Genomic profiling of microRNA target genes in colorectal cancer

Adewale Oluwaseun Fadaka1,2, Olalekan Olanrewaju Bakare2, Ashley Pretorius2 and Ashwil Klein3

Abstract
Colorectal cancer is the second and third most common cancer in men and women, respectively, worldwide. Alterations such as genetic and epigenetic are common in colorectal cancer and are the basis of tumor formation. The exploration of the molecular basis of colorectal cancer can drive a better understanding of the disease as well as guide the prognosis, therapeutics, and disease management. This study is aimed at investigating the genetic mutation profile of five candidate microRNAs (hsa-miR-513b-3p, hsa-miR-500b-3p, hsa-miR-500a-3p, hsa-miR-450b-3p, hsa-miR-193a-5p) targeted by seven genes (APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and GNAS)) using in silico approaches. Two datasets (dataset 1 from our previous study and dataset two (The Cancer Genome Atlas, Nature 2012) were considered for this study. Protein–protein interaction, expression analysis, and genetic profiling were carried out using STRING, FireBrowse, and cBioPortal, respectively. Protein–protein interaction network showed that epidermal growth factor receptor has the highest connection among the target genes and this can be considered as the hub gene. Relative to other solid tumors, in colorectal cancer, six of the target genes were downregulated and only CASP8 was upregulated. Genes with protein tyrosine kinases domain were frequently altered in colorectal cancer and the most common alteration in these genes/domain are missense mutation. These results could serve as a lead in the identification of driver genes responsible for colorectal cancer initiation and progression. However, the intense mechanism of these results remains unclear and further experimental validation and molecular approaches are the focal points in the nearest future.

Keywords
Molecular profiling, colorectal cancer, target genes, microRNA, mutation, cBioPortal, FireBrowse

Date received: 25 November 2019, accepted: 18 May 2020

Introduction
One of the fundamental processes driving the initiation and progression of colorectal cancer (CRC) is the accumulation of a variety of genetic and epigenetic changes in epithelial cells of the colorectum.1 Over the past decade, major advances have been made in the understanding of cancer epigenetics, particularly regarding aberrant DNA methylation, microRNA and noncoding RNA deregulation, and alterations in histone modification states. The number of driver events required for human tumorigenesis has remained one of the fundamental issues in cancer research.2 Gene mutations have long been known to be important in cancer formation.3 However, epigenetic alterations have only recently been recognized as significant contributors to cancer

1Department of Science and Technology/Mintek Nanotechnology Innovation Centre, Biolabels Node, Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Bellville, South Africa
2Bioinformatics Research Group, Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Bellville, South Africa
3Plant Omics Group, Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Bellville, South Africa

Corresponding author:
Adewale Oluwaseun Fadaka, Department of Science and Technology/Mintek Nanotechnology Innovation Centre, Biolabels Node, Department of Biotechnology, Faculty of Biotechnology, Faculty of Natural Sciences, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa.
Email: afadaka@uwc.ac.za
development. Genomic sequencing of malignancies from cohort studies will advance researchers’ understanding of the genomic landscape and insight into their cancer pathogenesis and treatment. However, limited information regarding the demographics, epidemiology, and clinical annotations for cancer cases has been a major setback for complex genomic exploration. Another notable drawback to cancer sequencing study is the limited statistical power to significantly identify mutated genes that have an intermediate or lower frequency of mutation (e.g. <5% frequency). Cancer driver genes can be extensively studied by increasing the sample number sequenced in each disease type, mostly in CRC. Genomic exploration and cancer molecular profiling are useful tools to unravel potential tumor biomarkers. The prevailing consensus suggests that epigenetic alterations in CRC occur early and manifest more frequently than genetic alterations. The expression of microRNAs has different effects on the molecular mechanism of carcinogenesis of the gastrointestinal tract from adenoma to carcinoma. The wide range of expression patterns of these microRNAs in CRC found in tumors and healthy tissues could be associated with diagnosis and prognosis. Circulating microRNAs have also been considered as diagnostic and prognostic biomarkers following their correlation with CRC.

From our previous study, 5829 genes were targeted by these microRNAs. These targets were identified using three different databases (miRDB, TargetScan, and mirDIP) according to the microRNA sequences (candidate and validated) and the following criteria: microRNA 3’ site, Conservation Status, and the seed region. The resulting lists generated 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 by CD-HIT. Furthermore, DAVID (Database for Annotation, Visualization, and Integrated Discovery) database was employed to identify genes associated with CRC from differentially expressed gene list. The genes were defined as differentially expressed genes that were significantly related to CRC at \( p < 0.05 \) (final gene list; \( p < 0.05 \)). Based on the prioritization and function assignment, the top seven genes, namely, APC, KRAS, TCF7L2, epidermal growth factor receptor (EGFR), insulin-like growth factor receptor (IGF1R), CASP8, and GNAS were considered for the genomic profiling study in this article. Different cells can express hundreds of copies of certain microRNA alone, and cancer cells are not excluded. Efforts have been made to identify CRC-specific microRNA transcripts as prospective diagnostic biomarkers which can be used to develop precancerous stage detection. In addition, advances in genomic technologies have led to the identification of a variety of specific epigenetic alterations as potential clinical biomarkers for CRC patients. This research briefly investigates the frequency of alterations of the microRNA target genes (APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and GNAS) in colorectal adenocarcinoma (TCGA, Nature 2012) patients/samples, and details of the field regarding the distribution of mutation across protein domains, and clinical usefulness of various alterations as biomarkers for the management of CRC patients.

**Materials and methods**

**Data selection**

In this study, two datasets were queried using the cBioPortal, with the first dataset being the differentially expressed microRNA target genes from our previous in silico studies. Seven of the genes targeted by our candidate microRNAs that showed significance (Figure 1) were profiled and their expression in CRC was further evaluated. The Colorectal Adenocarcinoma in The Cancer Genome Atlas (TCGA) Colorectal Cancer project (TCGA, Nature 2012) consisting of the whole-exome sequencing in colorectal carcinoma tumor/normal pairs was used as the second dataset.

**Analysis of microRNA target genes for cancer genomics**

cBioPortal is an open-access resource for exploring multidimensional genomics data. The database substantially reduces the drawback between complex genomic data and cancer researchers in order to provide high-quality access to molecular profiles and their clinical attributes. This further provides easy result interpretation of rich datasets into biologic insights and clinical applications. The cBioPortal database was used to query multiple genes (dataset 1) in a single dataset (dataset 2) (Figure 1(b)). The portal was accessed at http://www.cbioportal.org/index.do for Cancer Genomics. The genomic datasets were queried using the interactive web interface with the option to query a single cancer study to explore the relevant genomic profile in the microRNA target genes in CRC samples.

**Protein–protein interaction network**

STRING (Search Tool for the Retrieval of Interacting Genes) accessed at http://string-db.org/ is a unique tool, equipped to provide a comprehensive view of all the known and predicted interactions and associations among proteins. This database was used to construct a protein–protein interaction (PPI) network of the candidate microRNA target genes.
Expression analysis of the target genes

FireBrowse is a web-based expression analysis tool that provides easy and refined means to investigate the TCGA cancer data, using a powerful computational infrastructure, application programming interface (API), graphical tools, and online reports. This database sits above the TCGA-Genomic DNA Affinity Chromatography Firehose, one of the extensive and most effectively characterized web-based cancer datasets in the world. MicroRNA target gene expression in some solid tumors and in corresponding normal tissues was analyzed using data from TCGA. Data analysis was performed using FireBrowse (http://firebrowse.org/), which provides access to analyze data generated by TCGA. From the interface, colorectal adenocarcinoma was selected as the cohort after which microRNA target genes were submitted individually to generate an expression profile between a tumor and normal tissues among 12 solid tumors (Table 1). RNA-seq by expectation-maximization (RSEM) was selected as the data format with the expression sorted alphabetically.

### Statistical analysis

The parameter of interactions in STRING was set as the highest confidence >0.900. Degree of association of the candidate microRNA target genes was considered using both Spearman’s and Pearson’s correlation coefficient and test of goodness of fit ($R^2 > 0.250$). A $p < 0.05$ was considered to be statistically significant.

### Data availability

The datasets and the clinical data were obtained from the online databases as described above in the “Materials and methods” section. cBioPortal, STRING, and FireBrowse were accessed from the web-server.

### Results

#### Dataset identification

The seven microRNA target genes obtained from our previous study (Figure 1(a)) were used alongside data
from TCGA as query in order to explore the genomic profiles of these genes following the workflow presented in Figure 1(b). The criteria for target genes selection was based on gene prioritization, functional assignment, and association with CRC using bioinformatics approaches and the \( p \) value was set at 0.05 for significance. The final gene list considered in this study includes APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and GNAS.

The genetic landscape of the candidate microRNA target genes

It is crucial to correlate genomic mutations and gene expression patterns in clinical stratification of CRC patients. The cBioPortal cancer genomics tool was used to evaluate the genetic alterations of the microRNA target genes in CRC patient samples. Genetic alterations of APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and GNAS (e.g. mRNA expression (upregulation and downregulation) and copy number alterations (deletion and amplification)) were evaluated in the TCGA dataset (TCGA, 2012) to determine their association with colorectal pathogenesis and progression. From the oncoprint, 179 cases (92%) were alteration in at least one of the seven microRNA target genes, with the frequency of alteration in each of the seven genes as shown in Figure 2. It was observed that of the seven candidate genes, APC was the most frequently altered gene (77% with 150 cases altered) followed by KRAS (43% with 83 cases altered), TCF7L2 (18% with 36 cases altered), GNAS (13% with 26 cases altered), IGF1R (8% with 16 cases altered), EGFR (7% with 13 cases altered), and CASP8 (6% with 12 cases altered). The alterations observed include truncating mutations, deep deletion, missense mutation (putative driver), and amplification.

Genetic profiles of microRNA target genes in CRC using cBioPortal

The mRNA expression of the microRNA target genes was plotted against copy number alterations using cBioPortal (Figure 3). Inactivation of the microRNA target genes was observed to be due to mutation, amplification, and deletions. However, these alterations are not the only way to decrease the levels of functional protein in a cell. Colorectal adenocarcinoma (TCGA, Nature 2012) database was selected to observe the relationship between APC (Figure 3(a)), KRAS (Figure 3(b)), TCF7L2 (Figure 3(c)), EGFR (Figure 3(d)), IGF1R (Figure 3(e)), CASP8 (Figure 3(f)), and GNAS (Figure 3(g)) copy number alterations and mRNA levels. Five alterations (deep deletion, shallow deletion, diploid, gain, amplification) were detected in this study. Six out of seven (APC, KRAS, TCF7L2, EGFR, IGF1R, and CASP8) of the microRNA target genes present more gains and less deletions, and were mostly diploid in colorectal carcinoma. GNAS mRNA is increased in the samples in which GNAS is gained and amplified. These results indicated that the mRNA expressions of the target genes (APC, KRAS, TCF7L2, EGFR, IGF1R, and CASP8) are associated with genetic status (deletion or amplification) in CRC.

Co-expression/correlation analysis

The graphical representation and the statistical results of the co-expression analysis are presented in Figure 4 and Table 2. The graphs analyzed the mRNA expression of each gene in the colorectal adenocarcinoma samples and statistically ranked the degree of association of genes using the test of goodness of fit, Spearman’s rank, and Pearson’s correlation coefficient.
(Figure 4(a)–(g), Table 2). These assess the linear relationships and monotonic relationships between microRNA target genes and the co-expressed genes. The closer it is to zero, the weaker the association. The result identified seven co-expressed genes that are positively correlated with the target genes in our study (p < 0.05). Functions of the microRNA target genes in CRC may be further explored through APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and GNAS.

**Survival analysis**

The association between microRNA target genes and survival in CRC patients was studied using the survival
The feature in cBioPortal. The figure shows the two curves; cases with alteration(s) in query microRNA target genes and cases without alteration(s) in query genes. The survival curve indicated that CRC patients with microRNA target genes amplification had no significant association between the microRNA target genes and overall survival ($p = 0.964$, Figure 5) in 2 years. However, it could be seen from the figure that the survival rate was increased when the microRNA target genes were upregulated. Survival analysis is crucial in cancer research as the majority of cancer studies investigate time to event endpoints. This is equally used to define prognostic indices for mortality or recurrence of a disease, and to study the outcome of treatment. The survival plot was used to estimate the survival probability of the candidate microRNA target genes in CRC patients with overall survival as the endpoint. This was done by comparing the survival rate between cases with alteration in query genes and its counterpart at the end of the time curve of the survival analysis. Although the
association with survival analysis is not significant, it could be observed that patients with microRNA target genes alteration have a better survival.

MicroRNA target genes mutation hotspot analysis

The mutation hotspots of the target genes were analyzed using the cBioPortal web-based tool. Mutations are drawn as a lollipop along with the domain structure of the gene. The height of the lollipop reflects how many times the mutation was detected. Each lollipop shows the hotspot for specific gene mutations in CRC. The graphical summary showed the position and frequency of all the mutations in the context of Pfam protein domains encoded by the canonical gene isoform (Figure 6). The integration of heterogeneous evidence is an important step toward personalized cancer treatment to catalog mutational hotspots that are biologically and therapeutically relevant and thus represent where targeted therapy would likely be beneficial.20 However, existing methods do not sufficiently portray varying functionality of individual mutations within the same genes. The mutations hotspot provided details as a graphical illustration of all nonsynonymous mutations identified in each query gene.

PPI network

In order to establish the biological roles of the microRNA target genes, the STRING web-based tool was used to establish a PPI (Figure 7). The network consists of 7 nodes and 12 edges with an average node degree of 3.43 and the average local clustering coefficient of 0.767 at PPI enrichment p value of $5.25 \times 10^{-5}$. EGFR was considered as the hub gene connecting five nodes followed by KRAS and IGF1R, which were connected by four nodes. Other genes were connected by three nodes except TCF7L2 which interacted with only APC. The majority of the interactions are either experimentally validated or curated from the database. This information, therefore, suggests novel directions for further research and provides cross-species predictions for effective interaction mapping.

Gene expression analysis

To understand the mechanism underlying the dysregulation of the microRNA target genes in CRC, their expression levels were investigated using FireBrowse web-based tool (Figure 8(a)–(g)). Data mining showed that APC, KRAS, TCF7L2, and EGFR were significantly downregulated in COREAD than in normal colorectal tissue as opposed to the expression of CASP8 and GNAS, but the expression of IGF1R is relatively the same in both COREAD and normal tissue (Figure 8). Relative to other solid tumors, in CRC, the expressions of six of the microRNA target genes, namely, APC, KRAS, TCF7L2, EGFR, IGF1R, and GNAS, are relatively low. But GNAS expression in COREAD was significantly increased in CRC compared with other solid tumors.

Discussion

Colorectal carcinogenesis is a multistep process that results from the accumulation of numerous genetic
alterations. The activation of multiple signaling pathways plays an important role in regulating cell proliferation, angiogenesis, cell motility, and apoptosis.\textsuperscript{21,22} One of the major goals of cancer genomics is the identification of the driver genes responsible for tumor initiation and progression. The microRNAs targeted genes in this study were carried using the in silico approach to detect the frequency of mutation in each gene and investigate the frequency of mutation distribution across all protein domains present in the nucleotide sequences of the genes. The microRNA target genes were used as dataset due to the significance of the associated microRNAs from our previous study (hsa-miR-513b-3p, hsa-miR-500b-3p, hsa-miR-500a-3p, hsa-miR-450b-3p, hsa-miR-501-3p, hsa-miR-501-3p, and hsa-miR-513a-3p).\textsuperscript{8} After the assignment of function and the determination of the mechanism of action to these microRNAs and target genes (previous studies),\textsuperscript{15,16,23} their genomic profiling deserves closer attention in CRC.

The integration of heterogeneous evidence is an important step toward personalized cancer treatment to catalog mutational hotspots that are biologically and therapeutically relevant and thus represent where targeted therapy would likely be beneficial.\textsuperscript{20} However, existing methods do not sufficiently portray varying functionality of individual mutations within the same genes. The mutations hotspot provided details as a graphical illustration of all nonsynonymous mutations identified in each query gene. The graphical summary showed the position and frequency of all mutations in the context of Pfam protein domains encoded by the canonical gene isoform.

The mRNA expressions of the target genes are associated with genetic status (deletion or amplification) of APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image1.png}
\caption{Figure 6. Distribution of mutations of microRNA target genes across protein domains in CRC. Note: Black lollipop: Truncation, green: missense and brown: in-frame insertion and deletion.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image2.png}
\caption{Figure 7. Protein–protein interaction (PPI) network of the significant microRNA target genes. Nodes represent the genes while edges represent PPI. The color of the edges indicates varying PPI evidence.}
\end{figure}
GNAS in CRC. The cross-reference between the microRNA expression profiles and the expression profiles of its target genes can provide an effective viewpoint to understand the regulatory functions of the microRNA. Functions of the microRNA target genes in CRC may be further explored alongside the associated genes. Survival analysis is crucial in cancer research as the majority of cancer studies investigate time-to-event endpoints. This is equally used to define prognostic indices for mortality or recurrence of a disease, and to study the outcome of treatment. The survival plot was used to estimate the survival probability of the candidate microRNA target genes in CRC patients with overall survival as the endpoint. This was done by comparing the survival rate between cases with alteration in query genes and its counterpart at the end of the time curve of the survival analysis. Although the association with survival analysis is not significant, it could be observed that patients with microRNA target genes alteration have a better overall survival.

Although CRC is regarded as one of the most common malignant diseases worldwide, its involvement in signaling pathways and driven-genes is not completely elucidated. The integration of heterogeneous evidence is an important step toward personalized cancer treatment to catalog mutational hotspots that are biologically and therapeutically relevant and thus represent where targeted therapy would likely be beneficial. However, existing methods do not sufficiently portray varying functionality of individual mutations within the same genes. The mutations hotspot provided details as a graphical illustration of all nonsynonymous mutations identified in each query gene. From the mutation point of view, APC is a tumor suppressor gene involved in the Wnt signaling pathway. There are 30 missenses and 90 truncation mutations in the sample (not shown).

**Figure 8.** Differential plots of microRNA target genes in selected solid tumors in TCGA. Expressions of (a) APC, (b) KRAS, (c) TCF7L2, (d) EGFR, (e) IGF1R, (f) CASP8, and (g) GNAS. Sample number (N) = 626 for all tumor samples and 51 for all the normal tissues. Red bars indicate tumor, blue bars indicate normal, and white bars indicate missing samples. Note: the black box indicates the cancer subtype of interest chosen in this study (colorectal cancer). COADREAD: colon adenocarcinoma and rectum adenocarcinoma.
For the missense mutation, R1450 is commonly altered although the APC_basic domain appears to be frequently mutated. The APC R1450 is recurrently altered in CRC and it is likely to be oncogenic with loss of function as its biological effect with a somatic mutation frequency of 76.4% (percentage of sample with a somatic mutation in APC). KRAS, a GTPase that functions as an upstream regulator of the MAPK and PI3K pathways, is frequently mutated in a diverse range of cancers including colorectal cancers. G12D is the most common missense altered with the somatic mutation frequency of 41.0% and the Ras: Ras family (5–164) appeared to be frequently mutated. This mutated amino acid was identified as a recurrent hotspot (p < 0.05) and a 3D clustered hotspot in a population-scale cohort of tumor samples of various cancer types. The KRAS G12D mutation may be oncogenic with loss of function. The prognostic implications for KRAS mutations vary between cancer types but have been shown to be associated with poor outcome in CRC, non-small cell lung cancer, and others. Also, the frequency of alteration observed in the novel microRNA gene target is supported by Roa et al. who reported similar frequency of alteration in KRAS in CRC samples. TCF7L2 is regarded as a tumor suppressor and transcription factor. This gene is altered by mutation or deletion in various cancer types, most frequently in CRC. CTNNB1_binding and HMG-H domains are the most frequently mutated with K483_K485del and C486vfs*8 as the most common missense mutations (in-frame deletion) with other mutations such as in-frame deletion, non-stop, fusion, and frameshift deletion. The oncogenic function of K483_K485del mutation is considered unknown. This mutated amino acid was identified as a recurrent hotspot (statistically significant) in a population-scale cohort of tumor samples of various cancer types. EGFR plays a vital role in promoting cell growth. The protein tyrosine kinase is the most frequently mutated domain commonly mutated with L861Q as the most common missense mutation (in-frame deletion) with other mutations such as in-frame deletion, non-stop, fusion, and frameshift deletion. The oncogenic function of K483_K485del mutation is considered unknown. This mutated amino acid was identified as a recurrent hotspot (statistically significant) in a population-scale cohort of tumor samples of various cancer types.

IGF1R has been linked to CRC metastasis. Data suggested that IGF1R is an attractive therapeutic target for several tumor types including CRC. Shiratsuchi et al. also suggested that the expression of IGF1R may be linked to increasing tumor size in CRC. Other research also confirms the association of this gene with CRC. CASP8 plays a crucial role in the apoptotic pathway and its aberrant expression may lead to cancer. This gene also plays an important role in the defense mechanism against hyperproliferation and tumorigenesis. CASP8 mutation showed that the domain commonly mutated is peptidase_C14 (caspase domain) with the frequency of alteration found in R432*. CASP8, a tumor suppressor, and pro-apoptotic protein are inactivated by mutation or deletion in various cancer types. The CASP8 R432* mutation is likely oncogenic and its biological effect is the loss of function. Qiliu et al. through the meta-analysis study suggested that CASP8 could be a candidate gene for CRC susceptibility. Also, in the late stage of CRC, mutation in this gene may lead to loss of apoptotic function contributing to pathogenesis. The only mutation observed in this gene is missense along with the G-alpha domain. Mutations of this gene have been shown to activate the adenylate cyclase gene and lead to constitutive cAMP signaling. Zauber et al. also reported that GNAS mutations may function as an important driver mutation during certain phases of colorectal carcinogenesis, but may then be lost once the biological advantage gained by the mutated gene is no longer necessary to sustain or advance tumor development.

Based on expression on the target genes, the inactivation of APC is required as an early event to develop most adenomas and carcinomas in the colon and rectum. This gene has been investigated as the most frequently mutated and known driver gene in CRC. Several authors have also reported the expression analysis of this gene to be downregulated in CRC. Birnbaum et al. confirmed that APC status may not be used in the prediction of metastasis and for staging. Schell et al. demonstrated the prognostic role of APC and suggested that this gene may be a good potential therapeutic assignment for CRC. The RAS oncogene has a well-defined role in cell growth and regulation, and its protein product affects many cellular functions including cell proliferation, apoptosis, migration, fate specification, and differentiation. KRAS has been reported to be the most frequently altered gene, with mutations occurring in 17%–25% of all cancers, while its mutations in CRC lead to resistance to select treatment strategies. From this study, the expression KRAS was downregulated in CRC. This gene encodes for a guanosine triphosphate/guanosine diphosphate-binding protein that is downstream of EGFR in the RAS/RAF/MAPK pathway. Studies have also shown that approximately 40% of patients with CRC carry a KRAS mutation.
EGFR, IGF1R, CASP8, and GNAS are important for the diagnosis of cancer subtype. TCF7L2 is an essential element of the Wnt signaling pathway, which acts as a transcriptional factor in the nucleus. Its significance and protein expression in the progression of various malignancies is yet to be exploited. The expression analysis of IGF1R did not show any significant difference between normal colorectal tissue and the cancer subtype. IGF1R has been investigated to be a target by miR-448 in CRC and also, its expression was upregulated in CRC tissues and cell lines, which counter the result obtained by this study. Studies have suggested that mutation in GNAS may play a direct role in mucin production. The EGFR is a transmembrane glycoprotein and receptor tyrosine kinase that is encoded by the c-erbB-1 proto-oncogene. This gene is overexpressed in many types of cancers, specifically CRC. EGFR is estimated to be overexpressed in 60%–80% of tumors and is associated with a poor prognosis. The expression analysis of this gene showed that it was downregulated in CRC. Spano et al. revealed that EGFR remains a controversial prognostic factor, whose 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. Therefore, it could be a potential target for CRC treatment. Intestinal epithelium that possesses one of the highest renewal rates among human tissues, expresses IGF1R in addition to other genes, and the levels of these receptors are higher in CRC relative to colonic mucosa. Limited reports have been recovered from this gene to date in CRC. CASP-8 is one of the crucial factors (intrinsic signaling pathway) employed in apoptosis. The expression of this gene (CASP-8) has been reported to be associated with poor prognosis in stages II and III of CRC. Therefore, their functional polymorphism may alter cancer risk. Although several studies have reported that CASP8 may not contribute to the risk of CRC, this gene deserves a closer look. This information, therefore, suggests novel directions for further research and provide cross-species predictions for effective interaction mapping.

Conclusion

The analysis of the seven microRNA target genes in CRC from our previous study and data from TCGA with combined bioinformatics tools revealed that APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and GNAS were frequently altered in CRC. Furthermore, the expression of these genes was correlated with the genomic status. Genes with protein tyrosine kinase domains were frequently altered in CRC and the most common alteration observed in the domain of APC, KRAS, TCF7L2, EGFR, IGF1R, CASP8, and GNAS is missense mutation. The information provided by the expression, overall survival of the patients with alteration in the expression of the mRNA of the target genes, and their genetic mutation hotspots across protein domains in CRC could contribute to a better understanding of the molecular basis of colorectal carcinogenesis, and ultimately propel the transformation of genomic knowledge into clinical practice. These could be further exploited to potentially serve as a resource for explicitly selecting targets for diagnosis and management of CRC. However, the intense mechanism of these results remains unclear, and further experimental validation and molecular approaches are the focal points in the near future.

Acknowledgements

The authors acknowledge the Plant Omics Laboratory and the Bioinformatics Research Group of the University of the Western Cape for their usual support.

Authors’ contributions

All the authors have made significant contributions to the submission of the article. A.F. conceived the concept and the design of the manuscript, and A.K. and A.P. provided the necessary supports and software required for analysis. The analysis and data interpretations were done by both A.F. and A.P., while A.K. and O.O.B. drafted the rough draft and substantively revised the manuscript. Finally, all authors (A.F., O.O.B., A.K., and A.P.) read and approved the submitted version of the manuscript for publication.

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. Okugawa Y, Grady WM and Goel A. Epigenetic alterations in colorectal cancer: emerging biomarkers. Gastroenterology 2015; 149(5): 1204–1225.e12.
2. Tomasetti C, Marchionni L, Nowak MA, et al. Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc Natl Acad Sci USA 2015; 112: 118–123.
3. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002; 16: 6–21.

4. Garraway LA and Lander ES. Lessons from the cancer genome. *Cell* 2013; 153: 17–37.

5. Willett CG, Chang DT, Czitro BG, et al. Cancer genome atlas network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012; (5). *Int J Radiat Oncol Biol Phys* 2013; 86(1). https://doi.org/10.1016/j.ijrobp.2012.12.006

6. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014; 505: 495.

7. Liao X, Lochhead P, Nishihara R, et al. Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. *New Engl J Med* 2012; 367: 1596–1606.

8. Fadaka AO, Klein A and Pretorius A. In silico identification of microRNAs as candidate colorectal cancer biomarkers. *Tumour Biol* 2019; 41(11): 101042831983721.

9. Friedman RC, Farh Burge CB and Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009; 19(1): 92–105.

10. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007; 27: 91–105.

11. Nam J-W, Rissland OS, Koppstein D, et al. Global analysis of the effect of different cellular contexts on microRNA targeting. *Mol Cell* 2014; 53: 1031–1043.

12. 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: 15–20.

13. Gonzalez-Pons M and Cruz-Correa M. Colorectal cancer biomarkers: where are we now? *Biomed Res Int* 2015; 2015: 149014.

14. Ferracin M, Lupini L, Mangolini A, et al. Circulating non-coding RNA as biomarkers in colorectal cancer. *Adv Exp Med Biol* 2016; 937: 171–181.

15. Fadaka AO, Pretorius A and Klein A. Functional prediction of candidate microRNAs for CRC management using in silico approach. *Int J Mol Sci* 2019; 20: 5190.

16. Fadaka AO, Pretorius A and Klein A. MicroRNA assisted gene regulation in colorectal cancer. *Int J Mol Sci* 2019; 20: 4899.

17. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012; 487: 330.

18. Cerami E, Gao J, Dogrusoz U, et al. The eBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Philadelphia, PA: AACR, 2012.

19. BI-TCGA. Broad Institute TCGA Genome Data Analysis Center Analysis-ready standardized TCGA data from Broad GDAC Firehose 2016_01_28 run. Broad Institute of MIT and Harvard. Dataset 2016.

20. Chen T, Wang Z, Zhou W, et al. Hotspot mutations delineating diverse mutational signatures and biological utilities across cancer types. *BMC Genomics* 2016; 17: 394.

21. Engelman JA. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. *Nat Rev Cancer* 2009; 9(8): 550–562.

22. Samuels Y, Diaz LA Jr, Schmidt-Kittler O, et al. Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* 2005; 7(6): 561–573.

23. Fadaka AO, Pretorius A and Klein A. Identification of MicroRNAs as biomarkers in colorectal cancer diagnostics: an in silico approach. *FASEB J* 2020; 34: 1–1.

24. Bhandari A, Woodhouse M and Gupta S. Colorectal cancer is a leading cause of cancer incidence and mortality among adults younger than 50 years in the USA: a SEER-based analysis with comparison to other young-onset cancers. *J Investig Med* 2017; 65(2): 311–315.

25. 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: 722.

26. Birnbaum DJ, Laibe S, Ferrari A, et al. Expression profiles in stage II colon cancer according to APC gene status. *Transl Oncol* 2012; 5(2): 72–76.

27. Chang MT, Asthana S, Gao SP, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* 2016; 34(2): 155–163.

28. Gao J, Chang MT, Johnsen HC, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Medicine* 2017; 9: 4.

29. Prior IA, Lewis PD and Mattos C. A comprehensive survey of Ras mutations in cancer. *Cancer Res* 2012; 72: 2457–2467.

30. Roa I, Sánchez T, Majlis A, et al. Mutación del gen KRAS en el cáncer de colon y recto. *Rev Méd Chil* 2013; 141: 1166–1172.

31. 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.

32. Tsuchihashi Z, Khambata-Ford S, Hanna N, et al. Responsiveness to cetuximab without mutations in *EGFR*. *New Engl J Med* 2005; 353: 208–209.

33. 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: 133–141.

34. Pollak MN, Schernhammer ES and Hankinson SE. Insulin-like growth factors and neoplasia. *Nat Rev Cancer* 2004; 4: 505–518.

35. Gombos A, Metzger-Filho O, Dal Lago L, et al. Clinical development of insulin-like growth factor receptor-1 (IGF-1R) inhibitors: at the crossroad. *Invest New Drugs* 2012; 30(6): 2433–2442.

36. Hakam A, Yeatman TJ, Lu L, et al. Expression of the epidermal growth factor receptor: current treatment paradigms, the importance of diet, and the role of chemoprevention. *World J Clin Oncol* 2015; 6: 133–141.

37. Shiratsuchi I, Akagi Y, Kawahara A, et al. Expression of IGF-1 and IGF-1R and their relation to clinicopathological factors in colorectal cancer. *Anticancer Res* 2011; 31(7): 2541–2545.

38. Codony-Servat J, Cuatrechas M, Asensio E, et al. Nuclear IGF-1R predicts chemotherapy and targeted
therapy resistance in metastatic colorectal cancer. Br J Cancer 2017; 117: 1777–1786.

40. Sipos F, Székely H, Kis ID, et al. Relation of the IGF/IGF1R system to autophagy in colitis and colorectal cancer. World J Gastroenterol 2017; 23: 8109–8126.

41. Gangwar R, Mandhani A and Mittal RD. Caspase 9 and caspase 8 gene polymorphisms and susceptibility to bladder cancer in north Indian population. Ann Surg Oncol 2009; 16(7): 2028–2034.

42. Qiliu P, Xianjun L, Weizhong T, et al. CASP8 -652 6N Del polymorphism contributes to colorectal cancer susceptibility: evidence from a meta-analysis. PLoS ONE 2014; 9: e87925.

43. Kim HS, Lee JW, Soung YH, et al. Inactivating mutations of caspase-8 gene in colorectal carcinomas. Gastroenterology 2003; 125(3): 708–715.

44. Zauber P, Marotta SP and Sabbath-Solitare M. GNAS Del polymorphism contributes to colorectal cancer susceptibility: evidence from a meta-analysis. Br J Cancer 2013; 108: 951–958.

45. Laburthe M, Auger C, Rouyer-Fessard C, et al. Functional VIP receptors in the human mucus-secreting colonic epithelial cell line CL.16E. Am J Physiol 1989; 256(3 Pt. 1): G443–G450.

46. Yarden Y and Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2001; 2(2): 127–137.

47. Cohen RB. Epidermal growth factor receptor as a therapeutic target in colorectal cancer. Clin Colorectal Cancer 2003; 2: 246–251.

48. 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.

49. 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: 165.

50. Pollak M. Insulin and insulin-like growth factor signalling in neoplasia. Nat Rev Cancer 2000; 256: 58–66.

51. Saadeddin A, Babaei-Jadidi R, Spencer-Dene B, et al. The links between transcription, β-catenin/JNK signalling, and carcinogenesis. Mol Carcinog 2009; 7: 1189–1196.

52. Ishiguro H, Wakanuki T, Terasita Y, et al. Nuclear expression of TCF4/TCF7L2 is correlated with poor prognosis in patients with esophageal squamous cell carcinoma. Cell Mol Biol Lett 2016; 21: 5.

53. Li B, Ge L, Li M, et al. miR-448 suppresses proliferation and invasion by regulating IGF1R in colorectal cancer cells. Am J Transl Res 2016; 8: 3013–3022.