



UNIVERSIDADE ESTADUAL PAULISTA

“JÚLIO DE MESQUITA FILHO”

FACULDADE DE MEDICINA

Iael Weissberg Minutentag

Identificação de novos RNAs não codificadores (ncRNAs) associados com a carcinogênese colorretal

Tese apresentada à Faculdade de Medicina, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de Botucatu, para obtenção do título de Doutora em Cirurgia e Medicina Translacional.

Orientadora: Profa. Dra. Sandra Drigo Linde

Botucatu

2023

Iael Weissberg Minutentag

Identificação de novos RNAs não codificadores
(ncRNAs) associados com a carcinogênese
colorretal

Tese apresentada à Faculdade de
Medicina, Universidade Estadual Paulista
“Júlio de Mesquita Filho”, Câmpus de
Botucatu, para obtenção do título de
Doutora em Cirurgia e Medicina
Translacional

Orientadora: Profa. Dra. Sandra Drigo Linde

Botucatu

2023

M668i Minutentag, Iael Weissberg
Identificação de novos RNAs não codificadores
(ncRNAs) associados com a carcinogênese colorretal / Iael
Weissberg Minutentag. -- Botucatu, 2023
75 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista
(Unesp), Faculdade de Medicina, Botucatu
Orientadora: Sandra Drigo Linde

1. Câncer colorretal. 2. Genética molecular. 3.
MicroRNA. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Medicina, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Agradecimentos

À minha orientadora, Dra. Sandra A. Drigo Linde, pela dedicação, incentivo, organização, apoio, e suporte incondicional na realização desse trabalho. Por me tranquilizar quando parecia que nada ia dar certo. Por sempre fazer tudo prezando pela qualidade. E por ter me proporcionado aprendizado que vai além do laboratório.

À Dra. Patricia Reis, pela ajuda, pelos conselhos e ideias preciosas.

Ao Dr. Marcio de Carvalho, técnico do laboratório que sempre vai além de todas as expectativas e não mede esforços para nos ajudar a realizar nossos experimentos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa recebida e à Faculdade de Medicina de Botucatu e Unidade de Pesquisa Experimental (UNIPLEX), pela estrutura disponibilizada para a concretização desse trabalho

Aos meus amigos do laboratório Oncologia Molecular e Translacional e da UNIPLEX que participaram junto comigo do passo a passo da realização desse estudo.

Ao Departamento de Patologia da Faculdade de Medicina de Botucatu, em especial ao patologista Marcelo Padovani pela colaboração, dedicação e paciência para revisar todas as lâminas utilizadas nessa pesquisa.

Aos colaboradores desse trabalho Dr. Matheus Barros-Filho, Dr. Wan L. Lam e Dr. Fabio Marchi cuja ajuda foi fundamental para conclusão deste trabalho.

À minha querida amiga Michelle Paz com quem sempre tenho as melhores conversas e debates seja sobre ciência, política ou qualquer outro tema do universo, e está sempre disposta a me ajudar a organizar minhas ideias, por mais confusas que elas possam parecer.

A todos os funcionários da Biblioteca e Seção Técnica de Pós-Graduação, em especial a Márcia Fonseca Piagentini Cruz secretária do programa de Pós-Graduação em Bases Gerais da Cirurgia.

A todos que contribuíram muito ou pouco para a realização desse trabalho

Muito Obrigada!

Dedicatória

Aos pacientes, que aceitaram contribuir para esse estudo, fornecendo os seus dados e amostras, sem eles esse trabalho não teria sido possível.

À todas as pessoas que contribuíram direta ou indiretamente para a minha formação, todas as minhas professoras e professores de ciências e biologia que despertaram meu fascínio por entender como o mundo funciona.

À minha mãe que sempre respondeu todos os meus “porquês”, que criou minha paixão pela leitura todas as noites antes de dormir, que é minha amiga e companheira em todos os momentos, que nunca duvidou do meu potencial e sempre me incentivou a correr atrás dos meus sonhos.

À minha segunda mãe, minha vó, baba, não tenho palavras para descrever o quanto eu queria que você estivesse aqui do meu lado vendo finalmente eu me tornar “pressefora dotora”, sempre vou carregar comigo as memórias que construímos juntas, quando você lia todas as placas da rua quando saíamos para passear de carro, e sempre tinha um provérbio ou conselho pronto, você é uma parte fundamental de quem eu sou, espero que, onde quer que você esteja, você esteja sorrindo para mim .

Ao meu pai que sempre ofereceu o melhor conselho “Ficar triste ou nervosa não vai resolver seus problemas, levante e vai encontrar uma solução” e está sempre disposto a me ouvir reclamar simplesmente porque estou precisando.

Ao meu tio Daniel que pouco a pouco, sem que eu percebesse, me ensinou filosofia e pensamento científico, através de conversas e discussões que na época me pareciam confusas e só mais tarde fui entender a importância.

À minha tia Nancy que estava sempre disposta a me fazer contemplar novas possibilidades e caminhos

E por fim, dedico esse trabalho ao meu avô Marcos, que apesar de não estar mais entre nós, foi uma das pessoas que me inspirou a seguir na pesquisa em oncologia molecular para tentar melhorar a vida dos pacientes de câncer e encontrar novas soluções em terapia e diagnóstico.

“Ainda mais mistérios do universo permanecem ocultos. Sua descoberta espera os cientistas aventureiros do futuro. Eu gosto que seja assim”

Vera Rubin

“Como pesquisadora, os momentos inesquecíveis de minha vida são raros... Quando o véu que cobre os segredos da natureza parece erguer-se repentinamente”

Gerty Cori

Sumário

Resumo.....	1
Abstract	2
Prefácio	1
<i>Capítulo I</i>	2
Introdução.....	3
Hipótese	26
Objetivo	27
Materiais e Métodos	27
Referências.....	36
<i>Capítulo II</i>	45

Resumo

Os cânceres colorretais (CCRs) proximal e distal apresentam características moleculares e prognósticas diferentes. Os microRNAs (miRNAs) desempenham um papel relevante no câncer, sendo também associados com a carcinogênese colorretal. Novos miRNAs tecido-específicos e ainda não anotados foram relatados em diferentes tumores, indicando que novos mecanismos reguladores específicos de tecido podem contribuir para a carcinogênese. Assim, neste estudo investigamos a expressão de novos miRNAs em CCR, comparando CCR proximal e distal. Dados brutos de small-RNA-Seq foram obtidos de repositórios públicos e avaliados pela plataforma mirMaster para identificação dos novos miRNAs. Dados de 522 CRC e 11 tecidos normais do TCGA (grupo de descoberta), e 35 CRC e 20 normais do GEO (GSE89974, grupo de validação) foram analisados. Especificidade tecidual foi avaliada comparando os novos miRs identificados nos tecidos normais de cólon com os miRs identificados pela plataforma miRMaster a partir de dados de small-RNAseq de 5 tecidos normais do TCGA: bexiga, fígado, cérebro, mama e pulmão. Genes alvos dos novos miRNAs foram identificados pela ferramenta miRDB e seus padrões de expressão foram avaliados pelo UCSC Xena Browser. Após sucessivas análises e filtragens identificamos 15 novos miRs candidatos expressos diferencialmente entre CRC e amostras normais, detectados em ambos os grupos de descoberta e validação. Caracterização molecular dos novos miRs mostrou similaridade com miRs conhecidos tais como, conteúdo de GC, tamanho, localização genômica em regiões contendo miRs conhecidos e em regiões de sítios frágeis cromossômicos. Quatro novos miRNAs se mostraram expressos diferencialmente entre CRC proximal e distal, sendo 3 novos miRNAs com níveis mais altos nos (miR-13844-5p, nov-miR-7154-5p e nov-miR-590-5p) e 1 (nov-miR-5035-3p) com níveis diminuídos nos tumores distais comparados aos proximais. Na análise de especificidade tecidual 2 novos miRs (nov-miR-3345-5p e nov-miR-13172-3p) foram identificados apenas nos tecidos colorretais. Níveis mais altos de nov-miR-3345-5p e nov-miR-7154-5p foram associados com menor sobrevida em pacientes com tumores do cólon direito e reto, respectivamente. Análise de predição de alvos dos 15 novos miRNAs revelou 2.412 genes. Análise de correlação dos pares miRNA-gene alvo identificou correlação negativa significativa entre 8 novos miRNAs (6 com expressão aumentada e 2 diminuídos) e 81 genes alvos. A seguir, 47 dos 81 genes foram identificados com algum tipo de interação gênica, sendo esses genes regulados por 7 miRNAs candidatos. Os genes *CACNA1B* e *DLG2* mostraram o maior número de interações com outros genes sendo regulados pelo nov-miR-8861-5p e nov-miR-13172-3p, respectivamente. O nov-miR-8861-5p aparece como o miRNA regulando o maior número de genes (32 genes alvos) com interações entre si. Em resumo, identificamos 15 novos miRs expressos em tecidos de colorretal, sendo que três candidatos são expressos exclusivamente em tecidos colorretais. Além disso, a detecção da expressão diferencial de novos miRs e seus genes alvos entre tumores de cólon direito, cólon esquerdo e reto sugere alterações específicas de cada localização, que podem estar associadas com as diferenças clínicas e de progressão desses tumores. Esses resultados indicam a potencial

relevância desses novos miRs para a biologia dos tumores colorretais do CCR e possíveis aplicações futuras como biomarcadores em CRC e alvos terapêuticos.

Palavras-Chave: novos MicroRNAs, Adenocarcinoma Colorretal, Transcritoma, Lateralidade Tumoral

Abstract

Proximal and distal colorectal cancer (CRCs) have different molecular and prognostic characteristics. MicroRNAs (miRNAs) play an important role in cancer and are associated with colorectal carcinogenesis. Novel tissue-specific miRNAs have been reported in different tumors, indicating that new tissue-specific regulatory mechanisms may contribute to carcinogenesis. Therefore, in this study, we investigated the expression of novel miRNAs in CRC by comparing proximal and distal CRC. Raw small-RNA-Seq data were obtained from public repositories and evaluated using the mirMaster platform to identify novel miRNAs in CRC tissues. Data from 522 CRC and 11 normal tissues from TCGA (discovery group) and 35 CRC and 20 normal tissues from GEO (GSE89974, validation group) were analyzed. Tissue specificity was evaluated by comparing the new miRs identified in normal colon tissues with the miRs identified by the miRMaster platform from small-RNAseq data from five normal tissues from TCGA: bladder, liver, brain, breast, and lung. Target genes of the novel miRNAs were detected using the miRDB tool and their expression patterns were evaluated using the UCSC Xena Browser. After successive analysis and filtering, we identified 15 new candidate miRNAs that were differentially expressed between CRC and normal samples, detected in both the discovery and validation groups. Molecular characterization of the novel miRs showed similarities with known miRs, such as GC content, size, and genomic location in regions containing known miRs and in regions of chromosomal fragile sites. Four new miRNAs were differentially expressed between proximal and distal CRC, with three presenting higher levels (miR-13844-5p, nov-miR-7154-5p, and nov-miR-590-5p) and one (nov-miR-5035-3p) decreased levels in distal tumors compared to proximal tumors. In the tissue specificity analysis, two novel miRNAs (nov-miR-3345-5p and nov-miR-13172-3p) were identified only in colorectal tissues. Higher levels of nov-miR-3345-5p and nov-miR-7154-5p were associated with shorter survival in patients with right-sided colon and rectal tumors, respectively. Target prediction of the 15 candidate miRNAs revealed 2,412 genes. Correlation analysis of the miRNA-target gene pairs showed a significant negative correlation between eight novel miRNAs (six upregulated and two downregulated) and 81 target genes. Next, 47/81 target genes regulated by seven novel miRNAs showed genetic interactions. *CACNA1B* and *DLG2*, regulated by nov-miR-8861-5p and nov-miR-13172-3p, respectively, showed the highest number of interactions with other genes. The nov-miR-8861-5p had the highest number of genes (32 target genes). In summary, we identified 15 novel miRNAs expressed in colorectal tissues, with three candidate miRNAs expressed exclusively in colorectal tissues. In addition, the detection of novel miRNAs and their target genes differentially expressed between right-

sided colon, left-sided colon, and rectal tumors suggests anatomic-specific changes that might be associated with the clinical and progression differences of these tumors. These results indicate the potential relevance of these novel miRNAs in colorectal tumor biology and their possible future applications as CRC biomarkers and therapeutic targets.

.

Keywords: Novel microRNAs, Colorectal Adenocarcinoma, Transcriptome, Anatomical Location

1 Prefácio

2 Esta tese de doutorado está apresentada em dois capítulos:

3 1º Capítulo. Neste capítulo será apresentada uma revisão da literatura, a hipótese do estudo,
4 os objetivos e material e métodos detalhados, seguido das referências bibliográficas no padrão ABNT

5 2º Capítulo. Manuscrito publicado com os principais resultados do estudo
6 - <https://doi.org/10.3390/ncrna9060065>. Com sua própria lista de referências no padrão da revista

7 Non-Coding RNA

8

9

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

Capítulo I

Introdução

1. Epidemiologia E Fatores De Risco Do Câncer Colorretal

O câncer é a segunda causa de morte no mundo, ficando atrás apenas das doenças cardiovasculares, sendo o câncer colorretal (CCR) o terceiro tumor mais comum e a segunda causa de morte por câncer. (SUNG et al., 2021). No Brasil, o CCR é o segundo câncer mais incidente tanto em homens quanto em mulheres, sendo estimados para o ano de 2023 45.630 novos casos em todo o território nacional (INCA; INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA - INCA, 2022). Cerca de 20% dos pacientes apresentam metástases hepáticas sincrônicas ao diagnóstico (CHAN et al., 2022)

O CCR é considerado um dos marcadores mais claros de transição econômica em países ocidentais. Na última década houve um aumento da incidência e da mortalidade em países com médio a alto índice de desenvolvimento humano (IDH), particularmente na Europa Ocidental, Ásia e América do Sul (ARNOLD et al., 2017).

Uma tendencia preocupante é o aumento de casos de CCR em paciente com menos de 50 anos. Desde o início do monitoramento dos casos de câncer na década de 90, os casos de CCR em pacientes jovens de 20 a 49 anos dobrou, nos Estados Unidos os casos em pacientes jovens já representam 13.1 a cada 100.000 casos com resultados similares observados no Canadá, Austrália, Reino Unido, Japão, Hong Kong e Shanghai (Revisado em STOFFEL; MURPHY, 2020). Baseando-se na atual tendencia de crescimento dos casos de CCR estima-se que em 2030 a incidência de câncer de cólon e de reto terá aumentado em 90% e 124,2% para pacientes de 20 a 34 e em 27,7% e 46% para pacientes de 35 a 49 anos, respectivamente (BAILEY et al., 2015) .

Por muito tempo o câncer de cólon (CC) e o câncer de reto (CR) foram considerados como um único tipo tumoral devido a concepção errônea que o intestino grosso é igual em toda sua extensão. Características importantes como diferentes funções desempenhadas no trato intestinal como concentração do bolo fecal, reabsorção de fluídos, transporte das fezes e excreção não eram consideradas. Além disso, não só a incidência de CC e RC são diferentes (2,5:1, respectivamente), como diferenças nos fatores de risco entre os dois tumores já são conhecidas (Revisado em PASCHKE et al., 2018)

Os principais fatores de risco para os tumores de cólon proximal e transversal são ligados ao estilo de vida como sedentarismo, obesidade e consumo da dieta ocidental

caracterizada pelo alto consumo de carne vermelha e carnes processados, além de alimentos gordurosos e industrializados (Fig. 1), porém esses fatores não parecem afetar o risco de CR (Revisado em PASCHKE et al., 2018). Em contraste, os tumores de cólon distal e reto estão mais ligados a tabagismo, alcoolismo, susceptibilidade genética e fatores ambientais como ingestão de água contaminada por produtos agrícolas como compostos organofosforados, cloro (DENG, 2017) e metais pesados como mercúrio, cromo, bismuto e arsênico (Fig. 1) (ENTERLINE; DAY; MARSH, 1995; Revisado em JUNG; KIM; KIM, 2022; KIM et al., 2020; LEROYER et al., 2022).

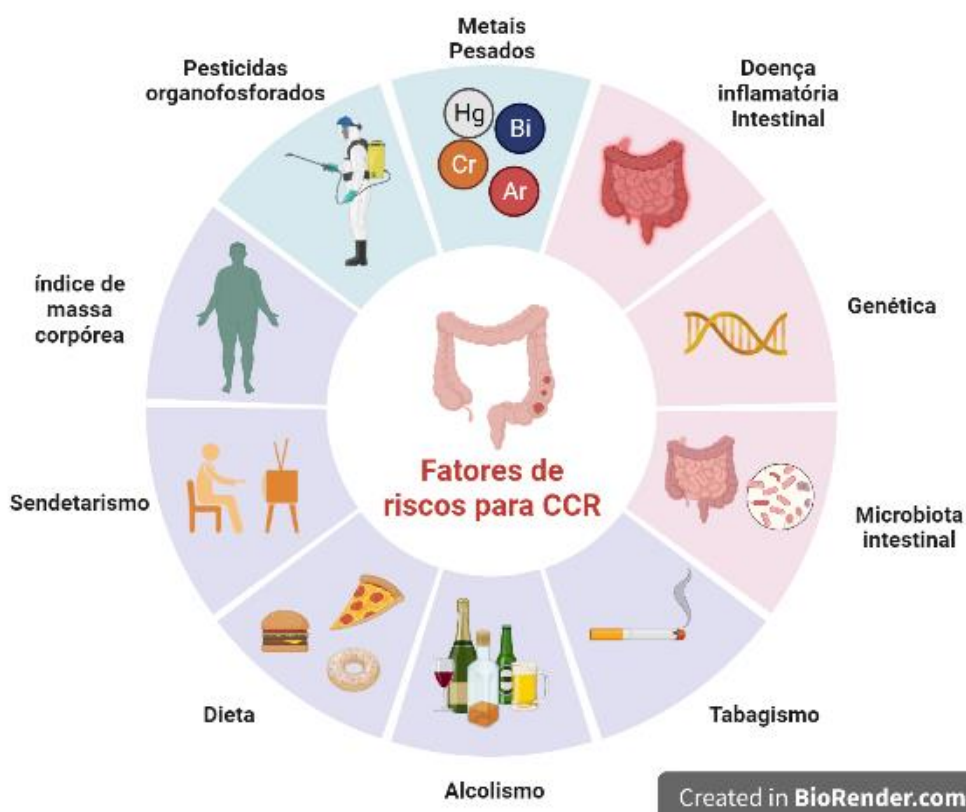


Figura 1. Fatores de risco de risco para CCR, em roxo estão os fatores relacionados ao estilo de vida, em verde os fatores ambientais e rosa fatores genéticos e intrínsecos do indivíduo

Esse dado é especialmente relevante para a população brasileira devido aos recentes desastres ambientais criados pelo rompimento das barragens de rejeitos de mineração nas cidades de Mariana-MG e Brumadinho-MG. Macêdo *et. al.* (2020) detectou a presença de cádmio, alumínio, ferro, manganês e mercúrio em amostras de água da região de Mariana-MG com concentrações até 9,1 vezes mais altas após o rompimento do que as encontrados antes do rompimento da barragem. Esses achados foram relacionados com alterações

morfológicas nas guelras e fígados de peixes da região. Na região de Brumadinho-MG foram analisadas amostras de urina e sangue da população de moradores locais e foram encontrados níveis aumentados de manganês, arsênio, chumbo, mercúrio e cádmio. Essas alterações podem levar a um maior risco de desenvolvimento de tumores nas populações afetadas (MOTA et al., 2022).

3. Microbioma no Câncer Colorretal

Vários fatores genéticos e epigenéticos foram listados como importantes na carcinogênese colorretal e seu impacto na heterogeneidade observada nesses tumores (Fig. 1), principalmente no que se refere às diferenças observadas associadas a lateralidade tumoral.

Um fator que vem se mostrando cada vez mais importante na tumorigênese e progressão dos CCRs é a microbiota intestinal do paciente que já foi apontada como responsável por diversas alterações no perfil transcritômico dos tumores, inclusive já tendo sido listada como um novo *hallmark* da carcinogênese (HANAHAN, 2022; SIDDIQUI et al., 2022). A microbiota intestinal é composta por cinco principais filos *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria* e *Verrucomicrobia* (Revisado em FEHILY et al., 2021) e já foi descrita como sendo composta por 10^{13} a 10^{14} bactérias, um número quase igual ao número de células animais no corpo (GILL et al., 2006). Muitos fatores podem afetar a microbiota intestinal como idade, região geográfica, hábitos alimentares, genética do paciente e pressão do sistema imune. Por outro lado, a microbiota intestinal pode afetar a fisiologia do hospedeiro de diversas maneiras (Revisado em FIDELLE et al., 2020).

No contexto do CCR já foi identificado enriquecimento dos gêneros *Fusobacterium*, *Porphyromonas*, *Peptostreptococcus* e *Prevotella* assim como alguns fungos. Esses achados foram relacionados com alterações na produção de bile e na degradação de aminoácidos (WIRBEL et al., 2019), porém a contribuição dessas alterações na carcinogênese ainda está sendo investigada. Em particular, o envolvimento da *Fusobacterium nucleatum* que emerge após a iniciação da carcinogênese já foi associada com CCRs proximais avançados, metástase, quimioresistência e prognóstico ruim (Revisado em FIDELLE et al., 2020).

3. Histopatologia e Aspecto Morfológicos do Câncer de Cólon e de Reto

Os CCR são tumores bastantes heterogêneos. Nos últimos anos, uma grande quantidade de evidências tem indicado que os CCR proximais são entidades biológicas distintas dos tumores distais não só por terem origens embriológicas distintas, perfil imunes característicos e alterações genéticas mais frequentes, mas também por apresentarem desfechos clínicos diferentes (YANG et al., 2022).

O cólon direito (cécum até a flexura esplênica) é derivado do intestino médio embrionário que forma o duodeno distal, jejuno, íleo, cécum, cólon ascendente e aproximadamente 2/3 do cólon transverso. Já o cólon esquerdo (flexura esplênica até o cólon sigmoide) é formado pelo intestino embrionário posterior que forma o terço distal do cólon transverso, o cólon descendente, o cólon sigmóide e 2/3 do canal ano-retal cuja porção final é formada pela membrana cloacal. Devido a essas diferentes origens o cólon proximal recebe suprimento sanguíneo da artéria mesentérica superior e o cólon distal da artéria mesentérica inferior (YANG; CHO; KIM, 2018). Como é de se esperar os tumores com origens nos diferentes sítios anatômicos também apresentam diferentes características clínicas e moleculares, começando pela incidência tumoral em cada sítio, cerca de 40% são tumores de cólon direito (CCD), 20% são tumores de cólon esquerdo (CCE) e 30% são tumores de reto (CR), 10% outros sítios (PASCHKE et al., 2018).

Os CCDs já foram associados com risco de morte significativamente maior em tumores com estadiamento III e IV, (Revisado em LEE; MENTER; KOPETZ, 2017). Esses tumores são mais comuns em pacientes mais velhos e do sexo feminino, frequentemente apresentando um crescimento tumoral local maior, metástases nos linfonodos regionais e tumores pouco diferenciados com maior frequência dos subtipos histológicos mucinoso e de células de anel de sinete (BENEDIX et al., 2010)

Os CCEs são tumores comumente diagnosticados em colonoscopias devido a sangramentos retais e mudanças nos hábitos intestinais ou ainda em cirurgias de emergência devido a complicações geradas pelo menor diâmetro da luz do cólon distal sendo ocupado pela massa tumoral; isso está diretamente relacionado com fato que pacientes com CCEs frequentemente já chegam ao serviço oncológico com diagnóstico confirmado por histologia e tumores em estadiamento inicial (MO YANG et al., 2019; PETRELLI et al., 2017)

No câncer inicial, histologicamente lesões da mucosa do tipo polipoide são mais frequentemente observadas no cólon do que no reto (51% CCD, 35% CCE e 14% RC) enquanto lesões da submucosa são mais frequentes no reto assim como lesões das mucosas com componentes vilosos (PASCHKE et al., 2018).

Existem 4 principais subtipos morfológicos de CCR, adenocarcinoma típico, adenocarcinoma mucinoso, carcinoma de células de anel de sinete, adenocarcinoma serrilhado. (JOHNCILLA; YANTISS, 2020)

Cerca de 85% dos CCR evoluem a partir de adenomas de alto grau e são adenocarcinomas típicos, esse tipo tumoral não tem preferência com relação a lateralidade tumoral e são ricos em mutações em *KRAS* (TAMAS et al., 2015)

Os adenocarcinomas mucinosos representam cerca de 10% do CCR, são caracterizados por terem um volume tumoral composto de pelo menos 50% de mucina, são carcinomas agressivos que frequentemente apresentam deficiência nos genes de reparo a danos ao DNA e instabilidade de microssatélites e são mais comuns no cólon proximal. (JOHNCILLA; YANTISS, 2020; TAMAS et al., 2015)

Da mesma forma o carcinoma de células de anel de sinete também é mais comum no cólon proximal, representa cerca de 1% dos CCRs e é mais encontrado em pacientes com mais de 40 anos e doença inflamatória intestinal. Aproximadamente 1/3 desses tumores tem deficiência nos genes de reparo a danos ao DNA e são associados com um pior desfecho clínico (JOHNCILLA; YANTISS, 2020)

O subtipo adenocarcinoma serrilhado é observado em 8% dos casos de CCR, ele é mais comum no cólon distal e reto; aproximadamente 20% tem mutação em *BRAF*, fenótipo metilador das ilhas CpG e deficiência nos genes de reparo a danos ao DNA. (JOHNCILLA; YANTISS, 2020).

4. Características Moleculares dos Tumores Cólon

O CCR é uma doença de progressão lenta, o modelo mais aceito para progressão do CCR é o modelo *multi-hit* onde o CCR seria derivado de pólipos intestinais benignos que ao longo do tempo acumulam mutações que levam a perda de função de genes supressores tumorais como *APC* e *TP53* e ativação de oncogenes como *KRAS* o que leva a malignização da

lesão e transformação do pólipo em um adenocarcinoma invasivo (Fig 2) (FEARON; VOGELSTEIN, 1990; LI et al., 2014; WANG; FAKIH, 2021).

Outro fator que diferencia os CCDs e CCEs é o perfil molecular. Três principais vias biológicas estão presentes nos CCRs, instabilidade cromossomal (CIN) (PINO; CHUNG, 2010), instabilidade de microssatélites (MSI) (BATTAGLIN et al., 2018) e fenótipo metilador das ilhas CpG (CIMP) (MOJARAD et al., 2013). Essas e outras características como mutações em *BRAF*, *KRAS*, *p53*, *APC* e *EGFR* também estão distribuídas em diferentes proporções nos tumores ao longo do trato intestinal (LOREE et al., 2018).

O consórcio *The Cancer Genome Atlas* (TCGA) reportou a caracterização molecular de 276 casos CCR, identificando dois grupos de tumores baseados na taxa de mutação. Os tumores hipermutados (16% dos casos), sendo maioria CCD, apresentavam mutações frequentes em *ACVR2A*, *APC*, *BRAF (V600E)*, *TGFBR2*, *MSH3*, *MSH6*, *SLC9A9* e *TCF7L2*; além de MSI, com hipermetilação do gene *MLH1*. Os tumores não-hipermutados (84% dos casos) apresentavam mutações em *APC*, *TP53*, *KRAS*, *PIK3CA*, *FBXW7*, *SMAD4*, *TCF7L2* e *NRAS*, sendo que os perfis de mutações eram semelhantes entre cólon e reto (THE CANCER GENOME ATLAS NETWORK, 2012) (Fig 3).

Em 2015, Guinney. *et al.* caracterizaram a partir de 18 conjuntos de dados públicos e privados, 4 subtipos moleculares (*consensus molecular subtype – CMS*) do CCR que reuniam as principais características moleculares presentes nesses tumores e em 2019 esses subtipos foram caracterizados de acordo com as características das populações imune infiltrantes de seus tumores tornando essa classificação ainda mais robusta (SOLDEVILLA et al., 2019).

O subtipo 1 (*CMS1*), também chamado de MSI imune, é encontrado com mais frequência no cólon direito e associado com melhor prognóstico. Esses tumores são hipermutados, comumente apresentando mutação em *BRAF* e *KRAS*. Também são hipermetilados com CIMP, com alta MSI e expressão de proteínas de reparo do DNA, e pouca prevalência de alterações somáticas no número de cópias. Apresentam forte ativação da resposta imune, com infiltração predominante de células T *helper* 1 e T citotóxicas com a ativação de diversas vias de escape imune como *PD-1*, *PD-L1*, *CTLA-4*, *LAG-3* e *IDO* (Chan & Siriwardena, 2022; Llosa et al., 2015).

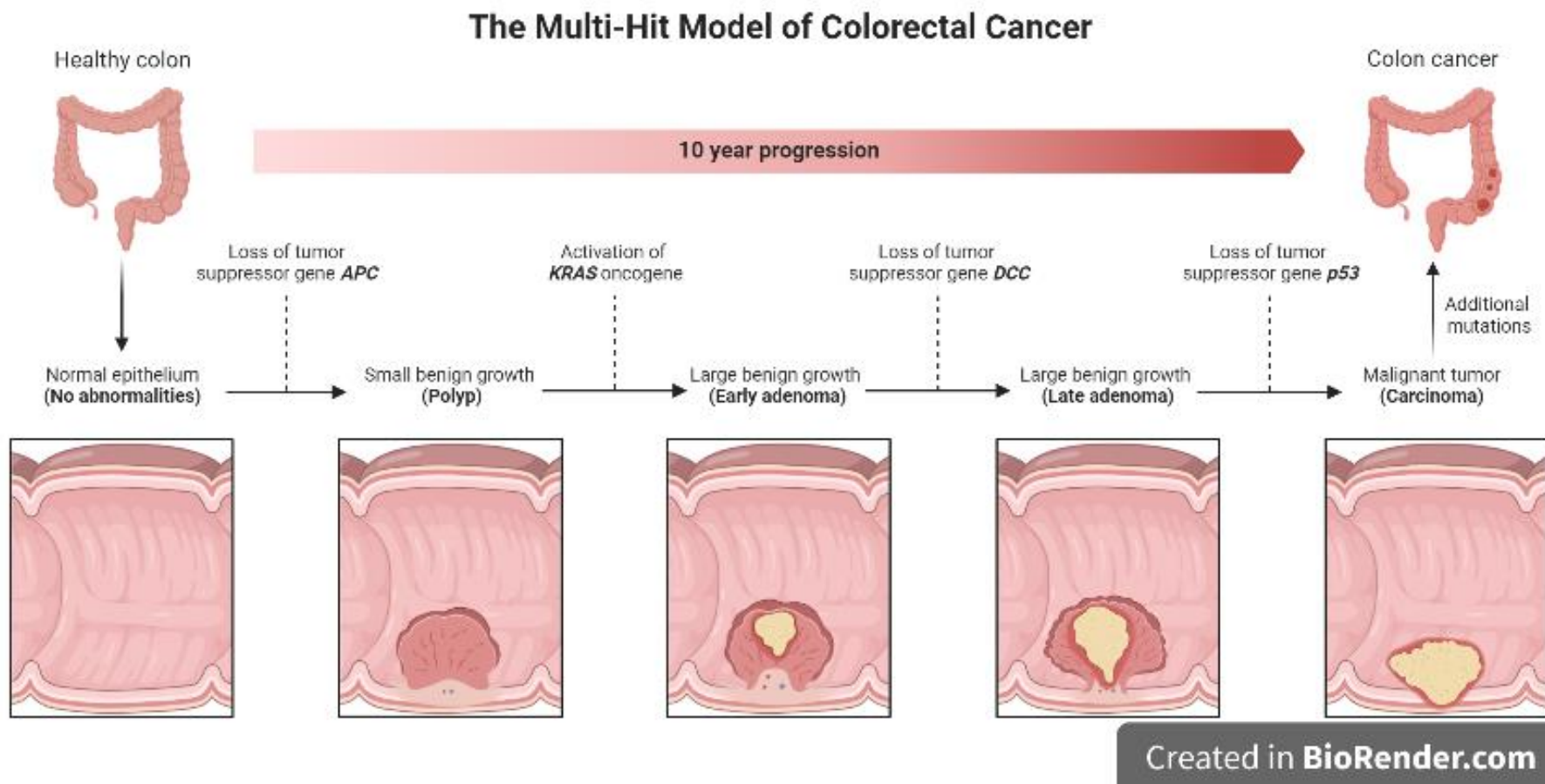


Figura 2. Modelo *Multi-Hit* de progressão do Câncer Colorretal ilustrando a progressão de um pólipó intestinal benigno para um adenocarcinoma invasivo

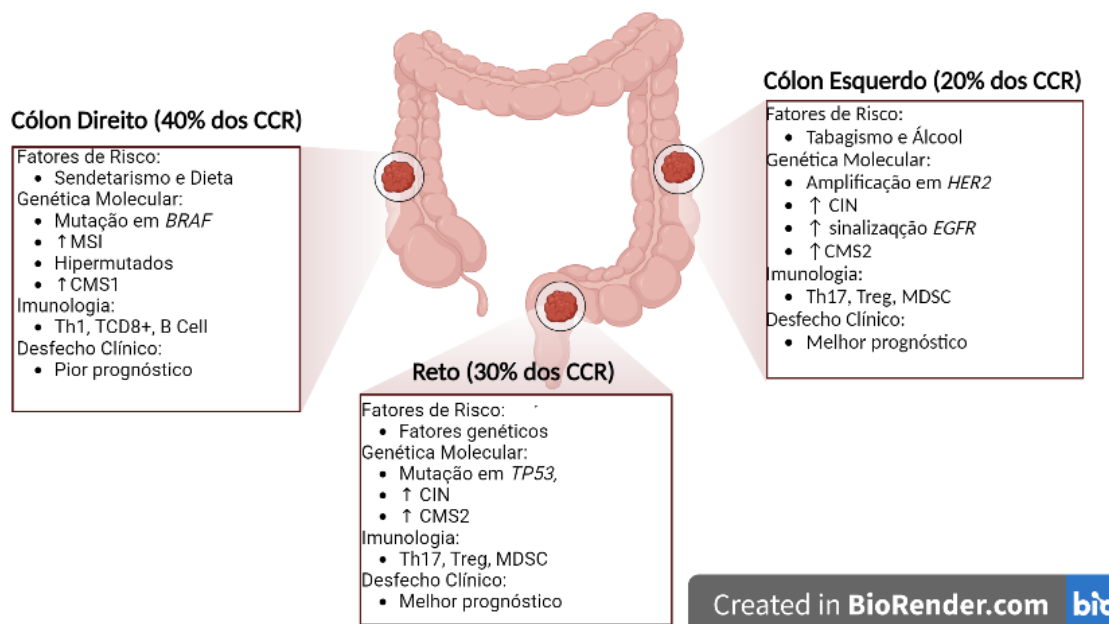


Figura 3. Resumo das diferentes características dos tumores de cólon esquerdo, cólon direito e reto.

O subtipo 2 (CMS2), chamado de canônico, é mais presente nos tumores do cólon distal. Nesse subtipo é possível observar amplificação do oncogene *HNF4A* que atua no desenvolvimento do intestino e na regulação da proliferação celular através da via WNT/ β -catenina (Ni et al., 2021), que também é fortemente ativada nesse subtipo tumoral. Os pacientes com esse subtipo tumoral mostram maior sobrevida pós recorrência (Guinney et al., 2015). Esse subtipo juntamente o subtipo 3 já foram descritos como imunologicamente “frios”, ou seja, com pouca ou nenhuma infiltração linfóide e mieloide e baixa expressão de genes do complexo MHC (SOLDEVILLA et al., 2019).

O outro subtipo considerado imunologicamente “frio” é o subtipo 3 (CMS3), também chamado de metabólico. Esse subtipo tumoral é mais raro, sendo encontrado em apenas 13% dos CCR e é caracterizado por apresentar desregulação de diversas vias metabólicas como a via MAPK e alta prevalência de mutações em *KRAS*, apresentando um perfil genômico semelhante ao subtipo estável descrito para cânceres gástricos (BASS et al., 2014; GUINNEY et al., 2015).

Por fim, o subtipo 4 (CMS4) são os tumores com pior prognóstico. Sua prevalência tem um discreto aumento no cólon distal, porém é estável ao longo do trato intestinal (LOREE et al., 2018). Esses tumores são chamados de mesenquimais por apresentarem aumento de expressão dos genes relacionados a transição epitélio-mesênquimal (TEM), juntamente com

uma forte associação a ativação de TGF- β , angiogênese e vias de remodelação da matriz extracelular(GUINNEY et al., 2015). O perfil imune deste subtipo é inflamatório com alta expressão de marcadores de linfócitos, monócitos e densidade de fibroblastos, sendo considerados como tumores que, potencialmente, poderiam ser beneficiados por imunoterapia (SVEEN et al., 2018) (Fig 3).

5. RNAs Não Codificadores

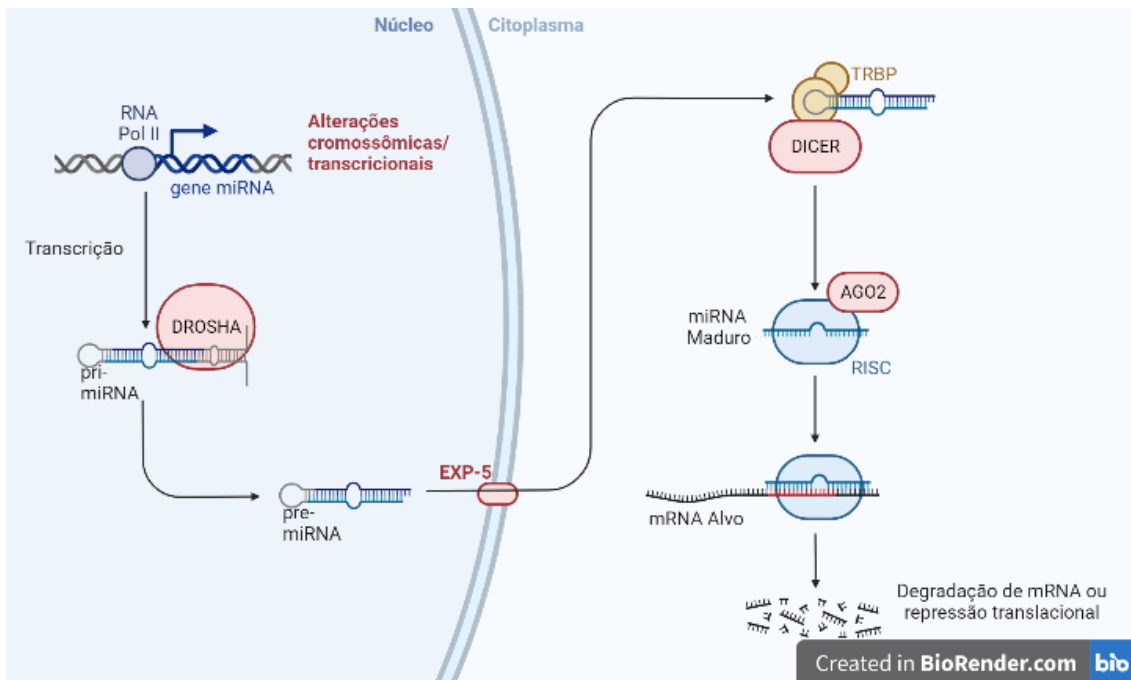
Historicamente, o RNA, com exceção do RNA ribossomal e transportador, foi considerado uma ponte entre o DNA e as proteínas e restrito a função codificadora, porém o nosso transcrito é composto por apenas 2% de RNA codificador, sendo que os 98% restantes são os RNA não codificadores (ncRNAs) que exercem diversas funções na regulação pós-transcricional (PALAZZO; LEE, 2015).

Os ncRNAs são categorizados de acordo com o número de bases de sua sequência: Os pequenos ncRNAs incluem os microRNAs (miRNAs), tsRNAs e piRNAs. Por outro lado, os RNAs não traduzidos com mais de 200 nt de comprimento são os RNAs longos não codificadores (lncRNAs), e incluem subclasses como pseudogenes e circRNAs. Os lncRNAs exercem diversas funções como alterar a arquitetura da cromatina regulando a expressão dos genes relacionados, modular a atividade de promotores e reguladores alterando a transcrição gênica, podem também se ligar as proteínas ligantes do RNA alterando o *splicing* do RNA mensageiro, atuar como “esponjas” capturando outros ncRNAs por complementariedade de base impedindo que esse exerçam sua função regulatória, além de poderem ser reprocessados em outras classes de ncRNA com sequências menores.(LULLI et al., 2022)

Os miRNAs ou miRs são o foco principal do presente trabalho. Os miRNAs são moléculas de RNAs não codificadores cujas sequências têm aproximadamente 23nt, são associados com silenciamento epigenético e altamente conservados evolutivamente, atuando em vários processos celulares como crescimento, diferenciação, desenvolvimento e apoptose (SALIMINEJAD et al., 2019a).

Os miRNAs maduros são sequências de fita simples, porém durante o seu processamento eles são excisados de moléculas precursoras de fita dupla com estrutura de *stem-loop*. De acordo com o banco de dados miRBase v22, 1.917 sequências de miRNAs precursores e 2654 miRs maduros identificados na espécie humana(KOZOMARA;

BIRGAOANU; GRIFFITHS-JONES, 2019). A maioria dos miRNAs são transcritos pela RNA polimerase II a partir do DNA em transcritos primários pri-miRNA, que são clivados pela RNase III (Drosha) e a *double-stranded RNA-binding protein* (DGCR8) em precursores pre-miRNA de aproximadamente 70 nucleotídeos e estrutura de fita dupla com formato de *loop*. O *stem-loop* é reconhecido pela proteína de transporte nuclear exportina-5 (Exp5) e transportado para o citoplasma, onde o pré-miRNA é reconhecido e capturado pelos domínios de RNase III da enzima DICER e da proteína de ligação TRBP que vão remover a estrutura de *loop*, criando um precursor de fita dupla que, por fim, são separadas em duas fitas funcionais denominadas 5' e 3' que se ligam a proteína argonauta (AGO) formando o complexo de silenciamento induzido por RNA (RISC) (Fig. 4) (Revisado em XIONG et al., 2022). O complexo RISC irá então se ligar, por complementariedade de bases parcial (em humanos), à região 3' não traduzida (3'-UTR) do RNA mensageiro (mRNA) alvo, levando a degradação do mRNA ou impedindo que este seja traduzido em uma proteína, promovendo o silenciamento gênico (Fig. 4) (FILIPOWICZ; BHATTACHARYYA; SONENBERG, 2008). Qualquer desbalanço nas etapas de biogênese dos miRNAs pode causar alterações na expressão destas moléculas, promovendo o



surgimento de doenças, incluindo o câncer (BJERKE; YI, 2020).

Figura 4 Síntese dos miRNAs a partir da RNA Pol II e ação silenciadora sobre o mRNA alvo.

5.1 RNAs Não Codificadores: os miRNAs no Câncer Colorretal

Como é de se esperar os miRNAs estão fortemente envolvidos na carcinogênese colorretal com possíveis aplicações na prática clínica já identificadas. Em 2018, foi proposta uma nova assinatura de miRNAs para predição da recorrência em pacientes com câncer colorretal nos estágios II e III. Foram avaliados dois conjuntos de dados públicos (TCGA e GSE29623 do GEO) incluindo 265 amostras de pacientes nos estágios II e III e foram selecionados 8 miRs (miR-744, miR-429, miR-362, miR-200b, miR-191, miR-30c2, miR-30b, miR-33a) que poderiam ser utilizados para predizer o risco de recorrência em ambos os estágios tumorais, tornando possível a estratificação de risco dos pacientes (KANDIMALLA et al., 2018).

Além de biomarcadores, os miRNAs estão associados a importantes processos associados a carcinogênese.

Em 2000, os pesquisadores Hanahan e Weinberg definiram os *hallmarks* da carcinogênese em seis processos biológicos fundamentais para o estabelecimento e sobrevivência de um câncer, sendo eles: autossuficiência de sinais de crescimento; insensibilidade a fator anti-crescimento; angiogênese sustentada; potencial replicativo ilimitado; evasão da apoptose e a capacidade de invadir e formar metástases em outros tecidos. Desde a publicação do primeiro estudo, os estudos da oncologia molecular avançaram a passos largos, levando a publicação de uma atualização em 2011, na qual foram adicionados quatro novos *hallmarks*: evasão do sistema imune; promoção de inflamação pró-tumoral; instabilidade genômica; e desregulação do metabolismo energético. (HANAHAN; WEINBERG, 2011).

Em janeiro de 2022 uma nova atualização foi publicada incluindo outros quatro *hallmarks* adicionais: a reprogramação epigenética não-mutacional; desbloqueio da plasticidade fenotípica; senescência celular e microbioma polimórfico, totalizando 14 *hallmarks*. Ao longo dessas atualizações é possível observar que para que possamos realmente entender a carcinogênese é necessário estudar não só as características das células tumorais, mas também como elas se relacionam e são influenciadas pelo hospedeiro (HANAHAN, 2022).

Os miRNAs atuam nos diferentes processos biológicos associados ao câncer, como crescimento celular, diferenciação e regulação da apoptose, e portanto, estão fortemente envolvidos na carcinogênese. Diversos estudos já demonstraram o envolvimento dessas moléculas nas diversas fases do desenvolvimento tumoral e nos *hallmarks* da carcinogênese sendo classificados como oncomirs, os microRNAs que aceleram o processo da carcinogênese; miRs supressores tumorais, responsáveis por desacelerar o desenvolvimento tumoral, além dos metastamirs que estão envolvidos particularmente com o processo metastático (WIGGINS et al., 2022; YAN; BU, 2021).

Os primeiros atributos chave das células tumorais são a proliferação celular constitutiva e evasão dos sinais de interrupção do crescimento celular. Muitos genes e proteínas estão envolvidos nesses processos especialmente quinases e seus receptores. As ciclinas dependentes de quinase (CKDs) são fundamentais na regulação do ciclo celular que estão frequentemente desreguladas no câncer levando ao aumento da proliferação celular, algumas como *CDK4* e *CDK6* são diretamente reguladas por diversos miRNAs.(DING et al., 2020; WHITTAKER et al., 2017). Alguns exemplos de miRs que atuam sobre *CDK4* e *CDK6* são miR-29a-3p, miR-34a-5p, miR-34b-5p, miR-124-3p, miR-145-5p, miR-195-5p, miR-545-3p, miR-497-5p, let-7a-5p, miR-129-2-3p e miR-200a-3p (VAN ROOSBROECK; CALIN, 2017).

Outro gene que também contribui para a desregulação do crescimento celular e são regulados por miRs são o fator de crescimento epidermal (*EGFR*). O *EGFR* é um receptor de tirosina quinase e é alvo terapêutico de diversas drogas utilizadas na prática clínica, como cetuximab utilizado no tratamento do CCD metastático e panitumumab utilizado no tratamento do CCR. Esse gene é alvo de diversos miRNAs como miR-7-5p, miR-27a-3p, miR-133b, miR-146a, miR-218-5p e miR-302a-5p. (CARVALHO et al., 2017; VAN ROOSBROECK; CALIN, 2017)

Outros 2 *hallmarks* da carcinogênese são a evasão da apoptose e indução de instabilidade genômica, muitos fatores pró e anti-apoptóticos atuam na regulação desse *hallmarks*, uma família de genes muito proeminente é a *Bcl-2*, o primeiro fármaco anti-apoptótico aprovado pelo FDA o venetoclax atuam sobre a proteína anti-apoptótica BCL2 (CROCE; REED, 2016). Especificamente no CCR, os miRNAs supressores tumorais que atuam na

regulação da BCL2 são miR-30b-5p, miR-143-3p, miR-195-5p e miR-1915-3p (VAN ROOSBROECK; CALIN, 2017)

Conforme a massa tumoral aumenta de tamanho a angiogênese é fundamental para que o crescimento se sustente, possibilitando o transporte de oxigênio e nutrientes para as células cancerígenas, assim como facilitando o processo de colonização de outros tecidos e formação de metástases. Alguns genes fortemente envolvidos nesse processo são os genes da família *VEGF*, *PDGF* e *FGF* (CARMELIET; JAIN, 2011). A família *VEGF* consiste de poucos membros com papéis não redundantes na angiogênese. Existem 5 ligantes do *VEGF* (*VEGFA* até *VEGFE*) e 3 receptores *VEGFR1*, *VEGFR2* e *VEGFR3*; entre os ligantes o *VEGF-A* é o mais importante sendo um forte estimulador da angiogênese causando proliferação celular, inibição da apoptose, aumento da permeabilidade vascular, vasodilatação e recrutamento de células imunes para o sítio tumoral. A proteína *VEGF-A* é secretada não apenas pelas células endoteliais, mas também por células tumorais, macrófagos, plaquetas, queratinócitos, entre outras, em resposta a deprivação de oxigênio (STANCA MELINCOVICI et al., 2018). Mais de 40 miRNAs já foram validados como reguladores de *VEGF*, sendo que no câncer colorretal já foram identificados os miRNAs miR-30b-5p, miR-143-3p, miR-148a-3p, miR-195-5p e miR-1915-3p (VAN ROOSBROECK; CALIN, 2017).

A evasão do sistema imune é outro *hallmark* do câncer, e ganhou maior notoriedade nos últimos anos devido a imunoterapia no câncer. O microambiente tumoral (TME) é composto por inúmeras células infiltrantes que, ainda que infiltrem o sítio tumoral com o papel de promover a erradicação tumoral, algumas vezes atuam de forma paradoxal promovendo o crescimento tumoral. A vigilância imune deficitária faz com que as células imunes infiltrantes forneçam às células tumorais fatores de crescimento e pró-angiogênicos. Esse processo é gerado em parte por subclones tumorais pouco imunogênicos e em parte por subclones que seriam imunogênicos, mas desenvolveram a capacidade de “desativar” componentes do sistema imune; por exemplo, as células tumorais podem inibir a ação do de linfócitos T e células NK através da secreção de TGF- β (HANAHAN; WEINBERG, 2011; REETESH; BELL; PAI, 2022).

A imunoterapia vem explorando esses processos de evasão imune para modular o próprio sistema imune do paciente para que esse volte a combater o câncer, os inibidores de

checkpoints imunes são alguns dos exemplos presentes na prática clínica. As células tumorais são capazes de expressar o ligante PD-L1 em sua superfície causando a inibição da ação citotóxica das células T ao se ligar no receptor de morte celular programada-1 (*PD-1*), esse fenômeno é semelhante ao que ocorre em outro *checkpoint* imune o CTLA-4. As células tumorais afetam as células apresentadoras de antígenos (APC) do microambiente tumoral e dos linfonodos que drenam a região afetada, fazendo com que as APCs tenham menos fatores co-estimulatórios; como resultado essas células não conseguem se ligar com o receptor estimulatório CD28, mas sim ao receptor inibidor CTLA-4 das células T, especialmente das células T CD4+. Isso evita a ativação e a expansão clonal das células T e, a longo prazo, pode induzir anergia, ou seja, falta de resposta das células T (BRUNI; ANGELL; GALON, 2020).

Esses processos de evasão imune também são regulados por miRNAs, tais como miR-15a, miR-15b, miR-16, miR-195, miR-424, miR-497 e miR-507, que regulam a expressão das moléculas PD-L1 e CD80. A baixa expressão do miR-424 está associada com alta expressão de CD80 e PD-L1 e quimioresistência. Esse miR também é capaz de bloquear a ligação CD80/CTLA-4 através da redução da expressão de CD80. O miR-138-5p com funções supressoras tumorais e expressão reduzida em CCR, inibe diretamente a expressão de PD-L1 ao seu ligar ao seu mRNA. A combinação de baixa expressão desse miR e alta expressão de PD-L1 foi relacionado com um desfecho clínico desfavorável em CCR (SMOLLE et al., 2017) .

Há inúmeros estudos mostrando a expressão de ncRNAs, incluindo miRNAs em CCR. No entanto, estudos comparando os padrões de expressão dessas moléculas de acordo com a lateralidade do tumor ou ainda com sua localização anatômica são relativamente escassos. Para o nosso conhecimento, há 8 trabalhos publicados que relatam os perfis de ncRNAs em CCR de acordo com a lateralidade do tumor (Tabela 1).

Em 2011 Slattery et al avaliaram os microRNAs diferencialmente expressos em tumores colorretais classificados de acordo com as localizações tumorais e características moleculares. Na comparação com amostras normais de cólon distal e proximal 287 miRNAs estavam diferencialmente expressos (DE), também na comparação com amostras normais, 143 miRNAs DE em tumores de cólon MSI+, 129 miRNAs em tumores de cólon CIMP+, 135 miRNAs tumores de cólon com mutação *KRAS2* e 139 miRNAs em tumores de cólon com mutação em *TP53*. Foram encontrados números similares de miRNAs DE em tumores de reto,

porém os miRNAs eram diferentes dos observados em cólon. Foram observados 129 miRNAs únicos para CIMP+, 143 miRNAs únicos para mutação *KRAS2* e 136 miRNAs únicos para mutação *TP53* em tumores de reto (SLATTERY et al., 2011).

Os microRNAs 146a e 147b foram sugeridos como possíveis biomarcadores para CCR, ambos foram mostrados significativamente mais expressos em CCE quando comparados com CCD (OMRANE et al., 2014). Outros microRNAs foram observados como alterados em CCE, como miR-221 e miR-30b que estavam com a expressão reduzida e miR-10b que apresentou expressão aumentada nestes tumores quando comparados com CCD (COEBERGH VAN DEN BRAAK et al., 2018).

Slattery et al (2016) avaliaram o perfil de miRNAs de acordo com o sítio anatômico do tumor, subtipo molecular e prognóstico dos pacientes com CRC. Os miRNAs miR-31-5p, miR-193b-3p, miR-214-3p, miR-28-5p e miR-99b-5 foram associados com sobrevida em câncer de cólon; enquanto miR-552, miR-3653, miR-520e, miR-636 e miR-192-3 associaram-se com sobrevida em câncer de reto (SLATTERY et al., 2016). O miR-31-5p juntamente com miR-224-5p e miR155-5p, foram indicados como marcadores para a progressão de CCD.

O estudo de Yang (2019) *et. al.* identificou 54 miRNAs diferencialmente expressos entre CCD e CCE, sendo que 22 miRNAs que poderiam ser utilizados como biomarcadores para CCD (L. Yang et al., 2019).

Uma análise ampla multi-ômica também identificou diferenças entre tumores de cólon direito e esquerdo, utilizando dados públicos do TCGA. Foi feita uma comparação do perfil de mutações, metilação do DNA, expressão gênica e miRNA. Além de 15 miRs diferencialmente expressos entre as localizações anatômicas, foi mostrado que sinalização cruzada de vias biológicas é mais prevalente em CCD, como a via PI3K que frequentemente influencia nas vias RAS e P53 de sinalização, sugerindo que tumores de cólon direito apresentam mais marcadores de agressividade biológica (W. Hu et al., 2018).

Em junho de 2020, Eneh (2020) et. al. compararam a expressão de microRNAs e suas vias biológicas relacionadas entre CCD e CCE, reforçando que além das grandes regiões anatômicas (cólon e reto), a expressão dos microRNAs também pode ser analisada de acordo com a lateralidade tumoral. Seis miRs foram validados como DE na comparação de CCD e CCE, sendo que os miRs miR-625-3p, miR-155-5p, miR- 625-5p, miR-31-5p e miR-330-5p estavam com a expressão aumentada em CCD, enquanto o miR-196b-5p estava com expressão

aumentada com CCE. Os miRNAs aumentados nos tumores do lado esquerdo regulam genes envolvidos nas vias de sinalização mTor, Wnt, PI3K-Akt, enquanto os miRs aumentados nos tumores do lado direito regulam genes envolvidos na via do TGF- β (ENEH et al., 2020).

Tabela 1. Trabalhos publicados que relatam os perfis de ncRNAs em CCR de acordo com a lateralidade do tumor

Referência	Descrição amostras	Plataforma utilizada	Resultados principais
Slattery (2011)	40 tumores de cólon e 30 de reto; 30 amostras normais (10 cólon proximal, 10 cólon distal, e 10 reto pareado com tumores)	Agilent Human miRNA V3.0 Microarrays	Foram identificados 143 miRNAs diferencialmente expressos (DE) em tumores de cólon MSI+, 129 miRNAs em tumores de cólon CIMP+, 135 miRNAs tumores de cólon com mutação <i>KRAS2</i> e 139 miRNAs em tumores de cólon com mutação em <i>TP53</i> . Números similares de miRNAs DE em tumores de reto foram detectados, porém os miRNAs eram diferentes dos observados em cólon. Foram observados 129 miRNAs únicos para CIMP+, 143 miRNAs únicos para mutação <i>KRAS2</i> e 136 miRNAs únicos para mutação <i>TP53</i> em tumores de reto
Omrane (2014)	25 amostras pareadas de CCR e mucosas normais, 11 amostras de cólon direito e 14 de cólon esquerdo	miScript II reverse transcription Kit (Qiagen).	Os miRs -21, -146a e -135a estavam mais expressos em tecidos tumorais. Os miRs 146a e 147b estavam mais expressos em CCE, esses dois miRs, e particularmente o miR-146a, são sugeridos como possíveis biomarcadores para CCR associados com a lateralidade do tumor.

Slattery (2016)	1893 amostras de CRC e normais pareadas	Agilent Human miRNA Microarray V19.0	Na comparação de pacientes com câncer de reto ou cólon em estágio 1 ou 2, nenhum miRNAs foi associado significativamente com estadio da doença. Dois miRNAs de expressão pouco frequentes foram associados com sobrevida em câncer de cólon. Nos pacientes diagnosticados com câncer de reto 201 miRNAs estavam associados com sobrevida. A análise de 105 miRNAs previamente associados com prognóstico mostrou que 4 influenciavam a sobrevida em câncer de cólon e 17 influenciavam em câncer de reto.
van den Braak (2018)	232 amostras de CCR	9-plex miRNA assay protocol e TaqMan miRNA assays	Expressão dos miRs 221 e 30b estava significativamente reduzida ao mesmo tempo que a expressão do miR-10b estava significativamente aumentada em CCD na comparação com CCE.
Hu (2018)	Dados públicos TCGA-COAD (grupo descoberta) Dados público GSE14333 (grupo validação)	limma package for R/Bioconductor, miRanda e TargetScan	15 miRNAs estavam diferencialmente expressos entre CCD e CCE. 8 estavam regulados negativamente em reto enquanto 7 estavam regulados positivamente.

Yang (2019)	300 amostras do TCGA: 151 tumores de cólon direito e 149 de cólon esquerdo (grupo descoberta) 14 amostras: 7 cólon direito e 7 cólon esquerdo (grupo validação)	R-bioconductor package DESeq2 qRT-PCR	54 miRNAs DE, sendo 22 regulados negativamente e 32 positivamente estavam diferencialmente expressos entre CCD e CCE. 22 miRNAs foram sugeridos como biomarcadores diagnósticos ótimos para CCD.
Xi (2019)	217 amostras com dados de RNA-seq depositados no banco de dados GDAC Firehose: 78 CCD, 98 CCED e 41 de reto	EdgeR TMM normalization (Version 3.4)	O prognóstico variou significativamente com relação ao sítio tumoral, quanto mais distal a localização tumoral, maior foi a sobrevida predita, 26 lncRNAs foram correlacionados com os sítios de CRC.

Eneh (2020)	24 amostras de CCD e CCE (grupo descoberta)	DESeq2 e DIANA/mirPath tool	Foram validados 6 miRs diferencialmente expressos entre CCD e CCE, sendo 5 com expressão aumentada em CCD e 1 com expressão aumentada CCE.
----------------	--	--------------------------------	---

Dados públicos miRSeq de
201 amostras de CRC do
TCGA (grupo validação).

1 5.2 Novos microRNAs

2 A caracterização inicial do transcrito de miRNA humano foi baseada
3 principalmente nas sequências mais abundantes de miRNA e/ou àquelas conservadas
4 nos vários tipos de tecidos, o que limitou a descoberta de miRNAs com expressões
5 restritas a tecidos individuais ou linhagens celulares (FRIEDLÄNDER et al., 2014).

6 Utilizando ferramentas avançadas de bioinformática em dados de
7 sequenciamento de nova geração (NGS) de RNAs pequenos (small-RNAseq) foram
8 reportados 2.469 novos candidatos de miRNA humanos, dos quais 1.098 foram
9 validados (FRIEDLÄNDER et al., 2014). Em 2015, Londin et. al. avaliaram o small-RNAseq
10 de 1.323 amostras representativas de 13 tipos distintos de células humanas e
11 identificaram 3.707 novos miRNAs humanos expressos em todo o genoma. Em conjunto
12 esses dados indicam uma maior diversidade de miRNAs humanos do que previamente
13 prevista e ainda não anotados em bancos de dados.

14 O algoritmo miRDeep2 é uma das ferramentas utilizadas para descoberta de
15 sequências de miRNAs não anotadas, que busca no transcrito novos candidatos à
16 miRNA e os compara com sequências conhecidas de miRNA (FRIEDLÄNDER et al., 2012).
17 Outras ferramentas foram desenvolvidas, como a ferramenta online miRMaster, que
18 além de realizar maior número de análises comparada ao miRDeep2, ainda permite
19 maior velocidade para obtenção dos dados resultantes das análises (FEHLMANN;
20 MEESE; KELLER, 2017). O mesmo grupo lançou em 2021 uma atualização do miRMaster,
21 miRMaster 2.0, com várias atualizações e novos recursos, que incluem a possibilidade
22 de análise de oito espécies (por exemplo, humano, camundongo, cachorro, etc), análise
23 de 10 classes de ncRNAs, tais como miRNAs, piRNAs, circRNAs, etc, além de novas
24 análises downstream, como análise de efeito em lote, bancos de dados de anotação
25 atualizados incluídos por padrão (miRBase, Ensembl, GtRNAdb), possibilidade de análise
26 de single-cell RNAseq, entre outras funcionalidades (FEHLMANN et al., 2021)

27 A última atualização do miRBase (v22.1) é de 2018, e com o crescente número
28 de novos possíveis miRNAs sendo descritos, outras plataformas com bancos de dados
29 de miR conhecidos e preditos têm sido necessárias (LUNA BUITRAGO; LOVERING;
30 CAPORALI, 2023)

1 Com o aumento considerável de dados públicos de RNAseq depositados no Gene
2 Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>), bem como de dados
3 de RNAseq de tumores de diferentes tecidos obtidos pelo The Cancer Genome Atlas
4 (TCGA, disponível em <https://portal.gdc.cancer.gov/>), estudos utilizando abordagens *in*
5 *silico* têm sido realizados para identificação de novos miRNAs em diferentes tecidos e
6 tumores (Tabela 2).

7 Minatel et al (2018) analisaram dados do TCGA de small-RNAseq de 47 tecidos
8 hepáticos normais adjacentes obtidos de pacientes com carcinoma hepatocelular
9 utilizando a ferramenta miRDeep2. Foram identificados 103 miRNA candidatos ainda
10 não descritos em tecidos hepáticos. Interessantemente, 83/103 miRNAs apresentaram
11 diminuição de expressão nas amostras pareadas de carcinomas hepatocelulares,
12 sugerindo que estes miRs devem ter um papel importante na homeostase do fígado.
13 Análise de predição de alvos mostrou que estes novos miRNAs potencialmente regulam
14 genes participantes de vias de regeneração tecidual e metabolismo de ácidos graxos,
15 importantes na biologia do fígado.

16 Em outro estudo, dados de small-RNAseq de 71 casos pareados de tecido não
17 maligno e de carcinoma de célula renal e 502 tumores do TCGA foram analisados
18 utilizando miRDeep2. Foram identificados 40 novos miRNAs não anotados nos tecidos
19 não malignos e 143 nos tumores. Na etapa de validação dos achados foram detectados
20 26/40 e 102/143 destes miRNAs avaliando o small-RNAseq em um painel de 8 linhagens
21 celulares renais. 59 destes miRNAs expressos em ambos tecidos (normal e tumoral)
22 foram reanalisados e destes, 30 apresentaram-se diferencialmente expressos entre
23 tumor e tecido normal, sugerindo um potencial papel desses novos miRNAs no câncer
24 renal. Notavelmente, dois miRNAs com altos valores de *fold change* (*downregulated* 28x
25 e *upregulated* 5x) nos tumores mostraram-se associados com pior sobrevida, reforçando
26 a importância desses novos miRNAs no desenvolvimento e progressão deste tumor
27 (SAGE et al., 2018). Novos miRNAs também foram identificados em câncer de cabeça e
28 pescoço (ROCK et al., 2019) e câncer de pulmão (MARTINEZ et al., 2019)

29

1 **Tabela 2.** Revisão dos estudos sobre descoberta de novos microRNA em tecidos
 2 humanos

Estudo	Tecido Analisado	Nº de miRs encontrado	Nº de miRs validados
RESHMI et al., (2011)	Câncer de colo uterino	13	4
LONDIN et al., (2015)	Plaquetas, tecido prostático, tecido mamário, pâncreas, células B, soro, células mononucleares do sangue periférico (PBMCs), tecido gástrico, linfócitos CD3+, tecido cerebral, tecido cutâneo	3707	-
BACKES et al., (2016)	Sangue	518	11
SAGE et al.,(2018)	Carcinoma Renal e Rim	183	30
MINATEL et al., (2018)	Fígado	103	-
BARROS-FILHO et al. (2019)	Carcinoma Papilar de Tireoide	234	9
PEWARCHUK et al., (2019)	Adenocarcinoma Gástrico	170	43
MARTINEZ et al., (2019)	Mesotelioma Pleural	154	-
ROCK et al (2019)	Carcinoma Espinocelular de Cabeça e Pescoço	146	-

3

4 Outros estudos utilizaram a plataforma miRMaster para identificação de novos
 5 miRNAs e detectaram novos miRNAs preditos em câncer gástrico (PEWARCHUK et al.,
 6 2019) e câncer de tireoide (BARROS-FILHO et al., 2019) (Tabela 2). Pewarchuk et al.
 7 (2019) utilizando dados públicos do TCGA (grupo de descoberta), identificaram 170
 8 novos miRNAs candidatos em tecidos gástricos, sendo que 143 DE nos tumores. Destes,
 9 43 novos miRNAs foram detectados no grupo de validação composto de 25 amostras
 10 gástricas. Em tecido não neoplásico de tireoide e em tumores papilíferos de tireoide
 11 foram detectados 234 novos miRNAs utilizando dados públicos do TCGA, sendo que 125
 12 novos miRs estavam diferencialmente expressos nos tumores. Nove novos miRs foram
 13 confirmados no grupo de validação (16 tumores e 8 tecidos normais, GSE63511)
 14 (BARROS-FILHO et al., 2019).

1 Desta forma, análises de dados de small-RNAseq utilizando esses algoritmos
2 foram capazes de detectar novos miRNAs que anteriormente não haviam sido descritos.
3 Além disso, foi possível observar que esses miRNAs recém-detectados eram tecido-
4 específicos (BACKES et al., 2016). A identificação de miRNAs anteriormente não
5 caracterizados pode representar novos reguladores da biologia tecidual com impacto na
6 patogênese de doenças, e ainda sendo úteis na clínica como potenciais biomarcadores
7 e alvos terapêuticos. Notadamente, não identificamos nenhum estudo investigando
8 novos miRNAs em CCR e relacionando a expressão desses candidatos a lateralidade
9 tumoral, como o proposto no presente estudo.

10 MicroRNAs se originam de diferentes regiões do genoma, a proporção de
11 distribuição de localizações genômicas parece se manter entre os miRNAs conhecidos e
12 novos. Em sua maioria a origem dos miRNAs são regiões intergênicas, seguida de regiões
13 intrônicas e regiões de alta repetição. Além disso, diversos novos miRs estão localizados
14 no braço do miRNA precursor que anteriormente era denominado de passageiro ou
15 estrela e que se acreditava ser sem função, porém já foi mostrada ser funcionalmente
16 relevante (LONDIN et al., 2015) Além disso, cerca de 34% dos miRNAs conhecidos estão
17 localizados em regiões frágeis dos cromossomos, estas regiões são *loci* hereditários
18 específicos suscetíveis a quebras e rearranjos cromossomais e representam cerca de
19 26,38% do genoma humano, e que contribuem para instabilidade genômica
20 característica do câncer (MATHELIER; CARBONE, 2013). Novos microRNAs já foram
21 mapeados nessas regiões e uma das hipóteses para recente descoberta dessas novas
22 moléculas é que elas seriam transcritas mais recentemente na evolução resultantes da
23 instabilidade genômica dos sites frágeis (CALIN; CROCE, 2007; FROMM et al., 2013; RESHMI
24 et al., 2011).

25

26 Hipótese

27 Existem ncRNAs, particularmente miRNAs, ainda não descritos que se expressam em
28 tumores colorretais humanos, podendo apresentar diferenças de expressão de acordo
29 com a lateralidade do tumor (cólon direito e esquerdo e reto). Estes miRNAs podem ser

1 modulados durante a carcinogênese e podem futuramente ser possíveis biomarcadores
2 em CCR.

3 Objetivo

4 Identificar novos ncRNAs, especificamente miRNAs, associados com a carcinogênese
5 colorretal em cólon direito, esquerdo e reto utilizando abordagens *in silico*.

6

7 Objetivos Específicos

8 1. Identificar miRNAs ainda não descritos a partir de dados de sequenciamento em
9 larga escala de RNA obtidos de amostras de adenocarcinoma de cólon direito, cólon
10 esquerdo e de reto disponíveis em bancos públicos do TCGA.

11

12 2. Associar os novos miRNAs com a lateralidade e localização anatômica dos tumores
13 colorretais

14

15 3. Identificar potenciais genes alvos desses novos miRNAs por meio de análise de
16 predição e redes de interação proteína-proteína baseada nos genes alvos
17 identificados.

18

19 Materiais e Métodos

20

21 1. Fonte de Dados de Expressão dos RNAs não Codificadores (ncRNAs)

22 1.1 Grupo de descoberta

23 Dados de sequenciamento de RNA (RNAseq de tecidos colorretais foram obtidos
24 do *The Cancer Genome Atlas* (TCGA) no *Cancer Genomics Hub*
25 (<https://portal.gdc.cancer.gov/>). Os dados disponíveis publicamente de small-RNAseq
26 são normalizados e agregados (nível 3). No entanto, para a identificação de novos
27 miRNAs é necessária obtenção dos dados de sequenciamento brutos e não
28 normalizados (nível 1). Assim foi obtida autorização para obtenção dos dados nível 1. A
29 plataforma utilizada pelo TCGA é Illumina.

1 Foram analisadas 522 amostras de tecido colorretal, sendo 11 amostras de
2 tecido normal adjacente aos tumores e 511 amostras de tecidos tumorais. As amostras
3 de CCR foram classificadas de acordo com as localizações anatômicas da tabela de dados
4 clínicos disponibilizados pelo TCGA. Desta forma, dentre os tumores foram analisadas
5 172 amostras de CCR direito (83 ceco, 60 cólon ascendente, 12 flexura hepática e 17
6 cólon transversal) e 257 CCR esquerdo (4 flexura esplênica, 12 cólon descendente, 98
7 sigmóide, 66 junção reto-sigmoide e 77 tumores de reto). Em 82 amostras de tumores
8 não consta a localização anatômica, apenas definidas como cólon, e essas amostras
9 foram excluídas das análises onde foi considerada lateralidade do tumor.

10 1.2 Grupo de validação

11 Com intuito de validar a presença de novos miRNAs em CCR, foram buscados
12 outros estudos de miRNAs em CCR utilizando o banco de dados *Gene Expression Omnibus*
13 (GEO) (<https://www.ncbi.nlm.nih.gov/gds>) para comparar com os resultados obtidos
14 com as amostras do TCGA. Foram selecionados estudos cujo sequenciamento foi
15 realizado com a plataforma Illumina small RNA sequencing (mesma utilizada pelo TCGA),
16 para evitar os vieses de plataformas diferentes. Além disso, foram selecionados apenas
17 estudos com número amostral mínimo de 50 amostras (mínimo de 25 CCR) e com mais
18 de 5Mb de *reads* totais.

19 Os critérios utilizados na busca foram:

20 **Keywords:** ((colon OR rectal OR colorectal OR CRC) AND (Cancer OR carcinoma OR
21 adenocarcinoma OR tumor OR tumour OR neoplasia)) AND "Homo sapiens"

22 **Study type:** Non-coding RNA profiling by high throughput sequencing

23 **Sample count:** From 50 to 999999999999

24

25 A busca no banco de dados GEO seguindo os critérios descritos resultou na
26 identificação de 12 datasets (Tabela 3), sendo que apenas dois estudos atendiam
27 inicialmente aos critérios de inclusão: GSE89974 e GSE63119. Destes, apenas os dados
28 do estudo GSE89974 foram utilizados nas análises no miRMaster, uma vez que o estudo
29 GSE63119 não tinha amostras de tecido normal para controle. A justificativa para
30 exclusão dos demais estudos está mostrada na Tabela 3

1 O estudo Wu, *et. al.* (2018), dados depositados sob o número de acesso
 2 GSE89974, utilizou 95 amostras de tecido colorretal, sendo 20 amostras normais, 26
 3 amostras de adenoma, 35 amostras de adenocarcinoma e 14 amostras de adenoma
 4 sésil serrilhado. Os dados das 20 amostras normais e 35 amostras de adenocarcinoma
 5 do GSE89974 disponíveis no formato FASTQ foram analisados na plataforma MirMaster
 6 seguindo os mesmos critérios utilizados para o grupo de descoberta (TCGA).

7

8 **Tabela 3.** Resultado da busca de estudos utilizando small RNAseq em CCR no banco de
 9 dados Gene Expression Omnibus (GEO).

Nº de Acesso	Atende aos critérios de inclusão	Justificativa
GSE118504	N	Sem amostras tumorais
GSE89974	S	Incluído
GSE110381	N	Sem amostras tumorais
GSE102855	N	Apenas amostras de linhagem celular
GSE87817	N	Apenas amostras de linhagem celular
GSE71008	N	Apenas amostras de sangue
GSE70408	N	Apenas amostras de linhagem celular
GSE66209	N	Apenas sete amostras de CCR
GSE63119	S	Sem amostras normais para controle
GSE28866	N	Apenas seis amostras de CCR
GSE16256	N	Apenas amostras de linhagem celular

10

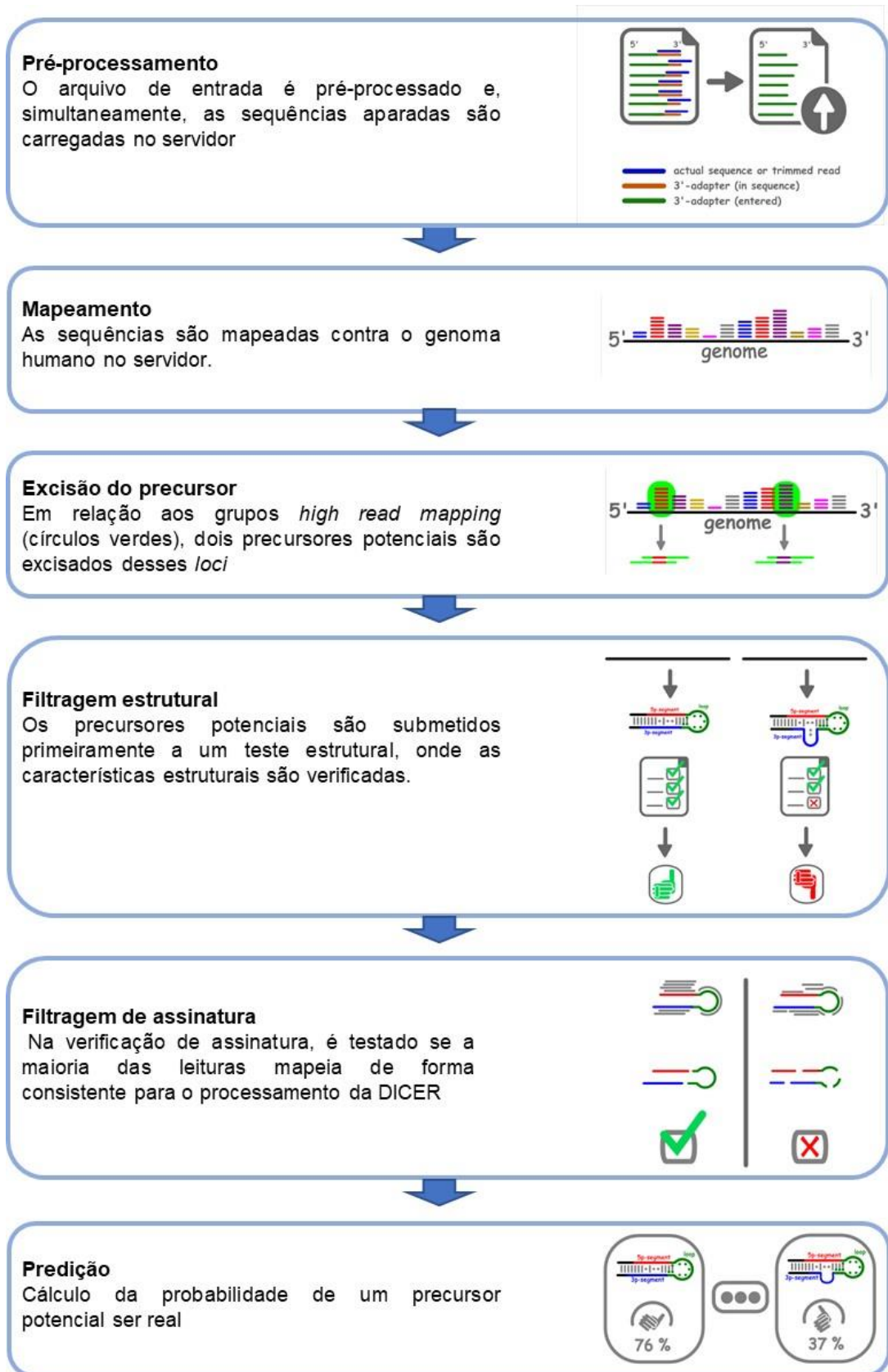
11

12

1 2. Métodos de Resgate de Dados e Análise

2 Todas as sequências brutas obtidas do TCGA e do GSE89974 foram processadas
3 para detecção de RNAs pequenos seguindo um pipeline previamente descrito
4 (MARTINEZ et al., 2015). Arquivos BAM brutos foram convertidos em FASTQ com a
5 ferramenta “SamToFastq” do pacote de análise Picard
6 (<http://broadinstitute.github.io/picard/>). Reads não alinhados foram aparados baseado
7 em seu escore de qualidade Phred (≥ 20) e alinhados de acordo com a versão mais
8 recente do genoma humano (hg38) utilizando o alinhador *Spliced Transcripts Alignment*
9 *to a Reference* (STAR)

10 Para predição de miRNAs previamente não anotados utilizamos a plataforma
11 miRMaster (<https://ccb-compute.cs.uni-saarland.de/mirmaster/tutorial/>), que, a partir
12 dos arquivos FASTQ realiza a busca de novos miRNAs candidatos, juntamente com a
13 quantificação da expressão e identificação de isoformas e variantes de miRNAs. Além
14 disso, o miRMaster também mapeia reads de sncRNAs de vírus e bactérias a partir do
15 banco de dados do NCBI RefSeq permitindo a identificação de contaminações, infecções
16 e miRNAs exógenos. É possível ainda quantificar a expressão de outras classes de
17 ncRNAs como os piRNAs, além dos rRNAs, tRNA, snRNAs, snoRNAs, scaRNAs e lincRNA.
18 (Fehlmann et al., 2017). Um fluxograma com as etapas de análise e identificação de
19 possíveis novos miRNAs está apresentado na Figura 5.



1

2 **Figura 5.** Fluxograma com as etapas de análise e identificação de possíveis novos
 3 miRNAs utilizando a ferramenta online miRMaster. Figura adaptada do site ([https://ccb-
 4 compute.cs.uni-saarland.de/mirmaster/about](https://ccb-compute.cs.uni-saarland.de/mirmaster/about)).

1 3. Expressão diferencial dos miRNAs candidatos

2 Para identificação de novos miRNAs diferencialmente expressos nos tumores foi
3 adotado o critério $p < 0,05$ e de *fold-change* (FC) $\geq 1,2$ para que o candidato fosse
4 considerado como mais expresso nas amostras tumorais em comparação com os tecidos
5 normais e $FC \leq 1,1$ para que o candidato fosse considerado como menos expresso nas
6 amostras tumorais em comparação com os tecidos normais. Os valores de FC adotados
7 nas análises se basearam no conhecimento que possíveis novos miRNAs apresentam
8 níveis de expressão baixos, de modo que esses candidatos não fossem excluídos nas
9 primeiras análises. Este critério foi adotado para o grupo de descoberta e de validação.
10 Expressão diferencial dos miRs conhecidos também foi realizada comparando amostras
11 tumorais e normais utilizando $p < 0,05$.

12 4. Validação *in silico* dos miRs candidatos

13 Os miRNAs identificados como diferencialmente expressos nos CCRs do grupo de
14 descoberta foram comparados aos miRNAs diferencialmente expressos nos CCRs do
15 grupo de validação. Foram considerados como validados apenas os miRNAs candidatos
16 identificados nos dois grupos e que apresentavam mesmo sentido de expressão nos
17 tumores em relação aos tecidos normais, ou seja, mais expresso nos tumores que nos
18 normais ou menos expressos nos tumores que nos tecidos normais tanto no grupo de
19 descoberta quanto de validação. A identificação dos novos miRNAs presentes nos dois
20 estudos (TCGA e GSE89974) aumenta a chance da descoberta de novos miRNAs
21 verdadeiros.

22 Análise de especificidade Tecidual dos miRNAs candidatos

23 Para análise de especificidade tecidual, os novos miRs identificados em CCR
24 foram comparados à tecidos normais de diferentes localizações. Foram incluídas nessas
25 análises apenas amostras colorretais normais dos grupos de descoberta e validação
26 (N=31) e comparadas às amostras de tecido normal adjacente de 5 sítios anatômicos
27 diferentes, conforme descrito a seguir: pulmão (n=89), mama (n=90), bexiga urinária
28 (n=19), fígado (n=59) e sistema nervoso central (n=5). Esses sítios foram escolhidos por
29 compartilharem a mesma origem embriológica do tecido colorretal (pulmão, bexiga
30 urinária, fígado) ou ser um sítio metastático típico dos tumores colorretais (SNC, pulmão

1 e fígado). Os dados brutos (nível 1) de small-RNAseq dessas amostras obtidas do TCGA
2 foram submetidos à análise na plataforma miRMaster para identificação de novos
3 miRNAs. Os níveis de expressão normalizados dos novos miRNAs candidatos
4 identificados nos tecidos normais colorretais foram comparados com os novos miRs
5 identificados nos 5 sítios anatômicos diferentes usando uma análise de redução de
6 dimensionalidade não linear t-Distributed Stochastic Neighbor Embedding (t-SNE) no
7 programa Orange Data Mining (DEMŠAR et al., 2013).

8 Um *heatmap* com o padrão de expressão dos novos miRs identificados em cada
9 amostra dos diferentes tecidos foi construído utilizando o pacote do R ComplexHeatmap
10 utilizando Z-score (GU; EILS; SCHLESNER, 2016) e finalizado no Inkscape
11 (<https://inkscape.org/>).

12 Análise de componente principal (PCA) dos novos miRNAs

13 A expressão dos miRNAs conhecidos e novos foi analisada para verificar se essas
14 moléculas poderiam ser utilizadas para distinguir amostras normais de amostras
15 tumorais, para esse fim utilizamos uma análise de componente principal (PCA), realizada
16 na plataforma ClustVis (<https://biit.cs.ut.ee/clustvis/>).

17 Associação da expressão dos novos miRNAs com localização anatômica

18 A expressão dos miRNAs candidatos validados foi então analisada de acordo com
19 o sítio anatômico de origem e estadiamento tumoral das amostras do TCGA, foram feitas
20 análises de acordo com a lateralidade do tumor (CCR direito, que inclui ceco, cólon
21 ascendente, flexura hepática e transverso; CCR esquerdo, incluindo flexura esplênica,
22 cólon descendente, sigmóide e junção retosigmoide; e câncer de reto). Também foram
23 analisados os níveis de expressão do miRNAs candidatos de acordo com a localização
24 anatômica dos tumores, buscando ilustrar as alterações de expressões do microRNAs ao
25 longo do trato intestinal.

26 A análise de associação só pôde ser realizada no grupo de descoberta (TCGA),
27 pois as amostras do grupo de validação não apresentavam dados da região anatômica
28 de origem. A comparação dos níveis de expressão dos novos microRNAs entre grupos
29 foi realizada utilizando os testes de Mann-Whitney e Kruskal-Wallis.

1 Associação da expressão dos novos miRNAs com sobrevida

2 As curvas de sobrevida câncer-específica para cada um dos miRNAs candidatos
3 validados foram construídas baseadas no padrão de expressão desses miRNAs nas
4 amostras tumorais do TCGA. Para dicotimização dos dados de expressão dos novos
5 miRs, cada amostra foi categorizada como *up-regulated* ou *down-regulated* de acordo
6 com z-score $>2,0$ ou $<-2,0$, respectivamente, em relação aos tecidos colorretais normais.
7 As amostras que não atendiam esse critério foram categorizadas como expressão
8 normal. A análise de sobrevida foi realizada pelo método de Kaplan-Meier, com o teste
9 de log-rank para a comparação das curvas. Foram considerados estatisticamente
10 significativos os resultados com $p<0,05$. Foi utilizado o programa SPSS v27 (*Statistical*
11 *Package for Social Science*, IBM).

12 Predição de alvos regulados pelos miRNAs candidatos

13 Considerando se tratar de novos miRNAs, para análise de predição dos genes
14 regulados por estes miRNAs candidatos foi necessária utilização de uma plataforma que
15 realiza a predição de alvos a partir das sequências dos novos miRs e a
16 complementariedade de bases entre essas sequências e os possíveis genes alvos. Para
17 tanto foi utilizada a plataforma miRDB(<http://mirdb.org/custom.html>) (CHEN; WANG,
18 2020), uma base de dados online para a previsão de alvos miRNA e anotações funcionais.

19 Todos os alvos no miRDB são preditos pela ferramenta, MirTarget (LIU; WANG,
20 2019), que foi desenvolvida analisando milhares de interações miRNA-alvo a partir de
21 experiências de sequenciamento de produção. A ferramenta miRDB atribui a todos os
22 alvos preditos um escore de predição de 50 a 100. Seguindo a recomendação dos
23 desenvolvedores foram considerados apenas os alvos que apresentavam escore igual
24 ou superior a 80, pois essas são as predições com maior chance de serem reais (CHEN;
25 WANG, 2020).

26 Análise de Expressão Gênica dos Genes Alvos

27 Para identificar genes alvos com sentido de expressão inversa aos seus possíveis
28 novos miRNAs reguladores, o próximo passo foi avaliar o padrão expressão gênica dos
29 alvos das mesmas amostras tumorais utilizando dados de RNA-seq obtidos através da
30 plataforma Xena Browser (<https://xenabrowser.net/>) (GOLDMAN et al., 2020). A análise

1 de expressão diferencial foi realizada comparando amostras tumorais e normais dos
2 projetos TCGA-COAD (n= 567) e TCGA-READ (n=97) para identificarmos os genes alvos
3 diferencialmente expressos nos tumores. As amostras de tumores de cólon foram
4 separadas em cólon esquerdo e cólon direito comparadas com a amostras normais da
5 mesma localização anatômica. Para definir o sentido de expressão dos genes alvos foi
6 utilizado $p < 0,05$ e $FC \geq 2$ para os alvos considerados *up-regulated* e $FC \leq -2$ para os alvos
7 considerados *down-regulated*. A seguir cruzamos esses dados com os nossos candidatos
8 e observamos quais duplas miR-alvo tinham direções opostas de FC, o que sugere que o
9 miRNA candidato está envolvido na regulação do respectivo alvo no sítio tumoral. A
10 seguir, todos os pares candidatos miRNAs-genes alvos foram submetidos a análise de
11 correlação de Spearman para identificar os pares que apresentavam correlação negativa
12 significativa ($p < 0,05$, $r \leq -0,1$) entre suas expressões, aumentando as chances de
13 identificação de pares candidatos miRNAs-genes alvos reais.

14 Construção das redes de interação entre os alvos

15 Após identificação dos genes alvos dos miRNAs candidatos (àqueles com
16 correlação negativa significativa), foi realizada análise de interações entre os alvos
17 utilizando a plataforma String (<https://string-db.org/>) (SZKLARCZYK et al., 2023). O
18 programa Cytoscape (SHANNON et al., 2003a) foi utilizado para construir as figuras das
19 redes de interação dos alvos juntamente com os respectivos miRs reguladores

20 Análise de Enriquecimento

21 Para a análise de enriquecimento dos alvos foram divididos em 3 grupos “Alvos
22 regulados em Cólon Esquerdo”, “Alvos regulados em Cólon Direito” e “Alvos regulados
23 em Reto”. Esses grupos foram analisados na plataforma Enrichr
24 (<https://maayanlab.cloud/Enrichr/>) utilizando 2 módulos de ontologias “GO Biological
25 Process 2021” e “GO Molecular Function 2021” e 2 módulos de vias biológicas “KEGG
26 Human 2021” e “Reactome 2022”.(CHEN et al., 2013; KULESHOV et al., 2016)

27

28

1 Referências

- 2 AHMED, Z.; MAL, C. Functional role of hub molecules in miRNA and transcription factor
3 mediated gene regulatory network of colorectal and lung cancer. **Gene Reports**, v. 23, 1 jun.
4 2021.
- 5 ARRIBAS, J. et al. NF-κB Mediates the Expression of TBX15 in Cancer Cells. **PLoS ONE**, v. 11, n.
6 6, 1 jun. 2016.
- 7 BACKES, C. et al. Prioritizing and selecting likely novel miRNAs from NGS data. **Nucleic Acids**
8 **Research**, v. 44, n. 6, p. 53, 2016.
- 9 BAILEY, C. E. et al. Increasing disparities in the age-related incidences of colon and rectal
10 cancers in the United States, 1975-2010. **JAMA surgery**, v. 150, n. 1, p. 17–22, 2015.
- 11 BARROS-FILHO, M. C. et al. Previously undescribed thyroid-specific miRNA sequences in
12 papillary thyroid carcinoma. **Journal of Human Genetics**, v. 64, n. 5, p. 505–508, 1 maio 2019.
- 13 BASS, A. J. et al. Comprehensive molecular characterization of gastric adenocarcinoma. **Nature**
14 **2014 513:7517**, v. 513, n. 7517, p. 202–209, 23 jul. 2014.
- 15 BATTAGLIN, F. et al. Microsatellite instability in colorectal cancer: Overview of its clinical
16 significance and novel perspectives. **Clinical Advances in Hematology and Oncology**, v. 16, n.
17 11, p. 735–747, 2018.
- 18 BENEDIX, F. et al. Comparison of 17,641 patients with right- and left-sided colon cancer:
19 Differences in epidemiology, perioperative course, histology, and survival. **Diseases of the**
20 **Colon and Rectum**, v. 53, n. 1, p. 57–64, jan. 2010.
- 21 BJERKE, G. A.; YI, R. Integrated analysis of directly captured microRNA targets reveals the
22 impact of microRNAs on mammalian transcriptome. **Rna**, p. rna.073635.119, 2020.
- 23 BRUNI, D.; ANGELL, H. K.; GALON, J. **The immune contexture and Immunoscore in cancer**
24 **prognosis and therapeutic efficacy. Nature Reviews Cancer**, 2020. Disponível em:
25 <<https://doi.org/10.1038/>>. Acesso em: 27 ago. 2020
- 26 CALIN, G. A.; CROCE, C. M. Chromosomal rearrangements and microRNAs: a new cancer link
27 with clinical implications. **Journal of Clinical Investigation**, v. 117, n. 8, p. 2059, 8 ago. 2007.
- 28 CARMELIET, P.; JAIN, R. K. Molecular mechanisms and clinical applications of angiogenesis.
29 **Nature**, v. 473, n. 7347, p. 298–307, 19 maio 2011.
- 30 CARVALHO, T. I. DE et al. Analysis of gene expression EGFR and KRAS, microRNA-21 and
31 microRNA-203 in patients with colon and rectal cancer and correlation with clinical outcome
32 and prognostic factors. **Acta Cirurgica Brasileira**, v. 32, n. 3, p. 243–250, 2017.
- 33 CHAN, A. K. C. et al. Management of Colorectal Cancer with Synchronous Liver Metastases: An
34 Inception Cohort Study (CoSMIC). **Annals of Surgical Oncology**, v. 29, 2022.

- 1 CHAN, A. K. C.; SIRIWARDENA, A. K. Practical Implications of KRAS Mutation Status and
2 Sidedness of Primary Tumour in Patients with Colorectal Cancer and Synchronous Liver
3 Metastases: A Subset Analysis of the CoSMIC Study. **Cancers**, v. 14, n. 19, 1 out. 2022.
- 4 CHANG, G. et al. Hypoexpression and Epigenetic Regulation of Candidate Tumor Suppressor
5 Gene CADM-2 in Human Prostate Cancer. **Clinical cancer research : an official journal of the
6 American Association for Cancer Research**, v. 16, n. 22, p. 5390, 15 nov. 2010.
- 7 CHEN, E. Y. et al. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis
8 tool. **BMC Bioinformatics**, v. 14, p. 128, 15 abr. 2013.
- 9 CHEN, L. et al. miR-203a-3p promotes colorectal cancer proliferation and migration by
10 targeting PDE4D. **American Journal of Cancer Research**, v. 8, n. 12, p. 2387, 2018.
- 11 CHEN, Y.; WANG, X. MiRDB: An online database for prediction of functional microRNA targets.
12 **Nucleic Acids Research**, v. 48, n. D1, p. D127–D131, 8 jan. 2020.
- 13 CHUNG, Y. et al. Nuclear Expression Loss of SSBP2 Is Associated with Poor Prognostic Factors in
14 Colorectal Adenocarcinoma. **Diagnostics**, v. 10, n. 12, 1 dez. 2020.
- 15 COEBERGH VAN DEN BRAAK, R. R. J. et al. Confirmation of a metastasis-specific microRNA
16 signature in primary colon cancer. **Scientific Reports**, v. 8, n. 1, p. 1–11, 2018.
- 17 CROCE, C. M.; REED, J. C. Finally, an apoptosis-targeting therapeutic for cancer. **Cancer
18 research**, v. 76, n. 20, p. 5914, 10 out. 2016.
- 19 CUI, X. et al. Homer1 is a Potential Biomarker for Prognosis in Human Colorectal Carcinoma,
20 Possibly in Association with G3BP1 Signaling. **Cancer Management and Research**, v. 12, p.
21 2899, 2020.
- 22 DAI, Y. et al. TCF21 functions as a tumor suppressor in colorectal cancer through inactivation of
23 PI3K/AKT signaling. **Oncotargets and therapy**, v. 10, p. 1603–1611, 14 mar. 2017.
- 24 DEBACKER, K.; FRANK KOOY, R. Fragile sites and human disease. **Human Molecular Genetics**, v.
25 16, n. R2, p. R150–R158, 15 out. 2007.
- 26 DEMŠAR, J. et al. Orange: Data Mining Toolbox in Python Tomaž Curk Matija Polajnar Laň
27 Zagar. **Journal of Machine Learning Research**, v. 14, p. 2349–2353, 2013.
- 28 DENG, Y. Rectal Cancer in Asian vs. Western Countries: Why the Variation in Incidence?
29 **Current Treatment Options in Oncology**, v. 18, n. 10, p. 64, 25 out. 2017.
- 30 DING, L. et al. The Roles of Cyclin-Dependent Kinases in Cell-Cycle Progression and Therapeutic
31 Strategies in Human Breast Cancer. **International Journal of Molecular Sciences**, v. 21, n. 6, 1
32 mar. 2020.
- 33 ENEH, S. et al. MicroRNAs associated with biological pathways of left- And right-sided
34 colorectal cancer. **Anticancer Research**, v. 40, n. 7, p. 3713–3722, 1 jul. 2020.

- 1 ENTERLINE, P. E.; DAY, R.; MARSH, G. M. Cancers related to exposure to arsenic at a copper
2 smelter. **Occupational and Environmental Medicine**, v. 52, p. 28–32, 1995.
- 3 FEARON, E. R.; VOGELSTEIN, B. A genetic model for colorectal tumorigenesis. **Cell**, v. 61, n. 5, p.
4 759–767, 1 jun. 1990.
- 5 FEHILY, S. R. et al. The gut microbiota and gut disease. **Internal Medicine Journal**, v. 51, n. 10,
6 p. 1594–1604, 1 out. 2021.
- 7 FEHLMANN, T. et al. Web-based NGS data analysis using miRMaster: A large-scale meta-
8 analysis of human miRNAs. **Nucleic Acids Research**, v. 45, n. 15, p. 8731–8744, 2017.
- 9 FEHLMANN, T. et al. miRMaster 2.0: multi-species non-coding RNA sequencing analyses at
10 scale. **Nucleic Acids Research**, v. 49, n. W1, p. W397–W408, 2 jul. 2021.
- 11 FEHLMANN, T.; MEESE, E.; KELLER, A. Exploring ncRNAs in Alzheimer’s disease by miRMaster.
12 **Oncotarget**, v. 8, n. 3, p. 3771–3772, 2017.
- 13 FIDELLE, M. et al. Resolving the Paradox of Colon Cancer Through the Integration of Genetics,
14 Immunology, and the Microbiota. **Frontiers in Immunology**, v. 0, p. 3209, 14 dez. 2020.
- 15 FILIPOWICZ, W.; BHATTACHARYYA, S. N.; SONENBERG, N. Mechanisms of post-transcriptional
16 regulation by microRNAs: Are the answers in sight? **Nature Reviews Genetics**, v. 9, n. 2, p.
17 102–114, 2008.
- 18 FRIEDLÄNDER, M. R. et al. MiRDeep2 accurately identifies known and hundreds of novel
19 microRNA genes in seven animal clades. **Nucleic Acids Research**, v. 40, n. 1, p. 37–52, 2012.
- 20 FRIEDLÄNDER, M. R. et al. Evidence for the biogenesis of more than 1,000 novel human
21 microRNAs. **Genome Biology**, v. 15, n. 4, p. 1–17, 2014.
- 22 FROMM, B. et al. Substantial Loss of Conserved and Gain of Novel MicroRNA Families in
23 Flatworms. **Molecular Biology and Evolution**, v. 30, n. 12, p. 2619–2628, 1 dez. 2013.
- 24 GILL, S. R. et al. Metagenomic Analysis of the Human Distal Gut Microbiome. **Science (New
25 York, N.Y.)**, v. 312, n. 5778, p. 1355, 2 jun. 2006.
- 26 GOLDMAN, M. J. et al. **Visualizing and interpreting cancer genomics data via the Xena
27 platform. Nature Biotechnology** Nature Publishing Group, , 22 maio 2020. Disponível em:
28 <<https://www.nature.com/articles/s41587-020-0546-8>>. Acesso em: 24 out. 2021
- 29 GU, Z.; EILS, R.; SCHLESNER, M. Complex heatmaps reveal patterns and correlations in
30 multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847–2849, 15 set. 2016.
- 31 GUINNEY, J. et al. The consensus molecular subtypes of colorectal cancer. **Nature Medicine**, v.
32 21, n. 11, p. 1350–1356, 2015.
- 33 HANAHAN, D. Hallmarks of Cancer: New Dimensions. **Cancer Discovery**, v. 12, n. 1, p. 31–46, 1
34 jan. 2022.

- 1 HANAHAN, D.; WEINBERG, R. A. The hallmarks of cancer. **Cell**, v. 100, n. 1, p. 57–70, 7 jan.
2 2000.
- 3 HANAHAN, D.; WEINBERG, R. A. Hallmarks of cancer: The next generation. **Cell**, v. 144, n. 5, p.
4 646–674, 2011.
- 5 HU, W. et al. Multi-omics approach reveals distinct differences in left- and right-sided colon
6 cancer. **Molecular Cancer Research**, v. 16, n. 3, p. 476–485, 1 mar. 2018.
- 7 INCA; INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA - INCA. **Incidência de**
8 **Câncer no Brasil 2023 - Neoplasia maligna do cólon e reto (taxas ajustadas) — Português**
9 **(Brasil)**. [s.l.: s.n.]. Disponível em: <[https://www.gov.br/inca/pt-](https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estimativa/por-neoplasia-taxas-ajustadas/colon-reto)
10 [br/assuntos/cancer/numeros/estimativa/por-neoplasia-taxas-ajustadas/colon-reto](https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estimativa/por-neoplasia-taxas-ajustadas/colon-reto)>. Acesso
11 em: 28 nov. 2022.
- 12 JOHNCILLA, M.; YANTISS, R. K. **Histology of Colorectal Carcinoma: Proven and Purported**
13 **Prognostic Factors. Surgical Pathology Clinics** W.B. Saunders, , 1 set. 2020.
- 14 JUNG, K. U.; KIM, H. O.; KIM, H. Epidemiology, Risk Factors, and Prevention of Colorectal
15 Cancer-An English Version. **Journal of the anus, rectum and colon**, v. 6, n. 4, p. 231–238, 2022.
- 16 KANDIMALLA, R. et al. Genome-wide discovery and identification of a novel miRNA signature
17 for recurrence prediction in stage II and III colorectal cancer. **Clinical Cancer Research**, v. 24, n.
18 16, p. 3867–3877, 2018.
- 19 KIM, H. et al. Dietary mercury intake and colorectal cancer risk: A case-control study. **Clinical**
20 **Nutrition**, v. 39, n. 7, p. 2106–2113, 1 jul. 2020.
- 21 KOZOMARA, A.; BIRGAOANU, M.; GRIFFITHS-JONES, S. miRBase: from microRNA sequences to
22 function. **Nucleic Acids Research**, v. 47, n. D1, p. D155–D162, 8 jan. 2019.
- 23 KULESHOV, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server
24 2016 update. **Nucleic Acids Research**, v. 44, n. W1, p. 90–97, 8 jul. 2016.
- 25 KUMAR, R. et al. HumCFS: A database of fragile sites in human chromosomes. **BMC Genomics**,
26 v. 19, n. 9, p. 1–8, 18 abr. 2019.
- 27 LAGANÀ, A. et al. Variability in the Incidence of miRNAs and Genes in Fragile Sites and the Role
28 of Repeats and CpG Islands in the Distribution of Genetic Material. **PLOS ONE**, v. 5, n. 6, p.
29 e111166, 2010.
- 30 LAI, S. H. et al. PDE4 Subtypes in Cancer. **Oncogene**, v. 39, n. 19, p. 3791, 7 maio 2020.
- 31 LEE, E.; CHEUNG, J.; BIALKOWSKA, A. B. Krüppel-like Factors 4 and 5 in Colorectal
32 Tumorigenesis. **Cancers**, v. 15, n. 9, 1 maio 2023.
- 33 LEE, M. S.; MENTER, D. G.; KOPETZ, S. Right versus left colon cancer biology: Integrating the
34 consensus molecular subtypes. **JNCCN Journal of the National Comprehensive Cancer**
35 **Network**, v. 15, n. 3, p. 411–419, 1 mar. 2017.

- 1 LEROYER, A. et al. Cancer mortality and chemical exposure in a retrospective zinc and lead
2 smelter cohort: A 48-year follow-up. **International Journal of Hygiene and Environmental**
3 **Health**, v. 242, p. 113955, 1 maio 2022.
- 4 LI, X. et al. Oncogenic transformation of diverse gastrointestinal tissues in primary organoid
5 culture. **Nature medicine**, v. 20, n. 7, p. 769, 2014.
- 6 LIANG, C. et al. Recent advances in the diagnostic and therapeutic roles of microRNAs in
7 colorectal cancer progression and metastasis. **Frontiers in Oncology**, v. 12, p. 911856, 13 out.
8 2022.
- 9 LIU, W.; WANG, X. Prediction of functional microRNA targets by integrative modeling of
10 microRNA binding and target expression data. **Genome Biology**, v. 20, n. 1, p. 1–10, 22 jan.
11 2019.
- 12 LLOSA, N. J. et al. The vigorous immune microenvironment of microsatellite instable colon
13 cancer is balanced by multiple counter-inhibitory checkpoints. **Cancer Discov**, v. 5, n. 1, p. 43–
14 51, 2015.
- 15 LONDIN, E. et al. Analysis of 13 cell types reveals evidence for the expression of numerous
16 novel primate- and tissue-specific microRNAs. **Proceedings of the National Academy of**
17 **Sciences**, v. 112, n. 10, p. E1106–E1115, 2015.
- 18 LOREE, J. M. et al. Classifying colorectal cancer by tumor location rather than sidedness
19 highlights a continuum in mutation profiles and consensus molecular subtypes. **Clinical Cancer**
20 **Research**, v. 24, n. 5, p. 1062–1072, 2018.
- 21 LULLI, M. et al. Role of Non-Coding RNAs in Colorectal Cancer: Focus on Long Non-Coding
22 RNAs. **International Journal of Molecular Sciences 2022, Vol. 23, Page 13431**, v. 23, n. 21, p.
23 13431, 3 nov. 2022.
- 24 LUNA BUITRAGO, D.; LOVERING, R. C.; CAPORALI, A. Insights into Online microRNA
25 Bioinformatics Tools. **Non-Coding RNA 2023, Vol. 9, Page 18**, v. 9, n. 2, p. 18, 6 mar. 2023.
- 26 MACÊDO, A. K. S. et al. Histological and molecular changes in gill and liver of fish (*Astyanax*
27 *lacustris* Lütken, 1875) exposed to water from the Doce basin after the rupture of a mining
28 tailings dam in Mariana, MG, Brazil. **Science of the Total Environment**, v. 735, 15 set. 2020.
- 29 MARTINEZ, V. D. et al. Unique somatic and malignant expression patterns implicate PIWI-
30 interacting RNAs in cancer-type specific biology. **Scientific Reports**, v. 5, 2015.
- 31 MARTINEZ, V. D. et al. Discovery of previously undetected microRNAs in mesothelioma and
32 their use as tissue-of-origin markers. **American Journal of Respiratory Cell and Molecular**
33 **Biology**, v. 61, n. 2, p. 266–268, 1 ago. 2019.
- 34 MATHELIER, A.; CARBONE, A. Large scale chromosomal mapping of human microRNA
35 structural clusters. **Nucleic Acids Research**, v. 41, n. 8, p. 4392–4408, 1 abr. 2013.

- 1 METSALU, T.; VILO, J. ClustVis: a web tool for visualizing clustering of multivariate data using
2 Principal Component Analysis and heatmap. **Nucleic Acids Research**, v. 43, n. W1, p. W566–
3 W570, 1 jul. 2015.
- 4 MINATEL, B. C. et al. Large-scale discovery of previously undetected microRNAs specific to
5 human liver. **Human Genomics**, v. 12, n. 1, p. 16, 2018.
- 6 MIRCETA, M. et al. Fragile sites, chromosomal lesions, tandem repeats, and disease. **Frontiers**
7 **in Genetics**, v. 13, 17 nov. 2022.
- 8 MO YANG, K. et al. Does the Different Locations of Colon Cancer Affect the Oncologic
9 Outcome? A Propensity-Score Matched Analysis. **Annals of Coloproctology**, p. 2287–9714,
10 2019.
- 11 MOJARAD, E. N. et al. The CpG island methylator phenotype (CIMP) in colorectal cancer.
12 **Gastroenterology and Hepatology from Bed to Bench**, v. 6, n. 3, p. 120–128, 2013.
- 13 MOTA, P. J. et al. Prevalence of metal levels above the reference values in a municipality
14 affected by the collapse of a mining tailings dam: Brumadinho Health Project. **Revista**
15 **Brasileira de Epidemiologia**, v. 25, n. suppl 2, 2022.
- 16 NI, Z. et al. Analysis of the HNF4A isoform-regulated transcriptome identifies CCL15 as a
17 downstream target in gastric carcinogenesis. **Cancer Biology & Medicine**, v. 18, n. 2, p. 530, 5
18 maio 2021.
- 19 NOETZEL, E. et al. Intermediate filament dynamics and breast cancer: aberrant promoter
20 methylation of the Synemin gene is associated with early tumor relapse. **Oncogene**, v. 29, n.
21 34, p. 4814–4825, 26 ago. 2010.
- 22 OMRANE, I. et al. MicroRNAs 146a and 147b biomarkers for colorectal tumor’s localization.
23 **BioMed Research International**, v. 2014, 2014.
- 24 PALAZZO, A. F.; LEE, E. S. Non-coding RNA: what is functional and what is junk? **Frontiers in**
25 **Genetics**, v. 6, n. JAN, 2015.
- 26 PASCHKE, S. et al. Are colon and rectal cancer two different tumor entities? A proposal to
27 abandon the term colorectal cancer. **International Journal of Molecular Sciences**, v. 19, n. 9,
28 2018.
- 29 PETRELLI, F. et al. Prognostic survival associated with left-sided vs right-sided colon cancer a
30 systematic review and meta-analysis. **JAMA Oncology**, v. 3, n. 2, p. 211–219, 1 fev. 2017.
- 31 PEWARCHUK, M. E. et al. Upgrading the repertoire of miRNAs in gastric adenocarcinoma to
32 provide a new resource for biomarker discovery. **International Journal of Molecular Sciences**,
33 v. 20, n. 22, p. 5697, 14 nov. 2019.
- 34 PINO, M. S.; CHUNG, D. C. The Chromosomal Instability Pathway in Colon Cancer.
35 **Gastroenterology**, v. 138, n. 6, p. 2059–2072, 2010.

- 1 REETESH, P.; BELL, P. D.; PAI, R. K. Immune Response in Colorectal Carcinoma: A Review of Its
2 Significance as a Predictive and Prognostic Biomarker. **Histopathology**, 27 jun. 2022.
- 3 RESHMI, G. et al. Identification and analysis of novel microRNAs from fragile sites of human
4 cervical cancer: Computational and experimental approach. **Genomics**, v. 97, n. 6, p. 333–340,
5 1 jun. 2011.
- 6 ROCK, L. D. et al. Expanding the Transcriptome of Head and Neck Squamous Cell Carcinoma
7 Through Novel MicroRNA Discovery. **Frontiers in Oncology**, v. 9, p. 1305, 27 nov. 2019.
- 8 SAGE, A. P. et al. Expanding the miRNA Transcriptome of Human Kidney and Renal Cell
9 Carcinoma. **International Journal of Genomics**, v. 2018, 2018.
- 10 SALIMINEJAD, K. et al. **An overview of microRNAs: Biology, functions, therapeutics, and**
11 **analysis methods. Journal of Cellular Physiology**Wiley-Liss Inc., , 1 maio 2019a.
- 12 SALIMINEJAD, K. et al. An overview of microRNAs: Biology, functions, therapeutics, and
13 analysis methods. **Journal of Cellular Physiology**, v. 234, n. 5, p. 5451–5465, 1 maio 2019b.
- 14 SEMBA, S. et al. Down-Regulation of PIK3CG, a Catalytic Subunit of Phosphatidylinositol 3-OH
15 Kinase, by CpG Hypermethylation in Human Colorectal Carcinoma. **Clinical Cancer Research**, v.
16 8, n. 12, 2002.
- 17 SHANNON, P. et al. Cytoscape: A software Environment for integrated models of biomolecular
18 interaction networks. **Genome Research**, v. 13, n. 11, p. 2498–2504, nov. 2003a.
- 19 SHANNON, P. et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular
20 Interaction Networks. **Genome Research**, v. 13, n. 11, p. 2498, nov. 2003b.
- 21 SIDDIQUI, R. et al. The Pivotal Role of the Gut Microbiome in Colorectal Cancer. **Biology 2022**,
22 **Vol. 11, Page 1642**, v. 11, n. 11, p. 1642, 9 nov. 2022.
- 23 SLATTERY, M. L. et al. MicroRNAs and colon and rectal cancer: differential expression by tumor
24 location and subtype. **Genes, chromosomes & cancer**, v. 50, n. 3, p. 196–206, 2011.
- 25 SLATTERY, M. L. et al. Site-specific associations between miRNA expression and survival in
26 colorectal cancer cases. **Oncotarget**, 2016.
- 27 SMOLLE, M. A. et al. Noncoding RNAs and immune checkpoints—clinical implications as cancer
28 therapeutics. **The FEBS Journal**, v. 284, n. 13, p. 1952–1966, 1 jul. 2017.
- 29 SOLDEVILLA, B. et al. The correlation between immune subtypes and consensus molecular
30 subtypes in colorectal cancer identifies novel tumour microenvironment profiles, with
31 prognostic and therapeutic implications. **European Journal of Cancer**, v. 123, p. 118–129, 1
32 dez. 2019.
- 33 STANCA MELINCOVICI, C. et al. Vascular endothelial growth factor (VEGF)-key factor in normal
34 and pathological angiogenesis. **Rom J Morphol Embryol**, v. 59, n. 2, p. 455–467, 2018.

- 1 STOFFEL, E. M.; MURPHY, C. C. Epidemiology and Mechanisms of the Increasing Incidence of
2 Colon and Rectal Cancers in Young Adults. **Gastroenterology**, v. 158, n. 2, p. 341–353, 2020.
- 3 SUNG, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality
4 Worldwide for 36 Cancers in 185 Countries. **CA: A Cancer Journal for Clinicians**, v. 71, n. 3, p.
5 209–249, maio 2021.
- 6 SVEEN, A. et al. Colorectal cancer consensus molecular subtypes translated to preclinical
7 models uncover potentially targetable cancer cell dependencies. **Clinical Cancer Research**, v.
8 24, n. 4, p. 794–806, 15 fev. 2018.
- 9 SZCZEPANEK, J.; SKORUPA, M.; TRETYN, A. MicroRNA as a Potential Therapeutic Molecule in
10 Cancer. **Cells**, v. 11, n. 6, 1 mar. 2022.
- 11 SZKLARCZYK, D. et al. The STRING database in 2023: protein–protein association networks and
12 functional enrichment analyses for any sequenced genome of interest. **Nucleic Acids Research**,
13 v. 51, n. D1, p. D638, 1 jan. 2023.
- 14 TAMAS, K. et al. Rectal and colon cancer: Not just a different anatomic site. **Cancer Treatment
15 Reviews**, v. 41, n. 8, p. 671–679, 2015.
- 16 THE CANCER GENOME ATLAS NETWORK. Comprehensive molecular characterization of human
17 colon and rectal cancer. **Nature**, v. 487, n. 7407, p. 330–337, 2012.
- 18 THOMAS, J. et al. MicroRNAs: Clinical relevance in colorectal cancer. **International Journal of
19 Molecular Sciences**, v. 16, n. 12, p. 28063–28076, 2015.
- 20 VAN ROOSBROECK, K.; CALIN, G. A. Cancer Hallmarks and MicroRNAs: The Therapeutic
21 Connection. Em: **Advances in Cancer Research**. [s.l.] Academic Press Inc., 2017. v. 135p. 119–
22 149.
- 23 WANG, C.; FAKIH, M. Targeting KRAS in Colorectal Cancer. **Current Oncology Reports**, v. 23, n.
24 3, p. 1–10, 1 mar. 2021.
- 25 WANG, Y. et al. SSBP2 is an in vivo tumor suppressor and regulator of LDB1 stability.
26 **Oncogene**, v. 29, n. 21, p. 3044, 27 maio 2010.
- 27 WANG, Y. et al. Integrated Analysis of the Functions and Prognostic Values of RNA-Binding
28 Proteins in Colorectal Cancer. **Frontiers in Cell and Developmental Biology**, v. 8, 5 nov. 2020.
- 29 WANG, Y. et al. MiR-17-5p Targets and Downregulates CADM2, Activating the Malignant
30 Phenotypes of Colon Cancer Cells. **Molecular Biotechnology**, v. 64, n. 12, p. 1388–1400, 1 dez.
31 2022.
- 32 WHITTAKER, S. R. et al. Inhibitors of Cyclin-Dependent Kinases as Cancer Therapeutics.
33 **Pharmacology & therapeutics**, v. 173, p. 83, 1 maio 2017.
- 34 WIGGINS, A. et al. MetastamiRs: The Role of MicroRNAs in the Metastatic Phenotype of
35 Prostate Cancer. Em: SEGI, C. M. (Ed.). **Metastasis**. [s.l.] Exon Publications, 2022.

1 WIRBEL, J. et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that
2 are specific for colorectal cancer. **Nature medicine**, v. 25, n. 4, p. 679, 1 abr. 2019.

3 WU, C. W. et al. A Comprehensive Approach to Sequence-oriented IsomiR annotation
4 (CASMiR): Demonstration with IsomiR profiling in colorectal neoplasia. **BMC Genomics**, v. 19,
5 n. 1, p. 401, 25 dez. 2018.

6 XI, Y. et al. Analysis of prognosis, genome, microbiome, and microbial metabolome in different
7 sites of colorectal cancer. **Journal of Translational Medicine**, v. 17, n. 1, p. 353, 29 out. 2019.

8 XIONG, Q. et al. Small Non-Coding RNAs in Human Cancer. **Genes 2022, Vol. 13, Page 2072**, v.
9 13, n. 11, p. 2072, 9 nov. 2022.

10 YAN, H.; BU, P. Non-coding RNA in cancer. **Essays in Biochemistry**, v. 65, p. 625–639, 2021.

11 YANG, C. Y. et al. Outcomes of right-sided and left-sided colon cancer after curative resection.
12 **Scientific Reports**, v. 12, n. 1, 1 dez. 2022.

13 YANG, L. et al. MiRNA and mRNA Integration Network Construction Reveals Novel Key
14 Regulators in Left-Sided and Right-Sided Colon Adenocarcinoma. **BioMed Research**
15 **International**, v. 2019, p. 1–9, 2019.

16 YANG, S. Y.; CHO, M. S.; KIM, N. K. **Difference between right-sided and left-sided colorectal**
17 **cancers: from embryology to molecular subtype. Expert Review of Anticancer Therapy**Taylor
18 and Francis Ltd, , 3 abr. 2018.

19 ZHU, X. et al. miR-152-3p Affects the Progression of Colon Cancer via the KLF4/IFITM3 Axis.
20 **Computational and Mathematical Methods in Medicine**, v. 2020, 2020.

21

22

23

24

25

26

27

28

29

30

31

32

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Capítulo II

1

2 Article

3 **Discovery of novel miRNAs in colorectal cancer:**
4 **potential biological roles and clinical utility.**5 **Iael W. Minutentag^{1,2,*}, Ana Laura Seneda^{1,2}, Matheus C. Barros-Filhos³, Márcio de Carvalho**
6 **⁴, Vanessa G. P. de Souza^{2,5}, Cláudia N. Hasimoto¹, Marcelo P. T. Moraes^{1,6}, Fabio A. Marchi**
7 **⁷, Wan L Lam⁸, Patricia P. dos Reis^{1,2}, Sandra A. Drigo.^{1,2*}**

8 São Paulo State University (UNESP), Medical School, Department of Surgery and Orthopaedics, Botucatu, SP, Brazil.

9 ² São Paulo State University (UNESP), Experimental Research Unity (UNIPLEX), Botucatu, SP, Brazil.10 ³ AC Camargo (Centro Internacional de Pesquisa (CIPE) - A. C. Camargo Cancer Center, São Paulo, SP, Brazil11 ⁴ São Paulo State University (UNESP), School of Veterinary Medicine and Animal Science, Botucatu, SP, Brazil12 ⁵ São Paulo State University (UNESP), Institute of Biosciences, Department of Genetics, Botucatu, SP, Brazil.13 ⁶ São Paulo State University (UNESP), Medical School, Department of Pathology, Botucatu, SP, Brazil.14 ⁷ Department of Head and Neck Surgery, Medical School and São Paulo State Cancer Institute (ICESP), University of São
15 Paulo (USP), São Paulo, SP, Brazil.16 ⁸ Department of Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, BC, Canada.

17

18 * Correspondence: SAD: sandra.d.linde@unesp.br

19 **Abstract:** Deregulated miRNAs are associated with colorectal cancer (CRC), with alterations depending on the
20 tumor location. Novel tissue-specific miRNAs have been identified in different tumors and are associated with
21 cancer. We used miRMaster to identify novel miRNAs in CRC from the TCGA and GEO data (discovery and
22 validation groups). We used TCGA data from five tissues to analyze miRNA tissue specificity. miRDB was used
23 to predict miRNA targets and the UCSC Xena Browser was used to evaluate target expression. After successive
24 analyses, we identified 15 novel miRNAs that were differentially expressed in CRC in both the discovery and
25 validation groups. Four molecules (nov-miR-13844-5p, nov-miR-7154-5p, nov-miR-5035-3p, and nov-miR-590-
26 5p) were differentially expressed in proximal and distal CRC. Nov-miR-3345-5p and nov-miR-13172-3p, which
27 are upregulated in tumors, are only expressed in colorectal tissues. These molecules have been linked to a worse
28 prognosis in right-sided colon and rectal carcinomas. Analysis revealed an association between eight novel
29 miRNAs and 81 targets, mostly cancer-related genes, with varying expression based on tumor location. These

30 findings provide new miRNAs with potential biological relevance, molecular
31 biomarkers, and therapeutic targets for CRC treatment.

Academic Editor: Firstname

Lastname

Received: date

Revised: date

Accepted: date

Published: date

32

33

34

35

32 **Keywords:** colorectal cancer; tumor location; novel microRNA; miRMaster; tissue
33 specificity; prognosis; small-RNAseq.

35 **1. Introduction**

36 Colorectal cancer (CRC) is the third most common cancer and the second
37 leading cause of cancer-related deaths (SUNG et al., 2021). In Brazil, CRC is the
38 second most common cancer in both men and women, with an estimated 45,630
39 new cases nationwide by the year 2023 (INCA; INSTITUTO NACIONAL DE
40 CÂNCER JOSÉ ALENCAR GOMES DA SILVA - INCA, 2022). CRC is a
41 heterogeneous disease with a large amount of evidence showing that proximal
42 and distal CRCs are distinct entities, with different embryological origins,

37 **Copyright:** © 2023 by the author.

38 Submitted for possible open-access

39 publication under the terms and

40 conditions of the Creative Commons

41 Attribution (CC BY) license

42 (<https://creativecommons.org/licenses/by/4.0/>).



1 characteristic immune profiles, genetic alterations, and different clinical
2 outcomes (THE CANCER GENOME ATLAS NETWORK, 2012)

3 Non-coding RNAs (ncRNAs) have historically been considered merely intermediaries
4 between DNA and proteins, except for ribosomal and transfer RNA. However, our transcriptome
5 is composed of only 2% coding RNA, whereas the remaining 98% consists of non-coding RNAs
6 that play diverse roles in post-transcriptional regulation (PALAZZO; LEE, 2015). MicroRNAs
7 (miRNAs or miRs) are small ncRNA molecules with approximately 22nt in length associated with
8 epigenetic silencing, participating in various cellular processes, such as growth, differentiation,
9 development, and apoptosis (SALIMINEJAD et al., 2019b) Numerous studies have provided
10 evidence of miRNAs involvement in colorectal carcinogenesis, including tumor initiation,
11 progression and metastatic spreading, revealing potential diagnostic and prognostic biomarkers
12 for CRC as well as therapeutic targets (LIANG et al., 2022)

13 Currently, nearly 2,800 human miRNAs are annotated in public repositories; however, more
14 recent studies have shown a higher number of miRNAs expressed in the human genome than
15 previously estimated. Human miRNA repertoire characterization was mostly limited to low-
16 coverage techniques used in early studies, detection of miRNA sequences that are conserved
17 across species and different tissue types, and the predominant detection of more abundant
18 transcripts (FRIEDLÄNDER et al., 2014; LONDIN et al., 2015). However, recent analyses using
19 next-generation sequencing (NGS) have revealed new human miRNAs with higher tissue
20 specificity and lower expression levels than known annotated miRNAs (LONDIN et al., 2015).
21 Using advanced bioinformatics tools for NGS of small RNAs (small-RNaseq), 2,469 new
22 candidates for human miRNAs have been reported, of which 1,098 have been validated
23 (FRIEDLÄNDER et al., 2014). Other studies have reported novel miRNAs associated with
24 different tumor types, such as head and neck cancer (ROCK et al., 2019), gastric adenocarcinoma
25 (PEWARCHUK et al., 2019), papillary thyroid carcinoma (BARROS-FILHO et al., 2019), and
26 cervical cancer (RESHMI et al., 2011). Taken together, these data indicate a greater diversity of
27 human miRNAs than previously anticipated, revealing tissue-specific regulatory networks
28 associated with these novel molecules relevant to tumor initiation and progression and extending
29 the miRNA repertoire for molecular biomarkers and drug targets in cancer.

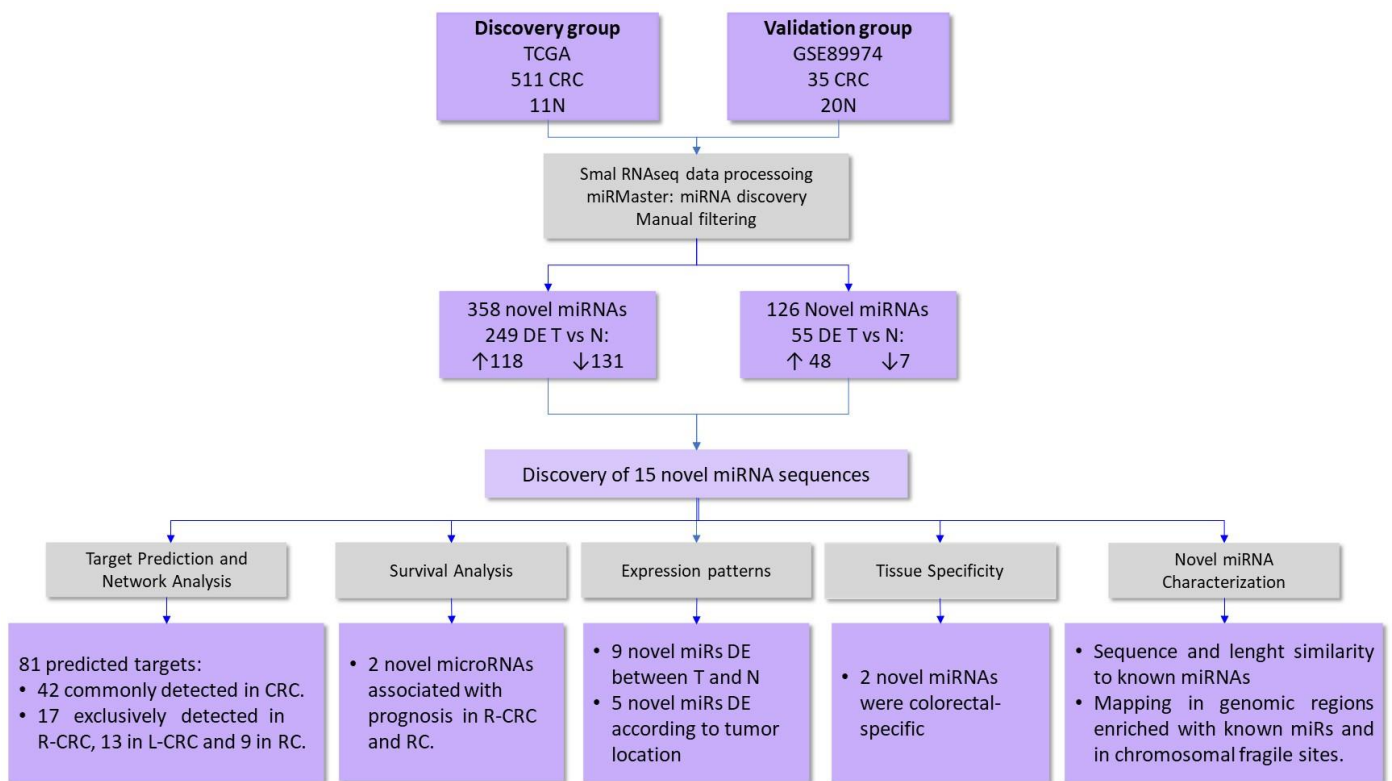
30 Owing to their high tissue specificity and role in tumorigenesis, miRNAs have also been
31 indicated as molecular biomarkers for diagnosing cancer and predicting patient outcomes.
32 Dysregulation of miRNAs has been reported in CRC and plays an important role in its
33 development and metastasis. Despite recent high-throughput and highly sensitive sequencing
34 methods that have allowed for the discovery of novel miRNAs in different tumors with potential
35 clinical utility, there is a paucity of studies addressing unannotated tissue-specific miRNAs in
36 CRC. Thus, in this study, we used large-scale analysis of high-throughput sequencing data to
37 discover novel miRNAs with a role in colorectal carcinogenesis and evaluated their expression
38 patterns according to tumor location, prognosis, and putative target genes.

39 2. Results

40 2.1. Identification of novel microRNAs

41 Using a custom discovery pipeline followed by miRMaster analysis, we identified 358 novel
42 microRNA candidates (without overlap with miRBase-annotated miRNAs) in a discovery group
43 composed by a large cohort of colorectal samples (N=522) from The Cancer Genome Atlas (TCGA
44 (Figure 1, Supplementary Table S1). By comparing CRC (N=511) and normal (N=11) tissues, 249
45 out of 358 unannotated miRNAs were detected as differentially expressed (DE) between tumor
46 and normal samples ($p < 0.05$) (Figure 1).

1
2
3



17 **Figure 1.** Flowchart describing the study design and main findings. Raw small RNA sequencing data from the
 18 discovery group (511 CRC and 11 N tissues) from the TCGA database and the validation group from the GEO database
 19 (35 CRC and 20 N, GSE89974) were processed and analyzed using the miRMaster tool. Curate analysis revealed 358
 20 and 126 miRNA candidates in the discovery and validation groups, respectively. By comparing tumor versus normal
 21 tissues, 249 and 55 DE miRNA candidates were detected in CRC samples from the discovery and validation groups,
 22 respectively. Fifteen novel DE miRNAs were detected in both groups with the same expression patterns (upregulated
 23 or downregulated in tumor samples). The biological significance of these novel miRNA candidates is supported by
 24 similarities in structural and genomic mapping with known miRNAs, colorectal tissue specificity, association with
 25 tumor location and prognosis, and correlation with predicted cancer-related target genes. CRC, colorectal cancer; T,
 26 tumor; N, adjacent normal tissues; DE, differentially expressed; R-CRC: Right-sided colorectal cancer; L-CRC, left-
 27 sided colorectal cancer; RC, rectal cancer.

28

29 After filtering, we evaluated the CG content, sequence length, folding structures, and
 30 genomic distribution of the miRNA candidates relative to annotated miRNAs to verify the
 31 likelihood of our candidates being real novel miRNAs. All miRNA candidates showed similar
 32 features to annotated miRNAs (Figures S1A and B), including chromosomal location within
 33 known miRNA clusters (Figure S1C), thus supporting their identity as true miRNA sequences.

34

35 2.1.1. In Silico Validation of the novel miRNAs

36 To validate the expression of miRNA candidates in CRC, an independent set of colorectal
 37 samples (N=55; GSE89974) was analyzed following the same pipeline and parameters used for

1 the discovery group. A total of 126 putative novel miRNAs were identified. Although the
 2 validation group consisted of a limited number of CRC samples (N=35), we were able to detect
 3 55 DE miRNAs in CRC tissues compared to normal tissues (N=20) (Figure 1). By comparing the
 4 miRNAs identified in TCGA and GSE89974 samples, 21 novel miRNAs were identified in both
 5 groups (Supplementary Table S1). Fifteen out of 21 candidate miRNAs showed similar expression
 6 patterns between tumor and normal tissues in both discovery and validation groups, with seven
 7 novel miRNAs showing increased expression and 8 decreased expression in tumors (Figure 1 and
 8 Table 1), suggesting that these candidates are likely to be true miRNAs with a role in colorectal
 9 carcinogenesis. Therefore, only 15 validated miRNAs were used in the subsequent analysis
 10 (Figure 1).

11
 12

13 **Table 1** Expression patterns of 21 candidates for new miRNAs identified in both discovery (TCGA) and validation
 14 (GSE89974) groups. In bold, 15 candidates with concordant fold changes (FC) in both groups.

Novel miRNA	FC		FC	
	Discovery Group (log(X+1.2))	p-value	Validation Group (log(X+1.2))	p-value
nov-miR-13844-5p	-1	4.15E-31	2,4	1.83E-01
nov-miR-766-3p	25,753	2.33E-07	1,5	9.21E-02
nov-miR-7154-5p	6,561	2.49E-05	3,8	1.12E-06
nov-miR-13172-3p	5,651	4.32E-04	1,9	2.12E-03
nov-miR-3345-5p	4,558	4.21E-05	1,5	2.72E-02
nov-miR-2199-5p	2,002	5.69E-01	3,2	1.18E-04
nov-miR-8861-5p	1,732	1.41E-01	2,6	5.26E-03
nov-miR-5035-3p	0,131	4.19E-05	0,6	2.74E-01
nov-miR-5065-5p	0,121	4.43E-04	0,7	2.23E-01
nov-miR-1156-3p	0,201	3.05E-03	1,0	9.41E-01
nov-miR-8999-3p	0,032	1.94E-02	0,4	8.76E-02
nov-miR-1156-5p	0,674	1.49E-01	1,0	9.41E-01
nov-miR-651-5p	0,671	7.35E-01	0,9	2.70E-01
nov-miR-590-5p	0,476	7.22E-01	0,6	5.45E-02
nov-miR-13996-5p	0,871	8.32E-01	0,6	5.07E-01
nov-miR-9377-5p	15,757	3.46E-08	1,1	5.89E-01
nov-miR-6348-3p	0,789	2.18E-01	5,0	2.10E-04
nov-miR-13634-5p	2,784	1.77E-03	0,8	4.09E-01
nov-miR-171-3p	0,450	4.80E-03	3,0	8.55E-02
nov-miR-4285-5p	0,172	3.42E-08	1,4	1.24E-01
nov-miR-2549-5p	0,054	1.34E-04	1,3	7.89E-01

15 ¹ The candidate nov-miR-13844-5p was not detected in adjacent normal samples from the
 16 discovery group, although it was detected in normal samples from the discovery group.

17
 18

19 2.2 Genomic Location and Tissue-Specific Expression Patterns of Novel miRNAs

20
 21

2.2.1 Genomic Location

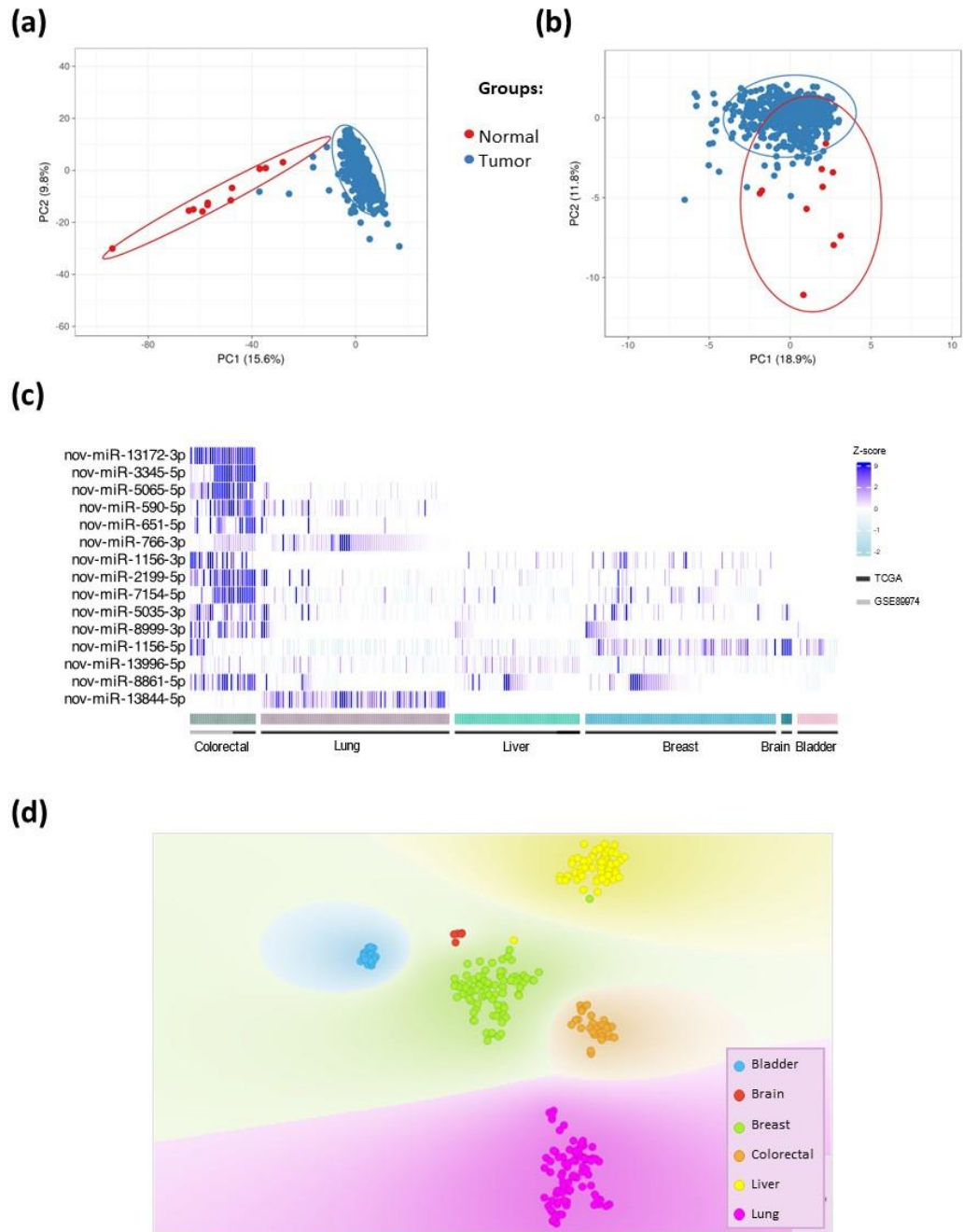
1 We evaluated the distribution of known miRNAs, candidate miRNAs, and predicted targets
2 in the genome by using a circular plot. Interestingly, some candidates were mapped to regions
3 rich in known miRNAs: nov-miR-13996-5p (chrX+:7187376–7187444), nov-miR-7154-5p (chr20-
4 :38425272–38425332), nov-miR-5035-5p (chr19+:12952630–12952689), nov-miR-1156-5p/3p (chr11-
5 :1477615–1477671), nov-miR-13172-3p (chr9+:115489101–115489152), and nov-miR-8861-5p (chr2-
6 :69520113–69520174) (Supplementary Figure S1 and Table S1). These data suggest that these
7 miRNA candidates might be members of known miRNA clusters.

8 We also assessed whether our 15 novel miRs were localized in chromosomal fragile sites;
9 three candidates are transcribed from these sites: nov-miR-8861-5p on fragile site FRA2E
10 (chr2p13: 68600001-75000000), nov-miR-766-3p on fragile site FRA11C (chr11p15.1: 16200001-
11 21700000) and nov-miR-3345-5p on fragile site FRA15A (chr15q22.2: 59100001-67500000). All
12 three fragile sites were classified as having a common frequency and can be induced by the
13 antibiotic aphidicolin.

14 2.2.2 Tissue-Specific Expression Patterns of Novel miRNAs

15 We investigated whether novel miRNA candidates can be used to differentiate tumors from
16 normal samples. Principal component analysis (PCA) was performed using known and novel
17 miRNAs identified in TCGA group (Figures 2A and 2B, respectively). Different expression
18 patterns of both known and novel miRNAs were observed between tumor and normal samples,
19 showing separate sample groups.
20

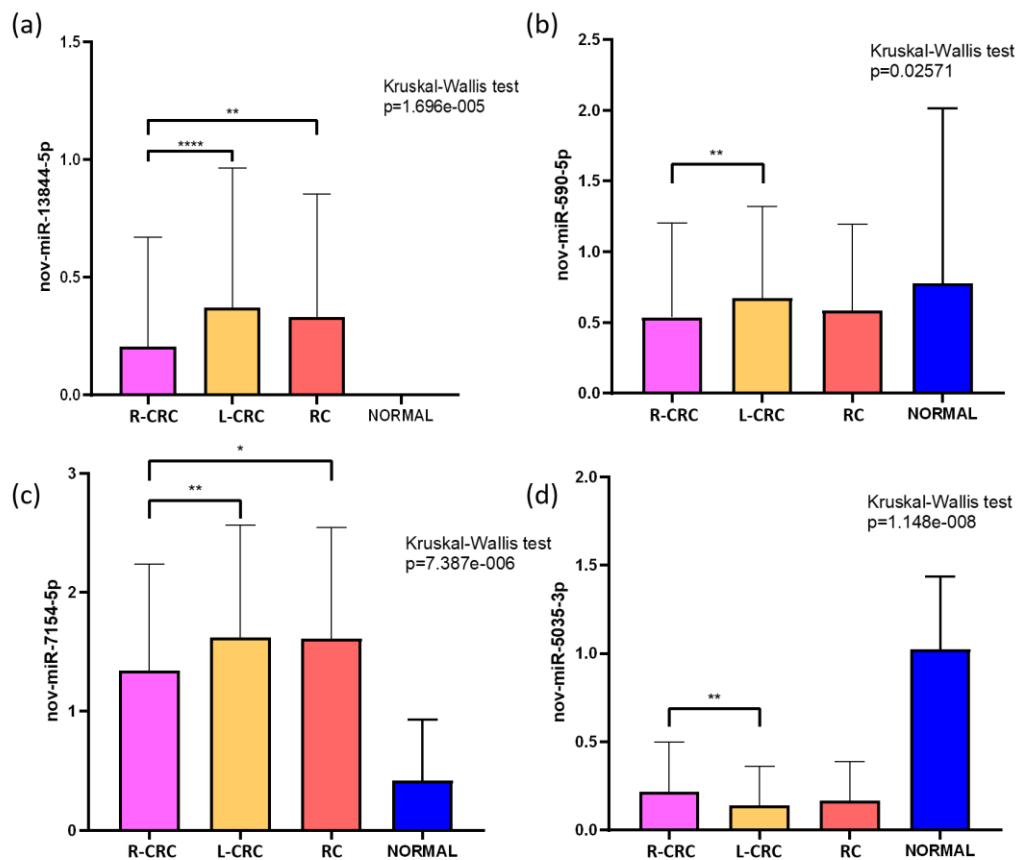
21 To determine whether the novel miRNAs exhibited organ-specific expression patterns,
22 we examined 15 validated novel miRNAs in normal colorectal samples from TCGA and
23 GSE89974 groups (N=31) and normal samples from five different tissues obtained from TCGA
24 (Figure 2C). Different expression patterns of the miRNA candidates were observed in the
25 samples. Interestingly, two out of 15 novel miRNAs were exclusively expressed in colorectal
26 tissues: nov-miR-13172-3p and nov-miR-3345-5p. Conversely, three novel miRNAs were
27 expressed in all tissues: nov-miR-1156-5p, nov-miR-8861-5p, and nov-miR-13996-5p (Figure 2C).
28 In addition, the expression patterns of the 15 novel miRs clearly distinguished colorectal samples
29 from normal samples derived from other tissues (liver, breast, urinary bladder, brain, and lung),
30 indicating tissue-specific expression patterns of these molecules (Figure 2D).
31



1
2 **Figure 2.** Tissue-Specific Expression Patterns of Novel miRNAs (a) PCA of known miRNAs showing a clear
3 separation between tumor and normal samples. (b) PCA showing tumor and normal samples separated according to
4 the expression of novel miRNAs. (c) Heatmap showing the expression patterns of the 15 novel miRNAs in colorectal
5 samples (left) and the five other non-malignant tissue types from TCGA. Two novel miRNAs (*nov-miR-13172-3p* and
6 *nov-miR-3345-5p*) were exclusively found in colorectal samples, whereas three novel miRNAs (*nov-miR-1156-5p*, *nov-*
7 *miR-13996-5p*, and *nov-miR-8861-5p*) were ubiquitously expressed in all tested samples. (d) t-distributed Stochastic
8 Neighbor Embedding (t-SNE) analysis of novel miRNAs in normal colorectal samples (N=31) from TCGA and
9 GSE89974 and normal tissues from five different anatomical sites obtained from TCGA: lung (n=89), breast (n=90),
10 urinary bladder (n=19), liver (n=59), and central nervous system (n=5). The distribution of the samples in the chart
11 demonstrates a cluster of colorectal samples (orange), clearly separated from the other non-malignant tissue types from
12 TCGA, according to the 15 *nov-miR* expression patterns.

1 Next, we investigated whether the novel miRNAs were DE according to the anatomical
 2 tumor location. We compared the expression of 15 novel miRNAs in samples from right-sided
 3 colorectal cancer (R-CRC), left-sided colorectal cancer (L-CRC), and rectal cancer (RC). Four novel
 4 miRNAs were DE according to tumor location. Two of them (nov-miR-13844-5p and nov-miR-
 5 7154-5p) were also significantly more expressed in tumors than in normal tissues (Figure 3),
 6 suggesting an oncogenic role for these novel miRNAs. The candidate miR nov-miR-13844-5p
 7 showed higher expression of L-CRC ($p=4.245e-005$) and RC ($p=0.002683$) than R-CRC samples
 8 (Figure 3A). Higher levels of nov-miR-590-5p were detected in L-CRC tissues than in R-CRC
 9 tissues, although it was not differentially expressed (DE) in tumors compared to normal tissues
 10 ($p=0.0257$, Figure 3B). Increased expression of nov-miR-7154-5p ($p=0.0028$) was detected in L-CRC
 11 ($p=4.245e-005$) and RC ($p=0.002683$) samples compared to that in L-CRC samples (Figure 3C). The
 12 nov-miR-5035-3p downregulated in tumors compared to normal samples showed lower levels in
 13 L-CRC tissues than in R-CRC tissues ($p=0.0294$, Figure 3D).

14



15

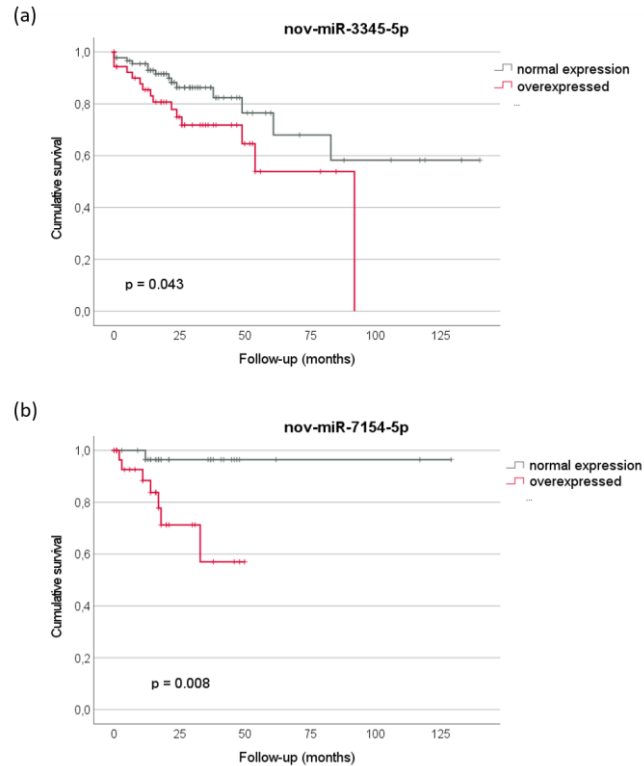
16 **Figure 3.** Expression patterns of the four novel microRNAs differentially expressed (DE) between colorectal
 17 anatomical sites of TCGA samples ($*$ = $p \leq 0.05$, $**$ = $p \leq 0.01$, $***$ = $p \leq 0.001$, $****$ = $p \leq 0.0001$) (a) up-regulation of nov-
 18 miR-13844-5p was detected in tumors compared to normal samples, with higher levels detected in L-CRC and RC
 19 samples. This miRNA was not detected in normal samples from TCGA cohort. (b) nov-miR-590-5p was more highly
 20 expressed in L-CRC; (c) overexpression of nov-miR-7154-5p was observed in tumors compared with normal sam1ples,
 21 with higher levels detected in L-CRC and RC compared to R-CRC tissues; (d) downregulation of nov-miR-5035-3p
 22 was observed in tumors compared to normal samples, with lower levels detected in L-CRC compared to R-CRC samples.
 23 R-CRC, right-sided colorectal cancer; L-CRC, left-sided colorectal cancer; RC, rectal cancer.

24

25

1 2.3. Prognostic relevancy

2 The prognostic value of the 15 novel miRNAs expressed in colorectal tumors was assessed
3 in the TCGA dataset. Two novel miRNAs were significantly associated with overall survival in
4 CRC patients according to the anatomical site (Figure 4). Higher levels of nov-miR-3345-5p in R-
5 CRC were associated with decreased survival ($p=0.043$) (Figure 4A). In addition, overexpression
6 of nov-miR-7154 in RC was associated with worse prognosis ($p=0.008$) (Figure 4B).
7



8
9 **Figure 4.** Cancer-specific survival curves according to the expression levels of (a) nov-miR-3345-5p in R-CRC, and
10 (b) nov-miR-7154-5p in RC, showing significantly decreased survival associated with overexpression of these novel
11 miRNAs in these tumors. log rank test. R-CRC: Right-sided colorectal cancer; RC: rectal cancer.
12

13 2.4. Target prediction and expression in CRC

14 Target prediction using the miRDB platform revealed 2412 target genes regulated by 15
15 novel miRNAs.

16 To refine our predicted target gene list, we evaluated the expression levels of target genes
17 in the discovery cohort available on the Xena Browser platform. Target genes were selected if
18 they were DE between tumor and normal samples, and if they presented opposite expression
19 patterns in relation to their novel miRNA regulators. This analysis revealed 176 target genes
20 regulated by 14 novel miRNAs; 24 genes were upregulated, and 152 were downregulated
21 (Figures 5A and 5B).

22 Next, we analyzed the DE target genes according to their anatomical location in R-CRC, L-
23 CRC, and RC (Figures 5A and B). R-CRC samples showed 141 DE-predicted targets, with 17
24 targets exclusively expressed in these tumors. In L-CRC, 150 DE targets were detected, 13 of
25 which were exclusively detected in L-CRC. In the RC samples, 67 were DE, with 9 exclusively
26 detected in these tumors (Figures 5A and B).
27
28

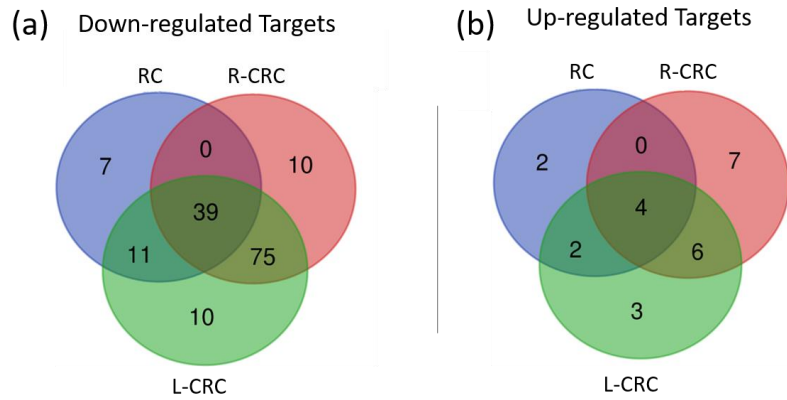


Figure 5. Venn Diagram showing (a) down-regulated and (b) up-regulated predicted targets for Right-sided colorectal cancer (R-CRC), left-sided colorectal cancer (L-CRC), and rectal cancer (RC).

To improve the target prediction, an integrative analysis was performed to select novel miRNA-target gene pairs with a statistically significant negative correlation ($r \leq -0.1$ e $p \leq 0.05$). 81 target genes exhibited a significant correlation with 9 out of 15 candidate miRNAs, with 6 upregulated miRNAs (nov-miR-13172-3p, nov-miR-13844-5p, nov-miR-3345-5p, nov-miR-7154-5p, nov-miR-766-3p, and nov-miR-8861-5p) and three downregulated miRNAs (nov-miR-1156-5p, nov-miR-590-5p, and nov-miR-5065-5p) (Supplementary Table 2).

It is noteworthy that nov-3345-5p and nov-miR-13172-3p, both upregulated in tumors and expressed exclusively in colorectal tissues, showed 21 putative targets (Supplementary Table S1). The upregulated nov-3345-5p, which is also associated with poor prognosis in R-CRC, showed a negative correlation with 10 targets. Among them, *BNC2*, *SYNM*, and *TCF21* target genes were detected in all tumor locations, and three targets (*GRIK2*, *KLHL14*, and *MS4A2*) were shared by R-CRC and L-CRC but not by RC. Candidate nov-miR-13172-3p showed 11 target genes. The correlation of *AFF3* and *DAAM2* with the candidate miRNAs was detected at all tumor locations.

This analysis also revealed 10 miRNA-target pairs exclusively present in R-CRC, 20 exclusively detected in L-CRC, and seven pairs in RC (Supplementary Table 2). Interestingly, nov-miR-8861-5p, which is upregulated in tumors, presented the highest number of targets in all three anatomic locations, showing both shared and exclusive targets for each location.

The candidate nov-miR-7154-5p was upregulated in CRC, particularly in distal tumors (L-CRC and RC), was associated with poor survival in RC, and showed six predicted targets, with the *EPHA7* gene exclusively detected in RC. Three other candidates (nov-miR-13844-5p, nov-miR-5035-3p, and nov-miR-590-5p) with differential expression between proximal and distal CRC also showed significant correlations with their predicted targets.

The downregulated candidates nov-miR-1156-5p, nov-miR-590-5p, and nov-miR-5065-5p showed different targets according to the tumor location. nov-miR-1156-5p showed correlation with *RNF183* exclusively in RC, and nov-miR-590-5p showed correlation with *HOMER1* and *TBX15* in R-CRC and L-CRC, respectively. The nov-miR-5065-5p was predicted to regulate *ERP27* only in L-CRC samples (Supplementary Table S2).

2.5 Interaction Networks

Interaction analysis including the 81 target genes showed that 47 genes presented functional interactions between them. (Figure 6) These genes are regulated by seven candidate miRNAs. *CACNA1B* and *DLG2* showed the highest number of interactions with other genes and were regulated by nov-miR-8861-5p and nov-miR-13172-3p, respectively. The candidate nov-miR-8861-5p emerged as the miRNA that regulates the highest number of genes (32 target genes) with interactions among them. Interestingly, this miRNA regulates unique genes from all three

1 part, by the different analysis strategies used in these studies and the larger number of samples
2 in the validation groups, which allowed the detection of a greater number of candidates. In our
3 study, despite the small sample size in the validation group (35 CRC and 20 normal samples)
4 compared with the discovery group (N=522), 21 miRNAs were identified, with 15 (71.43%)
5 showing consistent expression patterns when comparing tumors to normal tissue samples.

6 To support these novel miRNA transcripts as true positive miRNA sequences, molecular
7 characterization of the novel miRs was performed, which showed similarities with known miRs,
8 such as GC content, size, and genomic location in regions enriched with known miRs and
9 chromosomal fragile sites. We showed that both the known miRNAs and candidates exhibited
10 CG content above 50%, which is in agreement with a previous report showing that known
11 dysregulated miRNAs in CRC exhibit an average CG content above 50%, indicating that cancer-
12 regulated miRNAs might be of the non-conserved type and enriched in cytosine and guanine
13 (AHMED; MAL, 2021).

14 The validated miRNAs were mapped to chromosomes 2, 3, 9, 10, 11, 13, 15, 19, 20, and X,
15 most of which were located in regions rich in known miRNAs. Three novel miRNAs (nov-miR-
16 3345-5p, nov-miR-766-3p, and nov-miR-8861-5p) were detected in known genomic fragile sites,
17 FRA2E, FRA11C, and FRA15A, all of which are classified as common fragile sites (CFS) because
18 they are present in most of the human population (MIRCETA et al., 2022). Fragile sites are gaps,
19 constrictions, or breaks in metaphase chromosomes that arise when cells are exposed to
20 perturbation of the DNA replication process and are linked to regions of chromosomal
21 rearrangement in cancer (DEBACKER; FRANK KOOY, 2007). Fragile sites are also particularly
22 dense in miRNAs, many of which are located within genes commonly translocated in cancers
23 (LAGANÀ et al., 2010). Overall, these data provide additional evidence that these candidates are
24 true-positive miRNAs.

25 In addition, we observed that, similar to observations with known miRNAs, the 15 novel
26 miRNAs detected here could differentiate colorectal tumors from adjacent normal tissues, thus
27 enhancing the repertoire of miRNAs in CRC. Several miRNAs exhibit specific expression patterns
28 in tumors such as lung, breast, brain, liver, and colorectal cancers, making them potentially
29 valuable for the development of new therapies (SZCZEPANEK; SKORUPA; TRETYN, 2022). In
30 addition to this ability to differentiate tumors from normal tissues, it is highly desirable that
31 potential therapeutic targets present tissue specificity. Two out of the 15 candidates detected here,
32 nov-miR-13172-3p and nov-miR-3345-5p, were exclusively expressed in colorectal tissues and are
33 potential candidates for further experimental validation.

34 Among the 15 novel miRNAs detected in colorectal samples, 9 candidates (nov-miR-1156-
35 3p, nov-miR-13172-3p, nov-miR-13844-5p, nov-miR-3345-5p, nov-miR-5035-3p, nov-miR-5065-
36 5p, nov-miR-7154-5p, nov-miR-766-3p, and nov-miR-8999-3p) showed significant differential
37 expression between tumor and normal tissues. The nov-miR-13844-5p was not detected in normal
38 samples from the discovery group, although it was detected in normal samples from the
39 validation group. This suggests that this candidate miRNA may be highly specific to the tumor
40 context, potentially making it a promising target for diagnostic purposes (THOMAS et al., 2015).

41 Several miRNAs have been reported to have distinct expression profiles according to tumor
42 location (COEBERGH VAN DEN BRAAK et al., 2018; ENEH et al., 2020; OMRANE et al., 2014;
43 SLATTERY et al., 2011; YANG et al., 2019). In our study, four candidates (nov-miR-13844-5p, nov-
44 miR-590-5p, nov-miR-7154-5p, and nov-miR-5035-3p) were DE between anatomical sites.
45 Interestingly, similar expression patterns of these candidates were observed in L-CRC and RC
46 compared to R-CRC, corroborating previous reports showing that rectal tumors have a molecular
47 profile that is more closely aligned with L-CRC than with tumors from other intestinal sites
48 (LOREE et al., 2018).

1 To assess whether the 15 novel miRNAs could be associated with CRC prognosis, survival
2 curves were constructed based on their transcript levels in tumors compared to normal samples.
3 Two candidates, nov-miR-3345-5p and nov-miR-7154-5p, both upregulated in tumor samples,
4 were associated with poor prognosis in R-CRC and RC, respectively, suggesting that these novel
5 miRNAs could be considered promising prognostic biomarkers in CRC of different tumor
6 locations. Novel miRNAs were also associated with prognosis in thyroid cancer (BARROS-
7 FILHO et al., 2019), gastric cancer (PEWARCHUK et al., 2019), and head and neck tumors (ROCK
8 et al., 2019), reinforcing that unannotated miRNAs, such as those described here, may play a role
9 in tumor progression and can improve the prognostic risk stratification of colorectal tumors.

10 To better understand the biological role of the novel miRNAs discovered in CRC, target
11 prediction followed by integrative analysis evaluating novel miRNA-target gene pairs was
12 performed. Numerous genes strongly implicated in colorectal carcinogenesis were predicted to
13 be targets of our candidate miRNAs, which will be explored further in depth.

14 Candidates nov-3345-5p and nov-miR-13172-3p, detected here as colorectal-specific
15 miRNAs, showed 21 putative targets. *TCF21* gene showed a negative correlation with nov-3345-
16 5p in tumors from all anatomical locations. *TCF21*, a transcription factor that regulates cellular
17 differentiation and organogenesis, has been detected as a tumor suppressor in CRC, and reduced
18 *TCF21* expression is associated with poor prognosis. This gene mediates inactivation of the
19 PI3K/AKT pathway and inhibits the action of metalloproteinases in CRC (DAI et al., 2017).
20 Additionally, nov-miR-3345-5p was predicted to regulate *SYNM*, which encodes for synemin, a
21 type IV intermediate filament. This gene has been suggested to be a tumor suppressor in breast
22 cancer and is regulated by methylation (NOETZEL et al., 2010). *AFF3* was identified as a putative
23 target of nov-miR-13172-3p in CRC. *AFF3* encodes a nuclear transcriptional activator. *AFF3*
24 among others RNA-binding proteins has been investigated in CRC. Downregulation of *AFF3*
25 has been detected in CRC and is associated with poor prognosis (WANG et al., 2020). Further
26 experimental validation is required to prove the regulation of these genes by nov-miR-3345-5p
27 and nov-miR-13172-3p in colorectal tumors.

28 Four candidates (nov-miR-13844-5p, nov-miR-7154-5p, nov-miR-5035-3p, and nov-miR-590-
29 5p) showed different expression patterns according to tumor location. The candidate nov-miR-
30 13844-5p, identified as upregulated in distal tumors compared with proximal tumors, revealed
31 different targets according to these locations. Possible targets identified for nov-miR-13844-5p in
32 RC included *PDE4D*, whereas *SSBP2*, *KLF4*, and *GABRG2* were identified as potential targets in
33 R-CRC.

34 *PDE4D* belongs to the subfamily 4 of cyclic nucleotide phosphodiesterase and has been
35 implicated in the tumorigenesis of various neoplasms. Decreased expression of this gene has been
36 found up to 30 times lower in samples from patients with chronic lymphocytic leukemia than in
37 samples from healthy adults (LAI et al., 2020). *PDE4D* also plays a significant role in CRC, where
38 reduced expression has been associated with increased proliferation, colony formation,
39 apoptosis, invasion, and migration (CHEN et al., 2018). It has been demonstrated that
40 deregulation of *PDEA4* mediated by miR-203a-3p promotes tumor proliferation, growth, and cell
41 migration in CRC (CHEN et al., 2018). Thus, the newly identified miRNA nov-miR-13844-5p may
42 play a role in CRC by regulating the *PDE4D* gene, particularly in rectal tumors. Further functional
43 studies are required to confirm these findings.

44 *SSBP2* acts as a tumor suppressor by responding to DNA damage and promoting cell growth
45 arrest by blocking the Wnt signaling pathway (CHUNG et al., 2020). Reduced expression of this
46 gene has been associated with worse prognosis in CRC and various other tumors, such as the
47 esophagus, prostate, gallbladder, and acute leukemia (WANG et al., 2010). The candidate nov-
48 miR-13844-5p may play a regulatory role in the expression of this gene, particularly in R-CRC.

49 *KLF4*, a zinc finger transcription factor, plays a role in maintaining homeostasis in the
50 intestinal epithelium (LEE; CHEUNG; BIALKOWSKA, 2023; ZHU et al., 2020). *KLF4* is a tumor
51 suppressor gene associated with cell proliferation, migration, and invasion. The mechanisms

1 underlying *KLF4* deregulation include loss of heterozygosity, hypermethylation, and regulation
2 by miRNAs. Various miRNAs have been implicated in *KLF4* silencing in CRC cells.

3 Taken together, these data indicate that nov-miR-13844-5p might be involved in colorectal
4 carcinogenesis, making it a strong candidate for further validation.

5 The candidate nov-miR-7154-5p was significantly negatively correlated with *CADM2*,
6 *EPHA5*, *CPED1*, *EPHA7*, *BEND6*, and *SERTAD4*. Among these, we highlighted that the tumor
7 suppressor gene *CADM2*, which encodes a cell adhesion molecule, has been associated with
8 prostate cancer, and its main mechanism of dysregulation is mediated by gene hypermethylation
9 (CHANG et al., 2010; SEMBA et al., 2002). It has been reported that downregulation of *CADM2*
10 mediated by miR-17-5p promotes a malignant phenotype in colorectal cancer (WANG et al.,
11 2022). In the current study, *CADM2* was found to be dysregulated in both R-CRC and R-CRC but
12 not in RC.

13
14 The candidate nov-miR-590-5p was downregulated in CRC. Significantly lower levels of this
15 miRNA were detected in L-CRC than in R-CRC tissues. Upregulation of *HOMER1* in L-CRC and
16 *TBX1* in R-CRC, both predicted targets of this miRNA, was observed in these tumors.
17 Overexpression of *HOMER1* has been reported in various tumors, including CRC (CUI et al.,
18 2020). The newly identified miR nov-miR-590-5p may be involved in *HOMER1* deregulation in
19 CRC, particularly in right colon tumors, where a negative correlation was identified between this
20 gene and the new miR. *TBX15*, identified as a target gene in L-CRC, is naturally more highly
21 expressed during fetal development. In cancer, *TBX15* expression appears to be stimulated by
22 NF- κ B, exerting an anti-apoptotic effect and promoting cell proliferation (ARRIBAS et al., 2016).
23 Further studies are needed to ascertain whether *TBX15* acts as an oncogene in CRC, particularly
24 in L-CRC, and whether it is regulated by the newly identified miR nov-miR-590.

25 In conclusion, this study marks a pivotal advancement in our understanding of colorectal
26 cancer biology. The discovery of 15 novel microRNAs (miRNAs) through rigorous analysis of
27 small RNA sequencing data and the miRMaster algorithm underscores their potential
28 significance. Validated against an independent dataset, these miRNAs have emerged as
29 promising candidates for CRC treatment.

30 Their ability to differentiate between tumor and normal tissues offers new prospects for
31 diagnostics. Distinct expression patterns of miRNAs across different colon segments provide
32 nuanced insights into CRC complexity. Additionally, the correlation of specific miRNAs with
33 poor prognosis introduces potential prognostic biomarkers and refines patient care strategies.

34 Exploring miRNA-target gene interactions adds a layer of depth to elucidate potential
35 mechanisms. The predicted associations with key genes involved in colorectal carcinogenesis
36 suggest the impact of these miRNAs. This novel understanding sets the stage for innovative
37 therapeutic approaches.

38 The discovery of these miRNAs highlights the untapped potential of the human genome. As
39 these findings intersect with established cancer biology, they open new avenues for diagnosis,
40 prognosis, and treatment in colorectal cancer research and practice.

42 **4. Materials and Methods**

43 *4.1 Study Cohorts*

44 Discovery Group: Data from small-RNAseq sequencing of 522 colorectal samples (511
45 colorectal tumors and 11 adjacent normal tissues) were obtained from The Cancer Genome Atlas
46 (TCGA) using the Cancer Genomics Hub (<https://portal.gdc.cancer.gov/>; GDC project ID: 24725).
47 Clinical characteristics are summarized in Supplementary Table S3.

1 **Validation Group:** Publicly available small miRNA sequencing data from Gene Expression
2 Omnibus (GEO) DataSets (<http://www.ncbi.nlm.nih.gov/gds>) were used. The search was
3 performed using the following keywords: colorectal cancer, non-coding RNA profiling by high-
4 throughput sequencing, and Homo sapiens. The inclusion criteria were Illumina small RNA
5 sequencing, a minimum of 25 CRC samples, and 5 MB of total reads. The study GSE89974,
6 which followed all criteria, was included as the validation group and contained 35 CRC tissues
7 and 20 normal samples.

8 *4.2 Small-RNA Sequencing Processing and miRNA Discovery*

9 A previously described discovery pipeline was used to identify novel unannotated miRNAs
10 in CRC and normal samples (BARROS-FILHO et al., 2019; MARTINEZ et al., 2019;
11 PEWARCHUK et al., 2019). Briefly, all raw sequences obtained from TCGA and GSE89974
12 datasets were processed. Raw BAM files were converted to the FASTQ format using the
13 "SamToFastq" tool from the Picard analysis package (<http://broadinstitute.github.io/picard/>).
14 Unaligned reads were trimmed based on their Phred quality score (≥ 20) and aligned to the latest
15 version of the human genome (GRCh38) using spliced transcript alignment to a reference (STAR)
16 aligner.

17 The previously unannotated miRNAs were identified using the miRMaster platform
18 (<https://ccb-compute.cs.uni-saarland.de/mirmaster/tutorial/>), which, based on the FASTQ files,
19 searches for new candidate miRNAs, quantifies their expression, and identifies miRNA isoforms
20 and variants. The default parameters of miRMaster were used followed by manual filtering.
21 Sequences with low expression (RPM <1 and expressed in less than 10% of the samples) were
22 excluded. After filtering, the miRNA candidates were considered putative novel miRNAs. Both
23 the discovery group (TCGA) and the validation group (GSE89974) were subjected to the same
24 parameters and analysis (Figure 1).

25 For the identification of novel DE miRNAs in tumors, we considered a significance threshold
26 of $p < 0.05$, and fold change (FC) ≥ 1.2 for up-regulated miRNAs, and $FC \leq 1.1$ for down-regulated
27 miRNAs. We considered validated putative novel miRNAs identified in both cohorts (discovery
28 and validation groups) with the same expression pattern in CRC samples (upregulated or
29 downregulated).

30 Principal Component Analysis (PCA) was performed using the ClustVis platform
31 (METSALU; VILO, 2015) with both known and newly identified miRNAs.

32 *4.3 Novel microRNA Analysis*

33 The validated putative miRNA candidates were further subjected to five different analyses:
34 structural analysis, tissue specificity, differential expression according to tumor location, survival
35 analysis, and target prediction (Figure 1).

36 **Structural Analysis:** The CG content and sequence length of putative novel miRNAs were
37 evaluated using the Graph Pad Prism 8 Program, and the fragile sites database HumCFS
38 (KUMAR et al., 2019) was used to evaluate the genomic location of the miRNA candidates.

39 **Tissue specificity:** The newly identified miRNAs in CRC tissues were compared with those
40 detected in normal tissues from different locations. This analysis included normal samples from
41 TCGA and GSE89974, and adjacent normal tissues from five different anatomical sites obtained
42 from TCGA: the lung (n=89), breast (n=90), urinary bladder (n=19), liver (n=59), and central
43 nervous system (n=5). Novel miRNAs were detected in these samples following the same pipeline

1 as described above from the raw small RNA sequences retrieved from TCGA. A t-distributed
2 Stochastic Neighbor Embedding (t-SNE) nonlinear dimensionality reduction analysis was
3 performed using the Orange Data Mining program (DEMŠAR et al., 2013). A heatmap was
4 constructed using the R ComplexHeatmap package using Z-score (GU; EILS; SCHLESNER, 2016)
5 and finalized in Inkscape (<https://inkscape.org/>).

6 Expression of novel miRNAs according to tumor location: The expression of the validated
7 candidates was analyzed based on tumor location: right colon (cecum, ascending colon, hepatic
8 flexure, and transverse colon), left colon (splenic flexure, descending colon, sigmoid junction, and
9 rectosigmoid junction), and rectum. The Mann-Whitney and Kruskal-Wallis tests were applied.

10 Survival Analysis: Cancer-specific survival curves for each of the validated candidate
11 miRNAs were constructed based on the expression pattern of these miRNAs in tumor samples
12 from TCGA. Expression data were dichotomized into normal and upregulated or downregulated
13 expression according to z-score >2.0 or <-2.0 , respectively, relative to normal tissues. Survival
14 analysis was performed using the Kaplan-Meier method and log-rank test for curve comparison.
15 Analyses were performed using the SPSS v27 program.

16 Prediction of targets regulated by candidate miRNA target genes was performed using the
17 MirTarget tool (LIU; WANG, 2019) from the miRDB platform (CHEN; WANG, 2020) using a score
18 cutoff of ≥ 80 . The expression levels of the target genes were further evaluated in the same tumor
19 samples using RNA-seq data from the Xena Browser (<https://xenabrowser.net/>)(GOLDMAN et
20 al., 2020). Target genes with differential expression in TCGA-COAD (n=567) and TCGA-READ
21 (n=97) were identified by comparing tumor and normal samples, considering $p < 0.05$ and $FC \geq 2$
22 for upregulated and $FC \leq -2$ for downregulated targets. Subsequently, paired miRNA targets
23 were detected using opposite FC directions. In addition, Spearman correlation analysis was used
24 to refine the most probable miRNA-target pairs ($p < 0.05$, $r \leq -0.1$). Target prediction according to
25 tumor location (right colon, left colon, and rectum) was performed following the same analyses.

26 4.4 Gene Network

27 The STRING platform (<https://string-db.org/>) (SZKLARCZYK et al., 2023) was used to identify
28 interactions between target genes. The Cytoscape program (SHANNON et al., 2003b) was used
29 to construct interaction network figures showing target genes and their novel regulatory
30 miRNAs.

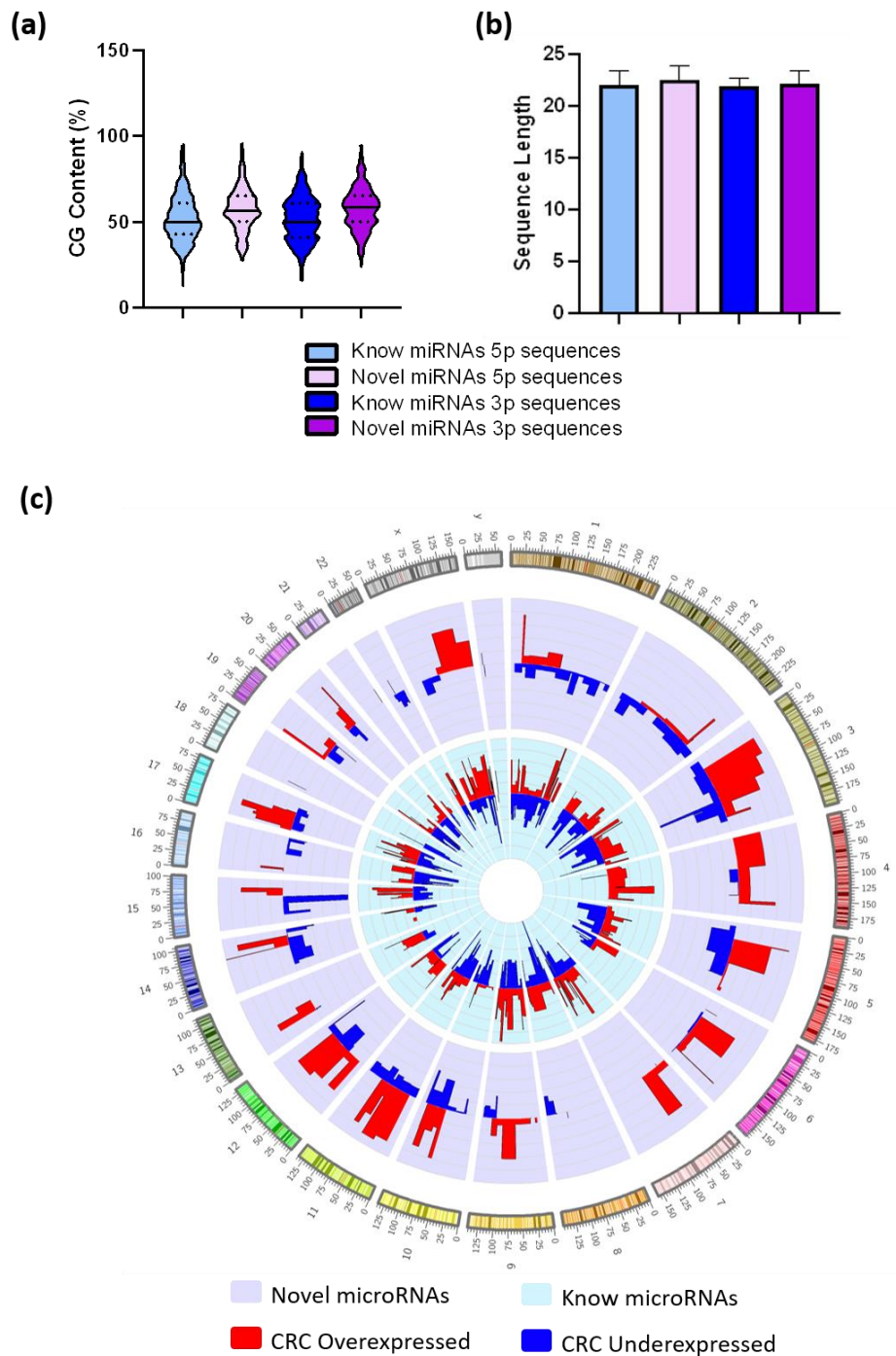
31 **Supplementary Materials:** The following supporting information can be downloaded at:
32 www.mdpi.com/xxx/s1, Figure S1: title; Table S1: title; Video S1: title.

33 **Author Contributions:** For research articles with several authors, a short paragraph specifying their
34 contributions must be provided. The following statements should be used “Conceptualization, SAD. and
35 IWM.; methodology, IWM, MCBF, WLL, FAM, and SAD.; software, IWM, ALS, MTCBF, and FAM.;
36 validation, IWM, MC, CNH, and MPTM; formal analysis, IWM, MCBF, VGPS, and FAM.; resources, PPR
37 and SAD.; data curation, IWM, MCBF, and SAD.; writing—original draft preparation, IWM and SAD;
38 writing—review and editing, IWM, and SAD; visualization, IWM and SAD.; supervision, SAD.; project
39 administration, SAD.; funding acquisition, PPR and SAD. All authors have read and agreed to the published
40 version of the manuscript.

41 **Funding:** This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível
42 Superior – Brasil (CAPES) – Finance Code 001.

43 **Institutional Review Board Statement:** “The study was conducted in accordance with the Declaration of
44 Helsinki, and was approved by the School of Medicine Research Ethics Committee under the # CAAE
45 65306922.3.0000.5411

- 1 **Informed Consent Statement:** Patient consent was waived due to all samples belonging to public databanks
2 in the small RNA-seq data samples.
3
- 4 **Data Availability Statement:** The datasets used in this study are available online.
5 TCGA Data-set: <https://portal.gdc.cancer.gov/repository>
6 GSE89974: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89974>
7
- 8 **Acknowledgments:** In this section, you can acknowledge any support given which is not covered by the
9 author contribution or funding sections. This may include administrative and technical support, or
10 donations in kind (e.g., materials used for experiments).
- 11 **Conflicts of Interest:** The authors declare no conflict of interest.
12



1
2 **Figure S1.** Structural and genomic localization characterization of novel miRNAs. (a) CG content of the sequences
3 identified as novel miRNAs compared to known miRNAs present in the discovery group. (b) Sequence length of
4 sequences identified as novel miRNAs compared with known miRNAs present in the discovery group. (c) Genomic
5 distribution of novel and known miRNAs identified in colorectal tissues and their respective expression levels (log₂
6 tumor/non-malignant ratio). The histogram represents the fold change values (log₂) from the miRNA expression in
7 colorectal adenocarcinoma (CRC) compared to non-malignant samples (bar widths of miRNA genomic positions were
8 adjusted to 10 MB for illustration). A circular illustration of the genomic localization of the novel and known miRNAs
9 was created using Clico (<https://cgdv-upload.persistent.co.in/cgdl/>).

10

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **2021**, *71*, 209–249, doi:10.3322/caac.21660.
2. INCA; Instituto Nacional de Câncer José Alencar Gomes da Silva - INCA *Incidência de Câncer No Brasil 2023 - Neoplasia Maligna Do Cólon e Reto (Taxas Ajustadas) – Português (Brasil)*; 2022;
3. The Cancer Genome Atlas Network Comprehensive Molecular Characterization of Human Colon and Rectal Cancer. *Nature* **2012**, *487*, 330–337, doi:10.1038/nature11252.Comprehensive.
4. Palazzo, A.F.; Lee, E.S. Non-Coding RNA: What Is Functional and What Is Junk? *Front Genet* **2015**, *6*, doi:10.3389/FGENE.2015.00002.
5. Saliminejad, K.; Khorram Khorshid, H.R.; Soleymani Fard, S.; Ghaffari, S.H. An Overview of MicroRNAs: Biology, Functions, Therapeutics, and Analysis Methods. *J Cell Physiol* **2019**, *234*, 5451–5465, doi:10.1002/jcp.27486.
6. Liang, C.; Yang, J.B.; Lin, X.Y.; Xie, B.L.; Xu, Y.X.; Lin, S.; Xu, T.W. Recent Advances in the Diagnostic and Therapeutic Roles of MicroRNAs in Colorectal Cancer Progression and Metastasis. *Front Oncol* **2022**, *12*, 911856, doi:10.3389/FONC.2022.911856/BIBTEX.
7. Friedländer, M.R.; Lizano, E.; Houben, A.J.S.; Bezdan, D.; Báñez-Coronel, M.; Kudla, G.; Mateu-Huertas, E.; Kagerbauer, B.; González, J.; Chen, K.C.; et al. Evidence for the Biogenesis of More than 1,000 Novel Human MicroRNAs. *Genome Biol* **2014**, *15*, 1–17, doi:10.1186/gb-2014-15-4-r57.
8. Londin, E.; Loher, P.; Telonis, A.G.; Quann, K.; Clark, P.; Jing, Y.; Hatzimichael, E.; Kirino, Y.; Honda, S.; Lally, M.; et al. Analysis of 13 Cell Types Reveals Evidence for the Expression of Numerous Novel Primate- and Tissue-Specific MicroRNAs. *Proceedings of the National Academy of Sciences* **2015**, *112*, E1106–E1115, doi:10.1073/pnas.1420955112.
9. Rock, L.D.; Minatel, B.C.; Marshall, E.A.; Guisier, F.; Sage, A.P.; Camargo Barros-Filho, M.; Stewart, G.L.; Garnis, C.; Lam, W.L. Expanding the Transcriptome of Head and Neck Squamous Cell Carcinoma Through Novel MicroRNA Discovery. *Front Oncol* **2019**, *9*, 1305, doi:10.3389/fonc.2019.01305.
10. Pewarchuk, M.E.; Barros-Filho, M.C.; Minatel, B.C.; Cohn, D.E.; Guisier, F.; Sage, A.P.; Marshall, E.A.; Stewart, G.L.; Rock, L.D.; Garnis, C.; et al. Upgrading the Repertoire of MiRNAs in Gastric Adenocarcinoma to Provide a New Resource for Biomarker Discovery. *Int J Mol Sci* **2019**, *20*, 5697, doi:10.3390/ijms20225697.
11. Barros-Filho, M.C.; Pewarchuk, M.; Minatel, B. de C.; Sage, A.P.; Marshall, E.A.; Martinez, V.D.; Rock, L.D.; MacAulay, G.; Kowalski, L.P.; Rogatto, S.R.; et al. Previously Undescribed Thyroid-Specific MiRNA Sequences in Papillary Thyroid Carcinoma. *J Hum Genet* **2019**, *64*, 505–508, doi:10.1038/s10038-019-0583-7.
12. Reshmi, G.; Chandra, S.S.V.; Babu, V.J.M.; Babu, P.S.S.; Santhi, W.S.; Ramachandran, S.; Lakshmi, S.; Nair, A.S.; Pillai, M.R. Identification and Analysis of Novel MicroRNAs from Fragile Sites of Human Cervical Cancer: Computational and Experimental Approach. *Genomics* **2011**, *97*, 333–340, doi:10.1016/J.YGENO.2011.02.010.
13. Sage, A.P.; Minatel, B.C.; Marshall, E.A.; Martinez, V.D.; Stewart, G.L.; Enfield, K.S.S.S.; Lam, W.L. Expanding the MiRNA Transcriptome of Human Kidney and Renal Cell Carcinoma. *Int J Genomics* **2018**, *2018*, doi:10.1155/2018/6972397.

- 1 14. Ahmed, Z.; Mal, C. Functional Role of Hub Molecules in MiRNA and Transcription
2 Factor Mediated Gene Regulatory Network of Colorectal and Lung Cancer. *Gene Rep*
3 **2021**, *23*, doi:10.1016/j.genrep.2021.101129.
- 4 15. Mirceta, M.; Shum, N.; Schmidt, M.H.M.; Pearson, C.E. Fragile Sites, Chromosomal
5 Lesions, Tandem Repeats, and Disease. *Front Genet* **2022**, *13*,
6 doi:10.3389/FGENE.2022.985975.
- 7 16. Debacker, K.; Frank Kooy, R. Fragile Sites and Human Disease. *Hum Mol Genet* **2007**,
8 *16*, R150–R158, doi:10.1093/HMG/DDM136.
- 9 17. Laganà, A.; Russo, F.; Sismeiro, C.; Giugno, R.; Pulvirenti, A.; Ferro, A. Variability in
10 the Incidence of MiRNAs and Genes in Fragile Sites and the Role of Repeats and
11 CpG Islands in the Distribution of Genetic Material. *PLoS One* **2010**, *5*, e11166,
12 doi:10.1371/JOURNAL.PONE.0011166.
- 13 18. Szczepanek, J.; Skorupa, M.; Tretyn, A. MicroRNA as a Potential Therapeutic
14 Molecule in Cancer. *Cells* **2022**, *11*, doi:10.3390/CELLS11061008.
- 15 19. Thomas, J.; Ohtsuka, M.; Pichler, M.; Ling, H. MicroRNAs: Clinical Relevance in
16 Colorectal Cancer. *Int J Mol Sci* **2015**, *16*, 28063–28076, doi:10.3390/ijms161226080.
- 17 20. Slattery, M.L.; Wolff, E.; Hoffman, M.D.; Pellatt, D.F.; Milash, B.; Wolff, R.K.
18 MicroRNAs and Colon and Rectal Cancer: Differential Expression by Tumor
19 Location and Subtype. *Genes Chromosomes Cancer* **2011**, *50*, 196–206,
20 doi:10.1002/gcc.20844.
- 21 21. Omrane, I.; Kourda, N.; Stambouli, N.; Privat, M.; Medimegh, I.; Arfaoui, A.;
22 Uhrhammer, N.; Bougatef, K.; Baroudi, O.; Bouzaienne, H.; et al. MicroRNAs 146a
23 and 147b Biomarkers for Colorectal Tumor’s Localization. *Biomed Res Int* **2014**, *2014*,
24 doi:10.1155/2014/584852.
- 25 22. Coebergh Van Den Braak, R.R.J.; Sieuwerts, A.M.; Lalmahomed, Z.S.; Smid, M.;
26 Wilting, S.M.; Bril, S.I.; Xiang, S.; Van Der Vlugt-Daane, M.; De Weerd, V.; Van
27 Galen, A.; et al. Confirmation of a Metastasis-Specific MicroRNA Signature in
28 Primary Colon Cancer. *Sci Rep* **2018**, *8*, 1–11, doi:10.1038/s41598-018-22532-1.
- 29 23. Yang, L.; Li, L.; Ma, J.; Yang, S.; Zou, C.; Yu, X. MiRNA and MRNA Integration
30 Network Construction Reveals Novel Key Regulators in Left-Sided and Right-Sided
31 Colon Adenocarcinoma. *Biomed Res Int* **2019**, *2019*, 1–9, doi:10.1155/2019/7149296.
- 32 24. Eneh, S.; Heikkinen, S.; Hartikainen, J.M.; Kuopio, T.; Mecklin, J.P.; Kosma, V.M.;
33 Mannermaa, A. MicroRNAs Associated with Biological Pathways of Left- And
34 Right-Sided Colorectal Cancer. *Anticancer Res* **2020**, *40*, 3713–3722,
35 doi:10.21873/ANTICANRES.14360.
- 36 25. Loree, J.M.; Pereira, A.A.L.; Lam, M.; Willauer, A.N.; Raghav, K.; Dasari, A.; Van
37 Morris, K.; Advani, S.; Menter, D.G.; Eng, C.; et al. Classifying Colorectal Cancer by
38 Tumor Location Rather than Sidedness Highlights a Continuum in Mutation Profiles
39 and Consensus Molecular Subtypes. *Clinical Cancer Research* **2018**, *24*, 1062–1072,
40 doi:10.1158/1078-0432.CCR-17-2484.
- 41 26. Dai, Y.; Duan, H.; Duan, C.; Zhu, H.; Zhou, R.; Pei, H.; Shen, L. TCF21 Functions as
42 a Tumor Suppressor in Colorectal Cancer through Inactivation of PI3K/AKT
43 Signaling. *Onco Targets Ther* **2017**, *10*, 1603–1611, doi:10.2147/OTT.S118151.
- 44 27. Noetzel, E.; Rose, M.; Sevinc, E.; Hilgers, R.D.; Hartmann, A.; Naami, A.; Knüchel,
45 R.; Dahl, E. Intermediate Filament Dynamics and Breast Cancer: Aberrant Promoter
46 Methylation of the Synemin Gene Is Associated with Early Tumor Relapse. *Oncogene*
47 **2010**, *29*, 4814–4825, doi:10.1038/ONC.2010.229.

- 1 28. Wang, Y.; Chen, Y.; Xiao, S.; Fu, K. Integrated Analysis of the Functions and
2 Prognostic Values of RNA-Binding Proteins in Colorectal Cancer. *Front Cell Dev Biol*
3 **2020**, *8*, doi:10.3389/FCELL.2020.595605/FULL.
- 4 29. Lai, S.H.; Zervoudakis, G.; Chou, J.; Gurney, M.E.; Quesnelle, K.M. PDE4 Subtypes
5 in Cancer. *Oncogene* **2020**, *39*, 3791, doi:10.1038/S41388-020-1258-8.
- 6 30. Chen, L.; Gao, H.; Liang, J.; Qiao, J.; Duan, J.; Shi, H.; Zhen, T.; Li, H.; Zhang, F.; Zhu,
7 Z.; et al. MiR-203a-3p Promotes Colorectal Cancer Proliferation and Migration by
8 Targeting PDE4D. *Am J Cancer Res* **2018**, *8*, 2387.
- 9 31. Chung, Y.; Kim, H.; Bang, S.; Jang, K.; Paik, S.S.; Shin, S.-J. Nuclear Expression Loss
10 of SSBP2 Is Associated with Poor Prognostic Factors in Colorectal Adenocarcinoma.
11 *Diagnostics* **2020**, *10*, doi:10.3390/DIAGNOSTICS10121097.
- 12 32. Wang, Y.; Klumpp, S.; Amin, H.M.; Liang, H.; Li, J.; Estrov, Z.; Zweidler-McKay, P.;
13 Brandt, S.J.; Agulnick, A.; Nagarajan, L. SSBP2 Is an in Vivo Tumor Suppressor and
14 Regulator of LDB1 Stability. *Oncogene* **2010**, *29*, 3044, doi:10.1038/ONC.2010.78.
- 15 33. Zhu, X.; Shen, Z.; Man, D.; Ruan, H.; Huang, S. MiR-152-3p Affects the Progression
16 of Colon Cancer via the KLF4/IFITM3 Axis. *Comput Math Methods Med* **2020**, *2020*,
17 doi:10.1155/2020/8209504.
- 18 34. Lee, E.; Cheung, J.; Bialkowska, A.B. Krüppel-like Factors 4 and 5 in Colorectal
19 Tumorigenesis. *Cancers (Basel)* **2023**, *15*, doi:10.3390/CANCERS15092430/S1.
- 20 35. Chang, G.; Xu, S.; Dhir, R.; Chandran, U.; O'Keefe, D.S.; Greenberg, N.M.; Gingrich,
21 J.R. Hypoexpression and Epigenetic Regulation of Candidate Tumor Suppressor
22 Gene CADM-2 in Human Prostate Cancer. *Clin Cancer Res* **2010**, *16*, 5390,
23 doi:10.1158/1078-0432.CCR-10-1461.
- 24 36. Semba, S.; Itoh, N.; Ito, M.; Youssef, E.M.; Harada, M.; Moriya, T.; Kimura, W.;
25 Yamakawa, M. Down-Regulation of PIK3CG, a Catalytic Subunit of
26 Phosphatidylinositol 3-OH Kinase, by CpG Hypermethylation in Human Colorectal
27 Carcinoma. *Clinical Cancer Research* **2002**, *8*.
- 28 37. Wang, Y.; Zhao, J.; Wang, Y.; Gao, J.; Yang, H.; Li, H. MiR-17-5p Targets and
29 Downregulates CADM2, Activating the Malignant Phenotypes of Colon Cancer
30 Cells. *Mol Biotechnol* **2022**, *64*, 1388–1400, doi:10.1007/S12033-022-00515-
31 Y/FIGURES/5.
- 32 38. Cui, X.; Liang, H.; Hao, C.; Jing, X. Homer1 Is a Potential Biomarker for Prognosis in
33 Human Colorectal Carcinoma, Possibly in Association with G3BP1 Signaling. *Cancer*
34 *Manag Res* **2020**, *12*, 2899, doi:10.2147/CMAR.S240942.
- 35 39. Arribas, J.; Cajuso, T.; Rodio, A.; Marcos, R.; Leonardi, A.; Velázquez, A. NF-KB
36 Mediates the Expression of TBX15 in Cancer Cells. *PLoS One* **2016**, *11*,
37 doi:10.1371/JOURNAL.PONE.0157761.
- 38 40. Martinez, V.D.; Marshall, E.A.; Anderson, C.; Ng, K.W.; Minatel, B.C.; Sage, A.P.;
39 Enfield, K.S.S.; Xu, Z.; Lam, W.L. Discovery of Previously Undetected MicroRNAs
40 in Mesothelioma and Their Use as Tissue-of-Origin Markers. *Am J Respir Cell Mol*
41 *Biol* **2019**, *61*, 266–268, doi:10.1165/rcmb.2018-0204LE.
- 42 41. Metsalu, T.; Vilo, J. ClustVis: A Web Tool for Visualizing Clustering of Multivariate
43 Data Using Principal Component Analysis and Heatmap. *Nucleic Acids Res* **2015**, *43*,
44 W566–W570, doi:10.1093/NAR/GKV468.
- 45 42. Kumar, R.; Nagpal, G.; Kumar, V.; Usmani, S.S.; Agrawal, P.; Raghava, G.P.S.
46 HumCFS: A Database of Fragile Sites in Human Chromosomes. *BMC Genomics* **2019**,
47 *19*, 1–8, doi:10.1186/S12864-018-5330-5/FIGURES/5.

- 1 43. Demšar, J.; Erjavec, A.; Hočevár, T.; Milutinovič, M.; Možina, M.; Toplak, M.; Umek,
2 L.; Zbontar, J.; Zupan, B. Orange: Data Mining Toolbox in Python Tomaž Curk
3 Matija Polajnar Laň Zagar. *Journal of Machine Learning Research* **2013**, *14*, 2349–2353.
4 44. Gu, Z.; Eils, R.; Schlesner, M. Complex Heatmaps Reveal Patterns and Correlations
5 in Multidimensional Genomic Data. *Bioinformatics* **2016**, *32*, 2847–2849,
6 doi:10.1093/BIOINFORMATICS/BTW313.
7 45. Liu, W.; Wang, X. Prediction of Functional MicroRNA Targets by Integrative
8 Modeling of MicroRNA Binding and Target Expression Data. *Genome Biol* **2019**, *20*,
9 1–10, doi:10.1186/s13059-019-1629-z.
10 46. Chen, Y.; Wang, X. MiRDB: An Online Database for Prediction of Functional
11 MicroRNA Targets. *Nucleic Acids Res* **2020**, *48*, D127–D131, doi:10.1093/nar/gkz757.
12 47. Goldman, M.J.; Craft, B.; Hastie, M.; Repečka, K.; McDade, F.; Kamath, A.; Banerjee,
13 A.; Luo, Y.; Rogers, D.; Brooks, A.N.; et al. Visualizing and Interpreting Cancer
14 Genomics Data via the Xena Platform. *Nat Biotechnol* **2020**, *38*, 675–678.
15 48. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.;
16 Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING Database in
17 2023: Protein–Protein Association Networks and Functional Enrichment Analyses
18 for Any Sequenced Genome of Interest. *Nucleic Acids Res* **2023**, *51*, D638,
19 doi:10.1093/NAR/GKAC1000.
20 49. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.;
21 Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated
22 Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13*, 2498,
23 doi:10.1101/GR.1239303.
24