. 2015;41(8):671-679.
94. Guastadisegni C, Colafranceschi M, Ottini L, Dogliotti E. Microsatellite instability as a marker of prognosis and response to therapy: a meta-analysis of colorectal cancer survival data. *Eur J Cancer*. 2010;46(15):2788-2798.
95. Lockhart-Mummery JP. Two hundred cases of cancer of the rectum treated by perineal excision. *British J Surg*. 1926;14(53):110-124.
96. Dukes CE. The classification of cancer of the rectum. *J Pathol Bacteriol*. 1932;35:323-332.

97. Dukes C. *Histological Grading of Rectal Cancer*. Thousand Oaks, CA: Sage; 1937.
98. Thebo JS, Senagore AJ, Reinhold DS, Stapleton SR. Molecular staging of colorectal cancer. *Dis Colon Rectum*. 2000;43: 155-159.
99. Hutter RV, Sobin LH. A universal staging system for cancer of the colon and rectum. Let there be light. *Arch Pathol Lab Med*. 1986;110(5):367-368.
100. Obrocea F, Sajin M, Marinescu EC, Stoica D. Colorectal cancer and the 7th revision of the TNM staging system: review of changes and suggestions for uniform pathologic reporting. *Rom J Morphol Embryol*. 2011;52(2):537-544.
101. Kantola T, Klintrup K, Väyrynen J, et al. Stage-dependent alterations of the serum cytokine pattern in colorectal carcinoma. *Br J Cancer*. 2012;107(10):1729-1736.
102. Barresi V, Reggiani LB, Ieni A, Caruso RA, Tuccari G. Histological grading in colorectal cancer: new insights and perspectives. *Histol Histopathol*. 2015;30(9):1059-1067.
103. Barresi V, Bonetti LR, Branca G, Di Gregorio C, De Leon MP, Tuccari G. Colorectal carcinoma grading by quantifying poorly differentiated cell clusters is more reproducible and provides more robust prognostic information than conventional grading. *Virchows Arch*. 2012;461(6):621-628.
104. Gospodarowicz MK, Brierley JD, Wittekind C. *TNM Classification of Malignant Tumours*. Hoboken, NJ: John Wiley & Sons, 2017.
105. Tang LH, Gonen M, Hedvat C, Modlin IM, Klimstra DS. Objective quantification of the Ki67 proliferative index in neuroendocrine tumors of the gastroenteropancreatic system: a comparison of digital image analysis with manual methods. *Am J Surg Pathol*. 2012;36(12):1761-1770.
106. Bujanda L, Cosme A, Gil I, Arenas-Mirave JI. Malignant colorectal polyps. *World J Gastroenterol*. 2010;16(25):3103-3111.
107. Ghazi S, Lindforss U, Lindberg G, et al. Analysis of colorectal cancer morphology in relation to sex, age, location, and family history. *J Gastroenterol*. 2012;47(6):619-634.
108. Kim JW, Shin MK, Kim BC. Clinicopathologic impacts of poorly differentiated cluster-based grading system in colorectal carcinoma. *J Korean Med Sci*. 2015;30(1):16-23.
109. Wu JS. Rectal cancer staging. *Clin Colon Rectal Surg*. 2007; 20(3):148-157.
110. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res*. 2012;5(1):19-27.
111. Matano M, Date S, Shimokawa M, et al. Modeling colorectal cancer using CRISPR-Cas9–mediated engineering of human intestinal organoids. *Nat Med*. 2015;21(3):256-262.
112. Ogino S, Nosho K, Irahara N, et al. Prognostic significance and molecular associations of 18q loss of heterozygosity: a cohort study of microsatellite stable colorectal cancers. *J Clin Oncol*. 2009;27(27):4591-4598.
113. Klaus A, Birchmeier W. Wnt signalling and its impact on development and cancer. *Nat Rev Cancer*. 2008;8(5):387-398.
114. Jin B, Robertson KD. DNA methyltransferases, DNA damage repair, and cancer. In: *Epigenetic Alterations in Oncogenesis*. Berlin, Germany: Springer, 2013:3-29.
115. Bunz F. Genetic instability and cancer. In: *Principles of Cancer Genetics*. 2008:125-172.
116. Fu T, Pappou EP, Guzzetta AA, et al. CpG island methylator phenotype positive tumors in the absence of MLH1 methylation constitute a distinct subset of duodenal adenocarcinomas and are associated with poor prognosis. *Clin Cancer Res*. 2012;18(17): 4743-4752. clincanres. 0707.2012.
117. Cummins JM, He Y, Leary RJ, et al. The colorectal microRNAome. *Proc Natl Acad Sci U S A*. 2006;103(10):3687-3692.
118. Volinia S, Calin GA, Liu C-G, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A*. 2006;103(7):2257-2261.
119. Luo X, Burwinkel B, Tao S, Brenner H. MicroRNA signatures: novel biomarker for colorectal cancer? *Cancer Epidemiol Biomarkers Prev*. 2011;20(7):1272-1286.
120. Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. *Cancer J (Sudbury, Mass.)*. 2012;18(3): 244-252.
121. Michael MZ, O'Connor SM, van Holst Pellekaan NG, Young GP, James RJ. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol Cancer Res*. 2003;1(12):882-891.
122. Wang J, Huang S-K, Zhao M, et al. Identification of a circulating microRNA signature for colorectal cancer detection. *PLoS One*. 2014;9(4):e87451.
123. Takahashi H, Takahashi M, Ohnuma S, et al. MicroRNA-193a-3p is specifically down-regulated and acts as a tumor suppressor in BRAF-mutated colorectal cancer. *BMC Cancer*. 2017; 17(1):723.
124. Yamazaki N, Koga Y, Taniguchi H, et al. High expression of miR-181c as a predictive marker of recurrence in stage II colorectal cancer. *Oncotarget*. 2017;8(4):6970-6983.
125. Pu XX, Huang GL, Guo HQ, et al. Circulating miR-221 directly amplified from plasma is a potential diagnostic and prognostic marker of colorectal cancer and is correlated with p53 expression. *J Gastroenterol Hepatol*. 2010;25(10):1674-1680.
126. Koga Y, Yasunaga M, Takahashi A, et al. MicroRNA expression profiling of exfoliated colonocytes isolated from feces for colorectal cancer screening. *Cancer Prev Res*. 2010;3(11). doi:10. 1158/1940-6207. CAPR-10-0036.
127. Link A, Balaguer F, Shen Y, et al. Fecal MicroRNAs as novel biomarkers for colon cancer screening. *Cancer Epidemiol Biomark*. 2010;19(7). doi:10.1158/1055-9965.EPI-10-0027.
128. Cheng Q, Feng F, Zhu L, et al. Circulating miR-106a is a novel prognostic and lymph node metastasis indicator for cholangiocarcinoma. *Sci Rep*. 2015;5:16103.

CHAPTER 4.0

IN SILICO IDENTIFICATION OF MICRORNAS AS CANDIDATE COLORECTAL CANCER BIOMARKERS

ADEWALE OLUWASEUN FADAKA*, ASHLEY PRETORIUS, AND ASHWIL KLEIN

Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa.

Accepted/ published: *Tumor Biology*, 1-15. (2019)
DOI: 10.1177/1010428319883721 **(Impact factor: 3.60)**

Statement of Contributions of Joint Authorship

Adewale Fadaka (PhD candidate)

Conceived the concept of the manuscript as well as the write up and compilation. Preparation of figures and tables.

Ashley Pretorius (Co-Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

Ashwil Klein (Principal Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

This Chapter is an exact copy of the journal paper below.

Original Article

Tumor Biology

Tumor Biology
November 2019: 1–15

DOI: 10.1177/1010428319883721
journals.sagepub.com/home/tub
SAGE

In silico identification of microRNAs as candidate colorectal cancer biomarkers

Adewale Oluwaseun Fadaka, Ashwil Klein and Ashley Pretorius

Abstract

The involvement of microRNA in cancers plays a significant role in their pathogenesis. Specific expressions of these non-coding RNAs also serve as biomarkers for early colorectal cancer diagnosis, but their laboratory/molecular identification is challenging and expensive. The aim of this study was to identify potential microRNAs for colorectal cancer diagnosis using in silico approach. Sequence similarity search was employed to obtain the candidate microRNA from the datasets, and three target prediction software were employed to determine their target genes. To determine the involvement of these microRNAs in colorectal cancer, the microRNA gene list obtained was used alongside with colorectal cancer expressed genes from gbCRC and CoReCG databases for gene intersection analysis. The involvement of these genes in the cancer subtype was further strengthened with the DAVID database. KEGG and Gene Ontology were used for the pathway and functional analysis, while STRING was employed for the interactions of protein network and further visualized by Cytoscape. The cBioPortal database was used to prioritize the target genes; prognostic and expression analysis were finally performed on the candidate microRNAs and the prioritized targets. This study, therefore, identified five candidate microRNAs, two hub genes (CTNNB1 and epidermal growth factor receptor), and seven significant target genes associated with colorectal cancer. The molecular validation studies are ongoing to ascertain the biological fitness of these findings.

Keywords

Colorectal cancer, diagnostics, early detection, in silico analysis, microRNA, biomarkers, BLAST, CD-HIT-EST-2D

Date received: 20 May 2019; accepted: 30 September 2019

Introduction

Colorectal cancer (CRC) is one of the commonly diagnosed cancers and the second cause of cancer mortality in the world. Jemal et al.[1] reported the rate of prevalence and mortality of CRC to be over 1 million and 600,000 cases per year, respectively, despite the diagnosis and treatment of this disease. The increasing rate of incidence and mortality of this disease is, therefore, a public health issue.[2] Immensely, patients' survival corresponds to tumor stage at the period of diagnosis and approximately 50% has been attributed to death due to metastasis.[3,4] Alterations, such as genetic and epigenetic, can alter tumor suppressor genes (*DCC*, *APC*, *SMAD4*, and *TP53*) and oncogenes (*KRAS*) in CRC.[5,6] One of the basic procedures driving the initiation and progression of CRC is the accumulation of a variety of genetic and epigenetic changes in epithelial cells of colorectum.[7] The significant challenge to the management of CRC is early detection worldwide, which makes the present treatment options to be administered so late after tumor metastasis. If tumors are detected early enough, and polyps are surgically resected, they could

Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Bellville, South Africa

Corresponding author:
Adewale Oluwaseun Fadaka, Department of Biotechnology, The Faculty of Natural Sciences, University of the Western Cape, Bellville 7535, South Africa.
Emails: silvernonferous@gmail.com; afadaka@uwc.ac.za

reduce both the incidence and death rate of CRC. Tools involved in the early detection are faced with numerous challenges thereby limiting the development of standardized biochemical diagnostic approaches which are non-invasive, more sensitive, and specific for CRC stages. To accomplish this, more non-invasive, selective, and specific diagnostic biomarker which can detect a tumor at an early stage is important.

MicroRNAs are attracting considerable interest, with expanding proof on the role of microRNAs' expression in CRC development and progression.[8]

MicroRNAs are involved in numerous biological and pathological processes, and dysregulation of these microRNAs has been strongly associated with CRC pathogenesis.[9] Since the expression pattern of microRNAs has been confirmed to be different between tissues and body fluids when compared to normal, they can be referred to as oncomiR and tumor suppressor microRNA and thus, they can be utilized as diagnostic, prognostic, and predictive biomarkers of CRC.[10] The identification of prognostic and predictive biomarkers is, therefore, crucial to certify the standard of purity in cancer genomics. The discovery of these microRNAs, which predominate at different stages of CRC, can permit a more comprehensive evaluation and understanding of their effects. This can additionally provide exciting opportunities for CRC pathogenesis, diagnosis, and treatment insights into CRC management.

Molecular approaches to identify these microRNAs for diagnostic and predictive biomarkers for CRC have suffered major restraints, such as the appropriate procedure to follow, time consumption, laborious, and expensive.[11] Since the development of a diagnostic microRNA biomarker to distinctly monitor colorectal carcinogenesis is important to enhance the diagnosis rate at latter stages, the in silico approach toward the identification of microRNAs and for diagnostics for CRC is largely required. The overall experimental approach is represented in Figure 1(a) and (b).

Materials and methods

Data selection

Total microRNA sequences were retrieved from miRBase[12] at http://www.mirbase.org/ as reference dataset, while microRNAs associated with CRC were obtained from dbDEMC 2.0 at http://www.picb.ac.cn/dbDEMC/,[13] miR2Disease at http://www.mir2disease.org/,[14] HMDD at http://www.cuilab.cn/hmdd, and miRCancer at http://mircancer.ecu.edu/[15] and pulled together for the query dataset. Furthermore, duplicates were removed using the script in Clustering Database at High Identity with Tolerance (CD-HIT) suit from each dataset to obtain unique sequences.

Sequence similarity search

For the identification of candidate microRNAs, sequence analysis was employed using standard available search tools, such as Basic Local Alignment Search Tool (BLAST) at https://blast.ncbi.nlm.nih.

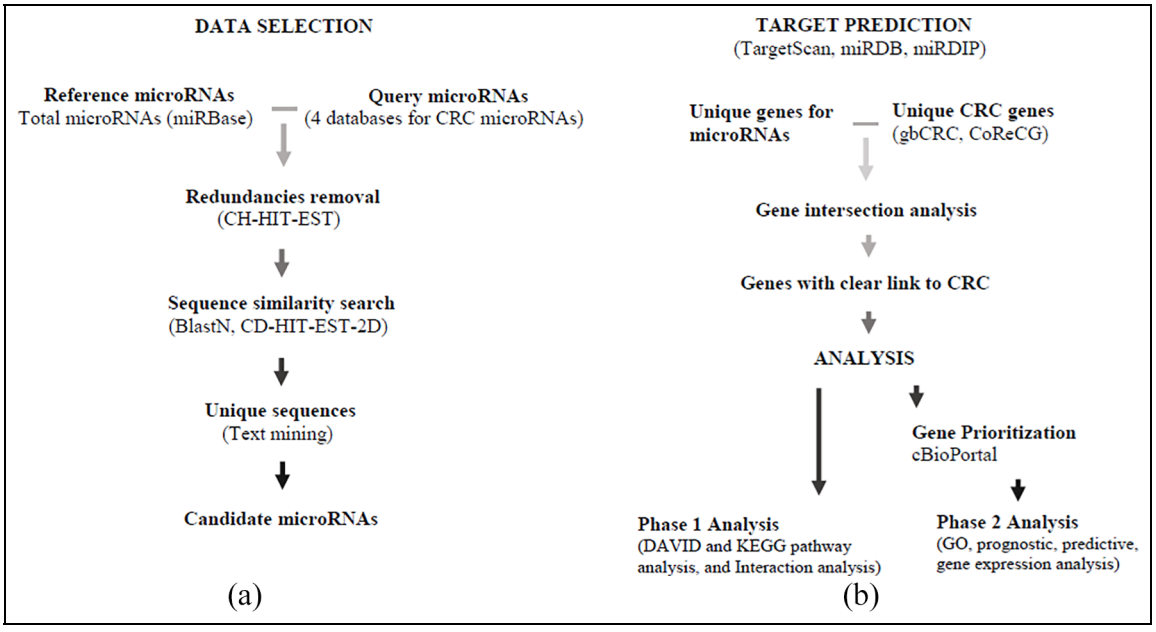


Figure 1. Experimental design of the study. (a) The data selection process for candidate microRNA. (b) Target prediction and further analysis.

gov/[16] and homology detection and CD-HIT.[17] BLASTN and CD-HIT-EST-2D were locally run between the two datasets (reference and query) using command lines with the following parameters: sequence identity threshold of 0.90, *e*-value of 1×10^{-3}, and word size of 7. The results of the search were further text mined to obtain the final list of candidate microRNAs together with the validated clusters.

MicroRNA target prediction

The targets of these microRNAs were identified using three different databases, namely miRDB at http://www.mirdb.org/index.html, TargetScan at http://www.targetscan.org/vert_72, and mirDIP at http://ophid.utoronto.ca/mirDIP/index.jsp, according to microRNA sequences and the following criteria: miRNA 3′ Site,[18,19] Conservation Status,[18,20] and the Seed Region.[18,21] Using these three algorithms, the two lists generated from the sequence similarity search (candidate and validated microRNAs) were queried separately and further analyzed by the intersection analysis with R-package (https://cran.r-project.org/) to obtain a unique gene list after the removal of redundancies.

Gene correlation

To discover the correlation of the gene list generated, it was further queried with CRC databases, namely gene browser for CRC (gbCRC) at http://gbcrc.bioinfo-minzhao.org/ and CRC for the gene database (CoReCG) at http://lms.snu.edu.in/corecg/gene. Intersection analysis between the genes obtained and the previous gene list provided a unique gene list and a lead of their association with CRC.

Identification of genes associated with CRC

DAVID database (Database for Annotation, Visualization, and Integrated Discovery) accessed at http://david.abcc.ncifcrf.gov/ gives a high-throughput and attractive data collection condition and merges functional genomic annotations with intuitive graphical representations encouraging the transition between genomic information and the biological meaning.[22] According to Liu et al.,[23] using this database, GENETIC_ASSOIATION_DB_DISEASE analysis was carried out to identify genes associated with CRC from differentially expressed genes. The genes associated with CRC were defined as differentially expressed genes that were significantly related to CRC (final gene list; $p < 0.05$).[24]

Analysis of protein–protein interaction network and identification of hub genes for CRC

STRING database (Search Tool for the Retrieval of Interacting Genes) accessed at http://string-db.org/ is a unique tool, equipped for providing a comprehensive view of all the known and predicted interactions and associations among proteins.[25] To clarify the interactions in the final gene list, STRING online software was used to construct a protein–protein interaction (PPI) network using the CRC genes, and the network was visualized using Cytoscape open-source software at http://www.cytoscape.org/.[26] In the interaction network, the genes in the network served as "nodes" and the link connecting two nodes represents a pairwise protein interaction. The degree of a node corresponds to the number of interactions that the protein is in possession of. The nodes with the highest degree of connection were considered the "hub" genes in the interaction network[23] at a confidence level of 0.9.

Gene prioritization

cBioPortal database accessed at http://www.cbioportal.org/index.do was used for Cancer Genomics exploration in this study. The genomic datasets were queried using cBioPortal with the option to query single cancer study and also to query the microRNA target genes across cancer studies with the aim of viewing the relevant genomic alterations in the microRNA target genes in CRC samples. All the 225 cancer samples in this database were selected first followed by all the six samples specific for CRC and finally, targeted sequencing of 1134 samples from metastatic CRC samples. Mutations and Putative copy-number alterations were also selected as the genomic profiles. The microRNA target gene list was queried across all the 225 samples present followed by the six CRC samples, while the final query was done against targeted sequencing of 1134 samples from metastatic CRC samples (MSK, Cancer Cell, 2018).[27]

Gene Ontology and KEGG pathway enrichment analysis

Gene ontology (GO) accessed at http://www.geneontology.org/ is a database that provides vocabularies and classifications in relation to the molecular and cellular structures and functions for biological annotations of genes.[28] GO terms consist of three categories: biological process (BP), cellular component (CC), and molecular function (MF). The KEGG database at https://www.genome.jp/kegg/pathway.html contains sufficient information regarding the known metabolic pathways

and regulatory pathways and accelerates the mapping of genes to KEGG pathways for systemic analysis of gene functions.[29] To provide an insight into the precise biological function and signaling pathways involved with the CRC genes identified in this study, GO and KEGG pathway enrichment analysis was performed for the differentially expressed gene list previously obtained. The *p*-value represents the probability that the number of genes selected for any of the three domains occurred by chance. In general, the lower the *p*-value, the greater the likelihood that the terms are significant. For this reason, the GO term selected was at false discovery rate (FDR) of <0.01 (*p*-value < 0.01) using Benjamini–Hochberg procedure,[30] and KEGG pathways with $p < 0.05$ were eliminated.

Prognosis and expression analysis

In cancer research, the relationship between a gene and clinical result proposes the underlying etiology of the disease and therefore can motivate further studies. PROGmiR accessed at http://xvm145.jefferson.edu/progmir/ is a database for identifying prognostic microRNA biomarkers in multiple cancers using publicly available data.[31] SurvMicro accessed at http://bioinformatica.mty.itesm.mx:8080/Biomatec/Survmicro.jsp is a database used for the assessment of microRNA-based prognostic signatures for cancer clinical outcomes by multivariate survival analysis,[32] SurvExpress accessed at http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp is an online biomarker validation tool and database for cancer gene expression data using survival analysis,[33] and PrognoScan accessed at http://dna00.bio.kyutech.ac.jp/PrognoScan/ provides a powerful platform for evaluating potential tumor markers and therapeutic targets in other to correlate cancer research. The database also serves as a tool for meta-analysis of the prognostic value of genes,[34] and dbDEMC accessed at http://www.picb.ac.cn/dbDEMC/ is a database used to assess the differential expression of microRNAs in human cancer. In this study, the prognostic and expression analysis were performed on all the candidate microRNAs and their target genes (generated from cBioPortal). PROGmiR and SurvMicro were used for microRNA prognostic analysis, while SurvExpress and dbDEMC were employed for their expression analysis. For the target genes, PrognoScan and FIREBROWSE were used for both the prognostic and the expression analysis.

Results

The following databases provide information regarding experimentally validated microRNAs as well as their involvement in CRC: miRCancer, miR2Disease, HMDD, and dbDEMC2. The databases were used to download all microRNAs associated with CRC. A total of 2024 microRNAs were retrieved from the databases and subjected to CD-HIT-EST for duplicate and redundancy removal. At the end of this analysis, a total of 125 unique microRNAs with involvement in CRC were obtained. These 125 microRNAs were used as the query dataset. A total of 2694 microRNA sequences were downloaded from miRBase (total microRNAs); after the removal of duplicates and redundancies, 2226 microRNAs were found to be unique and this list was used as the reference dataset. The two datasets were then saved as FASTA sequences for similarity search.

Sequence similarity search for candidate microRNAs

The significance of BLASTN is to discover regions of sequence similarity, which will yield functional and evolutionary clues about the structure and function of a novel sequence. Also, the polycistronic structure of microRNA cluster genes sets them apart from most protein-coding genes in animals. This confers upon them a unique ability and specificity for widespread gene regulation in the complex molecular networks for development and disease.[35] Clustering of microRNAs has been reported to relate functions by regulating processes in a biological environment.[36] In silico experiment relating functions through clustering has also been proven to be a useful tool in Biotechnology.[37,38] In this study, the result of BLASTN and CD-HIT-EST-2D produced 26 and 43 microRNAs, respectively, from 125 validated query sequences and 2226 total microRNA sequences as the reference microRNAs. The 26 microRNA list obtained from BLASTN was based on the parameters of the expected value of 1×10^{-3}, word size of 7, and similarity between 90% and 99%. The result of the CD-HIT-EST-2D obtained was based on a threshold of 0.90 and a word size of 7. After intersection analysis (Figure 2), a total of 16 microRNAs were found to be unique to both CD-HIT-EST-2D and BLASTN. This list was further prioritized by text mining and the novelty was confirmed in the PubMed database. The final result of five microRNAs was confirmed to be candidate microRNAs used for CRC (Table 1).

Target genes prediction

Five candidate microRNAs were submitted using their IDs alongside with their clusters (validated microRNAs) to three target prediction tools, namely human TargetScan, miRDB, and mirDIP, to generate lists of individually predicted genes. The common genes from both the candidate and the validated microRNAs obtained from intersection analysis were saved in an excel file after duplicate removal. From the result, a total of 6664 target genes were identified for the five microRNAs sharing 5829 unique genes (Figure 3).

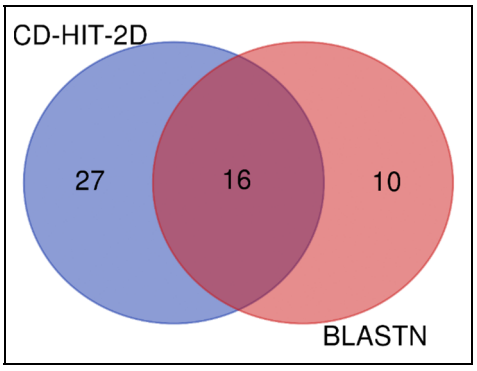


Figure 2. Number of microRNAs obtained using sequence similarity search with 90%–99% sequence similarities.

CRC gene association

The gene browser for CRC (gbCRC) and CRC database (CoReCG) were used to ascertain that there is a link/relationship between the predicted genes and the genes specific for CRC. A total of 2084 unique genes from these databases were used alongside the predicted microRNA target gene for intersection analysis. Figure 4 represents the analysis result. Intersected genes (96, 256, 60, 46, and 338 genes) were unique to the gene lists associated with microRNAs obtained above.

Enrichment analysis

Each of the lists of genes obtained (Figure 4) was submitted individually to DAVID to determine the enrichment in CRC. In total, 19 genes from the first microRNA were annotated to CRC, 28 genes were found in the second microRNA, while 8, 6, and 49 genes were enriched in CRC for third, fourth, and fifth candidate microRNAs. In summary, 110 genes were annotated to CRC and 82 of these genes were unique.

Gene prioritization

The microRNA target gene list was queried against all the 225 clinical data available in cBioPortal and the

Table 1. Candidate microRNAs and their clusters.

Candidate microRNA	Validated microRNA	FASTA sequences
miR-1	hsa-miR-193a-5p	>hsa-miR-193a-5p MIMAT0004614 UGGGUCUUUGCGGGCGAGAUGA
miR-2	hsa-miR-450b-3p	>hsa-miR-450b-3p MIMAT0004910 UUGGGAUCAUUUUGCAUCCAUA
miR-3	hsa-miR-501-3p	>hsa-miR-501-3p MIMAT0004774 AAUGCACCCGGGCAAGGAUUCU
miR-4	hsa-miR-501-3p	>hsa-miR-501-3p MIMAT0004774 AAUGCACCCGGGCAAGGAUUCU
miR-5	hsa-miR-513a-3p	>hsa-miR-513a-3p MIMAT0004777 UAAAUUUCACCUUUCUGAGAAGG

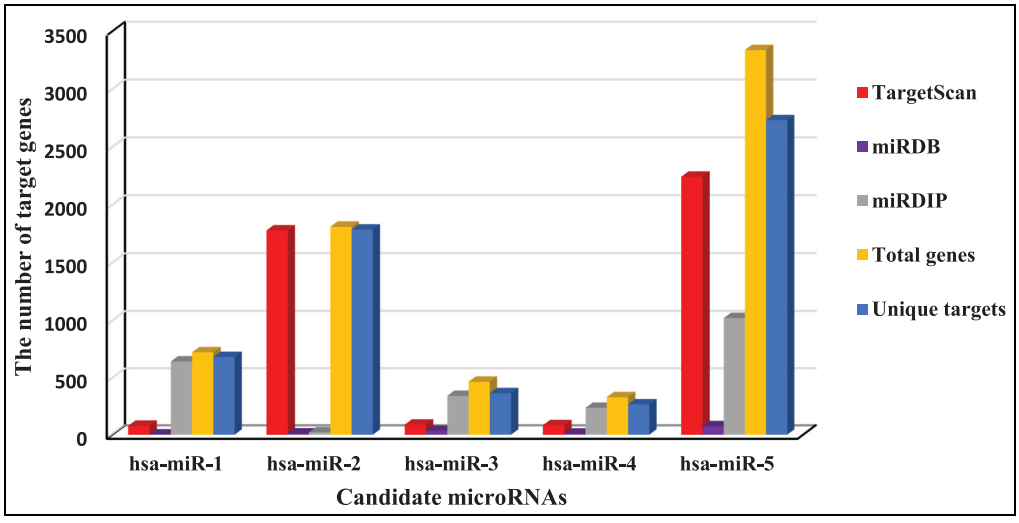


Figure 3. Number of genes targeted by the five microRNAs and relevant databases.

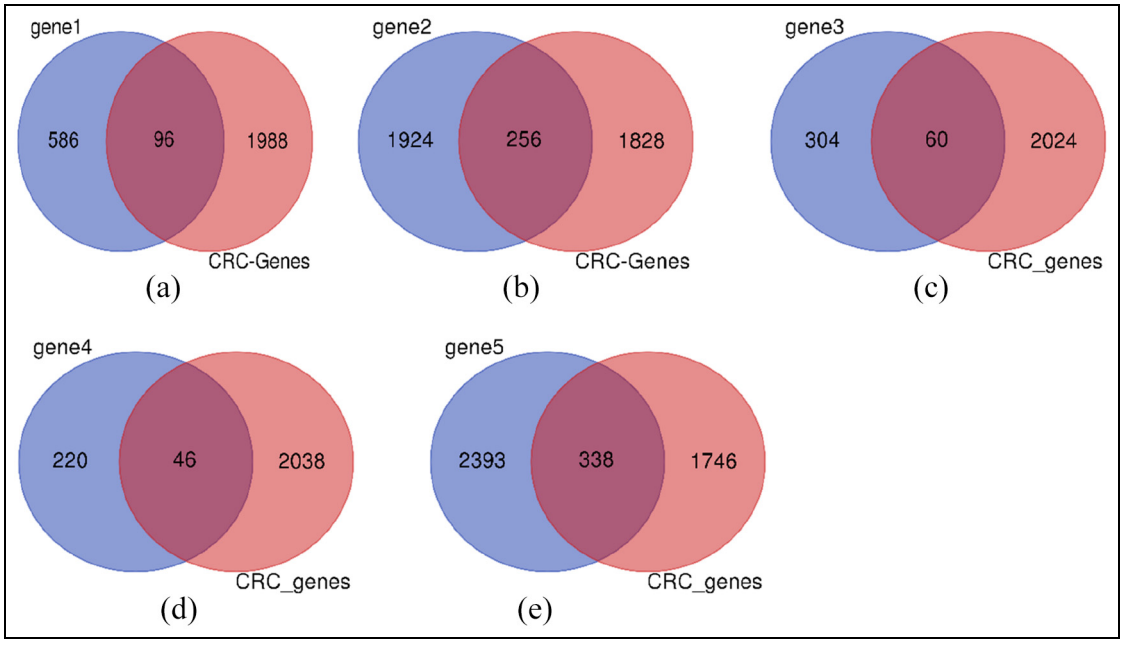


Figure 4. (a–e) Number of intersecting genes between the candidate microRNA target genes and CRC verified genes. Genes 1–5: target genes generated by the combination of the three target prediction tools. CRC_genes: unique genes expressed in CRC generated by CoReCG and gbCRC.

result was presented in a chart showing the frequency of alteration in descending order (Figure 7). It was further queried in the six CRC samples, while the final query was done against targeted sequencing of 1134 samples from metastatic CRC samples (MSK, Cancer Cell, 2018).[27] A list of 17 genes was obtained as the final gene list for the expression and prognostic analysis. These genes were selected based on the alteration frequency of $\geqslant 2\%$.

Prognostic and expression analysis of the candidate microRNAs

The five candidate microRNAs submitted to PROGmiRV2 and dbDEMC databases returned no result for prognostic and expression analysis, respectively, in CRC. Considering that these microRNAs are not associated with CRC as observed within the databases and text mining, the results in PROGmiRV2 strengthen the notion that these microRNAs are potential candidates for CRC. Has-mir-145 was further used as a control.

Expression analysis for microRNA target genes (SurvExpress)

The figure shows the result from the microRNA prioritized target gene list for CRC in SurvExpress database. Box plots across risk groups showed the expression levels and *p*-values resulting from a *t*-test of the different expression between high-risk (red) and low-risk (green) groups in CRC patients.

Prognostic analysis of the target gene list using PrognoScan

The data analysis utilizes the *p*-min approach to assess the cut-off point or the quantitative prognostic factor in continuous gene expression data measurement for patients grouping.[39] Patients are ordered by expression value of a given gene followed by the division of the expression group at all potential cut-off points into high and low, and the log-rank test estimates the risk difference of the two groups. The optimal cut-off point that gives the most pronounced *p*-min was chosen. Although this approach causes inflation of a type 1 error due to multiple correlated testing,[40–42] the *p*-value correction is conducted to reduce the error type rate.[43] For the microRNA target genes, the cut-off point determination and prognostic value assessment were applied to all possible combinations of dataset, probe, and endpoint.

Discussion

CRC still remains a major public health issue and a life-threatening disease.[44] The study was carried out to determine candidate microRNAs and their associated target genes that can serve as potential diagnostic biomarkers involved in CRC. A total of five candidate

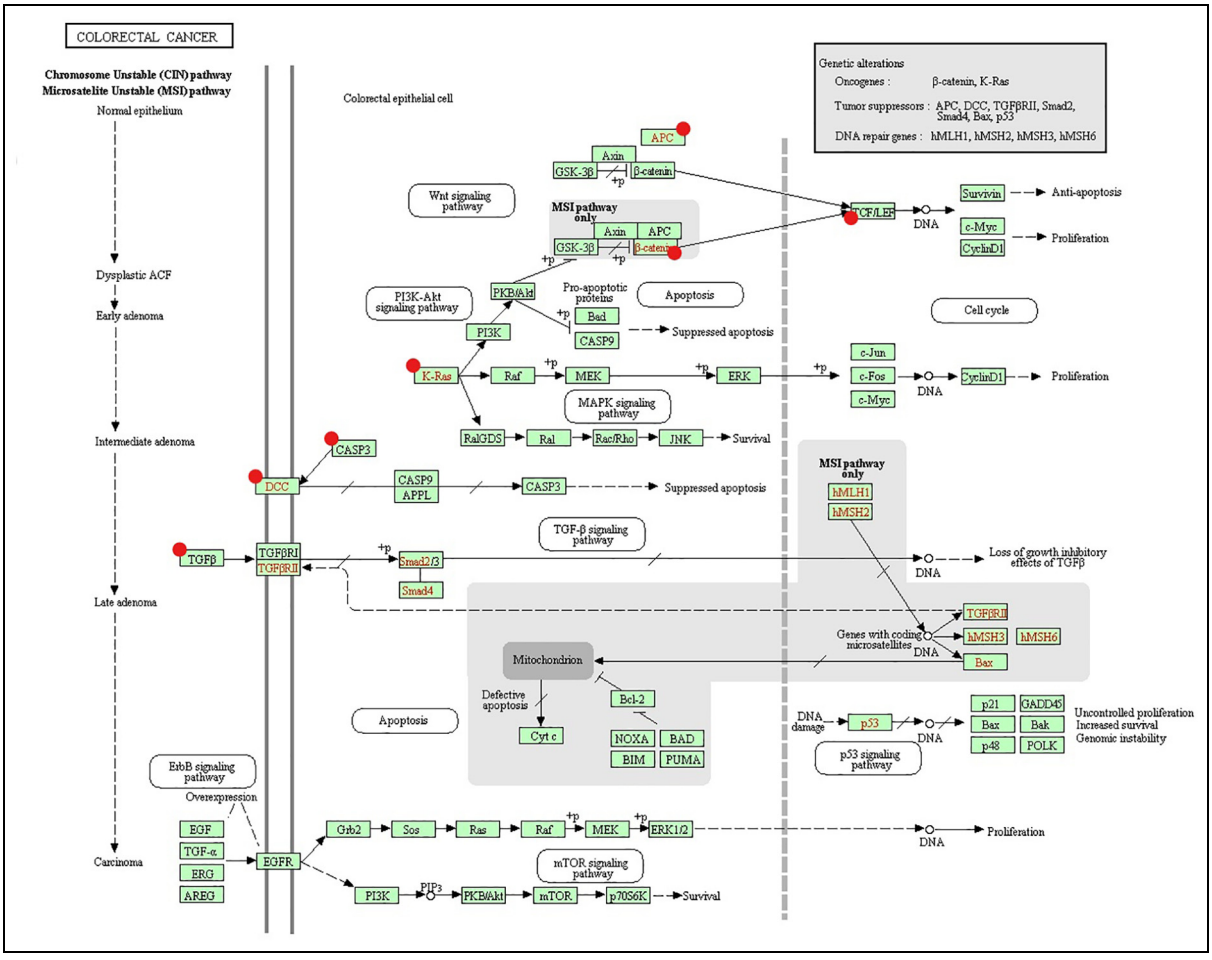


Figure 5. KEGG pathways enriched with CRC genes.
Genes indicated as red dots are candidate microRNA target genes involved in CRC.

microRNAs were screened out using sequence similarity search between four experimentally validated CRC microRNA databases and total microRNAs from miRBase (Figure 1 and Table 1). Using three microRNA target prediction tools, 5829 differentially expressed genes were discovered and further screened through two CRC databases and their enrichment was analyzed in DAVID to obtain a unique list of 82 genes (Figure 3). In a nutshell, the five candidate microRNAs obtained can regulate 82 differentially expressed genes in cancer subtype (Figure 4).

KEGG pathway analysis at FDR of 1.2×10^{-9} showed that the microRNA target genes were enriched in various pathways, including FoxO signaling pathway, p53 signaling pathways, apoptosis, microRNAs in cancer, CRC, pathways in cancer, proteoglycans in cancer, GnRH signaling pathway, Rap1 signaling pathway, p13k-Akt signaling pathway, and prostate cancer at $p < 0.01$. This suggests that abnormal pathways would be a significant cause of CRC[45–48] (Figure 5). In the PPI network, CTNNB1 (beta-catenin) and epidermal growth factor receptor (EGFR) were discovered as hub genes, which contain the highest number of interactions with other CRC genes (Figure 6).

CTNNB1 is a key downstream component of the canonical Wnt signaling pathway.[49] It binds with AXIN1, AXIN2, APC, CSNK1A1, and GSK3B to promote phosphorylation and ubiquitination of CTNNB1 through BTRC and its degradation by the proteasome in the absence of Wnt.[50] However, it is not ubiquitinated in the presence of Wnt and accumulates in the nucleus and acts as a coactivator for transcription factors of the T-cell factor/lymphoid enhancer-binding factor (TCF/LEF) family which activates Wnt response genes.[51] Dysregulation of the Wnt signaling pathway promotes cell survival, inhibits cell death and differentiation, and activates the development of gastrointestinal polyps and carcinoma.[52] *APC* gene is a negative regulator of CTNNB1, and up to 80% of CRCs have mutations in the *APC* gene resulting in a truncated protein.[53] The mutation of this gene occurs in various cancer types.[54] These mutations alter the

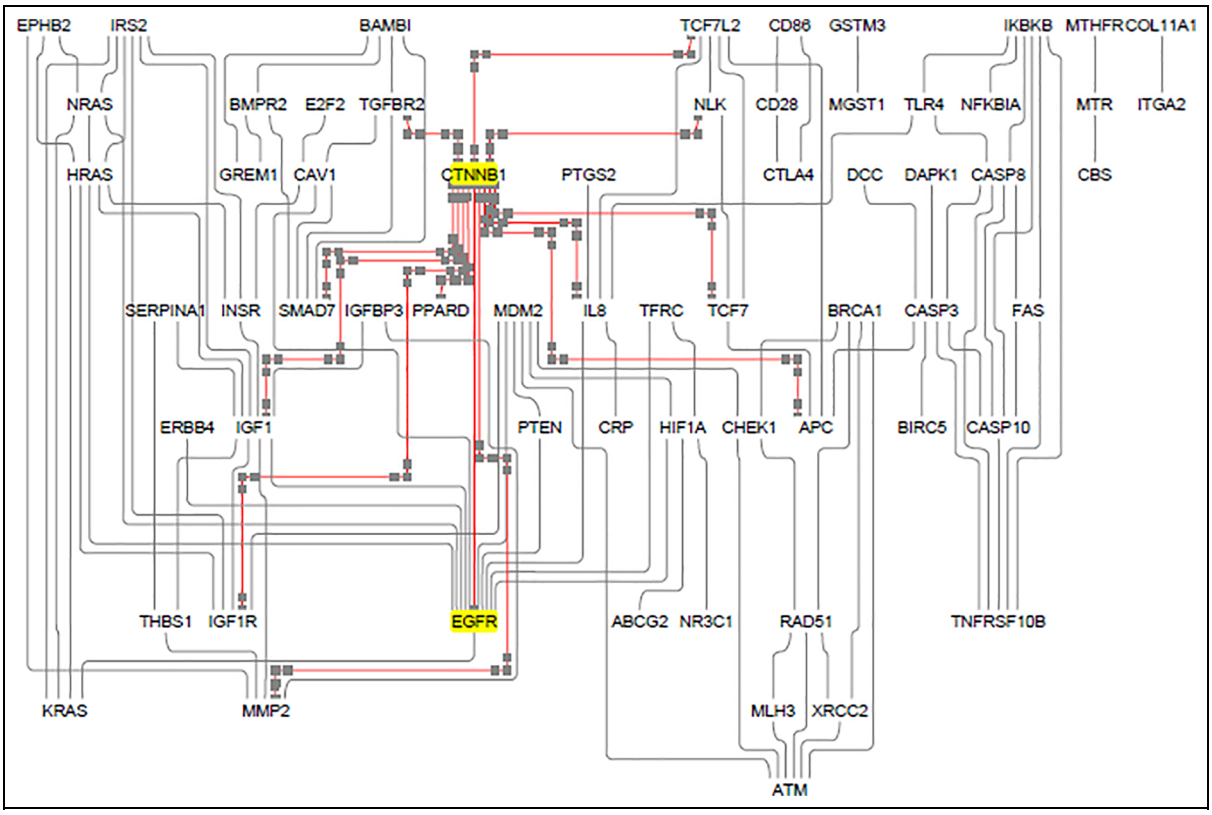


Figure 6. "Hub" genes of prioritized microRNA targets involved in CRC.
Yellow nodes represent the gene hubs. PPI represents protein–protein interaction. A gray line indicates the interaction between two genes.

spatial characteristics of the protein, leading to nuclear transcriptional network reprogramming. The result of the change includes increased cell proliferation, enhanced immunosuppression, and disruption of metabolic regulation.[54] Reports suggested that mutations in *APC* can also correlate with high expression levels of CTNNB1, whereas wild-type *APC* expression can reduce CTNNB1 levels in CRC cells.[55,56] Irrespective of the expression level of this gene, it is regarded as an important indicator of malignancy.[57,58]

The EGFR has been suggested to play a vital role in promoting cell growth.[59] The protein tyrosine kinase is the most frequently mutated domain with L861Q as the most common alteration. The major type of mutation found in this gene is missense. The EGFR L861Q mutation is known to be oncogenic and the biological effect is loss of function.[60] Existing literature reported that overexpression of EGFR is estimated to be 60%–80% of the tumor and is associated with poor prognosis in CRC.[61] The EGFR is a transmembrane glycoprotein and receptor tyrosine kinase that is encoded by the *c-erbB-1* proto-oncogene.[62] This gene is overexpressed in many types of cancers, specifically CRC.[61] EGFR is estimated to be overexpressed in 60%–80% of tumors and is associated with a poor prognosis.[63] The expression and prognostic analysis of this gene showed that it was downregulated in CRC and the prognostic value is significant. Spano et al.[64] revealed that EGFR remains a controversial prognostic factor; the expression may play an important role in a decision to initiate treatment. Another study also confirmed that the expression of this gene is implicated in CRC pathogenesis.[65] The "hub" genes of these candidate microRNAs, therefore, may be a potential target for CRC treatment.

Using the cBioPortal, the genomic exploration of microRNA target genes in CRC clinical data was queried against all the total cancer samples available and total CRC data. The 82 enriched genes were prioritized on the bases of genetic alteration/frequency of alteration in CRC to 17 genes (above 2%; Figure 7).

GO enrichment analysis of the prioritized target genes revealed the associations of the microRNA target genes in the three components of GO (CC, MF, and BP). Plasma membrane region is the most significant GO term for the microRNA target genes. The plasma membrane participates in the regulation of DNA methylation.[66] Most of the genes linked to the cell surface. The microRNAs targeting these genes may then be further proven to act as a good potential diagnostic biomarker. Also, five out of the seven genes (*APC, GNAS, IGF1R, CASP8*, and *EGRF*) are also located in the plasma membrane protein complex (integral

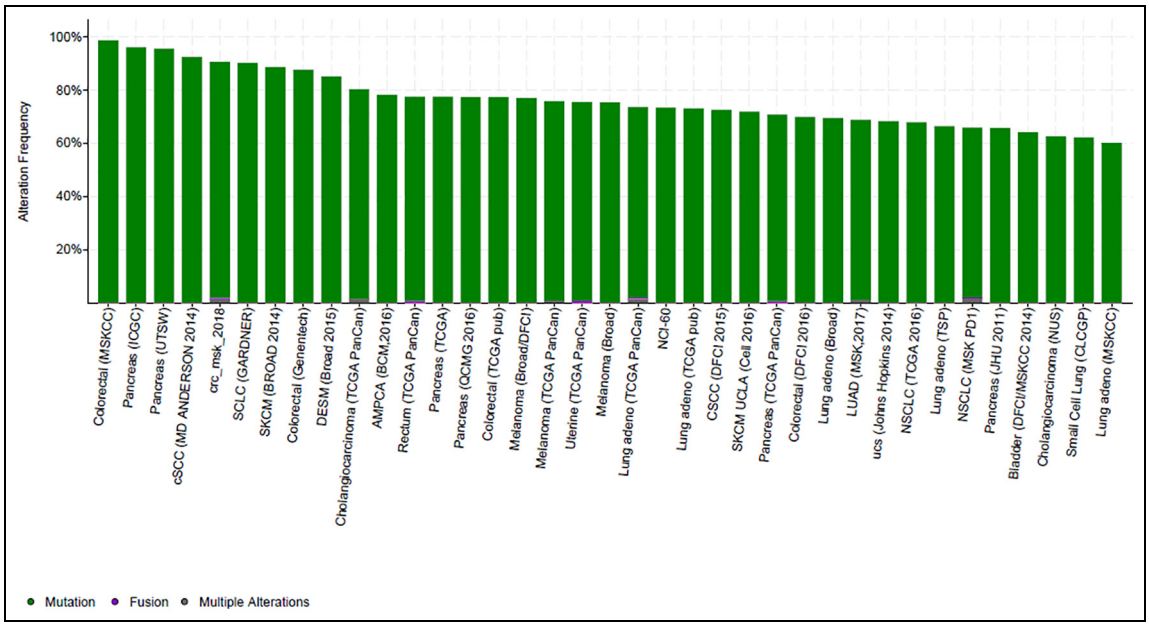


Figure 7. Cross-cancer alteration summary for prioritized microRNA target genes.

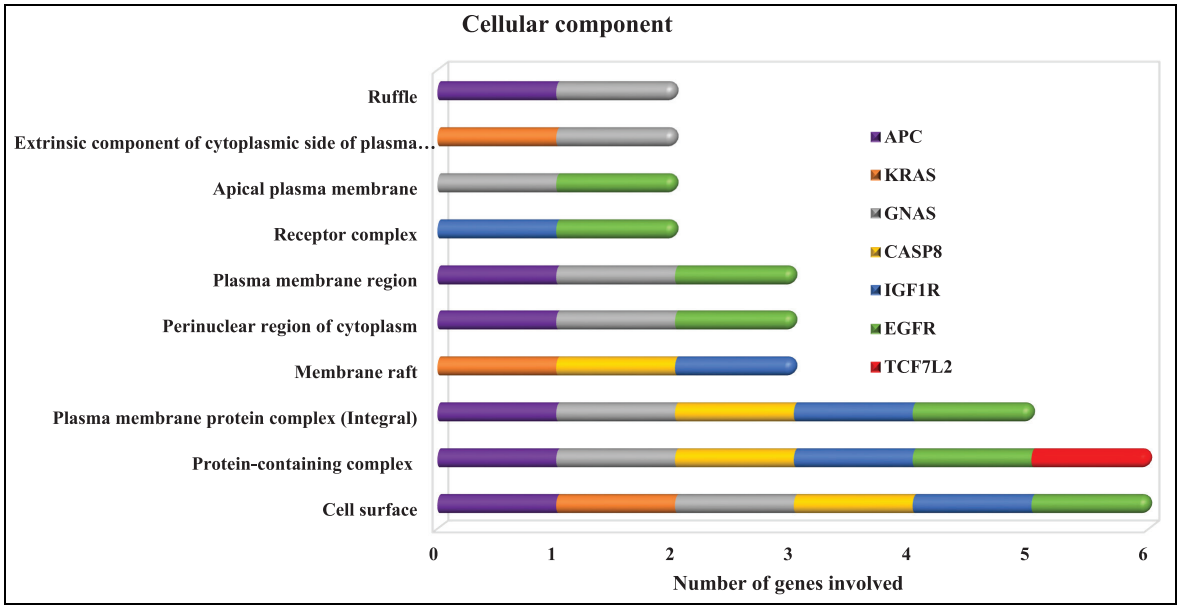


Figure 8. The result of the GO of microRNA target genes.
Cellular component: GO enrichment analysis results of microRNA target genes with $p < 0.01$. The *x*-axis represents the number of genes in the marked category; the *y*-axis indicates the descriptions (GO terms). Only functional categories with FDR <0.01 are shown. GO: Gene Ontology; FDR: false discovery rate.

protein). The microRNAs regulating these genes may also serve as therapeutic biomarkers for CRC. Other genes are found either in the receptor complex or the membrane raft. In the MF GO term, the genes were predominately associated with protein binding and receptor signaling binding. The microRNA target genes were also predominantly involved in a number of biological functions pertaining to regulations of MFs, signal transduction, cellular process, metabolic process, apoptotic process, cell differentiation, homeostatic process, cellular response, mitogen-activated protein (MAP) kinase activity, BP, peptidyl-tyrosine autophosphorylation, and catalytic activities. Wnt/CTNNB1 activation and malignant transformation of bowel

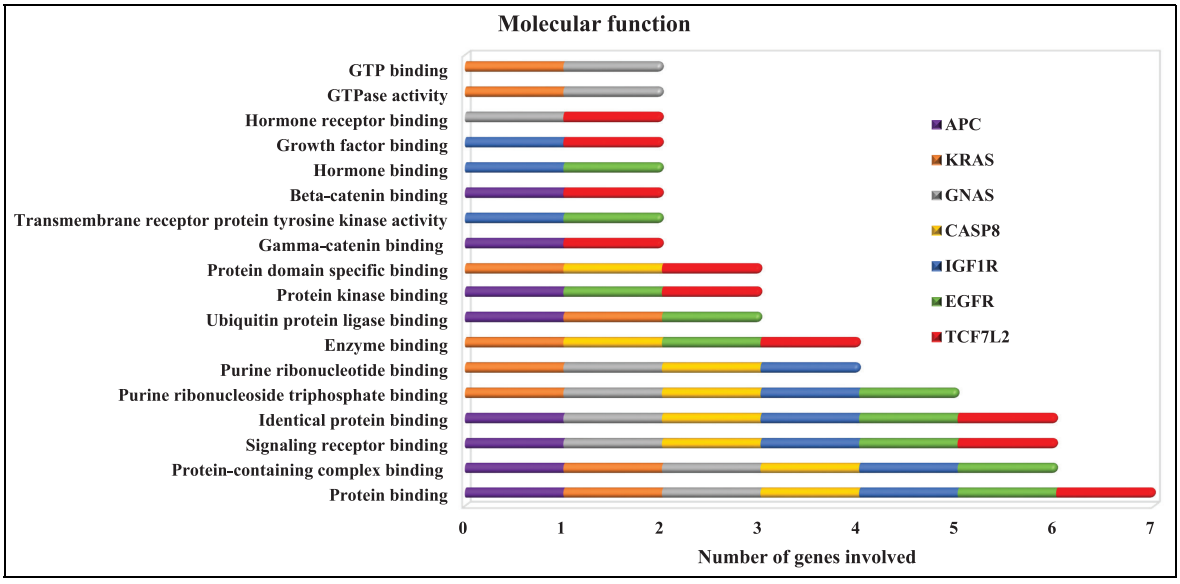


Figure 9. The result of the GO of microRNA target genes.
Molecular function: GO enrichment analysis results of microRNA target genes with $p < 0.01$. The *x*-axis represents the number of genes in the marked category; the *y*-axis indicates the descriptions (GO terms). Only functional categories with FDR <0.01 are shown. GO: Gene Ontology; FDR: false discovery rate.

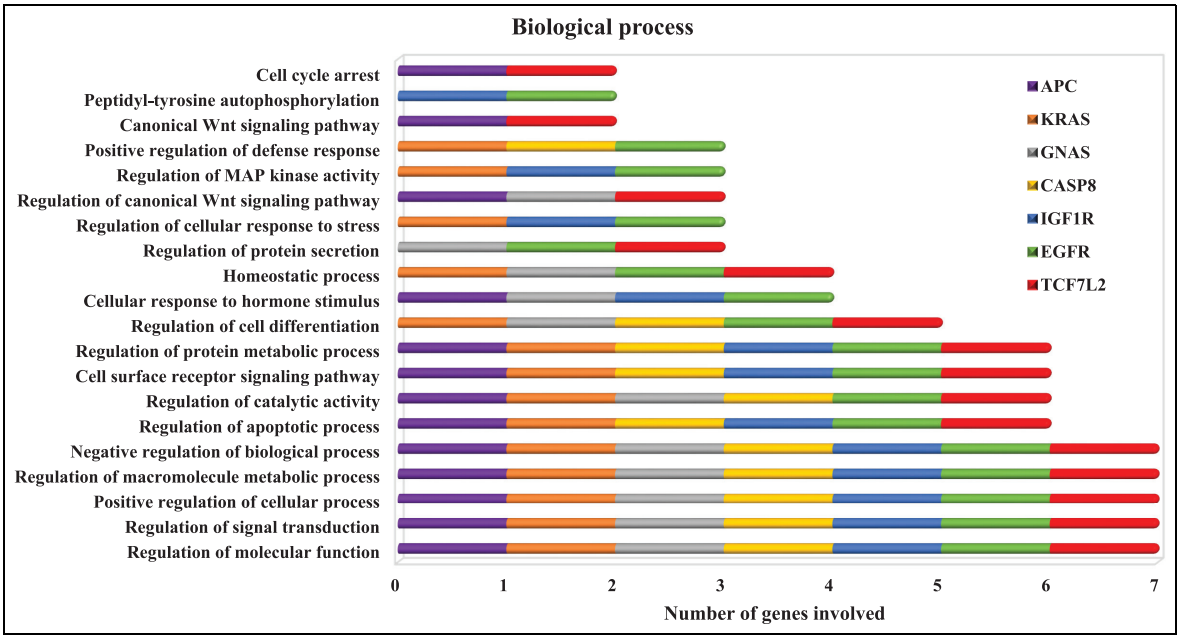


Figure 10. The result of the GO of microRNA target genes.
Biological process: GO enrichment analysis results of microRNA target genes with $p < 0.01$. The *x*-axis represents the number of genes in the marked category; the *y*-axis indicates the descriptions (GO terms). Only functional categories with FDR <0.01 are shown. GO: Gene Ontology; FDR: false discovery rate.

diseases are the two major causes of CRC. Both Wnt/CTNNB1[67,68] and inflammatory signaling pathway activation[69] can lead to intestinal epithelial disruption of homeostasis, for instance, if proliferation is increased, differentiation and apoptosis are decreased, in the intestinal tract[70] (Figures 8–10).

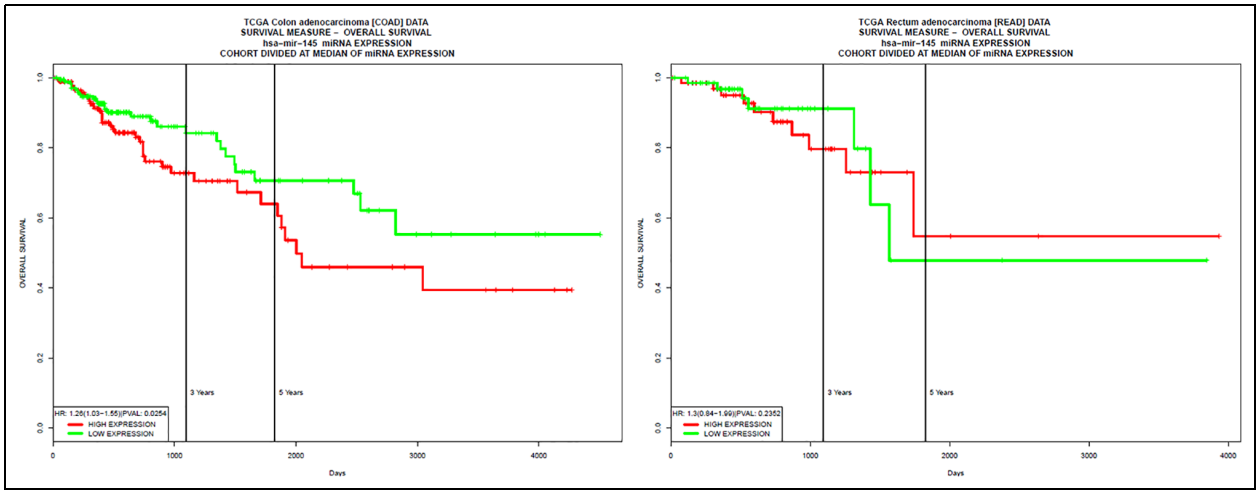


Figure 11. Kaplan–Meier plot for positive control microRNA generated by the PROGmiRV2 database for colon and rectum adenocarcinoma.

Prognostic and expression analysis of the candidate microRNA and their targets

Has-mir-145 is one of the microRNAs extensively studied in CRC as well as in other cancer types.[71–74] For this reason, it was used as a positive control for both colon and rectum adenocarcinoma in PROGmiRV2 to confirm the standard of the tool for prognostic analysis. The graphical output in the Kaplan–Meier plot for its expression in both colon and rectum was given in Figure 11. These results confirmed that the candidate microRNAs do not have any link with CRC and as such, they are potential novel microRNAs.

Validation of genes as biomarkers to predict the clinical outcome is a major issue for prognosis in cancer study. The availability of large cancer datasets publicly as an important information source for in silico validation is a step further for solving the aforementioned issue. Although evaluating genes prognostic performance with available datasets for biologists and physicians is a difficult task and tedious for statistics and bioinformatics experts, it is important to determine the correlations and validations of survival biomarkers for clinical outcomes. For this reason, the SurvExpress biomarker database was employed to predict the clinical result and prognostic value of CRC metastasis genes.

The gene expression by risk group (Figure 12) showed a box plot of gene expression level against the microRNA prioritized genes. It was assessed whether gene expression levels of the prioritized genes differ between the high-risk (red) and low-risk (green) groups as well as the level of significance of expression at $p < 0.05$ utilizing the *t*-test. The results indicated that genes *ERBB4*, *CASP8*, *BRCA1*, *GNAS*, *EGFR*, *INSR*, *NRAS*, *KRAS*, *CTNNB1*, *TCF7 L2*, *INHBA*, *TGFBR2*, *IGF1R*, and *IRS2* showed a distinct difference between the low-risk and the high-risk groups. The difference in expression between risks groups for these genes can be useful for prognostic and predictive values for CRC. Therefore, this gene panel can be used to determine a patient's survival if the survival is affected by the expression of these genes.

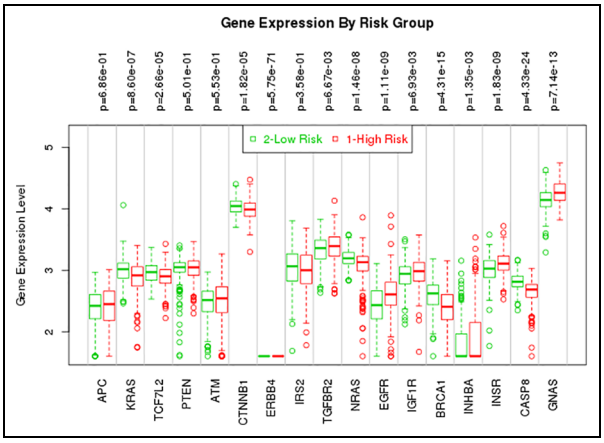


Figure 12. Result outputs of the SurvExpress database.

This section of the study assessed whether the expression of the prioritized list of the microRNA target gene predicted clinical outcome. Using the publicly available SurvExpress biomarker tool[33] that stratifies CRC patients into low-risk or high-risk groups based on differential gene expression, Kaplan–Meier survival curves were generated (Figure 13). Expression of the prioritized microRNA target genes significantly reduced recurrence-free survival in patients with CRC (Figure 12). This shows that the expression of the panel of genes is a prognostic indicator for survival in CRC patients.

Table 2. Prognostic analysis using PrognoScan (list of genes whose expression is associated with overall survival of patients with CRC).

Genes	MP-value	CP-value	Cox *p*-value	Probe ID (VMC)	Cut-off point
APC	0.001117	0.027970	0.032977	215310_at	0.11
KRAS	0.000063	0.002316	0.008760	204010_s_at	0.89
TCF7L2	0.000009	0.000421	0.002389	236094_at	0.15
EGFR	0.000061	0.002265	0.045184	211607_x_at	0.47
IGF1R	0.005914	0.107944	0.032731	243358_at	0.11
CASP8	0.012184	0.187135	0.030928	207686_s_at	0.58
GNAS	0.023589	1.57	0.029136	214548_x_at	0.89

CRC: colorectal cancer; MP-value: minimum *p*-value; CP-value: correlated *p*-value; VMC: Vanderbilt Medical Center.
Dataset: GSE17537; HRs (log2 ratio) with corrected *p*-value < 0.05 are shown. From the list of genes analyzed by this database, only the statistically significant genes at $p < 0.05$ were presented.

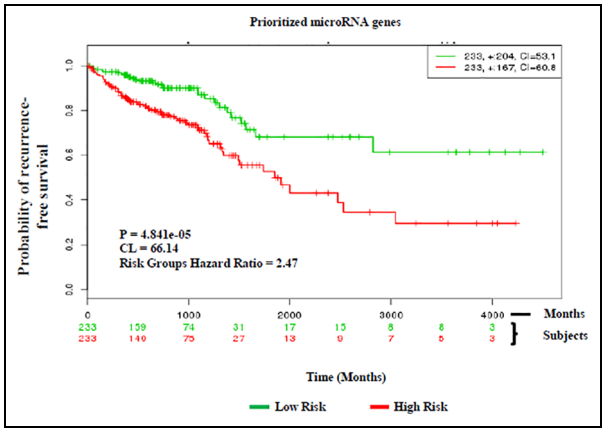


Figure 13. The expression of prioritized microRNA target genes predicts recurrence-free survival in patients with CRC. The Kaplan–Meier (KM) survival curves generated using the SurvExpress database showed the ability of the expression of the genes to predict recurrence-free survival outcome in patients with CRC. The insets in top right represent a number of individuals, number censored, and concordance index (CI) of each risk groups and " + " represents censoring samples. High-risk and low-risk groups were presented in red and green accordingly.

PrognoScan depicted the statistical significance of the expression of 7 out of 17 genes which were positively associated with CRC showing their contributions as well as prognosis to CRC (Table 2 and Figure 12). On a whole, prognostic and expression analysis of the candidate microRNAs confirmed that there is no link to CRC and could serve as potential candidate microRNAs while the prognostic and expression analysis of their target genes concluded that seven microRNAs, namely APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and GNAS, at $p < 0.05$ are statistically significant and showed good prognostic values with clear implications in CRC. The expression of these genes either collectively or individually discriminates between high-risk and low-risk CRC groups making the microRNAs potential biomarker in CRC diagnosis.

Conclusion

Using in silico approach, this study identified five candidate microRNAs alongside seven significant target genes. The patterns of expression obtained in their target genes relative to their microRNAs and their prognostic values could be inferred that patients with alterations in the microRNA prioritized target genes have significantly better overall survival than patients without these alterations. These could be further exploited and could potentially serve as a resource for explicitly selecting targets for diagnosis, drug development, and management of CRC. Although validation studies are ongoing to conclude the biological fitness of these findings, the study also indicated that the identified microRNAs and hub genes (CTNNB1 and EGFR) stimulate a better understanding of the molecular mechanisms underlying the development of CRC and might be used as molecular targets and potential diagnostic biomarkers for the treatment of the cancer subtype.

Acknowledgements

The authors would like to thank the Plant Omics Laboratory and the Bioinformatics Research Group of the University of the Western Cape for their usual support and also, Dr. Taiwo Akinsoji (MBBS, MPH) of University of Illinois at Springfield for co-proofreading the final manuscript.

Author contributions

All authors have made significant contributions to the submission of the article. A.O.F. conceived the concept and the design of the manuscript; A.K. and A.P. provided the necessary software required to carry out the analysis. The analysis and data interpretations were done by both A.O.F. and A.P., while A.K. drafted the rough draft and also substantively revised the manuscript. Finally, all authors read and approved the submitted version of the manuscript for publication. Also, they agreed to be personally accountable for their personal contributions and ensured that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately

investigated, resolved, and the resolution documented in the literature.

Data availability

The datasets and the clinical data were obtained from the online databases as described above in the methods, and their websites are as follows: http://www.mirbase.org/ for reference dataset, microRNAs associated with CRC at cancer were http://www.picb.ac.cn/dbDEMC/, http://www.mir2disease.org/, http://www.cuilab.cn/hmdd, and http://mircancer.ecu.edu/. The clinical data exploited were derived from cBioPortal database accessed at http://www.cbioportal.org/index.do.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Adewale Oluwaseun Fadaka https://orcid.org/0000-0002-3952-2098

References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin* 2011; 61: 69–90.
2. Fadaka AO, Pretorius A and Klein A. Biomarkers for stratification in colorectal cancer: microRNAs. *Cancer Control* 2019; 26(1): 862784.
3. Corté H, Manceau G, Blons H, et al. MicroRNA and colorectal cancer. *Digest Liver Dis* 2012; 44: 195–200.
4. Gonzalez-Pons M and Cruz-Correa M. Colorectal cancer biomarkers: where are we now? *Biomed Res Int* 2015; 2015: 149014.
5. Armaghany T, Wilson JD, Chu Q, et al. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res* 2012; 5: 19–27.
6. Fadaka AO, Ojo BA, Adewale OB, et al. Effect of dietary components on miRNA and colorectal carcinogenesis. *Cancer Cell Int* 2018; 18: 130.
7. Okugawa Y, Grady WM and Goel A. Epigenetic alterations in colorectal cancer: emerging biomarkers. *Gastroenterology* 2015; 149(5): 1204–1225.e12.
8. Reddy KB. MicroRNA (miRNA) in cancer. *Cancer Cell Int* 2015; 15: 38.
9. Slaby O, Svoboda M, Michalek J, et al. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. *Mol Cancer* 2009; 8: 102.
10. Madhavan D, Cuk K, Burwinkel B, et al. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. *Front Genet* 2013; 4: 116.
11. Radfar H. Computational prediction of target genes of microRNAs, 2014, https://tspace.library.utoronto.ca/bitstream/1807/44130/3/Hossein_Radfar_201406_PhD_thesis.pdf
12. Griffiths-Jones S. The microRNA registry. *Nucleic Acids Res* 2004; 32: D109–D111.
13. Cui H, Zhang Y and Ren F. dbDEMC2.0: a database of differentially expressed miRNAs in human cancers v2.0. *China J Mod Med* 2014; 24: 77–79.
14. Jiang Q, Wang Y, Hao Y, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res* 2008; 37: D98–D104.
15. Xie B, Ding Q, Han H, et al. miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics* 2013; 29(5): 638–644.
16. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990; 215: 403–410.
17. Li W and Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006; 22: 1658–1659.
18. Friedman RC, Farh Burge CB and Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009; 19(1): 92–105.
19. Grimson A, Farh Johnston WK, Garrett-Engele P, et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007; 27(1): 91–105.
20. Nam J-W, Rissland OS, Koppstein D, et al. Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol Cell* 2014; 53(6): 1031–1043.
21. Lewis BP, Burge CB and Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; 120(1): 15–20.
22. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003; 4(5): P3.
23. Liu F, Ji F, Ji Y, et al. In-depth analysis of the critical genes and pathways in colorectal cancer. *Int J Mol Med* 2015; 36(4): 923–930.
24. Huang DW, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2008; 4: 44–57.
25. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2012; 41: D808–D815.
26. Saito R, Smoot ME, Ono K, et al. A travel guide to Cytoscape plugins. *Nat Methods* 2012; 9(11): 1069–1076.
27. Yaeger R, Chatila WK, Lipsyc MD, et al. Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell* 2018; 33(1): 125–136.e3.
28. Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004; 32: D258–D261.
29. Kanehisa M, Goto S, Sato Y, et al. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2011; 40: D109–D114.
30. McDonald JH. *Handbook of biological statistics*. Baltimore, MD: Sparky House Publishing, 2009.
31. Goswami CP and Nakshatri H. PROGmiR: a tool for identifying prognostic miRNA biomarkers in multiple cancers using publicly available data. *J Clin Bioinforma* 2012; 2(1): 23.

32. Aguirre-Gamboa R and Trevino V. SurvMicro: assessment of miRNA-based prognostic signatures for cancer clinical outcomes by multivariate survival analysis. *Bioinformatics* 2014; 30(11): 1630–1632.
33. Aguirre-Gamboa R, Gomez-Rueda H, Martinez-Ledesma E, et al. SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. *PLoS ONE* 2013; 8(9): e74250.
34. Mizuno H, Kitada K, Nakai K, et al. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics* 2009; 2: 18.
35. Olive V, Jiang I and He L. mir-17-92, a cluster of miRNAs in the midst of the cancer network. *Int J Biochem Cell Biol* 2010; 42(8): 1348–1354.
36. Guo L, Yang S, Zhao Y, et al. Global analysis of miRNA gene clusters and gene families reveals dynamic and coordinated expression. *Biomed Res Int* 2014; 2014: 782490.
37. Dews M, Homayouni A, Yu D, et al. Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nat Genet* 2006; 38(9): 1060–1065.
38. Xu J and Wong C. A computational screen for mouse signaling pathways targeted by microRNA clusters. *RNA* 2008; 14(7): 1276–1283.
39. Abel U, Berger J and Wiebelt H. CRITLEVEL: an exploratory procedure for the evaluation of quantitative prognostic factors. *Methods Inf Med* 1984; 23(3): 154–156.
40. Altman DG, Lausen B, Sauerbrei W, et al. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994; 86: 829–835.
41. Mazumdar M and Glassman JR. Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Stat Med* 2000; 19(1): 113–132.
42. Hollander N and Schumacher M. On the problem of using "optimal" cutpoints in the assessment of quantitative prognostic factors. *Onkologie* 2001; 24(2): 194–199.
43. Miller R and Siegmund D. Maximally selected chi square statistics. *Biometrics* 1982; 38: 1011–1016.
44. Jemal A, Ward E and Thun M. Declining death rates reflect progress against cancer. *PLoS ONE* 2010; 5(3): e9584.
45. Kersten S, Desvergne B and Wahli W. Roles of PPARs in health and disease. *Nature* 2000; 405: 421–424.
46. Johnson SM, Gulhati P, Rampy BA, et al. Novel expression patterns of PI3K/Akt/mTOR signaling pathway components in colorectal cancer. *J Am Coll Surg* 2010; 210(5): 767–776.
47. Bienz M and Clevers H. Linking colorectal cancer to Wnt signaling. *Cell* 2000; 103(2): 311–320.
48. Huang Z, Yang Q and Huang Z. Identification of critical genes and five prognostic biomarkers associated with colorectal cancer. *Med Sci Monit* 2018; 24: 4625–4633.
49. MacDonald BT, Tamai K and He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* 2009; 17: 9–26.
50. Liu L-J, Xie S-X, Chen Y-T, et al. Aberrant regulation of Wnt signaling in hepatocellular carcinoma. *World J Gastroenterol* 2016; 22(33): 7486–7499.
51. Rao TP and Kuhl M. An updated overview on Wnt signaling pathways: a prelude for more. *Circ Res* 2010; 106(12): 1798–1806.
52. Hofacker IL. How microRNAs choose their targets. *Nat Genet* 2007; 39(10): 1191–1192.
53. Jacobsen A, Silber J, Harinath G, et al. Analysis of microRNA-target interactions across diverse cancer types. *Nat Struct Mol Biol* 2013; 20(11): 1325–1332.
54. Gao C, Wang Y, Broaddus R, et al. Exon 3 mutations of CTNNB1 drive tumorigenesis: a review. *Oncotarget* 2018; 9(4): 5492–5508.
55. Fodde R, Smits R and Clevers H. APC, signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer* 2001; 1(1): 55–67.
56. Enzo MV, Rastrelli M, Rossi CR, et al. The Wnt/β-catenin pathway in human fibrotic-like diseases and its eligibility as a therapeutic target. *Mol Cell Ther* 2015; 3: 1.
57. Polakis P. Wnt signaling in cancer. *Cold Spring Harb Perspect Biol* 2012; 4: a008052.
58. Xia J, Urabe K, Moroi Y, et al. Beta-catenin mutation and its nuclear localization are confirmed to be frequent causes of Wnt signaling pathway activation in pilomatricomas. *J Dermatol Sci* 2006; 41(1): 67–75.
59. Herbertson RA, Karapetis C, Price T, et al. Epidermal growth factor receptor (EGFR) inhibitors for metastatic colorectal cancer. *Cochrane Database Syst Rev* 2008; 6: CD007047.
60. Tsuchihashi Z, Khambata-Ford S, Hanna N, et al. Responsiveness to cetuximab without mutations in EGFR. *N Engl J Med* 2005; 353(2): 208–209.
61. Pabla B, Bissonnette M and Konda VJ. Colon cancer and the epidermal growth factor receptor: current treatment paradigms, the importance of diet, and the role of chemoprevention. *World J Clin Oncol* 2015; 6(5): 133–141.
62. Yarden Y and Sliwkowski MX. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* 2001; 2(2): 127–137.
63. Cohen RB. Epidermal growth factor receptor as a therapeutic target in colorectal cancer. *Clin Colorectal Cancer* 2003; 2: 246–251.
64. Spano J-P, Lagorce C, Atlan D, et al. Impact of EGFR expression on colorectal cancer patient prognosis and survival. *Ann Oncol* 2005; 16: 102–108.
65. Rego R, Foster N, Smyrk T, et al. Prognostic effect of activated EGFR expression in human colon carcinomas: comparison with EGFR status. *Br J Cancer* 2010; 102(1): 165–172.
66. Nautiyal S, Carlton VE, Lu Y, et al. High-throughput method for analyzing methylation of CpGs in targeted genomic regions. *Proc Natl Acad Sci USA* 2010; 107(28): 12587–12592.
67. Nelson WJ and Nusse R. Convergence of Wnt, β-catenin, and cadherin pathways. *Science* 2004; 303: 1483–1487.
68. Segditsas S and Tomlinson I. Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* 2006; 25(57): 7531–7537.
69. Grivennikov SI. Inflammation and colorectal cancer: colitis-associated neoplasia. *Semin Immunopathol* 2013; 35: 229–244.

70. Guo Y, Bao Y, Ma M, et al. Identification of key candidate genes and pathways in colorectal cancer by integrated bioinformatical analysis. *Int J Mol Sci* 2017; 18(4): 722.
71. Li S, Wu X, Xu Y, et al. miR-145 suppresses colorectal cancer cell migration and invasion by targeting an ETS-related gene. *Oncol Rep* 2016; 36(4): 1917–1926.
72. Sheng N, Tan G, You W, et al. MiR-145 inhibits human colorectal cancer cell migration and invasion via PAK4-dependent pathway. *Cancer Med* 2017; 6(6): 1331–1340.
73. Feng Y, Zhu J, Ou C, et al. MicroRNA-145 inhibits tumour growth and metastasis in colorectal cancer by targeting fascin-1. *Br J Cancer* 2014; 110(9): 2300–2309.
74. Akao Y, Nakagawa Y and Naoe T. MicroRNA-143 and -145 in colon cancer. *DNA Cell Biol* 2007; 26: 311–320.

CHAPTER 5.0

FUNCTIONAL PREDICTION OF CANDIDATE MICRORNAS FOR CRC MANAGEMENT USING *IN SILICO* APPROACH

ADEWALE OLUWASEUN FADAKA*, ASHLEY PRETORIUS, AND ASHWIL KLEIN

Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa.

Accepted/ published: *International Journal of Molecular Science* (2019).
DOI:10.3390/ijms20205190 **(Impact factor: 4.183)**

Statement of Contributions of Joint Authorship

Adewale Fadaka (PhD candidate).

Conceived the concept of the manuscript as well as the write up and compilation. Preparation of figures and tables.

Ashley Pretorius (Co-Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

Ashwil Klein (Principal Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

This Chapter is an exact copy of the journal paper below.

International Journal of
Molecular Sciences

MDPI

Article

Functional Prediction of Candidate MicroRNAs for CRC Management Using in Silico Approach

Adewale Oluwaseun Fadaka *, Ashley Pretorius and Ashwil Klein

Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Private Bag X17, Bellville, Cape Town 7535, South Africa; aspretorius@uwc.ac.za (A.P.); aklein@uwc.ac.za (A.K.)
* Correspondence: afadaka@uwc.ac.za; Tel.: +27-63-051-1928/+23-48-039-2420-52

Received: 29 August 2019; Accepted: 14 September 2019; Published: 19 October 2019

Abstract: Approximately 30–50% of malignant growths can be prevented by avoiding risk factors and implementing evidence-based strategies. Colorectal cancer (CRC) accounted for the second most common cancer and the third most common cause of cancer death worldwide. This cancer subtype can be reduced by early detection and patients' management. In this study, the functional roles of the identified microRNAs were determined using an in silico pipeline. Five microRNAs identified using an in silico approach alongside their seven target genes from our previous study were used as datasets in this study. Furthermore, the secondary structure and the thermodynamic energies of the microRNAs were revealed by Mfold algorithm. The triplex binding ability of the oligonucleotide with the target promoters were analyzed by Trident. Finally, evolutionary stage-specific somatic events and co-expression analysis of the target genes in CRC were analyzed by SEECancer and GeneMANIA plugin in Cytoscape. Four of the five microRNAs have the potential to form more than one secondary structure. The ranges of the observed/expected ratio of CpG dinucleotides of these genes range from 0.60 to 1.22. Three of the candidate microRNA were capable of forming multiple triplexes along with three of the target mRNAs. Four of the total targets were involved in either early or metastatic stage-specific events while three other genes were either a product of antecedent or subsequent events of the four genes implicated in CRC. The secondary structure of the candidate microRNAs can be used to explain the different degrees of genetic regulation in CRC due to their conformational role to modulate target interaction. Furthermore, due to the regulation of important genes in the CRC pathway and the enrichment of the microRNA with triplex binding sites, they may be a useful diagnostic biomarker for the disease subtype.

Keywords: microRNA; CRC; functional predictions; in silico analysis; triplex binding site

1. Introduction

With increasing incidence and mortality of cancers, colorectal cancer (CRC) remains one of the leading malignant cancers [1]. Currently, tumor resection followed by chemotherapy is one of the most effective treatments for CRC but recurrences are inevitable [2]. Approximately 45% of recurrences are in the first year after initial tumour removal, and over 90% of recurrences occur within four years [3]. In CRC, staging and pathological characteristics are one of the major predictors of diagnosis to enhance treatment options [4]. Thus far, there are limited biomarkers to predict CRC. The search for more biomarkers that are specific for staging is required.

Recent research suggests that microRNAs regulate many gene functions in human cancers [5], and these oligonucleotide sequences have been proposed as novel biomarkers for cancers [6]. Previous studies have shown that their expressions are altered in numerous types of cancers including CRC [7,8]. They also act as regulators of gene expression [9] through gene repression and/or mRNA degradation in many biological processes such as apoptosis, cell development, cell differentiation, cell proliferation,

and metabolism [10–12]. These biological processes are crucial in carcinogenesis [13]. Although several studies have reported an association between microRNAs and CRC development [14–17], the role of microRNAs forming triplexes with their targets to infer function in CRC has been barely explored. Several microRNAs have been reported to date but their laboratory functional determination in relation to CRC is challenging and time-consuming.

In silico approaches have been developed to connect sequences of microRNAs and their targets to infer function in Cancer studies. Biologists are mainly particular about the structural and functional properties of any newly derived sequence (protein or nucleotide). In silico predictions are therefore important for this discovery based on successful knowledge-based principles. These principles rely on the fact that the best way to predict the structure and/ or function is to find similar sequences in existing databases using the information about them to infer conclusions about properties of the new sequence. Algorithms have been developed and implemented as computer programs (local or web-based tools) to perform this function.

The genomic sequence of many higher eukaryotes is now complete, and the patterns of expression of thousands of genes under diverse conditions are known. This offers researchers the opportunity to identify and analyze the parts of a genome believed to be responsible for most transcription control, known as the promoters [18]. Short nucleotide sequences, most especially non-coding RNAs, have been reportedly used in the induction of DNA cleavage at specific sites with the aid of triplex formation between the RNA and the DNA [19,20]. The various triplex formation between these nucleotides has suggested a role in gene expression regulation [21,22]. The binding of DNA leading to the triplex bond formation through a hydrogen bond to the third strand has been experimentally discovered but the biological functions are still unknown [23]. The interaction of microRNAs with DNA have not been studied extensively [24,25]. Reports also suggested that the interactions between microRNAs and gene promoter regions may play a more direct role in regulating the efficiency of transcription of the gene involved [26–28].

One of the roles includes the methylation of CpG islands [29]. However, other mechanisms yet unidentified or not fully studied could be possible for the direct interaction of microRNAs with genes for transcription regulation. Since dsDNA is capable of forming triplex structures through interactions with DNA or RNA in the major groove of the DNA duplex, researchers postulated that microRNA may form triplex structures with duplex DNA through either Hoogsteen or reverse Hoogsteen hydrogen bonds, and thereby directly interacting with target DNA sequences in regulatory regions including gene promoters in the human genome, with the potential to alter gene function [30–32]. Bioinformatics approaches revealed that the mammalian genome is composed of several triplex formation binding sites [24,30,33,34]. This suggests that tethering RNA to specific genomic sites might guide RNA-associated regulatory proteins to establish an epigenetic landscape that facilitates or inhibits gene expression. In this study, the function of the candidate microRNAs was predicted for CRC management using an in silico approach. Given the ever-expanding number of microRNAs, understanding their functional aspects through sequences represents a promising research area.

2. Results

Secondary Structure of the Candidate MicroRNAs

The function of a given microRNA molecule may be determined by sequence or structure that it is most likely to fold into. It may also be governed by whether small sequence have the ability to fold into a particular substructure even if the substructure does not appear in any optimal fold.

3. Discussion

Genes implicated as targets for the candidate microRNAs have been experimentally validated and their associations with CRC confirmed by previous studies [35–43]. Also, research has confirmed that the regulators of these targets have been experimentally determined and validated in other cancer

subtypes but not in CRC [44–54]. Therefore, the interaction of both the microRNAs and targets could suggest their involvement in CRC. The prognostic, diagnostic, and expression analysis performed on these genes together with the gene ontology analysis from our previous study suggest that they could be used as good diagnostic and therapeutic biomarkers in the detection and management of CRC.

3.1. Secondary Structure and the Thermodynamic Energies of the Candidate MicroRNAs

MicroRNAs have intrinsic potential to form secondary and tertiary structures by folding through base pairing. This folding is given primarily by the fact that certain nucleotides have an affinity for binding another nucleotide, for example (G-C and A-U/). Each nucleotide can only bind with one partner at a time and long stretches of rich nucleotide as a binding partner elsewhere on the sequence are favoured over short stretches or individual non-sequential pairing. These pairings compete with each other since any G in the sequence could pair with any C and similarly, any A with any U. The arrangement in Figure 1 may be optimal in terms of maximizing the length of the concurrent run and pair of nucleotides. These arrangements are only probabilistic and outside forces such as temperature and pH may induce a different pairing arrangement. The molecule may even transition randomly between numerous newly optimal structures based on the fluctuation in available energy. The mature sequences of the five candidate microRNAs were subjected to secondary structure prediction by Mfold and inspected manually against filtering criteria, as indicated by default to check for any discrepancies.

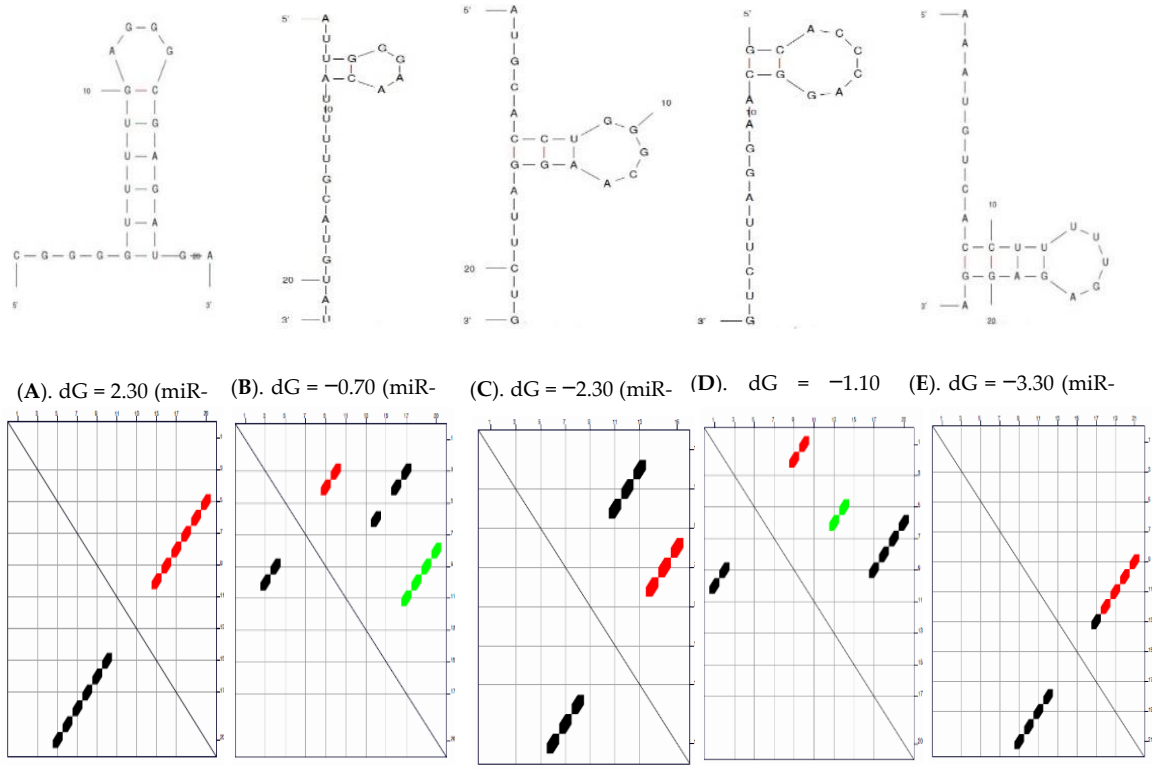


Figure 1. The predicted secondary structure of the five candidate microRNAs with their dot plot directly below showing the optimal energy. (**A**). microRNA miR-1 with its optimal energy of −2.30; (**B**). MicroRNA miR-2 with its optimal energy of −0.70; (**C**). microRNA miR-3 with its optimal energy of −2.30; (**D**). microRNA miR-4 with its optimal energy of −1.10; (**E**). microRNA miR-5 with its optimal energy of −3.30. For the dot plots, red, black, and green dots represent all the optimal foldings (superposition of all possible sub-optimal foldings) Therefore, each colour represents a potential folding configuration.

The prediction of microRNA secondary structure is a long-established problem of computational biology which has received a lot of attention in recent years due to mounting evidence that underscores

the significant of microRNA structure in a wide variety of biological processes [55–57]. The results showed that four of the five have the potential to form more than one secondary structure (folding configuration). Studies have reported numerous microRNAs as biomarkers of disease and potential therapeutic targets due to the fact that their secondary structures may give insight into orchestrated microRNA-dependent gene regulation and be a step forward to understand their functions and involvement in carcinogenesis, and improve therapeutic designs. The results obtained from the secondary structure prediction of the candidate microRNAs are consistent with previous in silico structural determination studies and have indicated that over 70% of human microRNAs may fold into a hairpin structure and almost 70% could potentially form self-aggregated homoduplexes [58–60]. All the predicted microRNAs showed hairpin structure which could be enhanced by the cellular environment.

Valid prediction of microRNA–mRNA binding energies is crucial for the understanding of interactions. The thermodynamics of microRNA interactions can be understood as the sum of the energy necessary to 'open' the binding site and the energy gained from hybridization [57]. The thermodynamic parameters of the secondary structure of the candidate microRNAs are represented in Table 1 above. Ronchieri et al. [61] reported that microRNA secondary structures contribute to target recognition because there is an energetic cost to freeing base-pairing interactions within mRNA in order to make the target accessible for microRNA binding [62–64]. The ΔG values of the candidate microRNAs were calculated using Mfold (Figure 1 and Table 1). The optimal sequences were predicted to have optimal folding in miR-1 and 5, while the sub-optimal sequences were predicted to have sub-optimal folding in miR-2 and 4. Understanding the optimal folding of base pairs is the least likely secondary structure formation during the reaction [65]. The lower the free energy the more stability of the microRNA. From Table 1, miR-5 is the most stable microRNA, followed by miR-3 and miR-1, respectively.

Table 1. Parameters and secondary structures of the microRNA sequences.

S/N	MicroRNAs	Length	δG (*kcal/mol*)	Initial ΔG (*kcal/mol*)	Stru^{C}
1	miR-1	22	0.0	−2.30	1
2	miR-2	22	0.7	−0.70 −0.40 0.00	3
3	miR-3	22	0.7	−2.80 −2.10	2
4	miR-4	20	0.8	−1.10 −0.80 −0.30	3
5	miR-5	22	0.6	−3.30 −2.70	2

δG: Free energy in plot profile; ΔG: Optimal energy of secondary structures (kcal/mol) at 37 °C with optimal and sub-optimal structures, respectively. Stru^{C}: Number of secondary structure calculated by Mfold.

3.2. CpG Island of the Promoter Sequences

The phenomenon of tumor alteration through epigenetic silencing associated with dense hypermethylation of CpG islands, and their complex interplay with modifications in histone structure, provides an alternate mechanism to genetic inactivation of tumor suppressor genes via loss or mutation [66]. Binding sites in the genome have a great regulatory impact on the gene activities in their neighborhood. Predictive tools are therefore essential for deciphering the overall regulatory potential of gene control regions such as promoters and enhancers. The CpG island analysis of the promoters of the prioritized microRNA target genes (7 genes) is shown in Table 2 with Min. GC% of 51% observed in APC while Max. GC% of 83% was observed in KRAS. The ranges of the observed/expected ratio of CpG dinucleotides of these genes are from 0.60 to 1.22. CpG islands have been reported to be found in approximately 50% of human promoters [67,68]. Promoters with CpG islands are associated with housekeeping genes [69,70] and are identified by three primary characteristics: (1) they are over 200

base pairs in length, (2) they have over 50% GC composition, and (3) they retain an observed/expected ratio of CpG dinucleotides greater than 0.6 [71]. In normal tissues, CpG islands associated with tumor suppressor genes are unmethylated but during tumor formation, they are often methylated. Evidence suggests that *de novo* methylation of CpG islands induces the silencing of associated tumor suppressor genes and may, in fact, be an important step in tumor formation [72,73]. The particular genes that are hypermethylated in tumor cells are strongly specific to the tissue of origin of the tumor [74]. This analysis, therefore, strengthens the proof that the sequences extracted from the genomes of these genes may be a true promoter sequence.

Table 2. Result of the CpG island assessed by sequence manipulator suite.

S/N	Gene_ID	Min. GC%	Max. GC%	Min. obs/exp	Max. obs/exp
1	APC	51.00	57.50	0.61	0.67
2	KRAS	59.50	83.00	0.78	1.14
3	TCF7L2	53.00	72.50	0.60	1.00
4	EGFR	56.00	57.00	0.62	0.90
5	IGF1R	53.00	81.50	0.61	1.22
6	CASP8	67.50	70.50	0.60	0.74
7	GNAS	67.50	70.50	0.61	0.78

Note: Min. GC%- Minimum GC content detected in each region of 1-200 base pairs; Max. GC%- Maximum GC content detected in each region of 1-200 base pairs. Min. obs/exp-Minimum observed/expected ratio of CpG dinucleotides. Max. obs/exp- Maximum observed/expected ratio of CpG dinucleotides.

3.3. Triplex Binding Interaction of the MicroRNAs and Target Genes

The specific binding properties of microRNA to proteins or (mRNA) involved in chromatin remodeling and several transcriptional regulations as broadly achieved in the literature demonstrate their multi-functionality [75]. The ability of RNA to participate in triplex formation according to hoogsteen base pairing rules is a less studied property of microRNA [76]. In this study, the binding of the candidate microRNAs to the TSS (600 bp upstream) of the promoter region of the target genes were determined by an in silico analysis using the Trident software. Pasquier et al. [76] reported that formative genes are profoundly represented in the TTS, which certifies the speculation of unexpected large-scale genome regulation mediated by the triplex DNA-RNA structure. The binding of microRNAs to mRNA directly for gene expression regulation has been experimentally validated. However, research has suggested the mechanism by which microRNA forms triplex with double-stranded DNA but their exact mechanism of interaction is less well understood [26,27,77]. Also, Blanco and Montoya [78] elucidated the transient DNA/RNA protein interactions but proposed a further study of triplex formation. From the triplex-forming oligonucleotide formation result represented in Table 3, it was observed that only three of the candidate microRNAs (miR-1, miR-2, and miR-5) are capable of forming multiple triplexes along with three of the target mRNAs (KRAS, TCF7L2, and EGFR). For miR-1, four binding sites were observed in KRAS, and EGFR, while nine binding sites were determined in TCF7L2. This microRNA (miR-1) reports no hit score for the genes APC and GNAS. MicroRNAs 2 and 4 reported no triplex interaction between the target genes showing negative correlation. In microRNA-3, only one reverse (indirect) hoogsteen pairing was observed with KRAS with hit energy of -168.93 (Figure 2). MicroRNA-5 showed two different triplex binding sites in TCF7L2 and one in both APC (indirect) and CASP8 (direct) (Figure 2). All the binding interactions in this study are of grade 5 (99 percentiles of triplex-forming interactions). Both the hit score and hit energy observed are greater than 140 and −140 respectively (Figure 2).

Table 3. Output of the results of trident showing different binding sites.

MicroRNA/Gene	KRAS	TCF7L2	APC	EGFR	CASP8	IGF1R	GNAS
miR-1	−4	−9	0	−3/+1	+1	+1	0
miR-2	0	0	0	0	0	0	0
miR-3	−1	0	0	0	0	0	0
miR-4	0	0	0	0	0	0	0
miR-5	0	−2	−1	0	+1	0	0

Legend: (0); No hit, (+); Direct Hoogsteen, (−); Reverse Hoogsteen, (Values); Number of hits; Grade: 5. Hit score: > 140; Hit energy: < −140. The heuristic score (hit score) represents Hoogsteen or Reverse Hoogsteen base pair complementarity and Thermodynamic Energy (hit energy) represents the binding energy of the triplex. The binding sites were categorized based on the number of hits relative to score and energy.

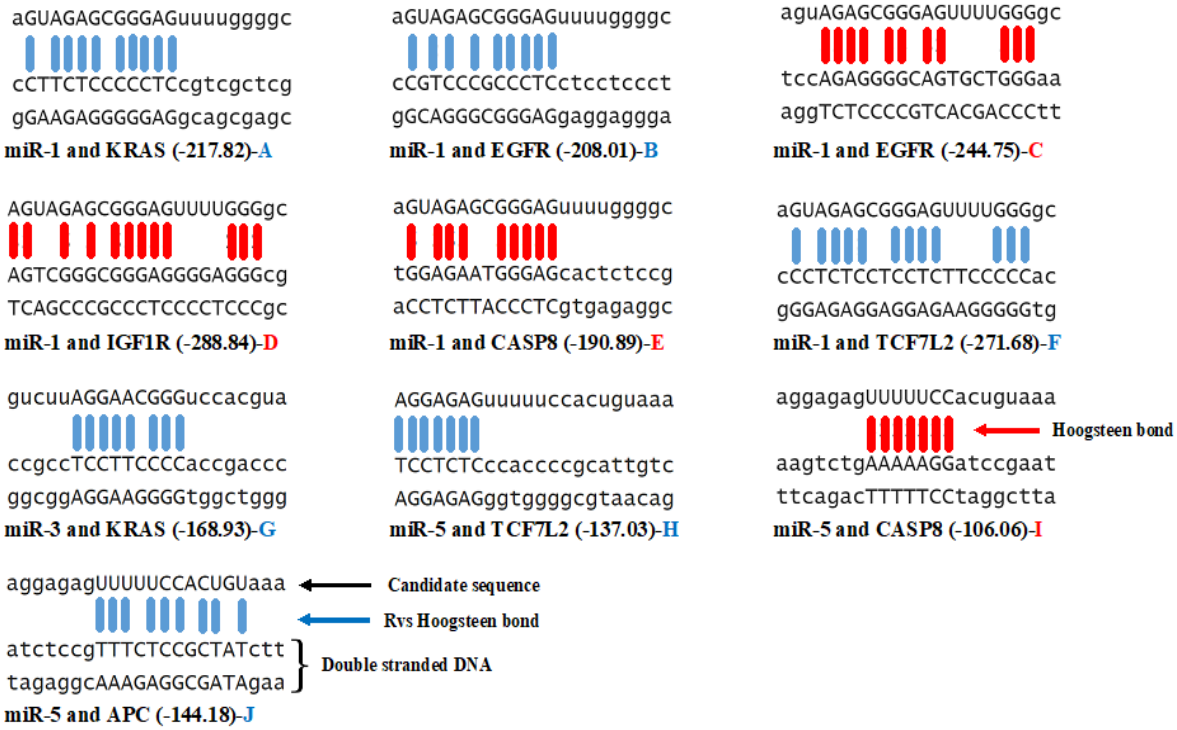


Figure 2. Structural determination of microRNA-DNA triplex formation. The first sequence in each structure (**A**–**J**) represent the candidate microRNA sequences involved in triplex binding while the two sequences without bond are the promoter sequence of the target gene. The blue bond indicates the indirect or reverse hoogsteen bond while the red bond is the direct or hoogsteen bond between the DNA and the microRNA. The negative values in the bracket are the hit energy of each binding.

From previous studies, it was recorded that microRNA with a 22-nucleotides sequence forms triplex structures with duplex DNA as documented by fluorescence resonance energy transfer, FRET, surface Plasmon resonance, electromobility shift assay, and Nuclear magnetic resonance [79,80]. Similar to transient protein–protein and DNA/RNA-protein interactions, Blanco and Montoya [78] suggested that transient formation of microRNA-duplex DNA triplexes may have as much biological importance as more stable interactions. Surprisingly, there are enzymes of the class helicase that possess the ability to unwound the intramolecular triplex DNA [30,81], it is feasible that this mode of action of these enzymes is similitude to which microRNA can mediate transcriptional activation [82]. On a whole, the formation of triplexes between the candidate microRNA (1, 3, and 5) and DNA double-stranded suggest that they are well conserved and crucial mechanism of transcription regulation. Comparing the thermodynamic energies of folding and triplex formation (Table 1 and Figure 2), it was observed that these microRNAs are better favoured for triplex binding to regulate gene expression in the cancer subtype than microRNA folding.

3.4. Somatic Event Evolution of the MicroRNA Target Genes

The evolutionary stage-specific and variant events of the microRNA target genes are presented in Figure 3a,b. Using the SEECancer database, all the genes involved in CRC initiation, progression, and metastasis were graphically represented and visualized using Cytoscape. From the network, four (KRAS, GNAS, APC, and EGRF) of the statistically significant genes targeted by the candidate microRNAs were indicated with yellow node. The database also reported that the other three genes (TCF7L2, CASP8, and IGF1R) are either a product of antecedent or subsequent event of the four genes implicated in CRC. TCF7L2 was seen from the database to be a subsequent event of APC mutation. Loss of heterozygosity (LOS) in the APC was shown to affect TP53 gene through mutation which in turn results in the mutation of TCF7L2. APC was found to be specific to early stage of CRC while GNAS was seen at the metastatic stage. Figure 3b shows that alteration, mutation, methylation aberrance, and loss of copy number of APC gene may result in CRC initiation, progression, or metastasis. Previous studies have reported that the loss of function, as well as methylation aberrance and mutation of the APC gene, is associated with early events in CRC [83–92].

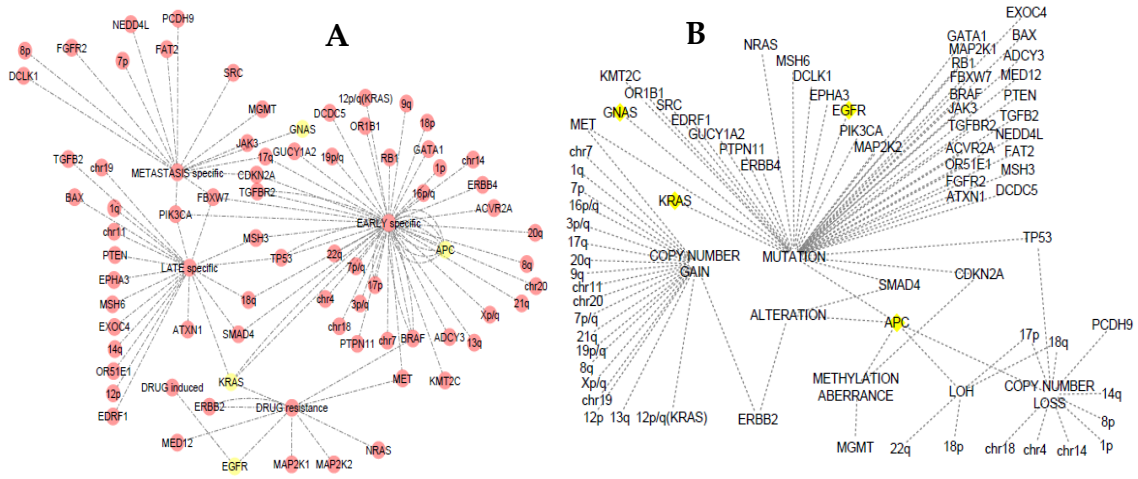


Figure 3. Evolutionary stage-specific somatic events in colorectal cancer (CRC). (**A**) stage-specific targets. (**B**) variant types and specific variants of the microRNA target genes. Yellow nodes indicate the microRNA target genes.

The antecedent event resulting in the mutation of KRAS was also detected to be from the mutation of CASP8 which in turn mutates PIK3CA gene. Mutation in the gene KRAS was revealed to be either at the early, late, or drug resistance stage. Burmer et al. [93] reported the early event of KRAS in CRC as detected by DNA amplification and oligonucleotide hybridization, as well as by RNase A mismatch cleavage analysis. Using a bioinformatics approach, Youn and Simon [94] also suggested that KRAS mutations occur as the first event with high probability for both colorectal and lung tumors. Genome-editing technology also revealed that mutations in this gene are sufficient to initiate tumour progression [95]. The report of Fumagalli et al. [96] using a high throughput experiment suggest that the initiating APC and KRAS mutations drives efficient proliferation and growth, whereas inactivating mutations in SMAD4 block differentiation during tumor progression. Mutation in EGFR was shown to be at the drug-induced and resistance stage event (Figure 3a,b). In metastatic CRC using high throughput experiment, Morelli et al. [97], concluded that acquired mutations of this gene in metastatic CRC patients were correlated with acquired resistance to anti-EGFR monoclonal antibodies. Mutation in other genes targeted by the candidate microRNAs may be involved in colorectal carcinogenesis.

3.5. Co-expression Analysis

GeneMANIA was used to generate a hypothesis about the candidate microRNA target genes' significance. The co-expression analysis of the candidate microRNAs result was represented in Figure 4. The output generated genes that are functionally similar, or have shared properties with microRNA target genes, displayed an interactive functional association network, illustrating the relationships among the genes and gene of interest.

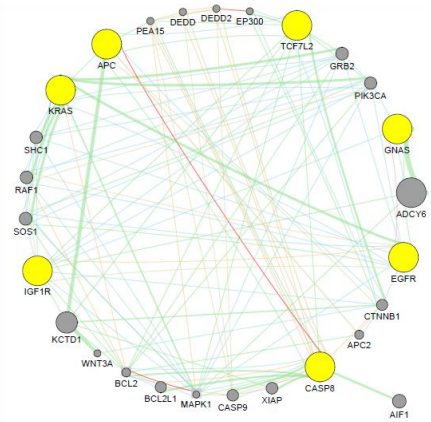


Figure 4. Co-expression analysis network of the prioritized microRNA target genes. The target genes are shown in yellow nodes while other genes in grey represent the associated genes. The red edges showed that the connected genes are co-expressed while the green edges are associated with genetic interactions. Other interactions include shared protein domains, physical interaction, pathway, and predicted interaction.

Genetic interaction of this network accounted for 43.83% of the total interactions, shared protein domain accounted for 26.28, physical interaction accounted for 14.23, co-expressed genes accounted for 10.14%. Only 3.89 and 2.62% are accounted for pathway and predicted interaction respectively. Within the microRNA target genes, only APC and CASP8 are co-expressed, although neighbouring genes found to be co-expressed are Bcl-2 and MAPK-1. Since this tool assigns a weighting feature to dataset to determine how genes in a list are well connected to each other or to determine which types of functional genomic data are most useful to retrieve for finding more genes similar to the query, KCTD1 and ADCY6 were assigned the highest weight due to size and are genetically connected with APC and GNAS respectively. Upregulation of ADCY6 activates the CREB pathway by increasing the tumorigenic potential of cells reported in gastric cancer [98]. Certain genes in the Wnt pathway influence KCTD1-mediated downregulation of β-catenin and suggested that KCTD1 functions as a novel inhibitor of Wnt signaling pathway by enhancing β-catenin degradation [99]. Inappropriate activation of this pathway has been observed in CRC [100,101]. These associated genes may then provide further a complete microRNA-gene network for CRC diagnosis and disease management.

4. Materials and Methods

The secondary structure, CpG island and triplex binding analysis, Somatic Events in Evolution of the target genes, and co-expression analysis were performed on both the candidate microRNAs and their targets using the following tools: Mfold at http://unafold.rna.albany.edu/?q=mfold, SMS at http://www.bioinformatics.org/sms2/translate.html, NCBI at https://blast.ncbi.nlm.nih.gov/Blast.cgi, Ensembl at http://www.ensembl.org/index.html, UCSC at https://genome.ucsc.edu/, Trident at http://trident.stjude.org/, SEECancer at http://biocc.hrbmu.edu.cn/SEECancer/index.jsp, GeneMANIA at http://genemania.org/, and Cytoscape at https://cytoscape.org/.

4.1. Datasets

Five candidate microRNAs decoded as miR-1 to 5 were identified using in silico approach and their targets genes were prioritized through three different target prediction tools (TargetScan, miRDB, and miRDIP) to generate seven genes namely: *APC*, *GNAS*, *EGFR*, *TCF7L2*, *KRAS*, *IGF1R*, and *CASP8*. The sequences of these microRNAs together with the promoter sequences of their targets were used for functional determination in this study.

4.2. Structural Determination of Candidate microRNA

The mature sequences of candidate microRNAs were submitted to Mfold program accessed at http://www.mfold.rna.albany.edu for predicting their secondary structure (s) and the free energy. MFold accepts a single nucleotide sequence as input mainly in FASTA format. The output file produced by MFold contains the calculated energy matrices that determine all optimal and suboptimal secondary structures for the folded nucleic acid molecule. This file is read by the companion program, PlotFold, which can display any of several different graphic representations of optimal and suboptimal secondary structures for the folded molecule.

4.3. Promoter Sequence Extraction

The database for Ensembl was accessed at http://www.ensembl.org/index.html. The seven prioritized microRNA target genes were used as input to generate their promoter sequences. From the display, the exons are highlighted in the pink background and red text, the sequence before exon 1 is the promoter sequence. Six-hundred base pairs 5′-flanking sequences were retrieved. To prevent disparity of the annotation concerning the transcription start site among databases, the obtained sequences were further analyzed in USCC using the BLAT tool at https://genome-euro.ucsc.edu/. Furthermore, CpG island feature and GC content were considered.

4.4. CpG Island Analysis

The sequence manipulator suites (SMS) report potential CpG island regions using the method described by [102]. The calculation was performed using a 200 base pairs window moving across the sequence at 1 base pairs intervals. CpG islands are defined as sequence ranges where the observed/expected value is greater than 0.6 and the GC content is greater than 50%. The expected number of CpG dimers in a window is calculated as the number of 'C's in the window multiplied by the number of 'G's in the window, divided by the window length. While the ration of the observed to the expected is calculated as Obs/Exp CpG = Number of CpG * N/(Number of C * Number of G), where N = length of the sequence. The CpG island suite was accessed at http://www.bioinformatics.org/sms2/cpg_islands.html for the identification of the CpG islands in the promoter sequences of the prioritized microRNA target genes.

4.5. Triplex Binding Analysis

Trident was accessed at http://trident.stjude.org/. This online prediction tool was used to identify triplex binding sites between the candidate microRNAs obtained in our previous study and the promoter region of the statistical significant target genes. Sequences of microRNAs and their targets were used as input. The output of this search was given based on the number of sites that the sequence of the candidate microRNA was able to bind on the promoter sequence. The grade, hit score, types of binding (direct or indirect), and hit energy were also considered.

4.6. Staging Analysis

The SEECancer database was accessed at http://biocc.hrbmu.edu.cn/SEECancer to explore the evolutionary stage-specific somatic events in CRC as well as their temporal order [103]. From the web interface, evolutionary stage and temporal order were selected individually to query the target genes

as well as the cancer subtype respectively. The result of these search generated a txt file as input in Cytoscape to generate a network of events based on variant types and evolutionary events in CRC.

4.7. Co-expression Analysis

A co-expression network identifies which genes have a tendency to show a coordinated expression pattern across a group of samples. This co-expression network can be represented as a gene–gene similarity matrix, which can be used in downstream analyses [104]. Seven prioritized genes target by candidate microRNA were used as input to generate the co-expressed genes using GeneMANIA http://www.genemania.org (http://apps.cytoscape.org/apps/genemania) plugin of Cytoscape. This database is used to generate hypotheses about gene function, analyzing gene lists and prioritizing genes for functional assays. The output is a network of genes with several interactions.

4.8. Statistical Analysis

For microRNA selection, the BlastN parameters were set at 1e-2 for expected value, 7.0 for word size, and 90–99% for similarity index. The CD-HIT-EST-2D parameters were set at 0.90 for threshold and 7.0 for word size. The genes considered in DAVID v6.8 were regarded statistically significant at p-value of 1.8×10^3 with the Benjamini score of 1.6×10^{-2}. A value of $p < 0.05$ was considered to indicate a statistically significant difference.

4.9. Data Availability

The datasets and the clinical data were obtained from the online databases as described above in the methods and their websites are as follows: http://www.mirbase.org/ for reference dataset, microRNA associated with CRC at cancer were http://www.picb.ac.cn/dbDEMC/, http://www.mir2disease.org/, http://www.cuilab.cn/hmdd, and http://mircancer.ecu.edu/. While Mfold, SMS, NCBI, Ensembl, UCSC, Trident, SEECancer, GeneMANIA and Cytoscape were accessed from the server.

5. Conclusions

Understanding microRNA's secondary structures, thermodynamic parameters, and targets may deliver greater promise towards their diagnostic potentials and mechanisms in the management of CRC. The secondary structure together with the thermodynamic parameters of the candidate microRNAs may, therefore, provide a valid result regarding their target even when the conservation of the microRNA is unknown. Also, the secondary structure of the candidate microRNAs suggests a conformational role to modulate target interactions and therefore can be used to explain the different degree of genetic regulation in CRC.

MicroRNAs that are capable of triplex formation with duplex DNA are more frequently positively correlated with gene transcripts. MiR-1, miR-3, and miR-5 are suggested to have significant enrichment of positive correlation with the target gene involved in the triplex structure. This analysis further confirmed that the targets of the candidate microRNAs are enriched with microRNA triplex binding sites. Furthermore, microRNA function may then depend not only on their sequences but also their structures and triplex binding interaction with their targets. Hypothetically speaking, microRNAs targeting these genes can be inferred to be an important regulator in the stage-specific events. Since miR-1, miR-2, and miR-3 are regulators of APC, it can be concluded that these microRNAs may be used as a diagnostic tool in the early detection of CRC. Also, miR-1 regulating GNAS may be involved in metastatic stage-specific event in CRC. MicroRNAs regulating KRAS may also be involved in early, late, or drug-resistant stage-specific events in CRC (MiR-1,3 and 5). However, further molecular experiments are on-going to confirm the function of these identified microRNAs alongside their targets in CRC.

Author Contributions: All authors have made significant contributions to the submission of the article. A.O.F. conceived the concept and the design of the manuscript, A.K. and A.P. provided the necessary software required

to carry out the analysis. The analysis and data interpretations were done by both A.O.F. and A.P. while A.K. drafted the rough draft and also substantively revised the manuscript. Finally, all authors (A.O.F., A.K., and A.P.) read and approved the submitted version of the manuscript for publication. Also, A.O.F., A.K., and A.P. agreed to be personally accountable for their personal contributions and ensured that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Funding: This research received no external funding.

Acknowledgments: Authors will like to acknowledge the Plant Omics Laboratory and the Bioinformatics Research Group of the University of the Western Cape for their usual support. Also, Taiwo Akinsoji (MBBS, M.PH) University of Illinois at Springfield for co-proof reading the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

miRNA	microRNA
CRC	Colorectal cancer
MRNA	Target genes
δG	Free enengy in plot profile
ΔG	Optinal energy of the secondery structure

References

1. Kuo, T.Y.; Hsi, E.; Yang, I.P.; Tsai, P.C.; Wang, J.Y.; Juo, S.H.H. Computational analysis of mRNA expression profiles identifies microRNA-29a/c as predictor of colorectal cancer early recurrence. *PLoS ONE* **2012**, *7*, e31587. [CrossRef] [PubMed]
2. Wolpin, B.M.; Meyerhardt, J.A.; Mamon, H.J.; Mayer, R.J. Adjuvant treatment of colorectal cancer. *Cancer J. Clin.* **2007**, *57*, 1681–1685. [CrossRef] [PubMed]
3. Longo, W.E.; Johnson, F.E. The preoperative assessment and postoperative surveillance of patients with colon and rectal cancer. *Surg. Clin. North Am.* **2002**, *82*, 1091–1108. [CrossRef]
4. Fadaka, A.O.; Pretorius, A.; Klein, A. Biomarkers for Stratification in Colorectal Cancer: MicroRNAs. *Cancer Control* **2019**, *26*. [CrossRef] [PubMed]
5. Schickel, R.; Boyerinas, B.; Park, S.; Peter, M. MicroRNAs: Key players in the immune system, differentiation, tumorigenesis and cell death. *Oncogene* **2008**, *27*, 5959. [CrossRef]
6. Hammond, S.M. MicroRNAs as tumor suppressors. *Nat. Genet.* **2007**, *39*, 582. [CrossRef]
7. Liu, W.; Mao, S.Y.; Zhu, W.Y. Impact of tiny miRNAs on cancers. *World J. Gastroenterol.* **2007**, *13*, 497. [CrossRef]
8. Iorio, M.V.; Croce, C.M. MicroRNAs in cancer: Small molecules with a huge impact. *J. Clin. Oncol.* **2009**, *27*, 5848. [CrossRef]
9. Bartel, D.P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **2004**, *116*, 2812–2897. [CrossRef]
10. Brennecke, J.; Hipfner, D.R.; Stark, A.; Russell, R.B.; Cohen, S.M. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **2003**, *113*, 253–256. [CrossRef]
11. Chen, C.Z.; Li, L.; Lodish, H.F.; Bartel, D.P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **2004**, *303*, 83–86. [CrossRef] [PubMed]
12. Fadaka, A.O.; Ojo, B.A.; Adewale, O.B.; Esho, T.; Pretorius, A. Effect of dietary components on miRNA and colorectal carcinogenesis. *Cancer Cell Int.* **2018**, *18*, 130. [CrossRef]
13. Lee, Y.S.; Dutta, A. MicroRNAs in cancer. *Annu. Rev. Pathol. Mech. Dis.* **2009**, *4*, 199–227. [CrossRef] [PubMed]
14. Motoyama, K.; Inoue, H.; Takatsuno, Y.; Tanaka, F.; Mimori, K.; Uetake, H.; Sugihara, K.; Mori, M. Over-and under-expressed microRNAs in human colorectal cancer. *Int. J. Oncol.* **2009**, *34*, 1069–1075. [PubMed]
15. Aslam, M.; Taylor, K.; Pringle, J.; Jameson, J. MicroRNAs are novel biomarkers of colorectal cancer. *Br. J. Surg.* **2009**, *96*, 702–710. [CrossRef] [PubMed]
16. Manne, U.; Shanmugam, C.; Bovell, L.; Katkoori, V.R.; Bumpers, H.L. miRNAs as biomarkers for management of patients with colorectal cancer. *Biomark. Med.* **2010**, *4*, 7617–7670. [CrossRef] [PubMed]

17. Xi, Y.; Edwards, J.R.; Ju, J. Investigation of miRNA biology by bioinformatic tools and impact of miRNAs in colorectal cancer—Regulatory relationship of c-Myc and p53 with miRNAs. *Cancer Inform.* **2007**, *3*. [CrossRef]
18. Ohler, U.; Niemann, H. Identification and analysis of eukaryotic promoters: Recent computational approaches. *Trends Genet.* **2001**, *17*, 56–60. [CrossRef]
19. Moser, H.E.; Dervan, P.B. Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **1987**, *238*, 645–650. [CrossRef]
20. Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.L.; Thuong, N.T.; Lhomme, J.; Héène, C. Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-(α]-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res.* **1987**, *15*, 77497–77760. [CrossRef]
21. Broitman, S.L.; Im, D.D.; Fresco, J.R. Formation of the triple-stranded polynucleotide helix, poly (AAU). *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 5120–5124. [CrossRef] [PubMed]
22. Wells, R.; Collier, D.; Hanvey, J.; Shimizu, M.; Wohlrab, F. The chemistry and biology of unusual DNA structures adopted by oligopurine. oligopyrimidine sequences. *FASEB J.* **1988**, *2*, 29392–29949. [CrossRef]
23. Buske, F.A.; Mattick, J.S.; Bailey, T.L. Potential in vivo roles of nucleic acid triple-helices. *RNA Biol.* **2011**, *8*, 427–439. [CrossRef] [PubMed]
24. Jalali, S.; Singh, A.; Maiti, S.; Scaria, V. Genome-wide computational analysis of potential long noncoding RNA mediated DNA: DNA: RNA triplexes in the human genome. *J. Transl. Med.* **2017**, *15*, 186. [CrossRef]
25. Li, Y.; Syed, J.; Sugiyama, H. RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem. Biol.* **2016**, *23*, 13251–13333. [CrossRef] [PubMed]
26. Janowski, B.A.; Younger, S.T.; Hardy, D.B.; Ram, R.; Huffman, K.E.; Corey, D.R. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat. Chem. Biol.* **2007**, *3*, 166. [CrossRef]
27. Check, E. RNA interference: Hitting the on switch. *Nature* **2007**, *448*, 855–858. [CrossRef]
28. Kanak, M.; Alseiari, M.; Balasubramanian, P.; Addanki, K.; Aggarwal, M.; Noorali, S.; Kalsum, A.; Mahalingam, K.; Pace, G.; Panasik, N. Triplex-forming MicroRNAs form stable complexes with HIV-1 provirus and inhibit its replication. *Appl. Immunohistochem. Mol. Morphol.* **2010**, *18*, 5325–5345. [CrossRef]
29. Schmitz, K.-M.; Mayer, C.; Postepska, A.; Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **2010**, *24*, 2264–2269. [CrossRef]
30. Goñi, J.R.; De La Cruz, X.; Orozco, M. Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res.* **2004**, *32*, 3543–3560. [CrossRef]
31. Jenjaroenpun, P.; Kuznetsov, V.A. In TTS mapping: Integrative WEB tool for analysis of triplex formation target DNA sequences, G-quadruplets and non-protein coding regulatory DNA elements in the human genome. *BMC Genom.* **2009**, S9. [CrossRef] [PubMed]
32. Jenjaroenpun, P.; Chew, C.S.; Yong, T.P.; Choowongkomon, K.; Thammasorn, W.; Kuznetsov, V.A. The TTSMI database: A catalog of triplex target DNA sites associated with genes and regulatory elements in the human genome. *Nucleic Acids Res.* **2014**, *43*, D110–D116. [CrossRef] [PubMed]
33. Buske, F.A.; Bauer, D.C.; Mattick, J.S.; Bailey, T.L. Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.* **2012**, *22*, 1372–1381. [CrossRef] [PubMed]
34. Wu, Q.; Gaddis, S.S.; MacLeod, M.C.; Walborg, E.F.; Thames, H.D.; DiGiovanni, J.; Vasquez, K.M. High-affinity triplex-forming oligonucleotide target sequences in mammalian genomes. *Mol. Carcinog* **2007**, *46*, 152–153. [CrossRef] [PubMed]
35. Batlle, E.; Bacani, J.; Begthel, H.; Jonkeer, S.; Gregorieff, A.; van de Born, M.; Malats, N.; Sancho, E.; Boon, E.; Pawson, T. EphB receptor activity suppresses colorectal cancer progression. *Nature* **2005**, *435*, 1126. [CrossRef] [PubMed]
36. Dow, L.E.; O'Rourke, K.P.; Simon, J.; Tschaharganeh, D.F.; van Es, J.H.; Clevers, H.; Lowe, S.W. Apc restoration promotes cellular differentiation and reestablishes crypt homeostasis in colorectal cancer. *Cell* **2015**, *161*, 15391–15552. [CrossRef] [PubMed]
37. Scholer-Dahirel, A.; Schlabach, M.R.; Loo, A.; Bagdasarian, L.; Meyer, R.; Guo, R.; Woolfenden, S.; Kristine, K.Y.; Markovits, J.; Killary, K. Maintenance of adenomatous polyposis coli (APC)-mutant colorectal cancer is dependent on Wnt/β-catenin signaling. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17135–17140. [CrossRef]

38. Misale, S.; Yaeger, R.; Hobor, S.; Scala, E.; Janakiraman, M.; Liska, D.; Valtorta, E.; Schiavo, R.; Buscarino, M.; Siravegna, G. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **2012**, *486*, 532. [CrossRef]
39. De Roock, W.; De Vriendt, V.; Normanno, N.; Ciardiello, F.; Tejpar, S. KRAS, BRAF, PIK3CA, and PTEN mutations: Implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol.* **2011**, *12*, 594–603. [CrossRef]
40. Bass, A.J.; Lawrence, M.S.; Brace, L.E.; Ramos, A.H.; Drier, Y.; Cibulskis, K.; Sougnez, C.; Voet, D.; Saksena, G.; Sivachenko, A. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nat. Genet.* **2011**, *43*, 964. [CrossRef]
41. Theodoropoulos, G.E.; Gazouli, M.; Vaiopoulou, A.; Leandrou, M.; Nikouli, S.; Vassou, E.; Kouraklis, G.; Nikiteas, N. Polymorphisms of caspase 8 and caspase 9 gene and colorectal cancer susceptibility and prognosis. *Int. J. Colorectal Dis.* **2011**, *26*, 1113–1118. [CrossRef] [PubMed]
42. Inno, A.; Di Salvatore, M.; Cenci, T.; Martini, M.; Orlandi, A.; Strippoli, A.; Ferrara, A.M.; Bagalà, C.; Cassano, A.; Larocca, L.M. Is there a role for IGF1R and c-MET pathways in resistance to cetuximab in metastatic colorectal cancer? *Clin. Colorectal Cancer* **2011**, *10*, 325–332. [CrossRef] [PubMed]
43. Idziaszczyk, S.; Wilson, C.; Smith, C.; Adams, D.; Cheadle, J. Analysis of the frequency of GNAS codon 201 mutations in advanced colorectal cancer. *Cancer Genet. Cytogenetics* **2010**, *202*, 676–679. [CrossRef] [PubMed]
44. Kojima, M.; Sudo, H.; Kawauchi, J.; Takizawa, S.; Kondou, S.; Nobumasa, H.; Ochiai, A. MicroRNA markers for the diagnosis of pancreatic and biliary-tract cancers. *PLoS ONE* **2015**, *10*, e0118220. [CrossRef]
45. Plieskatt, J.L.; Rinaldi, G.; Feng, Y.; Peng, J.; Yonglitthipagon, P.; Easley, S.; Laha, T.; Pairojkul, C.; Bhudhisawasdi, V.; Sripa, B. Distinct miRNA signatures associate with subtypes of cholangiocarcinoma from infection with the tumourigenic liver fluke Opisthorchis viverrini. *J. Hepatol.* **2014**, *61*, 8508–8558. [CrossRef]
46. Taylor, B.S.; Schultz, N.; Hieronymus, H.; Gopalan, A.; Xiao, Y.; Carver, B.S.; Arora, V.K.; Kaushik, P.; Cerami, E.; Reva, B. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **2010**, *18*, 11–22. [CrossRef]
47. Ng, S.-B.; Yan, J.; Huang, G.; Selvarajan, V.; Tay, J.L.-S.; Lin, B.; Bi, C.; Tan, J.; Kwong, Y.-L.; Shimizu, N. Dysregulated microRNAs affect pathways and targets of biologic relevance in nasal-type natural killer/T-cell lymphoma. *Blood* **2011**, *118*, 49194–49929. [CrossRef]
48. Zhang, P.; Bill, K.; Liu, J.; Young, E.; Peng, T.; Bolshakov, S.; Hoffman, A.; Song, Y.; Demicco, E.G.; Terrada, D.L. MiR-155 is a liposarcoma oncogene that targets casein kinase-1α and enhances β-catenin signaling. *Cancer Res.* **2012**, *72*, 17511–17762. [CrossRef]
49. Munding, J.B.; Adai, A.T.; Maghnouj, A.; Urbanik, A.; Zöllner, H.; Liffers, S.T.; Chromik, A.M.; Uhl, W.; Szafranska-Schwarzbach, A.E.; Tannapfel, A. Global microRNA expression profiling of microdissected tissues identifies miR-135b as a novel biomarker for pancreatic ductal adenocarcinoma. *Int. J. Cancer* **2012**, *131*, E86–E95. [CrossRef]
50. Zhang, N.; Wang, X.; Huo, Q.; Li, X.; Wang, H.; Schneider, P.; Hu, G.; Yang, Q. The oncogene metadherin modulates the apoptotic pathway based on the tumor necrosis factor superfamily member TRAIL (Tumor Necrosis Factor-related Apoptosis-inducing Ligand) in breast cancer. *J. Biol. Chem.* **2013**, *288*, 9396–9407. [CrossRef]
51. Biagioni, F.; Ben-Moshe, N.B.; Fontemaggi, G.; Canu, V.; Mori, F.; Antoniani, B.; Di Benedetto, A.; Santoro, R.; Germoni, S.; De Angelis, F. miR-10b*, a master inhibitor of the cell cycle, is down-regulated in human breast tumours. *EMBO Mol. Med.* **2012**, *4*, 12141–12229. [CrossRef] [PubMed]
52. Özata, D.M.; Caramuta, S.; Velázquez-Fernández, D.; Akçakaya, P.; Xie, H.; Höög, A.; Zedenius, J.; Bäckdahl, M.; Larsson, C.; Lui, W.-O. The role of microRNA deregulation in the pathogenesis of adrenocortical carcinoma. *Endocr. Relat. Cancer* **2011**, *18*, 6436–6455. [CrossRef] [PubMed]
53. Namløs, H.M.; Meza-Zepeda, L.A.; Barøy, T.; Østensen, I.H.; Kresse, S.H.; Kuijjer, M.L.; Serra, M.; Bürger, H.; Cleton-Jansen, A.-M.; Myklebost, O. Modulation of the osteosarcoma expression phenotype by microRNAs. *PLoS ONE* **2012**, *7*, e48086. [CrossRef] [PubMed]
54. Jung, H.M.; Phillips, B.L.; Patel, R.S.; Cohen, D.M.; Jakymiw, A.; Kong, W.W.; Cheng, J.Q.; Chan, E.K. Keratinization-associated miR-7 and miR-21 regulate tumor suppressor reversion-inducing cysteine-rich protein with kazal motifs (RECK) in oral cancer. *J. Biol. Chem.* **2012**, *287*, 29261–29272. [CrossRef]

55. Kretschmer-Kazemi Far, R.; Sczakiel, G. The activity of siRNA in mammalian cells is related to structural target accessibility: A comparison with antisense oligonucleotides. *Nucleic Acids Res.* **2003**, *31*, 44174–44424. [CrossRef]
56. Overhoff, M.; Alken, M.; Far, R.K.-K.; Lemaitre, M.; Lebleu, B.; Sczakiel, G.; Robbins, I. Local RNA target structure influences siRNA efficacy: A systematic global analysis. *J. Mol. Biol.* **2005**, *348*, 8718–8781. [CrossRef]
57. Mückstein, U.; Tafer, H.; Hackermüller, J.; Bernhart, S.H.; Stadler, P.F.; Hofacker, I.L. Thermodynamics of RNA–RNA binding. *Bioinform.* **2006**, *22*, 1177–1182. [CrossRef]
58. Belter, A.; Gudanis, D.; Rolle, K.; Piwecka, M.; Gdaniec, Z.; Naskręt-Barciszewska, M.Z.; Barciszewski, J. Mature miRNAs form secondary structure, which suggests their function beyond RISC. *PLoS ONE* **2014**, *9*, e113848. [CrossRef]
59. Maiti, M.; Nauwelaerts, K.; Lescrinier, E.; Schuit, F.C.; Herdewijn, P. Self-complementary sequence context in mature miRNAs. *Biochem. Biophys. Res. Commun.* **2010**, *392*, 5725–5776. [CrossRef]
60. Belter, A.; Naskret-Barciszewska, M. The correlation of structural features of mature miRNAs with their biological function. *Biotechnol. J. Biotechnol. Comput. Biol. Bionanotechnol.* **2014**, *95*, 187–191.
61. Ronchieri, E.; DAgostino, D.; Milanesi, L.; Merelli, I. microRNA-target interaction: A parallel approach for computing pairing energy. In *Proceedings of the 2016 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP), Heraklion, Greece, 17–19 February 2016*; IEEE Press: Piscataway, NJ, USA, 2016; pp. 535–540.
62. Witkos, T.M.; Koscianska, E.; Krzyzosiak, W.J. Practical aspects of microRNA target prediction. *Curr. Mol. Med.* **2011**, *11*, 93–109. [CrossRef] [PubMed]
63. Betel, D.; Koppal, A.; Agius, P.; Sander, C.; Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* **2010**, *11*, R90. [CrossRef]
64. Kertesz, M.; Iovino, N.; Unnerstall, U.; Gaul, U.; Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **2007**, *39*, 1278. [CrossRef]
65. Arif, M.; Ochoa-Corona, F. Comparative assessment of 5′ A/T-rich overhang sequences with optimal and sub-optimal primers to increase PCR yields and sensitivity. *Mol. Biotechnol.* **2013**, *55*, 172–176. [CrossRef] [PubMed]
66. Esteller, M. Epigenetics in cancer. *New Engl. J. Med.* **2008**, *358*, 1148–1159. [CrossRef] [PubMed]
67. Oberley, M.J.; Farnham, P.J. Probing chromatin immunoprecipitates with CpG-island microarrays to identify genomic sites occupied by DNA-binding proteins. *Methods Enzym.* **2003**, *371*, 577–596.
68. Curtin, K.; Slattery, M.L.; Samowitz, W.S. CpG island methylation in colorectal cancer: Past, present and future. *Pathol. Res. Int.* **2011**. [CrossRef] [PubMed]
69. Bird, A.P. Gene expression: DNA methylation—How important in gene control? *Nature* **1984**, *307*, 503. [CrossRef]
70. Bird, A.P. DNA methylation versus gene expression. *Development* **1984**, *83*, 31–40.
71. Takai, D.; Jones, P.A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 37403–37745. [CrossRef]
72. Gatter, K. Handbook of immunohistochemistry and in situ hybridisation of human carcinomas: Molecular genetics–lung and breast carcinomas. *Br. J. Cancer* **2005**, *93*, 1318. [CrossRef]
73. Portela, A.; Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **2010**, *28*, 1057. [CrossRef] [PubMed]
74. Esteller, M. CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future. *Oncogene* **2002**, *21*, 5427. [CrossRef] [PubMed]
75. Elliott, D.; Ladomery, M. *Molecular Biology of RNA*; Oxford University Press: Oxford, UK, 2017.
76. Pasquier, C.; Agnel, S.; Robichon, A. The Mapping of Predicted Triplex DNA: RNA in the Drosophila Genome Reveals a Prominent Location in Development-and Morphogenesis-Related Genes. *Genes Genomes Genet.* **2017**, *7*, 2295–2304.
77. Park, C.W.; Zeng, Y.; Zhang, X.; Subramanian, S.; Steer, C.J. Mature microRNAs identified in highly purified nuclei from HCT116 colon cancer cells. *RNA Biol.* **2010**, *7*, 606–614. [CrossRef] [PubMed]
78. Blanco, F.J.; Montoya, G. Transient DNA/RNA-protein interactions. *FEBS J.* **2011**, *278*, 16431–16650. [CrossRef] [PubMed]

79. Semerad, C.L.; James Maher III, L. Exclusion of RNA strands from a purine motif triple helix. *Nucleic Acids Res.* **1994**, *22*, 53215–53325. [CrossRef]
80. Roberts, R.W.; Crothers, D.M. Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. *Science* **1992**, *258*, 1463–1466. [CrossRef]
81. Maine, I.P.; Kodadek, T. Efficient unwinding of triplex DNA by a DNA helicase. *Biochem. Biophys. Res. Commun.* **1994**, *204*, 1119–1124. [CrossRef]
82. Paugh, S.W.; Coss, D.R.; Bao, J.; Laudermilk, L.T.; Grace, C.R.; Ferreira, A.M.; Waddell, M.B.; Ridout, G.; Naeve, D.; Leuze, M. MicroRNAs form triplexes with double stranded DNA at sequence-specific binding sites; a eukaryotic mechanism via which microRNAs could directly alter gene expression. *PLoS Comput. Biol.* **2016**, *12*, e1004744. [CrossRef]
83. Umetani, N.; Sasaki, S.; Watanabe, T.; Shinozaki, M.; Matsuda, K.; Ishigami, H.; Ueda, E.; Muto, T. Genetic alterations in ulcerative colitis-associated neoplasia focusing on APC, K-ras gene and microsatellite instability. *Jpn. J. Cancer Res.* **1999**, *90*, 1081–1087. [CrossRef] [PubMed]
84. Matkowskyj, K.A.; Chen, Z.E.; Rao, M.S.; Yang, G.-Y. Dysplastic lesions in inflammatory bowel disease: Molecular pathogenesis to morphology. *Arch. Pathol. Lab. Med.* **2013**, *137*, 338–350. [CrossRef] [PubMed]
85. Redston, M.S.; Papadopoulos, N.; Caldas, C.; Kinzler, K.W.; Kern, S.E. Common occurrence of APC and K-ras gene mutations in the spectrum of colitis-associated neoplasias. *Gastroenterology* **1995**, *108*, 3833–3892. [CrossRef]
86. Yashiro, M. Molecular alterations of colorectal cancer with inflammatory bowel disease. *Dig. Dis. Sci.* **2015**, *60*, 2251–2263. [CrossRef]
87. Molinaro, A.; Terdiman, J.P.; Willenbucher, R.; Chang, F.; Baretton, G.B.; Waldman, F.; Aust, A.E.; Loehrs, U. The APC/beta-catenin pathway in ulcerative colitis-related colorectal carcinomas: A mutational analysis. *Lancet* **2002**. [CrossRef]
88. Segditsas, S.; Rowan, A.; Howarth, K.; Jones, A.; Leedham, S.; Wright, N.; Gorman, P.; Chambers, W.; Domingo, E.; Roylance, R. APC and the three-hit hypothesis. *Oncogene* **2009**, *28*, 146. [CrossRef]
89. Kim, T.-M.; An, C.H.; Rhee, J.-K.; Jung, S.-H.; Lee, S.H.; Baek, I.-P.; Kim, M.S.; Lee, S.H.; Chung, Y.-J. Clonal origins and parallel evolution of regionally synchronous colorectal adenoma and carcinoma. *Oncotarget* **2015**, *6*, 27725. [CrossRef]
90. Sakai, E.; Fukuyo, M.; Matsusaka, K.; Ohata, K.; Doi, N.; Takane, K.; Matsuhashi, N.; Fukushima, J.; Nakajima, A.; Kaneda, A. TP 53 mutation at early stage of colorectal cancer progression from two types of laterally spreading tumors. *Cancer Sci.* **2016**, *107*, 8208–8227. [CrossRef]
91. Fodde, R.; Kuipers, J.; Rosenberg, C.; Smits, R.; Kielman, M.; Gaspar, C.; van Es, J.H.; Breukel, C.; Wiegant, J.; Giles, R.H.; et al. Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat. Cell Biol.* **2001**, *3*, 433–438. [CrossRef]
92. Smolle, M.A.; Pichler, M.; Haybaeck, J.; Gerger, A. Genetic markers of recurrence in colorectal cancer. *Pharmacogenomics* **2015**, *16*, 1315–1328. [CrossRef]
93. Burmer, G.C.; Rabinovitch, P.S.; Loeb, L.A. Analysis of c-Ki-ras mutations in human colon carcinoma by cell sorting, polymerase chain reaction, and DNA sequencing. *Cancer Res.* **1989**, *49*, 2141–2146. [PubMed]
94. Youn, A.; Simon, R. Estimating the order of mutations during tumorigenesis from tumor genome sequencing data. *Bioinformatics* **2012**, *28*, 15551–15561. [CrossRef]
95. Dienstmann, R.; Vermeulen, L.; Guinney, J.; Kopetz, S.; Tejpar, S.; Tabernero, J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* **2017**, *17*, 79. [CrossRef] [PubMed]
96. Fumagalli, A.; Drost, J.; Suijkerbuijk, S.J.; Van Boxtel, R.; De Ligt, J.; Offerhaus, G.J.; Begthel, H.; Beerling, E.; Tan, E.H.; Sansom, O.J. Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E2357–E2364. [CrossRef] [PubMed]
97. Morelli, M.; Overman, M.; Dasari, A.; Kazmi, S.; Mazard, T.; Vilar, E.; Morris, V.; Lee, M.; Herron, D.; Eng, C. Characterizing the patterns of clonal selection in circulating tumor DNA from patients with colorectal cancer refractory to anti-EGFR treatment. *Ann. Oncol.* **2015**, *26*, 7317–7336. [CrossRef]
98. Hong, S.H.; Goh, S.H.; Lee, S.J.; Hwang, J.A.; Lee, J.; Choi, I.J.; Seo, H.; Park, J.-H.; Suzuki, H.; Yamamoto, E. Upregulation of adenylate cyclase 3 (ADCY3) increases the tumorigenic potential of cells by activating the CREB pathway. *Oncotarget* **2013**, *4*, 1791. [CrossRef] [PubMed]

99. Li, X.; Chen, C.; Wang, F.; Huang, W.; Liang, Z.; Xiao, Y.; Wei, K.; Wan, Z.; Hu, X.; Xiang, S. KCTD1 suppresses canonical Wnt signaling pathway by enhancing β-catenin degradation. *PLoS ONE* **2014**, *9*, e94343. [CrossRef] [PubMed]
100. Cui, J.; Zhou, X.; Liu, Y.; Tang, Z.; Romeih, M. Wnt signaling in hepatocellular carcinoma: Analysis of mutation and expression of beta-catenin, T-cell factor-4 and glycogen synthase kinase 3-beta genes. *J. Gastroenterol. Hepatol.* **2003**, *18*, 2802–2887. [CrossRef]
101. Ewan, K.B.; Dale, T.C. The potential for targeting oncogenic wnt/β-catenin signaling in therapy. *Curr. Drug Targets* **2008**, *9*, 5325–5347. [CrossRef]
102. Gardiner-Garden, M.; Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **1987**, *196*, 2612–2682. [CrossRef]
103. Zhang, H.; Luo, S.; Zhang, X.; Liao, J.; Quan, F.; Zhao, E.; Zhou, C.; Yu, F.; Yin, W.; Zhang, Y. SEECancer: A resource for somatic events in evolution of cancer genome. *Nucleic Acids Res.* **2017**, *46*, 1018–1026. [CrossRef] [PubMed]
104. Van Dam, S.; Vosa, U.; van der Graaf, A.; Franke, L.; de Magalhaes, J.P. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* **2017**, *19*, 5755–5792. [CrossRef] [PubMed]

CHAPTER 6.0

MICRORNA ASSISTED GENE REGULATION IN COLORECTAL CANCER

ADEWALE OLUWASEUN FADAKA*, ASHLEY PRETORIUS, AND ASHWIL KLEIN

Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Cape Town, South Africa.

Accepted/ published: *International Journal of Molecular Science* (2019) **20**(**19**) 4899. DOI:10.3390/ijms20194899 **(Impact factor: 4.183)**

Statement of Contributions of Joint Authorship

Adewale Fadaka (PhD candidate)

Conceived the concept of the manuscript as well as the write up and compilation. Preparation of figures and tables.

Ashley Pretorius (Co-Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

Ashwil Klein (Principal Supervisor)

Supervised and assisted with manuscript compilation, editing and overall proofreading and co-author of the manuscript.

This Chapter is an exact copy of the journal paper below.

International Journal of
Molecular Sciences

MDPI

Article

MicroRNA Assisted Gene Regulation in Colorectal Cancer

Adewale O. Fadaka *, Ashley Pretorius and Ashwil Klein

Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Private Bag X17, Bellville, 7535 Cape Town, South Africa; aspretorius@uwc.ac.za (A.P.); asklein@uwc.ac.za (A.K.)

* Correspondence: afadaka@uwc.ac.za; Tel.: +27-630511928; +234-8039242052

Received: 3 August 2019; Accepted: 29 August 2019; Published: 3 October 2019

Abstract: Colorectal cancer (CRC) is the second-leading cause of cancer death and a major public health problem. Nearly 80% CRC cases are diagnosed after the disease have metastasized and are often too advanced for treatment. Small non-coding RNA guides argonaute protein to their specific target for regulation as the sole of RNA induced silencing complex for gene silencing. These non-coding RNA for example microRNA, are thought to play a key role in affecting the efficiency of gene regulation in cancer, especially CRC. Understanding the mechanism at the molecular level could lead to improved diagnosis, treatment, and management decisions for CRC. The study aimed to predict the molecular mechanism of gene regulation based microRNA-mRNA duplex as a lead in the silencing mechanism. Five candidate microRNAs were identified through the in silico approach. The MicroRNA target prediction and subsequent correlation, and prioritization were performed using miRTarBase, gbCRC and CoReCG, and DAVID databases respectively. Protein selection and preparation were carried out using PDB and Schrödinger suits. The molecular docking analysis was performed using PATCHDOCK webserver and visualized by discovery studio visualizer. The results of the study reveal that the candidate microRNAs have strong binding affinity towards their targets suggesting a crucial factor in the silencing mechanism. Furthermore, the molecular docking of the receptor to both the microRNA and microRNA-mRNA duplex were analyzed computationally to understand their interaction at the molecular level. Conclusively, the study provides an explanation for understanding the microRNAs-based gene regulation (silencing mechanism) in CRC.

Keywords: molecular interaction; microRNA; *in silico* prediction; target gene; gene expression; silencing; colorectal cancer

1. Introduction

Colorectal cancer (CRC) is considered as one of the most threatening diseases due to its incidence and mortality rate worldwide [1], and the most frequent cancers in western world [2]. Over 1.2 million individuals are diagnosed with this disease yearly, and over 600,000 mortalities are recorded [3]. Although the activation and inactivation of oncogenes and tumor suppressor genes respectively are known to be involved in CRC development at molecular level [4], the molecular mechanisms that lead to the development and progression of CRC remain unclear.

Despite the advances in the diagnosis, treatment, and management of patients with CRC, it is still a major public health problem globally [5]. Therefore, it is imperative to elucidate the mechanism of gene silencing in the tumorigenesis of CRC for better understanding.

The interactions between protein and nucleic acids play essential roles in various cellular and biological processes, including DNA replication, RNA transcription, the translation of polypeptides, RNA splicing, and the degradation of nucleic acids [6,7]. The errors in receptor-nucleic acid interactions are implicated in a number of diseases, ranging from neurological disorders to cancer [8]. RNA

binding proteins are mediators of RNA silencing processes, such as pathways in microRNA and RNA interference. Argonaute, a unique member of this family [9], forms the functional core of the RNA induced silencing complex (RISC) in humans [10]. The RISC complexed with AGO employs small molecules, such as microRNA, as a guide for target recognition and silencing through translational repression and/or degradation [11].

MicroRNAs are small non-coding RNAs with 18–22 nucleotide sequences possessing regulatory roles in both plants and animals. These non-coding RNAs are involved in different cellular processes [12–16] including human diseases [17], such as colorectal carcinogenesis [18]. Additionally, experiments revealed that these RNAs can act as oncomiR [19,20] and/or tumor suppressor microRNAs [21] and their differential expression between normal and abnormal tissue have been exploited in the diagnosis, treatment, and management of CRC [22]. The epigenetic regulation of gene expression at a transcriptional or post-transcriptional level is important as a mechanism of gene silencing.

Various experimental approaches [23] have been put forward to study the mechanism by which cancer genes are repressed, inactivated or silenced to prevent carcinogenesis, progression or metastasis of the involved gene. Recently, the microRNA binding proteins became a focal point in cancer research due to their involvement in microRNAs deregulation [24,25]. Argonaute utilizes microRNAs and RNA interference as sequence-specific guides in both transcriptional and posttranscriptional silencing mechanisms [26]. Several roles of AGO have been observed in translational regulation and RNA interference but their functions in human disease remain a top priority. Li, Yu, Gao and Li [23] studied the expression on AGO protein in colon cancer as a potential biomarker, Sun, et al. [27] reported the prognostic expression status of PIWIL1 in CRC and Völler, et al. [28] also studied their expression in cancer entitles. The information for the understanding of these processes is likely to improve as new structures of protein-nucleic acid complexes are solved and the structural details of the interactions are analyzed. However, experimental determination by high-resolution methods is a tedious and difficult process.

Molecular simulation has emerged as an efficient and cost-effective tool in binding analysis from lead identification to optimization and beyond [29]. The process of molecular interaction through a non-covalent bond with high affinity and specificity to form a specific complex is crucial to all processes in living organisms [30]. Protein functions are majorly determined based on their binding interaction with other molecules or ligands [31]. Therefore, understanding protein-ligand interactions are central to understanding molecular biology. Additionally, information regarding the mechanisms of target interaction of protein-ligand binding is also likely to promote the discovery of drugs, a better understanding of gene silencing, the treatment and management efficacy in various diseases, most especially in cancer and the CRC subtype. This study, therefore, insights into an improved understanding at the molecular level the microRNA-assisted target recognition and regulation of argonaute as a therapeutic modality against CRC. Molecular docking approaches of microRNA conformations adopted within the binding pocket of the Argonaute protein could also assist to estimate the residual amino acids, hydrogen bond interactions and binding free energy to provide information crucial to the intermolecular recognition mechanism.

2. Results

2.1. Identification of Candidate MicroRNA and Target Genes

Figure 1 represents the overall methodology employed in this study. The sequence similarity search was employed through the basic local alignment search tool for nucleic acids (BLASTN) and the Homology Detection and Clustering Database at High Identity with Tolerance (CH-HIT-EST-2D) between the total microRNAs from miRBase as reference microRNAs and microRNAs experimentally validated in 4 databases (DbDEMC at http://www.picb.ac.cn/dbDEMC/, miR2Disease at http://www.mir2disease.org/, HMDD at http://www.cuilab.cn/hmdd, and miRCancer at http://mircancer.ecu.edu/)

as the query set. With a similarity threshold of 0.90, the result was text-mined to obtain the final list of 5 candidate microRNAs together with their clusters associated with CRC (Table 1). MiRTarBase was used to predict the target genes of these microRNAs. Collectively, 44 genes alongside their minimum free energies (MFE) were identified to target candidate microRNAs after their intersection analysis with two CRC gene databases (CoReCG and gbCRC) (Table 2). The miR-1 targeted 12 genes, miR-2 targeted 10, miR-3 targeted 8, miR-4 targeted 6, and finally, miR-5 was associated with 8 genes. The combined targets were used as inputs in DAVID for the functional annotation as the first phase of gene prioritization. The result showed that 18 target genes were involved in cancer (GAD_DISEASE_CLASS) as shown in Table 3 with the *p*-value of 1.8E-3 and a Benjamini score of 1.6E-2. To further strengthen the involvement of the microRNAs in CRC and to further prioritize them for the candidate microRNAs, only genes that were enriched in CRC were considered. To finally select the genes of interest for the 5 microRNAs, the biological processes (Figure 2; Table 3), expression profile (Figure 3), MFE, and the number of experimentally validation methods were considered. The final list of microRNAs together with their target genes used for the docking analysis were shown in Table 4.

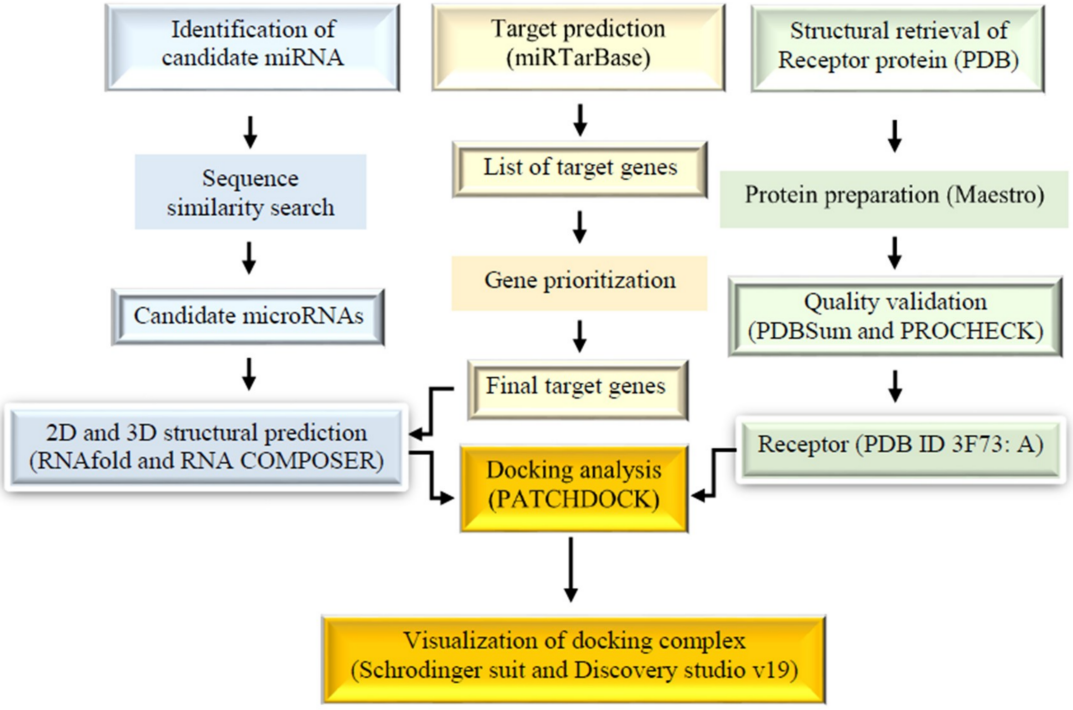


Figure 1. Schematic representation of the methodology.

Table 1. MicroRNAs and their clusters.

Candidate miRNA	Validated microRNA	Fasta Sequences
miR-1	hsa-miR-193a-5p	>hsa-miR-193a-5p MIMAT0004614 UGGGUCUUUGCGGGCGAGAUGA
miR-2	hsa-miR-450b-3p	>hsa-miR-450b-3p MIMAT0004910 UUGGGAUCAUUUUGCAUCCAUA
miR-3	hsa-miR-501-3p	>hsa-miR-501-3p MIMAT0004774 AAUGCACCCGGGCAAGGAUUCU
miR-4	hsa-miR-501-3p	>hsa-miR-501-3p MIMAT0004774 AAUGCACCCGGGCAAGGAUUCU
miR-5	hsa-miR-513a-3p	>hsa-miR-513a-3p MIMAT0004777 UAAAUUUCACCUUUCUGAGAAGG

Table 2. MicroRNA target genes associated with colorectal cancer (CRC) and their MFE (miRTarBase).

miR-1	MFE	miR-2	MFE	miR-3	MFE	miR-4	MFE	miR-5	MFE
A1CF	−19.10	BAMBI	−9.80	SOD2	−16.20	BARD1	−21.30	PDCD4	−11.80
PAQR3	−13.80	XIAP	−8.70	PAQR3	−10.80	SLC1A5	−17.30	VMP1	−12.50
STMN1	−19.90	BMP2	−8.70	SLC7A11	−20.60	WT1	−14.80	CDK4	−10.70
MACC1	−18.00	ZNF703	−13.80	MDM2	−11.90	CLMN	−16.40	TP53	−10.70
FGB	−12.90	PPM1D	−16.90	RAN	−14.70	REL	−19.80	CHEK1	−8.70
HOXB13	−23.90	BUB1	−8.00	LAMB1	−11.52	HDGF	−21.70	H2AFZ	−9.60
ALDOA	−19.20	LYN	−12.90	ORAI2	−19.50			RNF138	−18.20
CHAC1	−20.10	KLF8	−11.02	VAV3	−17.80			SLC7A5	−12.50
GSTK1	−18.10	FGF2	−14.60						
RPS19	−19.10	KMT2A	−17.02						
CRKL	−15.40								
VHL	−19.90								

MFE score based binding affinity between 5 miRNAs and 44 target genes associated with CRC as indicated by miRTarBase.

Table 3. Gene enrichment in cancer and their biological functions.

Gene	Function	miRNA	MFE
TP53	Cell cycle, Apoptosis, Cell proliferation, others	miR-5	−10.70
FGF2	Angiogenesis, Cell proliferation, others	miR-2	−14.60
CHEK1	Cell cycle, Apoptosis, other functions	miR-5	−8.70
WT1	Apoptosis, Cell proliferation, others	miR-4	−14.80
MDM2	Cell cycle, Cell proliferation, others	miR-3	−11.90
BARD1	Cell cycle, Apoptosis, others	miR-4	−21.30
BUB1	Cell cycle, others	miR-2	−8.00
XIAP	Apoptosis, others	miR-2	−8.70
BMP2	Cell proliferation, others	miR-2	−8.70
CDK4	Cell cycle, others	miR-5	−10.70
HOXB13	Angiogenesis, others	miR-1	−23.90
KMT2A	Apoptosis, others	miR-2	−17.02
VHL	Angiogenesis, others	miR-1	−19.90
BAMBI	Other functions	miR-2	−9.80
RAN	Other functions	miR-3	−14.70
REL	Other functions	miR-4	−19.80
RPS19	Other functions	miR-1	−19.10
SOD2	Other functions	miR-3	−16.20

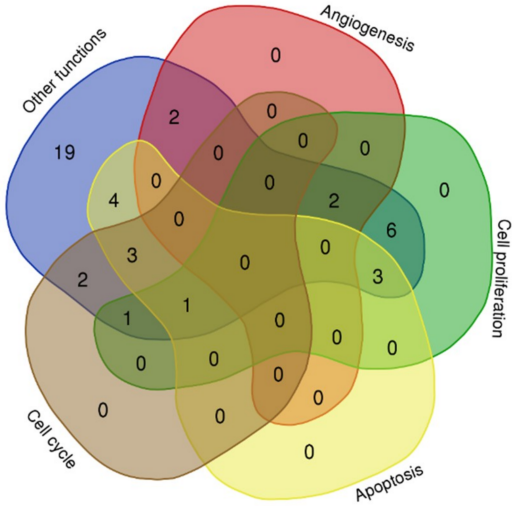


Figure 2. Biological processes of the microRNA target genes.

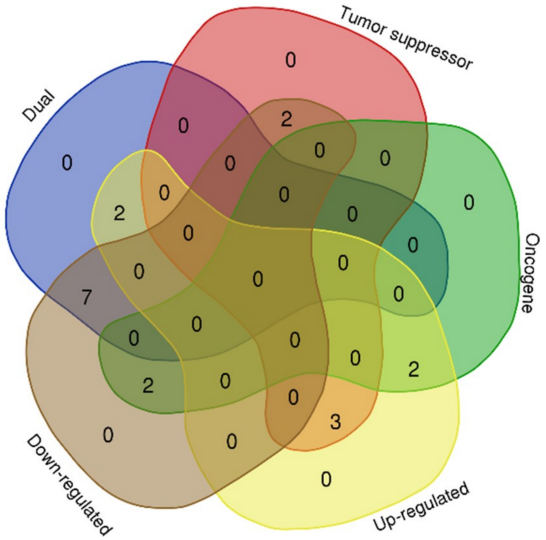


Figure 3. MicroRNA target genes involved in several cancer subtypes.

Table 4. The final list of microRNAs and target genes.

S/N	miRNAs	Target Gene
1	miR-1	HOXB13
2	miR-2	BAMBI
3	miR-3	SOD2
4	miR-4	BARD1
5	miR-5	TP53

2.2. MicroRNA Target Genes Associated with CRC and Their MFE (miRTarBase)

The table above (Table 2) showed the target genes of the five microRNAs discovered through a sequence similarity search implicated in CRC. The miRTarBase prediction tool was used to verify the target genes. These target genes have been experimentally validated by one or more of the following validation methods: Reporter assay, western blot, qPCR, microarray, NGS and pSILAC. Each of the genes was also confirmed by their minimum free energy (MFE) in kcal/mol.

2.3. Biological Processes of the MicroRNA Target Genes

The target genes and their involvement in different biological process plotted using a Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/). The numbers denoted in the plot indicated the number of target genes involved in each of the different biological process (Figure 2).

2.4. Gene Enrichment in Cancer and Their Biological Functions

The involvement of 18 genes from the 44 target genes after gene prioritization through the DAVID database are presented in Table 3 alongside their involvement in different biological functions as reported from UniProt (http://www.uniprot.org/) web server (Table 3).

2.5. Expressions MicroRNA Target Genes Involved in Several Cancer Subtypes

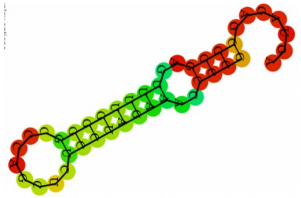
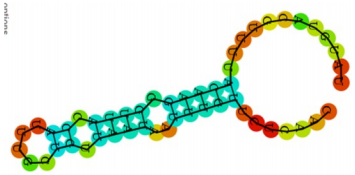
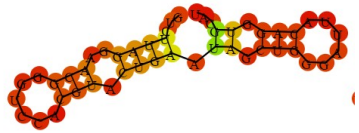
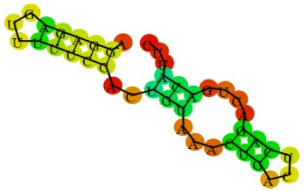
The expression profiles of the 18 genes (Table 3) were considered through an extensive literature search. The result of the search concluded that nine (FGF2, CHEK1, WT1, MDM2, BARD1, BMP2, CHEK4, BAMBI, and SOD2) of the genes have a dual role as oncogene and tumor suppressor genes. FGF2 [32,33] and CHEK4 [34] are up-regulated while the expression of CHEK1 [35,36], WT1 [37], MDM2 [38–41], BARD1 [42–44], BMP2 [45], BAMBI [46], and SOD2 [47–49], have been reported to be down-regulated in several cancer subtypes, including CRC. Furthermore, four (BUB1, RAN, REL, and

RPS19) of the target genes were strictly oncogenic in nature showing that BUB1 [50,51] and RPS19 [52] were up-regulated and RAN [53], and REL [54] were down-regulated in solid tumors including CRC (Figure 3). The remaining five target genes are tumor suppressive in nature. This shows that XIAP [55] and KMT2A [56,57] were up-regulated while TP53 [58], HOXB13 [59,60], and VHL [61,62] were reportedly downregulated in cancers, including CRC.

2.6. Binding Affinity and Structural Determination of MicroRNA and Duplex

The binding energy (BE in Kcal/mol) and minimum folding energy (MFE in kcal/mol) of the microRNA target genes were exploited with two web-based tools namely, miRTarBase and RNAfold respectively. The secondary structures of the duplexes (microRNA-mRNA) were also revealed through the latter webserver (Table 5). The minimum folding energy of all the duplexes is high enough to be regarded as a good binding affinity between the candidate microRNAs and their targets. Therefore, the target genes have strong binding affinity for their respective microRNAs (miR-1 and HOXB13, miR-2 and mRNA, miR-3 and SOD2, miR-4 and BARD1, miR-5 and TP53). These duplexes were finally subjected to the molecular docking interaction.

Table 5. Study of binding affinity between miRNA-mRNA duplex.

Gene	miRNA	Dot-Bracket Notation	2° Structure of Duplex	BE	MFE
HOXB13	miR-1	.....((((..(((((((((.....)))))))))..))))		−13.3	−23.9
BAMBI	miR-2	...........((((((.((((.((.....)).)))))..))))).....		−2.3	−9.6
SOD2	miR-3	...((((..(((.....))).)))).((.(((.....))).))...		−8.5	−16.2
BARD1	miR-4	...........((((((((((....))))))...)))))..		−12.8	−21.3
TP53	miR-5	.((((((...))))))..(((...(((...))))....)))...		−4.0	−10.7

Note: BE- Minimum binding energy in kcal/mol; MFE- Minimum free energy in kcal/mol.

2.7. Structural Model of MicroRNA-mRNA Duplexes

To reveal the three-dimensional structure of the microRNA-mRNA duplexes for the molecular ducking analysis, their binding sequences from the miRTarBase database were used as inputs in the RNAfold web server for a secondary folding pattern and dot-bracket. The secondary structure of the duplexes, their binding energy and the minimum folding energy in kcal/mol are reported in Table 5. The dot-bracket notation generated was also used as inputs in the prediction of the tertiary structure prediction [63] in RNA COMPOSER (Figure 4). The results of Table 5 (binding energies of the duplexes) and Figure 4 (the 3-D structure of the five microRNA-mRNA duplexes) suggested high binding affinity and strong molecular interaction between them.

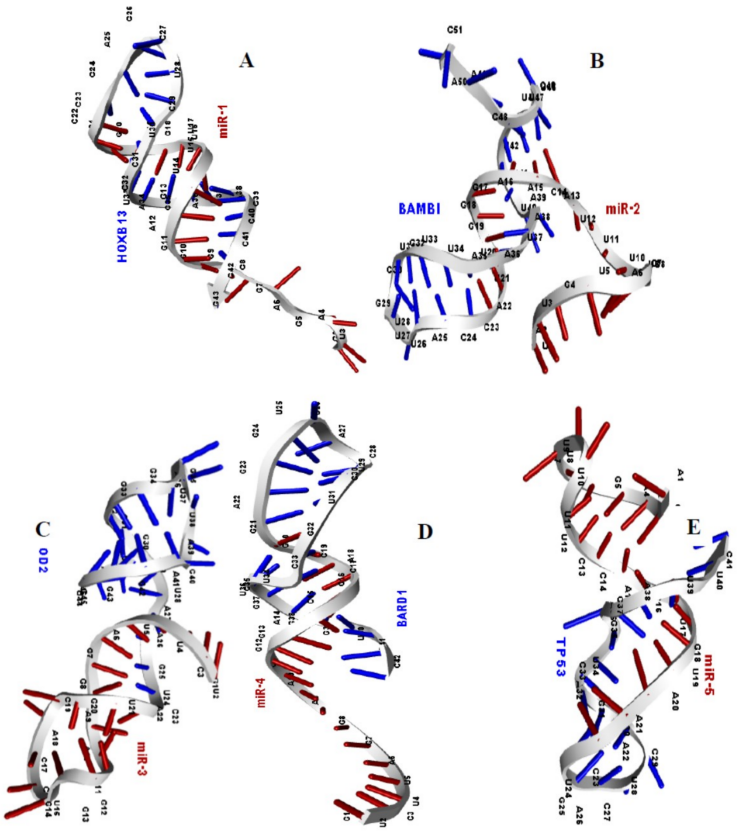


Figure 4. Structural model of miR-1 and mRNA of HOXB13 gene (**A**), miR-2 and mRNA of BAMBI gene (**B**), miR-3 and mRNA of SOD2 gene (**C**), miR-4 and mRNA of BARD1 gene (**D**), miR-5 and mRNA of TP53 gene (**E**), complexes are deciphered, respectively.

2.8. Extraction and Preparation of AGO Protein Structure

The 3D structure of the Argonaute protein was retrieved from the protein data bank (PDB ID: 3F73). In its raw state, AGO is a homodimer with two protein chains A and B, two nucleic acid groups, two molecules of co-factors Mg^{2+}, a molecule of phosphate group and 16 water molecules (chain A, B, C, H, X, and Y). The structural preparation and necessary corrections were carried out using the Maestro Molecular Modelling tool (2019-2), a product of Schrödinger, and discovery Studio v19.1.0. The AGO protein files from PDB were not suitable for immediate use in the molecular modeling calculation due to the fact that they contain heavy atoms which include co-crystallized ligands, water molecule, metal ions, and co-factors. Further, the structure is a homo-dimer with missing atoms and connectivity information. Therefore, protein preparation wizard in Maestro, Schrodinger was used for the preparation and finally, it was reduced to a single chain (A). For the optimization of the H-bond network, PROPKA was employed to re-orientate hydroxyl and thiol groups, water molecules, amide

groups of Asparagine and glutamine, and the imidazole ring of Histidine, as well as the prediction of the protonation states of histidine, aspartate, glutamate, and also the tautomeric states of histidine. The restrained minimization was finally performed to alleviate steric clashes and to relax side-chains (RMSD = 0.030 Å) and water molecules important to the binding receptor, was also maintained at 3. The AGO protein structure (raw and refined chain A) are depicted in Figures 5 and 6 respectively.

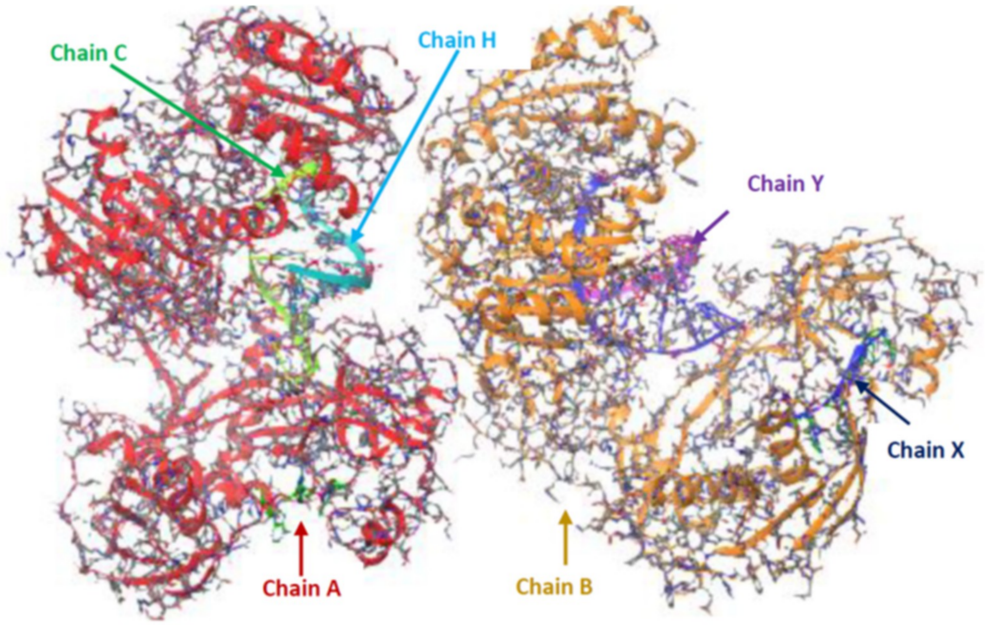


Figure 5. The receptor protein (3D-AGO protein) before preparation visualized by Maestro software.

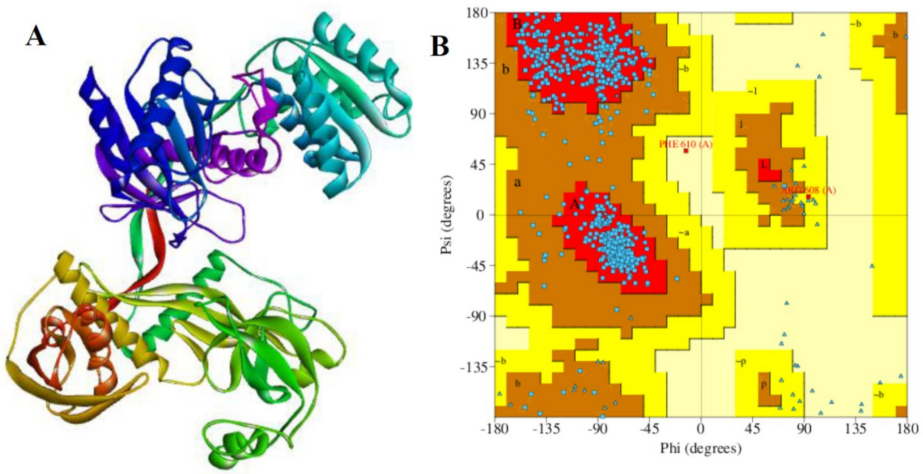


Figure 6. The structural details of the receptor. (**A**) Prepared chain A (Maestro, Discovery studio) and (**B**) its Ramachandran plot (PDBSum, PROCHECK). The quality of the prepared chain A was estimated by PDBSum server. The residues in most favored regions (A, B, L), the residues in additional allowed regions (a, b, l, p) and residues in generously allowed regions (~a, ~b ~l, ~p). The structural details of chain A (ID: 3F37: A) consist of 6 sheets, 9 gamma turns, 12 beta hairpins, 14 beta bulges, 33 strands, and 55 beta turns.

2.9. Validation of Chain A of Argonaute Protein

The quality of the processed chain A was evaluated and validated using PROCHECK, a program that relies on Ramachandran plot for structure verification [64]. As shown in Figure 6A,B, the results from the PROCHECK ascertained that the prepared chain A has 91.5% residues in the most favored

regions and 8.1% residues in the additional allowed regions. Further, 0.2% residues were found in the generously allowed regions and disallowed regions in each case. Therefore, the prepared protein is considered to be of high quality based on the percentage distribution of the amino acid residues. Furthermore, a G-factor that provides a measure of how unusual or conversely, how usual a given stereochemical property is [65], was also determined using this program. A G-factor of less than −0.5 is unusual and less than −1.0 indicates highly unusual. However, the generate G-factor for the prepared chain A of the receptor protein was −0.34 for dihedral angels, −0.04 for main chain covalent forces and −0.20 overall.

2.10. Docking Analysis Between Receptor Protein and MicroRNA

PatchDock as molecular docking method was used for the docking interaction between the microRNAs and the AGO protein. The PDB file of the AGO protein and each of the candidate microRNAs was used as inputs. The root mean square deviation (RMSD) is often used to measure the quality of reproduction of the correct pose and to validate the docking protocol. For a true binding pose to be good, the RMSD must be low, therefore, the clustering RMSD was adjusted to 1.5 Å. The method of PatchDock relies on the shape complementarity theory [66]. A previous study also reported the reliability and usability of the Patch dock tool in molecular docking analysis [67]. The result files generated for each of the microRNAs were ranked according to their geometric shape complementarity score. For the first round of docking, the result with the highest score (geometric shape complementary) was chosen as the best microRNA-AGO complex [67] for each of the five candidate microRNAs (Table 6). The strong binding affinity of these results was observed through their scores and the amino acid residues involved in the interaction between the microRNAs and the AGO protein. As evident, the presence of strong hydrophobic amino acids (mir-1: 21; mir-2: 20; mir-3: 27; mir-4: 22; and mir-5: 27) and amino acids with aromatic side chains (miR-1: 7; mir-2: 3; mir-3: 6; mir-4: 4; and mir-5: 7) within the distance of 3.5 Å (Figure 7; Table 7), and the hydrogen bond within the distance of 2.0 Å are supportive that gene regulation through the argonaute protein are driving by microRNA (Table 8).

Table 6. The docking scores between miRNA and AGO protein.

miRNA-mRNA and AGO	Score	Area	ACE
miR-1 -AGO	19544	3390.80	−258.22
miR-2-AGO	18618	2832.70	−22.43
miR-3-AGO	18420	2814.10	−151.43
miR-4-AGO	18024	2344.20	−131.18
miR-5-AGO	20.372	2913.20	−488.07

The score indicates the geometric shape complementary score and atomic contact energy (ACE) score generated for each miRNA and AGO complex. miRNA, microRNA; AGO, argonaute; ACE, atomic contact energy.

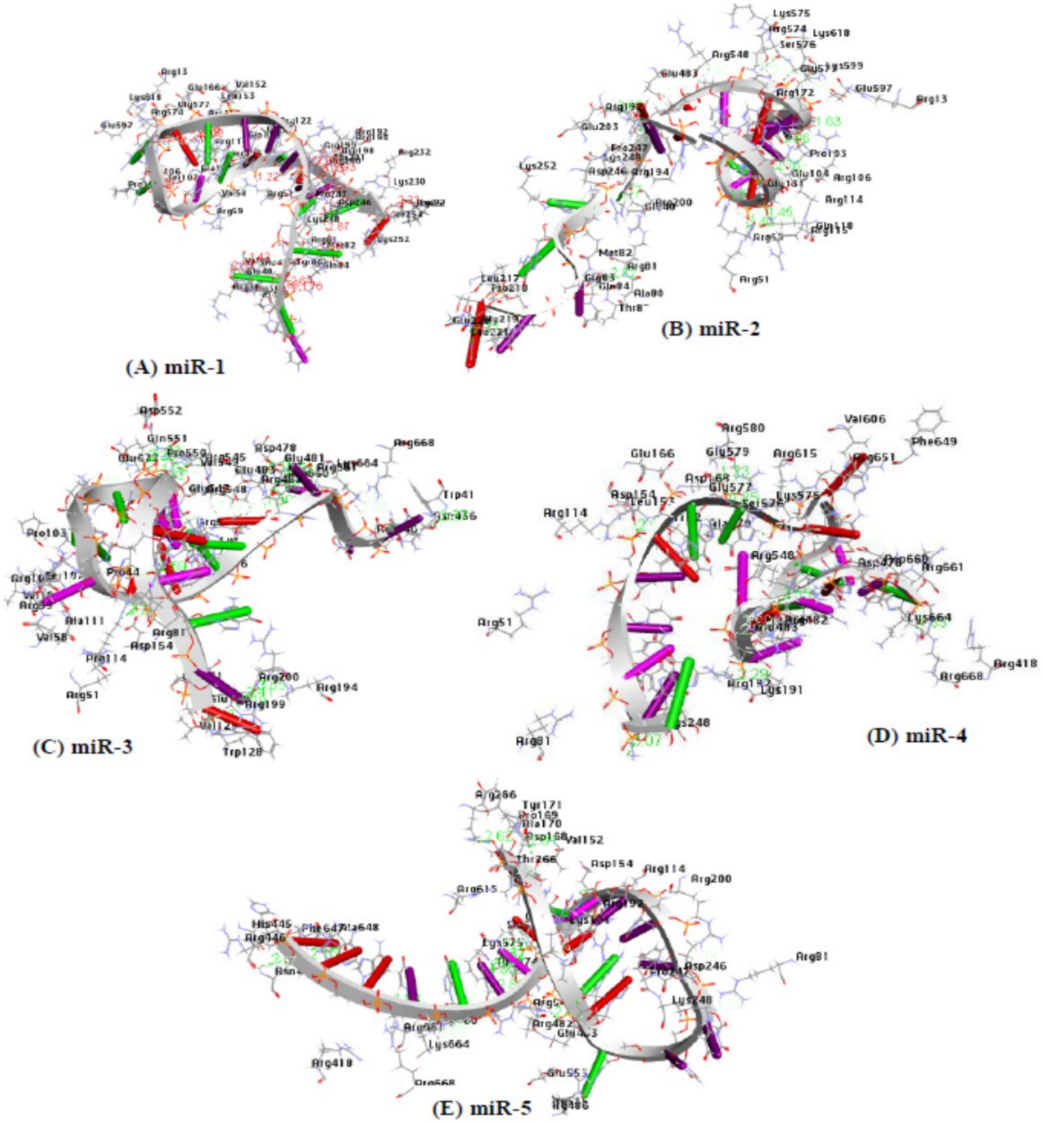


Figure 7. The amino acid residues of Argonaute protein participating in the interaction with each of the five candidate microRNAs within a distance of 3.5 Å are deciphered, respectively. (**A**) amino acids participating in miR-1-Agonaute protein duplex, (**B**) amino acids participating in mir-2-Agonaute protein duplex, (**C**) amino acids participating in mir-3-Agonaute protein duplex, (**D**) amino acids participating in mir-4-Agonaute protein duplex, and (**E**) amino acids participating in mir-5-Agonaute protein duplex.

Table 7. Molecular docking analysis results of microRNA and receptors with participating aa residues (3.5 Å).

miRNA	Hydrophobic AA	Aromatic AA	H-Bond
	(21 [a]), LEU45 [d], ALA47 [d], VAL58 [d], VAL108 [d], ALA111 [d], LEU112 [d], VAL129 [d], LEU132 [e], ALA133 [e], LEU217 [d], ALA245 [d], ILE254 [d], VAL264 [d], LEU596 [d]	(7b), TYR43 [d], TYR135 [d], TRP156 [e], TRP202 [d]	(25 [c]) ARG114 [d], ARG574 [d], GLY577 [d], LYS248 [d], ASP246 [d], ASP154 [d], ARG200 [d], GLY131 [d], PRO103 [d], LEU153 [d]
miR-2	(20 [a]), ALA47 [d], VAL58 [d], VAL108 [d], ALA111 [d], LEU112 [d], LEU132 [e], ALA133 [e], VAL152 [d], LEU153 [d], LEU217 [d], ALA245 [d], ILE254 [d], VAL264 [d], VAL549d [d], LEU596 [d], VAL620 [d]	(3 [b]), TYR43 [d], TRP156 [e], TRP202 [d]	(21 [c]) ARG114 [d], ARG574 [d], GLY577 [d], LYS248 [d], ASP246 [d], ARG548 [d], GLU483 [d], SER576 [d], ARG192 [d], LYS599 [d], ARG81 [d]
miR-3	(27 [a]), ALA47 [d], VAL58 [d], LEU64 [d], VAL108 [d], ALA111 [d], LEU112 [d], VAL129 [d], LEU132 [e], ALA133 [e], VAL152 [d], LEU153 [d], ALA450 [d], ALA479 [d], VAL549 [d], VAL620 [d], LEU652 [d], VAL663 [d]	(6 [b]), TYR43 [d], TRP156 [e], TRP447 [d]	(26 [c]) ARG114, ARG574, GLY577, ASP154, ARG548 GLU483, LYS664 ARG661, ARG200 GLY131 PRO103, LYS599 ARG81, ASP660,
miR-4	(22 [a]), ALA47 [d], LEU132 [e], ALA133 [e], ALA151 [d], VAL152 [d], LEU153 [d], ALA170 [d], ILE173 [d], VAL264 [d], LEU265 [d], LEU267 [d], LEU279 [d], ALA479 [d], VAL573 [d], ALA648 [d], LEU652 [d], LEU662 [d], VAL663 [d]	(4 [b]), TYR135 [d], TRP156e, PHE649 [d]	(15 [c]) ARG114, LYS248, ARG548 GLU483, SER576 ARG192, LYS664 ARG661, LEU153, THR266 LYS575 ARG482
miR-5	(27 [a]), LEU132 [e], ALA133 [e], ALA151 [d], VAL152d, LEU153 [d], ALA170 [d], ILE173 [d], VAL264 [d], LEU265 [d], LEU267 [d], LEU279 [d], ALA450 [d], ALA479 [d], VAL549 [d], VAL573 [d], ALA648 [d], LEU652 [d], LEU662 [d], VAL663 [d]	(7 [b]) TYR135 [d], TRP156 [e], TRP447 [d], PHE649 [d]	(17 [c]) ARG574 [d], ASP246 [d], ASP154 [d], SER576 [d], ARG192 [d], LYS664 [d], ARG661 [d], ASP660 [d], THR266 [d], LYS575 [d], ARG482 [d]

[a] Total number of residual hydrophobic amino acids involved in the interaction between the receptor and the candidate microRNAs; [b] Total number of aromatic amino acids involved in the interaction between the receptor and the candidate microRNAs; [c] Total number of hydrogen bond observed in the interaction between the receptor and the candidate microRNAs; [d] The residual amino acids of the receptor protein common to more than one interaction between microRNA binding to receptor; [e] The residual amino acids of the receptor protein common to all the microRNA binding to AGO.

Table 8. Hydrogen bond interaction between the amino acid residues of the receptor and the candidate microRNAs within the distance of 2.0 Å.

microRNA	AA Residues	Atoms	Distance	NA Residues
miR-1	GLN84	HE21-OP1	1.8	(G3)
	ARG574	HH11-O3′	1.7	(G12)
	ALA111	HA-O2′	1.9	(G16)
	PRO36	O-H4′	1.8	(G2)
	ASP154	OD1-H5′	2.0	(A11)
		O-H4′	2.0	(G2)
miR-2	GLY104	HA3-O6	1.8	(G15)
	ARG114	HD3-O4′	1.6	(G19)
	ARG574	HD3-OP1	1.5	(A10)
	GLU483	OE1-H5	2.0	(A8)
	ARG59	O-H4′	1.9	(A17)
miR-3	ARG548	HH11-O2′	1.9	(A17)
	ARG574	HH22-O4′	1.9	(G)
	VAL129	O-HO5′	1.8	(A1)
	ASP154	OD1-HO2′	1.6	(A13)
	PRO44	HA-O3′	2.0	(C4)
	GLY577	HA2-O2′	2.0	(A14)
	ARG661	HA-O2′	1.9	(U19)
	GLU622	OE1-H5′	1.8	(G9)
	ASP660	O-H2′	2.0	(U19)
miR-4	ARG668	HH12-O5′	1.8	(G1)
	ARG615	HD2-OP2	2.0	(A12)
	THR266	OG1-H5′	2.0	(G8)
miR-5	LYS575	HZ1-O2	1.9	(U6)
	ARG661	HE-O4′	1.9	(U4)
	ARG574	HD2-O4′	2.0	(A8)
	SER576	H-O2	2.0	(C7)

2.11. Hydrogen Bond Interaction

Hydrogen-bonds (H-bond) are an important interaction which dictate the specificity of ligand binding. Their important contribution is explicitly incorporated into the molecular simulation to enhance the binding of molecules to their receptors in an energetically favorable manner [68]. For protein-ligand interactions, hydrogen bonds have been thought to play some significant roles. These roles include the orientation of the binding molecule, ligand recognition, and binding affinity. The latter is one of the most important issues to be considered in protein-ligand interaction. The highest number of hydrogen bond interactions were found among the interacting atoms of miR-1, miR-3 and the residual amino acid of the receptor protein binding pocket with 45 H-bond and 35 H-bond respectively (Figure 8A,C). For miR-2, miR-5 and receptor protein, a total of 28 H-bonds were involved (Figure 9B,E), while the lowest number of hydrogen bonds (18 H-bonds) was observed among the interacting atoms of miR-4 and the receptor protein (Figure 8D). All the hydrogen bonds observed in Figure 8 are within the distance of 3.5 Å. Table 7 shows the residues of the amino acids involved in hydrogen bonding between the Argonaute protein and the microRNAs within the distance of ≤ 2.0 Å. The hydrogen bonds are key to the determination of the interaction (protein-ligand) therefore, they are fundamental to the biological process [69]. The results revealed that the higher the number of favorable interactions, the more the hydrogen bonds. This result may, therefore, support the mechanism by with microRNAs regulates gene expression through RISC.

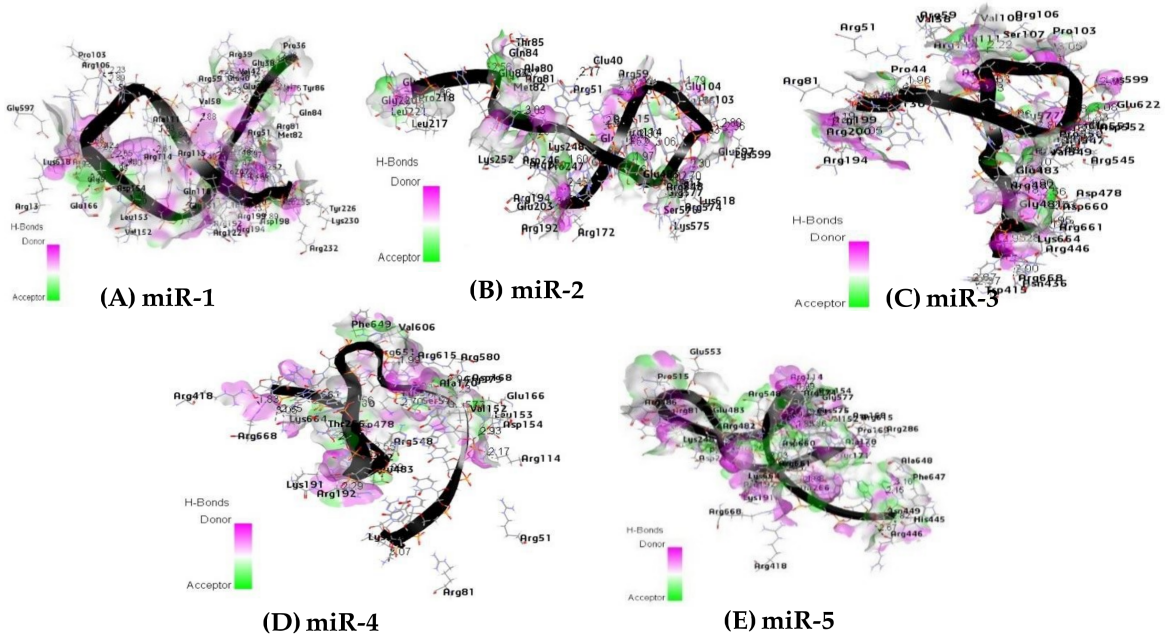


Figure 8. Hydrogen bond interaction between the residual amino acids of the receptor protein and the microRNAs (**A**) miR-1, (**B**) miR-2, (**C**) miR-3, (**D**) miR-4, and (**E**) miR-5 respectively (3.5 Å distance).

2.12. Docking Analysis between Argonaute Protein and MicroRNA-mRNA Complex

Similar to the docking analysis of microRNAs to the argonaute protein, the microRNA-mRNAs complexes between the candidate microRNAs and their target genes were further docked against the argonaute protein and possible binding interaction in terms of hydrophobicity, aromatic residual amino acids, and hydrogen bonding was analyzed. The docking was carried out on the argonaute protein (chain A) and miR-1-HOXB13; miR-2-BAMBI; miR-3-SOD2; miR-4-BARD1; and miR-5-TP53 separately in PATCHDOCK. Based on the geometric scoring analysis, the highest score for each of the complexes were reported in Table 9.

Table 9. Docking scores between miRNA-mRNA and AGO protein.

miRNA-mRNA and AGO	Score	Area	ACE
miR-1-HOXB13-AGO	24046	3962.90	−851.20
miR-2-BAMBI-AGO	24380	5528.70	−966.63
miR-3-SOD2-AGO	27570	3974.80	−652.52
miR-4-BARD1–AGO	24816	3524.00	−836.85
miR-5-TP53-AGO	23716	3402.30	−547.97

Score indicates the geometric shape complementary score and ACE score generated for each miRNA-mRNA and AGO complex. miRNA, microRNA; AGO, argonaute; ACE, atomic contact energy.

In nature, strong hydrophobic amino acids together with amino acids with aromatic side chains are important to binding interactions in terms of stability between the receptor and the ligand. Therefore, the binding interaction between the 5 complexes (miR-1-HOXB13, miR-2-BAMBI, miR-3-SOD2, miR-4-BARD1, and miR-5-TP53) and argonaute protein (chain A) was investigated by examining the residual amino acids in the binding pocket of the argonaute protein within the distance of 3.5 Å. The residual strong amino acids of the receptor (argonaute protein) VAL42, LEU45, VAL129, LEU132, ALA133, VAL147, ALA151, VAL152, LEU153, TRP156, ALA170, TYR171, ILE173, LEU174, VAL193, VAL264, LEU265, LEU267, LEU277, LEU279, LEU281, ALA331, ALA414, ILE434, ALA479, VAL549, VAL573, VAL606, LEU617, ALA644, LEU652, LEU658, VAL663, and VAL685 in miR-1 and HOXB13 complex; VAL42, LEU45, LEU46, ALA47, ALA50, VAL58, ALA111, LEU132, LEU189, LEU204, LEU205,

VAL264 LEU265, LEU267, LEU270, LEU277, ALA331, LEU389, LEU395, ALA414, LEU421, ALA423, LEU424, LEU435 LEU439, ALA479, LEU505, ALA508, LEU522, VAL549, ALA648, ALA659, LEU662, VAL663, VAL666, ILE671, LEU674, and VAL677 in miR-2 and BAMBI complex; VAL42, LEU45, LEU46, ALA47, VAL58, VAL108, ALA111, LEU112, LEU132, ALA133, VAL152, ALA170, ILE173, LEU215, LEU217, ILE254, LEU265, LEU267, LEU279, LEU281, VAL606, LEU617, ALA648, and LEU652 in miR-3 and SOD2 complex; VAL42, LEU45, VAL147, VAL152, ALA170, ILE173, VAL264, LEU265, LEU267, LEU277, ALA278, LEU279, LEU281, ILE434, LEU435, ALA450, LEU452, VAL573, VAL606, ALA644, ALA648, LEU652, and VAL685 in miR-4 and BARD1 complex; VAL42, LEU45, VAL147, VAL152, ALA170, ILE173, VAL264, LEU265, LEU267, LEU277, ALA278, LEU279, LEU281, ILE434, LEU435, ALA450, LEU452, VAL573, VAL606, ALA644, ALA648, LEU652, and VAL685 in miR-5 and TP53 complex (3.5 Å). Similarly, amino acids such as TYR43, TYR135, TRP202, TRP415, TYR642, PHE647, PHE649, and PHE684 in miR-1 and HOXB13 complex; TYR43, TRP182, TRP202, TRP415, TRP447, PHE485 PHE487, TRP503, PHE610, PHE647, PHE649, and PHE683 in miR-2 and BAMBI complex; TYR43, TYR86, TYR171, TRP202, TRP243, TRP415, PHE487, PHE647, and PHE649 in miR-3 and SOD2 complex; TYR43, TYR135, TYR171, TRP202, PHE360, TRP447, PHE487, PHE610, TYR642, and PHE649 in miR-4 and BARD1 complex; and TYR43, TYR135, TYR171, TRP202, TRP283, TRP447, PHE610, TYR642, PHE647, and PHE649 in miR-5 and TP53 with aromatic ring are also found as participating in the interaction within the binding pocket of the receptor protein (3.5 Å) (Table 8 and Figure 9).

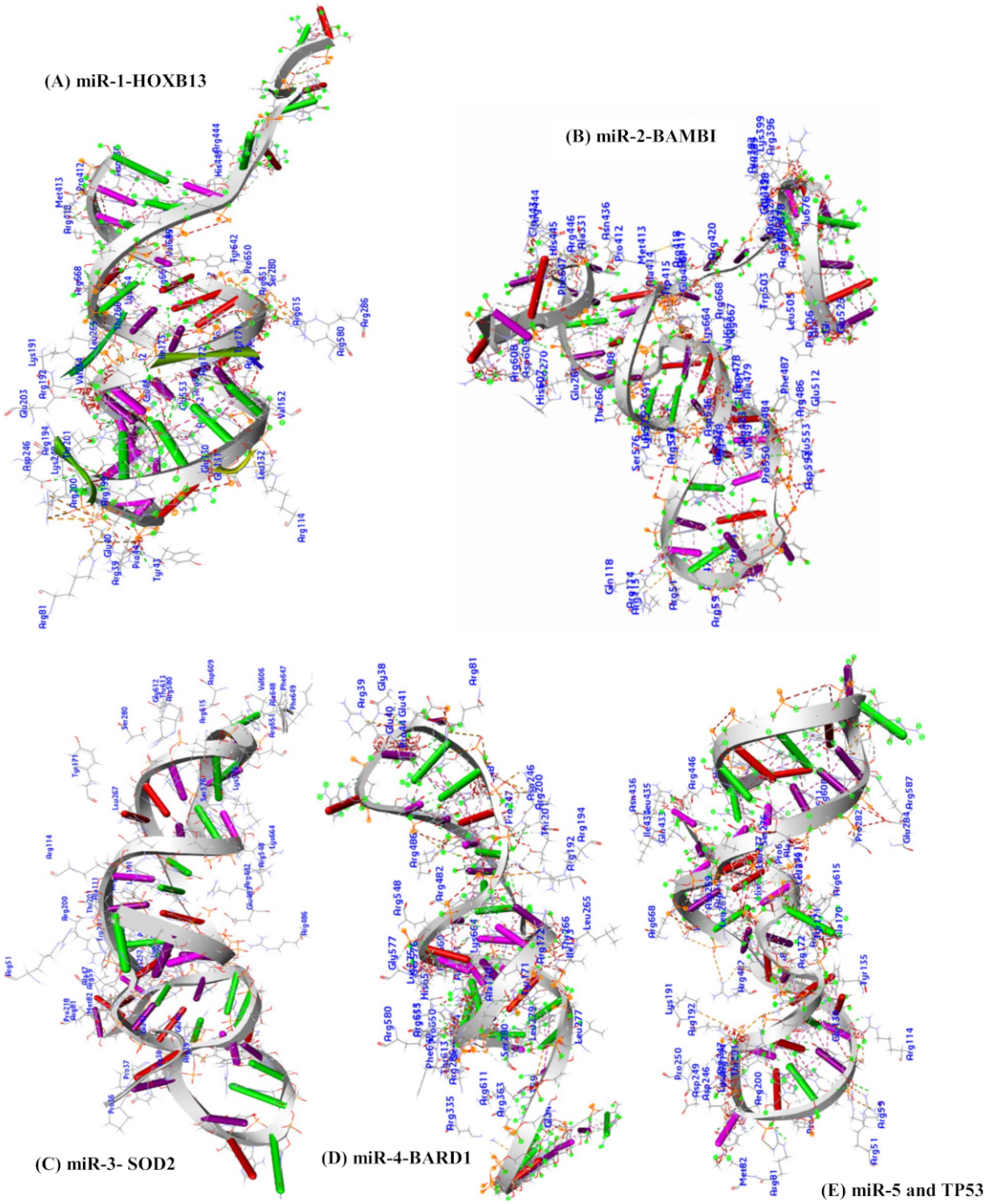


Figure 9. Docking complex results of microRNA-mRNA and the receptor (Chain A of argonaute protein). The amino acid residues participating in the interaction between the receptor and miR-1-HOXB13 (**A**); Residual amino acids participating in the interaction between the receptor and miR-2-BAMBI (**B**); miR-3-SOD2 (**C**); miR-4-BARD1(**D**); and miR-5- TP53 (**E**).

3. Discussion

The study aimed to predict the mechanism of gene regulation mediated through microRNAs involved in CRC using the in silico approach. Since the discovery of microRNA, several studies have reported their involvement in a variety of physiological and pathological processes and mutations affecting their normal expression which may be critical to their role in the development of human diseases [70–72], such as cardiovascular diseases [73,74], neurodegenerative diseases [75–77],

and several cancer subtypes [21,72,78,79]. Additionally, many studies have investigated the diagnostic [80,81] and therapeutic roles [82] of this non-coding RNAs in human diseases. These microRNAs are able to control gene expression in a sequence-specific manner, most especially in the mechanism of gene silencing by forming RISC comprising the argonaute protein [83]. Experimental approaches have been used to study the RNA induced silencing complex at both the molecular and atomic levels [84–86].

In light of this, the molecular interaction between chain A of the argonaute protein, microRNAs and the target genes were investigated in CRC with the in silico approach.

In this study, the results of BLASTN and CD-HIT-EST-2D were obtained from 125 validated query sequences and 2226 total microRNA sequences as the reference microRNAs. The microRNAs obtained from BLASTN were based on the parameters: (1) The expected value of 1e-2, (2) the word size of 7 and (3) a similarity index between 90%–99%. The result of the CD-HIT-EST-2D obtained was based on a threshold of 0.90 and word size of 7.

Five microRNAs (Table 1) were finally retained after a sequence similarity search based on their uniqueness after text-mining to show their non-involvement in CRC. These microRNAs targeted 44 genes (Table 2) which were further reduced to five based on the set criteria (involvement in CRC, expression pattern, MFE, biological processes, and their validation methods). The minimum free energy (kcal/mol) of the binding affinity of both the target genes and their microRNAs were further studied in other to verify their interaction strength. After studying the biological processes of these genes, their enrichment in cancer and CRC were also identified through the DAVID database.

The aberrant expression of HOXB13 is associated with CRC [87]. The expression of this gene in the early embryonic development of the intestine represents embryogenic phases in an important tissue-specific marker [88]. Therefore, it strongly correlates with lymph nodes metastasis (TNM) [89]. The over-expression of BAMBI has also been detected in colorectal cancer [90]. This gene has further been linked with late-stage (M) in CRC [91,92]. The expression of SOD2 is increased in pre-malignant (T and N) stages during colorectal carcinogenesis whereas SOD1 is expressed only in colorectal tumors [93]. The aberrant expression of BARD1 is associated with the drivers of various types of cancer [94]. TP53 is correlated with overall survival in stage II and III CRC patients [95].

The selected microRNA target genes were involved (HOXB13, BAMBI, SOD2, BARD1, and TP53) in various biological processes which are crucial to carcinogenesis in CRC. The minimum free energies of −23.9 kcal/mol for HOXB, −13, 9.6 kcal/mol for BAMBI, −16.2 kcal/mol for SOD2, −21.3 kcal/mol for BARD1, and −10.7 kcal/mol for TP53 confirmed that that the binding interaction between the candidate microRNAs and their target genes were energetically favorable, which can be confirmed by the binding energies of each duplex (Table 5). Additionally, to investigate the mechanism by which the candidate microRNAs miR-1, miR-2, miR-3, miR-4, and miR-5 bind argonaute protein in RNA induced silencing complex to target specific genes namely, HOXB13, BAMBI, SOD2, BARD1, and TP53, their molecular interactions were studied.

Prior to the molecular study, microRNAs and microRNA-mRNA duplexes were converted to PDB format. The argonaute protein (receptor) was also downloaded alongside from the protein data bank. As the raw structure is a homodimer consisting of heavy atoms (co-factors, water molecules, metal ions, and co-crystallized ligands) and is of limited resolution due to the x-ray crystallography experiment, the structure was checked (Figure 5).

Specifically, the protein preparation wizard in MAESTRO was used to optimize the hydrogen bond network (PropKa), and alleviate the steric clashes (restrained minimization) by force field: OPLS_2005, Epik was used to generate the het states and finally, missing atoms were fixed using PRIME. The prepared protein was validated using PROCHECK and PDBSum (Figure 6).

The docking algorithm (PATCHDOCK) was employed to computationally study the miRNA-protein and microRNA-mRNA-protein interactions. In order to estimate the strength of the interactions between the receptor and microRNAs, the molecular docking results (argonaute and

microRNA and microRNA-mRNA) were estimated by examining the structural components and binding affinity [96].

The general interactions between the receptors and ligand include hydrophobic, hydrogen, pi stacking, weak hydrogen bond, salt bridge, amide stacking, and cation pi. The molecular docking results of the receptor-microRNA interaction and receptor-(microRNA-mRNA) interaction indicated that the non-covalent interactions include hydrophobic interactions between the residual amino acids of the protein and specific atoms of the microRNA and or mRNAs. The hydrogen bonds and the pi stacking bonds, which are the most common interactions, are also observed in the binding analysis to prove that microRNA is crucial to gene regulation.

Rath et al. [97] reported that the presence of aliphatic amino acids such as, isoleucine, leucine, valine, and alanine, which are strong hydrophobic in nature, confer stability during molecular interaction in protein-ligand binding. Further, amino acid residues which are relatively hydrophobic with aromatic side chains such as tryptophan, tyrosine, and phenylalanine provide steadiness towards the binding stability within the binding pocket of a protein.

From the docking analysis of the receptor and the candidate microRNAs, the presence of strong and relatively strong hydrophobic residual amino acid and aromatic rings observed between the candidate microRNAs and the receptor protein (chain A of argonaute protein) together with hydrogen bond interactions within the distance of 3.5Å (Table 7, Figure 7) are proofs that the molecular interaction involved are favorable and stable at the atomic level respectively.

The amino acid residues of the receptor participating in the hydrogen bonding interaction with the candidate microRNAs at the molecular level within the distance of 2.0 Å are also reported in Table 8. This H-bonding interaction strongly assists in receptor stability through the candidate microRNAs during gene regulation (Figure 8). Previous studies have reported that hydrogen bonding between the interaction of two molecules, such as protein and ligand, are important interactions driving potent binding and selectivity [98] and stabilizing ligand conformation [99]. Furthermore, the presence of strong hydrophobic amino acids namely; LEU 45, LEU 265, and LEU 267 and aromatic rings of amino acids TYR 43, TRP 202 and PHE 649 (relatively strong hydrophobic) during molecular interaction of the receptor with miR-1 and HOXB13, miR-2 and BAMBI, miR-3 and SOD2 miR-4 and BARD1, and miR-5- TP53 complex (Table 9, Figure 9 and Table 10) are observed to be commonly participating in all the microRNAs together with their targets in the receptor-binding pocket within the distance of 3.5Å. The hydrophobic contacts are the most common interactions in protein-ligand complexes. The most common hydrophobic interaction is the one formed by an aliphatic carbon in the receptor and an aromatic carbon in the ligand. Leucine, followed by valine, isoleucine and alanine side-chains are the most frequently engaged in hydrophobic interactions [100].

Table 10. Amino acid residues of the binding pocket of the argonaute protein involved in the molecular interaction with the microRNA-mRNA complex (3.5 Å).

miRNA-mRNA	Residual Hydrophobic AA	Aromatic AA
miR1-HOXB13	(34 [a]), VAL42 [c], LEU45 [d], LEU132 [c], ALA133 [c], VAL147 [c], ALA151 [c], VAL152 [c], ALA170 [c], ILE173 [c], VAL264 [c], LEU265 [d], LEU267 [d], LEU277 [c], LEU279 [c], LEU281 [c], ALA331 [c], ALA414 [c], ILE434 [c], ALA479 [c], VAL549 [c], VAL573 [c], VAL606 [c], LEU617 [c], ALA644 [c], LEU652 [c], VAL663 [c], VAL685 [c]	(8 [b]), TYR43 [d], TYR135 [c], TRP202 [d], TRP415 [c], TYR642 [c], PHE647 [c], PHE649 [d]
miR-2-BAMBI	(38 [a]), VAL42 [c], LEU45 [d], LEU46 [c], ALA47 [c], ALA50 [c], VAL58 [c], ALA111 [c], LEU132 [c], VAL264 [c], LEU265 [d], LEU267 [d], LEU277 [c], ALA331 [c], ALA414 [c], LEU435 [c], LEU439 [c], ALA479 [c], VAL549 [c], ALA648 [c], VAL663 [c].	(12 [b]), TYR43 [d], TRP202 [d], TRP415 [c], TRP447 [c], PHE487 [c], PHE610 [c], PHE647 [c], PHE649 [d]
miR-3-SOD2	(24 [a]), VAL42 [c], LEU45 [d], LEU46 [c], ALA47 [c], VAL58 [c], ALA111 [c], LEU132 [c], ALA133 [c], VAL152 [c], ALA170 [c], ILE173 [c], LEU265 [d], LEU267 [d], LEU279 [c], LEU281 [c], VAL606 [c], LEU617 [c], ALA648 [c], LEU652 [c]	(9 [b]), TYR43 [d], TYR171 [c], TRP202 [d], TRP415 [c], PHE487 [c], PHE647 [c], PHE649 [d]
miR-4-BARD1	(23 [a]), VAL42 [c], LEU45 [d], VAL147 [c], VAL152 [c], ALA170 [c], ILE173 [c], VAL264 [c], LEU265 [d], LEU267 [d], LEU277 [c], ALA278 [c], LEU279 [c], LEU281 [c], ILE434 [c], LEU435 [c], ALA450 [c], VAL573 [c], VAL606 [c], ALA644 [c], ALA648 [c], LEU652 [c], VAL685 [c]	(10 [b]), TYR43 [d], TYR135 [c], TYR171 [c], TRP202 [d], TRP447 [c], PHE487 [c], PHE610 [c], TYR642 [c], PHE649 [d]
miR-5-TP53	(30 [a]), LEU45 [d], ALA47 [c], ALA50 [c], VAL58 [c], ALA111 [c], LEU132 [c], ALA133 [c], VAL147 [c], ALA151 [c], VAL152 [c], ALA170 [c], ILE173 [c], VAL264 [c], LEU265 [d], LEU267 [d], LEU277 [c], ALA278 [c], LEU279 [c], LEU281 [c], ILE434 [c], LEU435 [c], LEU439 [c], ALA450 [c], VAL606 [c], LEU617 [c], ALA644 [c], ALA648 [c], LEU652 [c], VAL685 [c]	(10 [b]), TYR43 [d], TYR135 [c], TYR171 [c], TRP202 [d], TRP447 [c], PHE610 [c], TYR642 [c], PHE647 [c], PHE649 [d]

AA- amino acid; [a] Total hydrophobic residual amino acid involved in docking interaction; [b] Total aromatic ring containing amino acid residues with aromatic rings; [c] The residual amino acid of the receptor involved in interaction common to more than one complex interaction; [d] The residual amino acids of the receptor protein common to all the microRNA-mRNA binding to AGO.

4. Materials and Methods

4.1. MicroRNA Identification

A sequence similarity search was employed to identify candidate microRNAs between the total microRNA sequences (obtained from miRBase) at http://www.mirbase.org/ and microRNAs associated with CRC obtained from four different experimentally validated CRC microRNA databases (dbDEMC, HMDD, miR2Disease, and miRCancer). The parameters for the command line include a sequence identity threshold of 0.90; an E-value of 1e-3; a word size of 7.

4.2. Target Prediction and Correlation to CRC

The miRTarBase is an experimentally validated microRNA-target interactions database at http://mirtarbase.mbc.nctu.edu.tw/php/index.php. Generally, this database is composed of targets experimentally validated through reporter assay, western blot, microarray, and next-generation sequencing experiments [101,102]. The gene browser for CRC (gbCRC) at http://gbcrc.bioinfo-minzhao.org/ and CRC for the gene database (CoReCG) at http://lms.snu.edu.in/corecg/gene are databases containing only validated CRC genes.

The miRTarBase, gbCRC, and CoReCG were used to identify and correlate the targets of the microRNAs.

Furthermore, the targets prioritization was carried out using the Database for Annotation, Visualization and Integrated Discovery (DAVID) at https://david.ncifcrf.gov/ for functional enrichment

in CRC [103,104]. To finally select the genes of interest, the expression profile, the biological processes, the minimum free energy score based binding affinity between the targets and the microRNAs (MFE), and the number of experimentally validation methods were considered.

An intersection analysis tool accessed at http://bioinformatics.psb.ugent.be/webtools/Venn/ was used to create Venn diagrams of the involvements of the target genes in the cancer subtypes and their functions.

4.3. Structural Prediction of Candidate MicroRNA and Target Complexes

To determine the secondary structure and the dot-bracket notation of both the microRNAs and their targets, the RNAfold web server was employed. This software at http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi is used to predict the secondary structure of single-stranded RNA or DNA sequences, including their folding energy. The dot-bracket annotations generated were therefore used as inputs in the RNA-COMPOSER (http://rnacomposer.cs.put.poznan.pl/) to generate the 3-dimensional structures of their duplex.

4.4. Protein Selection and Preparation

The sole component of the RNA induced silencing complex (RISC) (AGO protein) was retrieved in PDB format from the protein data bank with the ID: 3F73 (DOI: 10.2210/pdb3F73/pdb) at https://www.rcsb.org/structure/3f73. The molecule was further prepared and visualized using Schrödinger, 2019 suit and discovery studio v19. The protein preparation wizard in Maestro was to optimize the hydrogen bond network (PropKa), alleviation of steric clashes (restrained minimization) by force field: OPLS_2005, Epik was used to generate the het states and finally, missing atoms were fixed using PRIME. The prepared protein was validated using PROCHECK and PDB Sum.

4.5. Molecular Docking

In silico protein-ligand docking was performed using the webserver PATCHDOCK at https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php. The molecular docking between the receptor protein (chain A of the AGO protein) and the microRNA and the microRNA-mRNA complex with chain A of the AGO protein was carried out. The PATCHDOCK software is based on the shape complementarity of the interactions to generate the best candidate solution [67]. The clustering root-mean-square deviation (RMSD) was chosen to be 2.0 Å with the complex type protein-small ligand. The microRNAs, microRNA-mRNA, and AGO were all converted to PDB file formate and were used as inputs into the PATCHDOCK webserver. The results generated were presented in PDB based on the geometric shape complementary score, the approximate interface area (AI area), and the atomic contact energy with their transformation files. The pose with the highest score was considered as the best complex [67].

Finally, the interactions (including receptor surface (hydrogen bond and charge) and binding observed in the docked conformations in the PDB format were analyzed and inspected with Maestro and PDB sun and visualized using the discovery studio v19 software.

4.6. Statistical Analysis

For microRNA selection, the BlastN parameters were set at 1e-2 for expected value, 7.0 for word size, and 90–99% for similarity index. The CD-HIT-EST-2D parameters were set at 0.90 for threshold and 7.0 for word size. The genes considered in DAVID were regarded statistically significant at p-value of 1.8E-3 with the Benjamini score of 1.6E-2. The root-mean-square deviation (RMSD) was set at 0.030 Å for the protein preparation in Schrödinger (restrained minimization). The protein quality check at PROCHECK was also considered significant at 90% and above for residues in the most favored regions. In PatchDock, RMSD was adjusted to 1.5 Å. In discovery studio, the amino acid residues were considered within the distance of 3.5 Å, while hydrogen bonding was considered between 2–3.5 Å. A value of $p < 0.05$ was considered to indicate a statistically significant difference.

5. Conclusions

The study identified 5 microRNAs involved in CRC along with 5 target genes prioritized with some set criteria. The molecular docking analysis confirmed that these microRNAs could assist the RNA induced silencing complex (Argonaute protein as the sole) in targeting these genes for regulation. This was confirmed by the predominant hydrophobic interaction within the receptor pocket which made a substantial contribution in stability with microRNA-mRNA duplexes while hydrogen bonding and polar interactions assisted in the proper orientation of the binding interaction. These interactions at the molecular level are important in protein folding and structural stability and also in mediating the binding of the protein to their targets. This result may further serve as a lead to the experimental approach in understanding the molecular mechanism of action of gene regulation in CRC.

Author Contributions: All authors have made significant contributions to the submission of this manuscript. A.O.F. conceived the concept and the design of the manuscript, A.K. and A.P. provided the necessary software required to carry out the analysis. The analysis and data interpretations were done by both A.O.F. and A.P. while A.K. drafted the rough draft and also substantively revised the manuscript. Also, all authors also read and approved the submitted of the final version for publication. Finally, authors agreed to be personally accountable for their personal contributions and ensured that questions related to the accuracy or integrity of any part of the study, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Funding: Not applicable.

Acknowledgments: Authors are thankful to the Center for High-Performance Computing (CHPC), Department of Science and Technology, Republic of South Africa, for the access to their server. Also, to the National Agricultural Proteomics Research and Service Unit (NAPRSU), Plant Omics Laboratory, and the entire Bioinformatics Research Group, Department of Biotechnology, University of the Western Cape, South Africa, for the provision of work-station. During this study. Many thanks to Taiwo Akinsoji (MBBS, M.PH), The University of Illinois at Springfield for provision of materials related to the study.

Conflicts of Interest: The authors declare no conflicts of interest.

Data Availability : The datasets and the clinical data were obtained from specific databases as described in the methodology.

Abbreviations

AGO	Argonaute (Receptor)
miRNA	microRNA
mRNA	Target genes
AA	Amino acid
H-bond	Hydrogen bond
PDB	protein data bank
RISC	RNA induced silencing complex
NA	Nucleic acid
BE	Binding energy
MFE	Minimum folding energy

References

1. Torre, L.A.; Bray, F.; Siegel, R.L.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A. Global cancer statistics, 2012. *CA: A Cancer J. Clin.* **2015**, *65*, 87–108. [CrossRef] [PubMed]
2. Birkenkamp-Demtroder, K.; Christensen, L.L.; Olesen, S.H.; Frederiksen, C.M.; Laiho, P.; Aaltonen, L.A.; Laurberg, S.; Sørensen, F.B.; Hagemann, R.; Ørntoft, T.F. Gene expression in colorectal cancer. *Cancer Res.* **2002**, *62*, 4352–4363. [PubMed]
3. O'Keefe, S. Diet, microorganisms and their metabolites and colon cancer. *Nature Rev. Gastroenterol. Hepatol.* **2016**, *13*, 691–706. [CrossRef] [PubMed]
4. Kheirelseid, E.A.; Miller, N.; Chang, K.H.; Nugent, M.; Kerin, M.J. Clinical applications of gene expression in colorectal cancer. *J. Gastrointest. Oncol.* **2013**, *4*, 144. [PubMed]

5. Fadaka, A.O.; Pretorius, A.; Klein, A. Biomarkers for Stratification in Colorectal Cancer: MicroRNAs. *Cancer Control.* **2019**, *26*, 1073274819862784. [CrossRef] [PubMed]
6. Tuszynska, I.; Magnus, M.; Jonak, K.; Dawson, W.; Bujnicki, J.M. NPDock: A web server for protein–nucleic acid docking. *Nucleic Acids Res.* **2015**, *43*, W425–W430. [CrossRef]
7. Hudson, W.H.; Ortlund, E.A. The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 749–760. [CrossRef] [PubMed]
8. Cooper, T.A.; Wan, L.; Dreyfuss, G. RNA and Disease. *Cell* **2009**, *136*, 777–793. [CrossRef]
9. Höck, J.; Meister, G. The Argonaute protein family. *Genome Biol.* **2008**, *9*, 210. [CrossRef]
10. Schirle, N.T.; Kinberger, G.A.; Murray, H.F.; Lima, W.F.; Prakash, T.P.; MacRae, I.J. Structural analysis of human Argonaute-2 bound to a modified siRNA guide. *J. Am. Chem. Soc.* **2016**, *138*, 8694–8697. [CrossRef]
11. Elkayam, E.; Kuhn, C.-D.; Tocilj, A.; Haase, A.D.; Greene, E.M.; Hannon, G.J.; Joshua-Tor, L. The structure of human argonaute-2 in complex with miR-20a. *Cell* **2012**, *150*, 100–110. [CrossRef] [PubMed]
12. Das, J.; Podder, S.; Ghosh, T.C. Insights into the miRNA regulations in human disease genes. *BMC Genom.* **2014**, *15*, 1010. [CrossRef] [PubMed]
13. Silva Rodrigues, D.V.; Silva Monteiro, V.V.; Navegantes-Lima, K.C.; de Brito Oliveira, A.L.; de França Gaspar, S.L.; Gonçalves Quadros, L.B.; Monteiro, M.C. MicroRNAs in cell cycle progression and proliferation: Molecular mechanisms and pathways. *Non-Coding RNA Investig.* **2018**. [CrossRef]
14. Chivukula, R.R.; Shi, G.; Acharya, A.; Mills, E.W.; Zeitels, L.R.; Anandam, J.L.; Abdelnaby, A.A.; Balch, G.C.; Mansour, J.C.; Yopp, A.C. An essential mesenchymal function for miR-143/145 in intestinal epithelial regeneration. *Cell* **2014**, *157*, 1104–1116. [CrossRef] [PubMed]
15. Xin, M.; Olson, E.N.; Bassel-Duby, R. Mending broken hearts: Cardiac development as a basis for adult heart regeneration and repair. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 529. [CrossRef] [PubMed]
16. Hasuwa, H.; Ueda, J.; Ikawa, M.; Okabe, M. miR-200b and miR-429 function in mouse ovulation and are essential for female fertility. *Science* **2013**, *341*, 71–73. [CrossRef]
17. Giza, D.E.; Vasilescu, C.; Calin, G.A. Key principles of miRNA involvement in human diseases. *Discoveries* **2014**, *2*, e34. [CrossRef]
18. Moridikia, A.; Mirzaei, H.; Sahebkar, A.; Salimian, J. MicroRNAs: Potential candidates for diagnosis and treatment of colorectal cancer. *J. Cell. Physiol.* **2018**, *233*, 901–913. [CrossRef]
19. Nishida, N.; Nagahara, M.; Sato, T.; Mimori, K.; Sudo, T.; Tanaka, F.; Shibata, K.; Ishii, H.; Sugihara, K.; Doki, Y. Microarray analysis of colorectal cancer stromal tissue reveals upregulation of two oncogenic miRNA clusters. *Clin. Cancer Res.* **2012**, *18*, 3054–3070. [CrossRef]
20. Hirata, H.; Ueno, K.; Shahryari, V.; Tanaka, Y.; Tabatabai, Z.L.; Hinoda, Y.; Dahiya, R. Oncogenic miRNA-182–5p targets Smad4 and RECK in human bladder cancer. *PLoS ONE* **2012**, *7*, e51056. [CrossRef]
21. Fadaka, A.O.; Ojo, B.A.; Adewale, O.B.; Esho, T.; Pretorius, A. Effect of dietary components on miRNA and colorectal carcinogenesis. *Cancer Cell Int.* **2018**, *18*, 130. [CrossRef] [PubMed]
22. Zhang, B.; Pan, X.; Cobb, G.P.; Anderson, T.A. microRNAs as oncogenes and tumor suppressors. *Dev. Biol.* **2007**, *302*, 1–12. [CrossRef] [PubMed]
23. Li, L.; Yu, C.; Gao, H.; Li, Y. Argonaute proteins: Potential biomarkers for human colon cancer. *BMC Cancer* **2010**, *10*, 38. [CrossRef] [PubMed]
24. Jafarnejad, S.M.; Ardekani, G.S.; Ghaffari, M.; Martinka, M.; Li, G. Sox4-mediated Dicer expression is critical for suppression of melanoma cell invasion. *Oncogene* **2013**, *32*, 2131. [CrossRef] [PubMed]
25. Völler, D.; Meister, G.; Bosserhoff, A. Strong reduction of Ago2 expression in melanoma and cellular consequences. *Exp. Dermatol.* **2014**, *23*. [CrossRef] [PubMed]
26. Carmell, M.A.; Xuan, Z.; Zhang, M.Q.; Hannon, G.J. The Argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **2002**, *16*, 2733–2742. [CrossRef]
27. Sun, R.; Gao, C.-L.; Li, D.-H.; Li, B.-J.; Ding, Y.-H. Expression status of PIWIL1 as a prognostic marker of colorectal cancer. *Dis. Markers* **2017**, *2017*. [CrossRef] [PubMed]
28. Völler, D.; Linck, L.; Bruckmann, A.; Hauptmann, J.; Deutzmann, R.; Meister, G.; Bosserhoff, A.K. Argonaute family protein expression in normal tissue and cancer entities. *PLoS ONE* **2016**, *11*, e0161165. [CrossRef]
29. Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.* **2002**, *1*, 882. [CrossRef]
30. Demchenko, A.P. Recognition between flexible protein molecules: Induced and assisted folding. *J. Mol. Recognit.* **2001**, *14*, 42–61. [CrossRef]

31. Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into protein–ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* **2016**, *17*, 144. [CrossRef] [PubMed]
32. Zhang, H.; Ye, Q.; Du, Z.; Huang, M.; Zhang, M.; Tan, H. MiR-148b-3p inhibits renal carcinoma cell growth and pro-angiogenic phenotype of endothelial cell potentially by modulating FGF2. *Biomed. Pharmacother.* **2018**, *107*, 359–367. [CrossRef] [PubMed]
33. Huang, X.; Wang, L.; Liu, W.; Li, F. MicroRNA-497-5p inhibits proliferation and invasion of non-small cell lung cancer by regulating FGF2. *Oncol. Lett.* **2019**, *17*, 3425–3431. [CrossRef] [PubMed]
34. Lou, X.; Zhang, J.; Liu, S.; Lou, X.; Liao, D.J. The other side of the coin: The tumor-suppressive aspect of oncogenes and the oncogenic aspect of tumor-suppressive genes, such as those along the CCND–CDK4/6–RB axis. *Cell Cycle* **2014**, *13*, 1677–1693. [CrossRef] [PubMed]
35. Zhang, Y.; Hunter, T. Roles of Chk1 in cell biology and cancer therapy. *Int. J. Cancer* **2014**, *134*, 1013–1023. [CrossRef]
36. Gali-Muhtasib, H.; Ocker, M.; Kuester, D.; Krueger, S.; El-Hajj, Z.; Diestel, A.; Evert, M.; El-Najjar, N.; Peters, B.; Jurjus, A. Thymoquinone reduces mouse colon tumor cell invasion and inhibits tumor growth in murine colon cancer models. *J. Cell. Mol. Med.* **2008**, *12*, 330–342. [CrossRef]
37. Wallin, Å.; Francis, P.; Nilbert, M.; Svanvik, J.; Sun, X.-F. Gene expression profile of colon cancer cell lines treated with SN-38. *Chemotherapy* **2010**, *56*, 17–25. [CrossRef]
38. Zhao, Y.; Yu, H.; Hu, W. The regulation of MDM2 oncogene and its impact on human cancers. *Acta Biochim Biophys Sin. (Shanghai)* **2014**, *46*, 180–189. [CrossRef]
39. Wang, B.; Shen, Z.-L.; Gao, Z.-D.; Zhao, G.; Wang, C.-Y.; Yang, Y.; Zhang, J.-Z.; Yan, Y.-C.; Shen, C.; Jiang, K.-W. MiR-194, commonly repressed in colorectal cancer, suppresses tumor growth by regulating the MAP4K4/c-Jun/MDM2 signaling pathway. *Cell Cycle* **2015**, *14*, 1046–1058. [CrossRef]
40. Zhang, C.; Liu, J.; Wang, X.; Wu, R.; Lin, M.; Laddha, S.V.; Yang, Q.; Chan, C.S.; Feng, Z. MicroRNA-339-5p inhibits colorectal tumorigenesis through regulation of the MDM2/p53 signaling. *Oncotarget* **2014**, *5*, 9106–9117. [CrossRef]
41. Kaindl, U.; Morak, M.; Portsmouth, C.; Mecklenbräuker, A.; Kauer, M.; Zeginigg, M.; Attarbaschi, A.; Haas, O.; Panzer-Grümayer, R. Blocking ETV6/RUNX1-induced MDM2 overexpression by Nutlin-3 reactivates p53 signaling in childhood leukemia. *Leukemia* **2014**, *28*, 600. [CrossRef] [PubMed]
42. Sporn, J.C.; Hothorn, T.; Jung, B. BARD1 expression predicts outcome in colon cancer. *Clin. Cancer Res.* **2011**, *17*, 5451–5462. [CrossRef] [PubMed]
43. Zhang, Y.Q.; Pilyugin, M.; Kuester, D.; Leoni, V.P.; Li, L.; Casula, G.; Zorcolo, L.; Schneider-Stock, R.; Atzori, L.; Irminger-Finger, I. Expression of oncogenic BARD1 isoforms affects colon cancer progression and correlates with clinical outcome. *Br. J. Cancer* **2012**, *107*, 675–683. [CrossRef] [PubMed]
44. Irminger-Finger, I.; Pilyugin, M. Regulation of BARD1 expression by non-coding RNA. Google Patents US10273475B2, 2019.
45. Bach, D.-H.; Park, H.J.; Lee, S.K. The Dual Role of Bone Morphogenetic Proteins in Cancer. *Mol. Oncolytics* **2017**, *8*, 1–13. [CrossRef]
46. Yu, W.; Chai, H. Inhibition of BAMBI reduces the viability and motility of colon cancer via activating TGF-β/Smad pathway in vitro and in vivo. *Oncol. Lett.* **2017**, *14*, 4793–4799. [CrossRef] [PubMed]
47. Kim, Y.; Gupta Vallur, P.; Phaëton, R.; Mythreye, K.; Hempel, N. Insights into the Dichotomous Regulation of SOD2 in Cancer. *Antioxidants* **2017**, *6*, 86. [CrossRef]
48. Zhang, Z.; Lang, J.; Cao, Z.; Li, R.; Wang, X.; Wang, W. Radiation-induced SOD2 overexpression sensitizes colorectal cancer to radiation while protecting normal tissue. *Oncotarget* **2017**, *8*, 7791. [CrossRef] [PubMed]
49. Trimmer, C.; Sotgia, F.; Whitaker-Menezes, D.; Balliet, R.M.; Eaton, G.; Martinez-Outschoorn, U.E.; Pavlides, S.; Howell, A.; Iozzo, R.V.; Pestell, R.G.; et al. Caveolin-1 and mitochondrial SOD2 (MnSOD) function as tumor suppressors in the stromal microenvironment: A new genetically tractable model for human cancer associated fibroblasts. *Cancer Biol. Ther.* **2011**, *11*, 383–394. [CrossRef]
50. Feng, H.; Gu, Z.-Y.; Li, Q.; Liu, Q.-H.; Yang, X.-Y.; Zhang, J.-J. Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. *J. Ovarian Res.* **2019**, *12*, 35–35. [CrossRef]
51. Wenzel, E.S.; Singh, A.T. Cell-cycle checkpoints and aneuploidy on the path to cancer. *In Vivo* **2018**, *32*, 1–5.
52. Arthurs, C.; Murtaza, B.N.; Thomson, C.; Dickens, K.; Henrique, R.; Patel, H.R.; Beltran, M.; Millar, M.; Thrasivoulou, C.; Ahmed, A. Expression of ribosomal proteins in normal and cancerous human prostate tissue. *PLoS ONE* **2017**, *12*, e0186047. [CrossRef] [PubMed]

53. Fan, H.; Lu, Y.; Qin, H.; Zhou, Y.; Gu, Y.; Zhou, J.; Wang, X.; Fan, D. High Ran level is correlated with poor prognosis in patients with colorectal cancer. *Int. J. Clin. Oncol.* **2013**, *18*, 856–863. [CrossRef] [PubMed]
54. Flossbach, L.; Holzmann, K.; Mattfeldt, T.; Buck, M.; Lanz, K.; Held, M.; Möller, P.; Barth, T.F. High-resolution genomic profiling reveals clonal evolution and competition in gastrointestinal marginal zone B-cell lymphoma and its large cell variant. *Int. J. Cancer* **2013**, *132*, E116–E127. [CrossRef] [PubMed]
55. Li, Z.; Han, C.; Feng, J. Relationship of the expression levels of XIAP and p53 genes in hepatocellular carcinoma and the prognosis of patients. *Oncol. Lett.* **2017**, *14*, 4037–4042. [CrossRef] [PubMed]
56. Rao, R.C.; Dou, Y. Hijacked in cancer: The KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer* **2015**, *15*, 334–346. [CrossRef] [PubMed]
57. Fang, Y.; Zhang, D.; Hu, T.; Zhao, H.; Zhao, X.; Lou, Z.; He, Y.; Qin, W.; Xia, J.; Zhang, X. KMT2A histone methyltransferase contributes to colorectal cancer development by promoting cathepsin Z transcriptional activation. *Cancer Med.* **2019**. [CrossRef] [PubMed]
58. Badr, E.A.E.; Ali Assar, M.F.; Gohar, S.F.; Badr, M.H.; Hathout, R.M.; El-Kousy, S.M. The clinical impact of miRNA34a and P53 gene expression in colon cancer. *Biochem. Biophys. Rep.* **2018**, *16*, 88–95. [CrossRef]
59. Okuda, H.; Toyota, M.; Ishida, W.; Furihata, M.; Tsuchiya, M.; Kamada, M.; Tokino, T.; Shuin, T. Epigenetic inactivation of the candidate tumor suppressor gene HOXB13 in human renal cell carcinoma. *Oncogene* **2006**, *25*, 1733–1742. [CrossRef]
60. Maia, S.; Cardoso, M.; Pinto, P.; Pinheiro, M.; Santos, C.; Peixoto, A.; Bento, M.J.; Oliveira, J.; Henrique, R.; Jerónimo, C.; et al. Identification of Two Novel HOXB13 Germline Mutations in Portuguese Prostate Cancer Patients. *PLoS ONE* **2015**, *10*, e0132728. [CrossRef]
61. Xue, X.; Taylor, M.; Anderson, E.; Hao, C.; Qu, A.; Greenson, J.K.; Zimmermann, E.M.; Gonzalez, F.J.; Shah, Y.M. Hypoxia-inducible factor-2α activation promotes colorectal cancer progression by dysregulating iron homeostasis. *Cancer Res.* **2012**, *72*, 2285–2293. [CrossRef]
62. Heo, S.J.; Lee, C.-K.; Hahn, K.Y.; Kim, G.; Hur, H.; Choi, S.H.; Han, K.S.; Cho, A.; Jung, M. A Case of von Hippel-Lindau Disease with Colorectal Adenocarcinoma, Renal Cell Carcinoma and Hemangioblastomas. *Cancer Res. Treat.* **2016**, *48*, 409–414. [CrossRef] [PubMed]
63. Popenda, M.; Szachniuk, M.; Antczak, M.; Purzycka, K.J.; Lukasiak, P.; Bartol, N.; Blazewicz, J.; Adamiak, R.W. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **2012**, *40*, e112–e112. [CrossRef] [PubMed]
64. Chhabra, G.; Sharma, P.; Anant, A.; Deshmukh, S.; Kaushik, H.; Gopal, K.; Srivastava, N.; Sharma, N.; Garg, L.C. Identification and modeling of a drug target for Clostridium perfringens SM101. *Bioinformation* **2010**, *4*, 278. [CrossRef] [PubMed]
65. Laskowski, R.A.; Rullmann, J.A.C.; MacArthur, M.W.; Kaptein, R.; Thornton, J.M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **1996**, *8*, 477–486. [CrossRef] [PubMed]
66. Agrawal, P.; Singh, H.; Srivastava, H.K.; Singh, S.; Kishore, G.; Raghava, G.P. Benchmarking of different molecular docking methods for protein-peptide docking. *BMC Bioinform.* **2019**, *19*, 426. [CrossRef] [PubMed]
67. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367. [CrossRef] [PubMed]
68. Wade, R.C.; Goodford, P.J. The role of hydrogen-bonds in drug binding. *Prog. Clin. Biol. Res.* **1989**, *289*, 433–444.
69. Schiebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T.E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nat. Commun.* **2018**, *9*. [CrossRef] [PubMed]
70. Tüfekci, K.U.; Öner, M.G.; Meuwissen, R.L.J.; Genç, Ş. The role of microRNAs in human diseases. In *miRNomics: MicroRNA Biology and Computational Analysis*; Springer: Berlin, Germany, 2014; pp. 33–50.
71. Ha, T.-Y. MicroRNAs in human diseases: From cancer to cardiovascular disease. *Immune Netw.* **2011**, *11*, 135–154. [CrossRef]
72. Ardekani, A.M.; Naeini, M.M. The Role of MicroRNAs in Human Diseases. *Avicenna J. Med Biotechnol.* **2010**, *2*, 161–179.
73. Zhou, S.-S.; Jin, J.-P.; Wang, J.-Q.; Zhang, Z.-G.; Freedman, J.H.; Zheng, Y.; Cai, L. miRNAS in cardiovascular diseases: Potential biomarkers, therapeutic targets and challenges. *Acta Pharmacol. Sin.* **2018**, *39*, 1073. [CrossRef] [PubMed]

74. Silva, D.C.P.D.; Carneiro, F.D.; Almeida, K.C.D.; Fernandes-Santos, C. Role of miRNAs on the Pathophysiology of Cardiovascular Diseases. *Arq. Bras. De Cardiol.* **2018**, *111*, 738–746. [CrossRef] [PubMed]
75. Gui, Y.; Liu, H.; Zhang, L.; Lv, W.; Hu, X. Altered microRNA profiles in cerebrospinal fluid exosome in Parkinson disease and Alzheimer disease. *Oncotarget* **2015**, *6*, 37043. [CrossRef] [PubMed]
76. Shah, P.; Cho, S.K.; Thulstrup, P.W.; Bjerrum, M.J.; Lee, P.H.; Kang, J.-H.; Bhang, Y.-J.; Yang, S.W. MicroRNA biomarkers in neurodegenerative diseases and emerging nanosensors technology. *J. Mov. Disord.* **2017**, *10*, 18. [CrossRef]
77. Rajgor, D. Macro roles for microRNAs in neurodegenerative diseases. *Non-Coding RNA Res.* **2018**, *3*, 154–159. [CrossRef] [PubMed]
78. James, A.B.; Fadaka, A.O.; Magbagbeola, O.A.; Oturu, A.; KOLAWOLE, O.O.; Ogunjimi, A.; Oshodi, T.; Habeebu, M.; Onawoga, F.O.; Ajogbeje, E.O. Haematological and miRNAs (let-7g, miR-21, miR-141) expression modulation profile in serum samples of Human prostate cancer. *FASEB J.* **2017**, *31*, 712–757.
79. Ryan, B.M.; Robles, A.I.; Harris, C.C. Genetic variation in microRNA networks: The implications for cancer research. *Nat. Rev. Cancer* **2010**, *10*, 389. [CrossRef]
80. Luo, X.; Burwinkel, B.; Tao, S.; Brenner, H. MicroRNA signatures: Novel biomarker for colorectal cancer? *Cancer Epidemiol. Prev. Biomark.* **2011**, *20*, 1272–1286. [CrossRef]
81. Madhavan, D.; Cuk, K.; Burwinkel, B.; Yang, R. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. *Front. Genet.* **2013**, *4*, 116. [CrossRef]
82. Li, C.; Feng, Y.; Coukos, G.; Zhang, L. Therapeutic microRNA strategies in human cancer. *AAPS J.* **2009**, *11*, 747. [CrossRef]
83. O'Brien, J.; Hayder, H.; Zayed, Y.; Peng, C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol.* **2018**, *9*, 402. [CrossRef] [PubMed]
84. Liu, H.-S.; Xiao, H.-S. MicroRNAs as potential biomarkers for gastric cancer. *World J. Gastroenterol.* **2014**, *20*, 12007–12017. [CrossRef] [PubMed]
85. Gebert, L.F.R.; MacRae, I.J. Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 21–37. [CrossRef] [PubMed]
86. Lu, Y.; Leslie, C.S. Learning to predict miRNA-mRNA interactions from AGO CLIP sequencing and CLASH data. *PLoS Comput. Biol.* **2016**, *12*, e1005026. [CrossRef] [PubMed]
87. Kanai, M.; Hamada, J.-I.; Takada, M.; Asano, T.; Murakawa, K.; Takahashi, Y.; Murai, T.; Tada, M.; Miyamoto, M.; Kondo, S. Aberrant expressions of HOX genes in colorectal and hepatocellular carcinomas. *Oncol. Rep.* **2010**, *23*, 843–851.
88. Zacchetti, G.; Duboule, D.; Zakany, J. Hox gene function in vertebrate gut morphogenesis: The case of the caecum. *Development* **2007**, *134*, 3967–3973. [CrossRef] [PubMed]
89. Tatangelo, F.; Di Mauro, A.; Scognamiglio, G.; Aquino, G.; Lettiero, A.; Delrio, P.; Avallone, A.; Cantile, M.; Botti, G. Posterior HOX genes and HOTAIR expression in the proximal and distal colon cancer pathogenesis. *J. Transl. Med.* **2018**, *16*, 350. [CrossRef]
90. Sekiya, T.; Adachi, S.; Kohu, K.; Yamada, T.; Higuchi, O.; Furukawa, Y.; Nakamura, Y.; Nakamura, T.; Tashiro, K.; Kuhara, S. Identification of BMP and activin membrane-bound inhibitor (BAMBI), an inhibitor of transforming growth factor-β signaling, as a target of the β-catenin pathway in colorectal tumor cells. *J. Biol. Chem.* **2004**, *279*, 6840–6846. [CrossRef]
91. Togo, N.; Ohwada, S.; Sakurai, S.; Toya, H.; Sakamoto, I.; Yamada, T.; Nakano, T.; Muroya, K.; Takeyoshi, I.; Nakajima, T. Prognostic significance of BMP and activin membrane-bound inhibitor in colorectal cancer. *World J. Gastroenterol.* **2008**, *14*, 4880. [CrossRef]
92. Tang, J.; Gifford, C.; Samarakoon, R.; Higgins, P. Deregulation of negative controls on TGF-β1 signaling in tumor progression. *Cancers* **2018**, *10*, 159. [CrossRef]
93. Tiwari, A.K.; Joshi, S.; Brasky, J.T.; DeLaCruz, M.; Gibson, T.P.; Kunte, D.; Subramanian, P.; Wali, R.K.; Roy, H.K. Superoxide Dismutase 2 (Sod2) is Overexpressed at an Early Stage During Colorectal Carcinogenesis: A Putative Target for Celecoxib Chemoprevention. *Gastroenterology* **2011**, *140*, 343. [CrossRef]
94. Irminger-Finger, I.; Ratajska, M.; Pilyugin, M. New concepts on BARD1: Regulator of BRCA pathways and beyond. *Int. J. Biochem. Cell Biol.* **2016**, *72*, 1–17. [CrossRef] [PubMed]
95. Li, A.J.; Li, H.-G.; Tang, E.-J.; Wu, W.; Chen, Y.; Jiang, H.-H.; Lin, M.-B.; Yin, L. PIK3CA and TP53 mutations predict overall survival of stage II/III colorectal cancer patients. *World J. Gastroenterol.* **2018**, *24*, 631–640. [CrossRef] [PubMed]

96. Andrews, P.; Craik, D.; Martin, J. Functional group contributions to drug-receptor interactions. *J. Med. Chem.* **1984**, *27*, 1648–1657. [CrossRef] [PubMed]
97. Rath, S.N.; Das, D.; Konkimalla, V.B.; Pradhan, S.K. In silico study of miRNA based gene regulation, involved in solid cancer, by the assistance of argonaute protein. *Genom. Inform.* **2016**, *14*, 112. [CrossRef] [PubMed]
98. Alex, A.; Millan, D.S.; Perez, M.; Wakenhut, F.; Whitlock, G.A. Intramolecular hydrogen bonding to improve membrane permeability and absorption in beyond rule of five chemical space. *MedChemComm* **2011**, *2*, 669–674. [CrossRef]
99. Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular hydrogen bonding in medicinal chemistry. *J. Med. Chem.* **2010**, *53*, 2601–2611. [CrossRef]
100. Sahu, V.; Khan, A.; Singh, R.; Singh, P. Hydrophobic, polar and hydrogen bonding based drug-receptor interaction of tetrahydroimidazobenzodiazepinones. *Am. J. Immunol* **2008**, *4*, 33. [CrossRef]
101. Hsu, S.-D.; Lin, F.-M.; Wu, W.-Y.; Liang, C.; Huang, W.-C.; Chan, W.-L.; Tsai, W.-T.; Chen, G.-Z.; Lee, C.-J.; Chiu, C.-M. miRTarBase: A database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* **2010**, *39*, D163–D169. [CrossRef]
102. Chou, C.-H.; Shrestha, S.; Yang, C.-D.; Chang, N.-W.; Lin, Y.-L.; Liao, K.-W.; Huang, W.-C.; Sun, T.-H.; Tu, S.-J.; Lee, W.-H. miRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **2017**, *46*, D296–D302. [CrossRef]
103. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2008**, *37*, 1–13. [CrossRef] [PubMed]
104. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44. [CrossRef] [PubMed]

CHAPTER 7.0

GENERAL DISCUSSION

Discussion

CRC is the fourth most common cancer and the second leading cause of cancer-related death globally (Amin *et al.*, 2015). More than 1.2 million patients are diagnosed with CRC subtype annually, and more than 600000 people die from the disease (Brenner *et al.*, 2014). Recently, the disease was ranked the fourth most common and sixth in mortality of all deaths with approximately one-fourth of patients presenting synchronous metastatic diseases at the time of primary diagnosis in South Africa (Brand *et al.*, 2018). Despite the fact that chemotherapy is utilized in most stages of this disease, surgical resection of the primary tumor and metastases remains the best treatment option to accomplish prolong survival. To date, no information on microRNA diagnosis as a non-invasive CRC management have been published in South Africa. The long-term survival of patients with CRC remains poor regardless of the recent therapeutic advances in treatment. Therefore, there is an urgent and growing requirement for better understanding of the molecular pathogenesis, for identifying novel biomarkers for diagnosis and prognosis, and for exploring new and more effective treatment strategies for this disease.

Accumulating studies in past decades revealed that microRNAs have been implicated in numerous cellular processes, including differentiation, proliferation, autophagy, and apoptosis. Some oncogenic microRNAs have been demonstrated to be involved in the development of CRC (Okugawa *et al.*, 2014). However, the disease-specific mechanisms of most microRNAs involved in the progress of this cancer sub-type remains unknown (Y. Li *et al.*, 2017). MicroRNAs show potential use as diagnostic, prognostic and therapeutic resolutions for different cancers including CRC. Hence, the aim of this study was to investigate the potentials of candidate microRNA in the early detection and staging of CRC using *in silico* approaches. Bioinformatics data-mining of gene and non-coding RNA most especially microRNAs data is a useful tool for identifying novel significant genes and non-coding RNAs associated with the pathogenesis of diseases, providing valuable insights and a basis for further novel research (Kulasingam *et al.*, 2008; Zhang *et al.*, 2019). The advantages of these approaches are that they are rapid, cost-effective, and less labour intensive. Detailed discussion is as contained within the publications.

Summary of article published

Identification of candidate microRNAs and their target genes involved in CRC

The aim of this chapter (Chapter 3) was to identify candidate microRNAs as well as their target genes which can be used as diagnostic and prognostic biomarkers for CRC using *in silico* approaches. Several pre-clinical and clinical studies have been conducted to uncover the underlying mechanisms of CRC. In this chapter, 5 microRNAs and 82 of their target genes were identified. The candidate microRNAs were identified using two sequence similarity search tool namely BLASTn and CD-HIT-EST-2D combined. 10 microRNAs were identified with a potential novel association with CRC. These microRNAs were refined to six microRNAs through literature mining. One of the six microRNAs was also eliminated due to a lack of a gene target.

The target genes were identified to confirm the implication of these microRNAs in CRC using an *in silico* approach. A list of target genes was identified using three databases namely miRDB, miRDIP, and TargetScan, as well as a list of all genes implicated in CRC, was generated using the gbCRC and CoReCG databases. The intersection between these two lists was determined using Venn diagrams resulting in 796 genes targeted by the five microRNAs. Through functional annotation databases, this number was reduced to 82 unique genes. DAVID alongside STRING were used to conduct KEGG pathway enrichment analysis and protein protein-interaction analysis in conjunction with Cytoscape.

Through integrated bioinformatics analysis (DAVID), 82 genes were significantly enriched in CRC. In the KEGG pathway enrichment analysis, the microRNA target genes were primarily enriched in 104 pathways including 29 genes implicated in cancer pathway, 17 genes in proteoglycans in cancer, 16 genes in FoxO signaling pathway, 12 genes in P53 signaling pathways and 9 genes in CRC pathway with the FDR in the range of 1.25e-29 – 5.05e-8. The 9 hub genes associated with the CRC pathway in KEGG are: transcription factor 7 like 2 (TCF7L2), APC, Wnt signaling pathway regulator (APC), transcription factor 7 (T-cell specific, HMG-box) (TCF7), caspase 3 (CASP3), transforming growth factor beta receptor 2 (TGFBR2), DCC netrin 1 receptor(DCC), baculoviral IAP repeat containing 5 (BIRC5), KRAS proto-oncogene, GTPase (KRAS), and catenin beta 1 (CTNNB1) with a higher degree of connectivity and FDR of 5.05e-08. Furthermore, using the STRING online database and Cytoscape software, 82 microRNA target genes enriched in CRC obtained in DAVID were used as input for Protein-protein interaction (PPI) network complex and also to create a SIF

file for Cytoscape visualization. The PPI network contains 63 nodes and 113 edges, and 19 of the 82 genes were not included in the microRNA target gene network. The "hub" genes of these candidate microRNAs (CTNNB1 and EGFR) therefore, may be a potential target for CRC treatment.

Complex genomic exploration of microRNA target genes in CRC clinical data

In order to obtain more accurate results, molecular experiment with a large number of clinical samples is required to validate the present results and elucidate the underlying mechanisms of how these genes and microRNAs impact colorectal carcinogenesis and progression. In this chapter (Chapter 3), the integrative analysis of complex cancer genomics and clinical profiles were performed on the microRNAs target genes. Like other cancer diseases, CRC is caused about by the accumulation of genetic alterations in the cells that drives malignant development. These modifications are revealed by gene profiling and copy number alteration analysis. Moreover, recent evidence suggests that microRNAs have an important role in CRC development. Despite efforts to profile CRC, the genetic alterations and biological processes that correlate with the development and progression remain partially elusive. The identification of genes with high genetic alteration and co-expressed genes, that are functionally related, were exploited as the bases to prioritize and further identify a core network of genes associated with CRC for better management. Using the identified microRNAs target genes to explore complex genomics in CRC clinical samples using an *in silico* approach, 82 target genes were quarried across numerous clinical data. The result shows that 474 cases (77%) out of the 619 total cases sequenced in DFCI (Cell report, 2016) have at least one of the 82 genes with the frequency of alteration in each of the 82 selected genes. Also, the MSK, (Cancer Cell, 2018) was used to further query these microRNA target genes for integrative analysis. As the basis for prioritization, only genes with 2% and above of the percentage sample altered were considered for further analysis. Of the 82 genes queried, only 17 have a frequency of mutation alteration above 2%. This final list of 17 genes was further exploited for the distribution of mutation across protein domains in other to determine the domains susceptible to mutation and the type of mutation.

Finally, 17 genes were prioritized based on the genetic alteration and the distribution of mutation across protein domains in the cBioPortal database. The study revealed that genes with a protein tyrosine kinase domain are frequently altered in CRC and the most common alteration in these genes/ domain is missense mutations. The survival analysis of these

prioritized genes displayed by Kaplan-Meier plots revealed that patients with alterations in the microRNA prioritized target genes have significantly better overall survival than patients without these alterations. These could be further exploited and could potentially serve as a resource for explicitly selecting targets for diagnosis, drug development, and management of CRC.

Prognostic and expression analysis of the candidate microRNAs and target genes

Prognostic and expression analysis were performed on all the candidate microRNAs and their target genes using the available online databases namely, PROGmiRV2, SurvMicro, SurvExpress, PrognoScan, dbDEMC 2.0 and FIREBROWSE (Figure A.1) (Chapter 4).

The results of PROGmiRV2 and dbDEMC 2.0 databases for microRNAs prognostics and expression analysis respectively confirmed that the candidate microRNAs do not have any previously known link with CRC and as such, they are potentially novel microRNAs.

SurvExpress biomarker database was employed to predict the clinical outcome and prognostic value of CRC metastasis genes. The results indicated that genes ERBB4, CASP8, BRCA1, GNAS, EGFR, INSR, NRAS, KRAS, CTNNB1, TCF7L2, INHBA, TGFBR2, IGF1R, and IRS2 showed a distinct difference between the low and the high-risk groups (Figure A.19). The difference in expression between risks groups for these genes can be useful for prognostic and predictive outcomes for CRC. Therefore, this gene panel can be used to determine a patient's survival if the survival is affected by the expression of these genes. PrognoScan was also employed for the prognostic analysis (Figure A.20). The conclusive result of the PrognoScan analysis showed that all the genes were correlated or showed association to CRC and seven are significant considering their P Cox value. Therefore, these genes either collectively or singly have a good discriminatory value between high and low-risk CRC groups making them potential biomarker for CRC management.

Gene expression levels of these 17 microRNA target genes in different types of cancers and their healthy counterparts were analyzed by FIREBROWSE (Appendix A, Table A.1 and Figures A.2 – A.18). Of these genes, five were upregulated in tumor samples (CRC) and nine were upregulated in normal colorectal tissues and vice versa. The remaining three genes showed no differential expression (Table A.2). BRCA1, INHBA, CASP8, IRS2, and CTNNB1 were upregulated in CRC and downregulated in normal colorectal tissues while APC, KRAS, TCF7L2, PTEN, ERBB4, TGFBR2, NRAS, EGFR, and INSR are down-regulated in CRC and upregulated in normal colorectal tissue samples. IGF1R, ATM, and GNAS showed no

differential expression between tumor and normal colorectal tissues. Considering the pattern of these gene signatures, the candidate microRNAs could serve as a good diagnostic biomarker for CRC.

In silico structural identification and possible interactions between the candidate microRNA and target genes for functional determination

Mfold was used to determine the secondary structures of the candidate microRNAs (Chapter 5). Promoter sequence retrieval was performed on the 7 target genes using the available online databases namely; NCBI, Ensembl, and UCSC (Figure B.1, B.2 and Table B.1). The promoter sequences were further subjected to CpG island analysis using SMS tool, while the triplex binding interaction was determined using the Trident software tool.

Understanding microRNAs, their secondary structures, thermodynamic parameters, and targets may deliver greater promise towards their diagnostic potentials and mechanisms in the management of CRC. Since the secondary structure of microRNAs influences the efficiency of microRNA-mRNA interaction. The secondary structure of the candidate microRNAs was analyzed using software. The secondary structure together with the thermodynamic parameters of the candidate microRNAs may, therefore, provide a valid result regarding their target even when the conservation of the microRNA is unknown. Also, the secondary structure of the candidate microRNAs can have a conformational role ($Stru^{C}$) to modulate target interactions and therefore can be used to explain the different degree of genetic regulation in CRC. MicroRNAs act as post-transcriptional regulators of gene expression (D. P. Bartel, 2009). Genetic variation in microRNA-encoding sequences or their corresponding binding sites may influence the fidelity of the microRNA-messenger RNA interaction and subsequently alter risk of cancer development (Schmit *et al.*, 2015). Accurate identification of binding sites remains a major challenge in computational biology. Identification of such sites would facilitate the development of gene networks to model interactions that would help unravel important biological pathways.

Target genes binding of these microRNAs (Triplex formation) were predicted using online database (TRIDENT), and candidate microRNA-mRNA regulatory complexes were constructed, which were expected to serve vital roles in colorectal carcinogenesis and progression, and to thus serve as novel biomarkers for diagnosis and therapy of CRC. This study exploited the binding of the candidate microRNAs to the promoter region of the target genes. Trident was employed to identify the binding pattern. Promoters of these genes were analyzed to identify common and significant patterns with putative regulatory potential. The

results suggest that triplex formation is a general mechanism of RNA-mediated target-site recognition, and it is said to have a major impact on biological functions as previously reported. MicroRNAs that are capable of triplex formation with duplex DNA are more frequently positively correlated with gene transcripts, miR-1, miR-3 and miR-5 are suggested to have significant enrichment of positive correlation with the target gene involved in the triplex structure. This analysis further confirmed that the targets of the candidate microRNAs are enriched with microRNA triplex binding sites. Furthermore, microRNA function may then depend not on their sequence but also their structures and triplex binding interaction with their targets. This *in silico* analysis, therefore, merit further molecular study to determine the biological function of triplex interactions between mRNA and microRNA and to further classify hoogsteen (direct) and reverse hoogsteen (indirect) interaction based on biological function.

Determination of the biological significance of the microRNA target genes

Gene interaction analysis/ GO term, Somatic Events in Evolution of the target genes, and co-expression analysis were performed on the seven significant genes targeted by the candidate microRNAs obtained in Chapter 2 using the following databases; STRING SEECancer, and GeneMANIA in Cytoscape (Chapter 5).

Protein interaction was performed on all the seven genes targeted by the five candidate microRNAs. All these genes are interconnected with one another. To gain insight into the functions of the candidate microRNAs, network and enrichment analysis were performed on the statistically significant microRNA target genes using STRING database.

GO describes genes from three aspects, namely molecular function (MF), cellular component (CC) and biological process (BP). In the BP group, the microRNA target genes were significantly enriched in 994 processes including response to stress, regulation of cell proliferation, response to endogenous stimuli, regulation of signaling and positive regulation of protein metabolic process as the top five with false discovery rate (FDR) within 7.5e-20 – 9.25e-15. In the CC group, the microRNA target genes were primarily enriched 43 GO-terms including membrane raft, side of membrane, protein complex, cytosol, and receptor complex with FDR in the range of 2.41e-7 – 3.66e-6. In the MF group, the microRNA target genes were primarily enriched in 67 GO-terms with significant enrichment in protein binding, identical protein binding, protein binding complex, enzyme binding, receptor binding with FDR within the range of 6.0e-20 – 1.26e-7. The network consisted of 13 edges connecting 5 microRNAs

and 7 target genes as the nodes. Among these, miR-1 and miR-5 were the most common regulators, which were linked to 6 and 3 target genes, respectively. These candidate microRNAs obtained in this study may share several similar pathways. The evolutionary stage-specific and variant events of the microRNA target genes are presented were analyzed using the SEECancer database. The result suggested that genes together with their regulators can be used for CRC stratification. Finally, GeneMANIA was used to generate hypothesis about the candidate microRNA target genes' significance. Genetic interaction of this network accounted for 43.83% of the total interactions, shared protein domain accounted for 26.28, physical interaction accounted for 14.23, co-expressed genes accounted for 10.14%. Only 3.89 and 2.62% are accounted for pathway and predicted interaction. These associated genes may then provide further a complete microRNA-gene network for CRC diagnosis and disease management.

Molecular interaction between the candidate microRNAs and argonaute protein

The interactions between protein and nucleic acids play essential roles in various cellular and biological processes, including DNA replication, RNA transcription, the translation of polypeptides, RNA splicing, and the degradation of nucleic acids. The errors in receptor-nucleic acid interactions are implicated in a number of diseases, ranging from neurological disorders to cancer. The molecular docking analysis between the candidate microRNAs and five target genes were carry out using molecular docking tools.

Protein selection and preparation were carried out using PDB and Schrödinger suits. The molecular docking analysis was performed using PATCHDOCK webserver and visualized by discovery studio visualizer. The results of the study reveal that the candidate microRNAs have strong binding affinity towards their targets suggesting a crucial factor in the silencing mechanism. Furthermore, the molecular docking of the receptor to both the microRNA and microRNA-mRNA duplex were analyzed computationally to understand their interaction at the molecular level. Conclusively, the study provides an explanation for understanding the microRNAs-based gene regulation (silencing mechanism) in CRC.

CHAPTER 8.0

CONCLUSIONS, LIMITATION, AND FUTURE WORK

Conclusion

The discovery of microRNA over 25 years ago has ushered in a new era in biotechnology. Over 2000 microRNAs have been identified in humans and it is believed that they collectively regulate one-third of the genes in the genome. These non-coding RNAs have been linked to many human diseases including cancer and are being explored as clinical diagnostics and as therapeutic targets. Target identification for microRNAs is also crucial for several reasons. For biologists, identifying microRNA target genes is pivotal to understanding the mechanism of action. For scientists in the area of therapeutics, target validation provides the best biomarker(s) for the determination of the efficacy of a microRNA mimic or inhibitor.

The identification of microRNAs and their targets has followed three general approaches: computational/bioinformatics prediction algorithms, molecular isolation of microRNA/mRNA complexes, and transcriptomic/proteomic analysis. Roughly 50% of mammalian microRNAs have been identified in the transcription units. However, these number varies between species. MicroRNAs function by silencing or translationally repression of their targets when complexed with RNA induced silencing complex (RISC). Coupled with other functions, identification of microRNAs, targets and understanding their biological roles as well as the pattern of expression (tumor and normal cells) deserves closer attention.

This study, therefore, used several computational prediction algorithms to identify five candidate microRNAs specific for CRC, their targets and the biological roles/mechanism of action in the context of tumor histopathology.

The microRNAs obtained in this study could be suggested as both diagnostic, and prognostic biomarkers. The “hub” genes, therefore, may be a potential target for CRC treatment. However, their involvement in CRC was determined from bioinformatics perspective (mainly *in silico* approach). Finally, the molecular docking analysis confirmed that these microRNAs could assist the RNA induced silencing complex (Argonaute protein as the sole) in targeting these genes for regulation. This was confirmed by the predominant hydrophobic interaction within the receptor pocket which made a substantial contribution in stability with microRNA-mRNA duplexes while hydrogen bonding and polar interactions assisted in the proper orientation of the binding interaction. These interactions at the molecular level are important in protein folding and structural stability and also in mediating the binding of the protein to their targets.

This result may further serve as a lead to the experimental approach in understanding the molecular mechanism of action of gene regulation in CRC. In future works, selected microRNAs and their targets will be validated using real-time PCR and Western blotting.

Solving cancer problems with genomic data and various computational methods is a cost-effective approach in cancer research.

Limitations

Most experimental approaches for microRNA biomarker identification rely on previous predictions using *in silico* frameworks. Though experimental confirmation remains the highest form of novel microRNA validation, there are several limitations, for example, some experimental approaches may not be specific as expected. In addition, computational prediction suffers high rates of false positive prediction. Due to the shortcomings of both computational and experimental methods for novel microRNA identification, effective elucidation of novel microRNA and their target genes requires collaboration between these two approaches.

Future work

Molecular validation of the Identified microRNAs and their target genes

Following Duke's classification (Akkoca *et al.,* 2014) three CRC cell-lines will be obtained from America type culture collection (ATCC) (https://www.lgcstandards-atcc.org/About) namely HT-29, DLD-1 and SW-116, one normal (CRL-7418) and two other cell-lines MCF-7 (breast cancer), and GA-10 (blood cell-line). Cell lines will be cultured according to their growth requirements. Both the prioritized microRNAs as well as their targets will be tested for differential expression levels using qPCR across the various cell lines. Furthermore, human serum and tissue samples will be used to confirm the validation.

Design primers for the genes and microRNAs for qPCR

RT-qPCR would be applied for the detection of microRNAs and its precursor and expression analysis of microRNA host and targets genes. Related primers would be designed using NCBI Primer-blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), IDT oligo-analyzer and MWG online PCR primer design tools (Dokanehiifard *et al.,* 2017). Expression analysis will be performed on the LightCycler 480 instrument (Roche).

Construction of a lateral flow device

To combine the microRNA/s and/or gene/s that showed the high sensitivity and specificity for CRC in its early stages combined with nanotechnology to create a low-cost lateral flow device to be used within a clinical setting.

Integrated Lateral Flow Device Design

The lateral flow device will be designed according to the procedure of Williams *et al.* (2016) in conjunction with an industry partner.

REFERENCES

Adesanya, A., and da Rocha-Afodu, J. (2000). Colorectal cancer in Lagos: a critical review of 100 cases. *The Nigerian Postgraduate Medical Journal, 7*(3), 129-136.

Akkoca, A. N., Yanık, S., Özdemir, Z. T., Cihan, F. G., Sayar, S., Cincin, T. G., Çam, A., and Özer, C. (2014). TNM and Modified Dukes staging along with the demographic characteristics of patients with colorectal carcinoma. *International Journal of Clinical and Experimental Medicine, 7*(9), 2828.

Alwan, A. (2011). Global status report on noncommunicable diseases *2010*: *World Health Organization.*

Amin, M., and Lockhart, A. C. (2015). The potential role of immunotherapy to treat colorectal cancer. *Expert Opinion on Investigational Drugs, 24*(3), 329-344.

Angelo, N., and Dreyer, L. (2001). Colorectal carcinoma-a new threat to black patients? A retrospective analysis of colorectal carcinoma received by the Institute for Pathology, University of Pretoria. *South African Medical Journal, 91*(8), 689-693.

Ardekani, A. M., and Naeini, M. M. (2010). The role of microRNAs in human diseases. *Avicenna Journal of Medical Biotechnology, 2*(4), 161.

Baron, T. H., Smyrk, T. C., and Rex, D. K. (2013). Recommended intervals between screening and surveillance colonoscopies. Paper presented at the Mayo Clinic Proceedings.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell, 116*(2), 281-297.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell, 136*(2), 215-233.

Bartel, J.-R. M. B. D. (1953). B. 2006 MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology, 57*, 19-53.

Berezikov, E., and Plasterk, R. H. (2005). Camels and zebrafish, viruses and cancer: a microRNA update. *Human Molecular Genetics, 14*(suppl_2), R183-R190.

Bilen, J., Liu, N., Burnett, B. G., Pittman, R. N., and Bonini, N. M. (2006). MicroRNA pathways modulate polyglutamine-induced neurodegeneration. *Molecular Cell, 24*(1), 157-163.

Botteri, E., Iodice, S., Bagnardi, V., Raimondi, S., Lowenfels, A. B., and Maisonneuve, P. (2008). Smoking and colorectal cancer: a meta-analysis. *The Journal of the American Medical Association, 300*(23), 2765-2778.

Brand, M., Gaylard, P., and Ramos, J. (2018). Colorectal cancer in South Africa: An assessment of disease presentation, treatment pathways and 5-year survival. *South African Medical Journal, 108*(2), 118-122.

Brenner, H., Kloor, M., and Pox, C. P. (2014). Colorectal cancer. *Lancet, 383*, 1490 –1502.

Cannell, I. G., Kong, Y. W., and Bushell, M. (2008). How do microRNAs regulate gene expression? In: Portland Press Limited.

Catalanotto, C., Cogoni, C., and Zardo, G. (2016). MicroRNA in control of gene expression: an overview of nuclear functions. *International Journal of Molecular Science, 17*(10), 1712.

Croce, C. M., and Calin, G. A. (2005). miRNAs, cancer, and stem cell division. *Cell, 122*(1), 6-7.

Cronjé, L., Becker, P., Paterson, A., and Ramsay, M. (2009). Hereditary non-polyposis colorectal cancer is predicted to contribute towards colorectal cancer in young South African blacks. *South African Journal of Science, 105*(1-2), 68-72.

Cui, H., Zhang, Y., and Ren, F. (2014). dbDEMC2. 0: A database of differentially expressed miRNAs in human cancers v2. 0. *China Journal of Modern Medicine, 24*(3), 77-79.

Diederichs, S., Bartsch, L., Berkmann, J. C., Fröse, K., Heitmann, J., Hoppe, C., Iggena, D., Jazmati, D., Karschnia, P., and Linsenmeier, M. (2016). The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. *EMBO Molecular Medicine*, e201506055.

Dokanehiifard, S., Yasari, A., Najafi, H., Jafarzadeh, M., Nikkhah, M., Mowla, S. J., and Soltani, B. M. (2017). A novel microRNA located in the TrkC gene regulates the Wnt signaling pathway and is differentially expressed in colorectal cancer specimens. *Journal of Biological Chemistry, 292*(18), 7566-7577.

Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annual Review of Biochemistry, 79*, 351-379.

Fasanaro, P., Greco, S., Ivan, M., Capogrossi, M. C., and Martelli, F. (2010). microRNA: emerging therapeutic targets in acute ischemic diseases. *Pharmacology and Therapeutics, 125*(1), 92-104.

Fearon, E. R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell, 61*(5), 759-767.

Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer, 127*(12), 2893-2917.

Fredericks, E., Dealtry, G., and Roux, S. (2015). Molecular aspects of colorectal carcinogenesis: a review. *Journal of Cancer Biology and Research, 3*(1), 1057.

Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Enright, A. J., Thomson, J. M., Baskerville, S., Hammond, S. M., Bartel, D. P., and Schier, A. F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. *Science, 308*(5723), 833-838.

Grady, W. M. (2003). Genetic testing for high-risk colon cancer patients. *Gastroenterology, 124*(6), 1574-1594.

Graham, A., Davies Adeloye, L. G., Theodoratou, E., and Campbell, H. (2012). Estimating the incidence of colorectal cancer in Sub–Saharan Africa: A systematic analysis. *Journal of Global Health, 2*(2).

Griffiths-Jones, S. (2004). The microRNA registry. *Nucleic Acids Research, 32*(suppl_1), D109-D111.

Harfe, B. D., McManus, M. T., Mansfield, J. H., Hornstein, E., and Tabin, C. J. (2005). The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proceedings of the National Academy of Sciences, 102*(31), 10898-10903.

Hastings, M. L., and Krainer, A. R. (2001). Pre-mRNA splicing in the new millennium. *Current Opinion in Cell Biology, 13*(3), 302-309.

Hatfield, S., Shcherbata, H., Fischer, K., Nakahara, K., Carthew, R. W., and Ruohola-Baker, H. (2005). Stem cell division is regulated by the microRNA pathway. *Nature, 435*(7044), 974.

Irabor, D. O., Afuwape, O. O., and Ayandipo, O. O. (2014). The present status of the management of colon and rectal cancer in Nigeria. *Journal of Cancer Research, 2014*.

Jones, G. S. (2004). The microRNA registry. *Nucleic Acids Res, 32*(Database issue), D109-D111.

Kawamura, M., Toiyama, Y., Tanaka, K., Inoue, Y., and Kusunoki, M. (2014). Can circulating microRNAs become the test of choice for colorectal cancer? *urr. Colorectal Cancer Reports, 10*, 403-410.

Klein, M. E., Impey, S., and Goodman, R. H. (2005). Role reversal: the regulation of neuronal gene expression by microRNAs. *Current Opinion in Neurobiology, 15*(5), 507-513.

Kulasingam, V., and Diamandis, E. P. (2008). Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature Reviews Clinical Oncology, 5*(10), 588.

Lally, N., Nord, C. L., Walsh, V., and Roiser, J. P. (2013). Does excitatory fronto-extracerebral tDCS lead to improved working memory performance? *F1000Research, 2*.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell, 75*(5), 843-854.

Li, F.-y., and Lai, M.-d. (2009). Colorectal cancer, one entity or three. *Journal of Zhejiang University-Science B, 10*(3), 219-229.

Li, Y., Chen, M., Liu, J., Li, L., Yang, X., Zhao, J., Wu, M., and Ye, M. (2017). Upregulation of microRNA-18b contributes to the development of colorectal cancer through inhibiting CDKN2B. *Molecular and Cellular Biology*, 00391-00317.

Lu, J., and Clark, A. G. (2012). Impact of microRNA regulation on variation in human gene expression. *Genome Research, 22*(7), 1243-1254.

Mattick, J. S., and Makunin, I. V. (2005). Small regulatory RNAs in mammals. *Human Molecular Genetics, 14*(suppl_1), R121-R132.

Migliore, L., Migheli, F., Spisni, R., and Coppedè, F. (2011). Genetics, cytogenetics, and epigenetics of colorectal cancer. *Biomedical Research International, 2011*.

Morey, C., and Avner, P. (2004). Employment opportunities for non-coding RNAs. *FEBS letters, 567*(1), 27-34.

Mqoqi, N., Sitas, F., and Jula, M. (2004). Incidence of histologically diagnosed cancer in South Africa, 1998-1999: National Cancer Registry of South Africa. *National Health Laboratory Service*.

Muto, T., Bussey, H., and Morson, B. (1975). The evolution of cancer of the colon and rectum. *Cancer, 36*(6), 2251-2270.

Naguibneva, I., Ameyar-Zazoua, M., Polesskaya, A., Ait-Si-Ali, S., Groisman, R., Souidi, M., Cuvellier, S., and Harel-Bellan, A. (2006). The microRNA miR-181 targets the homeobox protein Hox-A11 during mammalian myoblast differentiation. *Nature cell biology, 8*(3), 278.

Okugawa, Y., Toiyama, Y., and Goel, A. (2014). An update on microRNAs as colorectal cancer biomarkers: where are we and what's next? *Expert Review of Molecular Diagnostics, 14*(8), 999-1021.

Osaki, M., Okada, F., and Ochiya, T. (2015). miRNA therapy targeting cancer stem cells: a new paradigm for cancer treatment and prevention of tumor recurrence. *Therapeutic Delivery, 6*(3), 323-337.

Pace, H. C., Hodawadekar, S. C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., Croce, C. M., and Brenner, C. (2000). Crystal structure of the worm NitFhit Rosetta

Stone protein reveals a Nit tetramer binding two Fhit dimers. *Current Biology, 10*(15), 907-917.

Poy, M. N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., MacDonald, P. E., Pfeffer, S., Tuschl, T., Rajewsky, N., and Rorsman, P. (2004). A pancreatic islet-specific microRNA regulates insulin secretion. *Nature, 432*(7014), 226.

Rao, C. V., and Yamada, H. Y. (2013). Genomic instability and colon carcinogenesis: from the perspective of genes. *Frontiers in Oncology, 3*.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature, 403*(6772), 901.

Saluja, S., Alatise, O. I., Adewale, A., Misholy, J., Chou, J., Gonen, M., Weiser, M., and Kingham, T. P. (2014). A comparison of colorectal cancer in Nigerian and North American patients: is the cancer biology different? *Surgery, 156*(2), 305-310.

Schmit, S. L., Gollub, J., Shapero, M. H., Huang, S.-C., Rennert, H. S., Finn, A., Rennert, G., and Gruber, S. B. (2015). MicroRNA polymorphisms and risk of colorectal cancer. *Cancer Epidemiology and Prevention Biomarkers, 24*(1), 65-72.

Seton-Rogers, S. (2013). Tumour microenvironment: means of resistance. *Nature Reviews Cancer, 13*(9), 607-607.

Tanase, C., Ogrezeanu, I., and Badiu, C. (2011). *Molecular pathology of pituitary adenomas*: Elsevier.

Thakur, V. S., Deb, G., Babcook, M. A., and Gupta, S. (2014). Plant phytochemicals as epigenetic modulators: role in cancer chemoprevention. *The American Association of Pharmaceutical Science, 16*(1), 151-163.

Tomaru, Y., and Hayashizaki, Y. (2006). Cancer research with non-coding RNA. *Cancer Science, 97*(12), 1285-1290.

Trang, P., Weidhaas, J., and Slack, F. (2008). MicroRNAs as potential cancer therapeutics. *Oncogene, 27*(2), S52-S57.

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Smits, A. M., and Bos, J. L. (1988). Genetic alterations during colorectal-tumor development. *New England Journal of Medicine, 319*(9), 525-532.

Wienholds, E., Kloosterman, W. P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, H. R., Kauppinen, S., and Plasterk, R. H. (2005). MicroRNA expression in zebrafish embryonic development. *Science, 309*(5732), 310-311.

Wilfred, B. R., Wang, W.-X., and Nelson, P. T. (2007). Energizing miRNA research: a review of the role of miRNAs in lipid metabolism, with a prediction that miR-103/107 regulates human metabolic pathways. *Molecular Genetics and Metabolism, 91*(3), 209-217.

Williams, M., Tincho, M., Gabere, M., Uys, A., Meyer, M., and Pretorius, A. (2016). Molecular validation of putative antimicrobial peptides for improved Human Immunodeficiency Virus diagnostics via HIV protein p24. *Journal of AIDS and Clinical Research, 7*, 571.

Winawer, S. J., and Zauber, A. G. (2002). The advanced adenoma as the primary target of screening. *Gastrointestinal Endoscopy Clinics of North America, 12*(1), 1-9.

Wu, G., D'Agati, V., Cai, Y., Markowitz, G., Park, J. H., Reynolds, D. M., Maeda, Y., Le, T. C., Hou Jr, H., and Kucherlapati, R. (1998). Somatic inactivation of Pkd2 results in polycystic kidney disease. *Cell, 93*(2), 177-188.

Xie, B., Ding, Q., Han, H., and Wu, D. (2013). miRCancer: a microRNA–cancer association database constructed by text mining on literature. *Bioinformatics, 29*(5), 638-644.

Xu, J., Zhao, J., and Zhang, R. (2016). Four microRNAs Signature for Survival Prognosis in Colon Cancer using TCGA Data. *Scientific Reports, 6*.

Zhang, T., Guo, J., Gu, J., Wang, Z., Wang, G., Li, H., and Wang, J. (2019). Identifying the key genes and microRNAs in colorectal cancer liver metastasis by bioinformatics analysis and in vitro experiments. *Oncology Reports, 41*(1), 279-291.

Appendix A

Prognostic and Expression analysis

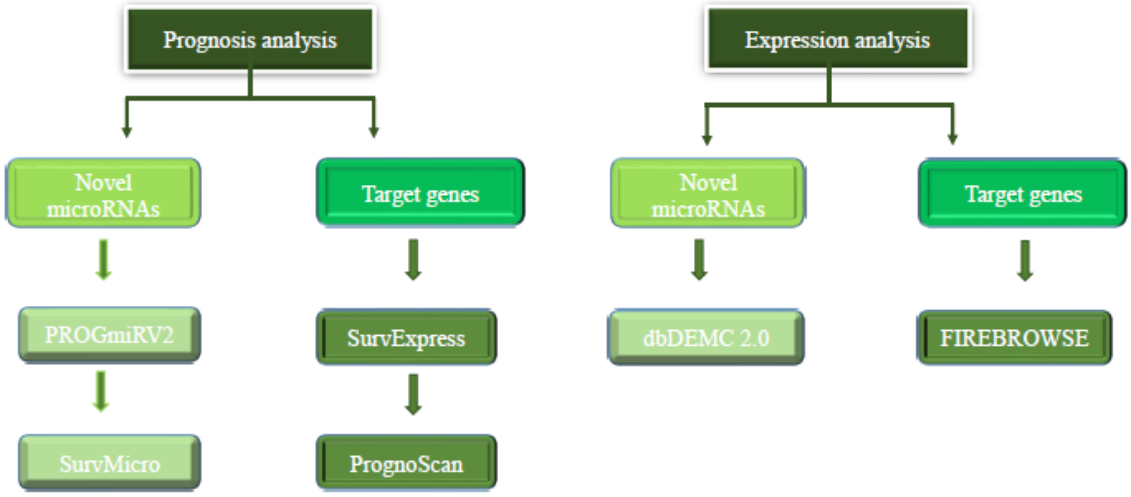


Figure A.1: Outline of the *in silico* methodology for prognostic and predictive analysis.

Table A.1: List of abbreviations used in graphical result from FIREBROWSE

S/N	Disease Name	Cohort	Cases
1	Adrenocortical carcinoma	ACC	92
2	Bladder urothelial carcinoma	BLCA	412
3	Breast invasive carcinoma	BRCA	1,098
4	Cervical and endocervical cancers	CESC	307
5	Cholangiocarcinoma	CHOL	51
6	Colon adenocarcinoma	COAD	460
7	Colorectal adenocarcinoma	COADREAD	631
8	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC	58
9	Esophageal carcinoma	ESCA	185
10	Glioblastoma multiforme	GBM	613
11	Glioma	GBMLGG	1,129
12	Head and Neck squamous cell carcinoma	HNSC	528
13	Kidney Chromophobe	KICH	113
14	Pan-kidney cohort (KICH+KIRC+KIRP)	KIPAN	973
15	Kidney renal clear cell carcinoma	KIRC	537
16	Kidney renal papillary cell carcinoma	KIRP	323
17	Acute Myeloid Leukemia	LAML	200

18	Brain Lower Grade Glioma	LGG	516
19	Liver hepatocellular carcinoma	LIHC	377
20	Lung adenocarcinoma	LUAD	585
21	Lung squamous cell carcinoma	LUSC	504
22	Mesothelioma	MESO	87
23	Ovarian serous cystadenocarcinoma	OV	602
24	Pancreatic adenocarcinoma	PAAD	185
25	Pheochromocytoma and Paraganglioma	PCPG	179
26	Prostate adenocarcinoma	PRAD	499
27	Rectum adenocarcinoma	READ	171
28	Sarcoma	SARC	261
29	Skin Cutaneous Melanoma	SKCM	470
30	Stomach adenocarcinoma	STAD	443
31	Stomach and Esophageal carcinoma	STES	628
32	Testicular Germ Cell Tumors	TGCT	150
33	Thyroid carcinoma	THCA	503
34	Thymoma	THYM	124
35	Uterine Corpus Endometrial Carcinoma	UCEC	560
36	Uterine Carcinosarcoma	UCS	57
37	Uveal Melanoma	UVM	80

Gene expression results from Firebrowse database

The list of prioritized microRNA target genes obtained in chapter 3 was submitted to FIREBROWSE for gene expression analysis among 37 cancer cohorts (Table 4.1).

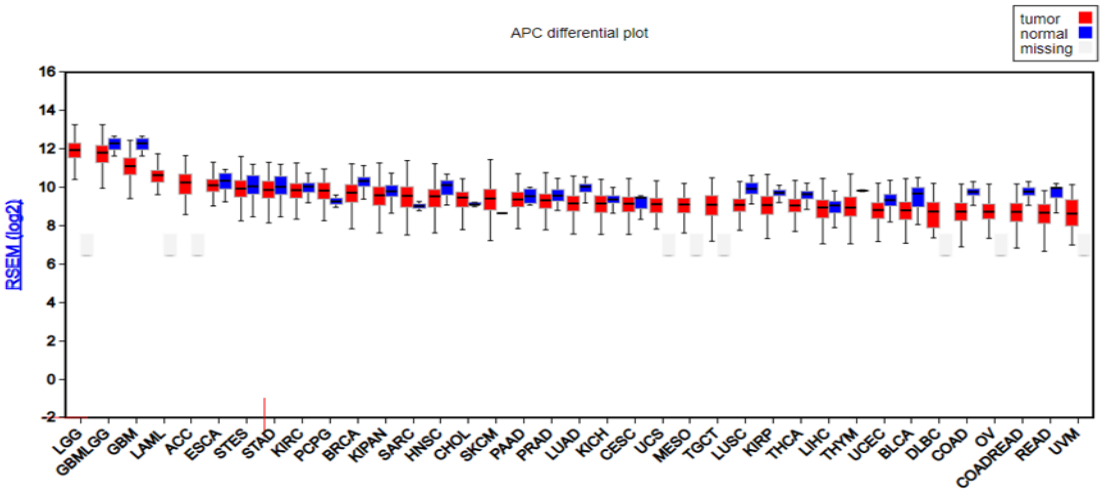


Figure A.2: Expression analysis of the APC gene in 37 cancers types with expression sorted from high to low.

From the expression analysis in Figure 4.6, high expression of the APC gene was presented in LGG while the overall expression of APC was found to be low in CRC. When comparing the expression levels between the normal and tumor colorectum, high expression of the gene was obtained in normal colorectal tissue relative to tumor sample.

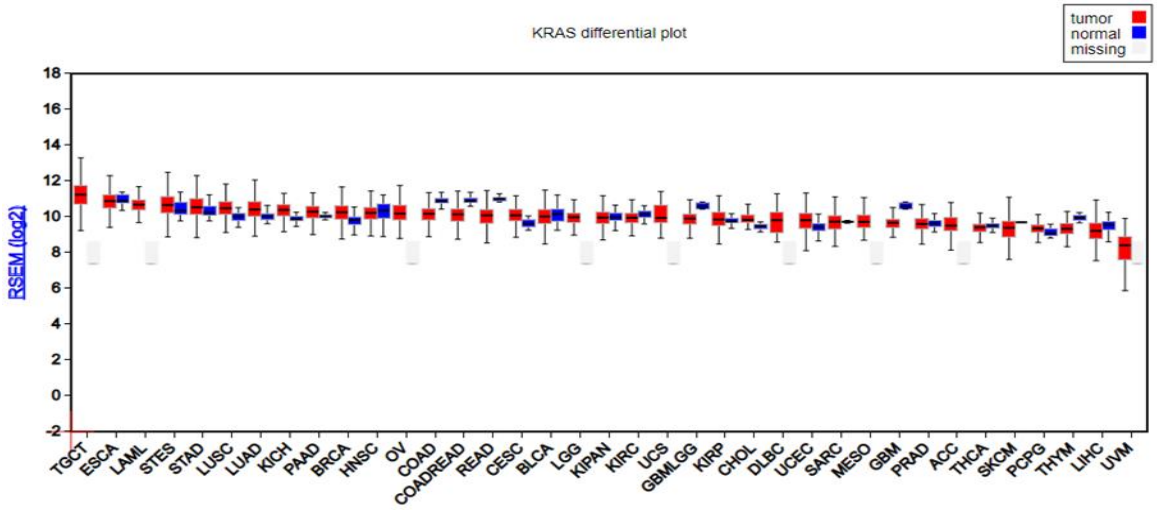


Figure A.3: Expression analysis of the KRAS gene in 37 cancers types with expression sorted from high to low.

From the cohort of the 37 cancer study, COADREAD appeared to be expressed in relatively high quantity (14th). The over-expression of this gene was found in TGCT followed by ESCA and LAML while the in UVM, the gene was least expressed.

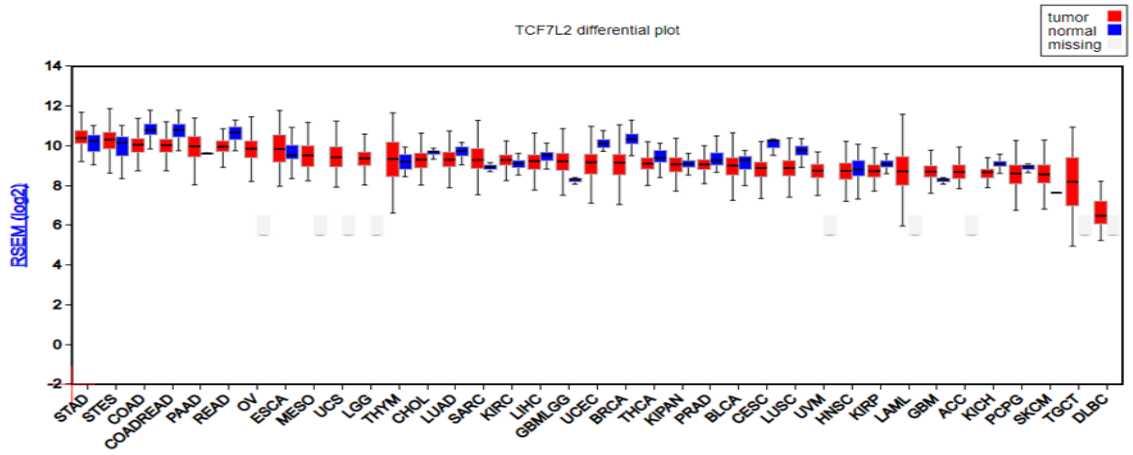


Figure A.4: Expression analysis of TCF7L2 in 37 cancers types with expression sorted from high to low.

The expression profile of TCF7L2 was analyzed using firebrowse as shown in Figure 4.8. This gene showed over-expression in STAD, STES, COAD and COADREAD. TCF7L2 is over-expressed in normal colorectal tissue when compared to the colorectal tumor. Also, its lowest expression was found in cancer type DLBC.

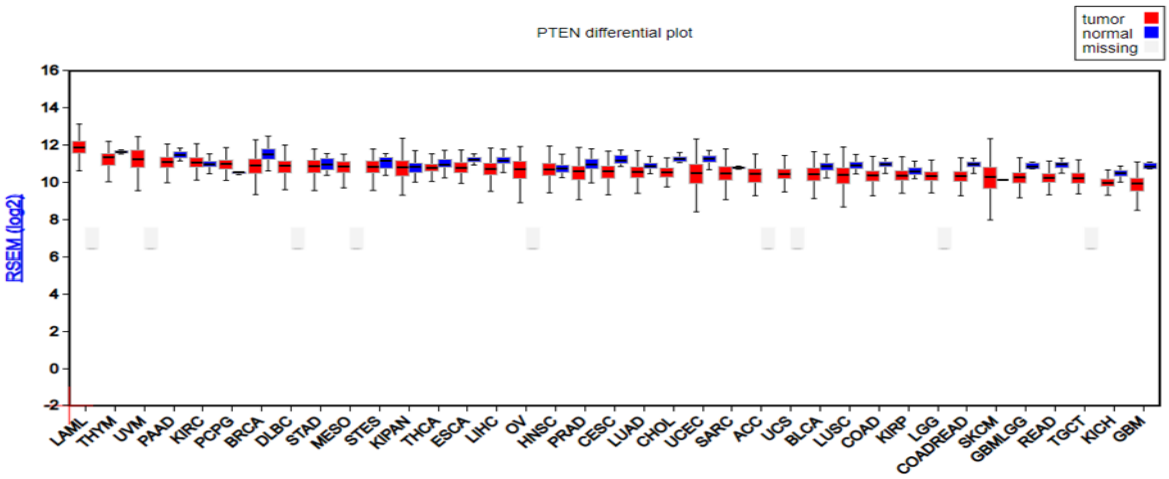


Figure A.5: Expression analysis of PTEN in 37 cancers types with expression sorted from high to low.

The expression analysis of PTEN among the 37 cancer cohort is presented in Figure 4.9. Over-expression of this gene was observed in LAML followed by THYM while low expression was obtained in COADREAD with GBM showing the least expression of PTEN. Expression of PTEN is high in normal colorectal tissue compared to the tumor tissue.

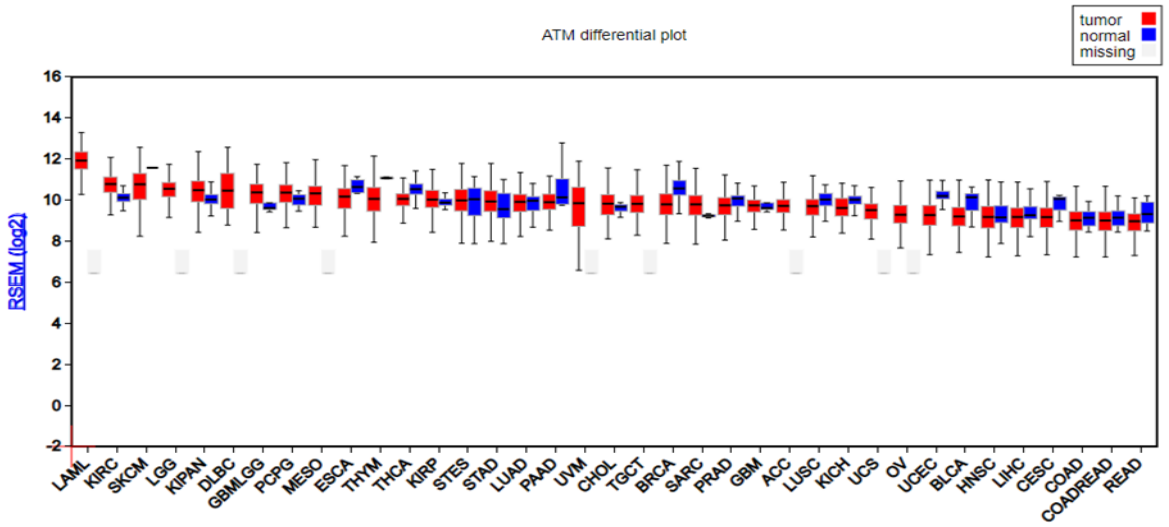


Figure A.6: Expression analysis of ATM in 37 cancers types with expression sorted from high to low.

From the expression analysis of ATM in Figure 4.10, high expression of APC gene was presented in LAML while the overall expression of APC expression was low in CRC after READ. When comparing the gene expression between the normal and tumor colorectal tissues, relatively high expression of the gene was obtained in normal colorectal tissue relative to tumor sample.

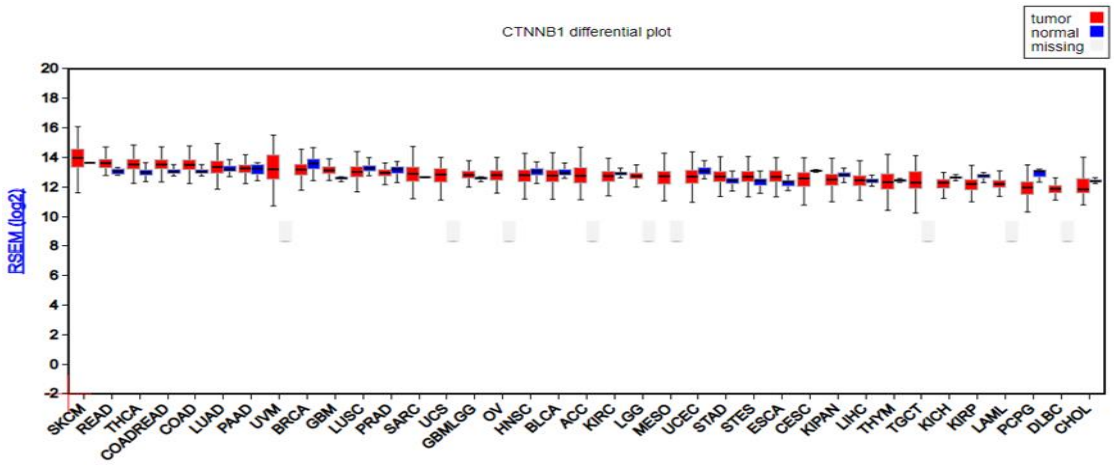


Figure A.7: Expression analysis of the CTNNB1 gene in 37 cancers types with expression sorted from high to low.

CTNNB1 is over-expressed in SKCM, READ, THCA, and COADREAD tissues, as shown in Figure 4.11, with CTNNB1 showing higher expression in tumor colorectal tissue when compared to colorectal tumor tissue.

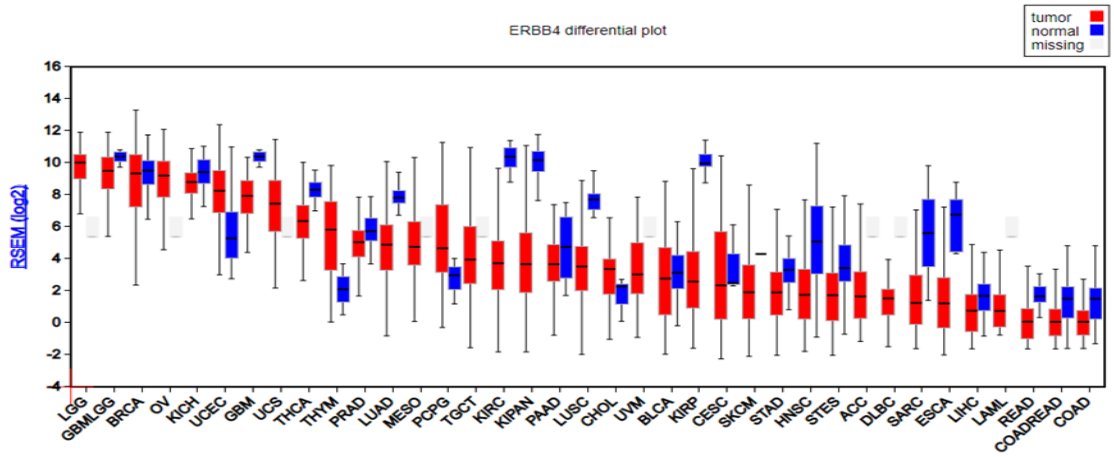


Figure A.8: Expression analysis of the ERBB4 gene in 37 cancers types with expression sorted from high to low.

ERBB4 shows very low expression in colorectal tumor tissue when compared to the other 37 tumor tissues with a higher expression in normal colorectal tissue compared to colorectal tumor tissue as shown in Figure 4.12.

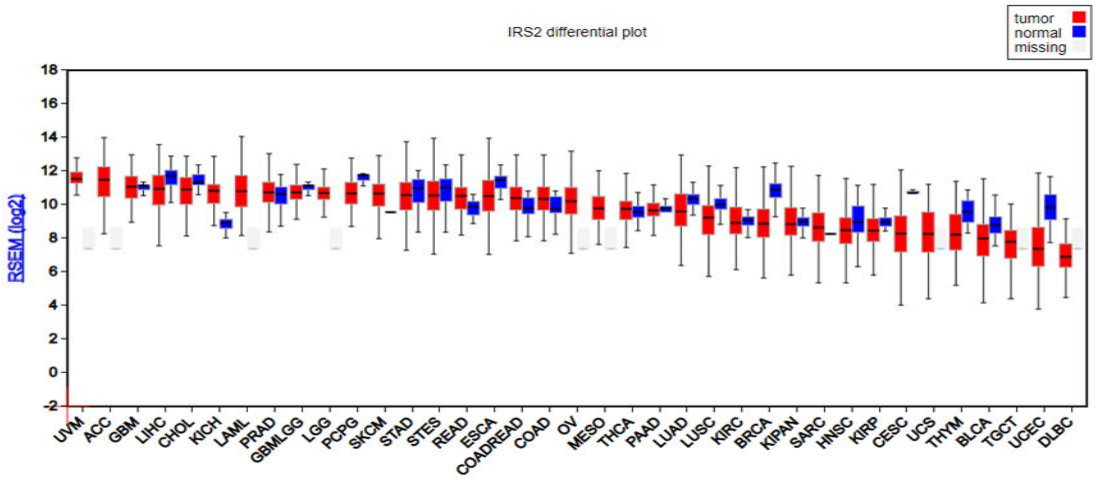


Figure A.9: Expression analysis of the IRS2 gene in 37 cancers types with expression sorted from high to low.

From the expression of IRS2 in the cancer cohort in Figure 4.13, it was observed that the gene expression in COADREAD is relatively low when compared to the overall expression analysis. Also, when compared with the normal COADREAD tissue, the expression of this gene in the tumor sample is relatively low.

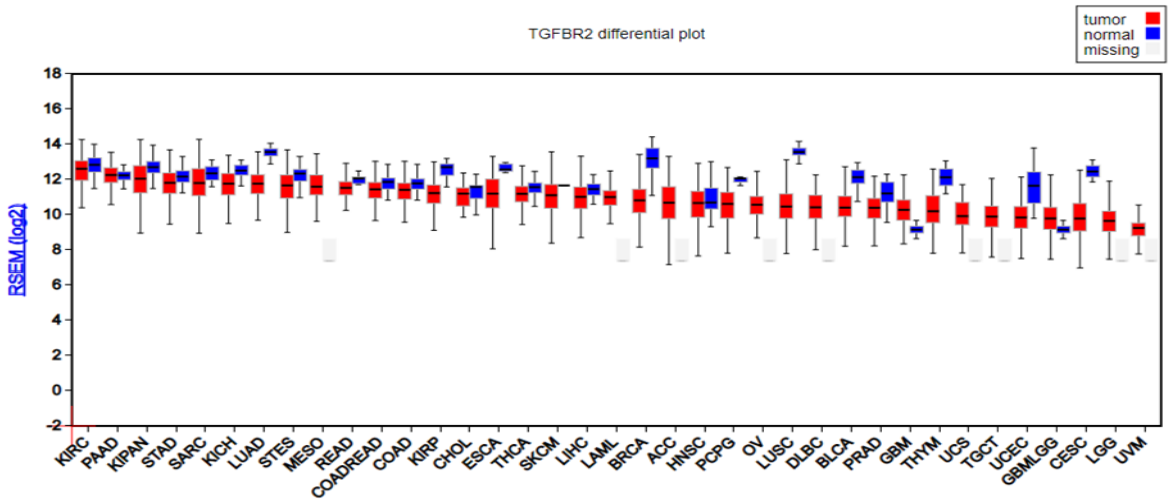


Figure A.10: Expression analysis of the TGFBR2 gene in 37 cancers types with expression sorted from high to low.

TGFBR2 shows relatively high expression in COADREAD when compared to the other 37 tumor tissues with a relatively higher expression in normal colorectal tissue compared to colorectal tumor tissue as shown in Figure 4.14.

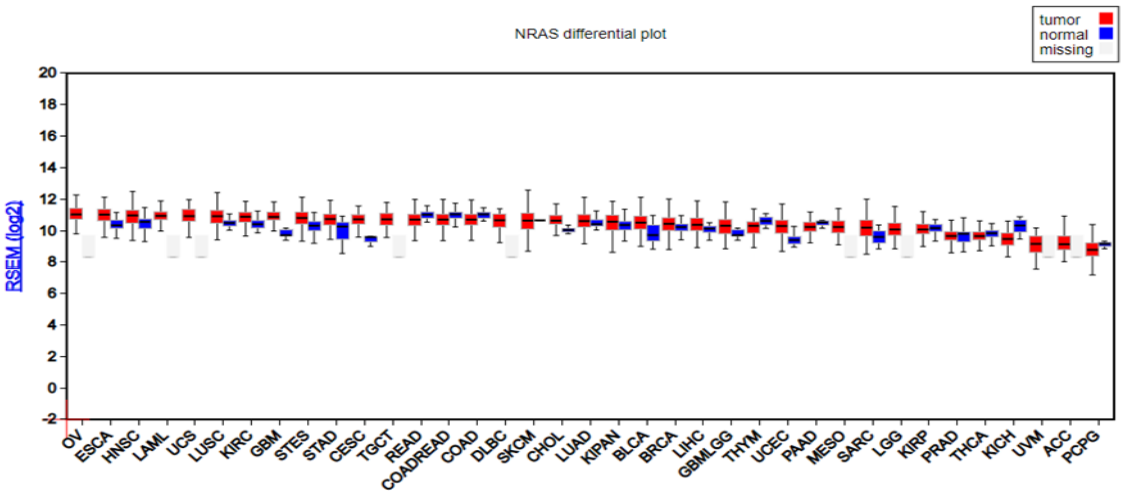


Figure A.11: Expression analysis of the NRAS gene in 37 cancers types with expression sorted from high to low.

NRAS shows relatively high expression in COADREAD when compared to the other 37 tumor tissues with a relatively higher expression in normal colorectal tissue compared to colorectal tumor tissue

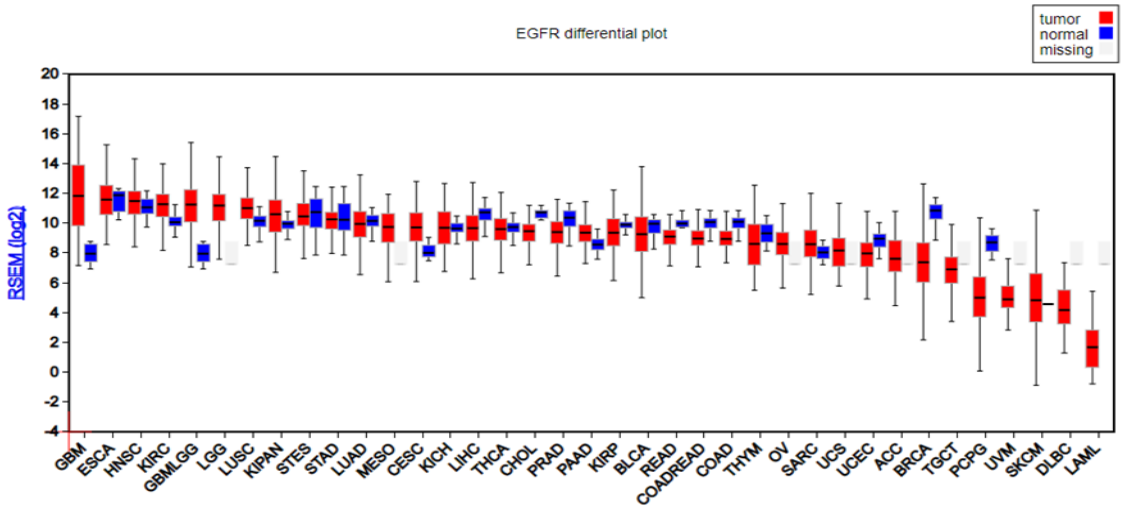


Figure A.12: Expression analysis of the EGFR gene in 37 cancers types with expression sorted from high to low.

From the expression of EGFR in the cancer cohort in Figure C.12, it was observed that the gene expression in COADREAD is relatively low when compared to the overall expression analysis. Also, when compared with the normal COADREAD tissue, the expression of this gene in the tumor sample is relatively low.

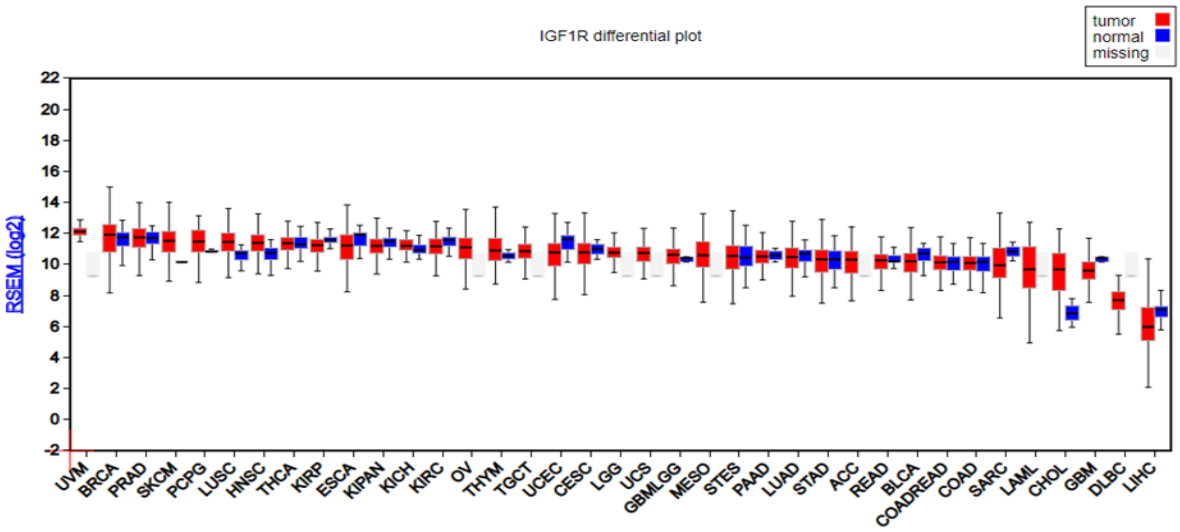


Figure A.13: Expression analysis of the IGF1R gene in 37 cancers types with expression sorted from high to low.

From the expression of IGF1R in the cancer cohort in Figure C.13, it was observed that the gene expression in COADREAD is relatively low when compared to the overall expression analysis. Also, when compared with the normal COADREAD tissue, the expression of this gene in the tumor sample is relatively the same.

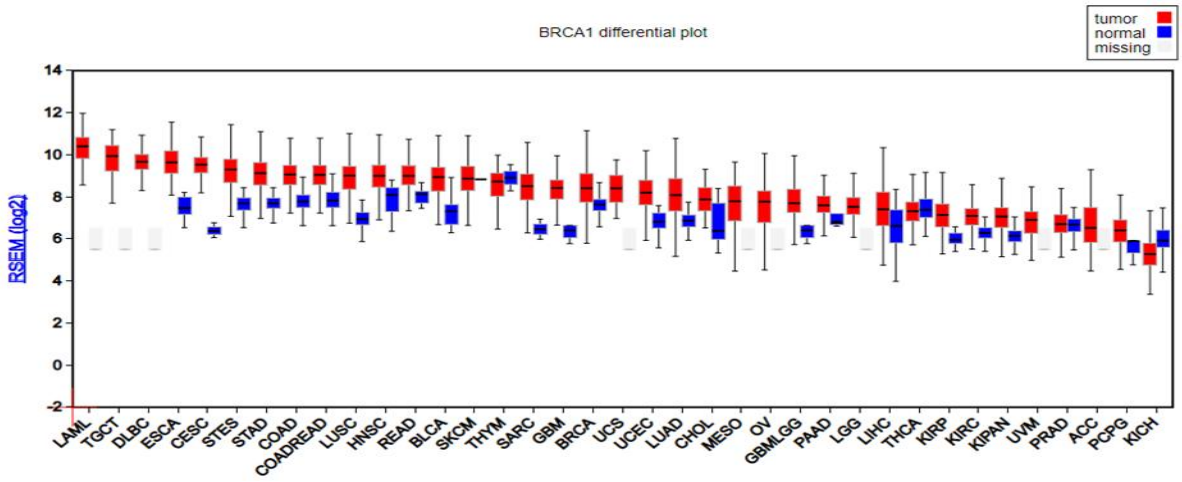


Figure A.14: Expression analysis of the BRCA1 gene in 37 cancers types with expression sorted from high to low. BRCA1 is over-expressed in LAML, TGCT, DLBC, ESCA, CESC, STES, STAD, COAD, and COADREAD tissues, as shown in Figure A.14, with BRCA1 showing higher expression in tumor colorectal tissue when compared to colorectal tumor tissue.

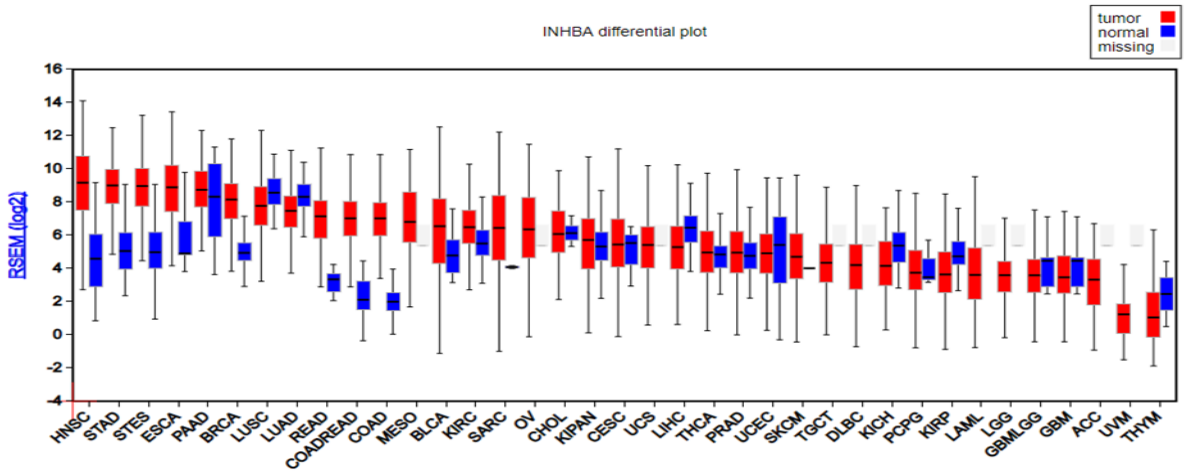


Figure A.15: Expression analysis of the INHBA gene in 37 cancers types with expression sorted from high to low. Figure A.15 illustrates the expression of INHBA in the 37 cancer types and was highly differentially expressed in HNSC, STAD, STES, ESCA, PAAD, BRCA, LUSC, READ, followed by COADREAD. All the other cancer types showed lower expression of the gene when compared to COADREAD. When compared with the normal colorectal tissue, expression of the INHBA gene was higher in the tumor tissue as indicated in the figure.

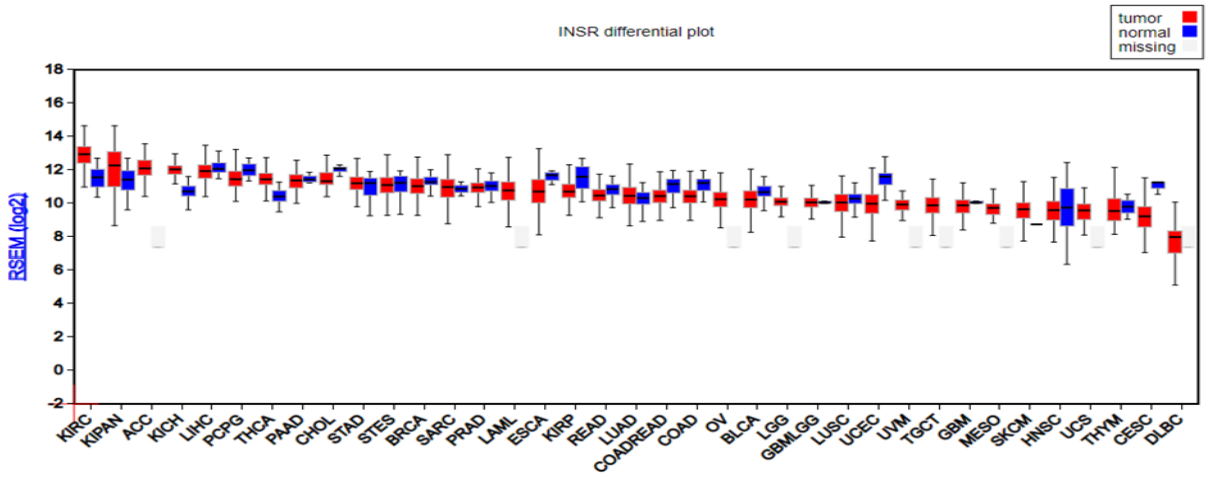


Figure A.16: Expression analysis of the INSR gene in 37 cancers types with expression sorted from high to low.

From the expression of INSR in the cancer cohort in Figure A.16, it was observed that the gene expression in COADREAD is relatively low when compared to the overall expression analysis. Also, when compared with the normal COADREAD tissue, the expression of this gene in the tumor sample is relatively low.

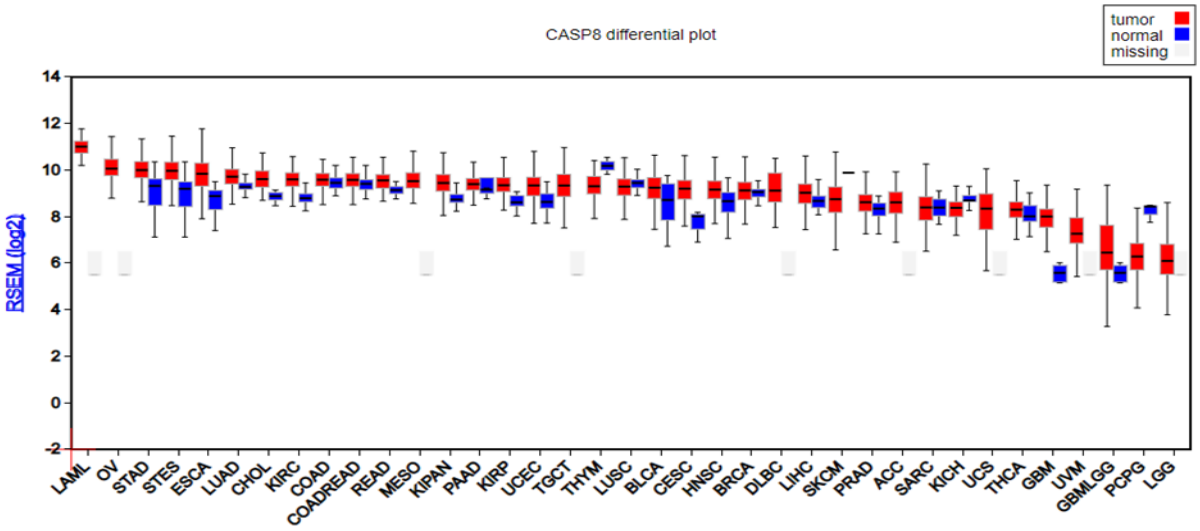


Figure A.17: Expression analysis of the CASP8 gene in 37 cancers types with expression sorted from high to low.

Figure A.17, illustrate the expression of CASP8 in the 37 cancer types and was highly differentially expressed in LAML, OV, STAD, ESCA, LUAD, CHOL, KIRC, COAD followed by COADREAD. All the other cancer types showed lower expression when compared to COADREAD. When compared with the normal colorectal tissue, expression of the gene was relatively higher in the tumor tissue as indicated in the figure**.**

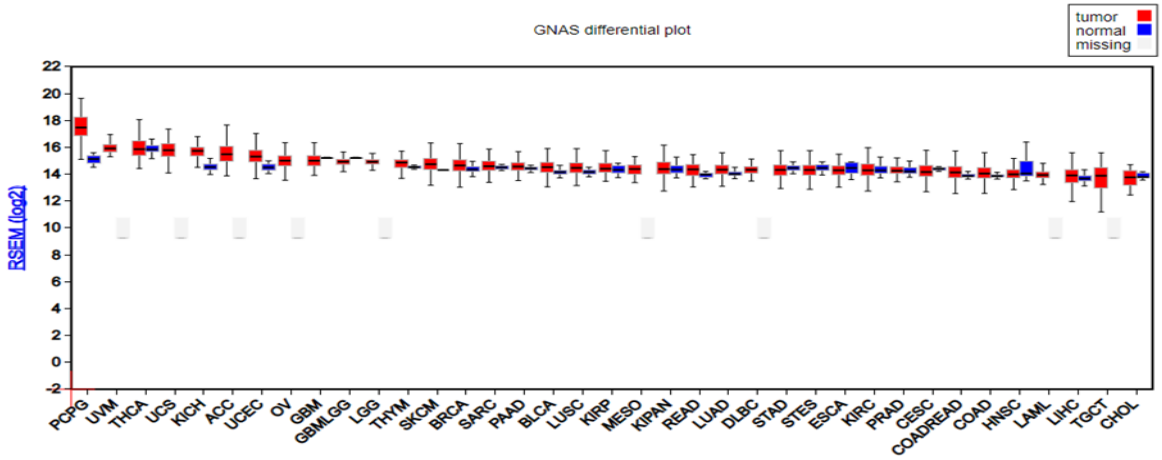


Figure A.18: Expression analysis of the GNAS gene in 37 cancers types with expression sorted from high to low.

From the expression of GNAS in the cancer cohort in Figure A.18, it was observed that the gene expression in COADREAD is relatively low when compared to the overall expression analysis. Also, when compared with the normal COADREAD tissue, the expression of this gene in the tumor sample is relatively high.

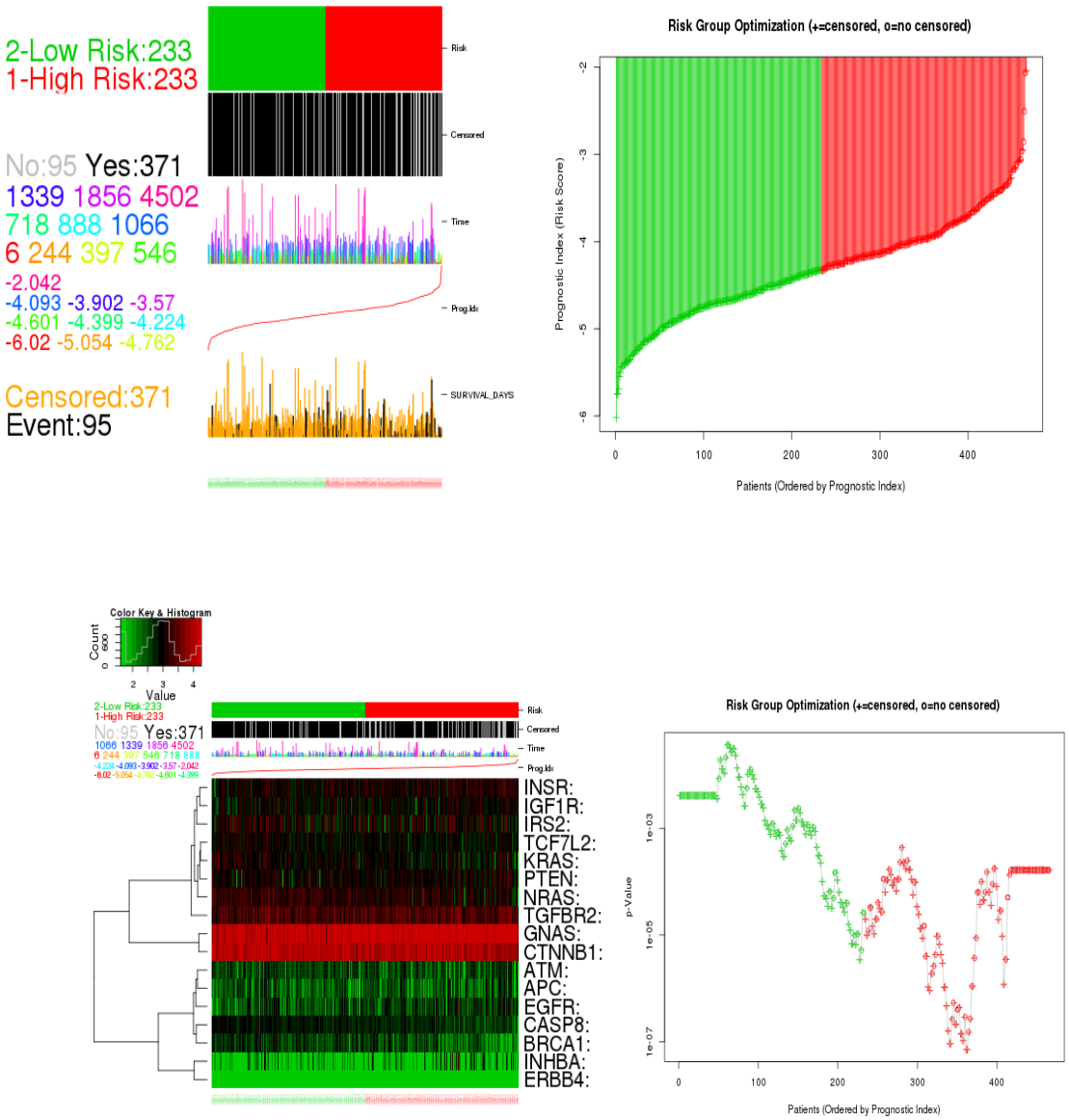


Figure A.19: Panel B shows a heat map representation of the gene expression values. Panel C shows the relation between risk groups and prognostic index. Panel A shows the clinical information available related to risk group, prognostic index, and outcome data.

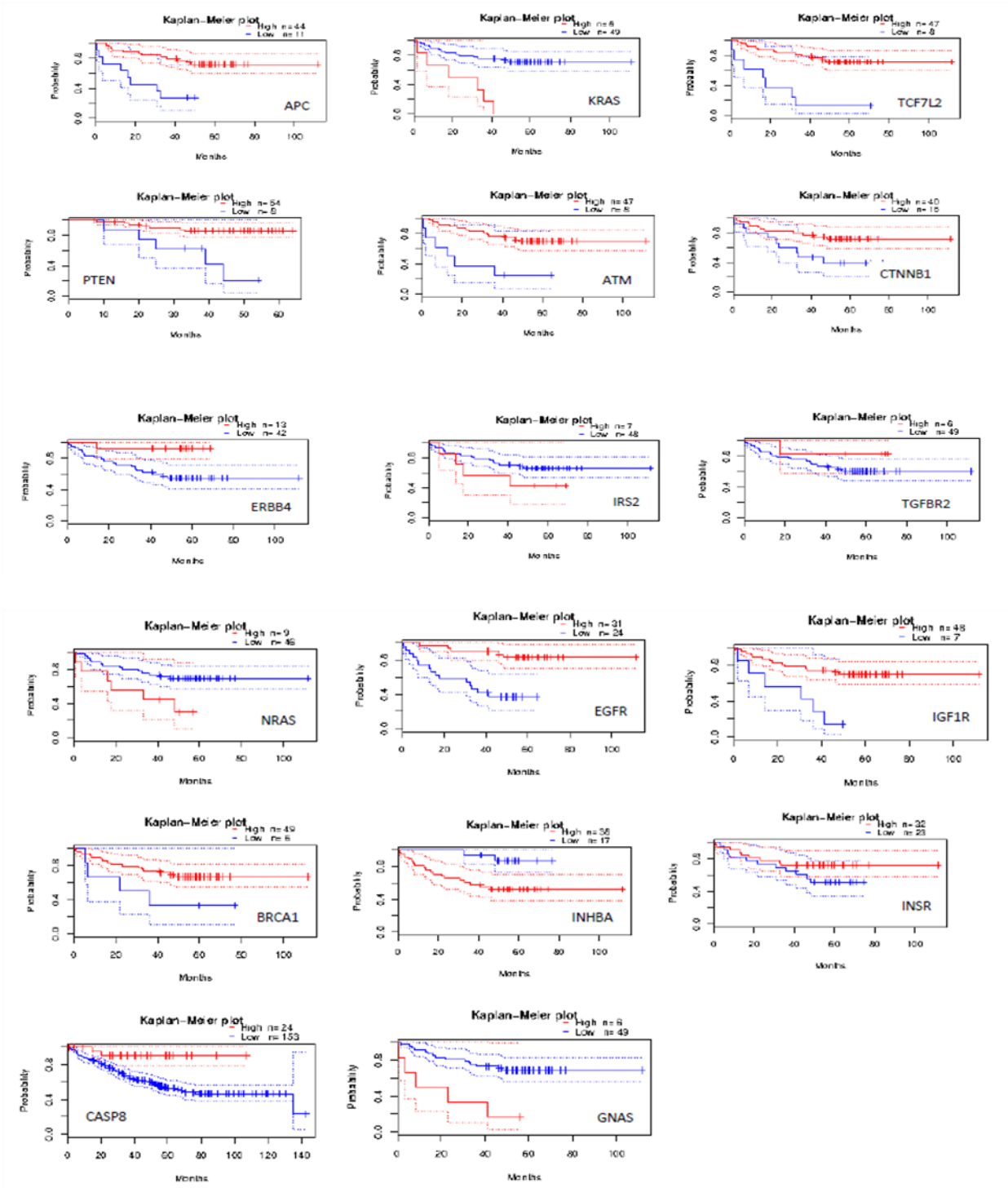
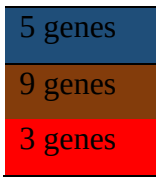


Figure A.20: The survival curves (Kaplan-Meier plots) for microRNA prioritized target gene list, showing the level of expression for the high-risk groups in red and for the low-risk groups in blue using the PrognoScan database. The X-axis represents time and the Y-axis represents survival rate. 95% confidence intervals for each group are also indicated by dotted lines.

Table A.2: Gene expression analysis of microRNA target gene in CRC using FIREBROWSE

S/N	Gene_ID	U*p*-regulation	Down regulation	No expression	Number among other subtype
1	CTNNB1				4
2	IRS2				17
3	CASP8				10
4	BRCA1				09
5	INHBA				10
6	APC				35
7	KRAS				14
8	TCF7L2				11
9	PTEN				31
10	ERBB4				36
11	TGFBR2				04
12	NRAS				14
13	EGFR				23
14	INSR				20
15	ATM				36
16	GNAS				31
17	IGF1R				30



Appendix B

Triplex binding analysis

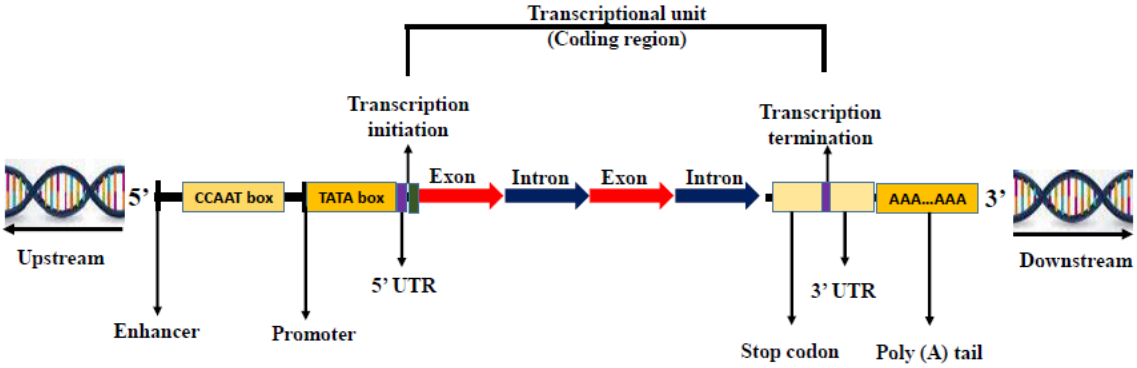


Figure B.1: The promoter region

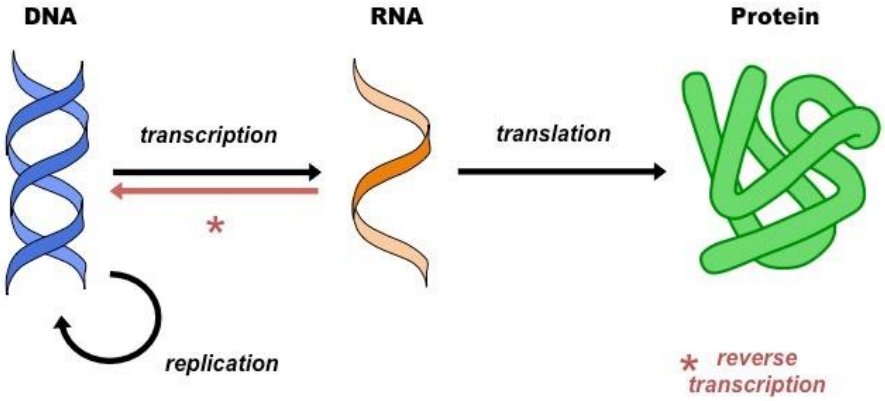


Figure B.2: Central dogma for molecular biology

Table B.1: The microRNA target genes' promoter regions

S/N	Gene_ID	Promoter Fasta Sequence
1	KRAS	>chromosome:GRCh38:12:25204189:25251536:-1 CCCGGTCTCCCTAAGTCCCCGAAGTCGCCTCCCACTTTTGGTGACTGCTTGTTTATTTA CATGCAGTCAATGATAGTAAATGGATGCGCGCCAGTATAGGCCGACCCTGAGGGTGGC GGGGTGCTCTTCGCAGCTTCTCTGTGGAGACCGGTCAGCGGGGCGGCGGGCCGCTCG CGGCGTCTCCCTGGTGGCATCCGCACAGCCCGCCGCGGTCCGGTCCCGCTCCGGGTC AGAATTGGCGGCTGCGGGGACAGCCTTGCGGCTAGGCAGGGGGCGGGCCGCCGCGTG GGTCCGGCAGTCCCTCCTCCCGCCAAGGCGCCGCCCAGACCCGCTCTCCAGCCGGCC CGGCTCGCCACCCTAGACCGCCCCAGCCACCCCTTCCTCCGCCGGCCCGGCCCCCGCT

		CCTCCCCCGCCGGCCCGGCCCGGCCCCCTCCTTCTCCCCGCCGGCGCTCGCTGCCTCC CCCTCTTCCCTCTTCCCACACCGCCCTCAGCCGCTCCCTCTCGTACGCCCGTCTGAAGA AGAATCGAGCGCGGAACGCATCGATAGCTCTGCCCTCTGCGGCCGCCCGGCCCCGAA CTCATCGGTGTGCTCGGAGCTC
2	**APC**	**>chromosome:GRCh38:5:112706898:112846839:1 TATTGTATTTCCTATTGCTGTAAATCTAGAGCCTGAAACAGTATCTAGTATATATCGGA AACCCAAATGATATTGTTCAATTAATGTTGAATGAAATGAAACTGAAGAGGTAATCCAA GGTAAAGAACTTATCAAACTGTGTGAGACAAACAAGGATTTCCCGGAAGAGGTGTTTT TAAATTAAAACTACAGGGTGGTAGTGAGAGGTGGTAGGGAAGCAGAGGTCGCAGCTA CAAAATAATAACTTGCTCATCTCCGTTTCTCCGCTATCTTATGAGCCCTTGTGGGGCTG GGACAGAATTTTATTCATCTTTCTATCATCAGCGTCTAGTACGGGGAGTAGCAAATAGT GAGCACTCGATAGATGTTTGCGGAATAATGGACTAGTGTGTGCAGAAGGATCTATTAA CTGGGCTGCAGCACAATTCAGAGAAGGCCAGTAAGTGCTGCAACTGAGACTCGGCTGC CTAGGCAGCAATGGCTCACGGGACAGAACAGCGAAGCAGTGCCCGGCAAGCGGAGCG CAGCACCCATTGCGCCTGCGCATAACAGGCTCTAGTCTCCGGGCTGTGGGAAGCCAGC AACACCTCTCACGCATGC**
3	**TCF7L 2**	**>chromosome:GRCh38:10:112949650:113168278:1 AAGCTCTCTGAAAGTGGAGGGTTGATCTTTCCTTTTGACCACTTTTGTCGCATCCCGTC TAAGGGTGGCTGATTTCACTGCTGAATTAACCACCAAGCACCCCCCCACCCCCTCCCC CAGCCACCACTTTCTCAAATACTACCCTTCCTTTTCCCCCTCCCTTAAGCTTATTTCTAA TATTTCTAAAGTGCACTGTTTTGGGCCTCTCCCCATCCCCGCCCCCCAAGTGGGCTTTC CTTCCGCCTTCCTCGGCTCGGATTCCTGACTTGGTCGCCACCCCCTTCTCCTCCTCTCC CACCCCGCATTGTCTTTCTGAAACCGCCCCCTCCCGGAGCAAGTCCCTGCACCCTCGC CCAGAATCCCGGGCTCGCACACACTCCGCGCAGGCCGCTCCCCCTGCACACCCTCCCT CCGTCTCCCCCCGGCTTCCCCGCCCCTCTCTTCCTCCTTCTTTCCCTCCTCCCTCTCCC GGCGCCCGAAAGGATCATTGTTAGCCGCCCCCGCCCCGCCCACCCCGGCTGTTTATTT ATGCACACGTCACTGGGCCGGCCCCGCCCTCCGGCATCTCATTAAGGCAGTGTGTTCC TCTCGCCCTGTC**
4	**EGFR**	**>chromosome:GRCh38:7:55018421:55212228:1 AGGACCCTCGGACTTTAGAGCACCACCTCGGACGCCTGGCACCCCTGCCGCGCGGGC ACGGCGACCTCCTCAGCTGCCAGGCCAGCCTCTGATCCCCGAGAGGGTCCCGTAGTGC TGCAGGGGAGGTGGGGACCCGAATAAAGGAGCAGTTTCCCCGTCGGTGCCATTATCC GACGCTGGCTCTAAGGCTCGGCCAGTCTGTCTAAAGCTGGTACAAGTTTGCTTTGTAA AACAAAAGAAGGGAAAGGGGGAAGGGGACCCTGGCACAGATTTGGCTCGACCTGGAC ATAGGCTGGGCCTGCAAGTCCGCGGGGACCGGGTCCAGAGGGGCAGTGCTGGGAACG CCCCTCTCGGAAATTAACTCCTCAGGGCACCCGCTCCCCTCCCATGCGCCGCCCCACT CCCGCCGGAGACTAGGTCCCGCGGGGGCCACCGCTGTCCACCGCCTCCGGCGGCCGC TGGCCTTGGGTCCCCGCTGCTGGTTCTCCTCCCTCCTCCTCGCATTCTCCTCCTCCTCT GCTCCTCCCGATCCCTCCTCCGCCGCCTGGTCCCTCCTCCTCCCGCCCTGCCTCCCCG CGCCTCGGCCCGCGCGAGCTAGAC**
5	**CASP8**	**>chromosome:GRCh38:2:201232843:201288311:1 TAAGCCTAGGAAAAAGTCTGAAAGAATCAAAATGTTAACAGCGGGGACCTCAAGGAAG CATTGAAGAGGCCATGGGAGAAGTTTTCACTTTGTTAAAAAATCAGTCCTTCAAATAAA TAAATACAGTGAGGCTTCCCCAGAAGCAGATGTCACTATGCTTCCTGTACAGCCTGTG GAACTGTGAGCCAGTTAAACCTCTTTTCTTTATAAATTATCCAGTCTTAGGATTTCTTT ATAACAGTGCTAGGATGAGCTGATACAGTTTCCTACACTGTAACCTAAGGCAATGCTTT GCACAAAGGGATGAGCCAGATTGCTTAGTAATTAAAACGCAAATACAAACCACAAGCA TATCCATTCATGAATTGGGGGGCTGCTTTGTGTGCATAGATAAGGTATATTTTTTAAAA AAATTATTT**

		TTCCAAGAAGAAAATAAACCAGTTAATAAACGACAACTCACAGTGCCAGGAAGTGAGA AACAAGTGTGTGATAAACGGTGGAGAATGGGAGCACTCTCCGCAGTGGGCGGGAGGA GACGAGGAGGGCGTTCCCTGGGGAGTGGCAGTGGTTGGAGCAAAGGTTTGGAGGAGG TAAGTCAT
6	**IGF1R**	**>chromosome:GRCh38:15:98647939:98965130:1 AGGTGCGCCCCTTCCGCCACGTTCGGGCTTTCCAGTACGCAGCGAAAAAAATGCCGCA TGCACGCATTTATTTATTTTGCAACAGCTGCAAGAAACAATGAAGCTTTTCAAGAACCG GGGAAACGCGCTTTCCAGCCGCGCTGTTGTTGTTTTCAATGAACCTCTCCCAGCCCCG CACTCCCCGCCCACCCCTCCCCTCTCCTGCCCACCCCTCCCCTGCCTAGCCTTTCCCTG GCTACCCACCCCTGCCCCGCCGAGACCGGACCGGCGGCGGGGGCATTGTTTTTGGAG TCGGGCGGGAGGGGAGGGCGCGTGCGGGGTGGCCGGCGCAGTGCGGTGGGGGCGGG AGCGGGTGGGCACGCGCGCGTGTCTCTGTGTGCGCGCGGGAGGCGGTGGGGCGGGA GATGGGGGCGGCGCCTCGCAGTCTCGCGCCCCACGCCCGGGCTCCGCTCCGCACGTC TTGGGGAACCGGGCTCCGGTTTTTTGCGCGCGCCGGCCTGGGCCGGGCCCTCGGCGC GCCGCTGCTGCGGCGGTGGCCGCTCGAGTGTGCGAGCGGGCGCGTGTGCGCGGGCCA GGGCGCGCGCGCGCGCGCGAGCCCCC**
7	**GNAS**	**>chromosome: GRCh38:20:58839118: 58911792:1 CAGTCCTCAGATGGGCTTGCAGACTTGCAAACTACTACCTCAACCGTGGGGGCCCTTG TTGAAGAAGATGAGGTAATTAGCCAATTCAGACTGTTTGAGCACGTATTGTTTGGAAA CTTTTCTAATTACAGTCACTATAACAGATCTGCTAAAGGTAGTTTCATGATTAATGGCT GCCCGAAGTTACCATCCTCAAACTATTAAGCCATTGGAAAATGCTAGAGGCTTAATACT CGTGTATAGACAGACCCTCATTCCCCTGAATATTTGCTATTTGGGGTTCTGGGTTTTTA GCTGCTCCCAGGCTGTGTTTATTTCCCCAATAACTGGGGAAACTGGGGAGGGGCACGA CCCCACGGGAGCCTGCGCCCACCTGCCCAAGTACTGGACCTGGGGCTAGCTTGCCGCT TGCTCCTTGCCACCCGCCAGGCCCCCCGCCCATCGCTTCGCCCAAATCCTTTCAAACA AGGTTCCCTCCTGCCACCTGCCGGCCCACTAGGGTCTGCGCCACAGGCTCGGCGCCAC CACGCAGCTCGCGGGGAGGTGGCCCCCACCTCCTTACTGCACATGCCCGGCAGAAGT CCGGGCGCGCAACTTCGC**