



UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA  
MOLECULAR

REPERCUSSÕES DAS ANÁLISES DE BIOINFORMÁTICA NOS  
RESULTADOS DE EXPRESSÃO DIFERENCIAL DE GENES E SEUS  
REGULADORES MIRNAS NO CÂNCER GÁSTRICO

Emily Vanessa Nascimento Araújo

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFPA como requisito parcial para obtenção do grau de Mestre em Genética e Biologia Molecular (Bioinformática).

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Ândrea Ribeiro-dos-Santos

Belém – Pará  
Outubro de 2018

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema de Bibliotecas da Universidade Federal do Pará  
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

---

A658r Araujo, Emily Vanessa Nascimento  
REPERCUSSÕES DAS ANÁLISES DE BIOINFORMÁTICA NOS RESULTADOS DE  
EXPRESSÃO DIFERENCIAL DE GENES E SEUS REGULADORES MIRNAS NO CÂNCER  
GÁSTRICO / Emily Vanessa Nascimento Araujo. — 2018  
74 f. : il. color

Dissertação (Mestrado) - , Instituto de Ciências Biológicas, Universidade Federal do Pará,  
Belém, 2018.

Orientação: Profa. Dra. Ândrea Ribeiro-dos-Santos

1. Câncer Gástrico. 2. TCGA. 3. miRNA. 4. Expressão Diferencial. 5. Z-Score. I. Ribeiro-dos-Santos, Ândrea , *orient.* II. Título

CDD 616.99433

---

## INSTITUIÇÕES PARTICIPANTES E FINANCIADORAS

### Instituições Participantes:

1. Laboratório de Genética Humana e Médica (LGHM) – Instituto de Ciências Biológicas/Universidade Federal do Pará (UFPA)
2. Núcleo de Pesquisas em Oncologia/Universidade Federal do Pará (UFPA)
3. *Bioinformatics Multidisciplinary Environment* (BioME) / Universidade Federal do Rio Grande do Norte (UFRN)
4. Instituto Metrópole Digital (IMD) /Universidade Federal do Rio Grande do Norte (UFRN)

### Instituições Financiadoras:

1. Coordenação de Aperfeiçoamento Pessoal de Ensino Superior (CAPES).
2. REDE DE PESQUISA EM GENÔMICA POPULACIONAL HUMANA.  
(CAPES, Bio Computacional 3381/2013).

## AGRADECIMENTOS

Agradeço primeiramente a Deus. Ele me deu saúde e discernimento para aproveitar esta oportunidade e principalmente, jamais me deixou desistir.

À minha orientadora Prof<sup>a</sup>. Dr<sup>a</sup>. Ândrea Kely Ribeiro dos Santos. A sua paciência, dedicação, amizade e a sua valorosa experiência acadêmica foram necessárias para a realização deste trabalho, assim como foram importantes para o desenvolvimento do meu conhecimento em bioinformática. À professora Ândrea o meu muito obrigada por tudo.

Ao Prof.Dr. Fabiano Cordeiro Moreira agradeço a sua disponibilidade em me auxiliar nos momentos de dúvida e também por compartilhar o seu respeitoso conhecimento em bioinformática, estes contribuíram para o desenvolvimento e enriquecimento deste trabalho.

Ao Prof.Dr. Jorge Estefano Santana de Souza. O seu conhecimento e experiência em bioinformática, o seu auxílio na análise dos dados durante e após o mestrado sanduíche e o convite para fazer parte do seu trabalho por meio do *R-periodot* foram essenciais para nortear o meu trabalho.

Agradeço ao Prof.Dr. Sandro José de Souza. A sua orientação durante o mestrado sanduíche, o apoio em Natal e a oportunidade de trabalhar em sua equipe me fizeram ser uma melhor pesquisadora da área de bioinformática.

Agradeço ao Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM) e ao Núcleo de Pesquisas em Oncologia (NPO), ambos da Universidade Federal do Pará (UFPA), pelas contribuições aos alunos do programa durante esta caminhada.

Agradeço o *Bioinformatics Multidisciplinary Environment* (BioME) e o Instituto Metr pole Digital (IMD), ambos da Universidade Federal do Rio Grande do Norte (UFRN). A acolhida durante o mestrado sandu che e todo o conhecimento que me ajudaram a adquirir foram indispens veis em minha pesquisa.

Agradeço aos meus amigos do Laborat rio de Gen tica Humana e M dica (LGHM) da Universidade Federal do Par  (UFPA) por compartilharem experi ncias, alegrias, tristezas e suas viv ncias profissionais, que sem d vidas colaboraram para que eu conseguisse chegar at  aqui.

Agradeço aos amigos que fiz no *Bioinformatics Multidisciplinary Environment* (BioME) da Universidade Federal do Rio Grande do Norte (UFRN). A demonstra o de paix o e orgulho em atuar na  rea de bioinform tica me fizeram acreditar que juntos podemos fazer muito pela pesquisa.

  minha fam lia, em especial   minha m e Marcia Helena Machado Nascimento,   minha av  (in Memoriam) Amazonita, ao meu tio padrinho Marcio,   minha tia Katia, aos meus "primos-irm os" Bruno, Reinaldo, Diego, Ricardo e Pedro. Agradeço por serem o meu incentivo de todos os dias e por sempre acreditarem que eu sou capaz. Graças a voc s, eu sou mais forte.

Agradeço ao meu namorado Leonardo e ao meu sogro Lauriston. As suas palavras de incentivo e apoio, as conversas sobre suas experi ncias profissionais, a credibilidade dada ao meu estudo e   minha capacidade intelectual me ajudaram a ser uma pessoa melhor, bem como foram importantes para que eu reafirmasse a import ncia da minha dedica o a este trabalho.

## SUMÁRIO

1. -	INTRODUÇÃO	10
1.1-	RNA-Seq	12
1.1.1 -	EXPRESSÃO GÊNICA	13
1.1.2 -	EXPRESSÃO DE MICRORNA	18
1.2 -	ANÁLISE DIFERENCIAL DE EXPRESSÃO GÊNICA	26
1.3 -	SOFTWARE R-PERIDOT	34
1.4 -	ANÁLISE FUNCIONAL	36
1.5 -	ANÁLISE Z-SCORE	37
2. -	OBJETIVOS	39
2.1 -	OBJETIVO GERAL	39
2.2-	OBJETIVOS ESPECÍFICOS	39
3. -	METODOLOGIA	40
3.1 -	ORIGEM DOS DADOS	41
3.2 -	PRÉ-PROCESSAMENTO	43
3.3 -	ANÁLISE DIFERENCIAL DE MIRNAS	43
3.4 -	FERRAMENTAS PARA SELEÇÃO E ANÁLISE FUNCIONAL DE GENES ALVOS	44
3.5 -	CÁLCULO DO Z-SCORE DE EXPRESSÃO DE mRNA	46
4. -	RESULTADOS	47
4.1 -	ANÁLISE DE MIRNAS NO R-PERIDOT	47
4.2 -	SELEÇÃO DOS GENES ALVOS	50
4.3 -	ANÁLISE FUNCIONAL DOS GENES ALVOS	51
4.4 -	ANÁLISE DAS VIAS KEGG E AS MARCAS DO CÂNCER	55
4.5 -	RESULTADOS PARA O Z-SCORE	56
5. -	DISCUSSÃO	57
6. -	CONCLUSÃO	64
7. -	REFERÊNCIAS	65

## LISTA DE FIGURAS E TABELAS

Figura 1	Novo esquema do Dogma Central da Biologia	14
Figura 2	Etapas envolvidas na expressão gênica de eucariotos	15
Figura 3	O Mecanismo da Biogênese de miRNA	20
Figura 4	Atuação dos microRNAs na região 3' da molécula do RNA mensageiro e suas conseqüências ao reprimir a tradução	22
Figura 5	Etapas da análise de miRNomas por meio de sequenciamento NGS	24
Tabela 1	Exemplo de dados de entrada para análise estatística por meio do DESeq	30
Tabela 2	Exemplo de dados de saída para análise estatística por meio do DESeq	31
Figura 6	Logotipo do software R-peridot	35
Figura 7	Interface do software R-peridot	36
Figura 8	Fluxograma linear da metodologia empregada neste estudo	40
Figura 9	Amostras de expressão de miRNA utilizadas no estudo	42
Figura 10	Amostras de expressão de genes utilizadas no estudo	42
Tabela 3	Parâmetros utilizados no R-peridot para o teste de análise de expressão de miRNAs no câncer gástrico com amostras pareadas	44
Figura 11	As 10 características principais (marcas) que levam ao desenvolvimento do Câncer	45
Figura 12	Fórmula para o cálculo do Z-score	46
Figura 13	Diagrama de Venn para demonstrar os 6 miRNAs diferencialmente expressos em comum aos pacotes EBSeq, edgeR e DESeq2	47
Tabela 4	miRNAs diferencialmente expressos no CG pelo <i>R-peridot</i> em comum aos pacotes DESeq2, EBSeq e edgeR, e os miRNAs encontrados para cada pacote estatístico	48
Tabela 5	Resultados pelo R-peridot dos miRNAs diferencialmente expressos no CG com parâmetros mais restritivos e empregando os pacotes DESeq2, EBSeq e edgeR separadamente	49
Figura 14	Genes alvos considerando dois ou mais miRNAs diferencialmente expressos de acordo com a plataforma miRTargetLink Human	50

Tabela 6	Vias KEGG identificadas no CG pela plataforma DAVID	51
Tabela 7	Ontologias identificadas no CG de acordo com a plataforma DAVID	52
Figura 15	Rede de interação entre os genes alvos dos miRNAs diferencialmente expressos no CG	53
Tabela 8	Vias KEGG identificadas pela base de dados STRING no CG	54
Tabela 9	Genes alvos identificados e associados com as 10 principais características do câncer	55
Tabela 10	Valores de Z-Score dos genes alvos nas amostras de CG do banco de dados TCGA	56
Tabela 11	57 miRNAs diferencialmente expressos no CG pelo pacote DESeq2 por meio do RStudio	60



## RESUMO

O câncer, de acordo com a Globocan, 2018, foi responsável por uma taxa de 8,2 milhões de mortes e 14,1 milhões de casos novos. No que diz respeito especificamente à incidência de câncer gástrico (CG), para o estado do Pará, estima-se no ano de 2018 um número de homens afetados de 23,30 e 11,45 de mulheres. No desenvolvimento do adenocarcinoma gástrico foram identificadas alterações moleculares significativamente aumentadas em vias metabólicas. Um desses elementos são os miRNAs, pequenos RNAs não codificantes que atuam na regulação da expressão gênica e podem agir como oncogenes ou supressores tumorais. O objetivo deste estudo foi investigar a possibilidade de genes das amostras tumorais depositadas no banco de dados Atlas Genômico do Câncer (TCGA) de serem modificados por miRNAs diferencialmente expressos. A análise foi realizada por plataforma de sequenciamento, RNA-Seq, e utilizou o *software R-peridot* na análise de expressão diferencial de microRNAs, o resultado identificou seis miRNAs alterados significativamente. Em seguida a plataforma miRTargetLink Human foi empregada e identificou oito genes alvos desses miRNAs entre amostras tumorais e seus correspondentes tecidos adjacentes no CG. As análises funcionais mostraram genes que participam de três vias metabólicas importantes associadas ao câncer: adesão focal, via de sinalização RAP1 e pela via de sinalização de cálcio. Por fim, a utilização do cálculo Z-Score confirmou os achados significantes.

Palavra-chave: Câncer Gástrico, TCGA, miRNA, mRNA, expressão diferencial, Z-Score.

## **ABSTRACT**

The cancer, according to Globocan, in 2018, was responsible for a rate of 8.2 million deaths and about 14.1 million new cases. Regarding the analysis of gastric cancer (GC), in the state of Pará, it was estimated that for the year 2018 the number of affected men would be 23.30 and 11.45 for women. In the development of gastric adenocarcinoma, several significantly increased molecular alterations were identified when compared to non-cancerous tissues. One of these elements are miRNAs, small non-coding RNAs that act on the regulation of gene expression and can act as oncogenes or tumor suppressors. The objective of this study was to investigate the possibility of genes from tumor samples (deposited in The Cancer Genome Atlas - TCGA database) to be modified by differentially expressed miRNAs. The analysis was performed by sequencing platform, RNA-Seq, and used the software R-peridot in differential expression analysis, whose result identified six miRNAs. Next, the miRTargetLink Human platform identified eight miRNA target genes between tumor samples and their corresponding adjacent CG tissues. Functional analyzes identified important genes that can be enriched by three pathways associated with cancer: focal adhesion, RAP1 signaling pathway and calcium signaling pathway. Finally, the use of the Z-Score calculation confirmed significant findings.

Key words: Gastric cancer, TCGA, miRNA, mRNA, differentially expressed, Z-Score

## 1. INTRODUÇÃO

O câncer, de acordo com a Globocan, 2018, é um importante obstáculo para o estabelecimento da saúde pública tanto em países desenvolvidos como em países em desenvolvimento. Em 2015, esta patologia foi responsável por uma taxa de 8,2 milhões de mortes e aproximadamente 14,1 milhões de casos novos (GLOBOCAN, 2018).

Existem catalogados mais de 200 tipos de câncer, dentre estes o mais freqüente é o que será abordado neste trabalho: o câncer gástrico. O câncer gástrico (CG) é uma patologia de etiologia multifatorial, em que os tumores se desenvolvem a partir de lesões da mucosa gástrica, ocasionadas pela ação e/ou interação de vários fatores de risco, ou seja, é um processo que inclui várias etapas sendo resultado da exposição a fatores endógenos (genéticos e epigenéticos) e exógenos (ambientais) por um longo período de tempo (CARVALHO *et al.*, 2010).

A incidência desta patologia varia de acordo com as diversas regiões do mundo, e no geral, é de duas a três vezes mais frequente nos países em desenvolvimento, além de ser mais comum em indivíduos do sexo masculino (FERLAY *et al.*, 2015). De acordo com o Instituto Nacional do Câncer (INCA) no biênio 2018-2019, para o Brasil, estima-se a ocorrência de 600 mil casos novos de câncer. Já com relação ao câncer gástrico, estimam-se 13.540 casos novos entre homens e 7.750 nas mulheres para cada ano do biênio 2018-2019.

Com relação a capital do estado do Pará, Belém, a taxa de incidência de câncer gástrico estimada para 2018 em homens foi de 23,30 e em mulheres,

11,45, sendo a segunda mais alta taxa entre os homens e a quarta mais alta entre as mulheres, respectivamente, para os tipos de câncer mencionados pelo INCA (INCA, 2018).

O câncer gástrico apresenta-se em três tipos histológicos principais, sendo o mais freqüente o adenocarcinoma, responsável por até 95% dos tumores diagnosticados. Ele é originado das glândulas da camada mucosa. O outro tipo histológico é denominado linfoma, este é originário do tecido linfóide e é responsável por 3% dos casos. O terceiro tipo, representando 1 a 2% dos casos, é o leiomiossarcoma, um tipo de câncer gástrico que se inicia em tecidos que dão origem aos músculos (ESPEJO *et al.*, 2003). Neste estudo, as citações ao câncer gástrico irão se referir exclusivamente ao tipo histológico adenocarcinoma.

O prognóstico do câncer gástrico é definido pelo estadiamento do tumor no momento do diagnóstico. Entretanto, em muitos países, mesmo em países desenvolvidos, a maioria dos casos de câncer gástrico é diagnosticada em um estágio avançado, o que limita a sobrevida (NAGINI, 2012). Ainda que haja avanços, o tratamento de câncer gástrico continua a ser um problema de saúde pública global (DURAES *et al.*, 2014), visto que nem todos os pacientes respondem de forma satisfatória às terapias direcionadas atualmente disponíveis. Todavia, novos alvos de tratamento e a descoberta de biomarcadores preditivos ainda são urgentemente necessários (STEFFEN *et al.*, 2012), já que aumentariam a capacidade de diagnóstico e de predição do comportamento do tumor, assim como de intervenção no estabelecimento da

neoplasia gástrica, além de gerar uma conduta terapêutica mais precisa e agressiva (Assumpção & Burbano, 2005).

Durante o desenvolvimento do adenocarcinoma gástrico foram identificadas várias alterações moleculares. A expressão aumentada do gene *Her3*, por exemplo, apresentou significativo aumento no câncer gástrico comparada com os tecidos estomacais normais, resultando em um prognóstico ruim (SHEN *et al.*, 2017; EMA *et al.*, 2014). Por meio de um estudo de ligação entre genes, AOKI *et al.*, (2005) identificaram que o cromossomo 2q33-35 possui um *locus* com potencial susceptibilidade para câncer gástrico proximal.

A introdução de novas tecnologias de sequenciamento de alto desempenho ou *Next Generation Sequencing* – NGS e o desenvolvimento de ferramentas mais avançadas possibilitaram novas descobertas acerca dos mecanismos envolvidos na carcinogênese, entre as quais destaco o controle da expressão celular.

### **1.1 RNA-Seq**

O emprego das plataformas NGS teve grande repercussão no mundo científico e principalmente no âmbito da genômica (LIU *et al.*, 2012). Em 1977 Sanger e colaboradores descreveram o primeiro método de sequenciamento de DNA. Em resumo, este método consiste em realizar a síntese de uma fita de DNA marcada que seja complementar à fita da qual se deseja determinar a sequência (SHENDURE & JI, 2008). Entretanto, esse método apesar de ser reconhecido como "padrão ouro" do seqüenciamento apresentava alto custo,

demanda elevado tempo, e tem um rendimento relativamente baixo. Em contraste, os sequenciadores NGS permitem análises paralelas, de elevado rendimento e um custo reduzido, trazendo grande aplicabilidade para as sequências genômicas (LIU *et al.*, 2012) por meio da produção de milhões de fragmentos de sequências em um curto tempo de processamento das amostras, dependendo da plataforma (GRADA; WEINBRECHT, 2013).

A análise de sequenciamento *High-Throughput* de RNA ou RNA-Seq está entre as metodologias existentes para o estudo do transcriptoma. Esta é uma ferramenta que permite acesso a informação, o entendimento detalhado, anotação e quantificação de genes, seus produtos e suas isoformas (GARBER *et al.*, 2011). Este tipo de análise de sequenciamento RNA-Seq se aperfeiçoa a cada dia com o desenvolvimento de novas ferramentas (GARBER *et al.*, 2011) e com os avanços computacionais (bioinformática). Com isso, essa técnica pode ser introduzida em estratégias distintas de pesquisa, como o sequenciamento de genomas completos, estudos de metagenômica e estudos relacionados à expressão diferencial (GRADA; WEINBRECHT, 2013).

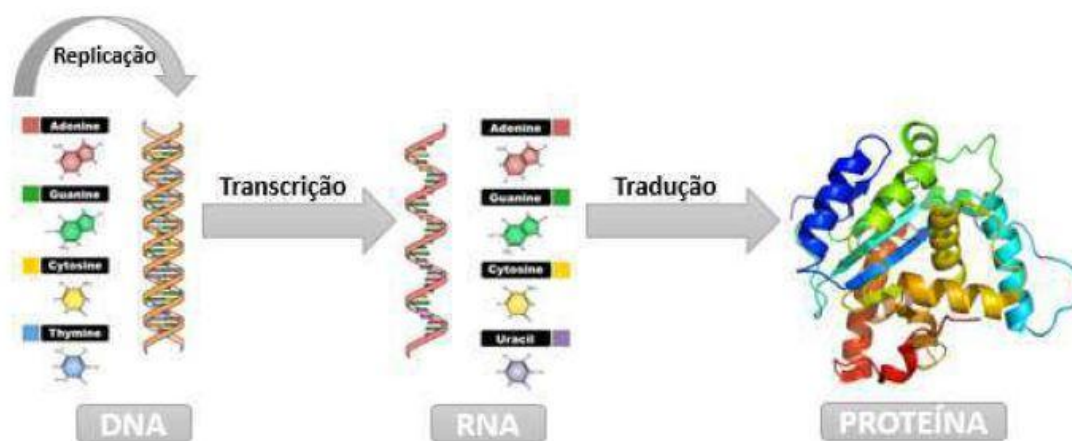
### **1.1.1 Expressão Gênica**

Este estudo irá abordar sobre a expressão de genes e seus reguladores nos mecanismos biológicos em nível molecular.

A molécula de DNA serve como molde para a produção de uma proteína, a partir de uma molécula de RNA, o que é conhecido como DOGMA CENTRAL DA BIOLOGIA, que coordena o fluxo de informação dentro de uma

célula e produz complexas formas de interações que originam o genoma, o transcriptoma e o proteoma (RUIVO, 2007). E baseado neste dogma central, a expressão gênica é unidirecional, a informação é propagada por meio da duplicação do DNA, e é traduzida com a contribuição de dois métodos: a transcrição e a tradução. Na transcrição a informação do DNA é convertida em uma fita de RNA e na tradução a informação contida no RNA é convertida em proteínas (RAMOS, 2018).

Partindo deste princípio, o fluxo da informação (expressão gênica) segue o sentido DNA/RNA/PROTEÍNAS, como é apresentado na Figura 1. Nesta figura, apresenta-se uma seqüência específica de nucleotídeos no DNA que será transcrito em uma molécula de RNA e conseqüentemente traduzido em uma seqüência de aminoácidos de uma proteína.

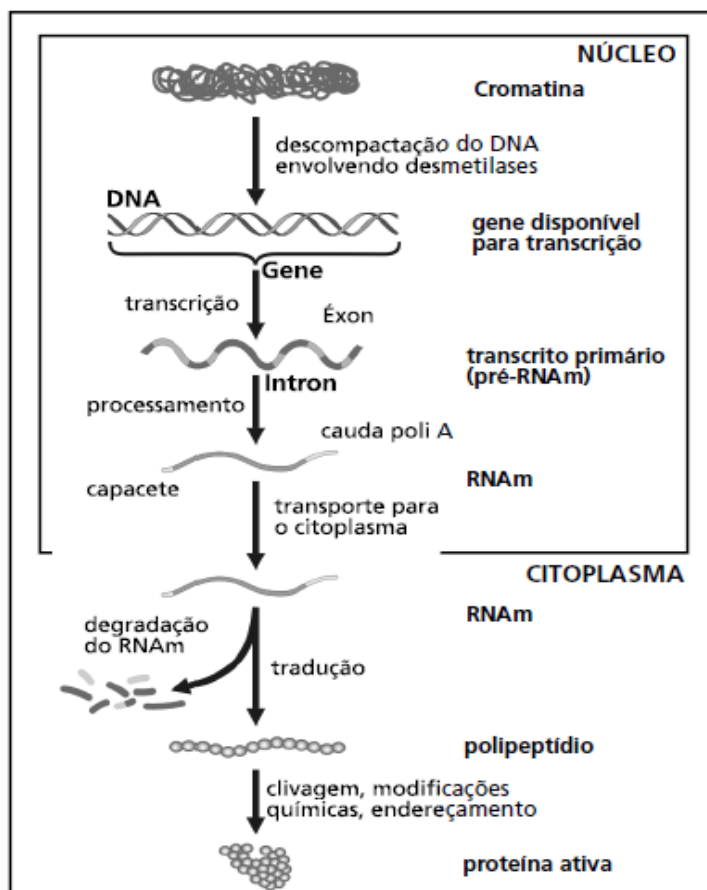


**Figura 1:** Novo esquema do Dogma Central da Biologia (RAMOS, 2018).

Para Lodish *et al.* (2014), um gene “é a parte do DNA que possui a informação gênica que irá especificar a síntese de uma cadeia polipeptídica única ou de um RNA funcional”.

Os genes levam as informações que irão formar as moléculas de RNAs, e conseqüentemente constituir as proteínas. Durante a síntese de uma proteína, as sequências nucleotídicas de uma molécula de mRNA possuem informações que especificam a ordem correta dos aminoácidos na construção da proteína (LODISH *et al.*, 2014). Neste contexto, a transcrição e a tradução são processos complexos relacionados aos diversos eventos de regulação (ALBERTS *et al.*, 2010). Na Figura 2, por exemplo, apresentam-se algumas etapas envolvidas na expressão gênica de eucariotos.





**Figura 2:** Etapas envolvidas na expressão gênica de eucariotos (Modificado de DE ROBERTIS & HIB, 2012; JORDE *et al.*, 2010).

De acordo com a Figura 2, para que os genes sejam transcritos primeiramente a cromatina deve ser descompactada. Em seguida o pré-mRNA (transcrito primário) é processado para que ocorra a síntese do mRNA maduro (transcrito secundário, sem íntrons), que irá ser transportado para o citoplasma e fará parte do mecanismo de tradução. Durante a tradução forma-se um polipeptídeo que passará por clivagens e alterações bioquímicas para formar a proteína com atividade biológica (DE ROBERTIS & HIB, 2012; JORDE *et al.*, 2010).

Neste contexto os fatores de transcrição (TF), que consistem na ligação de proteínas reguladoras de transcrição específicas, podem atuar como

ativadores ou repressores deste processo. Os TF são os principais responsáveis pelo controle da expressão de genes no decorrer dos mecanismos de desenvolvimento e diferenciação, e pela resposta das células a estímulos (KURISSIO, 2017). Assim, ao longo da regulação da expressão de um gene pode acontecer tanto regulação negativa ou *down-regulated*, na qual a proteína ligadora impede a transcrição diminuindo a expressão de um gene (e conseqüentemente gerando menos cópias de mRNA e menos proteínas), quanto uma regulação positiva ou *up-regulated* em que a proteína ligadora promove a transcrição aumentando a expressão do gene, produz mais cópias de mRNA e aumento da proteína correspondente (KURISSIO, 2017).

Distintas mutações têm sido descritas nos TF entre os principais macrogrupos geográficos humanos, como os Africanos, Asiáticos e Europeus. Foram identificadas várias mutações diferenciadas do tipo Inserções e Deleções (INDEL) em Sítio de Ligação de Fatores de Transcrição (IDTFBS) (RIBEIRO DOS SANTOS, 2014). Em Africanos, as mutações diferenciadas de IDTFBS estavam associadas a genes envolvidos em doenças complexas como câncer de bexiga e câncer de pele (JOHNSON *et al.*, 2007; MCELROY *et al.*, 2010; LANDGREN *et al.*, 2010; FERLAY *et al.*, 2013; WHO, 2013; TULL, 2013; VARDENY *et al.*, 2013; EWANG-EMUKOWHATE *et al.*, 2014). Em contrapartida, em Asiáticos as mutações estavam ligadas a genes envolvidos em doenças como diabetes, câncer de mama e adenocarcinomas (FERLAY *et al.*, 2013; WHO, 2013; TULL, 2013; VARDENY *et al.*, 2013; EWANGEMUKOWHATE *et al.*, 2014). Nos grupos Europeus as mutações diferenciadas de IDTFBS estavam relacionadas a genes envolvidos em doenças que possuem elevadas taxas de incidências nestas populações, como

câncer de pele, Parkinson e asma (SATAKE *et al.*, 2009; FERLAY *et al.*, 2013; WHO, 2013; LI *et al.*, 2013; PRICE *et al.*, 2014; MYERS *et al.*, 2014; LIU *et al.*, 2014).

### 1.1.2 Expressão de microRNA

O genoma é o conjunto completo de todas as informações e regiões contidas (lidas e não lidas) em um organismo, ou seja, a estrutura base do DNA. O transcriptoma consiste no conjunto completo das regiões transcritas: RNA mensageiro, RNA ribossômico, RNA transportador, e outros RNAs não codificantes como os microRNAs. Enquanto o proteoma se constitui da reunião de todas as proteínas traduzidas daquele organismo (RUIVO, 2007).

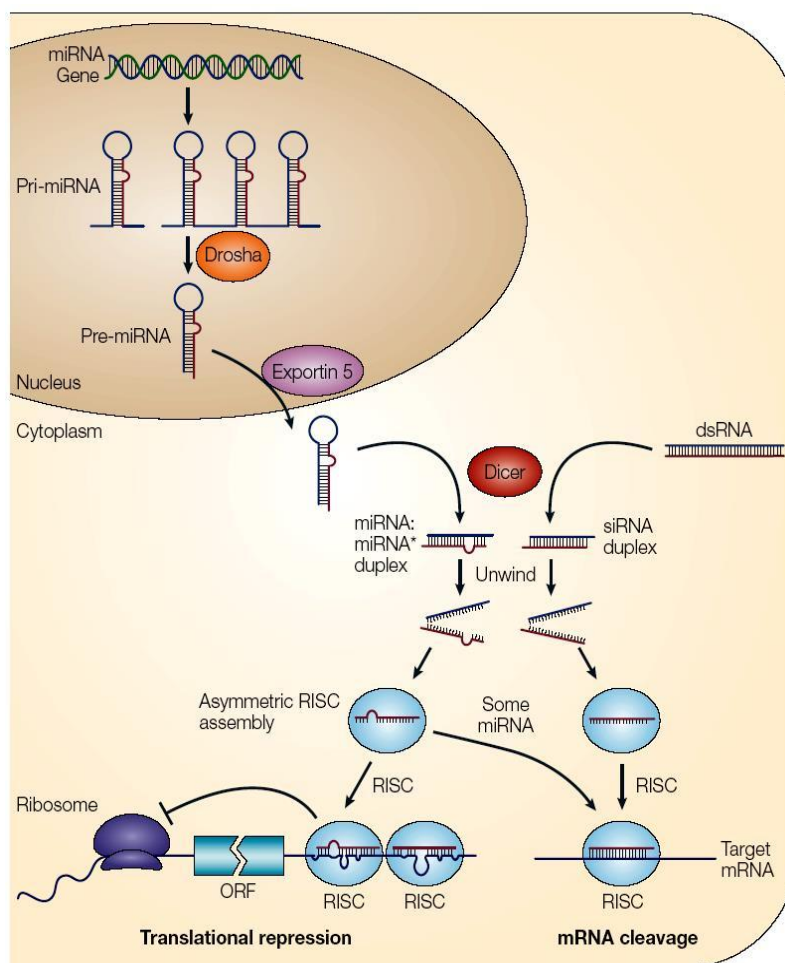
A análise de expressão, além da expressão gênica propriamente dita, envolve a expressão de um conjunto de outros elementos como os RNAs não codificantes (*non-coding RNA*), dentre os quais se destacamos miRNAs. Estes estão envolvidos diretamente no mecanismo de regulação das doenças, como as neoplasias (em especial o câncer gástrico) (DARNET *et al.*, 2015; MOREIRA *et al.*, 2015).

Os microRNAs são RNAs não codificantes que possuem entre 17 a 22 nucleotídeos de comprimento e que podem regular a expressão gênica por hibridização com a região 3' não traduzida de um alvo de um RNA mensageiro específico (RIBEIRO-DOS-SANTOS *et al.*, 2012; LEE *et al.*, 2017).

Estes pequenos RNAs não codificantes são produzidos de moléculas precursoras (pri-miRNAs), que podem ser traduzidas por RNA polimerase II de genes independentes ou podem representar íntrons de genes codificadores de proteínas. O pri-miRNAs ao se dobrar, atua como um substrato para Drosha e Dicer, que são da família de enzimas de RNaseIII (KROL *et al.*, 2010).

Um pré-miRNA de aproximadamente 70 nucleotídeos, produto da clivagem da Drosha, é exportado para o citoplasma por meio da ação da exportina 5. No citoplasma a Dicer processa o pré-miRNA em um duplex miRNA/miRNA de aproximadamente 20pb e uma destas fitas do duplex (miRNA maduro) é incorporada em um complexo de miRNA de silenciamento induzido - miRISC (Figura 3) (KROL *et al.*, 2010). Os miRNAs, localizados no miRISC, realizam pareamento de suas bases para alcançar mRNA e incentivar sua repressão translacional e degradação (KROL *et al.*, 2010).

Durante o processo de maturação do miRNA, as enzimas Drosha e Dicer são auxiliadas por numerosos cofatores ou proteínas acessórias, no qual algumas realizam uma função regulatória essencial. O mesmo ocorre durante a formação de miRISC, ou seja, há o envolvimento de muitos fatores adicionais (KROL *et al.*, 2010).



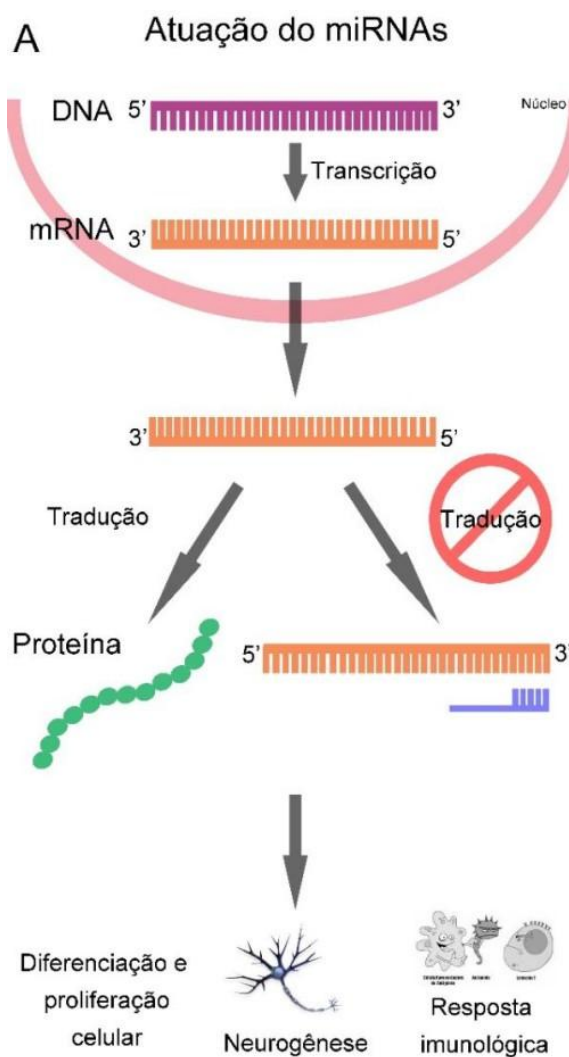
**Figura 3:** O Mecanismo da Biogênese de miRNA (AMERES & ZAMORE, 2013).

MiRNAs têm capacidade de regular alguns mecanismos biológicos e também está relacionado com o desenvolvimento de doenças inflamatórias crônicas ou neoplásicas, como é o caso do câncer gástrico que será abordado neste estudo. Um único miRNA, potencialmente, consegue aumentar ou diminuir a regulação de diversos alvos gênicos relacionados a mesma via metabólica (WU WK *et al.*, 2010).

Os miRNAs são importantes mecanismos de regulação pós-transcricional. Esta regulação pode ocorrer por meio de métodos distintos, dos quais dois estão bem estabelecidos: (i) o de clivagem do RNA mensageiro, e

(ii) o de repressão da tradução (RICARTE-FILHO & KIMURA, 2006; BARTEL, 2004). Se o pareamento do miRNA com a molécula de mRNA acontecer de forma imperfeita ocorrerá a repressão traducional, caso contrário, ocorrerá a clivagem, ou seja, o modo de atuação de um ou de outro método dependerá do tipo de pareamento miRNA - mRNA (RIBEIRO-DOS-SANTOS *et al.*, 2012).

A atuação dos miRNAs em nível pós-transcricional está representada na Figura 4. Esta atuação altera a expressão dos genes ao hibridizar com mRNAs e diminuir a sua tradução por processos que irão silenciar ou degradar o RNA mensageiro (mRNA) alvo. Entretanto, a intervenção dos miRNAs depende da sua concentração intracelular e da complementaridade das seqüências nucleotídicas entre miRNA e mRNA (AMBROS, 2004; DE SOUSA *et al.*, 2013; KINJO *et al.*, 2013). Quando ocorre a desregulação dos miRNAs, este fenômeno pode ocasionar redução ou excesso da síntese proteica, originando erros na diferenciação celular, neurogênese e resposta imunológica (WANG *et al.*, 2012).



**Figura 4:** Atuação dos microRNAs na região 3' da molécula do RNA mensageiro e suas conseqüências ao reprimir a tradução (MÓVIO *et al.*, 2015).

Neste sentido, tanto o processo de degradação de mRNA quanto o processo de repressão da tradução, proporcionam a regulação da expressão do mRNA pelo miRNA por meio da diminuição da expressão gênica (BARTEL, 2004). Como conseqüência, ocorrerá a diminuição do nível da proteína alvo, possibilitando ou não a redução dos níveis de RNA. De forma alternativa, caso haja a hipossupressão de um determinado miRNA, conseqüentemente haverá a

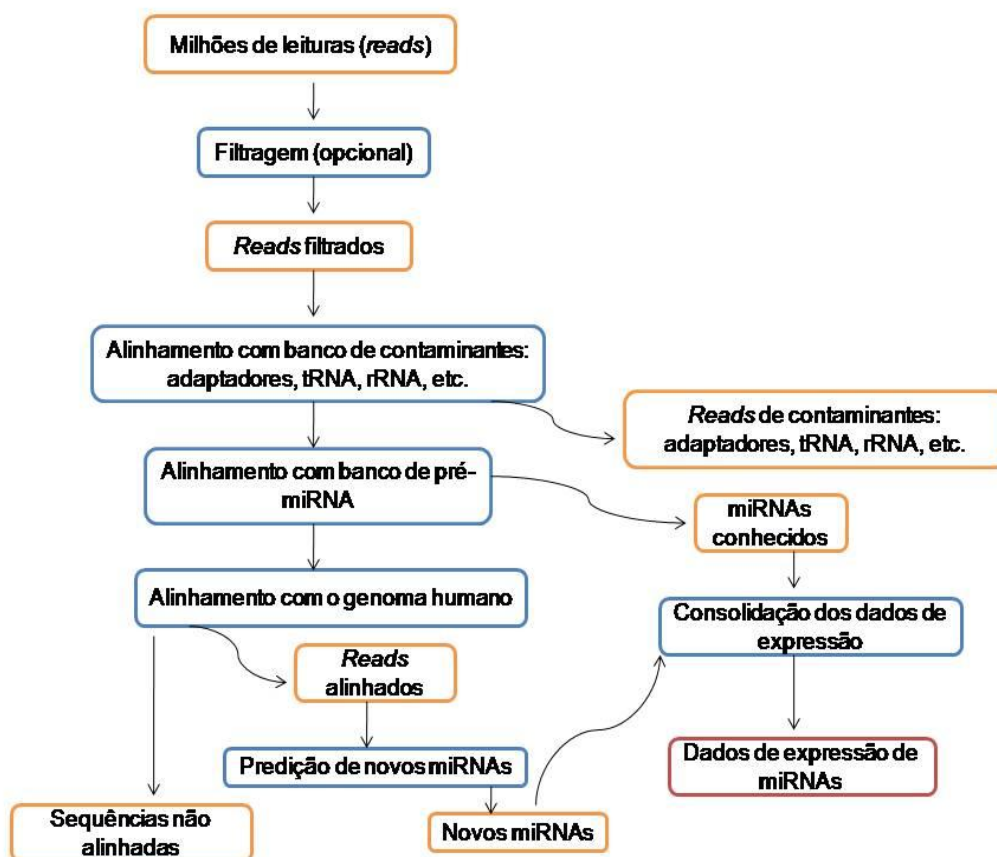
possibilidade de hiperexpressão do gene alvo, já que houve uma diminuição de expressão de seu mecanismo regulador (BEITZINGER *et al.*, 2007).

Diversas técnicas de análises de dados biológicos de grande porte foram desenvolvidas para verificar as expressões dos miRNAs. Entretanto, para que seja possível detectar a expressão de determinados miRNAs, o primeiro procedimento a ser realizado é o sequenciamento por meio de técnicas de NGS - *Next Generations Sequencing* ou Sequenciamento de Nova Geração. Posteriormente, cada molécula é sequenciada em plataforma de seqüenciamento *high-throughput*, tais quais as plataformas *Illumina GA*, *Life Technologies SOLID* e *Roche 454 LifeSciences*. Por meio dessas plataformas, serão geradas centenas de milhares a centenas de milhões de seqüências curtas, chamadas de *reads* (MARGUERAT, 2010; MARTIN, 2011; GERSTEIN e SNYDER, 2009).

Para análise de miRNomas, por meios de técnicas de NGS, algumas etapas são necessárias, como mostra a Figura 5. De acordo com Ribeiro-dos-Santos *et al.* (2012), apesar de que o esperado seja que os conjuntos de dados oriundos de experimentos de sequenciamento de bibliotecas cDNA identifiquem as seqüências de interesse, como os miRNAs, algumas outras seqüências também são identificadas, como os reagentes e contaminantes. O primeiro passo é realizar a identificação de moléculas conhecidas. Estas seqüências já identificadas são sistematicamente recolhidas dos dados para que sejam retirados também as seqüências contaminantes, *primers*, adaptadores e seqüências de pequenos RNAs conhecidos. Na próxima etapa, ocorre a detecção de miRNAs e alinha-se os dados remanescentes com um



banco de dados de pré-miRNA. Na etapa final, as sequências não detectadas são alinhadas com o genoma humano e posteriormente são investigadas por meio de algoritmos específicos na tentativa de detectar novos miRNAs, e será produzido um conjunto de sequências de miRNAs com suas respectivas contagens de *reads* para serem utilizados em análises de expressão.



**Figura 5:** Etapas da análise de miRNomas por meio de sequenciamento NGS (Ribeiro-dos-Santos *et al.*, 2012).

Neste contexto, para o estudo de perfis de miRNAs, habitualmente são utilizados três métodos: Quantificação por Reação em Cadeia da Polimerase em Tempo Real, ou RT-qPCR, Microarranjo (*Microarray*) e seqüenciamento de nova geração -NGS, como já citado anteriormente.

O método RT-qPCR, que se fundamenta na marcação fluorescente, possibilita a quantificação simultânea da expressão de um gene em diversas amostras quando se tem uma quantidade resumida de genes e, para miRNAs já conhecidos, permite um perfil mais individualizado para cada amostra (GALIVETI, *et al.*, 2010). Já o Microarranjo, esta é uma técnica em larga escala que possibilita realizar análises paralelas de centenas de miRNAs ao comparar o nível de expressão entre as amostras e os tecidos normais manipulados para a normalização (FARAZI, *et al.*, 2010). E a técnica NGS, seqüenciamento de nova geração, permite identificar novos miRNAs e detectar mutações, realizando a análise em várias amostras simultaneamente. Além disso, permite também estimar a quantidade de miRNAs com base na freqüência das seqüências lidas (FARAZI, *et al.*, 2010). Neste estudo, por exemplo, as amostras de miRNA retiradas do Atlas Genômico do Câncer foram seqüenciadas por técnicas de NGS, por meio da plataforma de seqüenciamento Illumina.

As vantagens da plataforma de seqüenciamento Illumina são o grande volume de seqüências que ela pode alcançar e o grande espectro de utilidade, mas principalmente o fato de ela ter o menor custo por base seqüenciada do mercado (GLENN, 2011).

As informações geradas nos experimentos de RNA-Seq proporcionam o estudo e a inferência dos mecanismos de regulação gênica e análise de expressão diferencial, até mesmo com os depósitos públicos do genoma do organismo estudado (KORPELAINEN *et al.*, 2015; MCGETTIGAN, 2012; VAN VERK *et al.*, 2013).

## 1.2 Análise Diferencial de Expressão Gênica

A análise diferencial de expressão gênica é outra forma de analisar os dados de RNA-Seq. Por meio dos dados de expressão diferencial de uma grande quantidade de genes em condições distintas, é possível reconstruir vias regulatórias dos genes; gerar previsões para identificar seus alvos; e assinalar novos genes associados a determinado processo (GONÇALVES, 2017).

A análise de expressão gênica diferencial em experimentos de RNA-Seq é realizada com pelo menos duas amostras, das quais uma é a referência (controle) e a outra o tratamento, sendo desejável existir mais de uma réplica para o controle e mais de uma para o tratamento. A utilização de réplicas aumenta a acurácia e garante a reprodutibilidade do experimento, pois será considerada a variação dentro dos tratamentos, bem como a variação entre tratamentos (AUER & DOERGE, 2010).

Neste estudo, é proposto o uso de *softwares* para análise de expressão diferencial de microRNAs (miRNAs) entre as amostras que foram seqüenciadas a partir de plataformas de RNA-Seq e que estão disponíveis em bases de dados públicas.

Neste sentido, quando as leituras saem do sequenciamento, a primeira etapa da análise diferencial é o mapeamento ou alinhamento contra um genoma ou um transcriptoma de referência. Essa etapa de mapeamento consiste em encontrar o lugar onde cada sequência alinhe de forma perfeita com a referência (OSHLACK et al., 2010; WILHELM e LANDRY, 2009). Após o alinhamento das *reads*, estas são sumarizadas de acordo com as características do objeto de estudo.

A próxima etapa consiste em uma lista de genes e uma tabela de contagens dos *reads*. A tabela de contagens pode ser gerada com as contagens de cada característica sumarizada para cada biblioteca sequenciada, no caso desse estudo, a tabela de contagem consiste em uma tabela com dados de expressão de genes e microRNAs para o caso de amostras cancerosas e amostras adjacentes ao tumor. Posteriormente, os dados são normalizados e testados estatisticamente para inferir a expressão diferencial (OSHLACK et al., 2010; WILHELM e LANDRY, 2009).

Por fim, possuindo a lista de genes diferencialmente expressos, pode-se utilizar abordagens de biologia de sistemas, como a produção de vias metabólicas, para gerar interpretações biológicas sobre os dados dos experimentos de RNA-Seq (OSHLACK et al., 2010).

Contudo, em se tratando de análise de expressão gênica por meio de RNA-Seq, a medida de expressão é uma variável discreta e se refere ao

número de *reads* alinhados a determinado gene. Então, já que a transformação dos dados de contagem não é bem aproximada por distribuições contínuas, alguns modelos estatísticos apropriados para dados de contagem foram sugeridos para retirar o máximo de informações a partir de dados provenientes do RNA-Seq (SOUZA, 2015).

Dentre as várias metodologias existentes para estudo da expressão diferencial, as que são baseadas em distribuição binomial (ROBINSON & SMYTH, 2008) se destacam, e por isso, são amplamente utilizadas. Tais metodologias, que na verdade são pacotes de *softwares*, se distinguem quanto aos conceitos adotados e aos princípios estatísticos empregados, como DESeq (ANDERS e HUBER, 2010), edgeR (ROBINSON *et al.*, 2010) e outros.

Neste estudo, utilizou-se a análise de expressão diferencial por meio dos pacotes DESeq2 (LOVE *et al.*, 2014), edgeR (ROBINSON *et al.*, 2010) e EBSec (LENG *et al.*, 2013). Esses pacotes são bibliotecas estatísticas desenvolvidas em R (R CORE TEAM, 2012) e distribuídas pelo projeto Bioconductor (GENTLEMAN *et al.*, 2004).

O pacote DESeq2, notoriamente, advém depois da criação do pacote DESeq (ANDERS & HUBER, 2010), que é baseado no princípio de que a contagem de *reads* que se encontram associadas a uma classe (genes e transcritos, por exemplo) está linearmente relacionada a abundância dessa classe (MORTAZAVI *et al.*, 2008). O seu objetivo é realizar testes estatísticos para analisar se o efeito de diferentes condições biológicas sobre a contagem de *reads* ocorreu ou não ao acaso (ANDERS & HUBER, 2010). Para isso, o pacote modela os dados utilizando a Distribuição Binomial Negativa, que oferece um modelo mais próximo da realidade do que a Distribuição de

Poisson, em se tratando da variação dos dados de contagem em experimentos de RNA-Seq (ROBINSON & SMYTH, 2007).

A Distribuição de Poisson possui apenas a média como parâmetro e esta, define sua variância e todas as suas outras propriedades, isso faz com que ocorra um fenômeno denominado superdispersão, no qual a variabilidade real dos dados é maior do que a distribuição de Poisson é capaz de prever. Ademais, com esta distribuição não é possível controlar o erro baseado na probabilidade de ocorrerem falsas descobertas (FDR) (ANDERS & HUBER, 2010; ROBINSON & SMYTH, 2007).

O modelo DESeq baseia sua modelagem sobre dados de contagem por meio de valores inteiros não negativos, os quais são viabilizados por uma tabela (Tabela1), onde as linhas representam as contagens em cada classe estudada e as colunas representam as distintas condições experimentais. Vale ressaltar que as condições experimentais são sempre duas, controle e tratamento, porém cada condição pode possuir várias réplicas. Na Tabela 1 o exemplo de dados é uma das tabelas de expressão de miRNAs utilizadas neste estudo, na qual a condição “01A” é expressão do tumor (caso) e a “11A” é a expressão do tecido adjacente (controle).

**Tabela 1:** Exemplo de dados de entrada para análise estatística pelo DESeq (Modificado de COVRE, 2012).

	TCGA.BR.6453. <b>01A</b>	TCGA.BR.6453. <b>11A</b>	TCGA.BR.6454. <b>01A</b>	TCGA.BR.6454. <b>11<sup>a</sup></b>
hsa-mir-101-1	92644	31740	56756	108570
hsa-mir-101-2	367	159	141	425
hsa-mir-103a-1	133063	41471	42043	39525
hsa-mir-103a-2	80	17	54	17
hsa-mir-106 <sup>a</sup>	58	39	116	34

Tendo em vista que os valores de contagem das classes são diretamente influenciados pela cobertura do sequenciamento, tratamentos com valores de contagem mais altos possivelmente não representam classes que foram diferencialmente expressas de fato. Para excluir esse tipo de tendência, o DESeq calcula para todos os tratamentos de fatores de tamanho que são estimativas do tamanho efetivo dos tratamentos (ANDERS & HUBER, 2010; ROBINSON & SMYTH, 2007).

Para realizar o cálculo da expressão diferencial, o DESeq fundamenta-se na estimativa da média e da variância entre as condições experimentais. Dado que o número de réplicas geralmente é limitado, este pacote foi desenvolvido para se ajustar ao desenho experimental do RNA-Seq, que normalmente é composto de poucas réplicas e muitas classes. Contudo, quando DESeq calcula os valores de média e variância para cada uma das

classes, ele escalona os valores de contagem com os fatores de tamanho (COVRE, 2012).

Na Tabela 2 apresenta-se o resultado gerado pelo DESeq contendo os valores de média para as condições experimentais (*baseMean*), o *foldchange* e seu logaritmo na base 2, o valor de *p* e esse mesmo valor ajustado pelo método de *Benjamini-Hochberg*. Explicando melhor esses valores: o valor de *foldchange* é calculado dividindo-se a média do tratamento pela média do controle para cada classe; o método de *Benjamini-Hochberg* é um ajuste usado em estudos que envolvem múltiplos testes de significância, possuindo assim o objetivo de controlar a taxa de falsas descobertas FDR - *False Discovery Rate* (proporção de erros causados pela rejeição equivocada da hipótese nula  $H_0$ ) (BENJAMINI & HOCHBERG, 1995).

**Tabela 2:** Exemplo de dados de saída para análise estatística por meio do DESeq (Modificado de COVRE, 2012).

Id	baseMean	baseMeanA	baseMeanB	Fold- Change	log2 foldCh ange	pval	padj
hsa-mir-1-2	1506,84	256,4167	2757,28188	10,74	3,42	2,8183E-21	1,0803E-19
hsa-mir-129-2	30,60	62,94680	50,3400206	4,72	2,23	2,5812E-10	1,799E-09
hsa-mir-133a-1	900,058	147,795	1652,32021	11,16	3,48	1,7428E-26	1,3361E-24
hsa-mir-139	530,52	151,361	909,686221	5,99	2,58	1,9213E-35	4,4191E-33
hsa-mir-143	2288743,4	909351,4	3668135,37	4,03	2,01	2,6681E-11	2,5569E-10



A Tabela 2 acima é um exemplo de dados utilizado neste estudo, onde se apresenta os miRNAs mais significativos como resultado de um dos testes de análise diferencial de expressão.

Neste contexto, após quatro anos do lançamento do pacote DESeq, LOVE *et al.*, apresentaram o pacote DESeq2. Este integra avanços metodológicos com vários recursos inovadores para facilitar uma análise mais quantitativa de dados comparativos de RNA-Seq. Dentre as vantagens estão melhor classificação e visualização de genes, testes de hipóteses acima e abaixo de um limiar, transformação de logaritmo regularizado para avaliação de qualidade e agrupamento de dados de contagem superdispersa, além de controlar a taxa de falsos positivos (LOVE *et al.*, 2014).

Diferente do DESeq, o DESeq2 utiliza um modelo linear generalizado para análise de expressão diferencial, além de, dentre outras características, fornecer uma matriz de fatores de normalização dependentes da amostra/gene, fazer filtragem independente automática na média de contagens normalizadas e detectar e manipular automaticamente valores muito elevados (LOVE *et al.*, 2014).

Por outro lado, o método edgeR (ROBINSON *et al.*, 2010), calcula a expressão diferencial por meio da estimativa de *Bayes* empírica e testes exatos fundamentados em um modelo binomial negativo. Este pacote foi desenvolvido para permitir análises de experimentos com um pequeno número de repetições.

O método empírico de Bayes é utilizado para moderar o grau de excesso de dispersão em todos os genes por meio do empréstimo de informações entre os genes. Um teste exato análogo ao teste exato de Fisher, porém adaptado a dados sobre-dispersos, é utilizado para verificar a expressão diferencial para cada gene. Com relação ao processo de normalização, por padrão o método de normalização TMM é realizado para reparar as diferentes profundidades de seqüenciamento entre as amostras. Enquanto isso, o procedimento de *Benjamini-Hochberg* é utilizado para controlar a FDR (BENJAMINI & HOCHBERG, 1995).

Em contrapartida, o modelo EBSeq (LENG *et al.*, 2013) foi desenvolvido principalmente para identificar isoformas diferencialmente expressas, mas também apresentou um resultado robusto nas análises de nível de genes. O pacote estima as probabilidades posteriores de expressão diferencial e expressão igual com a contribuição de métodos bayesianos empíricos, assumindo distribuição binomial negativa para os dados. Para tratar as diferentes profundidades de sequenciamento, um método de normalização mediano semelhante ao DESeq é usado como método padrão (ANDERS & HUBER, 2010).

Neste sentido, vários estudos analisaram o perfil de expressão em inúmeras situações por meio dos pacotes desenvolvidos em R. MASCARENHAS JUNIOR (2016) utilizou o pacote DESeq2 para analisar a expressão gênica diferencial entre o adenocarcinoma da junção esôfago-gástrica e o adenocarcinoma gástrico. Enquanto que SOBREIRO (2017) identificou genes diferencialmente expressos em resposta à baixa

disponibilidade hídrica em quatro espécies de ipês, também utilizando o modelo DESeq2.

Assim, para auxiliar na análise de expressão diferencial foi utilizado o *software R-peridot*, que contém todos estes pacotes mencionados anteriormente.

### **1.3 Software R-Peridot**

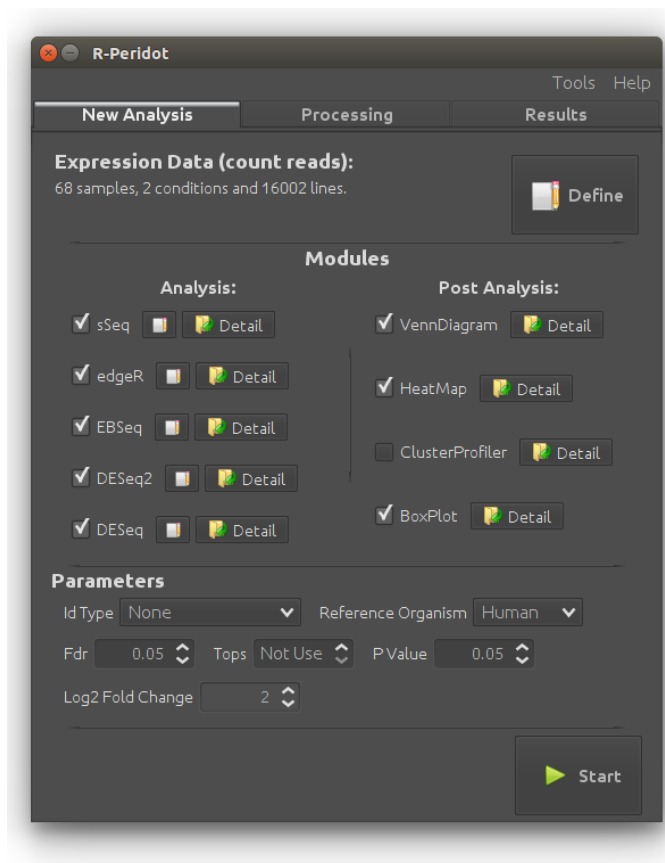
As etapas de análise diferencial dos dados utilizam meios computacionais para que se possam realizar inferências biológicas mais confiáveis possíveis (OSHLACK *et al.*, 2010). Diversos *softwares* podem ser utilizados em *pipelines* de análise de expressão diferencial em experimento por RNA-Seq (FONSECA *et al.*, 2012; GAO *et al.*, 2010; OSHLACK *et al.*, 2010; YENDREK *et al.*, 2012), um desses softwares denomina-se *R-peridot*.

O *R-peridot* é uma interface (Figura 6) que permite análises de expressão de RNASeq e para isso reúne vários pacotes do software R, como o DESeq2, DESeq, edgeR e outros. A escolha dos pacotes para análise de expressão fica a critério do usuário, podendo escolher entre um ou mais de um pacote, bem como definir quais parâmetros utilizar (Figura 7). Vale ressaltar que o *software* não analisa dados provenientes de *Microarray*.



**Figura 6:** Logotipo do software *R-peridot* (MARTINS, ALVES SOBRINHO, SOUZA, 2017)

Com a finalidade de identificar microRNAs (miRNAs) diferencialmente expressos entre tecido tumoral e adjacente, será utilizado o *software R-peridot* versão 1.0 desenvolvido por integrantes do Núcleo Multidisciplinar de Bioinformática (Bioinformatics Multidisciplinary Environment – BioME) pertencente a Universidade Federal do Rio Grande do Norte (UFRN) e que está disponível para *download* no endereço <http://work.bioinformatics-brazil.org/r-peridot/download.html>.



**Figura 7:** Interface do software *R-peridot* (MARTINS, ALVES SOBRINHO, SOUZA, 2017).

#### 1.4 Análise Funcional

Para entender a associação de genes envolvidos em funções semelhantes e nas mesmas vias metabólicas é necessário que se faça uma análise dos enriquecimentos funcionais, conseqüentemente essa análise resultará em inferências biológicas. Por meio de uma análise funcional, pode-se também prever funções de genes desconhecidos, como os RNAs não codificadores de proteínas (OSATO, 2018).

Neste estudo, foi realizada a análise funcional para obter conhecimento biológico (ontologia) acerca dos genes alvos de miRNAs diferencialmente expressos nos tecidos tumorais e adjacentes advindos do banco de dados TCGA. A anotação foi realizada também para descobrir quais as vias metabólicas dos genes alvos que estão relacionadas e se estão associadas com o processo de carcinogênese gástrica.

### 1.5 Análise Z-Score

Para cada um dos genes (tumor) identificados e associados com as 10 principais características do câncer, realizou-se o cálculo Z-Score para todos os indivíduos investigados. O Z-Score é um método habitual da estatística que relaciona um valor com a média de distribuição padrão em um grupo de valores, e neste trabalho, irá indicar o valor da expressão dos genes (tecidos tumorais) relacionado com a média de expressão das amostras de tecidos adjacentes.

O valor do Z-Score pode ser zero ou próximo de zero. Nesse caso, significa que esse valor está idêntico ao valor da média, ou seja, não possui variação. Por outro lado, o Z-Score pode apresentar valores positivos ou negativos, indicando que ele está acima ou abaixo da média, respectivamente. Assim, neste estudo, o valor de Z-score positivo identificará que a expressão de determinado gene está acima (*up regulated*) da média de distribuição padrão, enquanto que um Z-score negativo indicará que a expressão está abaixo (*down regulated*). Os valores Z-score positivos ou negativos indicam

também o número de desvios padrão em que este valor está acima ou abaixo da média.

A normalização das amostras pelo Z-Score, mostra mudanças significantes na expressão dos genes entre os diferentes tipos/fenótipos de amostras e condições (CHEADLE, 2003). Assim, neste estudo, o cálculo de Z-Score foi utilizado para verificar se existem diferenças significantes (alterações significantes) na expressão dos genes entre os distintos tipos de amostras tumorais e adjacentes e se os alvos encontrados possuem expressão regulada pelos miRNAs.

## 2 OBJETIVOS

### 2.1 Objetivo geral

Analisar, identificar e realizar associação de genes e microRNAs reguladores diferencialmente expressos em tecidos de câncer gástrico, no banco de dados TCGA, com o auxílio do *software R-peridot*.

### 2.2 Objetivos específicos

- Realizar as análises diferenciais de expressão de miRNAs pelo *software R-peridot* em amostras de tecidos de câncer gástrico e seus tecidos adjacentes;

- Analisar quais são os genes alvos dos miRNAs diferencialmente expressos por meio do miRTargetLink Human;

- Investigar quais as vias metabólicas envolvidas no processo com o auxílio da plataforma DAVID e String;

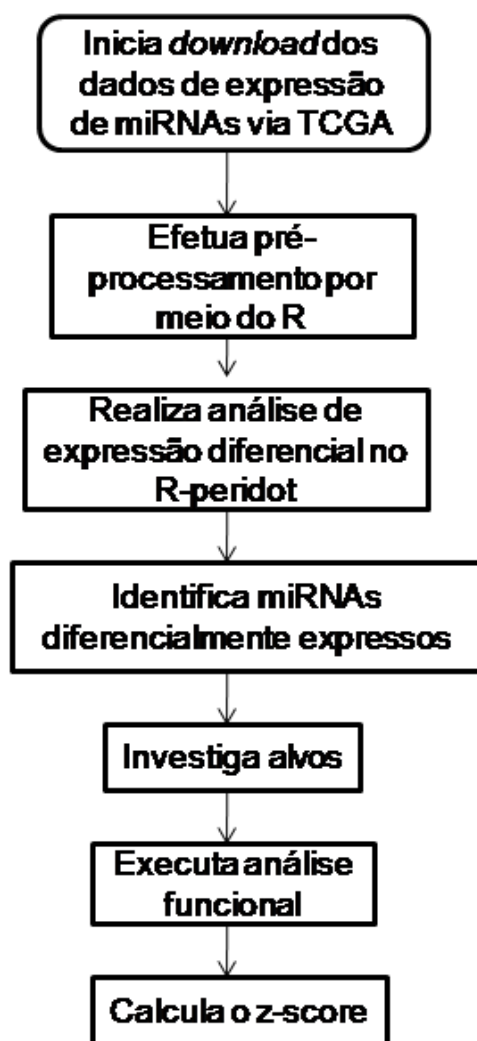
- Verificar por meio do Z-Score se os genes alvos são modificados pelos miRNAs diferencialmente expressos;

- Comparar os resultados de expressão diferencial obtidos pelo *software R-peridot* com outros softwares já estabelecidos na literatura científica.



### 3 METODOLOGIA

Para facilitar o entendimento e o desenvolvimento deste projeto e da metodologia empregada, a Figura 8 é apresentada resumindo passo a passo das etapas que o constituem.



**Figura 8:** Fluxograma linear da metodologia empregada neste estudo.

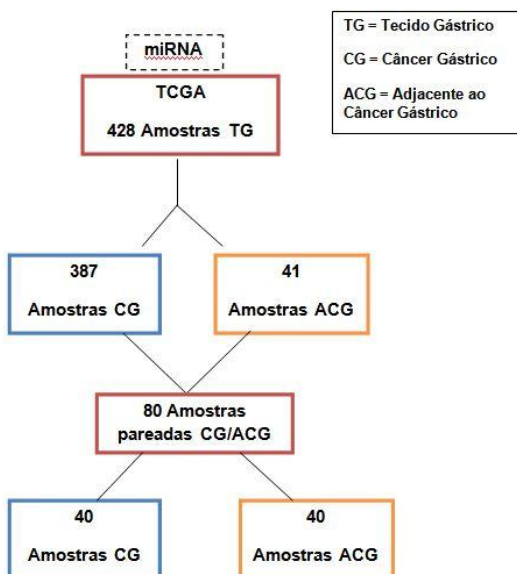
### 3.1 Origem dos Dados

Neste estudo, as amostras de expressão de microRNAs e genes no câncer gástrico, utilizadas para análise de expressão diferencial, são provenientes do Atlas Genômico do Câncer (TCGA).

O TCGA é um banco de dados público que congrega amostras sem tumor e amostras de tumor, deixando acessível ao público 37 tipos de dados genômicos e clínicos de 33 tipos distintos de câncer (ROBBINS *et al.*, 2013). Ressalta-se que as amostras denominadas como tecido adjacente ao câncer gástrico estão descritas como “tecido normal” pelo TCGA.

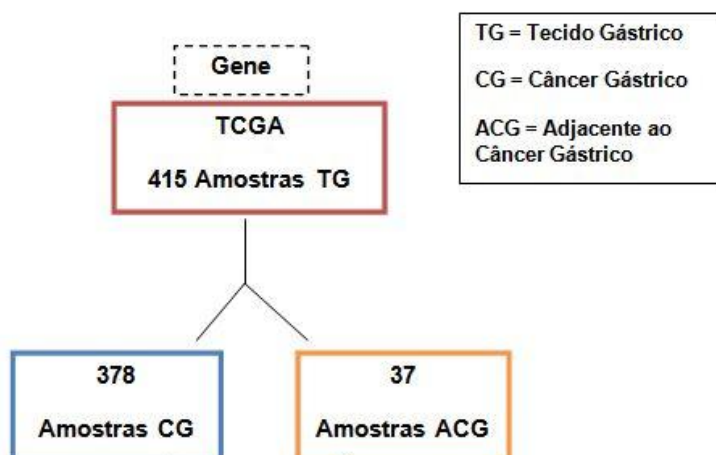
A plataforma de dados TCGA está disponível no site <https://portal.gdc.cancer.gov/> e é um artifício que tem como finalidade ser útil no desenvolvimento de novos mecanismos para a prevenção, diagnóstico e tratamento do câncer (CACCIATORE, 2014).

Os dados de miRnoma (expressão de miRNAs) de amostras de câncer gástrico (Figura 9) foram retirados do banco TCGA por meio da ferramenta *GDC Data Transfer Tool*. Estes dados estão medidos em dados de contagem (*read\_count*). As amostras utilizadas para análise de expressão diferencial foram 80 amostras pareadas (tumor e tecido adjacente) de miRNA.



**Figura 9:** Amostras de expressão de miRNA utilizadas no estudo.

Os dados de expressão gênica do câncer gástrico (Figura 10) também foram retirados do TCGA por meio da ferramenta *GDC Data Transfer Tool*. Para esta análise foram aplicadas todas as 415 amostras de genes.



**Figura 10:** Amostras de expressão de genes utilizadas no estudo.

### 3.2 Pré-Processamento

No passo que segue, todas as amostras passaram pela sincronização dos dados (do inglês, *merge*) em uma única tabela (neste caso, uma para miRNA e outra para gene) com todas as amostras tumorais e adjacentes e as suas expressões de miRNAs e de gene no câncer gástrico. Este *merge* foi efetuado pelo *script* em R, uma linguagem de programação muito utilizada na bioinformática.

O outro procedimento necessário para análise diferencial no *R-peridot* foi a criação do arquivo de condições incluindo a característica, se é tumor ou adjacente, para cada paciente. O arquivo de condições também foi gerado por um *script* em R.

Posteriormente, como intuito de aumentar a acurácia sobre a distribuição da variabilidade dos dados de expressão de miRNA no câncer gástrico, removemos os miRNAs que apresentavam baixa expressão, mais especificamente aqueles que estavam com expressão igual a zero. A expressão baixa pode dificultar a análise das diferenças de expressões e também prejudicar a avaliação dos resultados.

### 3.3 Análise Diferencial de miRNAs

Neste trabalho, a análise de expressão diferencial foi realizada por meio do *software R-peridot*. Os dados de entrada para o *software* foram as 80 amostras de expressão de miRNAs pareadas (40 amostras CG e 40 amostras ACG), para isso foram utilizados os pacotes EBSeq, edgeR e DESeq2. Os miRNAs diferencialmente expressos encontrados, são os miRNAs em comum aos três pacotes. É necessário ressaltar que foram feitos vários testes de

análise de expressão, utilizando parâmetros diferentes, entretanto, o teste mencionado neste trabalho foi o que apresentou melhor precisão.

Como explicitado anteriormente, para a análise da expressão diferencial de microRNAs (miRNAs) no câncer gástrico, apresenta-se dados pareados. Neste contexto, a Tabela 3 mostra quais parâmetros utilizados.

**Tabela 3:** Parâmetros utilizados no *R-peridot* para o teste de análise de expressão de miRNAs no câncer gástrico com amostras pareadas.

Pacotes	FDR	P-value	log2foldChange
DESeq2, EBSeq e edgeR	0.05	0.05	2

### 3.4 Ferramentas para Seleção e Análise Funcional de Genes Alvos

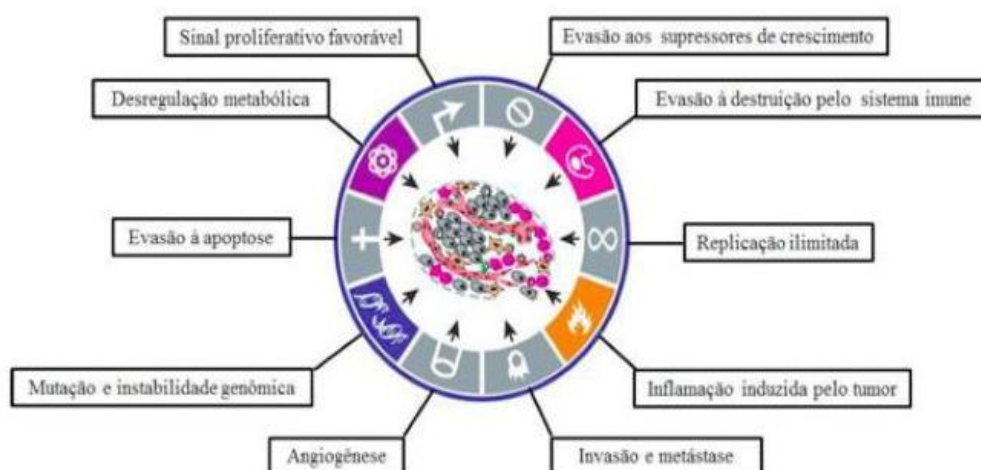
Após a realização da análise diferencial, pelo *software* R-peridot, obteve-se como resposta quais miRNAs foram diferencialmente expressos entre os dois tipos de tecidos, o tumoral e o adjacente. A partir deste resultado, foi realizada uma análise em busca dos genes alvos dos miRNAs diferencialmente expressos que já estão validados na literatura. Esta investigação foi efetuada por meio da plataforma miRTargetLink Human (HAMBERG *et al.*, 2016), disponibilizada no site <https://ccb-web.cs.uni-saarland.de/mirtargetlink/>.

Por conseguinte, com o objetivo de avaliar o enriquecimento funcional e encontrar as vias metabólicas do KEGG envolvidas no processo de cancerização, os alvos dos miRNAs foram aplicados na plataforma DAVID versão 6.8 (HUANG *et al.*, 2009a; 2009b) disponível em

<https://david.ncifcrf.gov/> e na base de dados STRING (<https://string-db.org/>) versão 10.5 (SZKLARCZYK *et al.*, 2017).

As vias encontradas foram posteriormente analisadas principalmente para averiguar a relação com as 10 principais características associadas ao desenvolvimento do câncer (HANAHAN & WEINBERG, 2000). De acordo com esses autores, à maioria dos tumores cancerígenos estão associados a um pequeno número de propriedades moleculares, bioquímicas e celulares.

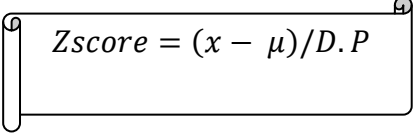
Neste sentido, os identificou-se seis (6) variações importantes (evasão aos supressores de crescimento, autossuficiência de sinais proliferativos, fuga à apoptose, potencial de replicação ilimitado, indução à angiogênese, ativação da invasão de tecidos e metástase) na fisiologia celular que contribuem para atingir um fenótipo maligno (HANAHAN & WEINBERG, 2000). Em 2011, os mesmos autores descreveram mais quatro (4) alterações fundamentais: desregulação energética celular, evasão à destruição pelo sistema imune, instabilidade genômica e mutação e inflamação induzida pelo tumor, configurando assim as 10 marcas do câncer (Figura 11).



**Figura 11:** As 10 características principais (marcas) que levam ao desenvolvimento do Câncer. (Modificado de HANAHAN & WEINBERG, 2011).

### 3.5 Cálculo do Z-Score de expressão de mRNA

O cálculo do Z-Score foi realizado por meio da fórmula:


$$Zscore = (x - \mu) / D.P$$

**Figura 12:** Fórmula para cálculo do Z-score.

Na fórmula acima,  $x$  é a expressão da amostra tumor,  $\mu$  é a média aritmética da expressão da amostra adjacente e D.P é o desvio padrão da amostra adjacente.

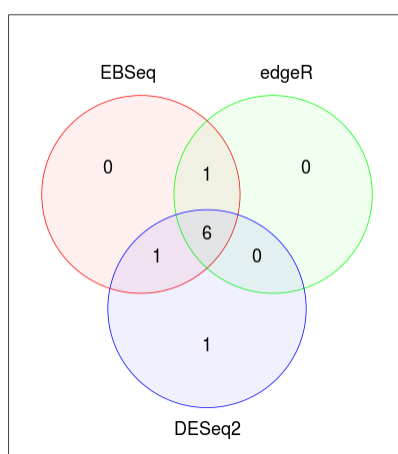
## 4 RESULTADOS

O principal objetivo no decorrer deste estudo foi analisar microRNAs (miRNAs) e genes que se apresentam diferencialmente expressos no câncer gástrico (CG), com o auxílio do *software R-peridot*.

### 4.1 Análise de miRNAs no *R-Peridot*

Em relação aos miRNAs, após a realização da análise diferencial, pelo *software R-peridot*, obteve-se como resposta quais estavam diferencialmente expressos entre os dois tipos de tecidos, o tumoral e o adjacente.

A análise de expressão de miRNAs das 80 amostras pareadas pelo *R-peridot*, identificou seis (6) miRNAs diferencialmente expressos, como demonstrado no Diagrama de Venn da Figura 13, entre os tecidos tumorais e seus adjacentes, em comum para os pacotes DESeq2, EBSeq e edgeR.



**Figura 13:** Diagrama de Venn para demonstrar os 6 miRNAs diferencialmente expressos em comum aos pacotes EBSeq, edgeR e DESeq2.



Na Tabela 4 pode-se verificar os 6 miRNAs grafados em vermelho como os miRNAs em comum aos pacotes estatísticos abordados e ainda a análise individual para cada pacote, demonstrando os miRNAs adicionais, ou seja, os que foram encontrados além dos 6 identificados anteriormente. Por exemplo, somente com o pacote DESeq2 apresenta-se os miRNAs adicionais ***hsa-mir-143*** e ***hsa-mir-129-5p***; para o EBSeq os miRNAs adicionais ***hsa-mir-653*** e ***hsa-mir-143***; e para o edgeR o miRNA ***hsa-mir-653*** como miRNA adicional.

**Tabela 4:** miRNAs diferencialmente expressos no CG pelo *R-peridot* em comum aos pacotes DESeq2, EBSeq e edgeR, e os miRNAs encontrados para cada pacote estatístico.

MiRNAs	DESeq2		EBSeq		edgeR	
	Log2FC	p-value	Log2FC	p-value	Log2FC	p-value
<b><i>hsa-mir-1-5p</i></b>	3,43	2,81E-21	-3,42	0	3,32	2,45E-18
<b><i>hsa-mir-133a-3p</i></b>	3,48	1,74E-26	-3,48	0	3,40	2,96E-23
<b><i>hsa-mir-139-5p</i></b>	2,58	1,92E-35	-2,58	0	2,46	2,76E-32
<b><i>hsa-mir-145-5p</i></b>	2,42	3,70E-20	-2,42	0	2,37	1,00E-18
<b><i>hsa-mir-183-5p</i></b>	-2,01	1,11E-10	2,019	1,53E-09	-2,06	3,48E-10
<b><i>hsa-mir-196b-5p</i></b>	-4,40	3,57E-27	4,41	0	-4,45	8,73E-22
<b><i>hsa-mir-143</i></b>	2,01	2,66E-11	-2,01	1,36E-10		
<b><i>hsa-mir-129-5p</i></b>	2,23	2,58E-10				
<b><i>hsa-mir-653</i></b>			2,03	0,0004	-2,00	1,00E-05

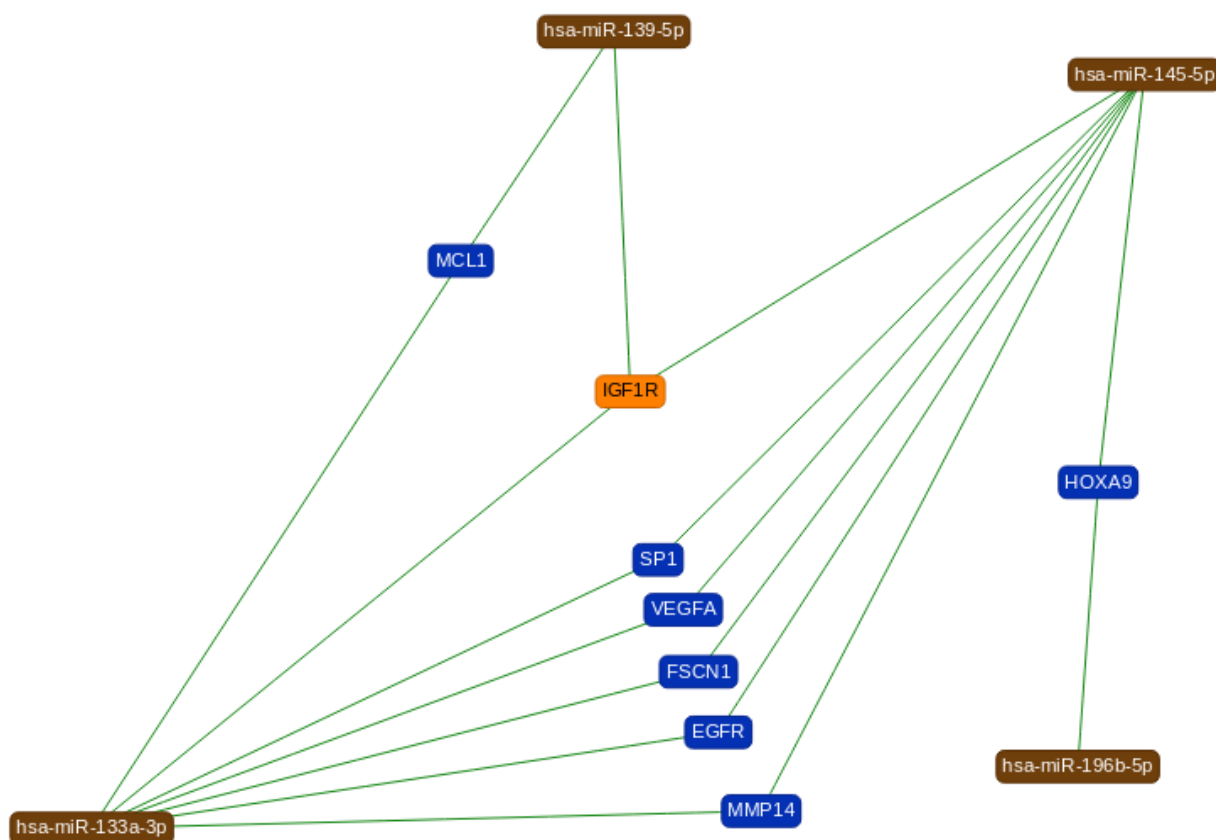
Posteriormente, foi realizada uma nova análise diferencial com os dados tratados, também por meio do *software R-peridot*, usando parâmetros mais restritivos, os mesmos empregados dentro do grupo de Bioinformática do LGHM. Os novos parâmetros retiraram da análise aqueles miRNAs os quais a soma dos *reads* era menor que 10. Com os dados tratados, efetuou-se uma nova análise diferencial. O resultado do *software R-peridot*, não conseguiu identificar miRNAs diferencialmente expressos considerando aos três pacotes estatísticos em conjunto. Desta forma, realizou-se uma nova rodada de análise utilizando os pacotes DESeq2, EBSeq e edgeR separadamente. Os resultados estão apresentados na Tabela 5 e indica os miRNAs diferencialmente expressos no CG.

**Tabela 5:** Resultados pelo R-peridot dos miRNAs diferencialmente expressos no CG com parâmetros mais restritivos e empregando os pacotes DESeq2, EBSeq e edgeR separadamente.

MiRNAs	DESeq2		EBSeq		edgeR	
	<i>Log2FC</i>	<i>p-value</i>	<i>Log2FC</i>	<i>p-value</i>	<i>Log2FC</i>	<i>p-value</i>
<i>hsa-mir-184</i>	3,02	1,23E-07	-2,81	0		
<i>hsa-mir-548y</i>			-3,48	0,015		
<i>hsa-mir-4728</i>			-2,04	1,47E-06		
<i>hsa-mir-4652</i>			-2,01	7,45E-07		
<i>hsa-mir-105-3p</i>			-2,32	2,15E-09		
<i>hsa-mir-105-5p</i>			-2,31	1,047E-09		
<i>hsa-mir-1269a</i>			-2,53	3,52E-13		
<i>hsa-mir-1224</i>			-4,23	0	3,96	4,07E-07
<i>hsa-mir-1269b</i>			-3,30	0		

## 4.2 Seleção dos Genes Alvos

Em busca dos genes alvos, ao todo foram identificados oito genes alvos para os seis miRNAs diferencialmente expressos pela plataforma *miRTargetLink Human*, todos genes validados e alvos de dois ou mais miRNAs investigados no presente trabalho (Figura 14).



**Figura 14:** Genes alvos considerando dois ou mais miRNAs diferencialmente expressos de acordo com a plataforma *miRTargetLink Human*.

Como apresentado na Figura 14, os miRNAs *miR-133a-3p* e *miR-139-5p* possuem um alvo em comum, o gene *MCL1*. Os miRNAs *miR-133a-3p*, *miR-139-5p* e *miR-145-5p* também possuem um alvo em comum, o gene *IGF1R*. Enquanto os miRNAs *miR-133a-3p* e *miR-145-5p* têm cinco alvos em comum, são eles os genes *SP1*, *VEGFA*, *FSCN1*, *EGFR* e *MMP14*. Por fim, os *miR-145-5p* e *miR-196b-5p* possuem o gene *HOXA9* em comum.

### 4.3 Análise funcional dos Genes Alvos

Em um primeiro momento, os oito genes alvos identificados (*MCL11*, *IGF1R*, *SP1*, *VEGFA*, *FSCN1*, *EGFR*, *MMP14* e *HOXA9*) foram submetidos a uma análise funcional por meio da plataforma DAVID, que encontrou as vias KEGG (Tabela 6) e a ontologia desses genes (Tabela 7).

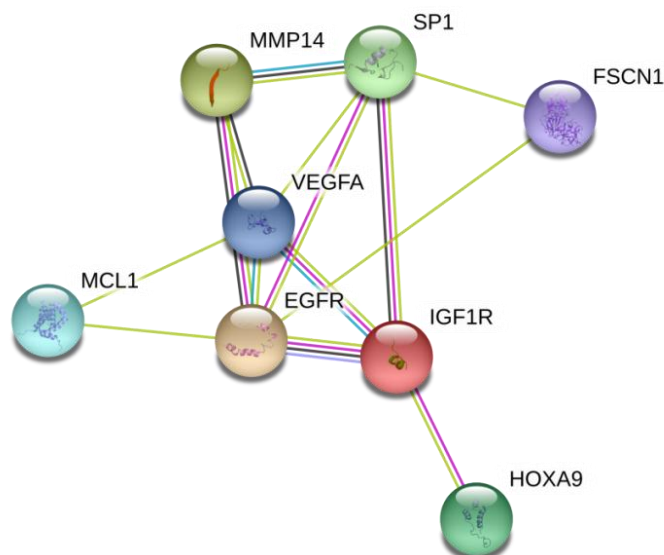
**Tabela 6:** Vias KEGG identificadas no CG pela plataforma DAVID.

ViaKEGG	p-value	Genes Alvos
Via de sinalização PI3K-Akt	0,00114445	<i>VEGFA</i> , <i>IGF1R</i> , <i>MCL11</i> , <i>EGFR</i>
Via de sinalização HIF-1	0,00193638	<i>VEGFA</i> , <i>IGF1R</i> , <i>EGFR</i>
Proteoglicanosem câncer	0,007868867	<i>VEGFA</i> , <i>IGF1R</i> , <i>EGFR</i>
Adesão Focal	0,008334598	<i>VEGFA</i> , <i>IGF1R</i> , <i>EGFR</i>
Via de sinalização Rap1	0,008652042	<i>VEGFA</i> , <i>IGF1R</i> , <i>EGFR</i>
Via de sinalização Ras	0,009977009	<i>VEGFA</i> , <i>IGF1R</i> , <i>EGFR</i>
MicroRNAs em câncer	0,015606886	<i>VEGFA</i> , <i>MCL1</i> , <i>EGFR</i>
Vias em câncer	0,028768445	<i>VEGFA</i> , <i>IGF1R</i> , <i>EGFR</i>
Câncer de Bexiga	0,029325564	<i>VEGFA</i> , <i>EGFR</i>
Glioma	0,046169831	<i>IGF1R</i> , <i>EGFR</i>
Câncer de Pâncreas	0,046169831	<i>VEGFA</i> , <i>EGFR</i>

**Tabela 7:** Ontologias identificadas no CG de acordo com a plataforma DAVID.

<b>Ontologia Gênica (GO)</b>	<b>p-value</b>	<b>Genes Alvos</b>
Desenvolvimento de estrutura anatômica	0,041109397	VEGFA, HOXA9, FSCN1, MCL1, MMP14, EGFR
Regulação da biogênese do componente celular	0,041237934	VEGFA, FSCN1, MMP14
Desenvolvimento imunológico	0,04256176	VEGFA, HOXA9, MMP14
Regulação da resposta ao estímulo	0,043690133	VEGFA, IGF1R, MCL1, MMP14, EGFR
Diferenciação celular feminina	0,044169747	VEGFA, MMP14 VEGFA, HOXA9, MCL1,
Diferenciação celular	0,044344322	MMP14, EGFR
Cascata de MAPK	0,0444822	VEGFA, IGF1R, EGFR
Regulação positiva da proliferação celular	0,045259912	VEGFA, IGF1R, EGFR VEGFA, IGF1R, MCL1,
Resposta ao estresse	0,045378996	MMP14, EGFR
Processo de desenvolvimento	0,046433247	VEGFA, HOXA9, FSCN1, MCL1, MMP14, EGFR
Proliferação de células endoteliais	0,046573774	VEGFA, MMP14
Regulação da replicação do DNA	0,047773843	IGF1R, EGFR
Regulação do processo de desenvolvimento	0,048009413	VEGFA, HOXA9, MMP14, EGFR
Transdução de sinal pela fosforilação de proteínas	0,048024212	VEGFA, IGF1R, EGFR
Organização de junção de aderentes	0,048573169	VEGFA, MMP14 VEGFA, HOXA9, SP1,
Regulação do processo biológico	0,04921599	FSCN1, IGF1R, MCL1, MMP14, EGFR
Diferenciação do cardiócito	0,04977108	VEGFA, EGFR
Resposta ao composto de nitrogênio	0,049835537	SP1, IGF1R, EGFR

Posteriormente, na base de dados STRING, repetiu-se a análise funcional com os mesmos genes alvos, e pôde-se encontrar a rede de interação formada entre estes genes (Figura 15): (a) os nós coloridos significam algum tipo de associação entre os genes; (b) as associações conhecidas são representadas pela linha em cor azul mais claro (interações oriundas de bancos de dados validados) e pela linha em cor rosa (interações determinadas experimentalmente); (c) as associações preditas são simbolizadas pela linha verde mais escura (interações relacionadas à vizinhança dos genes), pela linha vermelha (interações relacionadas à fusão dos genes) e pela linha azul mais escura (interações com co-ocorrência gênica); e (d) as outras interações são caracterizadas pela linha verde clara (interações com mineração de texto), pela linha preta (interações com co-expressão) e pela linha lilás (interações com homologia genética).



**Figura 15:** Rede de interação entre os genes alvos dos miRNAs diferencialmente expressos no CG.

No que diz respeito às vias KEGG, estas foram encontradas pela base de dados STRING e estão apresentadas na Tabela 8.

**Tabela 8:** Vias KEGG identificadas pela base de dados STRING no CG.

<b>Via KEGG</b>	<b>Número de Genes</b>	<b>FDR</b>	<b>Genes</b>
MicroRNAs em câncer	4	0,000045	<i>EGFR,FSCN1,MCL1,VEGFA</i>
Via de sinalização HIF-1	3	0,000652	<i>EGFR,IGF1R,VEGFA</i>
Via de sinalização PI3K-Akt	4	0,000652	<i>EGFR,IGF1R,MCL1,VEGFA</i>
Via de sinalização Ras	3	0,00267	<i>EGFR,IGF1R,VEGFA</i>
Via de sinalização Rap1	3	0,00267	<i>EGFR,IGF1R,VEGFA</i>
Adesão Focal	3	0,00267	<i>EGFR,IGF1R,VEGFA</i>
Proteoglicanos em câncer	3	0,00267	<i>EGFR,IGF1R,VEGFA</i>
Câncer de bexiga	2	0,00318	<i>EGFR,VEGFA</i>
Vias em câncer	3	0,00632	<i>EGFR,IGF1R,VEGFA</i>
Câncer de pâncreas	2	0,00632	<i>EGFR,VEGFA</i>
Glioma	2	0,00632	<i>EGFR,IGF1R</i>
Junção aderente	2	0,00704	<i>EGFR,IGF1R</i>
Melanoma	2	0,00704	<i>EGFR,IGF1R</i>
Via de sinalização GnRH	2	0,00942	<i>EGFR,MMP14</i>
Câncer de Próstata	2	0,00942	<i>EGFR,IGF1R</i>
Via de sinalização de estrogênio	2	0,0101	<i>EGFR,SP1</i>
Via de sinalização FoxO	2	0,0158	<i>EGFR,IGF1R</i>
Desregulação transcrricional no câncer	2	0,0276	<i>IGF1R,SP1</i>
Endocitose	2	0,0357	<i>EGFR,IGF1R</i>

#### 4.4 Análise das Vias KEGG e as Características do Câncer

Com a posse das vias metabólicas envolvidas neste processo, foi analisado quais dos oito genes alvos estão associados com as 10 características principais do câncer (Tabela 9).

**Tabela 9:** Genes alvos identificados e associados com as 10 principais características do câncer.

<b>10 Marcas do Câncer</b>	<b>Genes Alvos</b>
Auto-suficiência de sinais proliferativos	<i>VEGFA, IGFR1, EGFR</i>
Evasão aos supressores de crescimento	<i>VEGFA, HOXA9, MCL1, MMP14, EGFR, FSCN1</i>
Fuga à apoptose	<i>VEGFA, IGF1R, EGFR</i>
Potencial de replicação ilimitado	<i>IGF1R, EGFR</i>
Indução à angiogênese	<i>EGFR, IGF1R, VEGFA</i>
Ativação da invasão de tecidos e metástase	<i>VEGFA, MCL1, EGFR, FSCN1, HOXA9, MMP14</i>
Desregulação energética celular	<i>IGF1R, EGFR, VEGFA</i>
Evasão à destruição pelo sistema imune	<i>VEGFA, HOXA9, MMP14, IGF1R</i>
Instabilidade genômica e mutação	<i>VEGFA, EGFR, IGF1R</i>
Inflamação induzida pelo tumor	<i>VEGFA, HOXA9, MCL1, MMP14, EGFR, IGF1R</i>



#### 4.5 Resultados para o Z-Score

Nesta análise, o esperado é que o valor do Z-Score para o gene esteja o mais distante possível do zero, denotando que a sua expressão possivelmente é regulada pelos miRNAs investigados. Caso o contrário, pode ser que o alvo não esteja associado diretamente a regulação do miRNA.

Seguindo este princípio, o resultado mostrou que os genes *IGFR1*, *EGFR* e *FSCN1* apresentaram valores de Z-Score negativos e baixos, enquanto que os genes *VEGFA*, *MMP14*, *MCL1* e *SP1* apresentaram valores bem distantes do zero para a maioria dos indivíduos (Tabela 10). Ressalta-se que o gene *HOXA9* não foi encontrado nas amostras tumorais para o cálculo do Z-Score e por isso não é apresentado na Tabela 10.

**Tabela 10:** Valores de Z-Score dos genes alvos nas amostras de CG do banco de dados TCGA.

Gene	TCGA.AB46.01A	TCGA.AB47.01A	TCGA.4267.01A	TCGA.4280-01A
<b>VEGFA</b>	<b>21274</b>	<b>6471</b>	<b>26275</b>	<b>17759</b>
<i>IGFR1</i>	-0,736767558	-0,264935822	0,271976153	-0,107373138
<i>EGFR</i>	-0,84652243	-0,706954502	-0,099881166	-0,101885097
<b>MMP14</b>	<b>17719</b>	<b>12524</b>	<b>70095</b>	<b>12949</b>
<b>MCL1</b>	<b>27184</b>	<b>40983</b>	<b>65012</b>	<b>33591</b>
<i>FSCN1</i>	-0,130503598	0,577451092	0,28017464	-0,392047322
<b>SP1</b>	<b>17887</b>	<b>15736</b>	<b>39060</b>	<b>19150</b>

## 5 DISCUSSÃO

O adenocarcinoma gástrico engloba cerca de 10% dos novos casos de câncer diagnosticados, e é a segunda principal causa de morte por câncer (CUNNINGHAM *et al.*, 2006; LIU *et al.*, 2018). Por isso, é necessário entender os mecanismos moleculares envolvidos durante a neoplasia gástrica, bem como encontrar novos alvos terapêuticos para contribuir no melhor prognóstico da doença.

Neste estudo, foram encontrados seis (6) miRNAs diferencialmente expressos (*hsa-mir-1-5p*, *hsa-mir-133a-3p*, *hsa-mir-139-5p*, *hsa-mir-145-5p*, *hsa-mir-183-5p* e *hsa-mir-196b-5p*) associados a oito (8) genes alvos, entre as amostras tumorais e seus correspondentes tecidos adjacentes no câncer gástrico, depositados no banco de dados TCGA.

As análises das vias funcionais KEGG identificaram que alguns desses genes importantes são enriquecidos pelas vias associadas ao câncer e miRNAs em câncer, adesão focal, via de sinalização RAP1 e consequentemente pela via de sinalização de cálcio.

De acordo com CANCE *et al.* (2013), a via de adesão focal está relacionada com algumas características como invasão, metástase, angiogênese, manutenção de células-tronco cancerígenas e favorecimento de células tumorais, o que sugere o seu envolvimento com a progressão do câncer. Ainda existem poucos estudos que abordam a participação da via de sinalização RAP1 no câncer (NEAPOLITAN *et al.* 2015; ZHANG *et al.* 2017), entretanto, este estudo sugere esta via como importante e deve ser observada, principalmente na neoplasia gástrica. No que tange à via de sinalização de

cálcio, esta é uma fundamental via pela qual o acontecimento de distúrbios dos canais ou receptores de cálcio, se for identificado, levarão a doenças como o câncer (ELSHOLZ *et al.*, 2014). Além disso, ela pode regular a maioria dos processos fisiológicos, perpassando pela proliferação celular até a apoptose (CARAFOLI, 2003).

Neste sentido, o presente estudo obteve oito genes alvos (*MCL1*, *IGF1R*, *SP1*, *VEGFA*, *FSCN1*, *EGFR*, *MMP14* e *HOXA9*) de miRNAs diferencialmente expressos, no qual quatro destes genes: *MCL1*, *SP1*, *VEGFA* e *MMP14* estão associados a processos funcionais importantes na carcinogênese, o que sugere a participação destas vias no adenocarcinoma gástrico.

A realização do cálculo do Z-Score, para as amostras tumorais mostraram que quatro (4) destes genes (*MCL1*, *SP1*, *VEGFA* e *MMP14*) se destacaram de forma significativa. Portanto, no presente estudo, confirma-se que estes genes de amostras tumorais de câncer gástrico foram modificados por miRNAs diferencialmente expressos e estão presentes em vários trabalhos que os associam a esta neoplasia (HUANG *et al.*, 2018; LEE *et al.*, 2015; QIU *et al.*, 2014).

Adicionalmente, alguns trabalhos já publicados pelo nosso grupo ratificam a atuação dos miRNAs na regulação do câncer gástrico. Vidal *et al.* (2017) identificaram cinco miRNAs candidatos mais expressos e potencialmente regulados por cinco circRNAs em tecidos gástricos. Dentre eles, o *hsa-mir-1* e *hsa-mir-145-5p*, encontrados também nesta pesquisa. PEREIRA, A. (2018) ao realizar a análise de expressão diferencial entre tecidos de câncer gástrico e seus correspondentes tecidos adjacentes, observou

somente o miRNA *hsa-mir-196-5p* como diferencialmente expresso. Neste mesmo trabalho, também foi identificado que o miRNA *hsa-mir-145-5p* apresentou elevada expressão em tecidos adjacentes ao câncer gástrico.

Por outro lado, destaca-se a importância de se utilizar as análises de bioinformática corretas, ou seja, o mais próximo da realidade, para validação dos dados. Neste estudo, utilizou-se o software R-peridot para análise diferencial de expressão de miRNAs no câncer gástrico e notou-se que ao utilizar os pacotes DESeq2, EBSeq e edgeR em conjunto, os miRNAs encontrados se diferenciavam de quando se utilizava os pacotes separadamente, tanto em número de miRNAs quanto no miRNA em si.

Da mesma forma, em outro teste, quando houve a normalização dos dados de expressão de miRNAs, utilizando parâmetros de bioinformática mais restritivos, o R-peridot não identificou miRNAs diferencialmente expressos em conjunto aos três pacotes, mas quando analisados separadamente, cada pacote encontrou um número de miRNAs e miRNAs diferentes da análise anterior (*hsa-mir-184*, *hsa-mir-548y*, *hsa-mir-4728*, *hsa-mir-4652*, *hsa-mir-105-3p*, *hsa-mir-105-5p*, *hsa-mir-1269a*, *hsa-mir-1224* e *hsa-mir-1269b*).

Como método de comparação e validação dos resultados obtidos pelo *R-peridot* dos seis (6) miRNAs diferencialmente expressos entre amostras de câncer gástrico e amostras adjacentes, utilizou-se os mesmos parâmetros adotados pelo grupo da Bioinformática do Laboratório de Genética Humana e Médica (LGHM) da Universidade Federal do Pará (UFPA). Para isso foi aplicado somente o pacote DESeq2 no RStudio, para análise de expressão

diferencial, ou seja, sem a utilização do programa *R-peridot*. Esta análise mostrou 57 miRNAs diferencialmente expressos como segue na Tabela 11.

Pode-se observar ainda, que dentre os 57 microRNAs identificados pelo pacote DESeq2, cinco (5) microRNAs (*hsa-mir-1-5p*, *hsa-mir-133a-3p*, *hsa-mir-139-5p*, *hsa-mir-145-5p* e *hsa-mir-196b-5p*) dos seis (6) miRNAs encontrados anteriormente pelo *R-peridot* com os pacotes DESeq2, EBSeq e edgeR (Tabela 4) estavam presentes.

**Tabela 11:** 57 miRNAs diferencialmente expressos no CG pelo pacote DESeq2 por meio do RStudio.

<b>MiRNAs</b>	<b>Log2FC</b>	<b>p-value</b>	<b>Padj</b>
<i>hsa-mir-196a-1</i>	4,74579771	5,4773E-36	3,1111E-33
<i>hsa-mir-139</i>	-2,6455244	2,3016E-35	6,5365E-33
<i>hsa-mir-196b</i>	4,408271	1,1613E-30	2,1988E-28
<i>hsa-mir-133a-1</i>	-3,56396735	5,4434E-27	7,7296E-25
<i>hsa-mir-1-2</i>	-3,49452693	4,4851E-22	3,1844E-20
<i>hsa-mir-145</i>	-2,50809974	1,4971E-21	9,4484E-20
<i>hsa-mir-552</i>	5,49824734	4,2495E-20	2,4137E-18
<i>hsa-mir-135b</i>	3,16140121	1,0588E-19	5,4674E-18
<i>hsa-mir-133b</i>	-2,9317548	1,4088E-18	6,6681E-17
<i>hsa-mir-196a-2</i>	3,88085502	3,0873E-18	1,3489E-16
<i>hsa-mir-129-1</i>	-2,43819299	1,5348E-17	6,2267E-16
<i>hsa-mir-133a-2</i>	-2,90153097	2,6738E-17	1,0125E-15
<i>hsa-mir-1269<sup>a</sup></i>	6,4067544	7,2926E-16	2,3012E-14
<i>hsa-mir-490</i>	-4,41125353	2,3742E-15	6,1296E-14
<i>hsa-mir-129-2</i>	-2,20080192	2,7184E-15	6,5272E-14
<i>hsa-mir-767</i>	5,63695807	5,6322E-14	1,0664E-12
<i>hsa-mir-5683</i>	-3,16256133	1,8618E-13	3,4113E-12
<i>hsa-mir-6720</i>	-2,20095256	4,7508E-13	8,1771E-12
<i>hsa-mir-615</i>	2,58542634	9,2029E-12	1,4128E-10

<i>hsa-mir-23c</i>	-2,19912143	9,5794E-12	1,4319E-10
<i>hsa-mir-143</i>	-2,07900464	2,1831E-11	3,1E-10
<i>hsa-mir-204</i>	-2,0910232	4,4934E-11	5,9354E-10
<i>hsa-mir-937</i>	2,20999993	1,8476E-10	2,1618E-09
<i>hsa-mir-935</i>	2,42524017	6,902E-10	6,8778E-09
<i>hsa-mir-548f-1</i>	3,44974865	1,7633E-09	1,6154E-08
<i>hsa-mir-3648</i>	2,40074668	2,4426E-09	2,1678E-08
<i>hsa-mir-105-1</i>	4,22215358	3,6895E-09	2,9938E-08
<i>hsa-mir-3687</i>	2,67571529	3,6855E-09	2,9938E-08
<i>hsa-mir-1258</i>	-2,03762517	7,1276E-09	5,2577E-08
<i>hsa-mir-184</i>	3,61068425	7,9299E-09	5,7746E-08
<i>hsa-mir-885</i>	-2,56718751	9,0502E-09	6,4256E-08
<i>hsa-mir-4652</i>	4,88352706	2,0551E-08	1,3265E-07
<i>hsa-mir-1-1</i>	-2,84042551	6,8941E-08	4,122E-07
<i>hsa-mir-548ba</i>	-2,84805818	8,908E-08	5,2162E-07
<i>hsa-mir-509-1</i>	2,92121679	4,0511E-07	2,2125E-06
<i>hsa-mir-549<sup>a</sup></i>	2,90707349	5,2241E-07	2,7223E-06
<i>hsa-mir-9-3</i>	-2,11116326	6,7046E-07	3,3701E-06
<i>hsa-mir-187</i>	-2,2682189	9,1386E-07	4,4365E-06
<i>hsa-mir-105-2</i>	3,76525445	2,2079E-06	1,0279E-05
<i>hsa-mir-663<sup>a</sup></i>	3,00780089	5,8539E-06	2,427E-05
<i>hsa-mir-3176</i>	2,21723935	8,328E-06	3,4031E-05
<i>hsa-mir-6507</i>	-2,72193724	9,4299E-06	3,772E-05
<i>hsa-mir-3662</i>	2,36915455	1,2359E-05	4,8082E-05
<i>hsa-mir-509-2</i>	2,30900881	4,6024E-05	0,00016545
<i>hsa-mir-509-3</i>	2,26325863	6,2241E-05	0,00021557
<i>hsa-mir-4466</i>	2,01715802	8,8521E-05	0,00028897
<i>hsa-mir-4705</i>	-2,29761825	0,00013307	0,00041304
<i>hsa-mir-1911</i>	4,08202973	0,00014345	0,00044042
<i>hsa-mir-3189</i>	2,10304156	0,00020388	0,00060002
<i>hsa-mir-573</i>	2,30224358	0,00030232	0,00084589
<i>hsa-mir-5589</i>	3,26694454	0,00047653	0,00127075
<i>hsa-mir-6510</i>	-4,00410136	0,00086997	0,00218649
<i>hsa-mir-934</i>	2,26303968	0,00092129	0,00230525
<i>hsa-mir-944</i>	-2,07736651	0,0009342	0,00231715
<i>hsa-mir-7641-1</i>	2,38812933	0,00182525	0,00431777
<i>hsa-mir-4713</i>	2,29701241	0,00346002	0,00755881
<i>hsa-mir-1269b</i>	4,47977452	0,00362298	0,0078544

---

Neste contexto, nota-se que quando se utiliza os três pacotes em conjunto para análise diferencial, o resultado torna-se mais restritivo, os pacotes tornam-se mais robustos e por isso tem-se um número menor de miRNAs, tendo em vista que cada pacote possui a sua forma de tratar e analisar os dados estatisticamente. Outro ponto de vista que vale ressaltar é a importância de verificar a versão do *software* R e de seus pacotes estatísticos no momento da análise diferencial de expressão, já que este pode ser um fator que irá influenciar no resultado.

A relevância da validação dos dados de bioinformática para qualquer análise biológica humana se dá principalmente pelo fato de se tratar de uma vida, no sentido de o resultado da análise servir de alvo terapêutico futuro para o paciente com neoplasia gástrica, como é o caso deste estudo. A maioria dos alvos encontrados neste trabalho são possíveis tratamentos para o câncer gástrico, bem como os miRNAs, porém foram necessárias vários testes estatísticos para que isto pudesse ser confirmado, por isso o paciente deve ser visto como “ator” principal durante a realização das análises de bioinformática.

Adicionalmente, a identificação correta e precisa dos miRNAs e seus genes alvos é indispensável, visto que cada miRNA e gene possui uma característica que irá atuar positiva ou negativamente no tratamento da neoplasia gástrica. Como exemplo, no estudo de Liu *et al.* (2015), ao comparar tecidos tumorais e seus correspondentes adjacentes no câncer gástrico, foi

detectado que o *miR-133b/ a-3p* estava diferencialmente expresso e que atuava como supressor tumoral regulando as moléculas *MCL-1* e *BCL-XL* por meio de experimentos *in vitro* e *in vivo*. Outros estudos propõem que a rede regulatória miRNAs-IGF1R/IRS1 realiza uma atuação fundamental na carcinogênese e progressão da neoplasia (GONG *et al.*, 2015).

Independente, os cinco (5) miRNAs encontrados anteriormente somente com o pacote DESeq2, corroboram a significância estatística dos miRNAs identificados e sua relação com o câncer gástrico. No presente estudo, confirma-se o papel importante desempenhado pelos miRNAs no mecanismo da carcinogênese durante a atuação de seus genes alvos. A validação destes genes foi efetuada por meio do cálculo que identificou os valores de Z-Score “significantes” (valores positivos e acima da média) para os genes *VEGFA*, *MMP14*, *MCL1* e *SP1*. Estes valores propõem que estes últimos, foram modificados por miRNAs diferencialmente expressos no momento em que são positivamente regulados, o que os torna possíveis alvos terapêuticos que auxiliarão no prognóstico da doença .



## 6 CONCLUSÃO

Conclui-se que na análise de expressão de amostras tumorais e amostras adjacentes, foram descobertos seis miRNAs diferencialmente expressos (*hsa-mir-1-5p*, *hsa-mir-133a-3p*, *hsa-mir-139-5p*, *hsa-mir-145-5p*, *hsa-mir-183-5p* e *hsa-mir-196b-5p*), e oito genes alvos (*MCL11*, *IGF1R*, *SP1*, *VEGFA*, *FSCN1*, *EGFR*, *MMP14* e *HOXA9*). Destes, quatro são essenciais no processo da carcinogênese gástrica (*VEGFA*, *MMP14*, *MCL1* e *SP1*) e estão relacionados a três miRNAs já validados em trabalhos anteriores (*hsa-mir-133a-3p*, *hsa-mir-139-5p* e *hsa-mir-145-5p*). Vale ressaltar a importância das análises de bioinformática na validação dos dados para um melhor prognóstico da neoplasia.

## 7 REFERÊNCIAS

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WATSON, J. D., WALTER, P. *Biologia Molecular da Célula*. 5 ed. Porto Alegre: ArtMed, 2010.

AMBROS V, LEE RC (2004) Identification of microRNAs and other tiny non-coding RNAs by cDNA cloning. *Methods Mol Biol* 265: 131–158.

AMERES, S.L; ZAMORE, P.D. Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* 2013; 14:475-488.

ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biology**, 11:R106. 2010.

ASSUMPÇÃO PP, BURBANO RR. Genética e câncer gástrico. In: Linhares E, Laércio L and Takeshi S (eds). *Atualização em Câncer-Gástrico*. 1st ed. Tecmed Editora, São Paulo, 2005.

AUER, P. L.; DOERGE, R. W. Statistical design and analysis of RNA sequencing data. *Genetics*, v. 185, n. 2, p. 405-16, 2010.

BARTEL, David P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, v. 116, n. 2, p. 281-297, 2004.

BEITZINGER, Michaela et al. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA biology*, v. 4, n. 2, p. 76-84, 2007.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society**, v. 57, n. 1, p. 289-300, 1995.

BETEL D, WILSON M, GABOW A, MARKS DS, SANDER C (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res* 36: D149–53.D.

CACCIATORE, PAUL. In-Depth Review: The Cancer Genome Atlas Reports On Landmark Analysis of High-Grade Serous Ovarian Cancer.

CANCE WG, KURENOVA E, MARLOWE T, GOLUBOVSKAYA V. *Disrupting the scaffold to improve focal adhesion kinase-targeted cancer therapeutics. Sci Signal*. 2013;6(268):e10.

CARAFOLI E (2003) The calcium-signalling saga: tap water and protein crystals. *Nat Rev Mol Cell Biol* 4,326–332.

CHEADLE C, VAWTER MP, FREED WJ, BECKER KG. Analysis of microarray data using Z score transformation. *J Mol Diagn* (2003) 5:73–81.

COVRE, R. F. Wedring: Pipeline para Análise de Expressão Diferencial em Experimentos de RNA-Seq. 2013. 88 f. Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Educação Profissional e Tecnológica, Programa de Pós-Graduação em Bioinformática. Defesa: Curitiba, 14/05/2013.

CUNNINGHAM SC, KAMANGAR F, KIM MP, HAMMOUD S, HAQUE R, LACOBUZIO-DONAHUE CA, MAITRA A, ASHFAQ R, HUSTINX S, HEITMILLER RE et al. (2006) Claudin-4, mitogen-activated protein kinase kinase 4, and stratifin are markers of gastric adenocarcinoma precursor lesions. *Cancer Epidemiol Biomarkers Prev* 15, 281–287.

DE ROBERTIS, E. M. F; HIB, José. De Robertis bases da biologia celular e molecular. 4. ed. Rio de Janeiro: Guanabara Koogan, 2012. xii, 418 p.

DE SOUSA, E. et al. (2013). Developmental and Functional Expression of miRNA-Stability Related Genes in the Nervous System. *PLoS One*, 8(5):e56908.

DURÃES, C. et al. Biomarkers for gastric cancer: prognostic, predictive or targets of therapy?. *Virchows Archiv*,[s.l.], v. 464, n. 3, p.367-378, 1 fev. 2014.

ELSHOLZ F, HARTENECK C, MULLER W and FRIEDLAND K (2014) Calcium – a central regulator of keratinocyte differentiation in health and disease. *Eur J Dermatol* 24,650–661.

EMA, A; YAMASHITA, K; USHIKU, H; KOJO, K; MINATANI, N; KIKUCHI, M; MIENO, H; MORIYA, H; HOSODA, K; KATADA, N. Immunohistochemical analysis of RTKs expression identified HER3 as a prognostic indicator of gastric cancer. *Cancer Sci*. 2014; 105(12):1591–1600.

EWANG-EMUKOWHATE, M.; PERERA, D.; WIERZBICKI, A. S. Dyslipidaemia Related to Insulin Resistance and Cardiovascular Disease in South Asian and West African Populations. **Current Pharmaceutical Design**, 20 jun. 2014.

FARAZI TA, SPITZER JI, MOROZOV P, TUSCHL T. miRNAs in human cancer. *J Pathol*. 2010 Oct 14. PubMed PMID: 21089039. Epub 2010/11/20. Eng.

FERLAY, J.; SOERJOMATARAM, I.; ERVIK, M.; DIKSHIT, R.; ESER, S.; MATHERS, C.; REBELO, M.; PARKIN, D.; FORMAN, D.; BRAY, F. **GLOBOCAN 2018 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase. No. 11 [Internet]**. Disponível em: <<http://globocan.iarc.fr>>. Acesso em: 01 fev. 2018.

FONSECA, N. A.; RUNG, J.; BRAZMA, A.; MARIONI, J. C. Tools for mapping high throughput sequencing data. **Bioinformatics (Oxford, England)**, v. 28, n. 24, p.3169-3177, 2012.

GALIVETI CR, ROZHDESTVENSKY TS, BROSIUS J, LEHRACH H, KONTHUR Z. Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR. *RNA*. 2010 Feb;16(2):450-61. PubMed PMID: 20040593. Pubmed Central PMCID: 2811673. Epub 2009/12/31. eng.

GAO, D. et al. A survey of statistical software for analysing RNA-seq data. **Human genomics**, v. 5, n. 1, p. 56-60, 2010.

GARBER, M; GRABHERR, M.G; GUTTMAN, M; TRAPNELL, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods*. 2011; 8(6): 469–77.

GENTLEMAN, R. et al. Bioconductor: Open software development for computational biology and bioinformatics. **Genome Biology**, v. 5, p. R80, 2004. Disponível em: <<http://genomebiology.com/2004/5/10/R80>>.

GINGERAS TR. 2007. Origin of phenotypes: Genes and transcripts. *Genome Res* 17: 682–690.

GLENN T. C. Field guide to next-generation DNA sequencers. **Mol. Ecol. Resour.**, v. 11, n. 5, p. 759-69. 2011.

GONÇALVES, Leilane Oliveira. Análise de dados de sequenciamento de RNA voltados à comparação de expressão gênica visando a um melhor entendimento dos mecanismos de resistência aos antimoniais. 2017. 93 p. Dissertação (Mestrado em Ciências - Concentração Biologia Celular e Molecular)-Fundação Oswaldo Cruz. Instituto René Rachou. Programa de Pós - Graduação em Ciências da Saúde. Belo Horizonte. 2017.

GONG Y, REN J, LIU K, TANG L-M. Tumor suppressor role of miR-133a in gastric cancer by repressing IGF1R. *World Journal of Gastroenterology : WJG*. 2015;21(10):2949-2958. doi:10.3748/wjg.v21.i10.2949.

GONZÁLEZ, I. Statistical analysis of RNA-Seq data. Tutorial. 2014. Disponível em: <http://www.nathalievilla.org/doc/pdf/tutorial-rnaseq.pdf>. Acesso em: 19/12/2017.

GRADA, A.; WEINBRECHT, K. Next-Generation Sequencing: Methodology and Application. *Journal of Investigative Dermatology*, v. 133, n. 8, p. e11, ago. 2013.

HAMBERG M., BACKES C., FEHLMANN T., HART M., MEDER B., MEESE E., KELLER A. MiRTargetLink–miRNAs, genes and interaction networks. *Int. J. Mol. Sci*. 2016; 17:564.

HANAHAN D, WEINBERG RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646–674.

- HOLT, R. A.; JONES, S. J. M. The new paradigm of flow cell sequencing. *Genome research*, v. 18, n. 6, p. 839-46, 2008.
- HUANG, D. W.; SHERMAN, B. T.; LEMPICKI, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, v. 37, n. 1, p. 1–13, jan. 2009a.
- HUANG, D. W.; SHERMAN, B. T.; LEMPICKI, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, v. 4, n. 1, p.44–57, jan. 2009b.
- HUANG Y, LUO H, LI F, YANG Y, OU G, YE X, LI N. LINC00152 down-regulated miR-193a-3p to enhance MCL1 expression and promote gastric cancer cells proliferation. *Biosci Rep*. 2018;38:pii: BSR20171607
- INCA - INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Incidência de câncer no estado do Pará**, Rio de Janeiro, 2017.
- JOHNSON, D. S.; MORTAZAVI, A.; MYERS, R. M.; WOLD, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, v. 316, n. 5830, p.1497–502, 8 jun. 2007.
- JORDE, Lynn B.; CAREY, John C; BAMSHAD, Michael J. *Genética médica*. 4. ed. Rio de Janeiro: Elsevier, 2010. 350 p.
- KINJO, E. R. et al. (2013). A Possible New Mechanism for the Control of miRNA Expression in Neurons. *Experimental Neurobiology*, 248:546-58.
- KORPELAINEN, E. et al. *RNA-seq Data Analysis A Practical Approach*. [s.l.] CRC Press, 2015.
- KROL, J; LOEDIGE, I; FILIPOWICZ, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, 11 (2010), pp. 597–610.
- KUEHN H, LIBERZON A, REICH M, MESIROV JP (2008) Using GenePattern for gene expression analysis. *Curr Protoc Bioinformatics Chapter 7: Unit 7.12.H*.
- KURISSIO, J.K. Análise da expressão gênica celular e viral durante a infecção in vitro pelo herpesvírus canino 1. Botucatu, 2017,109 p. Defesa de Tese – Faculdade de Medicina Veterinária e Zootecnia, Campus de Botucatu, Universidade Estadual Paulista.
- LANDGRAF P, RUSU M, SHERIDAN R, SEWER A, LOVINO N, et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129: 1401–1414.P.
- LANDGREN, O.; RAJKUMAR, S. V.; PFEIFFER, R. M.; KYLE, R. A.; KATZMANN, J. A.; DISPENZIERI, A.; CAI, Q.; GOLDIN, L. R.; CAPORASO, N. E.; FRAUMENI, J. F.; et al. Obesity is associated with an increased risk of

monoclonal gammopathy of undetermined significance among black and white women. **Blood**, v. 116, n. 7, p.1056–9, 19 ago. 2010.

LAPPALAINEN T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501:506–11.

LEE, J.W; KIM, N; PARK, J.H, et al. Differential MicroRNA Expression Between Gastric Cancer Tissue and Non-cancerous Gastric Mucosa According to *Helicobacter pylori* Status. *Journal of Cancer Prevention*. 2017;22(1):33-39.

LEE, W.S., KIM, N., PARK, Y.R., OH, H.H., MYUNG, E., KIM, S.H. et al. (2015) Myeloid cell leukemia-1 promotes epithelial-mesenchymal transition of human gastric cancer cells. *Oncol. Rep.*34,1011– 1016, <https://doi.org/10.3892/or.2015.4040>

LENG N, DAWSON JA, THOMSON JA, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments, *Bioinformatics*, 2013, vol.29 (pg.1035-43).

LI, H.; TEO, Y. Y.; TAN, E. K. Patterns of linkage disequilibrium of LRRK2 across different races: implications for genetic association studies. **PloS One**, v. 8, n. 9, p. e75041, jan.2013.

LIN LIU, YINHU LI, SILIANG LI, et al., “Comparison of Next-Generation Sequencing Systems,” *Journal of Biomedicine and Biotechnology*, vol. 2012, Article ID 251364, 11 pages, 2012. doi:10.1155/2012/251364.

LIU J, LIU F, SHI Y, TAN H, ZHOU L. Identification of key miRNAs and genes associated with stomach adenocarcinoma from The Cancer Genome Atlas database. *FEBS Open Bio*. 2018;8(2):279-294. doi:10.1002/2211-5463.12365.

LIU, Y.; POPP, B.; SCHMIDT, B. CUSHAW3: sensitive and accurate base-space and colorspace short-read alignment with hybrid seeding. **PloS One**, v. 9, n. 1, p. e86869, jan. 2014.

LIU, Y., ZHANG, X., ZHANG, Y., HU, Z., YANG, D., WANG, C., ... CAI, Q. (2015). Identification of miRNomes in human stomach and gastric carcinoma reveals miR-133b/a-3p as therapeutic target for gastric cancer. *Cancer Letters*, 369(1), 58–66. doi:10.1016/j.canlet.2015.06.028

LODISH, H.; BERK, A.; KAISER, C. A.; KRIEGER, M.; BRETSCHER, T.; PLOEGH, H.; AMON, A. *Biologia celular e molecular*. 7. ed. Porto Alegre: Artmed, 2014.

LOVE MI, HUBER W, ANDERS S. 2014. Moderated estimation of foldchange and dispersion for RNA-Seq data with DESeq2. *bioRxiv* doi: 10.1101/002832.

MARGUERAT, S.; BÄHLER, J. RNA-seq: from technology to biology. *Cellular and molecular life sciences: CMLS*, v. 67, n. 4, p. 569-79, 2010.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. *NATURE REVIEWS*, v. 12, p. 671-682, 2011.

MARTINS, D. L. ALVES SOBRINHO, P. de A. SOUZA, J. E. S. de. *SoftwareR-peridot*. Natal, 2017. 5,7 MB. Conjunto de programas.

MASCARENHAS JUNIOR, Rui Wanderley. **Análise da Expressão Gênica Diferencial entre o Adenocarcinoma da junção Esôfago-Gástrica e o Adenocarcinoma Gástrico**. 72 f. Dissertação - Instituto de Ciências Biológicas, Universidade Federal do Pará, 2016.

MCELROY, J. P.; CREE, B. A. C.; CAILLIER, S. J.; GREGERSEN, P. K.; HERBERT, J.; KHAN, O. A.; FREUDENBERG, J.; LEE, A.; BRIDGES, S. L.; HAUSER, S. L.; et al. Refining the association of MHC with multiple sclerosis in African Americans. *Human Molecular Genetics*, v. 19, n. 15, p. 3080–8, 1 ago. 2010.

MCGETTIGAN, P. A. Transcriptomics in the RNA-seq era. *Current Opinion in Chemical Biology, Omics*. v. 17, n. 1, p. 4–11, fev. 2013.

METZKER, M. L. Sequencing technologies-the next generation. *Nature reviews. Genetics*, v. 11, n. 1, p. 31-46, 2010. Nature Publishing Group.

MOREIRA, F. C. Análise *in Silico* da Expressão Diferencial de microRNAs no Câncer Gástrico. 2014. 133 f. Tese (Doutorado em Genética e Biologia Molecular) – Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, 2014.

MOREIRA FC, Assumpção M, Hamoy IG, Darnet S, Burbano R, et al. (2014) MiRNA Expression Profile for the Human Gastric Antrum Region Using Ultra-Deep Sequencing. *PLoS ONE* 9(3): e92300. doi:10.1371/journal.pone.0092300.

MORTAZAVI, A.; WILLIAMS, B. A.; MCCUE, K.; SCHAEFFER, L.; WOLD, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, v. 5, n. 7, p. 621-8, 2008.

MÓVIO, M. MOREIRA, C. NISHIO, L. WALTER, L.T. KIHARA, A.H. **OS EFEITOS MALÉFICOS DA POLUIÇÃO PODEM SER MEDIADOS POR microRNAs**. Laboratório de Neurogenética / Núcleo de Cognição e Sistemas Complexos / Centro de Matemática, Computação e Cognição / Universidade Federal do ABC. Edição Vol. 2, N. 17, 8 de Setembro de 2015. Disponível em: <http://www.nanocell.org.br/os-efeitos-maleficos-da-poluicao-podem-ser-mediados-por-micrnas/>. Acesso em: 15/01/18

MYERS, R. A.; SCOTT, N. M.; GAUDERMAN, W. J.; QIU, W.; MATHIAS, R. A.; ROMIEU, I.; LEVIN, A. M.; PINO-YANES, M.; GRAVES, P. E.; VILLARREAL, A. B.; et al. Genome-wide interaction studies reveal sex-specific asthma risk alleles. *Human Molecular Genetics*, 13 maio 2014.

NAGINI, S. Carcinoma of the stomach: A review of epidemiology, pathogenesis, molecular genetics and chemoprevention. *World Journal Of Gastrointestinal Oncology*, [s.l.], v. 4, n. 7, p.156-169, 2012.

National Cancer Institute. NCI Dictionary of Cancer Terms, 2018. Disponível em: <https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=45693>. Acesso em: 15/01/18.

NEAPOLITAN R, HORVATH CM, JIANG X. Pan-cancer analysis of TCGA data reveals notable signaling pathways. *BMC Cancer*. 2015;15:516. doi: 10.1186/s12885-015-1484-6.

OSATO N. Characteristics of functional enrichment and gene expression level of human putative transcriptional target genes. *BMC Genomics*. 2018;19(Suppl 1):957. doi:10.1186/s12864-017-4339-5.

OSHLACK, A.; ROBINSON, M. D.; YOUNG, M. D. From RNA-seq reads to differential expression results. *Genome biology*, v. 11, n. 12, p. 220, 2010.

PALMA, Flávio Romero. Caracterização de Genes Análogos de Resistência (RGAS) em *Elaeis guineensis* e *Elaeis oleifera* contrastantes em resistência a *Fusarium oxysporum* f.sp. *elaedis*. 2011. 115 f., il. Dissertação (Mestrado em Fitopatologia)—Universidade de Brasília, Brasília, 2011.

PEREIRA, A. ONCOGENÔMICA: microRNAs NO CÂNCER GÁSTRICO. Belém, 2018, 127 p. Defesa de Tese – Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Pará.

PRICE, D.; FLETCHER, M.; VAN DER MOLEN, T. Asthma control and management in 8,000 European patients: the REcognise Asthma and LInk to Symptoms and Experience (REALISE) survey. **NPJ Primary Care Respiratory Medicine**, v. 24, p. 14009, jan. 2014.

QIU T, ZHOU X, WANG J, DU Y, XU J, HUANG Z, ZHU W, SHU Y, LIU P. MiR-145, miR-133a and miR-133b inhibit proliferation, migration, invasion and cell cycle progression via targeting transcription factor Sp1 in gastric cancer. *FEBS Lett*. 2014;588:1168–1177. doi: 10.1016/j.febslet.2014.02.054.

RAMOS, T. de A. R. DESENVOLVIMENTO E USO DO CORAZON: FERRAMENTA PARA NORMALIZAÇÃO E AGRUPAMENTO DE DADOS DE EXPRESSÃO GÊNICA. 2018. 158 f. Dissertação (Mestrado em Bioinformática) – Instituto Metr pole Digital, Universidade Federal do Rio Grande do Norte, Natal, 2018.

R CORE TEAM. R: A language and environment for statistical computing. 2012. Vienna, Austria: R Foundation for Statistical Computing. Disponível em: <http://www.r-project.org/>.

RIBEIRO DOS SANTOS, Andr  M. Diversidade de Indels em S cios de Liga o de Fatores de Transcri o Humanos. 2014. 123 f., Disserta o (Mestrado em



Genética e Biologia Molecular) - Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, 2014.

RIBEIRO-DOS-SANTOS, Ândrea et al. The Deep Sequencing of MicroRNA in Cancer: Expression Profiling and its Applications. **Regulatory RNAs**. Springer Berlin Heidelberg, 523-546. 2012.

RIBEIRO-DOS-SANTOS Â, KHAYAT AS, SILVA A, ALENCAR DO, LOBATO J, LUZ L, et al. (2010) Ultra-Deep Sequencing Reveals the microRNA Expression Pattern of the Human Stomach. *PLoS ONE* 5(10): e13205. <https://doi.org/10.1371/journal.pone.0013205>.

RICARTE FILHO, Júlio; KIMURA, Edna Teruko. MicroRNAs: novel class of gene regulators involved in endocrine function and cancer. **Arquivos Brasileiros de Endocrinologia & Metabologia**, v. 50, n. 6, p. 1102-1107, 2006.

ROBBINS, D.E; GRUNEBERG, A; DEUS, H.F; TANIK, M.M; ALMEIDA, J.S. A self-updating road map of The Cancer Genome Atlas. *Bioinformatics*. 2013;29(10):1333–40. pmid: 23595662; PubMed Central PMCID: PMC3654710.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, 26:139–140. 2010.

ROBINSON, M. D.; SMYTH, G. K. Moderated statistical tests for assessing differences in tag abundance. **Bioinformatics (Oxford, England)**, v. 23, n. 21, p. 2881-7, 2007.

ROBINSON, M. D.; SMYTH, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. **Biostatistics**, 9, 2, pp. 321–332, 2008.

RUIVO, H.M. **Análise Integrada de Dados Ambientais Utilizando Técnicas de Classificação e Agrupamento de Microarranjos de DNA**. 98 f. Dissertação-Instituto Nacional de Pesquisas Espaciais-INPE, São José dos Campos, 2007.

SATAKE, W.; NAKABAYASHI, Y.; MIZUTA, I.; HIROTA, Y.; ITO, C.; KUBO, M.; KAWAGUCHI, T.; TSUNODA, T.; WATANABE, M.; TAKEDA, A.; et al. Genomewide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. **Nature Genetics**, v. 41, n. 12, p. 1303–7, dez. 2009.

SHEN, L; ZHAO, L; TANG, J. *et al.* Key Genes in Stomach Adenocarcinoma Identified via Network Analysis of RNA-Seq Data. *Pathol. Oncol. Res*, 2017. SHENDURE J., JI H. Next generations DNA sequencing. *Nature Biotechnology*, v. 26, n.10, p. 1135–1145. 2008.

SHILIN ZHAO, YAN GUO, QUANHU SHENG, and YU SHYR, “Advanced Heat Map and Clustering Analysis Using Heatmap3,” *BioMed Research International*, vol. 2014, Article ID 986048, 6 pages, 2014. doi:10.1155/2014/986048.

SOBREIRO, M. B. Expressão gênica diferencial de quatro espécies da Aliança Tabebuia em resposta ao déficit hídrico. 2017. 55 f. Dissertação (Mestrado em Genética e Biologia Molecular) - Universidade Federal de Goiás, Goiânia, 2017.

SOUZA, Pâmela Tamiris Caldas Serra de. Comparação de metodologias para identificação de genes diferencialmente expressos em experimentos de RNA-Seq de suínos. 2015. 37 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa. 2015.

STEFFEN, J. S. et al. LGR4 and LGR6 are differentially expressed and of putative tumor biological significance in gastric carcinoma. *Virchows Archiv*, [s.l.], v. 461, n. 4, p.355-365, 2 ago. 2012.

SZKLARCZYK D, MORRIS JH, COOK H, KUHN M, WYDER S, SIMONOVIC M, SANTOS A, DONCHEVA NT, ROTH A, BORK P, JENSEN LJ, VON MERING C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research*. 2017;45:D362–D368. doi: 10.1093/nar/gkw937.

TRAPNELL, C.; PACHTER, L.; SALZBERG, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, v. 25, n. 9, p. 1105–11, 2009.

TULL, E. S. Assessment of the ability of the triglyceride to high density lipoprotein cholesterol ratio to discriminate insulin resistance among Caribbean-born black persons with and without Hispanic ethnicity. **The West Indian Medical Journal**, v. 62, n. 2, p. 109–13, fev. 2013.

VALIENTE, A. E. F. Avaliação da imunoexpressão de LGR5 e LGR6 em células-tronco no câncer gástrico primário, metástases linfonodais e mucosa gástrica histologicamente normal. 2017. 78 f. Dissertação (Mestrado em Patologia) - Faculdade de Medicina, Universidade Federal do Ceará, Fortaleza, 2017.

VAN VERK, M. C. et al. RNA-Seq: revelation of the messengers. *Trends in Plant Science*, v. 18, n. 4, p. 175–179, abr. 2013.

VARDENY, O.; GUPTA, D. K.; CLAGGETT, B.; BURKE, S.; SHAH, A.; LOEHR, L.; RASMUSSEN-TORVIK, L.; SELVIN, E.; CHANG, P. P.; AGUILAR, D.; et al. Insulin resistance and incident heart failure the ARIC study (Atherosclerosis Risk in Communities). **JACC. Heart Failure**, v. 1, n. 6, p. 531–6, dez. 2013.

VÊNCIO, Ricardo ZN et al. Bayesian model accounting for within-class biological variability in Serial Analysis of Gene Expression (SAGE). **BMC bioinformatics**, v. 5, n. 1, p. 119, 2004.

VIDAL AF. et al. **The comprehensive expression analysis of circular RNAs in gastric cancer and its association with field cancerization.** *Sci Rep.* 2017 Nov 6;7(1):14551. doi: 10.1038/s41598-017-15061-w.

WANG, F. et al. (2012). Modulation of MicroRNA Expression by Volatile Organic Compounds in Mouse Lung. *Wiley Periodicals*, 29(6):679-689.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, v. 10, n. 1, p. 57-63, 2009.

WHO. **WHO | Global Health Observatory (GHO)**. Disponível em: <<http://www.who.int/gho/en/>>. Acesso em: 01 fev. 2018.

WILHELM, B. T.; LANDRY, J.-R. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. **Methods**, v. 48, p. 249-257, 2009.

WU, W.K; LEE, C.W; CHO, C.H; FAN, D; WU, K; YU, J et al. MicroRNA dysregulation in gastric cancer: a new player enters the game. *Oncogene* 2010; 29: 5761–5771.

YENDREK, C. R.; AINSWORTH, E. A; THIMMAPURAM, J. The bench scientist's guide to statistical analysis of RNA-Seq data. **BMC Research Notes**, v. 5, n. 1, p.506, 2012. *BMC Research Notes*.

YU D, HUBER W, VITEK O. 2013. Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size. *Bioinformatics* 29: 1275–1282.

ZHAN Y-L, WANG R-C, CHENG K, RING BZ, SU L. Roles of Rap1 signaling in tumor cell migration and invasion. *Cancer Biology & Medicine*. 2017;14:90–99. doi: 10.20892/j.issn.2095-3941.2016.0086.