

Universidade de São Paulo
Programa Interunidades em Bioinformática

Vanessa Galdeno Freitas

Perfil de expressão dos circRNAs no câncer Colorretal

São Paulo
2021
Universidade de São Paulo
Programa Interunidades em Bioinformática

Vanessa Galdeno Freitas

Perfil de expressão dos circRNAs no câncer Colorretal

Versão original

Dissertação de Mestrado
apresentada ao Programa Interunidades de
Pós-Graduação em Bionformática da Universidade
de São Paulo para obtenção do título de Mestre em
Ciências.

Área de concentração: Bioinformática

Orientador: Prof. Dr. Pedro Alexandre Favoretto
Galante

São Paulo

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

F672 Freitas, Vanessa Galdeno
Perfil de expressão dos circRNAs no câncer colorretal / Vanessa Galdeno
Freitas, orientador Pedro Alexandre Favoretto Galante. -- São Paulo : 2021.
106 p.

Dissertação (Mestrado) -- Universidade de São Paulo
Orientador: Prof. Dr. Pedro Alexandre Favoretto Galante
Programa Interunidades de Pós-Graduação em Bioinformática
Área de concentração: Bioinformática
Versão Original

1. RNAs Circulares. 2. circRNAs. 3. Expressão Gênica. 4. Bioinformática.
5. Câncer Colorretal. 6. Progressão Tumoral. 7. Metástase. 8. Epigenética. I.
Galante, Pedro Alexandre Favoretto, orientador. II. Universidade de São Paulo.
III. Título.

CDD - 572.8

Agradecimentos

“Descobri que há uma harmonia maravilhosa nas verdades complementares da fé e da ciência. O Deus da Bíblia é também o Deus do genoma. Deus pode ser encontrado na catedral e no laboratório. Investigando a criação incrível e majestosa de Deus, a ciência pode na verdade ser uma forma de louvor” - Francis Collins (1960-), diretor do Instituto Nacional de Saúde dos EUA e diretor do Projeto Genoma Humano.

Primeiramente, gostaria de agradecer a Deus.

Agradeço ao meu orientador Pedro, que sempre apoiou minhas decisões e encorajou as minhas idéias, espero poder continuar essa jornada em colaboração com o melhor pesquisador.

Agradeço a minha família, a minha mãe Terezinha e a minha irmã Júlia, que sempre me mostraram o lado feliz e sincero da vida.

Agradeço ao Rubens e família, Marylaine, Walmir e João Pedro.

Agradeço aos nossos *petz*, Astolfo e Pikachu, pois os animais tem esse poder de nos fazer sentir vivos e ultrapassar a barreira das espécies por amor e amizade.

Agradeço a todos os colegas e amigos de laboratório, embora o mundo e o trabalho nos fizeram nos distanciar socialmente (aliás, sobrevivemos a uma pandemia em 2020, temos que comemorar), a positividade e apoio sempre serão lembrados.

Resumo

Freitas, V.G. (2021) . Perfil de expressão dos circRNAs no câncer Colorretal. (Dissertação de Mestrado). Programa de Pós-Graduação Interunidades em Bioinformática, Universidade de São Paulo, São Paulo.

Introdução: Os RNAs circulares (circRNAs) são uma nova classe de RNA que forma laços (*loops*) contínuos fechados covalentemente. Como os RNAs longos e não codificadores, as funções efetivas dos circRNAs dependem principalmente das características de suas sequências e estruturas e há um crescente corpo de papéis funcionais atribuídos aos circRNAs. Por exemplo, os circRNAs atuam como esponjas de microRNAs que modulam indiretamente a expressão gênica de genes codificadores de proteínas. Além disso, a desregulação da expressão de circRNAs têm sido associada a várias patologias humanas, como o câncer. Em termos de expressão, estudos relataram centenas de circRNAs como mais abundantes do que seus correspondentes mRNAs lineares, não apenas nos tecidos, mas também no sangue. O câncer colorretal (CRC) é o terceiro câncer mais diagnosticado e a quarta principal causa de mortes relacionadas ao câncer no mundo. Estudos em linhas de células CRC e tecidos CRC mostram uma redução geral na abundância de circRNA em comparação com tecidos saudáveis, permitindo que as células CRC sejam inesperadas e apresentem proliferação descontrolada, por exemplo. **Metodologia:** Caracterizamos e estudamos o perfil de expressão dos circRNAs em linhagens de células CRC com o objetivo de entender melhor o papel desses RNAs na tumorigênese e progressão do câncer. Primeiro, os dados de sequenciamento de RNA (*RNA-seq*) de duas linhas celulares comerciais do mesmo paciente primário e metastático de CRC (SW480 e SW620,

respectivamente) foram obtidos a partir de uma plataforma Illumina NextSeq. Em seguida, foram utilizados dois métodos de preparação da biblioteca de sequenciamento de RNA: i) o protocolo padrão sugerido por Illumina; ii) e um protocolo interno para melhorar a detecção de circRNAs. Finalmente, todos os dados de *RNA-seq* foram alinhados ao genoma de referência (GRCh38) e usados como entrada para identificar circRNAs por meio do algoritmo CircExplorer e outros pipelines de computação internos. As metodologias RSEM e Kallisto também foram usadas para gerar o perfil de expressão gênica de todas as linhagens celulares e preparações de bibliotecas. O pacote do R, edgeR, foi utilizado para normalização e seleção dos genes diferencialmente expressos. Resultados: Nossos resultados mostraram que nosso protocolo interno de preparação da biblioteca de sequenciamento de RNA detectou 70% mais circRNAs do que uma preparação padrão comumente usada. Em seguida, avaliando o número de circRNAs expressos nas linhagens celulares primária (SW480) e metastática (SW620), encontramos 3.991 circRNAs expressos diferencialmente (únicos, $FDR < 0,05$; $\log_2 \text{fold change} > 2$ ou < -2), 1.936 e 2.055 circRNA super e sub expressos, respectivamente, na linhagem de tumor primário versus a metastática. Ao avaliar as vias celulares dos genes que geram circRNAs, identificamos vários candidatos envolvidos na tumorigênese e na progressão do câncer por vias epigenéticas, mas também, diversos circRNAs que parecem estar agindo em mecanismos e vias de supressão tumoral, em ambas as linhagens. Portanto, em resumo, este trabalho traz contribuições importantes para entendermos melhor o papel dos circRNAs no câncer colorretal e também suas contribuições na progressão tumoral de tumor primário à metástase, e como eles podem ser importantes como biomarcadores para essa patologia.

Palavras-chave: RNAs circulares, expressão gênica, bioinformática, câncer colorretal, progressão tumoral, metástase, epigenética.

Abstract

Freitas, V.G. (2021). CircRNAs Expression Profile in Colorectal Cancer. (Masters dissertation). Interunit Graduate Program in Bioinformatics, University of São Paulo, São Paulo.

Introduction: *Circular RNAs (circRNAs) are a new class of RNA that form continuous covalently closed loops. Like long, non-coding RNAs, the actual functions of circRNAs depend mainly on the characteristics of their sequences and structures, and there is a growing body of functional roles assigned to circRNAs. For example, circRNAs act as microRNA sponges that indirectly modulate gene expression of protein-coding genes. Furthermore, dysregulation of circRNA expression has been associated with several human pathologies, such as cancer. In terms of expression, studies have reported hundreds of circRNAs as more abundant than their corresponding linear mRNAs, not just in tissues but also in blood. Colorectal cancer (CRC) is the third most diagnosed cancer and the fourth leading cause of cancer-related deaths worldwide. Studies in CRC cell lines and CRC tissues show an overall reduction in circRNA abundance compared to healthy tissues, allowing CRC cells to be unexpected and proliferate uncontrolled, for example.*

Methodology: *We characterized and studied the expression profile of circRNAs in CRC cell lines in order to better understand the role of these RNAs in tumorigenesis and cancer progression. First, RNA sequencing data (RNA-seq) from two commercial cell lines from the same primary and metastatic CRC patient (SW480 and SW620, respectively) were obtained from an Illumina NextSeq platform. Then, two methods of preparing the RNA sequencing library were used: i) the standard protocol suggested by Illumina; ii) and an internal protocol to improve*

detection of circRNAs. Finally, all RNA-seq data were aligned to the reference genome (GRCh38) and used as input to identify circRNAs through the CircExplorer algorithm and other internal computation pipelines. RSEM and Kallisto methodologies were also used to generate the gene expression profile of all cell lines and library preparations. The R package, edgeR, was used for normalization and selection of differentially expressed genes.

Results: Our results showed that our in-house RNA sequencing library preparation protocol detected 70% more circRNAs than a commonly used standard preparation. Then, evaluating the number of circRNAs expressed in the primary (SW480) and metastatic (SW620) cell lines, we found 3,991 differentially expressed circRNAs (single, FDR <0.05; log₂ fold change > 2 or <-2), 1,936 and 2,055 super and under expressed circRNA, respectively, in the primary versus metastatic tumor lineage. By evaluating the cellular pathways of genes that generate circRNAs, we identified several candidates involved in tumorigenesis and cancer progression through epigenetic pathways, but also several circRNAs that seem to be acting on tumor suppression mechanisms and pathways, in both strains. Therefore, in summary, this work makes important contributions to better understand the role of circRNAs in colorectal cancer and also their contributions in tumor progression from primary tumor to metastasis, and how they can be important as biomarkers for this pathology.

Keywords: circular RNAs, gene expression, bioinformatics, colorectal cancer, tumor progression, metastasis, epigenetics.

Lista de Abreviaturas

circRNA	<i>Circular RNA</i>
CRC	<i>Colorectal cancer</i>
SW480	<i>Commercial lineage of CRC model cells</i>
SW620	<i>Commercial lineage of CRC model cells</i>
RNA	<i>Ribonucleic Acid</i>
DNA	<i>Deoxyribonucleic Acid</i>
NGS	<i>Next Generation Sequencing</i>
RNA-Seq	<i>RNA Sequencing</i>
miRNA	<i>Micro RNA</i>
mRNA	<i>Messenger RNA</i>
ncRNA	<i>Non-coding RNA</i>
EV	<i>Extracellular Vesicle</i>
ciRNA	<i>Intronic long non-coding circular RNA</i>
ElciRNA	<i>Exon–intron circular RNA</i>
GRCh38 or hg38	<i>Genome Reference Consortium Human Build 38</i>
hg19	<i>Genome Reference Consortium Human Build 37</i>
PCR	<i>Polymerase chain reaction</i>
qRT-PCR	<i>Real-time polymerase chain reaction</i>
siRNA	<i>Small interfering RNA</i>
lncRNAs	<i>long non coding RNAs</i>
RBP	<i>RNA-binding protein</i>
ElciRNAs	<i>retained intron circRNAs</i>
exo-circRNAs	<i>exosomal circRNAs</i>
QKI	<i>Quaking protein</i>
dsDNAs	<i>double stranded DNAs</i>
f-circRNA	<i>fusion circRNAs</i>
OAS	<i>2'-5'-oligoadenilato sintetase</i>
ceRNA	<i>competing endogenous RNA</i>
RT-qPCR	<i>Real time quantitative PCR</i>
ATCC	<i>American Type Culture Collection</i>
SAV	<i>Sequencing Analysis Viewer</i>
RTA	<i>Real-Time Analysis</i>
DNMT1	<i>DNA Methyltransferase 1</i>
TGM2	<i>Transglutaminase 2</i>
TSPAN4	<i>Transmembrane 4 Superfamily</i>

EPAS1 *Endothelial Pas Domain Protein 1*
UBE3C *Ubiquitin Protein Ligase E3c*
ST18 *Suppression Of Tumorigenicity 18*
WNT5A *Wingless-Type Mmtv Integration Site Family, Member 5a*
GLI3 *Gli-Kruppel Family Member 3*
LRP5 *Low Density Lipoprotein Receptor-Related Protein 5*
SMURF2 *Smad-Specific E3 Ubiquitin Protein Ligase 2*
RUNX1 *Runt-Related Transcription Factor 1*
CCAR1 *Cell Division Cycle And Apoptosis Regulator 1*
FOXP1 *Forkhead Box P1*
KMT2C *Lysine-Specific Methyltransferase 2c*
ATRX *ATRX Chromatin Remodeler*
EMT *Epitelial-Mesenquimal*
PPI *Protein-Protein-Interaction*
GICs *Glioma Initiator Cells*
CHC *Hepatocellular Carcinoma*
TNM *Classification of Malignant Tumours*
LUAD *Lung adenocarcinoma*
CECO *Oral Squamous Cell Carcinoma*
H3K4 *Histone 3 Lysine 4*
HCC *Hepatocellular Carcinoma*

Lista de Figuras

Figura 1- Formação, Função e Degradação dos circRNAs.	6
Figura 2: Fluxo esquemático do <i>pipeline</i> do CIRCexplorer.	26
Figura 3: Gráficos das métricas de qualidade do sequenciamento gerados pelo SAV.	37
Figura 4: <i>Venn diagram</i> da comparação dos diferentes preparos de biblioteca no R.	43
Figura 5: <i>Heatmap</i> da expressão diferencial dos 1612 circRNAs em comum nos sequenciamentos apresentados no R.	44
Figura 6: <i>Volcano plot</i> da comparação da expressão dos genes entre as linhagens celulares no R.	45
Figura 7: Rede PPI dos genes <i>down-regulated</i> no STRING.	49
Figura 8: Rede PPI dos genes <i>down-regulated</i> no STRING.	52
Figura 9: Rede PPI dos genes <i>up-regulated</i> no STRING.	53
Figura 10: Rede PPI dos genes <i>up-regulated</i> no STRING.	56
Figura 11: Vias KEGG <i>up-regulated</i> no WebGestalt.	57
Figura 12: Via KEGG <i>up-regulated</i> “ <i>Proteólise mediada por Ubiquitinação</i> ” no ShinyGO com pacote Pathview (com legenda dos ícones da imagem gerada acima, usada para todas as imagens a seguir que utilizaram o mesmo <i>software</i> e pacote).	59
Figura 13: Via KEGG <i>up-regulated</i> “ <i>Via de sinalização do mTOR</i> ” no ShinyGO com pacote Pathview.	60

Figura 14: Via KEGG <i>up-regulated</i> "Via da resistência ao inibidor de tirosina quinase de EGFR" no ShinyGO com pacote Pathview.	61
Figura 15: Via KEGG <i>up-regulated</i> "Biogênese do ribossomo nos eucariotos" no ShinyGO com pacote Pathview.	63
Figura 16: Via KEGG <i>up-regulated</i> "Degradação do RNA" no ShinyGO com pacote Pathview.	64
Figura 17: Via KEGG <i>up-regulated</i> "Adesão focal" no ShinyGO com pacote Pathview.	65
Figura 18: Via KEGG <i>up-regulated</i> "Transporte Núcleo-citoplasma" no ShinyGO com pacote Pathview.	66
Figura 19: Vias KEGG <i>down-regulated</i> no WebGestalt.	67
Figura 20: Via KEGG <i>down-regulated</i> "Ciclo celular" no ShinyGO com pacote Pathview.	68
Figura 21: Via KEGG <i>down-regulated</i> "Vias do câncer" no ShinyGO com pacote Pathview.	69
Figura 22: Via KEGG <i>down-regulated</i> "Proteólise mediada por ubiquitinação" no ShinyGO com pacote Pathview.	70
Figura 23: Via KEGG <i>down-regulated</i> "Via de sinalização WNT" no ShinyGO com pacote Pathview.	71
Figura 24: Via KEGG <i>down-regulated</i> "Via de sinalização Hedgehog" no ShinyGO com pacote Pathview.	72
Figura 25: Via KEGG <i>down-regulated</i> "Via da resistência ao inibidor de tirosina quinase de EGFR" no ShinyGO com pacote Pathview.	73

Figura 26: Via KEGG <i>down-regulated</i> "Câncer de pulmão de pequenas células" no ShinyGO com pacote Pathview.	74
Figura 27: Via KEGG "Proteólise mediada por ubiquitinação" no ShinyGO com pacote Pathview. Comparação entre a mesma via com os valores de <i>fold change</i> positivos e negativos.	75
Figura 28: Termos GO <i>up-regulated</i> para "Processos biológicos" no ShinyGO.	76
Figura 29: Termos GO <i>up-regulated</i> para "Componentes celulares" no ShinyGO.	77
Figura 30: Termos GO <i>up-regulated</i> para "Funções Moleculares" no ShinyGO.	77
Figura 31: Termos GO <i>up-regulated</i> para "Processos biológicos" no WebGestalt.	78
Figura 32: Termos GO <i>up-regulated</i> para "Componentes celulares" no WebGestalt.	79
Figura 33: Termos GO <i>up-regulated</i> para "Funções Moleculares" no WebGestalt.	80
Figura 34: Termos GO <i>up-regulated</i> no GOrilla com pacote Revigo no R (para "Processos biológicos, Componentes celulares e Função Molecular").	81
Figura 35: Termos GO <i>down-regulated</i> para "Processos biológicos" no ShinyGO.	82
Figura 36: Termos GO <i>down-regulated</i> para "Componentes celulares" no ShinyGO.	83
Figura 37: Termos GO <i>down-regulated</i> para "Funções Moleculares" no ShinyGO.	83
Figura 38: Termos GO <i>down-regulated</i> para "Processos biológicos" no WebGestalt.	84

Figura 39: Termos GO <i>down-regulated</i> para "Componentes celulares" no WebGestalt.	84
Figura 40: Termos GO <i>down-regulated</i> para "Funções Moleculares" no WebGestalt.	85
Figura 41: Termos GO <i>down-regulated</i> no GOrilla com pacote Revigo no R.	85
Figura 42: Relação dos 50 primeiros genes da lista <i>up-regulated</i> ordenados do maior para o menor de acordo com o valor de <i>fold change</i> e expressão nas linhagens celulares comerciais para câncer pelo GSEA.	87
Figura 43: Relação dos 50 primeiros genes da lista <i>down-regulated</i> ordenados do maior para o menor de acordo com o valor de <i>fold change</i> e expressão nas linhagens celulares comerciais para câncer pelo GSEA.	88

Lista de Tabelas

Tabela 1: Modelo de arquivo referência do CIRCexplorer.	27
Tabela 2: Modelo do arquivo de <i>output</i> do resultado final do CIRCexplorer.	28
Tabela 3: Valores das métricas de qualidade da corrida obtidos através da análise pelo SAV.	36
Tabela 4: Métricas do alinhamento dos <i>reads</i> das bibliotecas padrão (SW480/SW620-ribo) e preparação feita pelo nosso grupo (SW480/SW620-circ).	38
Tabela 5: Comparação entre as linhagens e entre os preparos de biblioteca para detecção dos circRNAs.	39
Tabela 6: Comparação da contagem total dos transcritos (Kallisto) e genes (RSEM) identificados nas bibliotecas e linhagens celulares.	40
Tabela 7: Total de genes diferencialmente expressos edgeR/Kallisto.	41
Tabela 8: Total de genes diferencialmente expressos edgeR/RSEM.	41
Tabela 9: Total de genes diferencialmente expressos para circRNAs edgeR/Kallisto.	42

Anexos

Anexo 1: Relatório FASTQC	107
Anexo 2: Relatório RNASEQC	119

Sumário

Introdução	1
1.1 Biogênese dos circRNAs	3
1.2 Metabolismo dos circRNAs	7
1.3 Regulação da expressão gênica pelos circRNAs	9
1.4 circRNAs e o câncer	11
1.5 microRNAs e circRNAs no câncer	13
1.6 circRNAs e o câncer colorretal	15
1.7 Linhagens	18
Objetivos	20
Objetivo geral	20
Objetivos específicos	20
Materiais e Métodos	21
3.1- Cultura de células e extração de RNA	21
3.1.1- Lise das amostras e separação de fases	21
3.1.2- Precipitação, lavagem e ressuspensão do RNA	22
3.2- Rna Sequencing	23
3.3- Qualidade da corrida e métricas	23
3.4- Alinhamento	25
3.5- Detecção dos circRNAs	25
3.6- Counts	29
3.6.1- Montagem do arquivo FASTA de circRNAs	30
3.6.2- Kallisto	30
3.6.3- RSEM	31
3.7- Expressão diferencial	31
3.8- Gráficos	33
3.9- Vias e Redes de interação	33
3.10- Bancos de dados	34
Resultados e Discussão	36
Conclusão	103
Referências	122

Introdução

Os RNAs circulares (circRNAs) são uma nova classe de RNAs não codificantes de fita simples que formam laços (*loops*) contínuos fechados de forma covalente (Wang et al., 2017). Eles variam em comprimento de algumas centenas a milhares de nucleotídeos (Guo et al., 2014). O primeiro RNA circular foi encontrado em RNA viral em 1976 (Kolakofsky, 1976). Quando identificados em plantas, eles foram considerados patogênicos devido à sua similaridade estrutural com vírus (Ence, 1979; Sanger et al., 1976). Nos mamíferos, os circRNAs foram relatados pela primeira vez há mais de 20 anos como RNAs longos não codificadores gerados por erros ocasionais no mecanismo de *splicing* (Cocquerelle, Daubersies, Majerus, & Kerckaert, 1992; Hsu MT, 1979; Ruppert & Kinzler, 1991). Assim permaneceram até pouco tempo quando houve o advento do sequenciamento em larga escala de segunda geração (também definido, em inglês, como *Next Generation Sequencing, NGS*), mais especificamente o *RNA-Sequencing (RNA-seq)*, que desencadeou a descoberta de mais de 30.000 circRNAs até o presente momento. Esses circRNAs são descritos no envolvimento das mais diversas funções em múltiplos processos biológicos e organismos, sugerindo assim que essas moléculas de RNA são evolutivamente conservadas e, portanto, não são artefatos incorretos do mecanismo de *splicing* (Ashwal et al., 2014; Wang et al., 2014).

Como RNAs não codificantes longos (lncRNAs), as funções dos circRNAs podem depender das características de suas sequências e estruturas, amplamente envolvidos em processos fisiológicos e patológicos. A desregulação dos circRNAs podem influenciar em uma ampla variedade de doenças, especialmente o câncer, doenças cardiovasculares e doenças neurológicas (Dong et al., 2017; Nair et al., 2016; Szabo, 2016, Zhang et al., 2018). Por exemplo, em um estudo seminal, Hansen e colaboradores (Hansen et al., 2011) descrevem que o miR- 671 direcionou a clivagem de um transcrito circular antisenso do *locus* da Proteína 1 (CDR1) relacionada à degeneração cerebral de uma maneira dependente de AGO2. Dessa forma, este foi o primeiro relato na literatura de que transcritos circulares antissenso não codificantes podem atuar como alvos diretos de microRNAs (miRNA), sugerindo-se um novo mecanismo regulador de genes que envolve uma correlação positiva entre os níveis de microRNA e os RNAs circulares antissenso (Hansen et al., 2011). Além de atuarem como ‘esponjas’ de miRNA (Hansen et al., 2013), os circRNAs são descritos como reguladores de RNA ligantes de proteína (*RNA-binding protein- RBP*) na modulação da expressão das proteínas e codificação de genes (Li et al., 2016). Além disso, há relatos de circRNAs envolvidos na transcrição de peptídeos (Granados-riveron & Aquino-jarquín, 2016), circRNAs retrotranscritos e circRNAs funcionando como RNA competitivo (Dong et al., 2016).

Os circRNAs são altamente estáveis e geralmente exibem expressão específica de tecido ou tipo de célula, desta maneira, eles podem servir como potenciais alvos terapêuticos, principalmente para tratamentos que se destinam ao câncer (Enuka et al., 2016; Ji et al., 2019; Salzman et al., 2013) já que podem

sinalizar progressão tumoral e estar envolvidos em vias de sinalização, tais como as vias Wnt / β -catenina, interagirem com miRNAs (Yang et al., 2017) e serem facilmente detectados no plasma sanguíneo de pacientes (Wu et al., 2019).

1.1 Biogênese dos circRNAs

Os circRNAs são transcritos (expressos) em tecidos ou estágios de desenvolvimento específicos (Memczak et al., 2013). CircRNAs se distinguem dos mRNAs (lineares) na medida que lhe faltam a cauda poli(A) e o CAP na extremidade 5' (Jeck et al., 2013; Zhang et al., 2014). Assim, são mais estáveis que os mRNAs lineares associados *in vivo* quanto a sua resistência à atividade da RNase (Jeck et al., 2013; Memczak et al., 2013). Os circRNAs são mais de 14% dos genes transcritos ativamente em fibroblastos humanos e, em alguns casos, a abundância de moléculas circulares excede em mais de 10 vezes a do RNA mensageiro (mRNA) linear associado (Jeck et al., 2013).

Em termos do compartimento celular que eles atuam, os circRNAs são predominantemente citoplasmáticos (ciRNAs), mas alguns parecem atuar também no núcleo, Figura 1A (Memczak et al., 2013; Z. Wang et al., 2015).

Além dos RNAs circulares estarem localizados no núcleo e citoplasma, estes recentemente foram descritos como conteúdo presente nos exossomos (Figura 1C). Os exossomos foram uma descoberta importante para a biologia celular que contribuiu para o entendimento de diversos processos. Estes são uma

classe de vesículas extracelulares secretadas por quase todos os tipos de células que circulam nos fluidos corporais, como sangue, urina e saliva (Vakhshiteh et al., 2019). Os conteúdos exossômicos podem variar, e são compostos por diferentes proteínas, lipídios e ácidos nucleicos, RNAs longos não codificantes e circRNAs (Yoon et al., 2014; Kim et al., 2017). Quando os exossomos são liberados pelas células-mãe no fluido corporal, os circRNAs exossômicos (exo-circRNAs) iniciam sua circulação e suas funções biológicas. Em artigos recentes, exo-circRNAs demonstraram desempenhar um papel importante na iniciação, progressão e resistência à terapia do câncer, incluindo o câncer colorretal (Su et al., 2019; Ma et al., 2020; Papatsirou et al., 2020).

A localização dos circRNAs pode sugerir uma multiplicidade de papéis (Figura 1B) e possíveis funções epigenéticas (Jeck et al., 2013; Memczak et al., 2013; Z. Wang & CAS-MP, 2015; Y. Zhang et al., 2013). A modificação pós-transcricional é vital para a vida. Diversas evidências indicam que os circRNAs podem ser modificados por N-6 metilação (m6A) em muitos aspectos, como biogênese, exportação nuclear, tradução e degradação. Atualmente, m6A trata-se de uma modificação de RNA encontrada em muitas espécies. Muitos pesquisadores publicaram que m6A regula mRNAs de várias maneiras, como promoção da tradução e degradação de mRNAs (Zhou et al., 2015; Shi et al., 2017; Lee et al., 2020), contribuindo para o desenvolvimento de várias doenças em humanos. Embora os estudos sobre a modificação pós-transcricional de circRNAs sejam escassos, a modificação m6A nos circRNAs está se tornando um tema de pesquisa cada vez mais explorado (Mei et al., 2021).

Os circRNAs podem ser gerados por *splicing* canônicos e não-canônicos (*back-splicing*) (Dong et al., 2017). Até agora, três tipos de circRNAs produzidos a partir de diferentes mecanismos de ciclagem foram identificados por NGS: i) circRNAs exônicos (Jeck et al., 2013), que possuem sequências repetidas especiais, como os retroelementos *Alu* e outras sequências curtas, em seus dois lados (Figura 1A) (Jeck et al., 2013; Zheng et al., 2016); ii) circRNAs intrônicos (ciRNAs; Figura 1B) (Y. Zhang et al., 2013), dependentes de sequências ricas em GU perto do local de união de 5' e sequências ricas em C perto do ponto de derivação (Gall, 2014; Zhang et al., 2013); iii) circRNAs de introns retidos (ElciRNAs) (Salzman, Chen, Olsen, Wang, & Brown, 2013). Durante a formação de circRNAs, as sequências específicas são conservadas e as sequências de regiões *upstream* e *downstream* são ligadas entre si em uma direção inversa, o que contribui para a maturação dos circRNAs por um mecanismo chamado de *back-splicing* (Vicens et al., 2014). No entanto, detalhes dos mecanismos de biogênese de circRNAs ainda permanecem pouco claros.

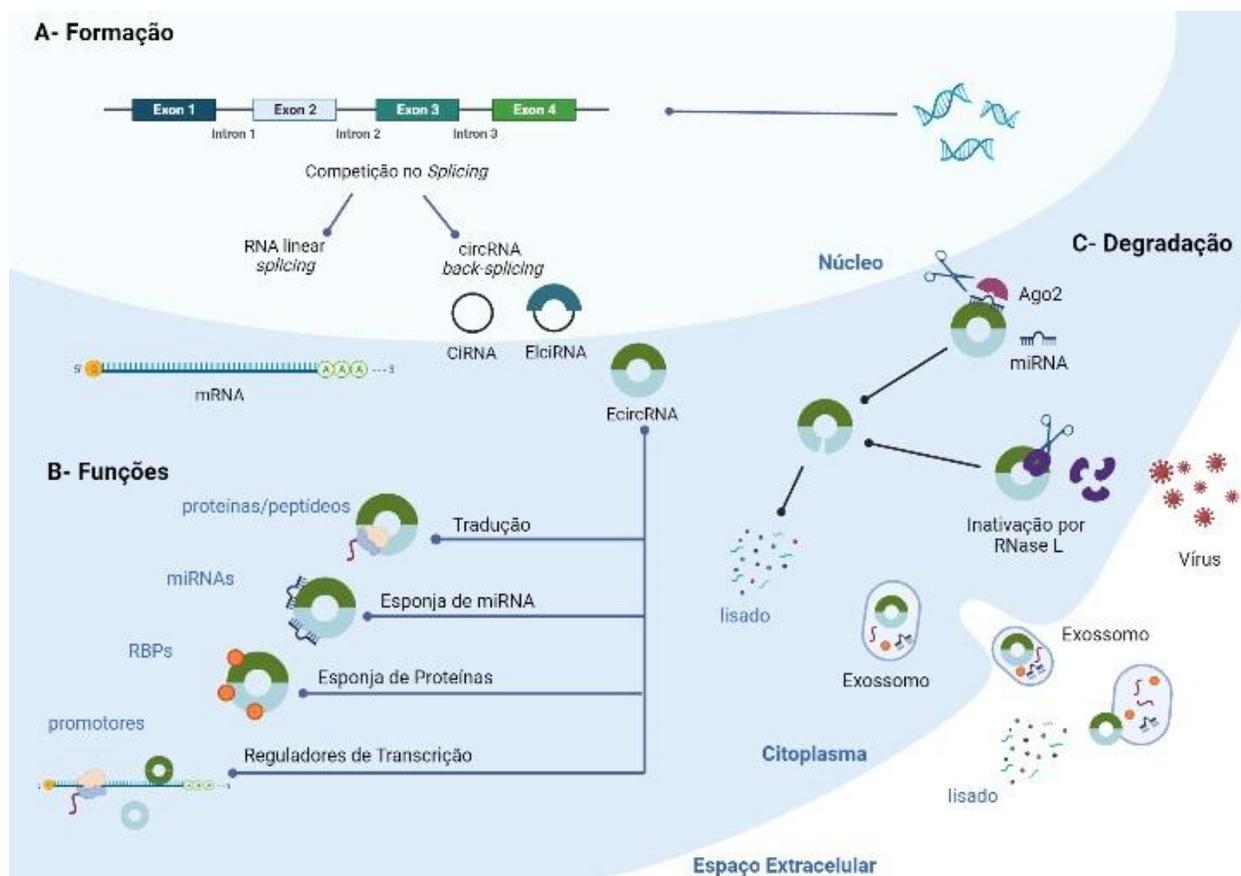


Figura 1- Formação, Função e Degradação dos circRNAs. **(A)** Formação dos circRNAs. O splicing de pré-mRNA e remoção de íntrons para formar RNA linear. O *Back-splicing* de pré-mRNA para formar circRNA competindo contra splicing linear de RNA. Os três tipos de circRNA: Exon circRNA, exon-intron circRNA e intron circRNA. **(B)** Funções dos circRNAs. Interações de circRNA na regulação da transcrição e ligação da região do promotor, esponja de miRNA, tradução, estrutura de proteína e proteínas de ligação de RNA (RBPs) (Li et al., 2021). **(C)** Degradação de circRNAs. miR-671 está localizado no núcleo e direciona a clivagem dependente do fatiador Ago2 de circCDR1. RNase L pode ser ativada quando as células são invadidas pelo vírus. Posteriormente, a RNase L ativada medeia a degradação global dos circRNAs, que é essencial para ativar a PKR no estágio inicial da resposta imune inata. A excreção celular de circRNAs para o espaço extracelular por meio de vesículas extracelulares (EVs) pode ser um mecanismo para a depuração celular de circRNAs (Tao et al., 2021) (Imagem adaptada de Li et al., 2021; Tao et al., 2021 pelo Biorender.com).

Semelhante ao RNA linear, a biogênese dos circRNAs é regulada por vários tipos de proteínas, ribonucleoproteína e complexos celulares, tais como as Proteínas de Ligação a RNAs (do Inglês, RBPs) o *RNA spliceosome*. O *spliceosome* de RNA é uma enzima à base de RNAs com um núcleo U5 constituído por 5 *small nuclear RNAs* (snRNAs) e uma série de proteínas. Sob o efeito dos fatores U6 e U2, muitas proteínas processam cooperativamente pré RNAs (Perriman et al., 1998). Semelhante à formação de mRNA, circRNAs podem ser formados a partir de precursores de RNA lineares pelos spliceossomos. Além disso, uma molécula de RNA linear pode ser processada em diferentes tipos de RNAs, incluindo mRNA, circRNA e ncRNA, através de diferentes eventos de *splicing* (Bartel, 2004). Registros indicam uma certa competição entre os fatores que realizam o *splicing* dessas moléculas (Chen et al., 2016), que pode ser um mecanismo de regulação de circRNAs. Outra proteína regulatória é a proteína Quaking (QKI), uma molécula representativa. Ela pode se ligar a regiões *downstream* específicas na sequência linear do pré mRNA e promover a ciclagem e *back-splicing* dos circRNAs (Jeck et al, 2008).

1.2 Metabolismo dos circRNAs

Os circRNAs são altamente estáveis, mas não significa que os circRNAs não sejam degradáveis. Suas formas de degradação em relação ao mRNAs se diferem por causa das diferenças em sua estrutura. Além disso, os mecanismos por quais

as células degradam e/ou removem os circRNAs ainda não são bem conhecidos (Tao et al., 2021).

Recentemente, os pesquisadores levantaram a possibilidade de que células possam eliminar circRNAs através da liberação de vesículas, como exossomos e microvesículas (Figura 1C). Em alguns estudos, para três linhagens celulares, os autores mostraram que os circRNAs examinados são enriquecidos em relação aos seus homólogos lineares nas preparações de vesículas extracelulares (EV) quando comparadas às células produtoras, o que sugere que a expulsão das células para o espaço extracelular pela liberação de EV pode ser um mecanismo pelo qual as células reduzem os circRNAs (Lasda & Parker, 2016).

Ainda existem evidências que demonstram que circRNAs podem ser degradados por outras maneiras, incluindo RNase-L, degradação Ago-dependente ou independente (Figura 1C) (Tao et al., 2021).

A degradação por RNase L dos circRNAs é relatada de várias maneiras (Han et al. 2014; Liu et al. 2019). Por exemplo, inicialmente a RNase L é codificada pelo *locus* hereditário do câncer de próstata 1 (HPC1) e funciona como uma endoribonuclease. Em segundo lugar, ele pode ser ativado por moléculas 2-5A da fórmula $[p\hat{x}5'A (2'p5'A)_n; \times 1 \text{ } 3; n > 2]$. O 2-5A é produzido a partir de ATP quando a 2'-5'-oligoadenilato sintetase (OAS) é ativada por DNAs de fita dupla (dsDNAs), a RNase L finalmente promove a degradação do vírus invasivo (Chakrabarti et al. 2011). É relatado que circRNAs são globalmente degradados por RNase L quando as células são infectadas pelo vírus, o que é essencial para a ativação de PKR no estágio inicial da resposta imune inata (Liu et al. 2019b).

Outro exemplo, é degradação de circRNAs dividida em degradação Ago-dependente e Ago-independente. Alguns estudos indicam que alguns miRNAs podem causar a degradação de circRNAs (Kleaveland et al. 2018; Hansen et al. 2011). Um exemplo desse tipo é o miR-671, que está localizado no núcleo e direciona a clivagem dependente de Ago2 de circCDR1 (Hansen et al. 2011).

1.3 Regulação da expressão gênica pelos circRNAs

Estudos anteriores forneceram evidências de que os circRNAs formam um grande número de reguladores pós-transcricionais (Figura 1B) (Memczak et al., 2013). Zhang et al, abordaram que circRNAs contendo íntrons, como ciRNAs e ElciRNAs, permanecem retidos no núcleo celular e podem interferir na transcrição e processamento de diversos transcritos, incluindo aqueles derivados dos mesmos loci genômicos (parentais) que os circRNAs. Os pesquisadores relataram após novos estudos que os ElciRNAs, como o circEIF3J e o circPAIP2, podem interagir com o U1 da pequena ribonucleoproteína nuclear U1 (U1snRNP) e a RNA polimerase II (Pol II) (Zhang et al., 2013). A interação direta desses ElciRNAs com o U1 formam um complexo ElciRNA-U1 que pode recrutar a RNA polimerase II (Pol II) para a região promotora de um gene, estimulando assim o início da transcrição (Li et al., 2016). Padrão semelhante também foi observado no circRNA PAIP2 e seu gene parental (Perriman et al., 1998). Deste modo, a ativação da transcrição pode ser uma função geral de RNAs circulares contendo sequências intrônicas e suas funções reguladoras podem explicar sua abundância no núcleo.

Assim, pode-se considerar que uma redução desses circRNAs abundantes no núcleo com pouco enriquecimento para os sítios alvo de miRNA podem levar, conseqüentemente, à redução da expressão de seus genes parentais (Memczak et al., 2013; Y. Zhang et al., 2013).

Outras evidências mostram o importante papel da regulação da expressão gênica dos circRNAs no câncer, como o Circ-CSPP1 que atua como esponja do miR-1236-3p e prejudica o efeito inibitório de miR-1236-3p em ZEB1, que subsequentemente promove a transição epitelial-mesenquimal e o desenvolvimento de câncer de ovário (Li et al., 2019). Também o circEGFR pode atuar como esponja para miR-125a-3p, modulando assim a expressão de Fyn e pode reproduzir um papel crítico nas células da granulosa ovariana de camundongos (Jia et al., 2018).

Outro importante papel dos circRNAs e sua função reguladora de expressão gênica é discutido em estudos na leucemogênese, principalmente na leucemia mielóide aguda. No estudo de Gaffo *et al*, é descrita a expressão de circRNAs em células B, células T e monócitos de indivíduos saudáveis. Eles confirmaram os dados por qRT-PCR para circRNAs específicos de células B (circPAX5, circAFF3, circIL4R e circSETBP1) ou células T (circIKZF1, circTNIK, circTXK e circFBXW7) e para circRNAs intrônicos (circBCL2) e regiões intergênicas que foram superexpressas em linfócitos. Além disso, no exame direcionado identificaram expressão impressionante e generalizada de circPAX5, circPVT1 e circHIPK3 na leucemia linfoblástica aguda do precursor B pediátrico e revelaram circRNAs com expressão variável em subtipos citogenéticos (Gaffo et al., 2019).

1.4 circRNAs e o câncer

Até o momento, há evidências robustas que mostram que circRNAs possuem estreita associação com doenças humanas, especialmente com diversos tipos de cânceres e podem ser usados como novos biomarcadores em amostras clínicas, inclusive de sangue padrão de pacientes oncológicos (Memczak, Papavasileiou, Peters, & Rajewsky, 2015).. No estudo de Bachmayr-Heyda *et al.* há o relato de que amostras de tumor colorretal têm menor número de circRNAs em comparação com a mucosa do cólon normal correspondente (Bachmayr-Heyda *et al.*, 2015). Em outro estudo, Wang *et al.* investigou a expressão do circRNA no câncer colorretal (CRC), o hsa_circ_001988. Os resultados mostraram que a expressão de hsa_circ_001988 foi diminuída nos tecidos tumorais, o que sugere que hsa_circ_001988 pode ser um novo alvo de tratamento e um potencial marcador biológico de câncer colorretal [24]. Outros estudos mostram que a expressão de cir-ITCH era tipicamente desregulada no CRC, e o cir-ITCH poderia aumentar o nível de ITCH, que está envolvido na inibição da via Wnt / β -catenina. Assim, o cir-ITCH pode desempenhar um papel no CRC regulando a via Wnt / β -catenina [25]. Deste modo, os estudos citados acima ilustram que os circRNAs são biomarcadores promissores para o CRC.

Artigos publicados demonstram que o hsa_circ_002059, foi significativamente reduzido em tecidos de câncer gástrico em comparação com tecidos não tumorais adjacentes pareados. Este estudo demonstrou claramente

as implicações clínicas de hsa_circ_002059 como biomarcador, uma vez que sua expressão é capaz de distinguir a mucosa gástrica adjacente normal do tecido do câncer gástrico. Além disso, este circRNA também pode ser usado como um biomarcador molecular na avaliação da eficácia da ressecção gástrica (P. Li et al., 2015).

Estudos relacionados a leucemia promielocítica aguda mostraram que translocações cromossômicas associadas ao câncer bem estabelecidas deram origem a circRNAs de fusão (f-circRNA) produzidos a partir de exons transcritos de genes distintos afetados pelas translocações. Assim, os pesquisadores analisaram a presença de f-circRNA (f-circPR e f-circM9) na linhagem celular NB4 derivada de leucemia promielocítica aguda (APL). Os dados sustentam a noção de que o f-circRNA, quando associado a outros estímulos oncogênicos, desempenha um papel ativo em favorecer a progressão da leucemia in vivo (Guarnerio et al., 2016). Alhasan et al. descobriram que os circRNAs são enriquecidos em plaquetas humanas em média de 17 a 188 vezes em relação aos tecidos nucleados. Pesquisas anteriores mostraram que o sequenciamento de mRNA de plaquetas sanguíneas com formação de tumor poderia distinguir pacientes com câncer de uma população saudável com 96% de precisão, e a localização do tumor primário foi corretamente identificada com precisão de 71% em seis tipos diferentes de tumor (Best et al, 2015). Esses resultados sugerem que os circRNAs e sua alteração de quantidade nas plaquetas podem desempenhar um papel proeminente e perene no diagnóstico e tratamento de diversos tipos de cânceres.

1.5 microRNAs e circRNAs no câncer

MicroRNAs (miRNA) são RNAs endógenos não codificantes com tamanho aproximado de 22 nucleotídeos que regulam a expressão do gene ligando-se às sequências complementares nos mRNAs, inibindo a síntese de seus produtos proteicos (Duursma et al., 2008; Lytle et al., 2007; Rigoutsos, 2009) e estimulando uma degradação acelerada desses mRNA. Vários miRNAs exibem padrões de expressão anormais em tumores humanos com consequente alteração de oncogenes alvo ou genes supressores de tumores (Medina and Slack, 2008). Devido às suas capacidades moduladoras de “ajuste fino”, os miRNAs são reguladores-chave em várias vias de sinalização envolvidas no desenvolvimento e na progressão do câncer (Malumbres, 2008).

Ser ‘esponja’ de miRNAs é o principal mecanismo de atuação dos circRNAs em células tumorais. Ainda não se entende como isso ocorre exatamente, mas o fato é que existem muitos sítios de ligação de miRNA em circRNAs. Assim, os circRNAs podem ser alvo de miRNAs, funcionando como uma esponja para captar esses RNA curtos e regulando (aumentando), de maneira indireta, a expressão de genes originalmente alvo de tais miRNAs. Como descrito anteriormente, nesse mecanismo, os circRNAs são definidos como ceRNA (*competing endogenous RNA*) e podem, por exemplo, contribuir para progressão ou manutenção tumoral (Y. Wang et al., 2017) por serem alvo de certos miRNAs.

Um dos exemplos é o ciRS-7, também chamado de CDT1as, um dos primeiros circRNAs conhecidos por atuar como um ceRNA em células tumorais (Peng, Yuan, & Li, 2015). ciRS-7 sofre o *splicing* a partir da transcrição anti-sentido do gene CDR1. Inicialmente, foi constatado que ele era sobre expresso em linhagens celulares tumorais malignas, como as células HeLa. Em seguida, em outros estudos, encontraram mais de 70 sítios de ligação de miRNAs em ciRS-7 (Memczak et al., 2013). Mais tarde, descobriu-se que o ciRS-7 pode ser alvo do miR7, funcionando como uma esponja e minimizando a quantidade livre de miR-7 (Xu, Guo, Li, & Yu, 2015). No entanto, o mecanismo de ação pode ser ainda mais complexo, e a interação de CDR1as com miR-7 pode ser direcionada por outro miRNA (miR-671) que é totalmente complementar e pode induzir a clivagem mediada por AGO deste circRNA (Chou et al., 2010; Hansen et al., 2011). Isto sugere que: i) o miR-671 diminui a inibição da expressão de miR-7 através de ciRS-7 aumentando o nível de miR7 nas células tumorais; ii) contribui para o aumento dos oncogenes alvos *downstream*, como EGFR e XIAP; iii) e promove vascularização, metástase e reprodução de células tumorais (Hansen et al., 2013; Xu et al., 2015) .

Outro exemplo é o circHIPK3, sua estrutura genômica mostra que o circHIPK3 contém um grande segundo exon (1.099 pb) do gene HIPK3, flanqueado por longos íntrons de cada lado. Huang encontrou altos níveis de circRNA, incluindo circHIPK3, no sequenciamento de RNA de sete tipos de amostras de tumor e seis tipos de tecidos normais (Zheng et al., 2016). Foi mostrado que: i) o silenciamento de circHIPK3 diminui a velocidade de proliferação das células tumorais (Zheng et al., 2016); ii) o circHIPK3 é alvo de

dezoito sítios de ligação de nove miRNAs (Zheng et al., 2016); iii) o circHIPK3 inibe a atividade do miR-124 trabalhando como esponja deste miRNA (Zheng et al., 2016). Em uma pesquisa, o miR-124 reduziu a expressão de iASPP e bloqueou o crescimento celular da linhagem celular de câncer de próstata PC3. Assim, miR-124 é um alvo potencial para a terapia do câncer (Chen et al., 2014).

1.6 circRNAs e o câncer colorretal

O câncer colorretal (CRC) é o terceiro câncer mais comumente diagnosticado no mundo (1,4 milhão em 2012) e a quarta maior causa de mortes por cânceres em todo o mundo (694.000 óbitos para ambos os sexos) (World Health Organization, 2012; Bosman et al, 2014). Em termos de incidência, representa quase 10% dos diagnósticos globais de câncer. Mais diagnósticos ocorrem em regiões desenvolvidas, mas as regiões menos desenvolvidas têm maiores taxas de mortalidade, refletindo uma taxa de sobrevivência mais pobre nesses países (World Health Organization, 2012; Ferlay et al., 2015). Sessenta e cinco por cento dos novos casos ocorrem em países desenvolvidos, incluindo os Estados Unidos, Canadá, Austrália e Europa. (Bosman et al., 2014; American Cancer Society, 2015). Também em 2016, era esperado que CRC causasse mais de 49.190 mortes apenas nos Estados Unidos (American Cancer Society, 2015). Em contrapartida, as taxas de incidência e mortalidade de CRC na Ásia aumentaram continuamente (Sung et al., 2015). Além da Ásia, em muitas regiões da Europa e a América do Norte, as taxas de incidência e mortalidade por CRC diminuíram

tanto para homens como para mulheres, devido a novos métodos de triagem que permitem o diagnóstico e tratamento precoce (American Cancer Society, 2015; American Cancer Society, 2016). Esses testes de triagem podem impedir o desenvolvimento do CRC, já que alguns têm o potencial de detectar pólipos antes que possam progredir em câncer. Nem todos os pólipos irão progredir em tumores de câncer, mas sua remoção pode prevenir a doença (American Cancer Society, 2016). Em 2020, segundo os dados publicados pelo INCA, o CRC é o segundo câncer mais prevalente na população masculina no Brasil, com 20.540 novos casos, e na população feminina, com 20.470 novos casos. É o terceiro com maior mortalidade, com 10.191 óbitos no sexo masculino e 10.385 óbitos no sexo feminino (INCA, 2021).

Diversas revisões relataram que em linhagens celulares de CRC e tecidos de CRC, observa-se uma redução global da abundância de circRNA, em comparação com tecido saudável, permitindo a proliferação de células do CRC (Arvelo et al., 2015; Yang et al., 2016). Bachmayr-Heyda et al (Bachmayr-Heyda et al., 2015) relatou uma redução global da abundância de circRNA em linhagens celulares de CRC e amostras de tumores, em comparação com a mucosa normal em pacientes com CRC. Usando *RNA-seq* e o algoritmo descrito por Memczak et al (Memczak et al., 2013), confirmados por qRT-PCR, foram encontrados cerca de 39 circRNAs diferencialmente expressos na mucosa do cólon normal e amostras de CRC, sendo que 11 estavam sobre (*upregulated*) e 28 sub (*downregulated*) regulados. Adicionalmente, o grupo de Bachmayr descobriu que a expressão de circRNAs foi reduzida em amostras de tumores (expressão de 27,8%) em comparação com amostras normais de mucosa do cólon (expressão de 78,1%).

Evidências semelhantes foram encontradas em 11 linhagens celulares do CRC, com intervalos ainda maiores nas proporções de expressão (Bachmayr-Heyda et al., 2015). Para validar os circRNAs *downregulated* em CRC em comparação com amostras normais de mucosa, eles enriqueceram circRNAs através da digestão com RNase R e subsequente *deep sequencing*, relatando 21.653 *back spliced junctions*. Desta forma, este método revelou que é impossível garantir a detecção de todos os circRNA, e que o número real de tipos de circRNA é muito maior. Isso também foi corroborado em outras linhagens celulares em uma doença neoproliferativa não cancerígena, mostrando uma correlação negativa semelhante à do índice de circRNA e à proliferação observada em tecidos do cólon e linhagens celulares (Bachmayr-Heyda et al., 2015). Os autores hipotetizaram que as maquinarias de *back-splice* responsáveis pela circularização dos RNAs são disfuncionais em células tumorais se devia a uma degradação aumentada por oncomiRNAs (Bachmayr-Heyda et al., 2015; Ragusa et al., 2015).

Em outro estudo, 45 amostras pareadas de pacientes com CRC, foram analisadas e a expressão de cir-ITCH foi significativamente reduzida nos tecidos do CRC em comparação com o tecido peritumoral, confirmando que cir-ITCH estava envolvido na inibição da via Wnt / beta-catenina (Huang et al., 2015). O ciRS-7, uma esponja potencial de miR-7, foi significativamente regulado nos tecidos de CRC em comparação com a mucosa normal correspondente e emergiu como fator de risco independente para a sobrevivência global. Assim, CiRS-7 foi eleito um biomarcador prognóstico promissor e serviu como um alvo terapêutico para reduzir a atividade de EGFR-RAF1 em pacientes com câncer colorretal (Weng et al., 2017).

Os circRNAs exônicos de soro têm sido usados para distinguir pacientes com CRC de controles saudáveis (Li et al., 2015). Foram analisados mais de 1.800 RNAs circulares. A proporção de isoformas de RNA circular a linear foi significativamente menor em tumores em comparação com amostras normais de cólon e muito mais baixas nas linhagens celulares de câncer colorretal, além de serem correlacionadas negativamente com o índice de proliferação (Bachmayr-Heyda et al., 2015). Utilizando 89 amostras pareadas de HCC e tecidos de fígado adjacentes, hsa_circ_0001649 foi significativamente reduzido e seu nível de expressão foi estreitamente correlacionado com o tamanho do tumor (p-valor = 0,045) e a ocorrência de êmbolo tumoral (p-valor = 0,017). O estudo concluiu que hsa_circ_0001649 esteve envolvido na regulação da tumorigênese e metástase de HCC e pode servir como um novo biomarcador potencial para HCC humano (Qin et al., 2016).

Muitos outros estudos mostram a importância dos circRNAs como biomarcadores diretos ou indiretos do câncer colorretal, é imprescindível o enfoque nessa área, assim podemos avançar no prognóstico e tratamento dessa patologia com ocorrência tão acentuada.

1.7 Linhagens

A invasão e progressão do tumor é um processo de etapas múltiplas que requer células tumorais para sobreviver na circulação, extravasar e colonizar locais distantes (Jin et al., 2017), esse processo é denominado metástase, sendo a

principal causa de morte para pacientes com câncer (Chaffer and Weinberg, 2011). Os processos gerais da metástase são semelhantes entre a maioria dos tipos de tumor, embora a metástase para diferentes tecidos parece exigir conjuntos distintos de reguladores do hospedeiro. Muitos genes que envolvem metástases tumorais foram identificados (Jin et al., 2017), incluindo genes codificadores de proteínas e não codificadores, como miRNAs.

As linhagens celulares SW480 e SW620 foram estabelecidas a partir de um paciente do sexo masculino com CRC (Leibovitz et al., 1976). As células SW480 são derivadas do tumor primário, enquanto as células SW620 são derivadas de uma metástase linfonodal retirada do mesmo indivíduo (Provenzani et al., 2006). Como esperado, as células SW620 apresentam propriedades de crescimento e migração em placa (mimetizando a metástase) mais acentuadas do que as células SW480 (Bergmann- Leitner et al., 2000; Hewitt et al., 2000; Provenzani et al., 2006). Por exemplo, a tendência ao crescimento autônomo de clones de células SW620 é três vezes maior que nas células SW480 (Provenzani et al., 2006). Consequentemente, as células SW620 têm sido rotineiramente usadas para estudar metástase e metástase associados ao CRC (Provenzani et al., 2006; Adler et al., 2014; Bachmayr-Heyda et al., 2015; Paschall et al., 2016; Rasmussen et al., 2016).

Objetivos

Objetivo geral

Caracterizar a presença e estudar o perfil de expressão de RNAs circulares (circRNA) em duas linhagens de células, de tumor primário (câncer colorretal) e metastático do mesmo paciente, com o intuito de entender o papel dos circRNAs na tumorigênese e progressão tumoral.

Objetivos específicos

- Identificar os circRNAs transcritos nas linhagens SW480 (tumor primário) e SW620 (tumor metastático).
- Identificar os circRNAs diferencialmente expressos nas linhagens celulares SW480 e SW620.
- Selecionar os circRNAs diferencialmente expressos em SW620 e avaliar as possíveis relações destes circRNAs com o processo metastático, tal como mobilidade celular e invasão tecidual.
- Avaliar de forma global o potencial dos circRNAs aqui encontrados terem algum papel na origem, manutenção e progressão do câncer colorretal.

Materiais e Métodos

3.1- Cultura de células e extração de RNA

As linhagens celulares de tumor do cólon SW480 e SW620 foram obtidas da *American Type Culture Collection* (ATCC) e cultivadas de acordo com as recomendações. Ambas foram mantidas em meio L15 Leibovitz (Gibco®) suplementado com soro fetal bovino (10%) (Gibco®) a 37 °C, em uma estufa livre de CO₂. O teste de autenticação das linhagens foi realizado usando GenePrint® (Promega). O teste para atestar ausência de contaminação por micoplasma foi feito por técnica de *PCR* convencional, usando-se primers específicos.

A extração RNA foi feita em confluência de 70-80%, Invitrogen™ Trizol da Thermo Fisher (de acordo com as recomendações do fabricante) (Invitrogen, 2016). Foi feita extração do RNA das células utilizando protocolo padrão de extração por Invitrogen™ Trizol da Thermo Fisher (Invitrogen, 2016).

3.1.1- Lise das amostras e separação de fases

Foi removido o meio de crescimento das células. Adicionou-se 0,3-0,4 mL de reagente TRIZOL™ por 1×10^5 - 10^7 células diretamente à placa de cultura para lisar as células. O lisado foi pipetado para cima e para baixo várias vezes para

homogeneizar. Foi incubado por 5 minutos para permitir a dissociação completa. Adicionou-se 0,2 mL de clorofórmio por 1 mL de reagente TRIzol™ usado para lise. Foi incubado por 2 a 3 minutos. Em seguida, centrifugou-se a amostra por 15 minutos a $12.000 \times g$. A mistura se separa em 3 fases, a mais baixa é vermelha, há a intermediária, e a superior é uma fase aquosa incolor, que contém o RNA. Essa foi transferida para um novo tubo (Invitrogen, 2016).

3.1.2- Precipitação, lavagem e ressuspensão do RNA

Adicionou-se 0,5 mL de isopropanol à fase aquosa, por 1 mL de reagente TRIzol™ usado para lise. Foi incubado por 10 minutos. Após a incubação, foi feita uma centrifugação por 10 minutos a $12.000 \times g$. O precipitado de RNA total forma um grânulo branco tipo gel no fundo do tubo, assim o sobrenadante foi descartado com um micropipeta (Invitrogen, 2016).

O sedimento foi ressuspendido em 1 mL de etanol a 75% por 1 mL de reagente TRIzol™ usado para lise. Depois, foi feita uma agitação breve no vórtex da amostra e, em seguida, centrifugou-se por 5 minutos a $7500 \times g$ a 4°C . O sobrenadante foi descartado com uma micropipeta. O sedimento de RNA foi mantido em repouso para secar ao ar cerca de 5 a 10 minutos (Invitrogen, 2016).

O sedimento foi ressuspendido em 20–50 μL de água livre de Rnase. Após isso, foi incubado em banho-maria a $55\text{--}60^{\circ}\text{C}$ por 10-15 minutos. O RNA foi armazenado a -70°C até o uso (Invitrogen, 2016).

3.2- Rna Sequencing

Após a extração, foi determinada a concentração do RNA através do método de fluorescência e aparelho Qubit™ (Invitrogen). Para aceitação do *RNA integrity number* (RIN) o valor deveria estar entre 9,6 a 10. Foi utilizado um valor de input da amostra para início do preparo da biblioteca de 10 µg por condição. Foi utilizado o kit *TruSeq Stranded mRNA* da Illumina® para o preparo da biblioteca. No experimento de *RNA-Seq* ribossômico, foi utilizado tratamento *Ribo Minus Eukaryote kit for RNA-Seq* (Invitrogen- A10837-08) para depleção do RNA ribossômico. No experimento de *RNA-Seq* para favorecimento da captura de circRNAs, foi feito o tratamento com RNase R (NEB). A quantificação da biblioteca foi feita por *KAPA Library Quantification* (Roche). A avaliação do tamanho dos fragmentos foi feita utilizando o kit e aparelho para Bioanalyzer 2100 (Agilent). O sequenciamento foi feito na plataforma NextSeq 500 Illumina®, com kit 300 ciclos, *paired end* e *single index*.

3.3- Qualidade da corrida e métricas

A qualidade da corrida foi avaliada inicialmente pelo *Sequencing Analysis Viewer* (SAV) versão 1.8.37 Software Illumina® (Illumina, 2014) que permite visualizar em tempo real métricas importantes de qualidade geradas pelo *software Real-Time Analysis* (RTA) nos sistemas de sequenciamento Illumina. Em seguida, foi utilizado o *Illumina bcl2fastq Conversion Software* versão 1.8.4

(Illumina, 2013) para demultiplexar os dados de sequenciamento e convertê-los de arquivos de chamada básica (BCL) para arquivos FASTQ, arquivo o qual é requerido em diversos programas como *input*. Sempre que um ciclo de execução de sequenciamento termina, o software *Real-Time Analysis* (RTA) gera um arquivo BCL contendo chamadas básicas e índices de qualidade associados (*Q-scores*).

Além disso, foi utilizado o programa FastQC versão 0.11.7 (Andrews, 2010) para gerar as métricas a partir dos arquivos FASTQ antes de prosseguir com as demais análises. O FastQC tem como objetivo fornecer uma maneira simples de realizar verificações de controle de qualidade em dados de sequência bruta provenientes de *pipelines* de sequenciamento de alto rendimento. Ele fornece um conjunto modular de análises, como tabelas, gráficos e relatórios, para dar uma rápida impressão se os dados apresentam algum problema antes de fazer qualquer análise adicional.

Para complementar a etapa de análise de qualidade, após o alinhamento das amostras ao genoma referência, foi utilizado o RNA-SeQC versão 1.1.8 (DeLuca et al., 2012), junto a referência GRCh38 (Schneider et al., 2016) e o Gencode versão 27 (Harrow et al., 2012). O RNA-SeQC é um programa escrito em java que calcula uma série de métricas de controle de qualidade para dados de *RNA-seq*. A entrada pode ser um ou mais arquivos BAM. A saída consiste em relatórios HTML e arquivos delimitados por tabulação de dados de métricas. Esse programa avalia a qualidade do sequenciamento em diferentes amostras ou experimentos para

avaliar diferentes parâmetros experimentais (DeLuca et al., 2012). Este foi aplicado aos arquivos BAM de cada condição estudada por *lane*.

3.4- Alinhamento

Após o *demultiplexing* das amostras, os arquivos FASTQ foram concatenados por *lane* de replicata técnica por cada condição e em seguida foram alinhados ao genoma referência GRCh38 (Schneider et al., 2016) pelo programa STAR versão 2.5.3 (Dobin et al., 2016). O *software* STAR é capaz de gerar alinhamentos com altos níveis de precisão e velocidade. Além de detectar *fusion junctions* anotadas e novas, o STAR é capaz de descobrir arranjos de sequência de RNA mais complexos, como quimerismo e circRNAs. Foi utilizado o parâmetro `-chimSegmentMin` para detecção de *fusion junctions*, necessário para detecção dos circRNAs posteriormente. Após o alinhamento, esses arquivos ".Chimeric.out.junction" foram convertidos em arquivos ".fusion_junction.txt" que serviram de input para etapa de detecção dos circRNAs. Para o experimento de avaliação do *RNA-Seq*, os arquivos BAM provenientes do alinhamento seguiram para o pipeline para gerar *counts*. Para o experimento de circRNAs, uma etapa de detecção destes foi realizada.

3.5- Detecção dos circRNAs

Para a detecção dos circRNAs, foi utilizado o programa CircExplorer versão 1.1.10 (Zhang et al., 2014).

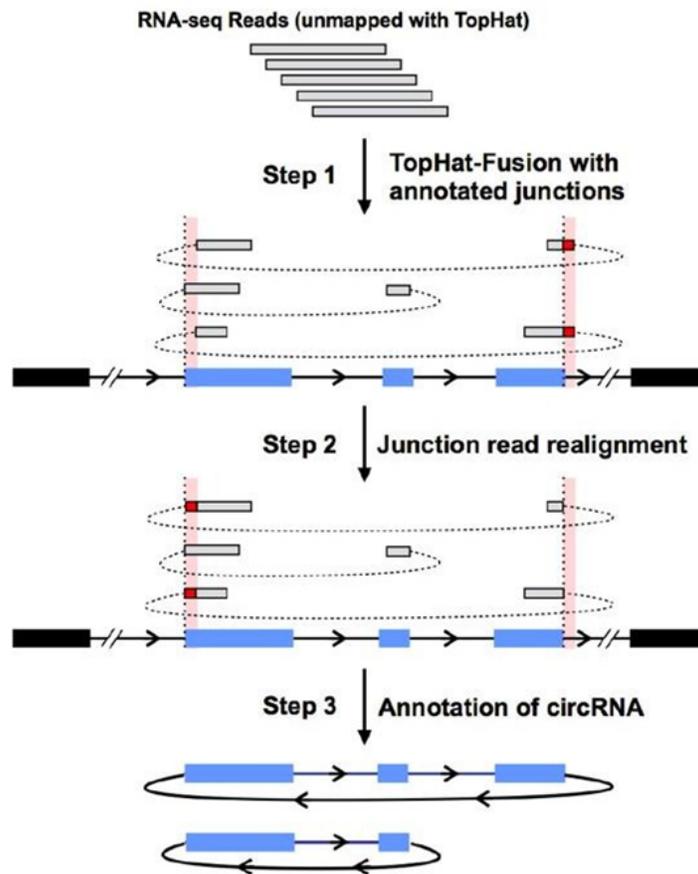


Figura 2: Fluxo esquemático do *pipeline* do CircExplorer. O CircExplorer é um *pipeline* computacional para leitura de *fusion junction* posterior, chamando para anotar com precisão RNAs circulares. Na versão utilizada, o algoritmo STAR foi aplicado para chamar leituras de junção (barras finas crescidas, etapa 1). Devido à semelhança de sequência no final de certos exons (barras azuis), as leituras de junção mapeadas foram divididas (linhas do arco do traço) em locais incorretos, levando ao mapeamento para locais incorretos (barra fina vermelha) (Zhang et al., 2014). Esses erros de mapeamento foram ajustados com um algoritmo personalizado (etapa 2). Os RNAs circulares previstos a partir de exons com emenda posterior (barras azuis) foram somados com as leituras de

junção previstas (etapa 3) na tabela disponibilizada no artigo publicado do pipeline (Zhang et al., 2014). Os exons que não retornam emendas são indicados como barras cinza.

Os arquivos “.fusion_junction.txt” gerados pelo STAR na etapa de alinhamento foram utilizados como *input* para o *pipeline* do CIRCexplorer, juntamente com o genoma referência GRCh38 (Schneider et al., 2016) e o arquivo referência montado e modificado a partir do arquivo de coordenadas “.gtf” do GRCh38, o *download* foi feito a partir do *link* do site oficial a seguir (<http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/refGene.txt.gz>), seguindo o padrão recomendado pelo CIRCexplorer (Tabela 1).

Tabela 1: Modelo de arquivo referência do CIRCexplorer.

<i>string name</i>	"Name of gene"
<i>string chrom</i>	"Chromosome name"
<i>char[1] strand</i>	"+ or - for strand"
<i>uint txStart</i>	"Transcription start position"
<i>uint txEnd</i>	"Transcription end position"
<i>uint cdsStart</i>	"Coding region start"
<i>uint cdsEnd</i>	"Coding region end"
<i>uint exonCount</i>	"Number of exons"

<i>uint[exonCount] exonStarts</i>	"Exon start positions"
<i>uint[exonCount] exonEnds</i>	"Exon end positions"

Após o processamento, foi gerado um arquivo *output* com o seguinte *layout* (Tabela 2).

Tabela 2: Modelo do arquivo de *output* do resultado final do CIRCexplorer.

Campo	Descrição
<i>chrom</i>	<i>Chromosome</i>
<i>start</i>	<i>Start of junction</i>
<i>end</i>	<i>End of junction</i>
<i>name</i>	<i>Circular RNA/ Junction reads</i>
<i>score</i>	<i>Flag to indicate realignment of fusion junctions</i>
<i>strand</i>	<i>+ or - for strand</i>
<i>thickStart</i>	<i>No meaning</i>
<i>thickEnd</i>	<i>No meaning</i>
<i>itemRgb</i>	<i>0,0,0</i>

<i>exonCount</i>	<i>Number of exons</i>
<i>exonSizes</i>	<i>Exon sizes</i>
<i>exonOffsets</i>	<i>Exon offsets</i>
<i>readNumber</i>	<i>Number of junction reads</i>
<i>circType</i>	<i>'circRNA' or 'ciRNA'</i>
<i>geneName</i>	<i>Name of gene</i>
<i>isoformName</i>	<i>Name of isoform</i>
<i>exonIndex/intronIndex</i>	<i>Index (start from 1) of exon (for circRNA) or intron (for ciRNA) in given isoform (newly added in 1.1.6)</i>
<i>flankIntron</i>	<i>Left intron/ Right intron</i>

3.6- Counts

Foram utilizados dois programas e duas metodologias para gerar os *counts*, o Kallisto versão 0.43.1 (Bray et al., 2016) e o RSEM versão 1.3.0 (Li and Dewey, 2011) descritos abaixo. Em um primeiro momento, os *RNA-Seqs* foram analisados seguindo-se pipeline para genes totais utilizando a referência GRCh38 (Schneider et al., 2016) com objetivo de validar os tipos de preparos de bibliotecas, com e sem tratamento para circRNAs. Em um segundo momento, as coordenadas do arquivo de *output* do CIRCexplorer foram utilizadas como

referência (processo de desenvolvimento descrito a seguir) para montagem do *index* de cada *pipeline* de contagem escolhido.

3.6.1- Montagem do arquivo FASTA de circRNAs

Após processamento dos dados com o CIRCexplorer, os arquivos de *output* com as coordenadas dos circRNAs foram utilizados para montagem dos arquivos de referência que foram utilizados para gerar o *index* dos programas de contagem.

Esses arquivos de coordenadas foram transformados em arquivos BED. Foi utilizado o comando *getfasta* do pacote de utilidades Bedtools versão 2-2.26.0 (Quinlan and Hall, 2010). Este pipeline extrai sequências de um arquivo FASTA para cada um dos intervalos definidos em um arquivo BED / GFF / VCF. Foi utilizado como referência o arquivo FASTA do genoma referência GRCh38 (Schneider et al., 2016).

3.6.2- Kallisto

Para a etapa de *counts*, foi utilizado o Kallisto versão 0.43.1 (Bray et al., 2016). O Kallisto é um programa para quantificar a abundância de transcritos a partir de dados de *RNA-Seq*. Baseia-se na ideia de pseudo-alinhamento para determinar

rapidamente a compatibilidade das leituras com os alvos, sem a necessidade de alinhamento. Como input, foram utilizados os arquivos FASTQ concatenados por lane de cada condição. Para gerar o index do Kallisto, para o primeiro experimento foi utilizado o arquivo FASTA de referência GRCh38 (Schneider et al., 2016). Para o segundo experimento, os FASTAs gerados por condição pelo Bedtools com as coordenadas do circRNAs foram concatenados e utilizados para gerar o *index* do Kallisto.

3.6.3- RSEM

Também para gerar os *counts* foi utilizado o programa RSEM versão 1.3.0 (Li and Dewey, 2011). O RSEM é um pacote de *software* para estimar os níveis de expressão de genes e isoformas a partir de dados de *RNA-Seq*. O pacote RSEM fornece uma interface amigável, suporta *threads* para computação paralela, dados de leitura de *single end* e *paired end*, índices de qualidade, leituras de comprimento variável e estimativa de *RSPD*. Além disso, fornece estimativas de média posterior e intervalo de credibilidade de 95% para os níveis de expressão. O *index* do RSEM foi gerado utilizando o FASTA referência GRCh38 (Schneider et al., 2016) e o arquivo GTF de anotação do Gencode para GRCh38 versão 27 (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_27/).

3.7- Expressão diferencial

Para a expressão diferencial, foi utilizado o programa edgeR versão 3.24.3 (Robinson et al., 2010) e o limma versão 3.38.3 (Ritchie et al., 2015) no *software* R versão 3.5.1 (R Core Team, 2014). O edgeR provém análises estatísticas da expressão diferencial sobre rotinas de *RNA-Seq*. Uma característica particular da funcionalidade do edgeR, é o método empírico de Bayes que permite a estimativa de variação biológica específica de um gene, mesmo para experimentos com níveis mínimos de replicação biológica, por isso a escolha deste algoritmo para o atual projeto. O edgeR pode ser aplicado à expressão diferencial no gene, exon, transcritos ou *tag level*, ou seja, os *counts* podem ser resumidos por qualquer recurso genômico. As análises edgeR no nível do exon são facilmente estendidas para detectar *junctions* diferenciais ou expressão diferencial específica de isoformas (Robinson et al., 2010). O limma é um pacote de software R/Biocondutor que fornece uma solução integrada para analisar dados de experimentos de expressão gênica. Ele contém recursos avançados para lidar com projetos experimentais complexos e para superar o problema de falta de replicatas. O limma tem sido uma escolha popular para a descoberta de genes através de análises de expressão diferencial de *microarray* e dados de *PCR* de alto rendimento. O pacote contém facilidades fortes para ler, normalizar e explorar esses dados. Os recursos do limma foram significativamente expandidos em duas direções importantes. Primeiro, o pacote pode executar análises de expressão diferencial e de emenda diferencial de dados de sequenciamento de *RNA-seq*. Segundo, ele é capaz de passar pelas análises tradicionais de expressão por genes de várias maneiras, analisando perfis de expressão em termos de

conjuntos de genes co-regulados ou em termos de assinaturas de expressão *upstream* (Ritchie et al., 2015).

3.8- Gráficos

Os gráficos foram feitos no *software* R versão 3.5.1 e R versão 4.1.2 (R Core Team, 2014) utilizando respectivos pacotes para cada tipo de gráfico construído.

3.9- Vias e Redes de interação

As vias de enriquecimento KEGG foram feitas por dois algoritmos distintos, com acesso online disponível. O primeiro foi a plataforma WebGestalt 2019 (Liao et al., 2019) que utilizou como referência o genoma humano, e a plataforma ShinyGO (Ge et al., 2020), com seleção da análise com referência à espécie humana. Para os termos *Gene Ontology* (GO) foi utilizada a plataforma DAVID versão 6.8 (Huang et al., 2009), a plataforma GOrilla (Eden et al., 2009) e a plataforma WebGestalt 2019 (Liao et al., 2019). Para demonstração dos termos enriquecidos GO foi utilizado o *software* REVIGO (Supek et al., 2011) utilizando *p-value* menor ou igual a 0.05 para os termos enriquecidos do DAVID e um *threshold* do *p-value* de 10^{-3} para os termos enriquecidos do GOrilla, no *software* R versão 3.5.1 e 4.1.2 (R Core Team, 2014). Para demonstração das vias KEGG enriquecidas pelo *log2foldchange* foi utilizado o *software* Pathview (Luo

and Brouwer, 2013). Para complementação da visualização através de gráficos das vias e linhagens de câncer associadas ao alvo foi utilizado o GSEA (Mootha et al., 2003) versão online. Para as redes de *Protein-Protein Interaction* (PPI) foi utilizado o *software* STRING (Jensen et al., 2009).

3.10- Bancos de dados

Com a aplicação de sequenciamento de próxima geração e ferramentas de bioinformática, muitos circRNAs estão sendo descobertos e integrados aos diversos bancos de dados de circRNAs (Szabo, 2016). Os bancos de dados Cirbase e Deepbase 2.0 contêm mais de 150.000 mil circRNAs descritos e sequências de RNA maduro de diferentes espécies, incluindo humanos, moscas da fruta e nemátodo (Glazar et al., 2014; J. H. Li, Liu, Zhou, Qu, & Yang, 2014; L. L. Zheng et al., 2016). CircRNADb, contém 32.914 circRNAs exônicos humanos cuidadosamente selecionados de fontes diversificadas, apresenta diversos detalhes do circRNA, incluindo informações genômicas, splicing de exon, sequência do genoma, local de entrada do ribossomo interno (IRES), estrutura de leitura aberta (ORF) e referências (Chen et al., 2016). O banco de dados Cir2Traits contém circRNAs relacionados a várias doenças (Ghosal, Das, Sen, Basak, & Chakrabarti, 2013). Circnet e circRNABase contêm variantes alternadas e seus padrões de expressão em diferentes tecidos e as possíveis redes reguladas por miRNA (J. H. Li et al., 2014). O banco de dados CirInteractome contém sequências de genoma correspondentes, *primers* para qRT-PCR e siRNAs de circRNAs maduros. Também

pode ser usado para prever os locais de ligação RBP (Dudekula et al., 2016). Os bancos de dados CIRCpedia e TSCD contêm dados *RNA-seq* de várias linhas de células humanas e amostras (Zhang et al., 2016; Xia et al., 2016). Assim, esses diversos bancos de dados ajudam a anotar e elucidar funções dos circRNAs em diversos estudos, tecidos e condições.

Resultados e Discussão

Primeiramente, foram analisadas as métricas da corrida e qualidade do sequenciamento antes de se prosseguir com as demais análises do *RNA-Seq*. Nos resultados do SAV Illumina, a clusterização atingiu 177 mm², dentro do esperado e permitido segundo as boas práticas da documentação da Illumina. O Q30 se mostrou acima dos 75% indicados no manual do fabricante (80%) para uma corrida de 300 ciclos no NextSeq. O *Cluster Passing Filter* atingiu 85%. O número total de *reads* foi de 459.564.242, com um total de *reads* mapeados de 390.629.606 (Tabela 3). A distribuição dos *reads* entre e ao longo das 4 lanes se mostrou homogênea (Figura 3C). Os demais parâmetros também estavam de acordo com o manual (Figura 3) (Gruber et al., 2010) *Illumina NextSeq Specification Parameters*. Assim, concluiu-se que a corrida apresentou ótimos valores para os parâmetros analisados de acordo com as especificações do fabricante, retirando a possibilidade de erro de processamento e qualidade da corrida das bibliotecas sequenciadas, nos habilitando a prosseguir com as análises.

Tabela 3: Valores das métricas de qualidade da corrida obtidos através da análise pelo SAV.

Critérios	Valor
-----------	-------

<i>clusters/mm2</i>	177
<i>Cluster passed</i>	85%
<i>Q>30</i>	80%
<i>Total reads</i>	459,564,242
<i>Mapped reads</i>	390,629,606

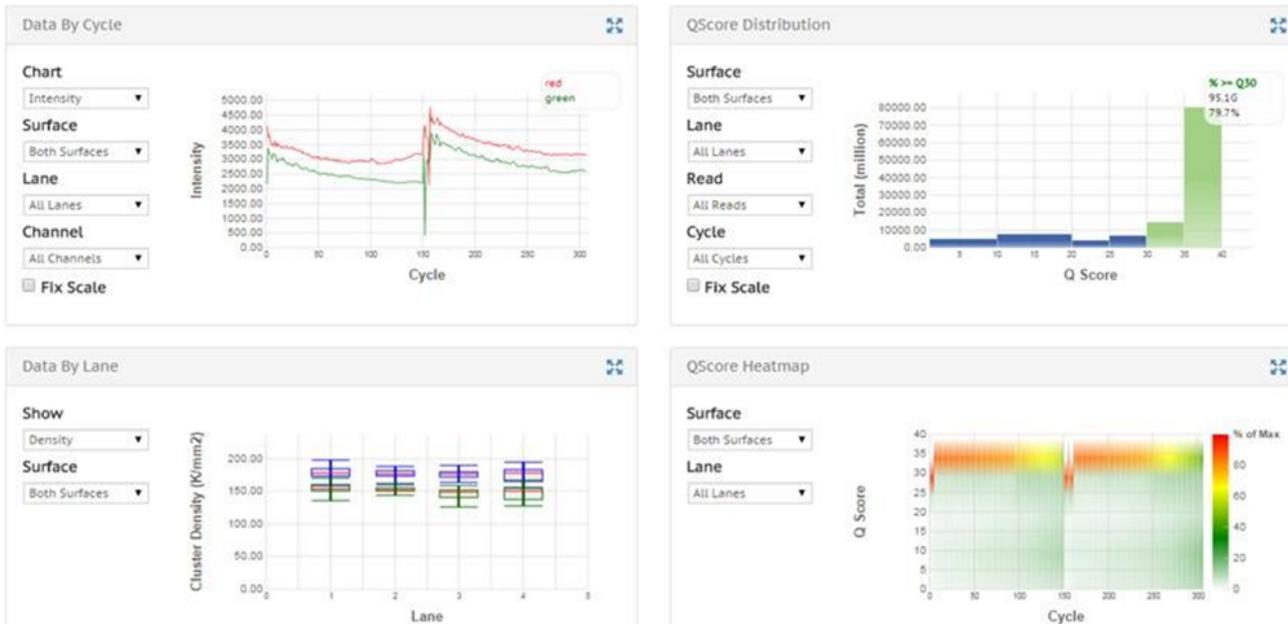


Figura 3: Gráficos das métricas de qualidade do sequenciamento gerados pelo SAV. **A)** Histograma da intensidade de leitura por ciclo. **B)** Bar-plot da análise do Q30. **C)** Box-plot da distribuição dos reads representados pelo Cluster Density por lane. **D)** Heatmap da análise do Q30 por ciclo.

Após o *demultiplex* das amostras pelo *bcl2fastq*, foi feita uma análise de qualidade dos arquivos do sequenciamento FASTQ, pelo programa FastQC com o objetivo de realizar algumas verificações de controle de qualidade nesses arquivos de dados de sequência bruta (Anexo 2). A média do total de sequências detectadas foi de 22.000.000 para cada arquivo (R1 e R2), com um tamanho médio de fragmento de 35-150 e uma porcentagem de GC de 57, não havia nenhuma contaminação de adaptadores e uma qualidade por base sequenciada dentro do padrão de qualidade, tanto para os arquivos R1 quanto para R2. Neste segundo *check-point* de qualidade, todos os resultados obtidos possuíram valores dentro do critério aceitável. Vale ressaltar, que isso foi observado tanto no preparo padrão da biblioteca de *RNA-Seq* quanto na biblioteca para detecção de circRNAs.

Os resultados das métricas da etapa de alinhamento, indicam uma boa porcentagem de *reads* mapeados, sendo por volta de 84% para ambas as linhagens celulares com a preparação padrão e por volta de 80% para as bibliotecas enriquecidas para os circRNAs (Tabela 4). Essa pequena diferença (84% x 80%) de *reads* mapeados entre as duas preparações foi o primeiro indicativo de que as bibliotecas customizadas (circ) apresentam mais *reads* reportando RNAs circulares, o que certamente dificulta o trabalho de mapeamento contra o genoma.

Tabela 4: Métricas do alinhamento dos *reads* das bibliotecas padrão (SW480/SW620-ribo) e preparação feita pelo nosso grupo (SW480/SW620-circ).

	SW480-ribo	SW480-circ	SW620-ribo	SW620-circ
Total de reads	104.533.007	88.537.286	91.448.989	97.363.769
Mapeados unicamente	59.728.108 (57,14%)	25.841.982 (29,19%)	49.150.678 (53,75%)	34.535.217 (35,47%)
Mapeados múltiplas vezes	28.425.022 (27,19%)	43.903.798 (49,59%)	26.804.980 (29,31%)	43.927.829 (45,12%)
Total mapeado	84,33%	78,78%	83,06%	80,59%

Foi feita a quantificação dos genes expressos para cada linhagem celular e diferentes preparos de biblioteca, padrão (ribo) e para identificação de circRNAs (circ). Enquanto o preparo padrão foi capaz de detectar entre 3 mil e 4 mil circRNAs, o preparo feito pelo nosso grupo detectou entre 11,7 mil e 14,7 mil, um enriquecimento de quase 4 vezes, indicando o sucesso experimental do protocolo adaptado (Tabela 5). Em seguida, foi feita a análise do número de circRNAs identificados entre as duas linhagens celulares, SW480 e SW620. A linhagem metastática (SW620) apresentou aproximadamente 30% mais circRNAs expressos do que a linhagem de tumor primário (SW480). Curiosamente, os circRNAs intrônicos estão sutilmente mais presentes na SW480.

Tabela 5: Comparação entre as linhagens e entre os preparos de biblioteca para detecção dos circRNAs.

	circRNAs (total)	EcircRNAs (éxons)	IcircRNAs (íntrons)
SW480ribo	3.096	3.093	3
SW480circ	11.718	11.621	97
SW620ribo	4.322	4.317	5
SW620circ	14.727	14.672	55

Seguiu-se para a quantificação dos transcritos (lineares) e genes expressos em cada linhagem celular e preparação de biblioteca. Foram utilizados os algoritmos Kallisto e RSEM, que apresentam métodos de contagem distintos (Tabela 6): pseudo-alinhamento e agrupamento baseado no alinhamento contra o genoma, respectivamente. Na Tabela 7, podemos observar, que as bibliotecas construídas com o protocolo padrão foram mais eficientes para detectar os genes lineares, conforme o esperado. Além disso, detectou-se aproximadamente 3 vezes mais genes na análise pelo RSEM. Outro fato marcante é a maior variedade de transcritos presentes na construção padrão do Kallisto (Tabela 6).

Tabela 6: Comparação da contagem total dos transcritos (Kallisto) e genes (RSEM) identificados nas bibliotecas e linhagens celulares.

	SW480-ribo	SW480-circ	SW620-ribo	SW620-circ
Kallisto*	48.294	5.728	43.273	5.956

RSEM**	14.134	5.449	13.424	7.946
---------------	--------	-------	--------	-------

TPM > 0.5 como corte mínimo de expressão. * Kallisto: número de transcritos. ** RSEM: número de genes.

O próximo passo, foi buscar pelos genes com expressão diferencial pelo edgeR para o Kallisto (Tabela 7) e para o RSEM (Tabela 8).

Tabela 7: Total de genes diferencialmente expressos edgeR/Kallisto.

Comparação	UP	DOWN
SW480ribo x SW620ribo	1709	2054
SW480circ x SW620circ	1646	1852
SW480ribo x SW480circ	1402	2012
SW620ribo x SW620circ	686	131

Tabela 8: Total de genes diferencialmente expressos edgeR/RSEM.

Comparação	UP	DOWN
SW480ribo x SW620ribo	1623	2009
SW480circ x SW620circ	1723	1781

Nas Tabelas 7 e 8 pode-se notar que a expressão diferencial se mostra surpreendentemente bem similar entre os métodos Kallisto e RSEM. Assim, decidiu-se seguir com o Kallisto para a análise de expressão diferencial específica de circRNAs, onde foram utilizadas as coordenadas dos circRNAs detectados como referência para gerar o *index* da ferramenta.

Tabela 9: Total de genes diferencialmente expressos para circRNAs edgeR/Kallisto.

Comparação	<i>UP</i>	<i>DOWN</i>
SW480ribo x SW620ribo	2257	2078
SW480circ x SW620circ	1936	2055
SW480circ x SW480ribo	4319	2542
SW620circ x SW620ribo	3479	1849

Na Tabela 9, observa-se que o número de genes diferencialmente expressos se mantêm parecido entre as bibliotecas e linhagens celulares, exceto pela comparação SW620circ x SW620ribo, que apresentou maior número na análise de circRNAs, como era de se esperar devido a um número maior de posição cromossômica de detecção dos circRNAs nos genes. Para as análises de expressão diferencial, foi usada a ordem de diferença da linhagem SW480 sobre a SW620, assim em todos os resultados, a expressão *up-regulated* se refere aos

genes superexpressos na linhagem SW480 (tumor primário) e a expressão *down-regulated*, se refere aos genes subexpressos na SW480, o seja, superexpressos na linhagem SW620 (tumor metastático).

Prosseguiu-se com a comparação da expressão entre os preparos de biblioteca diferentes e foi feito um *Venn Diagram* para obter a lista de interseção dos genes em comum dos sequenciamentos (Figura 4).

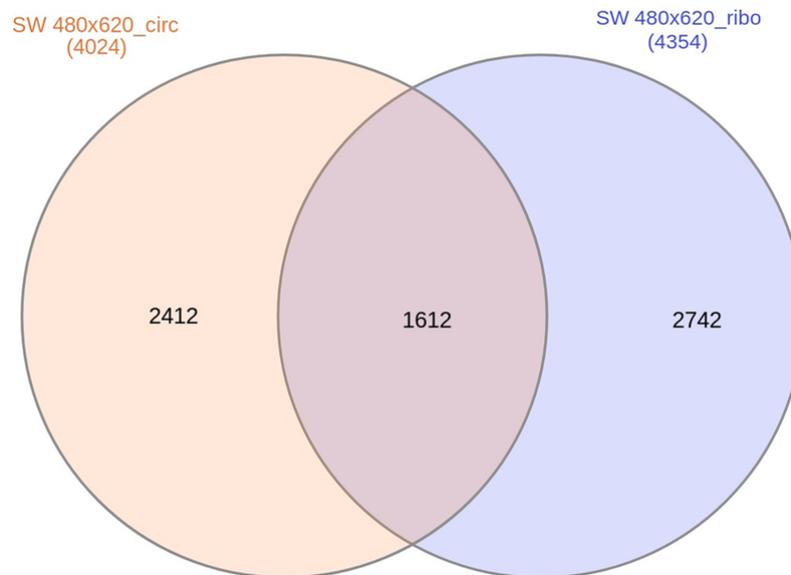


Figura 4: *Venn diagram* da comparação dos diferentes preparos de biblioteca no R.

Dos aproximadamente 4.100 genes de cada condição de preparo, apenas 1.612 circRNAs foram comuns em ambos os sequenciamentos. Com esses 1.612 genes em comum, foi feito um *Heatmap* (Figura 5) para avaliar a similaridade de expressão dos circRNAs em ambos os preparos.

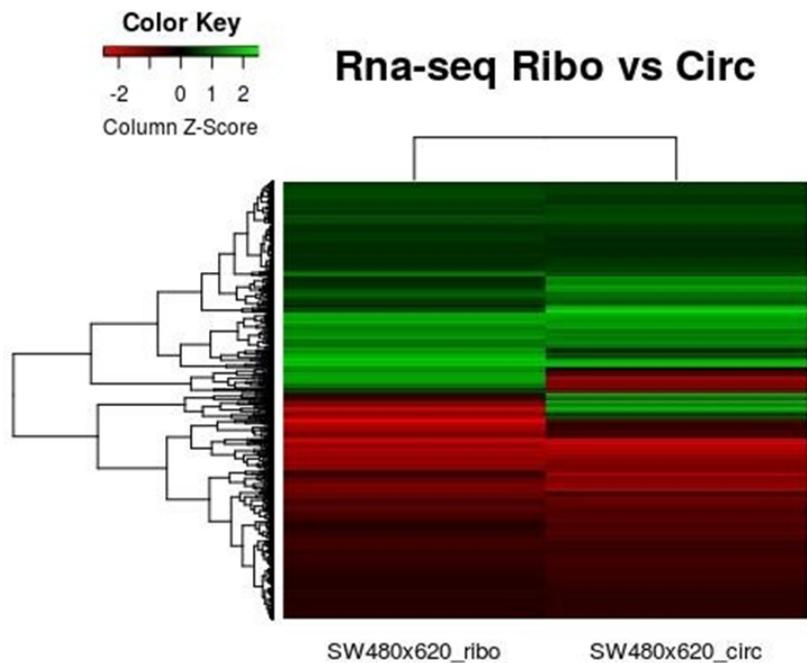


Figura 5: *Heatmap* da expressão diferencial dos 1612 circRNAs em comum nos sequenciamentos apresentados no R. Comparação da expressão diferencial entre os experimentos de *RNA-Seq* com preparos de biblioteca diferentes, padrão e para detecção dos circRNAs. Ordem de comparação entre as linhagens SW480 contra SW620. Filtro de *fold change* < -2 ou >2 e FDR <0.05.

De uma maneira geral, o nível de expressão diferencial entre os preparos de biblioteca foi similar (por volta de 4 mil genes), porém a concordância não foi grande, em torno de 40%. Além daqueles genes que só apresentaram expressão diferencial em uma das bibliotecas, também foram identificados alguns genes com expressão oposta, pois as coordenadas dos circRNAs analisados, em algumas vezes coincidem com a mesma linhagem para o mesmo gene em vias diferentes e expressão oposta. Além disso, tem o fato do preparo de biblioteca ser

diferente, levando a hipótese que um preparo comum pode mascarar a real expressão dos circRNAs quando comparado a um preparo específico para detecção destes.

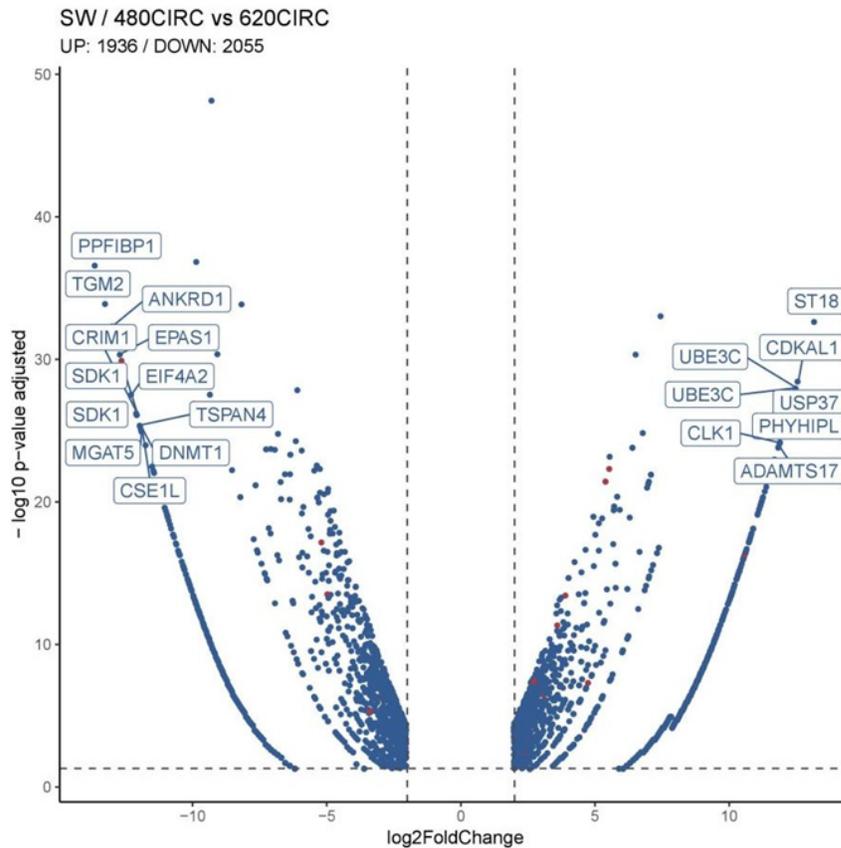


Figura 6: *Volcano plot* da comparação da expressão dos genes entre as linhagens celulares no R. Filtro de *fold change* < - 2 ou >2 e FDR <0.05.

Na figura 6, observa-se um grupo de genes *down-regulated*. Pode-se observar entre os maiores valores negativos de *fold change* que alguns circRNAs destacam-se, como o DNMT1 (*DNA Methyltransferase 1*). DNMT1 é codificador da proteína que é a principal enzima responsável pela manutenção dos padrões de

metilação após a replicação do DNA (OMIM 126375). Nos bancos de dados, ele é referenciado como o hsa_circ_28013 (Chen et al., 2016). Sua localização é apresentada no cromossomo 19, com a coordenada 10274001-10288043 no hg19 e 10163325-10177367 no hg38 (Song et al., 2016). O circ DNMT1 apresenta as coordenadas da linhagem celular SW620 (as quais estão disponíveis a partir da análise que o circExplorer resulta em seu output e é utilizado como referência na análise de expressão diferencial com os algoritmos utilizados e descritos na seção de materiais e métodos), deste modo, ele se encontra *up-regulated* nessa linhagem.

O TGM2 (*Transglutaminase 2*) codifica uma proteína que atua como monômero, além de ser induzida pelo ácido retinóico, parece estar envolvida na apoptose (OMIM 190196). O hsa referente a esse circRNA com as coordenadas 36775118-36779459 no hg19 e 38146716-38151057 no hg38 no cromossomo 20 não foi encontrado em nenhum depósito nos bancos de dados analisados. O circ TGM2 apresenta as coordenadas da linhagem celular SW480 na análise de expressão diferencial, portanto seguindo a ordem de comparação, ele está *down-regulated* na linhagem de tumor primário.

O TSPAN4 (*Transmembrane 4 Superfamily*), pertence à grande família de tetraspaninas que são frequentemente expressas na superfície celular em associação entre si e com outras moléculas, como as integrinas, e funcionam para regular a adesão, migração, proliferação e diferenciação celular (OMIM 602644). O circ TSPAN4 é referenciado como o hsa_circ_18995 (Chen et al., 2016). Sua coordenada é 847200-850367 no hg19 e 847200-850367 no hg38 no cromossomo

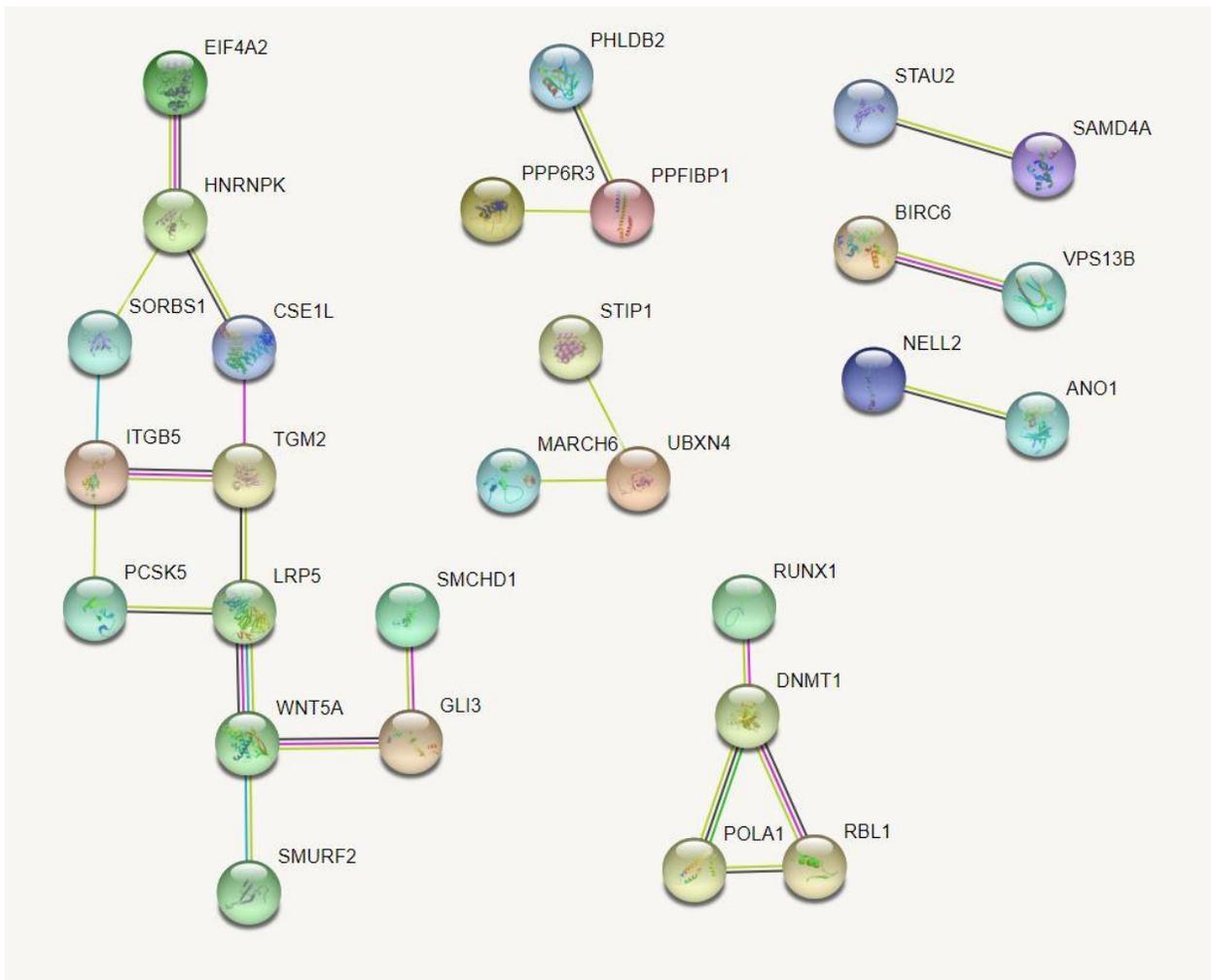
11 (Zhang et al., 2014). Ele está *down-regulated* na linhagem SW480 seguindo as coordenadas relatadas nas análises.

EPAS1 (*Endothelial Pas Domain Protein 1*), é um importante fator de transcrição que pode representar um importante regulador da vascularização, talvez envolvendo a regulação da expressão do gene da célula endotelial em resposta à hipóxia (OMIM 603349). O circ EPAS1 não é referenciado em nenhum banco de dados com as coordenadas para o cromossomo 2, 46583289-46588229 no hg19 e 46356150-46361090 no hg38. Seguindo as coordenadas do circRNA, ele está *down-regulated* na linhagem SW480.

Para os *up-regulated* destacam-se UBE3C (*Ubiquitin Protein Ligase E3c*) que pertence à família das enzimas E3. Estas aceitam transitoriamente a ubiquitina do E2 em uma ligação éster tiol antes de transferir a ubiquitina para a lisina alvo (OMIM 614454). Nos bancos de dados ele é referenciado como o hsa_circ_12017 (Chen et al., 2016). Está localizado no cromossomo 7, com a coordenada 156976571-156994501 no hg19 e 156976571-156994501 no hg38 (Jeck et al., 2013). O circRNA UBE3C apresenta a mesma coordenada em ambas as linhagens na expressão diferencial, assim ele está *up-regulated* na SW480 e *down-regulated* na SW620.

O circ ST18 (*Suppression Of Tumorigenicity 18*) pode ser observado com uma das expressões mais diferenciais (*Suppression Of Tumorigenicity 18*). ST18 é uma proteína *zinc finger* do tipo C2HC que se prevê ligar ao elemento de consenso principal AAAGTTT nas regiões promotoras dos genes-alvo e reprimir a transcrição gênica (OMIM 617155). ST18 é descrito nos bancos de dados

referenciado no cromossomo 8, como o hsa_circ_13413 (Chen et al., 2016). Sua coordenada genômica no hg19 é 53028834-53085143 ou no hg38 52116274-52172583 (Song et al., 2016). Sua expressão diferencial apresentada se refere a coordenada da linhagem celular SW620, assim este está *down-regulated* nesta linhagem, sendo o circRNA *top 1* da expressão diferencial quando analisado pelo critério *fold change*.



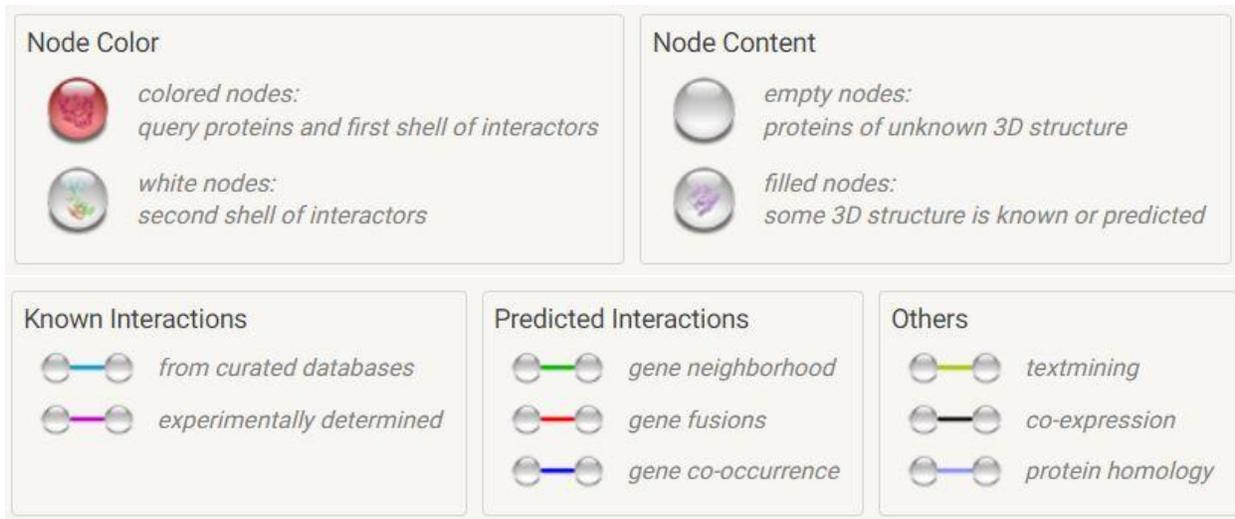


Figura 7: Rede PPI dos genes *down-regulated* no STRING. Filtro de *fold change* < -10 ou >10. Confidência média 0.4.

Na análise de interação de proteínas (redes *PPI*) feita pelo STRING para o set de genes *down-regulated* (Figura 7), observa-se a correlação com os resultados encontrados de expressão. Em uma média confidência, é clara a conexão da rede modulada por TGM2 sobre ITGB5 (Integrin, Beta-5), pertencente à família das integrinas que são glicoproteínas heterodiméricas que mediam interações célula-célula e matriz célula-extracelular (OMIM 147561). Antes, em 2019, essa via era mediada pelo TSPAN4, mas com novas evidências, a rede mudou e TGM2 antes sem interação com ITGB5, atualmente se liga fortemente. As coordenadas de ITGB5 são apresentadas em ambas as linhagens, em alguns circRNAs com expressão opostas.

Na mesma interação, observa-se WNT5A (*Wingless-Type Mmtv Integration Site Family, Member 5a*). Os WNTs compreendem uma grande classe de

proteínas secretadas que controlam processos essenciais de desenvolvimento, como padrões embrionários, crescimento celular, migração e diferenciação. A via de sinalização Wnt canônica bem conhecida envolve a ligação de proteínas Wnt a receptores *Frizzled* (por exemplo, FZD1; OMIM 603408) que induzem estabilização e entrada de beta-catenina (CTNNB1; OMIM 116806) no núcleo, onde afeta a transcrição de genes. O gene WNT5A codifica uma proteína Wnt envolvida nas vias de sinalização canônica e não-canônica, dependendo do contexto do receptor (OMIM 164975) (Mikels and Nusse, 2006). Também é possível observar a ação de circ GLI3 (*Gli-Kruppel Family Member 3*) com WNT5A. GLI3 codifica um fator de transcrição *zinc finger* que funciona na via de transdução de sinal *hedgehog* (SHH, 600725), que depende dos cílios primários em muitos tecidos. GLI3 sofre maturação pós-tradução em 2 isoformas com atividades transcricionais antagônicas. A sinalização de SHH causa o processamento proteolítico de GLI3 em um peptídeo, GLI3A, com função ativadora da transcrição. Na ausência de sinalização SHH, GLI3 é processado para um peptídeo mais curto, GLI3R, com função repressora transcricional. O processamento de GLI3 para qualquer uma das formas requer o cílio primário (OMIM 165240) (Laclef et al., 2015). Analisando as coordenadas, WNT5A e GLI3 são circRNAs diferencialmente expressos da linhagem SW480, eles estão *down-regulated* para esse tumor primário. O circ GLI3 é referenciado como hsa_circ_08687 (Chen et al., 2016). Suas coordenadas no cromossomo 7 são 42187824-42188067 para o hg19 e 42148225-42148468 para o hg38 (Jeck et al., 2013).

Além disso, observa-se a interação de circ TGM2 com circ LRP5 (*Low Density Lipoprotein Receptor-Related Protein 5*). LRP5 é membro da família de

receptores de lipoproteínas de baixa densidade (*LDLR*) que são proteínas da superfície celular que se ligam e internalizam ligantes no processo de endocitose mediada por receptores. (OMIM 603506). O circ LRP5 diferencialmente expresso, está associado às coordenadas da linhagem SW620, mas o circ PCSK5 *upstream* a LRP5, está expresso na linhagem SW480. Embora as coordenadas do circ LRP5 não coincidam com a via subexpressa na linhagem SW480, esta pode ter ligação através do GLI3 que liga as proteínas ITGB5, TGM2 e WNT5A quando a confiança da via é baixa.

Outro importante ligação com circ WNT5A é o circ SMURF2 (*Smad-Specific E3 Ubiquitin Protein Ligase 2*). A referência do circRNA para SMURF2 não foi descrito em nenhum dos bancos de dados analisados. SMURF2 se apresenta como um dos genes mais diferencialmente expressos na linhagem SW480.

Colaborando com os resultados encontrados na análise de expressão diferencial, foi observada a presença de DNMT1 na rede de interação, superexpresso na linhagem SW620, *upstream*, pode-se observar o fator de transcrição RUNX1 (*Runt-Related Transcription Factor 1*) (OMIM 151385). A coordenada do circ RUNX com expressão diferencial analisada na via *down-regulated* não foi encontrada em nenhum banco de dados, apenas o *start* dela, mas o final foi diferente, tratando-se de um circRNA muito maior que os depositados.

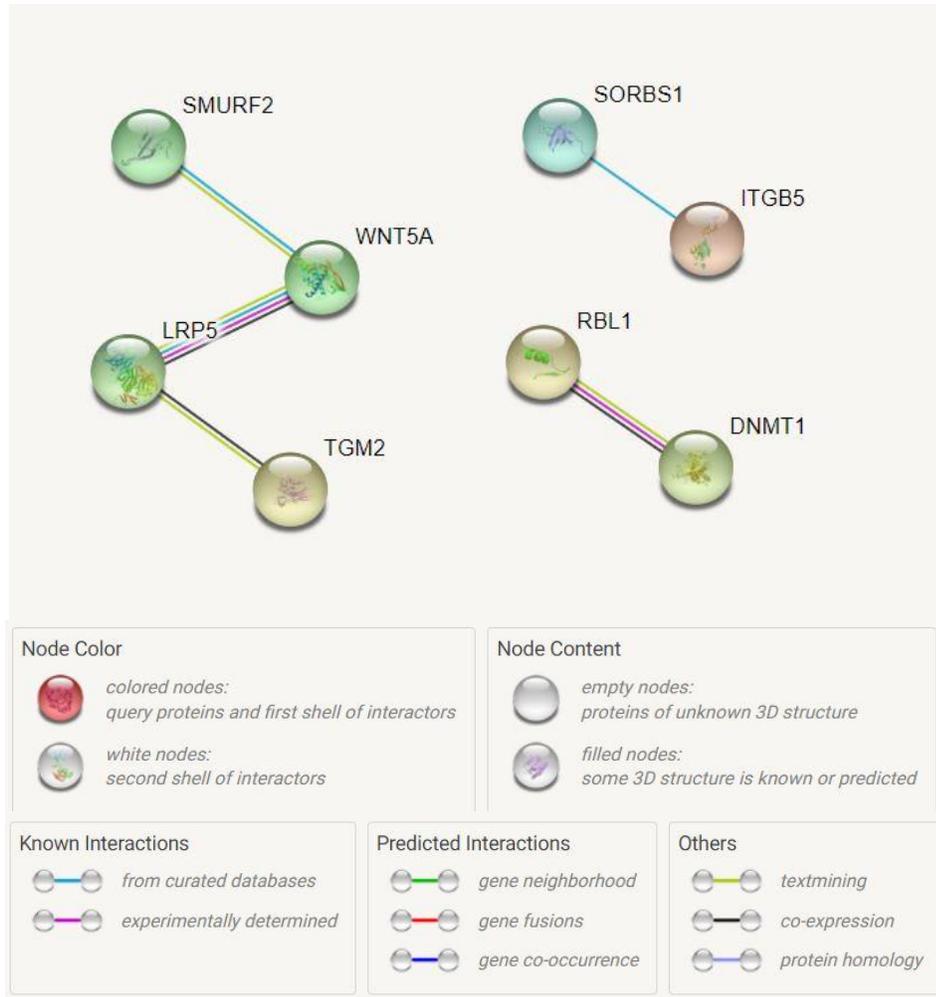


Figura 8: Rede PPI dos genes *down-regulated* no STRING. Filtro de fold change < -10 ou >10. Confiabilidade de 0.7.

Na análise de rede com alta confiabilidade (0.7), as vias relacionadas a WNT5A e DNMT1 se mantiveram, levando-se a considerar que essas vias são robustas e não apenas um ruído de análise (Figura 8).

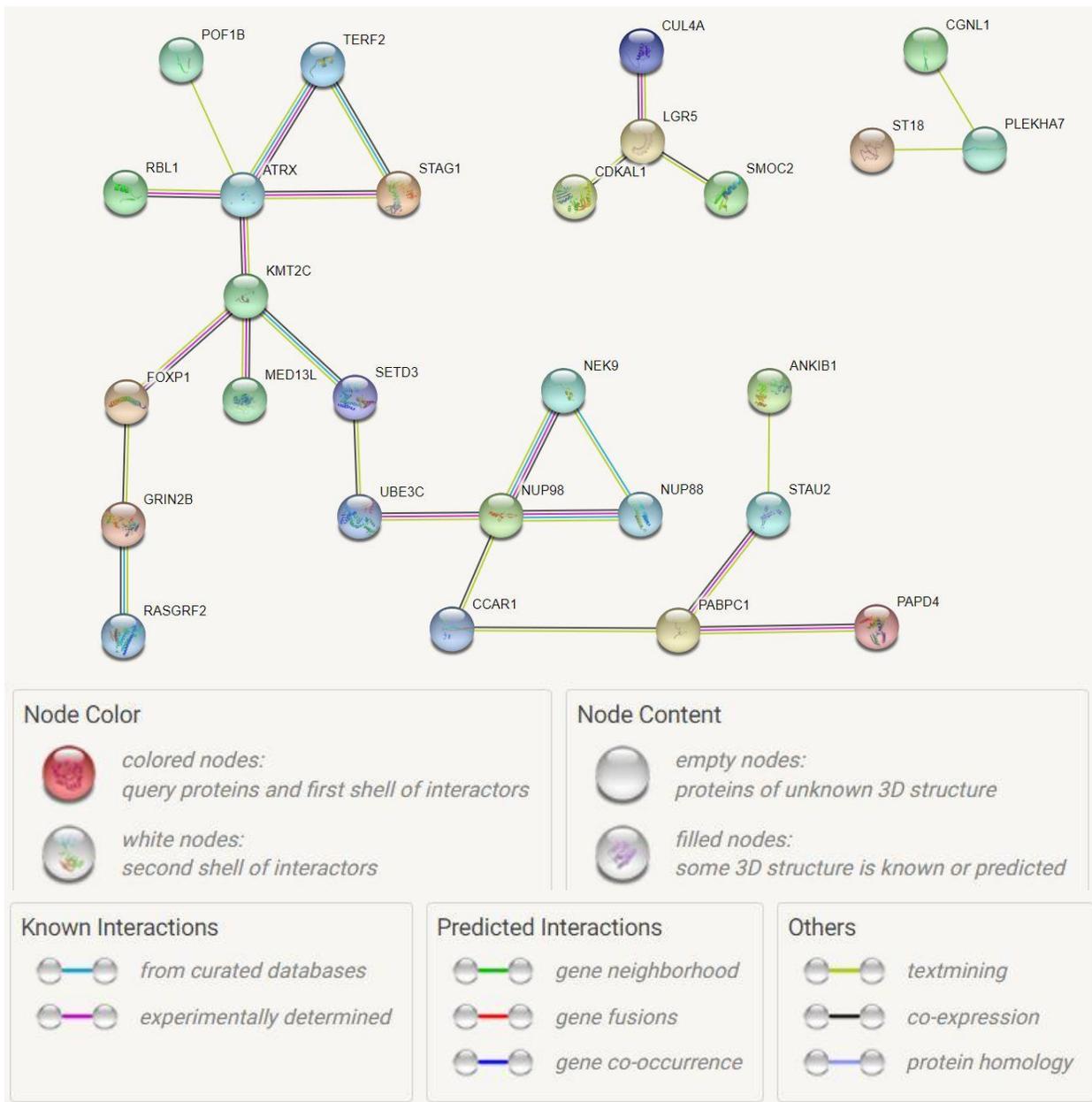


Figura 9: Rede PPI dos genes *up-regulated* no STRING. Filtro de fold change < -10 ou >10. Confidência média 0.4.

Na rede *up-regulated* do STRING (Figura 9), confirma-se a presença de UBE3C citado na expressão diferencial. O circ UBE3C como apresentado anteriormente, desenvolve dois papéis, cada um em uma via e em uma linhagem diferente. O UBE3C com expressão diferencial *up-regulated*, apresenta as mesmas coordenadas na linhagem SW480 e na linhagem SW620. É interessante visualizar na Figura 9, como ele desempenha o papel na via *up-regulated* na SW480 através da ligação com NUP98, e como ele desempenha função na via *down-regulated* na SW620 através da ligação com SETD3 (hsa_circ_13006 (Chen et al, 2016), localizado no cromossomo 14, coordenadas 99924615-99929915 no hg19 e 99458278-99463578 no hg38 (Jeck et al., 2013)), que se liga a KMT2C e ATRX (hsa_circ_0140609 (Dudekula et al., 2016), localizado no cromossomo X, 76907603-76920267 no hg19 e 77652113-77664778 no hg38), ambos com coordenadas correspondentes a SW620, portanto *down-regulated* no tumor metastático.

Nota-se a presença de mais 2 vias, a primeira com CCAR1 (*Cell Division Cycle And Apoptosis Regulator 1*) (OMIM 612569). O CD437 é um adamantil-retinóide seletivo ao RAR-gama (RARG; 180190) que causa a parada do ciclo celular e apoptose em vários tipos de células.

A segunda via é a interação de FOXP1 (*Forkhead Box P1*). O circ FOXP1 está associado às coordenadas da SW620, linhagem metastática, e está *down-regulated* nessa linhagem. O circ FOXP1 é referenciado como o hsa_circ_11346 (Chen et al, 2016). Suas coordenadas no cromossomo 3 são 71090478-71102924 no hg19 e chr3:71041327-71053773 no hg38 (Jeck et al, 2013).

Outra presença marcante, é o gene KMT2C (Lysine-Specific Methyltransferase 2c). O gene KMT2C codifica uma histona metiltransferase que regula a transcrição genética, modificando a estrutura da cromatina (OMIM 606833). Assim, como há o circRNA de DNMT1 envolvido em modulações epigenéticas na linhagem SW620 nas vias *up-regulated*, temos circRNA KMT2C relacionado aos eventos epigenéticos nessa linhagem nas vias *down-regulated*.

O ATRX (*ATRX Chromatin Remodeler*) faz parte da via que se forma acima junto ao KMT2C e FOXP1. Nos bancos de dados, ele é referenciado como o hsa_circ_0140609 (Dudekula et al., 2016). Suas coordenadas são da linhagem SW620 (cromossomo X, 76907603-76920267 no hg19 e 77652113-77664778 no hg38), portanto este está *down-regulated* no tumor metastático.

Nota-se o ST18, novamente *down-regulated* na SW620 em uma via isolada, com menos evidências que a via epigenética comentada acima.

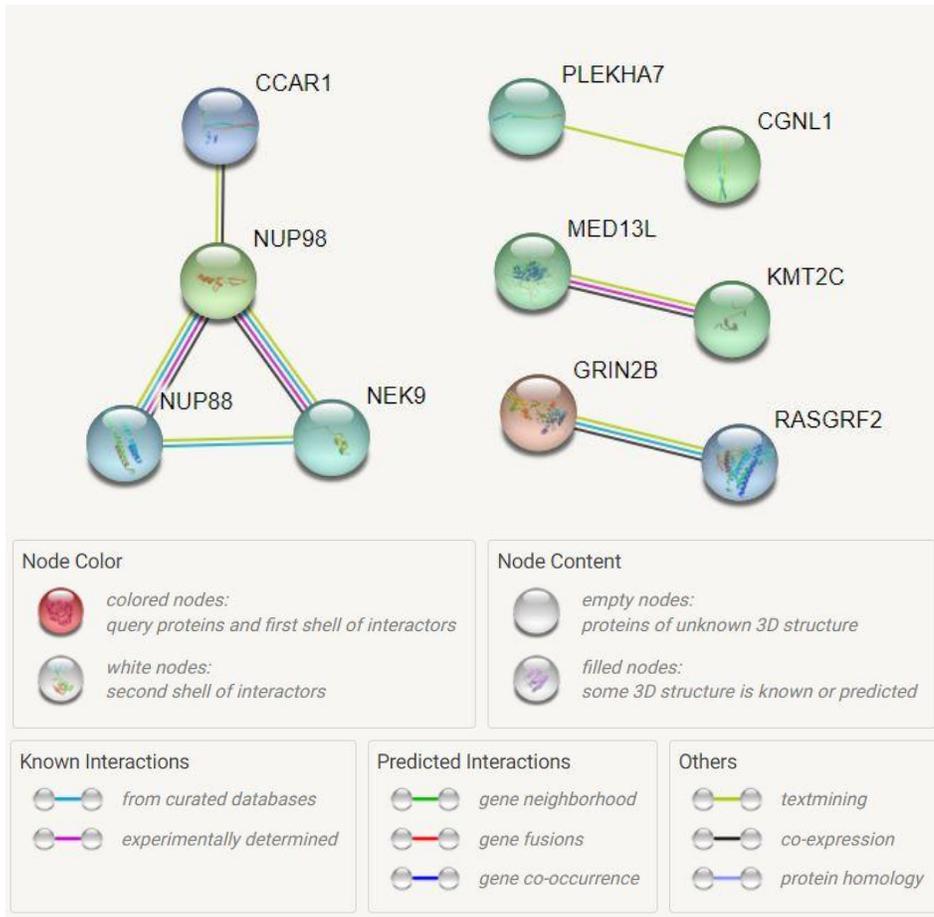


Figura 10: Rede PPI dos genes *up-regulated* no STRING. Filtro de fold change < -10 ou >10. Confiabilidade de 0.7.

Em uma análise de rede com alta confiabilidade (0.7), a via relacionada a CCAR1 e KMT2C se manteve, levando a conclusão de que ambas as vias são robustas (Figura 10).

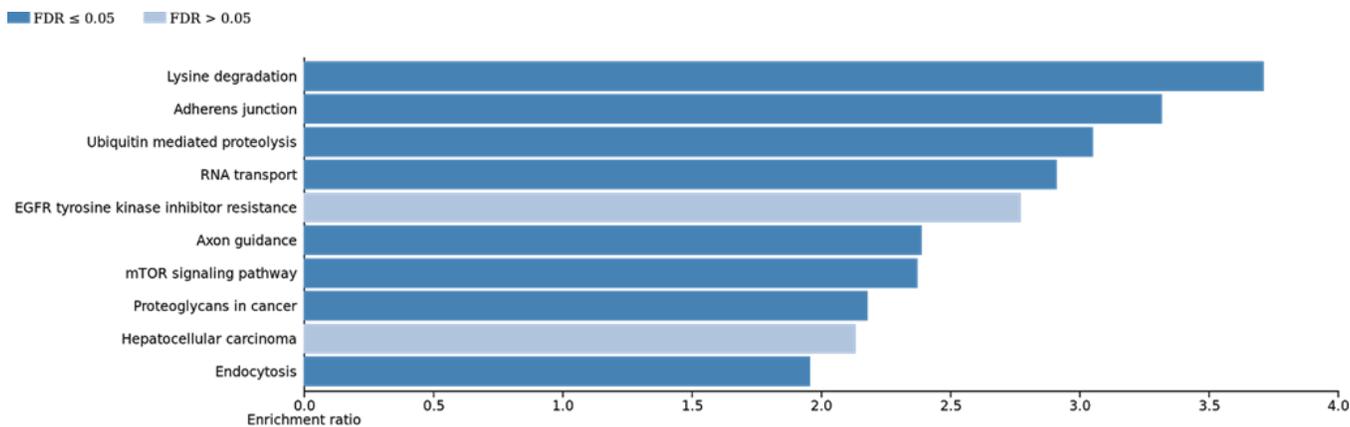
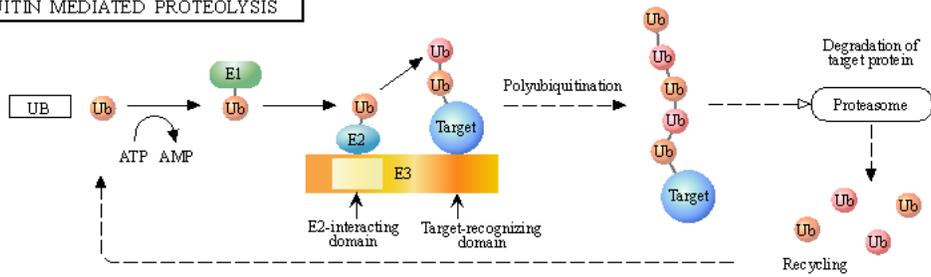


Figura 11: Vias KEGG *up-regulated* no WebGestalt.

Foi feita uma pré-análise das vias envolvidas na progressão tumoral das linhagens SW480 e SW620 (Figura 11). Conforme observado nas análises anteriores, apresentam-se vias *up-regulated* relacionadas a degradação da lisina (KMT2C e SET3D, igual na relação PPI), proteólise mediada por ubiquitinação (SMURF2 e UBE3C), endocitose (SMURF2) entre outras vias interessantes mediadas por outros genes importantes para tumorigênese como PTEN, BRAF E AKT3 (via da resistência ao inibidor de tirosina quinase de EGFR).

UBIQUITIN MEDIATED PROTEOLYSIS



E1
(Ubiquitin-activating enzyme)

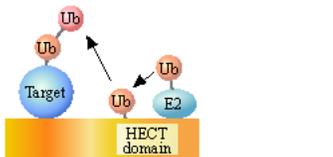
UBE1	UBLE1A	UBLE1B	UBE1C
------	--------	---------------	-------

E2
(Ubiquitin-conjugating enzyme)

UBE2A	UBE2B	UBE2C	UBE2D	UBE2E	UBE2F	UBE2G1	UBE2G2	UBE2H
UBE2I	UBE2J1	UBE2J2	UBE2L3	UBE2L6	UBE2M	UBE2N	UBE2O	
UBE2Q	UBE2R	UBE2S	UBE2U	UBE2W	UBE2Z	HIP2	UPL1CN	

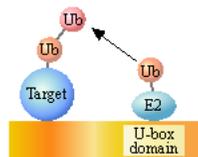
E3
(Ubiquitin ligase)

HECT type E3



E6AP	UBE3B	UBE3C	<i>Smurf</i>	<i>Itch</i>
WWP1	WWP2	TRIP12	NEDD4	ARF-BP1
EDD1	HERC1	HERC2	HERC3	HERC4

U-box type E3



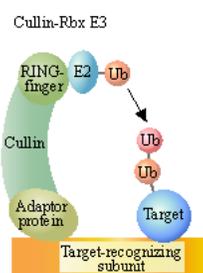
UBE4A	UBE4B	CHIP
CYC4	PRP19	UIP5

single RING-finger type E3

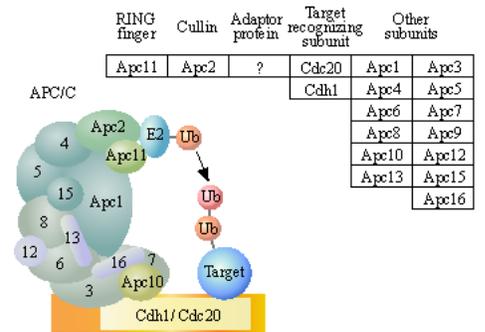


Mdm2	CBL	Parkin	SIAH-1	PML	TRAF6	MEKK1
COPI	PIRH2	cIAPs	PIAS	SYVN	NHLRC1	AIRE
MGRN1	BRCA1	FANCL	MID1	Trim32	Trma37	

multi subunit RING-finger type E3



	RING finger	Cullin	Adaptor protein	Target recognizing subunit
SCF complex	RBX1	Cul1	Skp1	F-box
BCV complex	RBX1	Cul2	EloB EloC	VHLbox
Cul3 complex	RBX1	Cul3		BTB
Cul4 complex	RBX1	Cul4	DDB1	DCAF
ECS complex	RBX2	Cul5	EloB EloC	SOC3box
Cul7 complex	RBX1	Cul7	Skp1	Fbxw8



RING finger	Cullin	Adaptor protein	Target recognizing subunit	Other subunits
Apc11	Apc2	?	Cdc20	Apc1 Apc3
			Cdh1	Apc4 Apc5
				Apc6 Apc7
				Apc8 Apc9
				Apc10 Apc12
				Apc13 Apc15
				Apc16

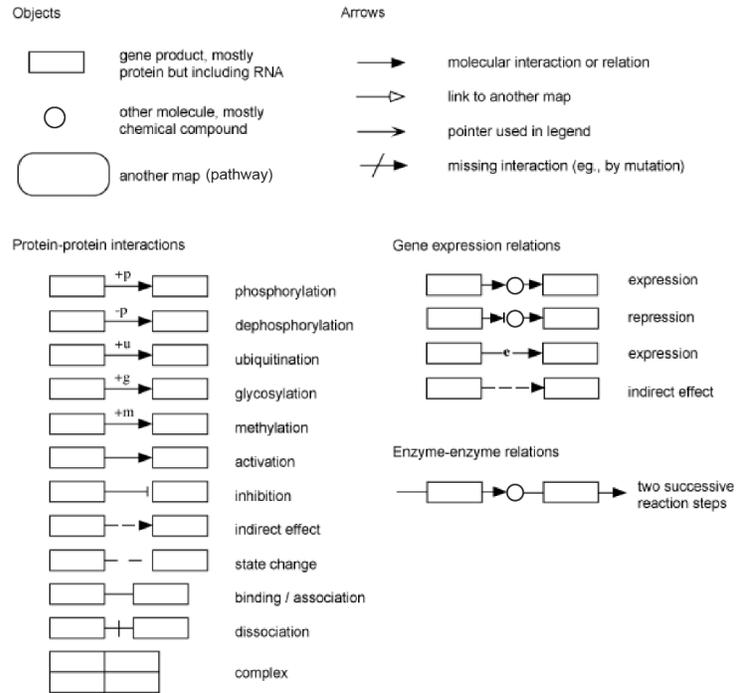


Figura 12: Via KEGG *up-regulated* “*Proteólise mediada por Ubiquitinação*” no ShinyGO com pacote Pathview (com legenda dos ícones da imagem gerada acima, usada para todas as imagens a seguir que utilizaram o mesmo *software* e pacote).

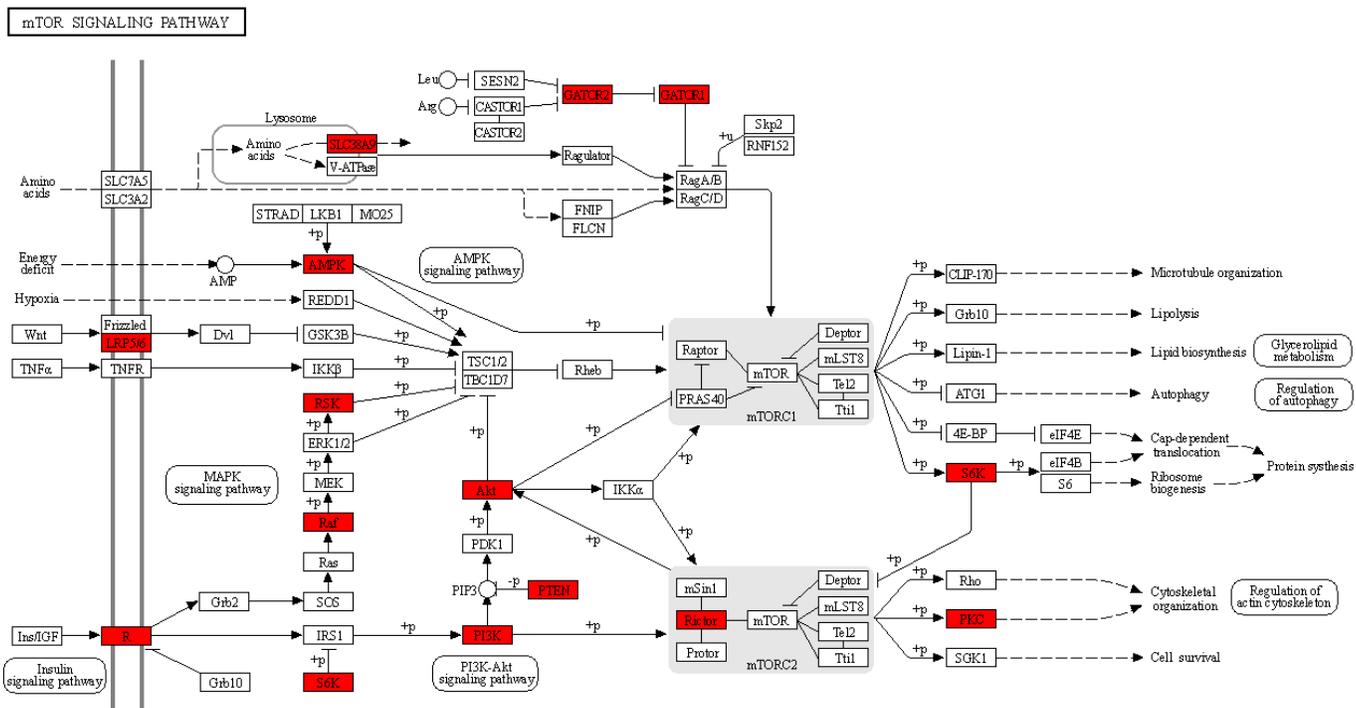


Figura 13: Via KEGG up-regulated “Via de sinalização do mTOR” no ShinyGO com pacote Pathview.

EGFR TYROSINE KINASE INHIBITOR RESISTANCE

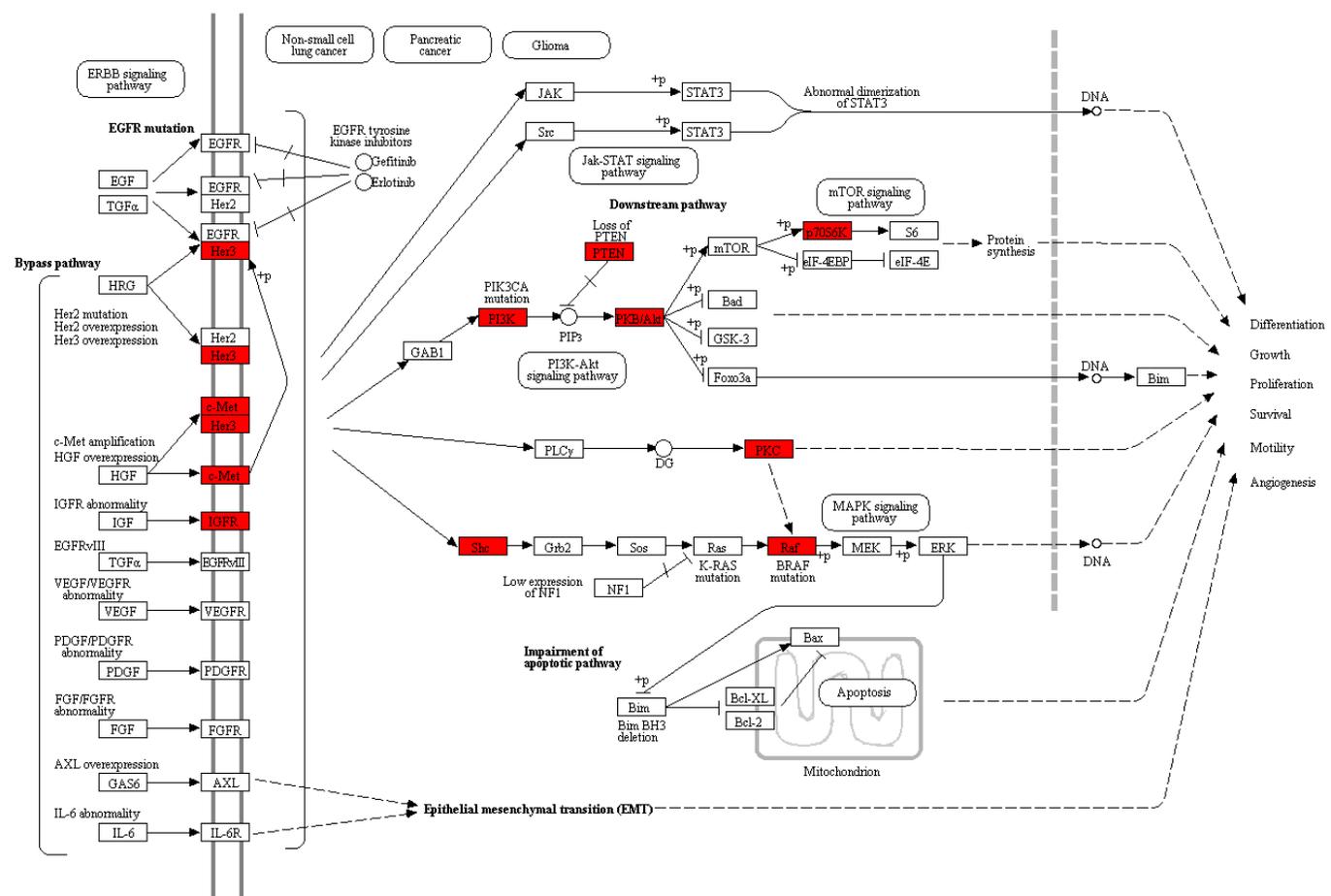


Figura 14: Via KEGG up-regulated "Via da resistência ao inibidor de tirosina quinase de EGFR" no ShinyGO com pacote Pathview.

Para verificar os detalhes de como os circRNAs estavam em cada via KEGG, foi feita uma análise mais focada em cada via já anotada no WebGestalt no ShinyGO com o pacote Pathview (Figuras 12, 13 e 14).

RIBOSOME BIOGENESIS IN EUKARYOTES

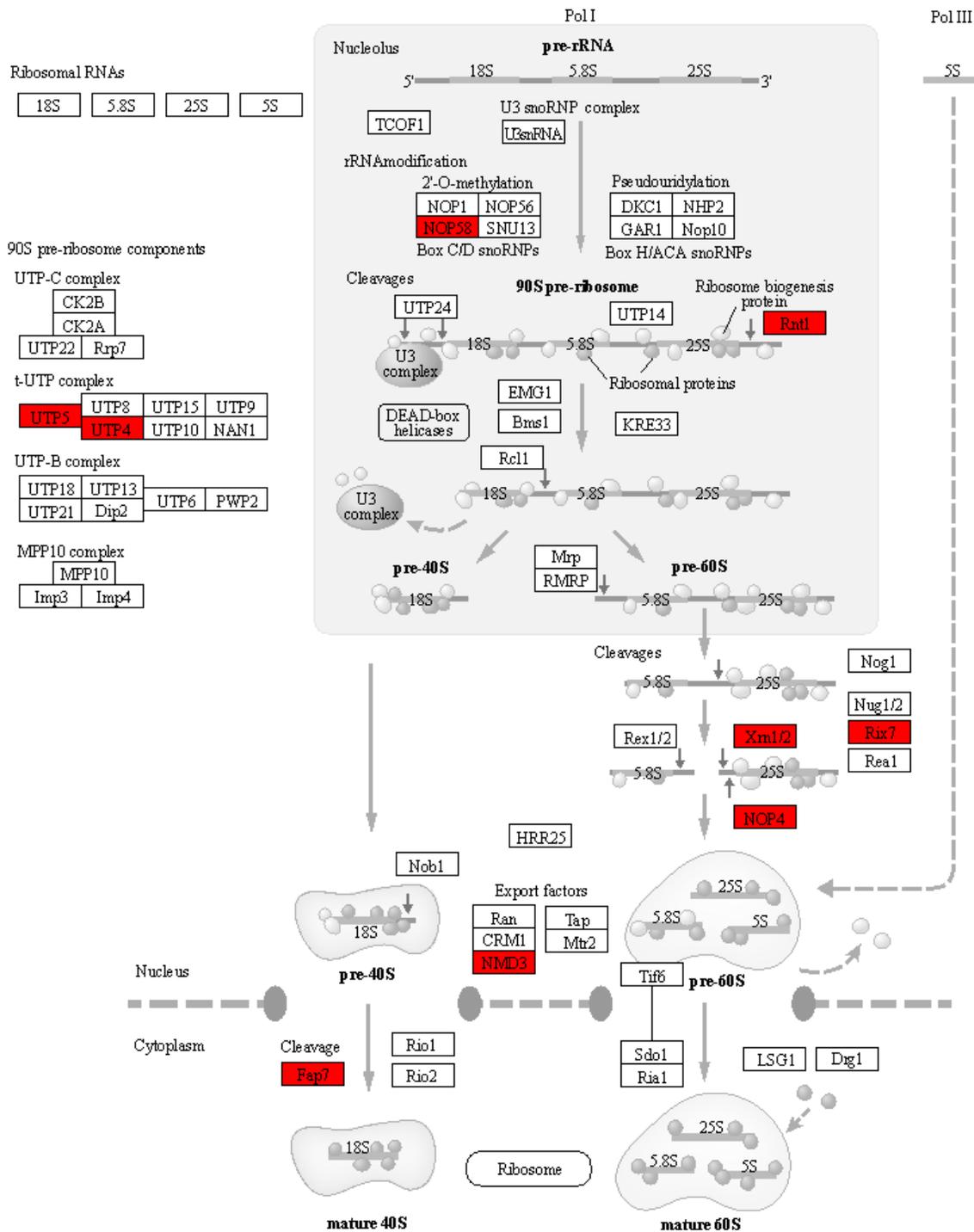


Figura 15: Via KEGG *up-regulated* “Biogênese do ribossomo nos eucariotos” no ShinyGO com pacote Pathview.

RNA DEGRADATION

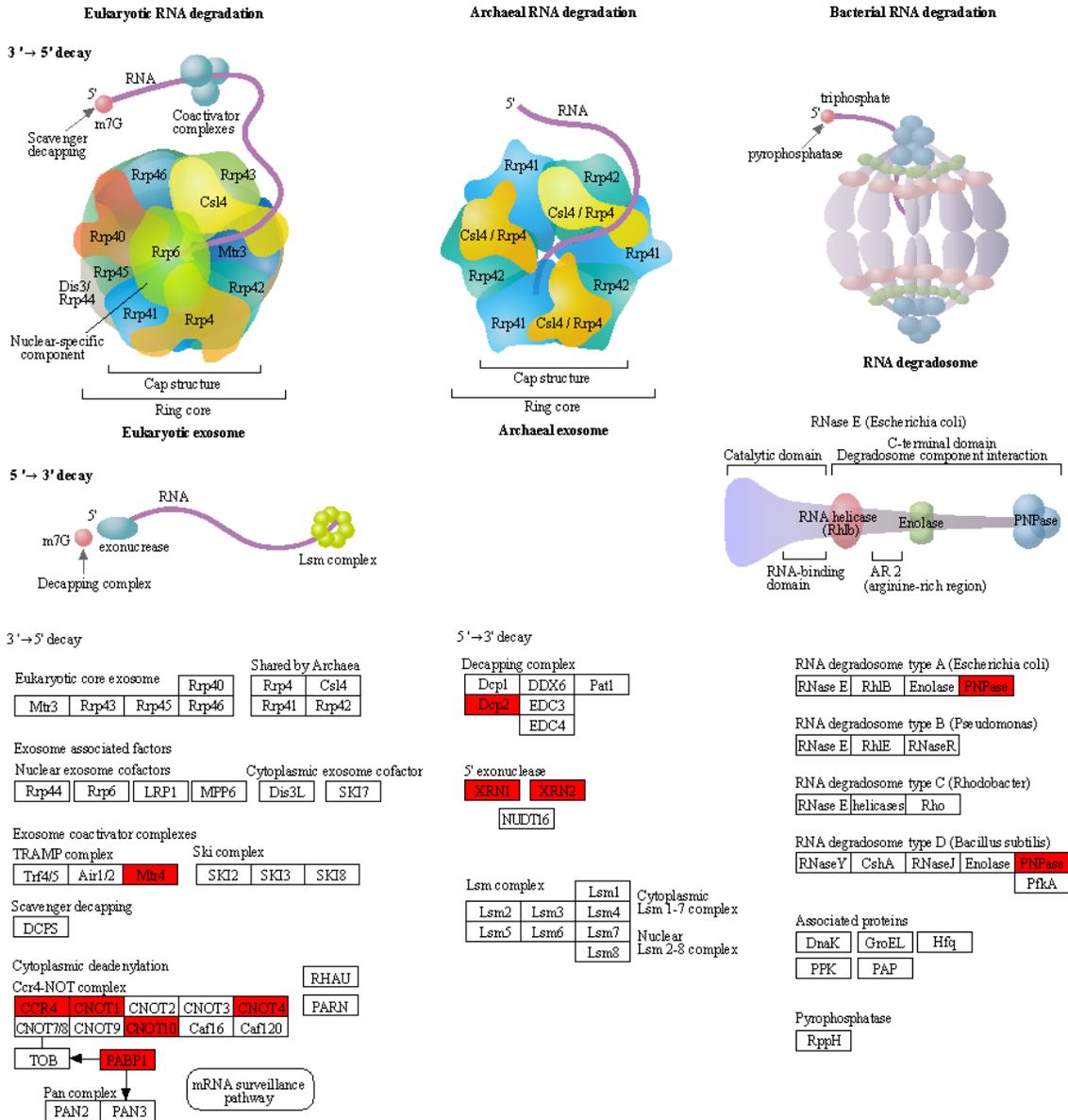


Figura 16: Via KEGG up-regulated "Degradação do RNA" no ShinyGO com pacote Pathview.

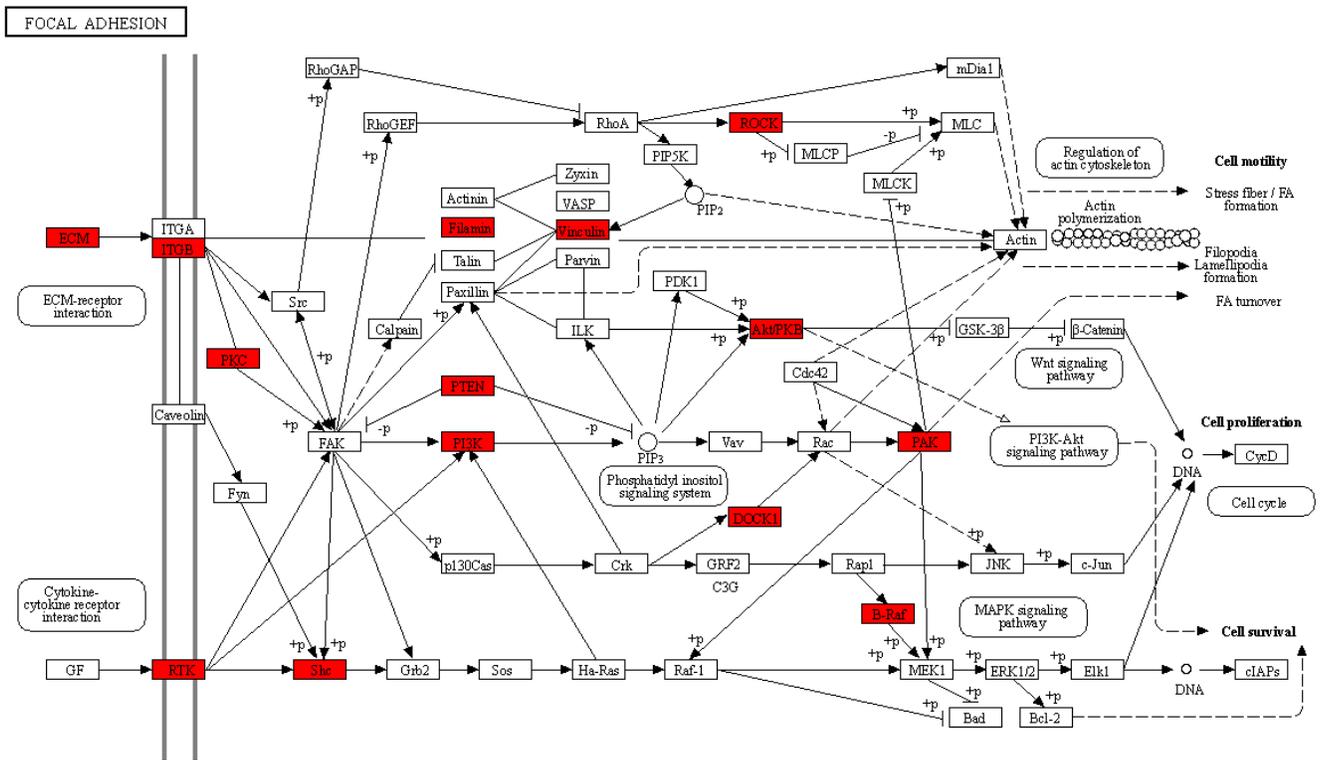


Figura 17: Via KEGG up-regulated "Adesão focal" no ShinyGO com pacote Pathview.

NUCLEOCYTOPLASMIC TRANSPORT

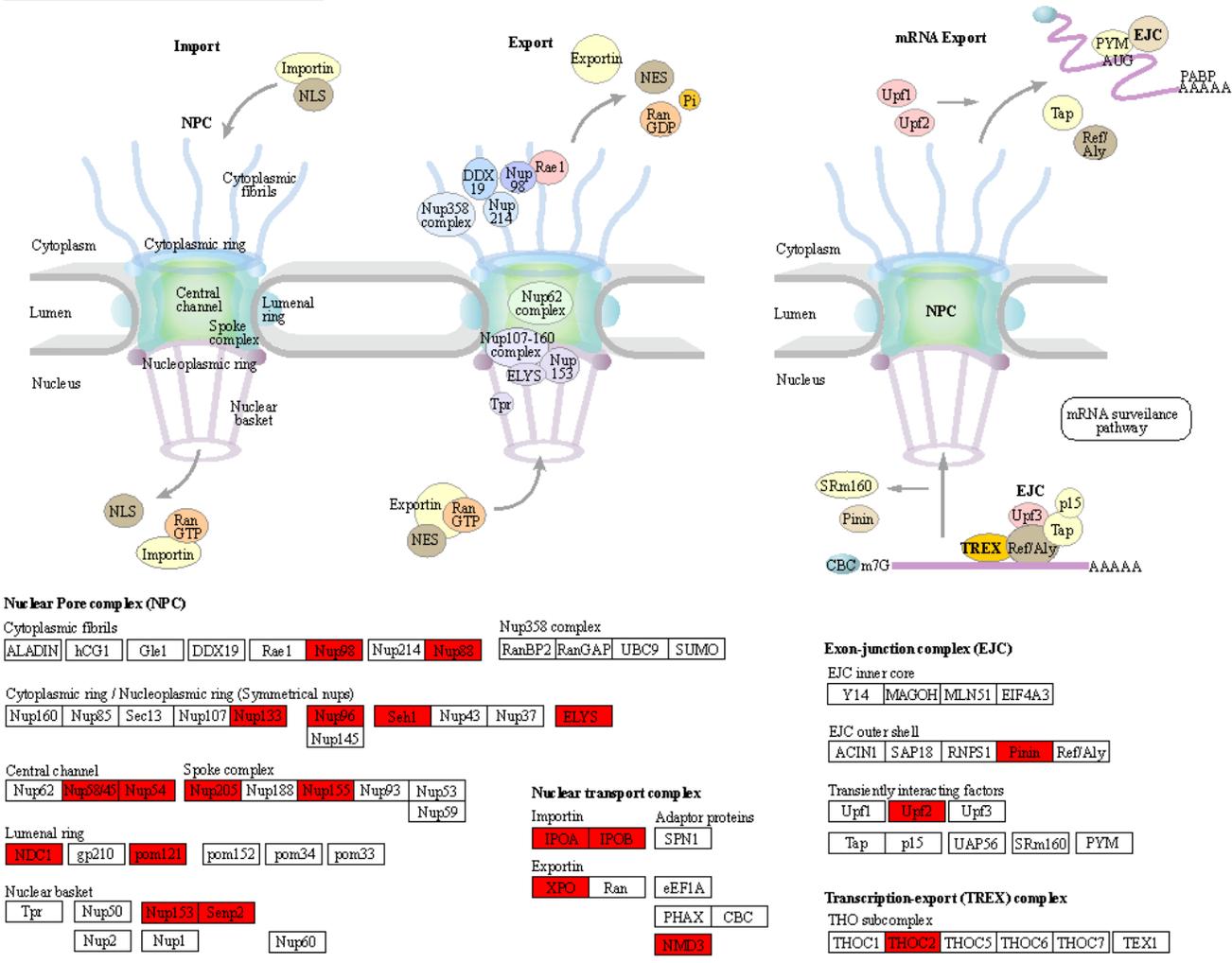


Figura 18: Via KEGG *up-regulated* “Transporte Núcleo-citoplasma” no ShinyGO com pacote Pathview.

É interessante visualizar que além das vias encontradas no software do WebGestalt, o ShinyGO também enriqueceu outras vias (Figuras 15, 16, 17 e 18).

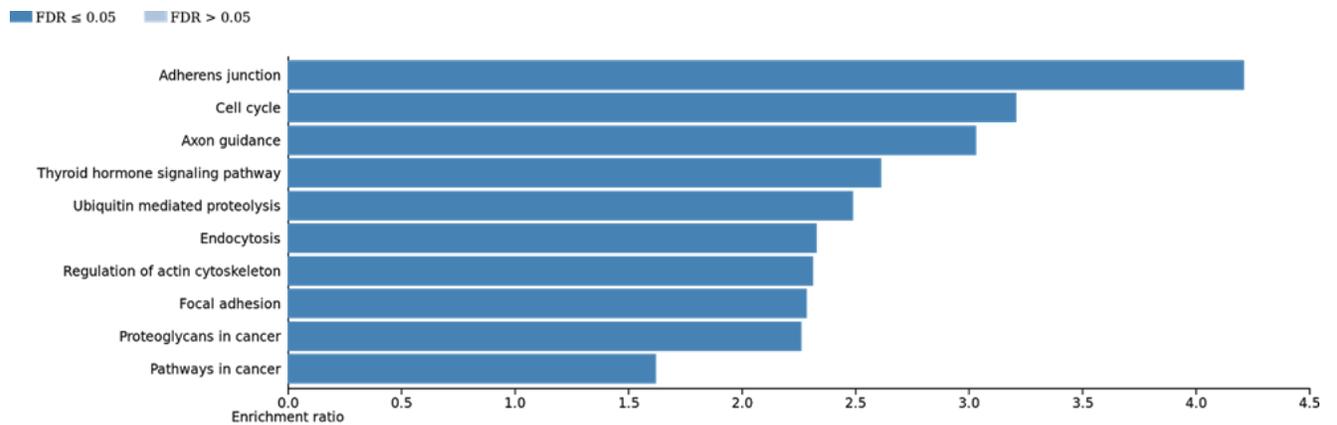


Figura 19: Vias KEGG *down-regulated* no WebGestalt.

Já nas vias *down-regulated* (Figura 19), observa-se a de “Axon guidance” (WNT5A), endocitose (SMURF2), regulação da actina no citoesqueleto e adesão focal (ITGB5), via dos proteoglicanos em câncer (WNT5A e ITGB5), vias de câncer (GLI3, RUNX1, WNT5A) e outras vias interessantes como proteólise mediada por ubiquitinação (BRCA1, UBE3C e SMURF2).

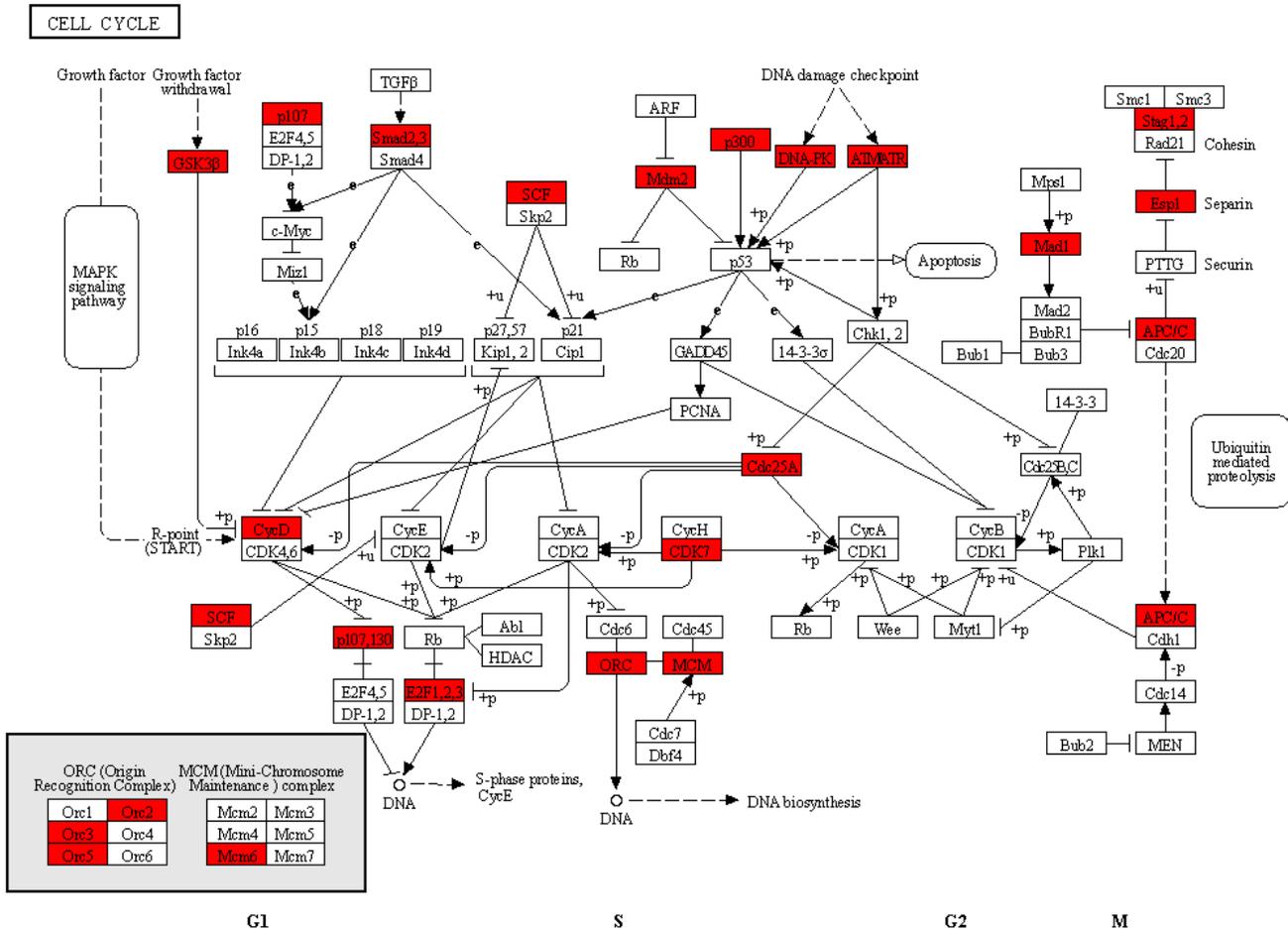


Figura 20: Via KEGG down-regulated "Ciclo celular" no ShinyGO com pacote Pathview.

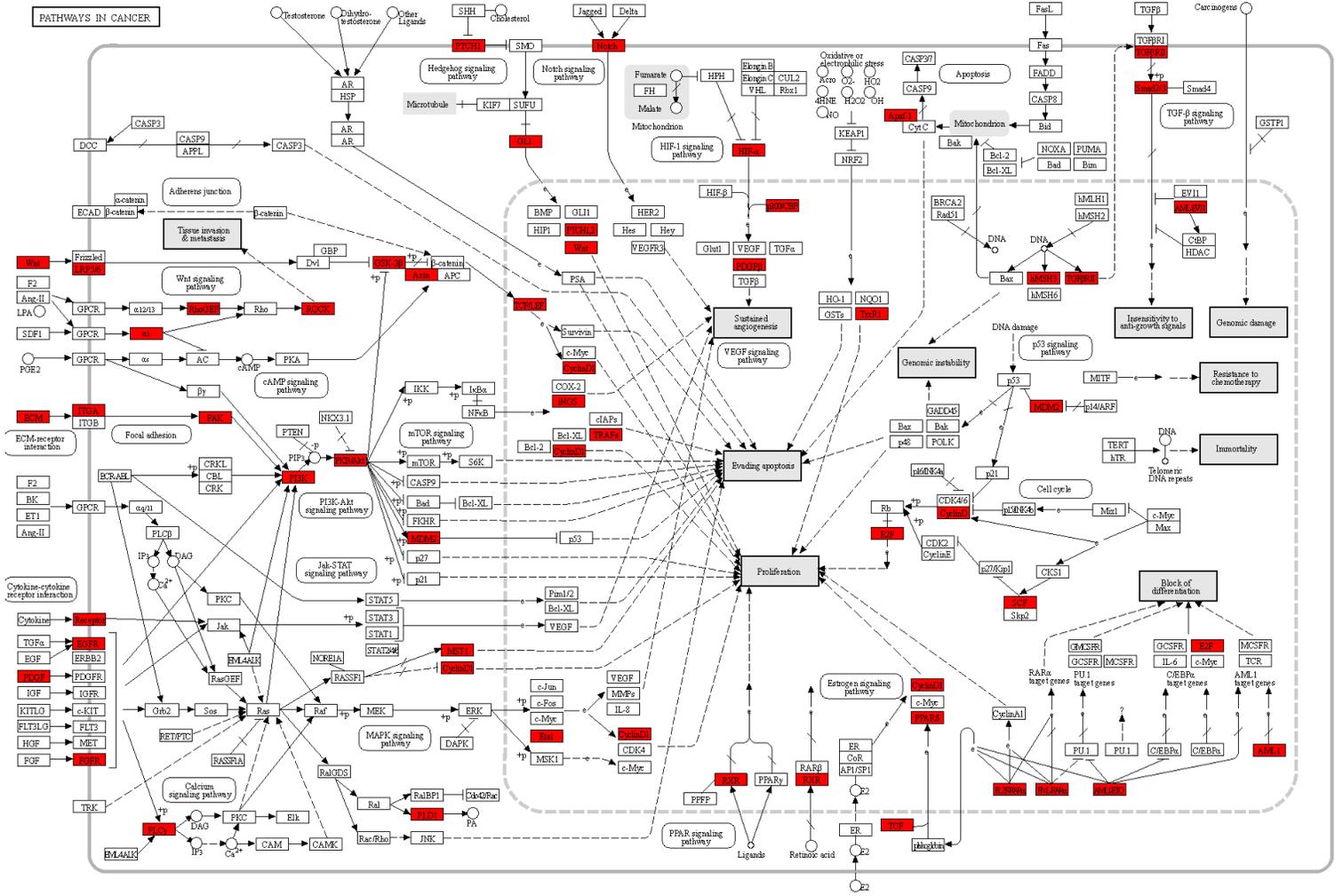


Figura 21: Via KEGG down-regulated “Vias do câncer” no ShinyGO com pacote Pathview.

UBIQUITIN MEDIATED PROTEOLYSIS

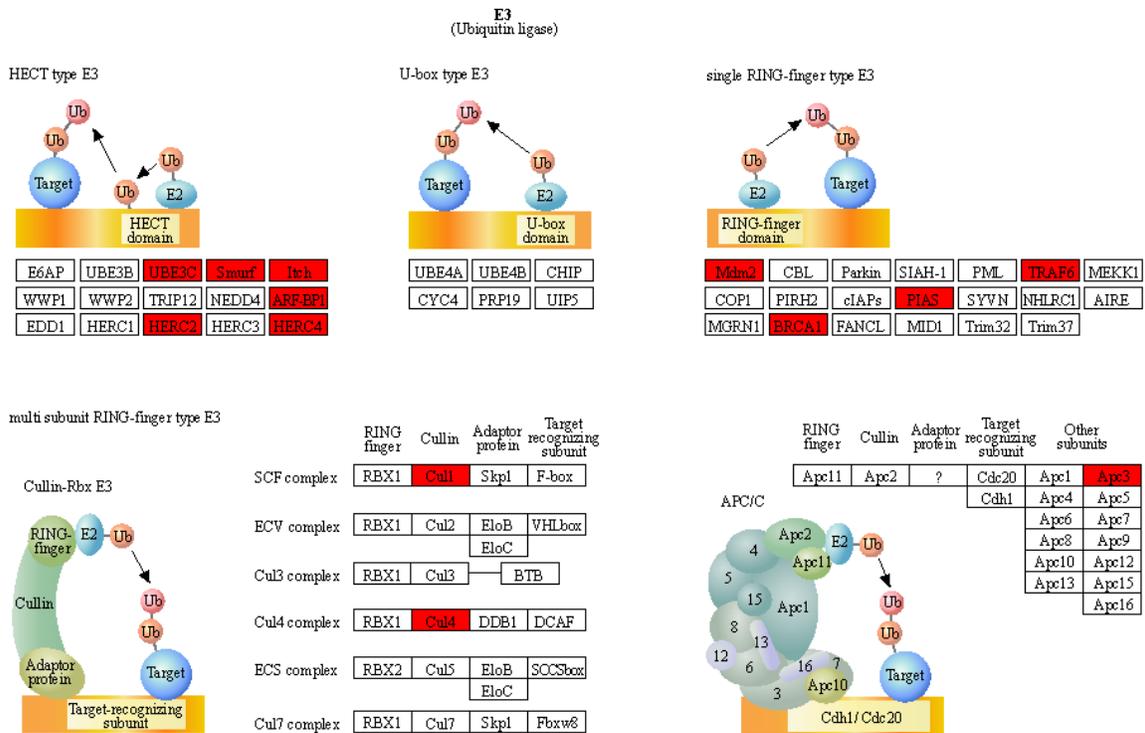
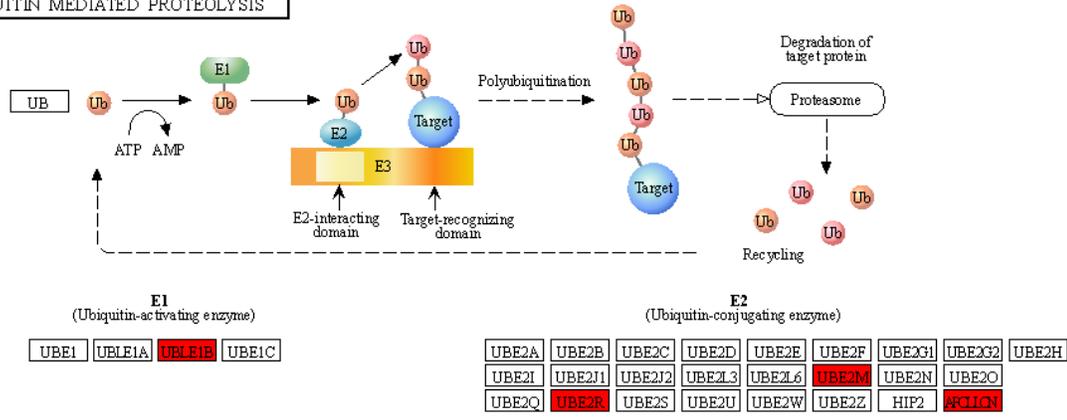


Figura 22: Via KEGG *down-regulated* “Proteólise mediada por ubiquitinação” no ShinyGO com pacote Pathview.

Para verificar os detalhes de como os circRNAs estavam em cada via KEGG *down-regulated*, foi feita uma análise mais focada em cada via já anotada no WebGestalt no ShinyGO com o pacote Pathview (Figuras 20, 21 e 22).

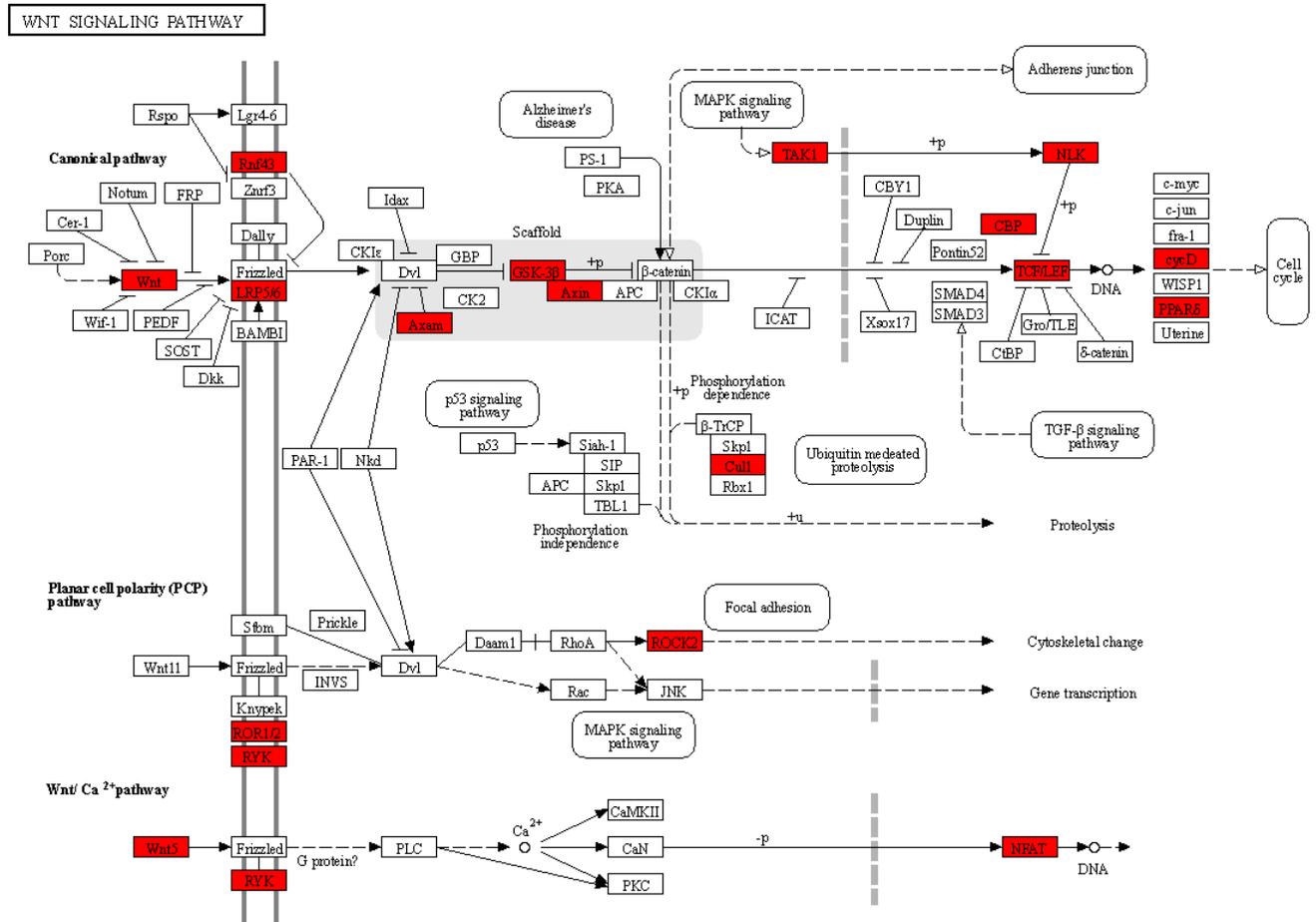


Figura 23: Via KEGG *down-regulated* “Via de sinalização WNT” no ShinyGO com pacote Pathview.

HEDGEHOG SIGNALING PATHWAY

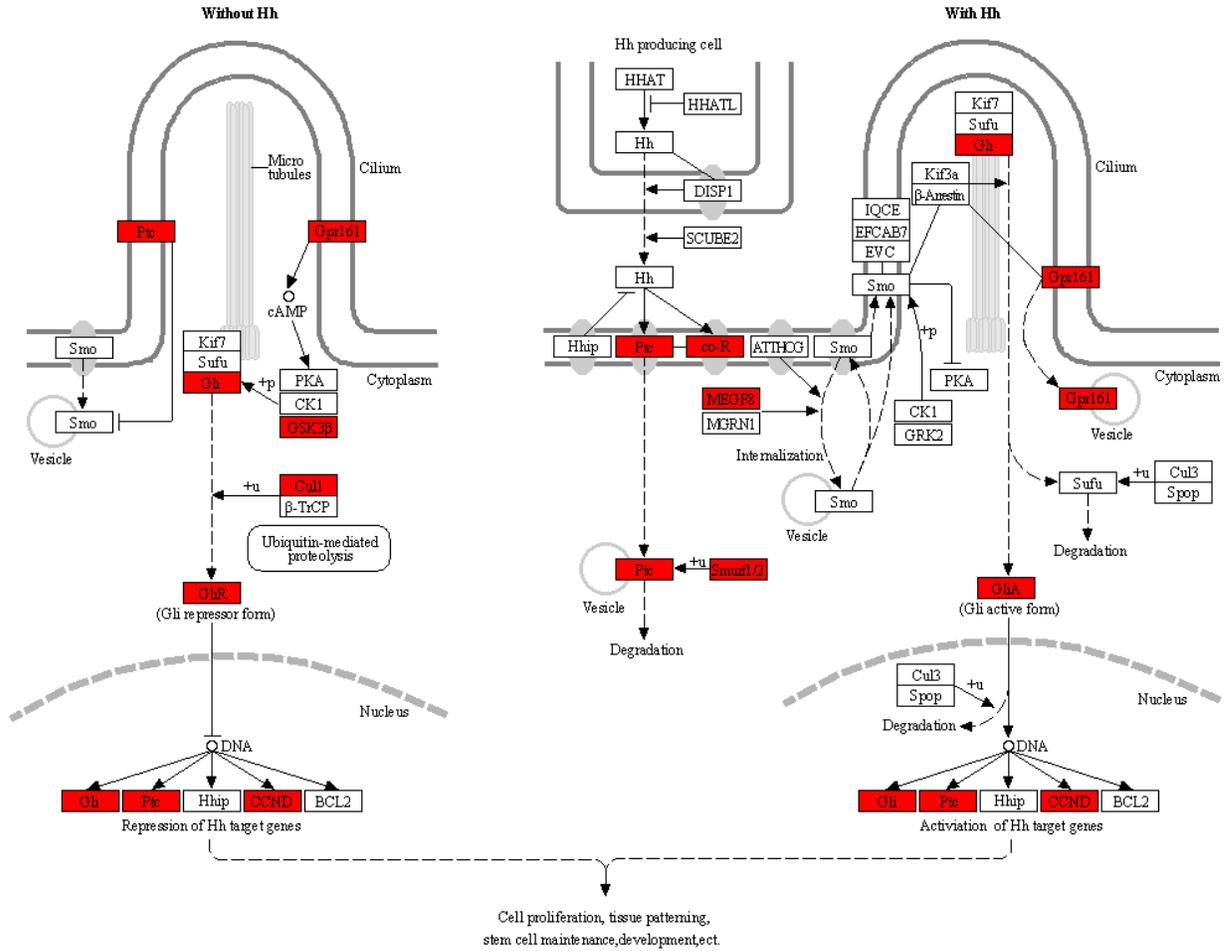


Figura 24: Via KEGG *down-regulated* "Via de sinalização Hedgehog" no ShinyGO com pacote Pathview.

EGFR TYROSINE KINASE INHIBITOR RESISTANCE

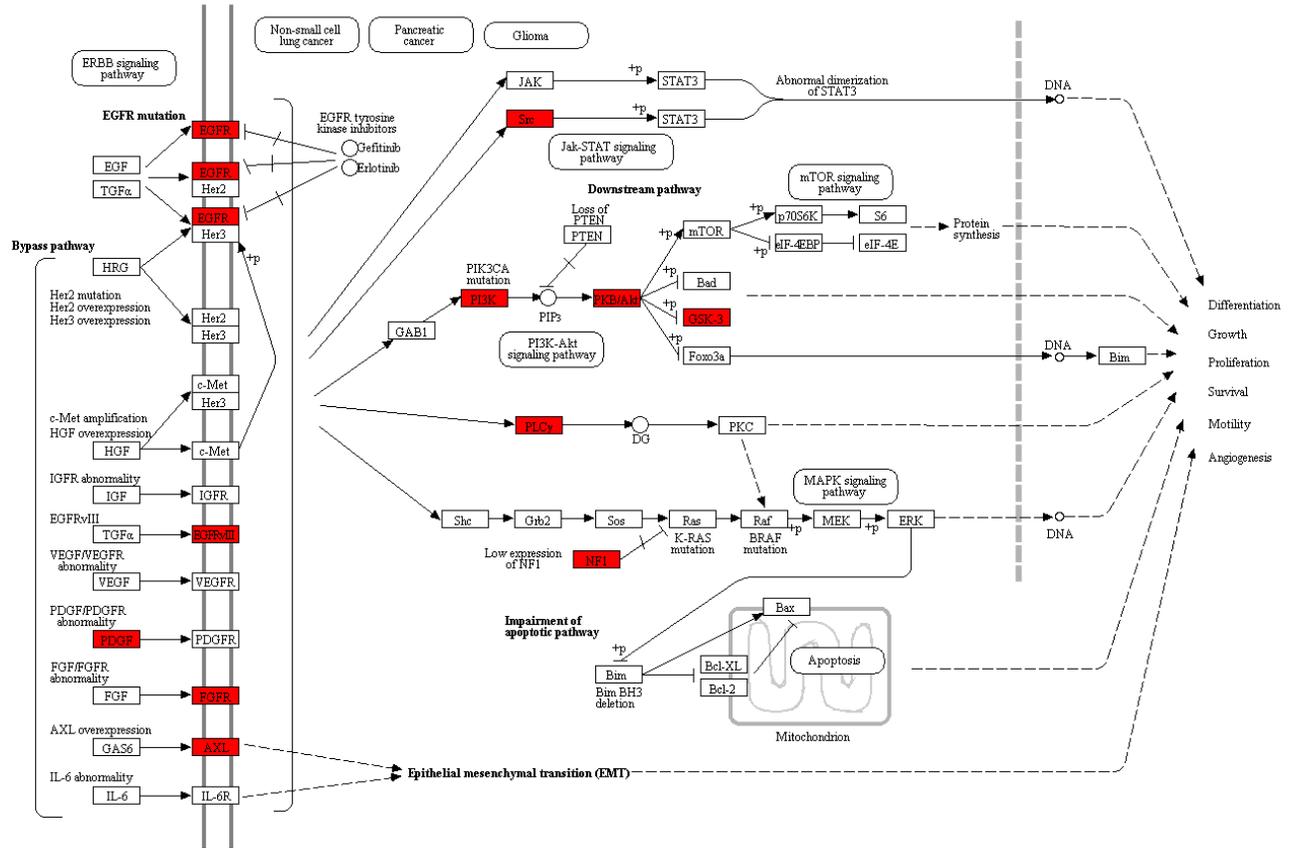


Figura 25: Via KEGG down-regulated “Via da resistência ao inibidor de tirosina quinase de EGFR” no ShinyGO com pacote Pathview.

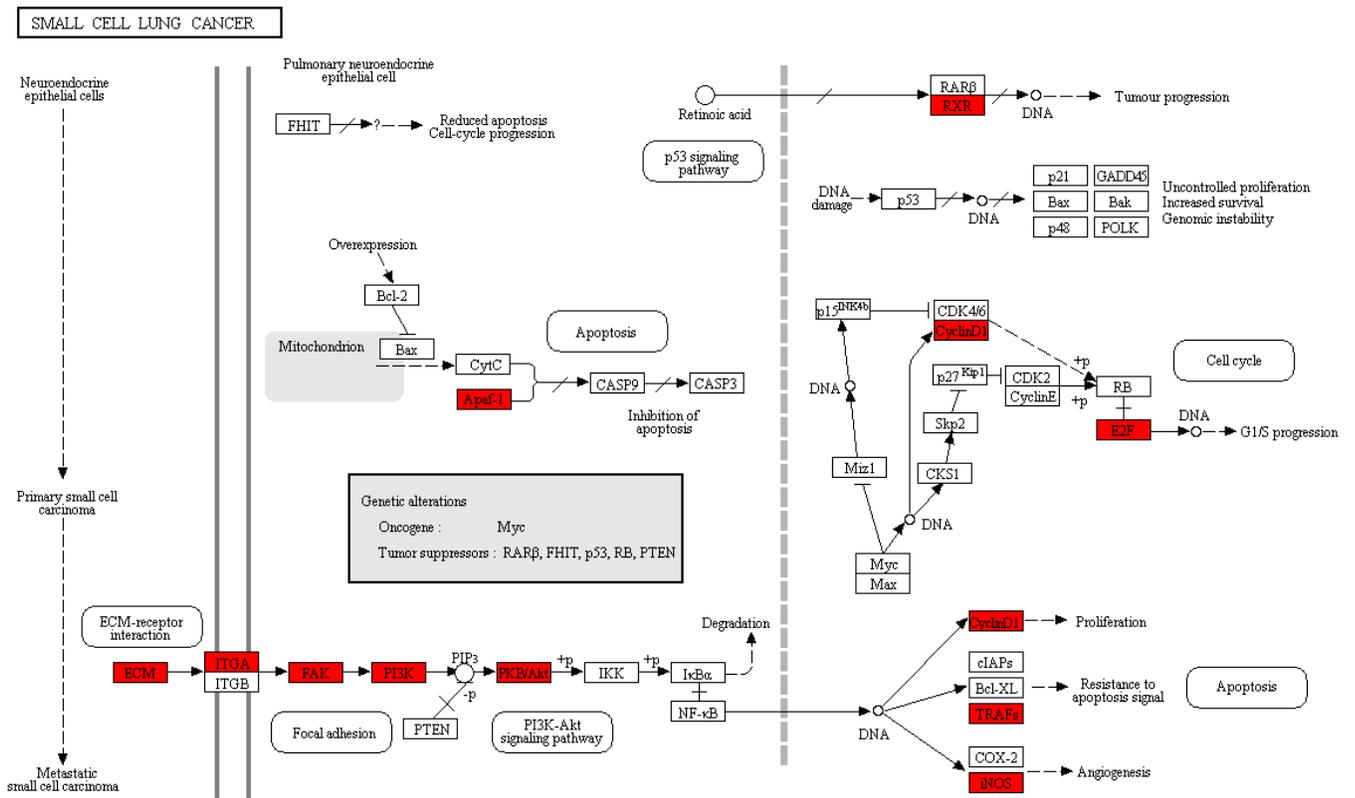


Figura 26: Via KEGG *down-regulated* “Câncer de pulmão de pequenas células” no ShinyGO com pacote Pathview.

É interessante visualizar que além das vias encontradas no *software* do WebGestalt, o ShinyGO também enriqueceu outras vias *down-regulated*, assim como no caso das vias *up-regulated* (Figuras 23, 24, 25 e 26).

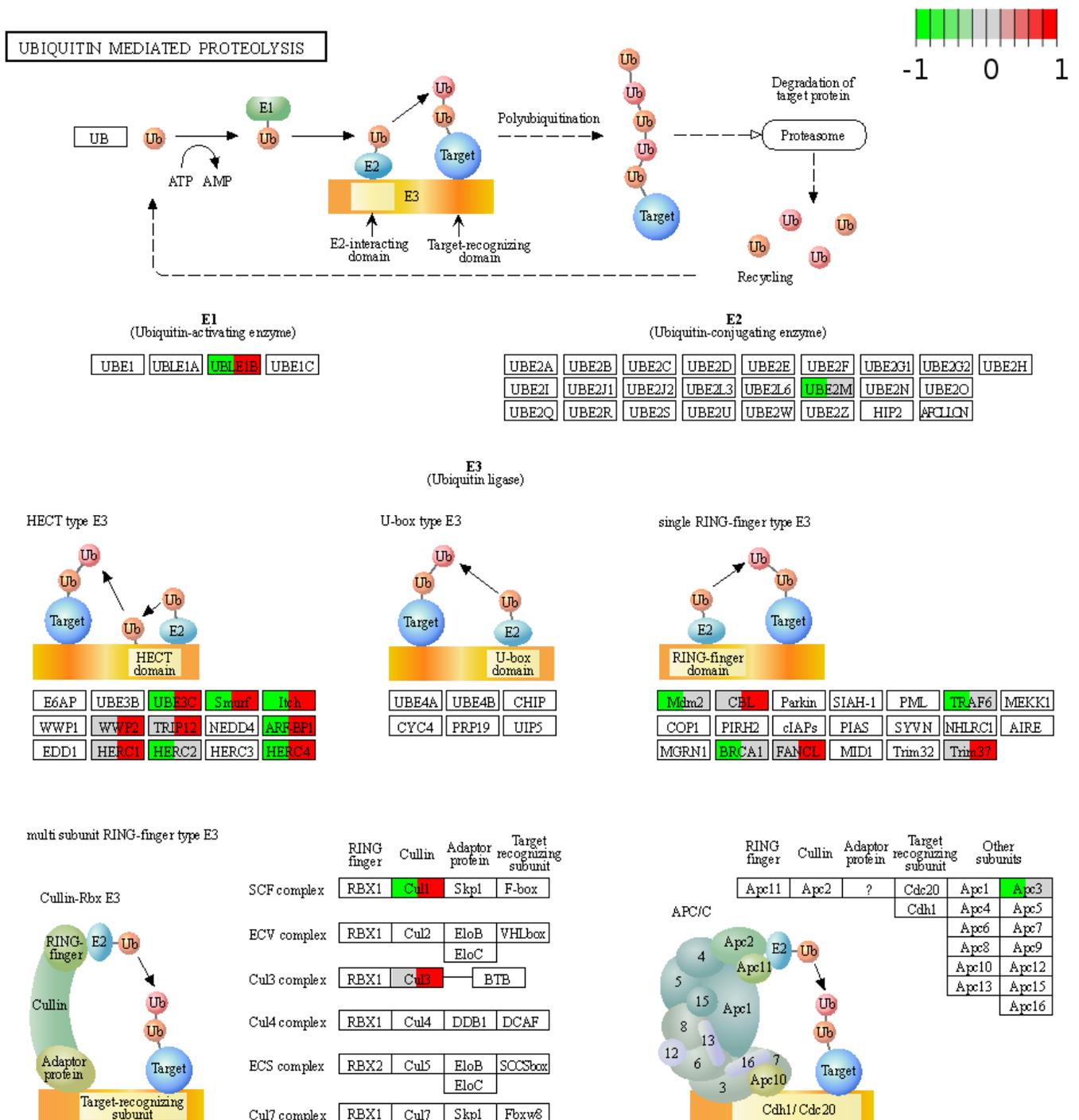


Figura 27: Via KEGG “Proteólise mediada por ubiquitinação” no ShinyGO com pacote Pathview. Comparação entre a mesma via com os valores de *fold change* positivos e negativos.

A via da proteólise mediada por ubiquitinação (Figura 27), aparece em ambas as análises do KEGG, em dois *softwares* online diferentes com algoritmos de análise distintos. Deste modo, foi feita uma imagem comparativa dessa via com o Pathview enriquecida com os valores de *fold change* apresentados nas análises para melhor visualização de como cada circRNA pode estar influenciando ou sendo influenciado.

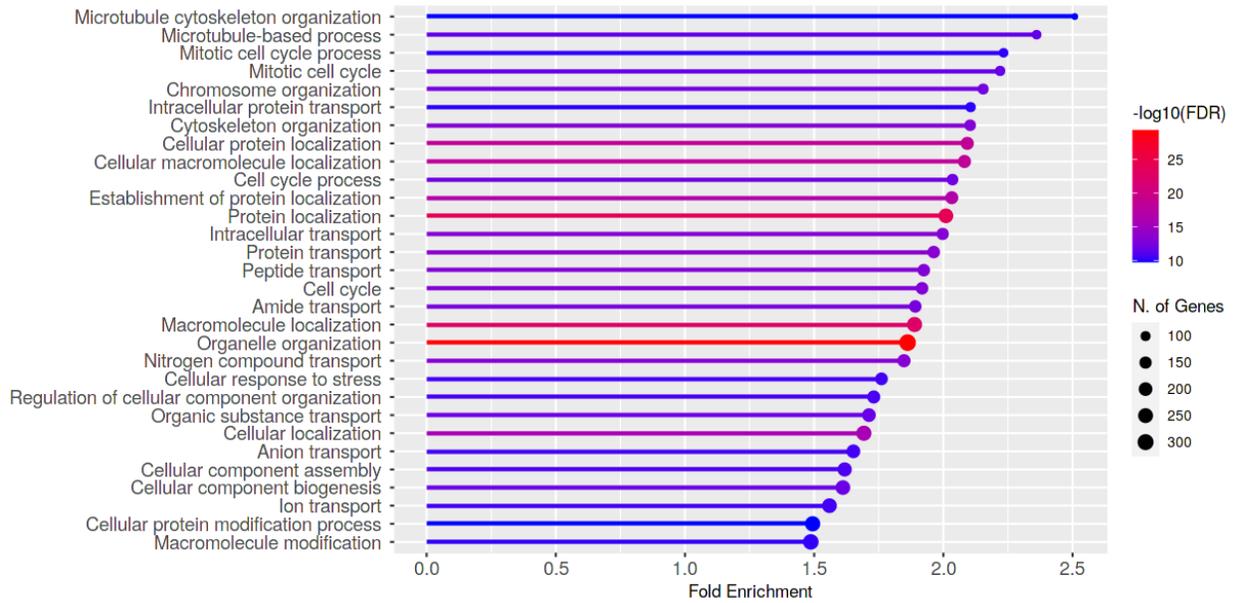


Figura 28: Termos GO *up-regulated* para “Processos biológicos” no ShinyGO.

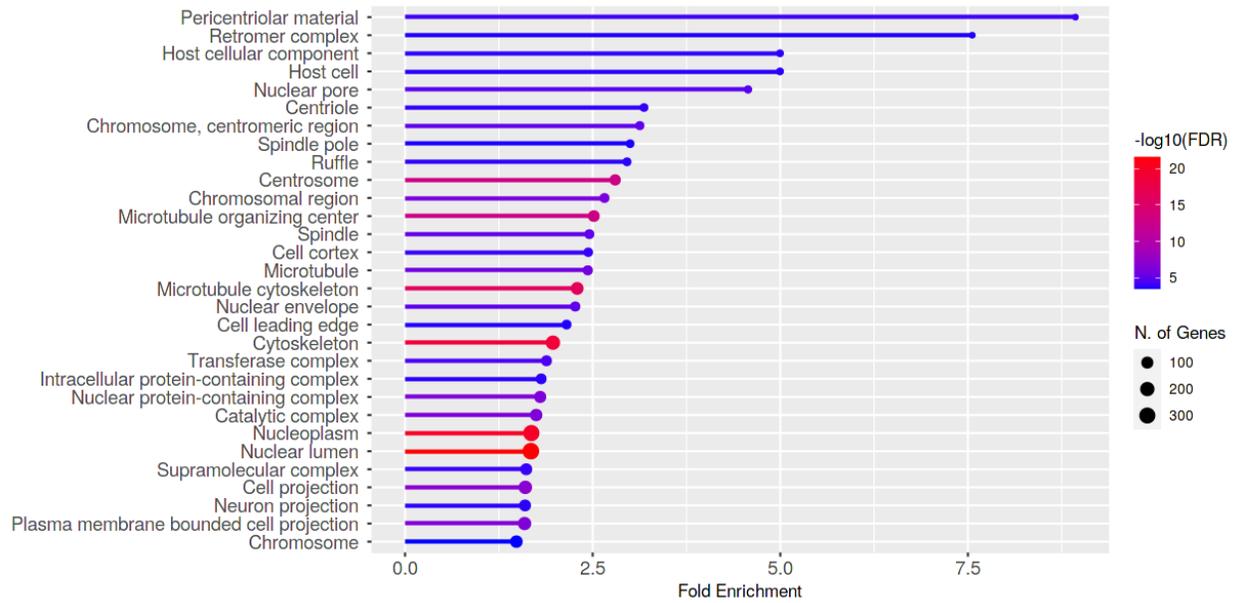


Figura 29: Termos GO *up-regulated* para "Componentes celulares" no ShinyGO.

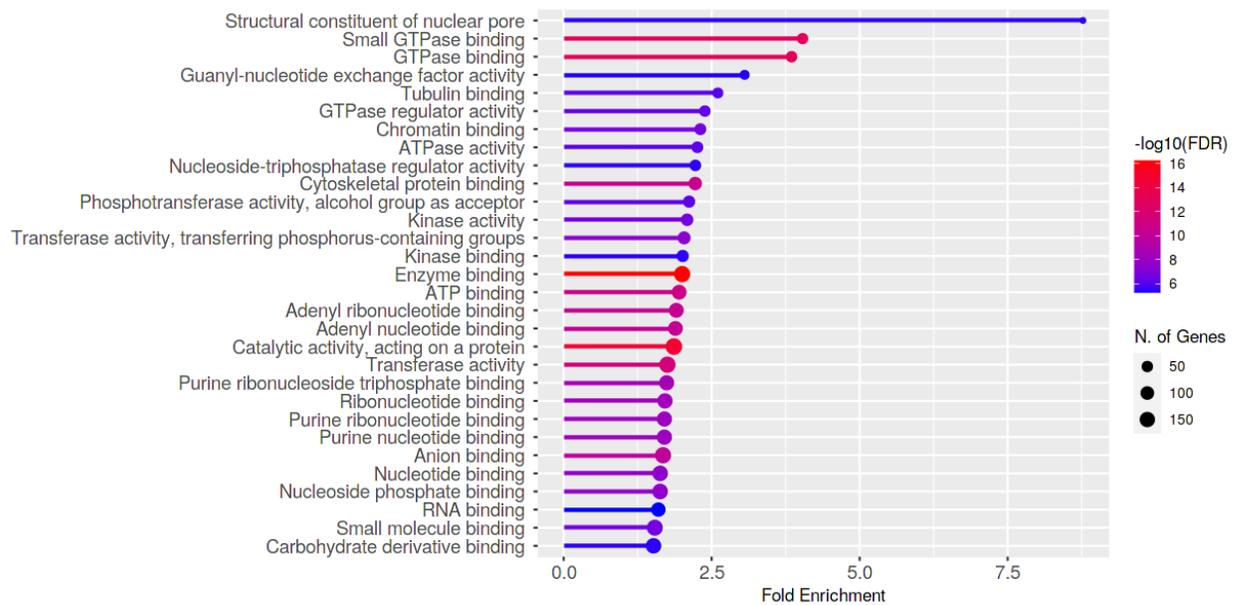


Figura 30: Termos GO *up-regulated* para "Funções Moleculares" no ShinyGO.

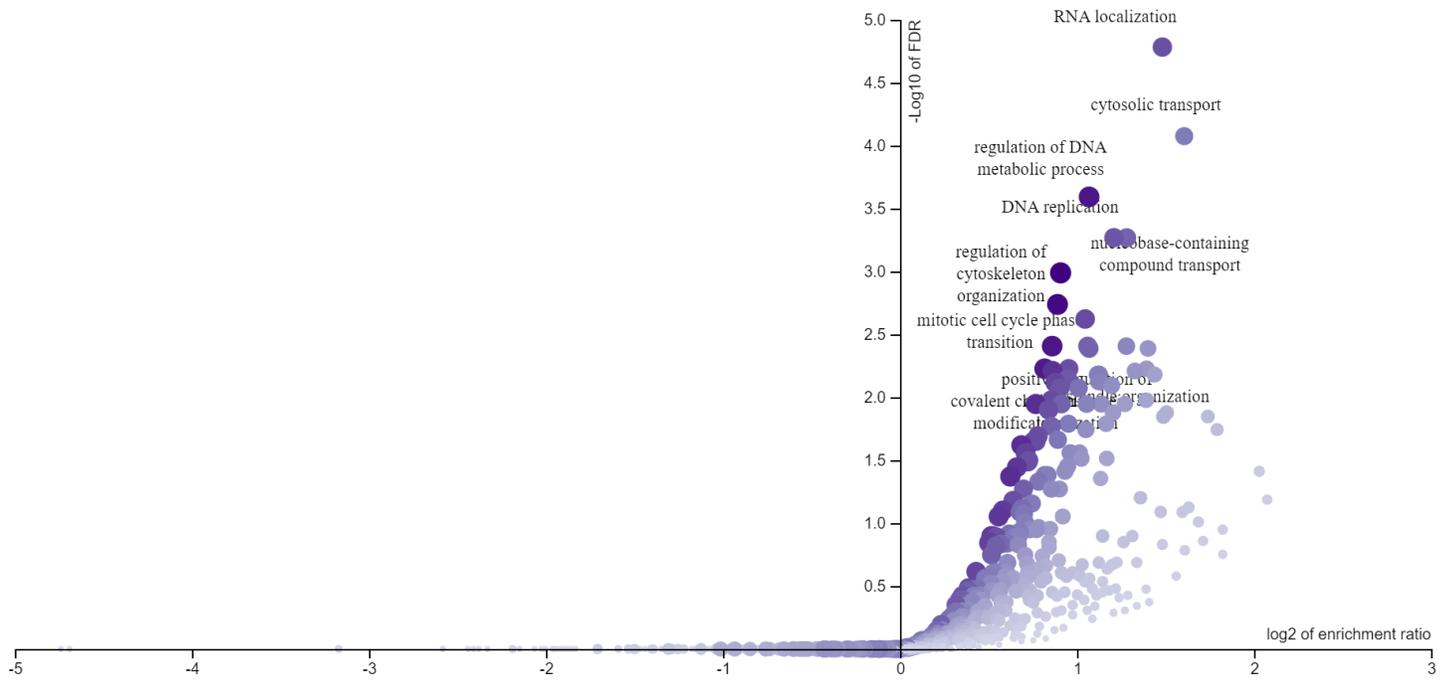
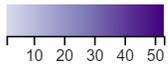


Figura 31: Termos GO *up-regulated* para "Processos biológicos" no WebGestalt.

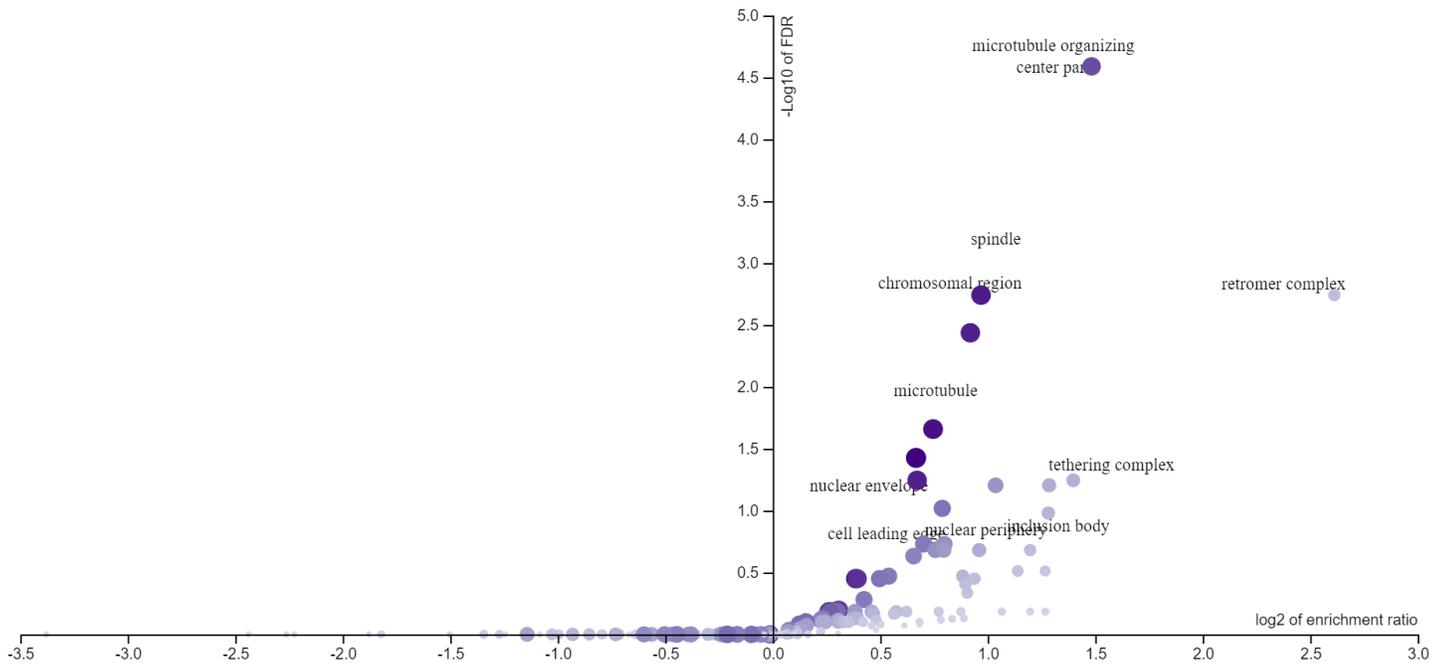
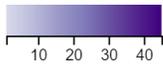


Figura 32: Termos GO *up-regulated* para “Componentes celulares” no WebGestalt.

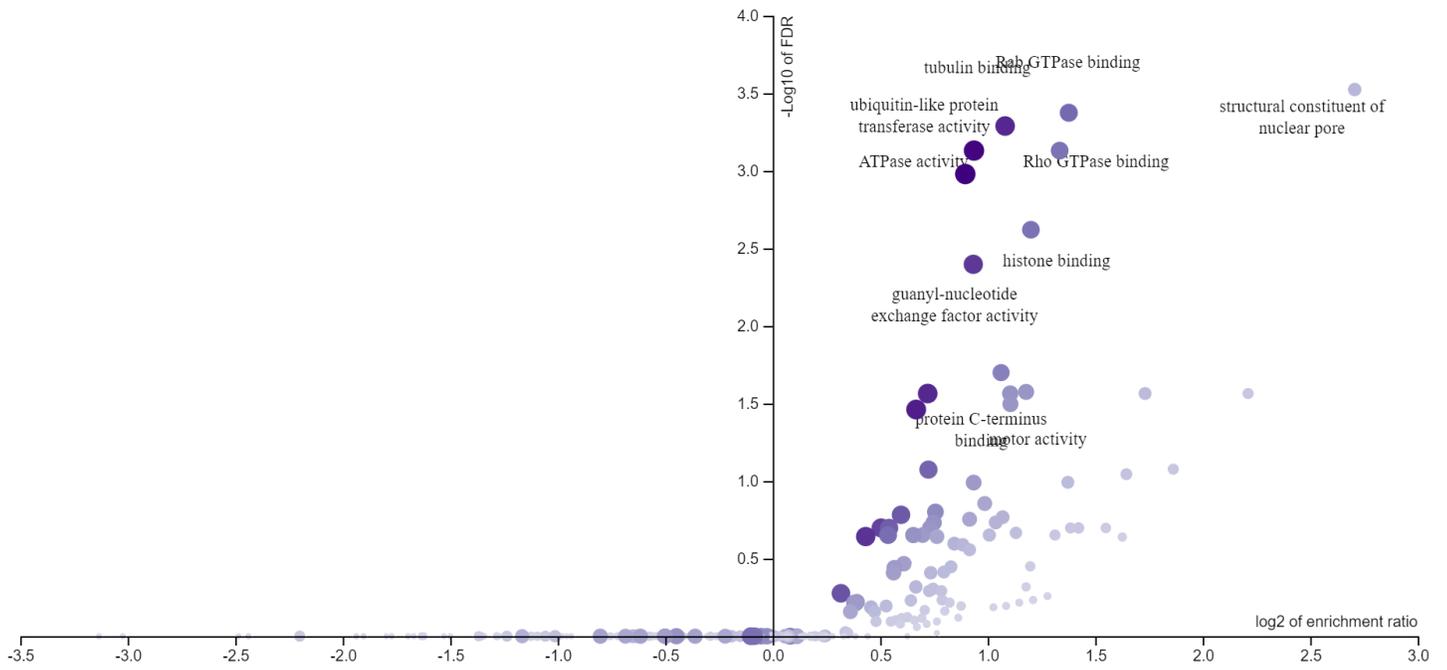
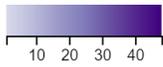


Figura 33: Termos GO *up-regulated* para "Funções Moleculares" no WebGestalt.

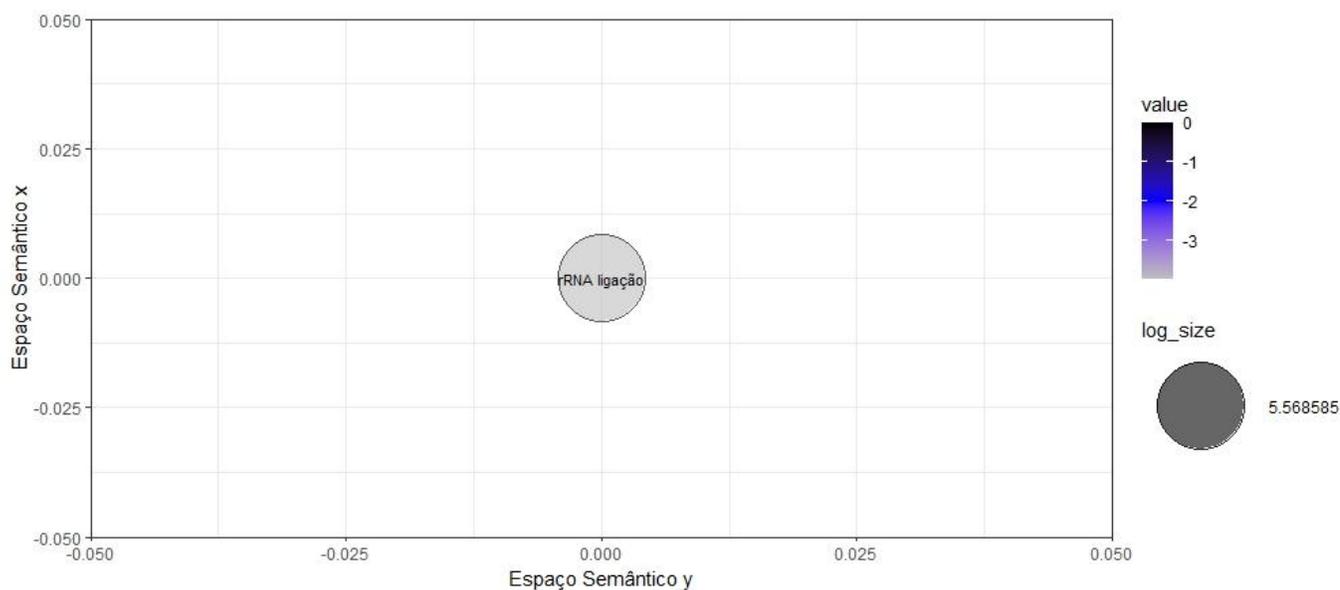


Figura 34: Termos GO *up-regulated* no GOrilla com pacote Revigo no R (para “Processos biológicos, Componentes celulares e Função Molecular”).

Foram feitas análises dos termos GO da lista de genes *up-regulated* por softwares diferentes: o ShinyGO, WebGestalt e GOrilla (Figuras 28, 29, 30, 31, 32, 33 e 34). O ShinyGO e o WebGestalt apresentaram resultados similares. Em contrapartida, o GOrilla enriqueceu apenas 1 termo, diferente dos demais apresentados pelos outros 2 algoritmos. Esse fato dos resultados do ShinyGO e WebGestalt serem parecidos, pode estar associado ao tipo de análise utilizada, chamada de ORA (*Over Representation-Analysis*) que ambos apresentam, enquanto o GOrilla usa a estatística do tipo mHG. Nos termos apresentados, ATRX está presente no termo mais enriquecido na Figura 27 “*Microtubule cytoskeleton organization*”, onde o mesmo termo também é encontrado entre o top 10 termos

mais enriquecidos na Figura 29. O ATRX também pode ser visto em outra via chamada “*Organelle organization*” na figura 27 juntamente com KMT2C e SET3D, todos *down-regulated* na linhagem celular SW620. No termo “*Cellular response to stress*” na figura 27, ATRX está junto com FOXP1, *down-regulated* na linhagem metastática.

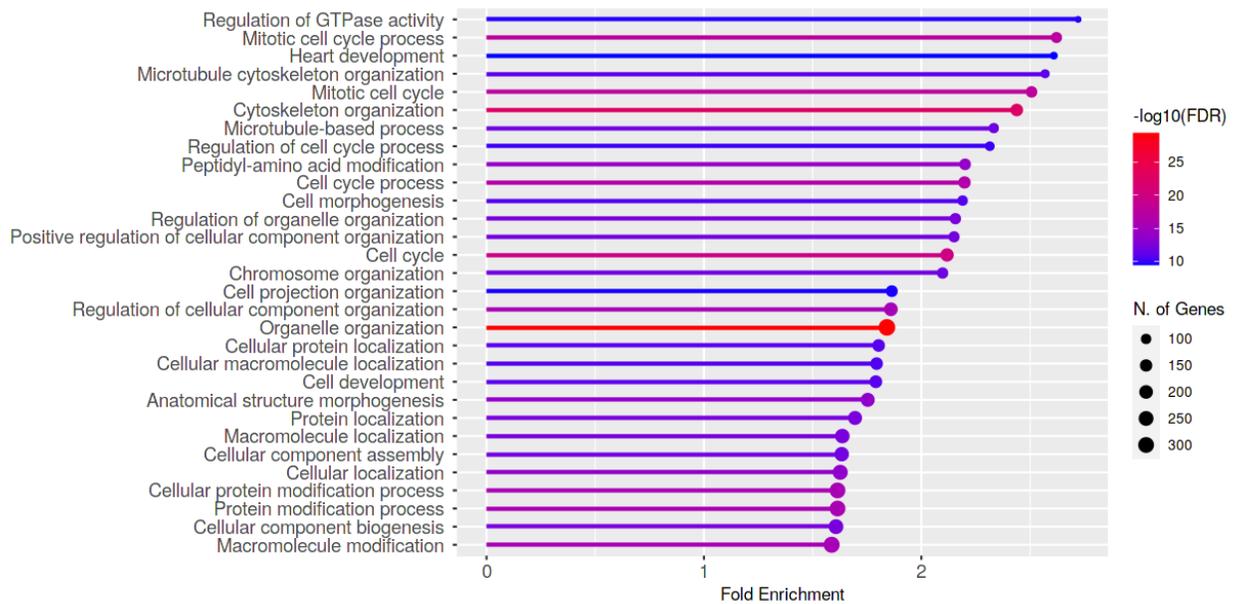


Figura 35: Termos GO *down-regulated* para “Processos biológicos” no ShinyGO.

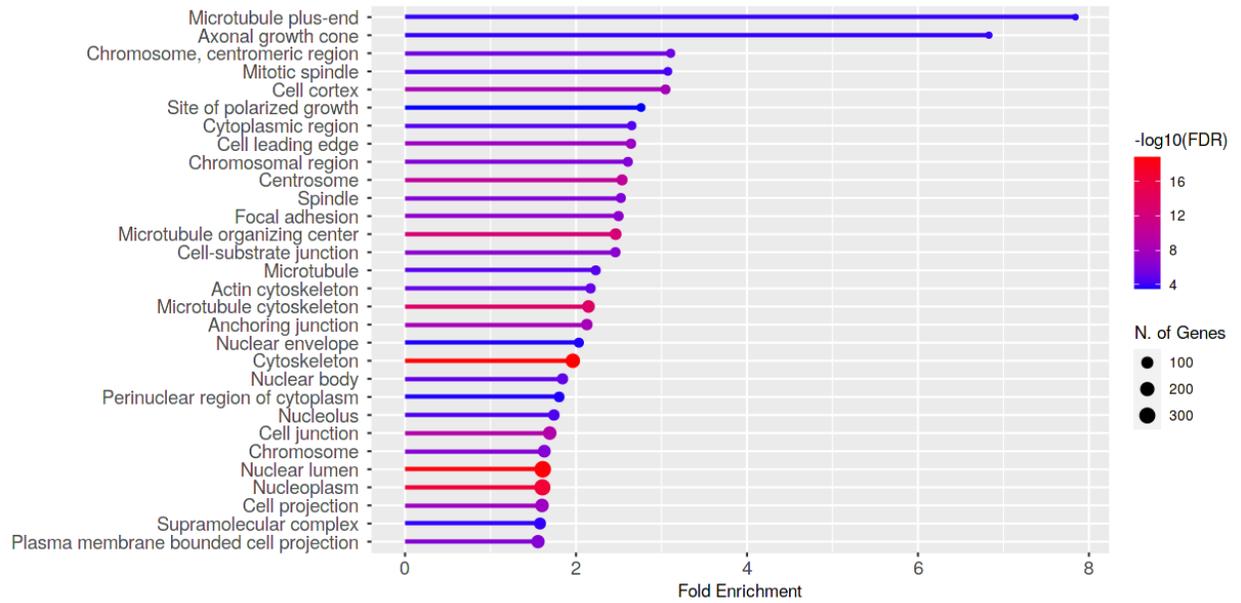


Figura 36: Termos GO *down-regulated* para "Componentes celulares" no ShinyGO.

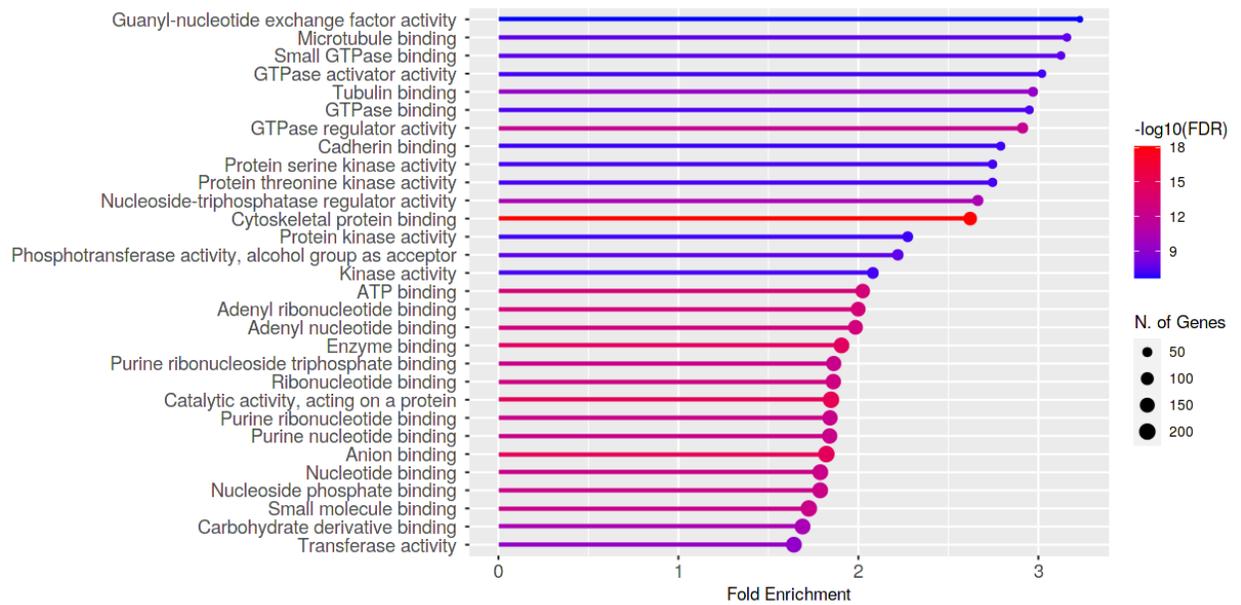


Figura 37: Termos GO *down-regulated* para "Funções Moleculares" no ShinyGO.

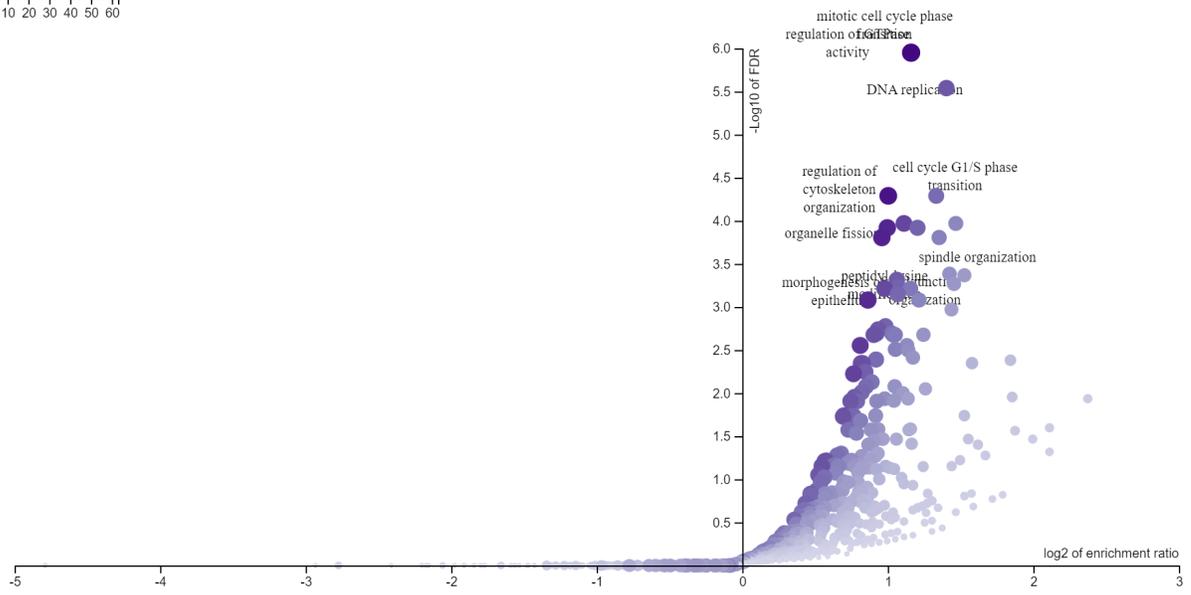
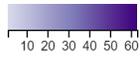


Figura 38: Termos GO *down-regulated* para “Processos biológicos” no WebGestalt.

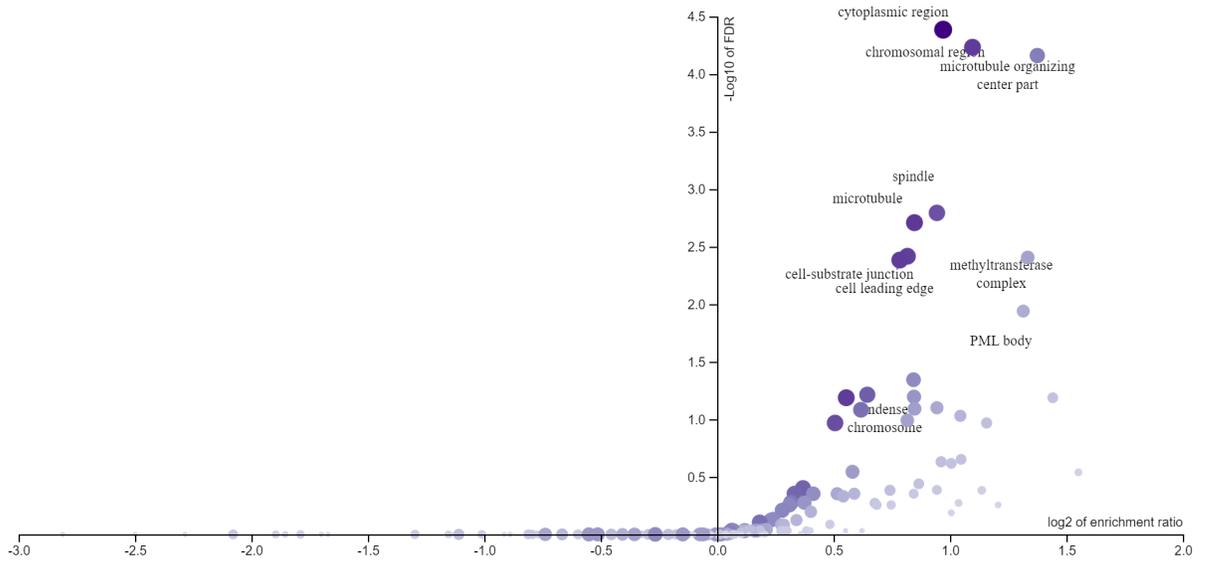
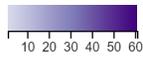


Figura 39: Termos GO *down-regulated* para “Componentes celulares” no WebGestalt.

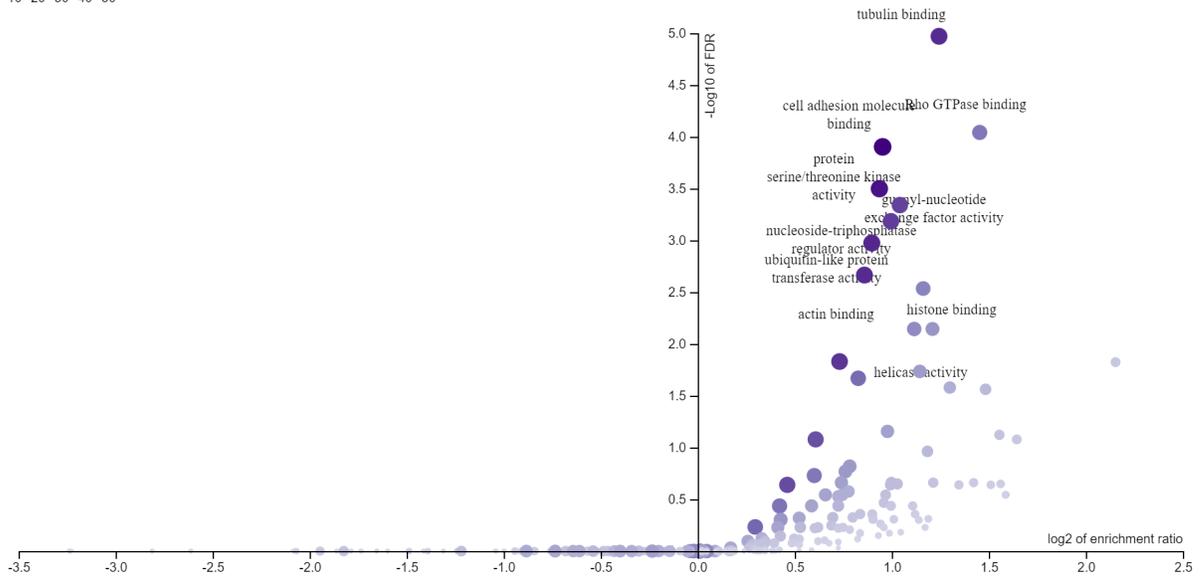
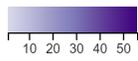


Figura 40: Termos GO *down-regulated* para “Funções Moleculares” no WebGestalt.

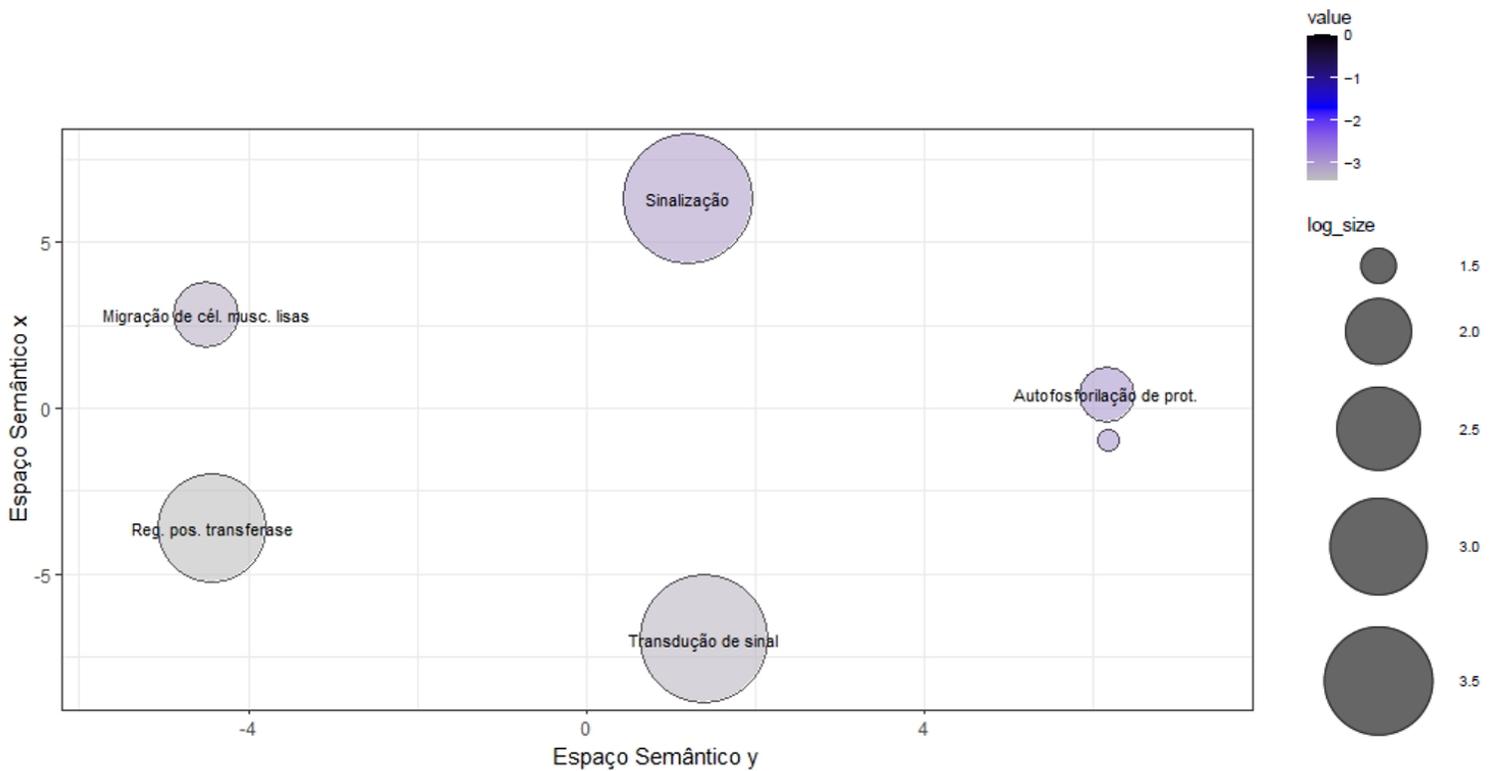


Figura 41: Termos GO *down-regulated* no GOrilla com pacote Revigo no R.

Foram feitas análises dos termos GO da lista de genes *down-regulated* por *softwares* diferentes: o ShinyGO, WebGestalt e GOrilla (Figuras 35, 36, 37, 38, 39, 40 e 41). O ShinyGO e o WebGestalt apresentaram resultados similares novamente. Nos termos apresentados na figura 34, WNT5A está presente no termo "*Cellular component biogenesis*", junto a TSPAN4, ambos *down-regulated* na linhagem celular SW480. O WNT5A também pode ser visto em uma via chamada "*Organelle organization*" juntamente com EPAS1, *down-regulated* na linhagem celular SW480.

Para a linhagem SW620, DNMT1 e RUNX estão presentes juntos em diversas termos na figura 34, como "*Regulation of cellular component organization*", "*Cellular component biogenesis*", "*Cellular component assembly*" e "*Positive regulation of cellular component organization* ", ambos estão *up-regulated* na SW620.

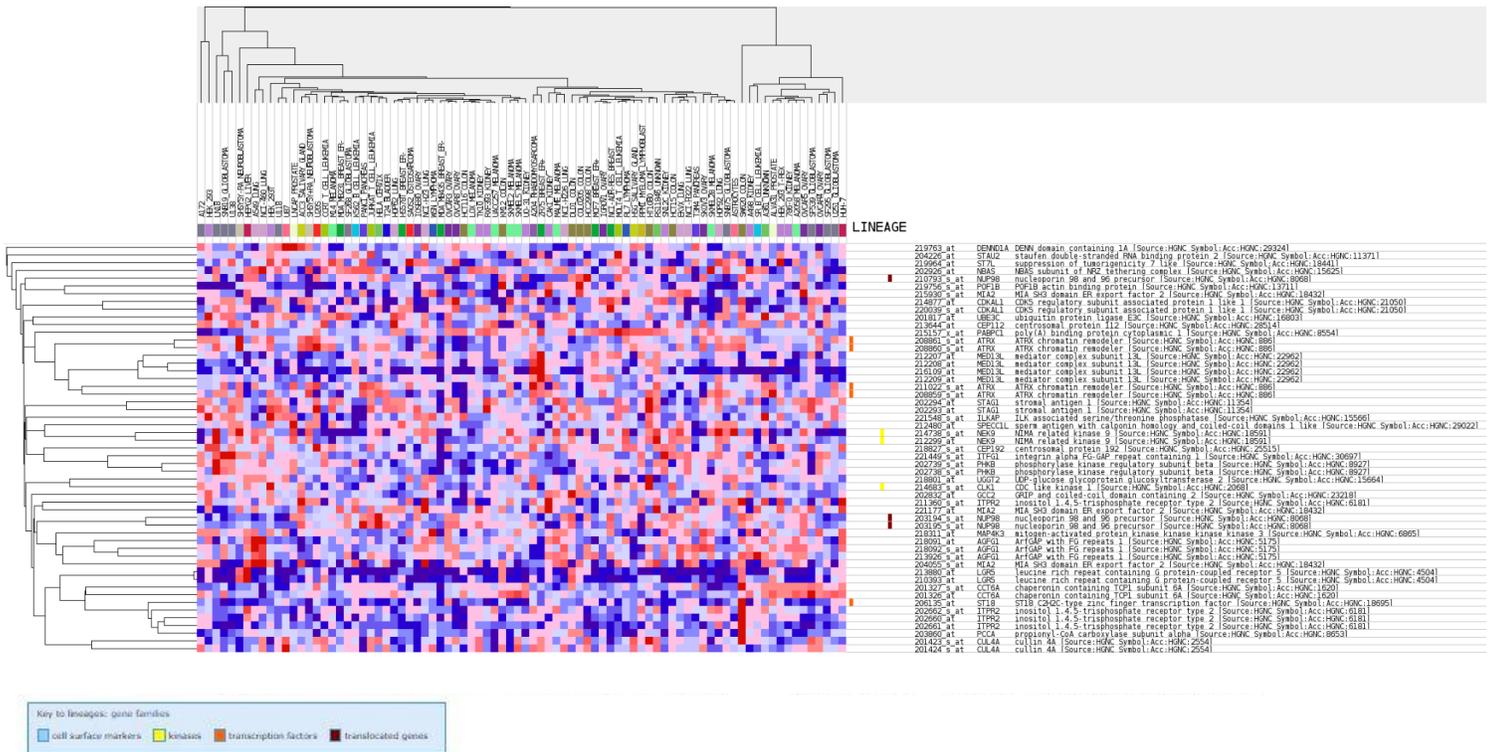


Figura 42: Relação dos 50 primeiros genes da lista *up-regulated* ordenados do maior para o menor de acordo com o valor de *fold change* e expressão nas linhagens celulares comerciais para câncer pelo GSEA.

Acima, na figura 42, observa-se uma análise feita pelo GSEA para as linhagens celulares e sua relação com os genes da lista *up-regulated*. O ATRX, já descrito como um fator de transcrição, tem sua expressão baixa na linhagem SW620, concordando com os resultados encontrados nas análises de expressão diferencial, que o circ ATRX está *down-regulated* na linhagem SW620.

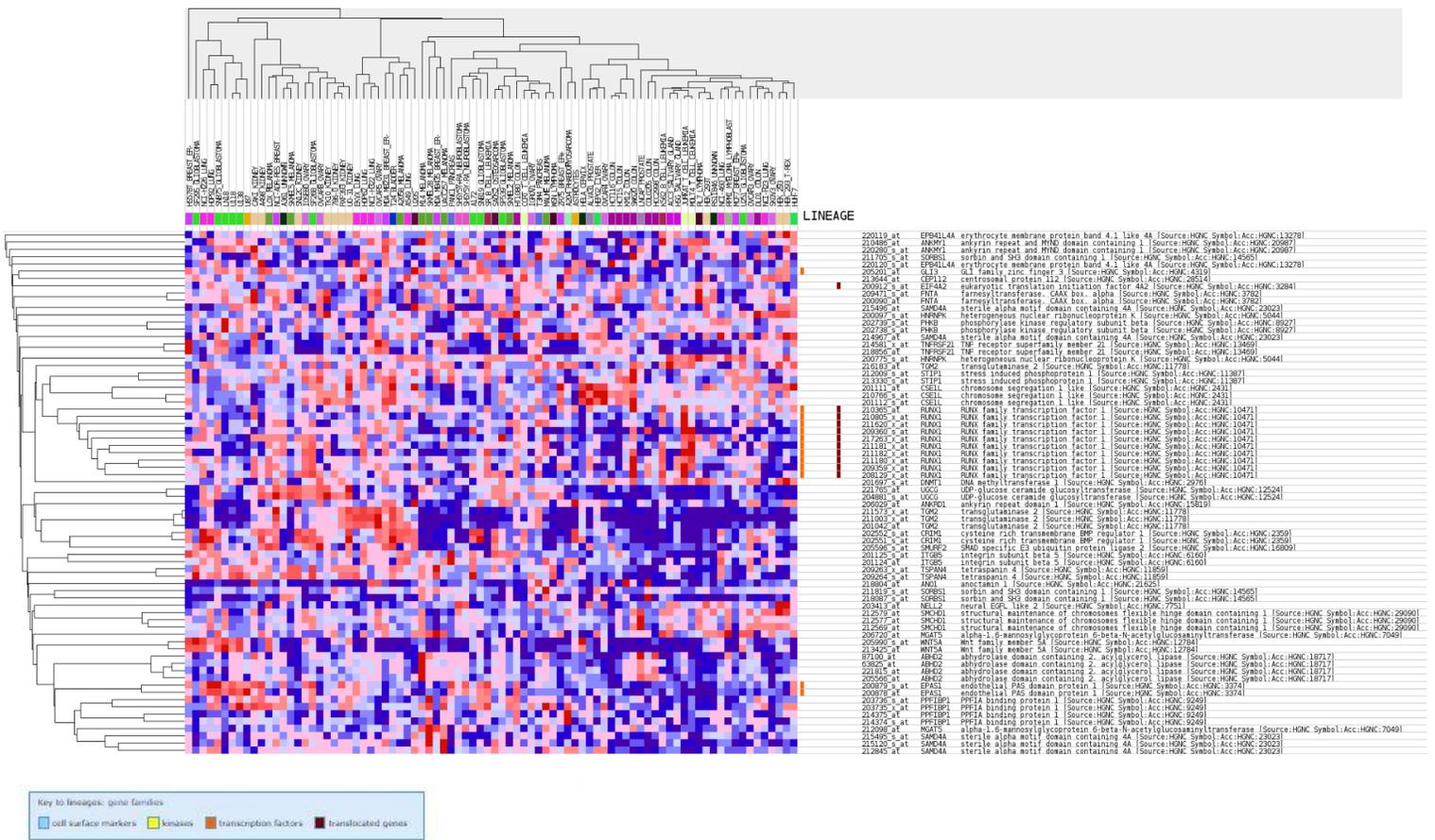


Figura 43: Relação dos 50 primeiros genes da lista *down-regulated* ordenados do maior para o menor de acordo com o valor de *fold change* e expressão nas linhagens celulares comerciais para o câncer pelo GSEA.

Na figura 43, observa-se uma análise feita pelo GSEA para as linhagens celulares e sua relação com os genes da lista *down-regulated*. O DNMT1, analisado como um circRNA *up-regulated* na linhagem SW620 anteriormente pela expressão diferencial, apresenta-se no GSEA superexpresso nessa linhagem, corroborando com os resultados já apresentados.

Embora existam diversos algoritmos de detecção para circRNAs provenientes de experimentos de *RNA-Seq* comum ou específico com tratamento da RNase R, poucos mostraram eficiência na detecção correta dos RNAs circulares e isso continua sendo uma área em desenvolvimento e melhoramento contínuo. Diversos artigos fazem a comparação de mais de 5 desses algoritmos, cada um com um método diferente seja na detecção ou na análise de expressão diferencial, mas nenhum comprovou a eficiência real na eliminação de falsos positivos detectados ou expressos com a intersecção de vários métodos ou a escolha de apenas um. Nesse projeto, foram feitos diversos testes, e foi escolhido o método de análise baseado na identificação de circRNAs pelo circExplorer e análise de expressão diferencial utilizando um *pipeline in house* adaptado e baseado na literatura, que usa programas já bem estabelecidos para análise tradicional de expressão de transcritos, de acordo com a revisão de artigos e tentativas de utilização de outros pacotes como Circompara (Gaffo et al., 2017) e find-circ (Memczak et al., 2013) para a detecção dos circRNAs, e algoritmos como DEBKS (Zelin et al., 2021) e SRCP (Rabin et al., 2021) para expressão diferencial. Para a identificação, circExplorer foi diversas vezes descrito nos artigos como o mais atual e curado banco de detecção e método de identificação de circRNAs, apenas com a desvantagem de precisar de mais horas e requerir maior poder de processamento (Szabo et al., 2016; Zhang et al., 2020; Li et al., 2020; Annie et al., 2021). Para a análise de expressão diferencial, os algoritmos testados como DEBKS e SCRCP apresentaram diversos erros nos *scripts*, necessitando de modificações em seu código fonte, e não apresentaram dados confiáveis. Talvez uma atualização em sua versão com essas correções

deva levá-los a serem métodos interessantes em se testar no futuro, já que todos são muito recentes, e estão publicados e testados apenas em seus artigos de origem.

Com os resultados obtidos, é possível observar um perfil de circRNAs expressos nas linhagens do tumor primário e metastático do CRC.

Na linhagem SW480, do tumor primário, os circRNAs diferencialmente expressos com fortes evidências quando se foi analisada a correlação das redes PPI, KEGG e termos GO foram TGM2, TSPAN4 (hsa_circ_18995), WNT5A, GLI3 (hsa_circ_08687), SMURF2 e UBE3C (hsa_circ_12017).

Um grupo de pesquisadores publicou em 2019, um estudo sobre o miR-532-3p. No painel de expressão de microRNA plasmático, o miR-532-3p foi definido como um biomarcador valioso para adenoma colorretal (CRA). No entanto, seu padrão de expressão e função no CRC permaneceram obscuros. Nesse estudo publicado, foi investigado os níveis de expressão de miR-532-3p e descobriu que ele era regulado para baixo *in situ* tanto no CRA quanto no CRC. Ele também funcionou como um sensibilizador para quimioterapia em CRC, induzindo a parada do ciclo celular e apoptose precoce por meio de seus efeitos de ativação no p53 e nas vias de sinalização apoptótica. No entanto, descobriu-se que miR-532-3p restringe o crescimento celular, metástase e fenótipo de transição epitelial-mesenquimal (EMT) do CRC. O estudo sobre o mecanismo por trás desses efeitos revelaram que miR-532-3p se liga diretamente às regiões 3'UTR de ETS1 e TGM2, reprimindo a via canônica de sinalização Wnt/ β -catenina. TGM2 serviu como um alvo funcional vital do miR-532-3p na supressão da progressão

do CRC (Gu et al., 2019). Em outro estudo, foram estudados os processos angiogênicos e antiapoptóticos, críticos para o crescimento do tumor CRC (Song et al., 2019). Relataram o efeito do TGM2 na angiogênese e apoptose do CRC via ativação da sinalização Wnt / β -catenina (Yang et al., 2019). A regulação negativa de TGM2 inibiu a angiogênese e a expressão de β -catenina, Wnt3a, MMP-2, MMP-9 e ciclina D1, enquanto sua regulação negativa promoveu a expressão de TIMP-1 e caspase-3 e apoptose em TGM2 nas células CRC transfectadas com siRNA. Esses resultados indicaram que o TGM2 é capaz de promover efetivamente a angiogênese e inibir a apoptose por meio da ativação da via de sinalização Wnt / β -catenina em células CRC (Ji et al., 2021). Nas análises desse projeto, TGM2 está relacionado a WNT5A nas redes PPI, ambos *down-regulated* na linhagem SW480. Como visto no estudo acima publicado, sua expressão baixa, pode influenciar positivamente, influenciando na não progressão do CRC. Pode-se hipotetizar, que o circ TGM2 pouco expresso pode estar influenciando também na via Wnt do KEGG, que se encontra *down* regulada, com circ WNT5A e circ NFAT, suprimindo essa via, e colaborando na não progressão tumoral.

Em 2020, circ KIF4A e WNT5A foram regulados positivamente e miR-139-3p regulado negativamente em células e tecidos de glioma. O circKIF4A promoveu a expressão de Wnt5a por esponja miR-139-3p. O *knockdown* de circKIF4A inibiu a capacidade de formação de colônias, migração e invasão e promoveu a apoptose de células de glioma pela regulação do miR-139-3p. O *knockdown* de circKIF4A inibiu a via de sinalização Wnt / β -catenina e o sinal relacionado à proliferação via miR-139-3p. Além disso, o *knockdown* de circKIF4A ou a superexpressão de miR-139 suprimiu a capacidade de formação de esferas de GICs e inibiu a via de

sinalização Wnt/ β -catenina e o sinal relacionado à proliferação em células iniciadoras de glioma (GICs). A depleção de circ KIF4A diminuiu o nível de expressão de WNT5A e Ki-67, e inibiu a tumorigênese nos modos de xenoinxerto (Huo et al., 2020). Neste projeto, o circ WNT5A está *down-regulated* na linhagem SW480, e pode estar associado a uma tentativa de inibição da tumorigênese.

O circ UBE3C (hsa_circ_12017) na linhagem SW480 encontra-se *up-regulated*. No estudo a seguir, os autores mostraram que o UBE3C promoveu a progressão do carcinoma hepatocelular (CHC) ao regular a transição epitelial-mesenquimal das células do CHC. Clinicamente, um estudo de microarray de tecido de uma coorte contendo 323 pacientes com CHC revelou que a superexpressão de UBE3C em tecidos primários de CHC se correlacionou com a diminuição da sobrevida (razão de risco [HR] = 1,657, intervalo de confiança de 95% [CI] = 1,220-2,251, P = 0,001) e recorrência tumoral precoce (HR = 1,653, IC 95% = 1,227-2,228, P = 0,001) em pacientes com CHC pós-operatório. Portanto, os resultados indicaram que o UBE3C é um oncogene candidato envolvido no desenvolvimento e progressão do tumor.

A ubiquitinação controla várias funções celulares marcando proteínas para degradação ou incorporação proteasomal em outros complexos reguladores. SMURF2 é uma ubiquitina ligase do domínio HECT E3 envolvida na degradação de SMADs (OMIM 601595), receptor de TGF-beta (TGFB1; OMIM 190181) e outros substratos. O SMURF2 também funciona na regulação da polaridade das células neuronais e planares, indução de senescência e supressão de tumores (OMIM 605532) (Blank et al., 2012). Também foi implicado na ubiquitinação de

reguladores negativos (Axin e GSK-3 β) e positivos (β -catenina) da via de sinalização canônica Wnt (Han et al., 2006; Wu et al., 2009). Estudos com ratos SMURF 1 / 2 com nocaute duplo revelaram que o SMURF2 regulou a sinalização Wnt não-canônica através da ubiquitinação de PRICKLE1 (Narimatsu et al., 2009). Outros estudos mostraram que o SMURF2 ubiquitina Id1 (Kong et al., 2011) foi envolvido em diversos processos celulares, incluindo proliferação, diferenciação e angiogênese (Perk et al., 2005). Assim, foi possível observar como é importante o estudo dessa via a fim de se determinar se a ubiquitinação mediada por SMURF2 de um desses ou de outros substratos proteicos ainda não identificados, além da regulação da senescência, desempenham um papel importante em sua função como supressor de tumor (Ramkumar et al., 2012). O circ SMURF2 está *up* e *down-regulated* na SW480, e pode agir como um supressor de tumor. Porém, devido às suas coordenadas, este pode desempenhar dois possíveis papéis reguladores. Quando *up-regulated*, embora as evidências estejam fracas quando analisadas as redes PPI e vias KEGG, ele atua junto na via da ubiquitinação com circ UBE3C (hsa_circ_12017), facilitando a progressão tumoral. Quando *down-regulated*, ele atua na via Wnt seguindo a interação PPI com WNT5A, como um supressor de tumor, podendo se relacionar com os demais circRNAs como TGM2 e GLI3.

Em uma meta-análise dos perfis de expressão de circRNA em 1.430 pacientes com CRC (Li et al., 2021), circRNAs diferencialmente expressos no CRC correlacionaram-se com o diâmetro do tumor, diferenciação, estágio TNM, invasão, linfonodo e metástase à distância. Em particular, os circRNAs regulados para cima no CRC correlacionaram-se com uma pior sobrevida geral, enquanto os

circRNAs regulados para baixo no CRC se correlacionaram com a sobrevida geral prolongada. A comparação entre células SW480 primárias e células SW620 metastáticas revelou 623 circRNAs diferencialmente expressos, dos quais 275 foram regulados para cima e 348 foram regulados para baixo em células SW620. Muitos circRNAs, como aqueles derivados dos genes GLI3 e RAPGEF5, foram regulados para níveis indetectáveis em células SW620, apesar da alta expressão em células SW480. Curiosamente, nas análises apresentadas, não houve detecção na linhagem SW620, conforme a literatura acima, mas na linhagem SW480, circ GLI3 está *down-regulated*, associado a circ WNT5A. Porém, a transição epitelial-mesenquimal (EMT) é uma marca registrada do câncer que promove a progressão maligna, especialmente durante o processo de invasão do câncer. Uma melhor compreensão da EMT ajudaria a elucidar o mecanismo molecular subjacente à metástase do CRC e poderia fornecer novos *insights* sobre a identificação de potenciais biomarcadores e alvos terapêuticos. Nesse artigo publicado em 2021, GLI3 foi identificado como um regulador chave nas redes de coexpressão e interação proteína-proteína (PPI) envolvidas na EMT. As análises de bioinformática indicaram que o GLI3 teve uma alta correlação com os marcadores EMT no CRC. Experimentos *in vitro* mostraram que o *knockdown* de GLI3 atenuou as capacidades migratórias e invasivas das células CRC por meio da influência da propriedade EMT, especialmente pela regulação da fosforilação da via de sinalização ERK. Além disso, a maior expressão de GLI3 prediz pior prognóstico em pacientes com CRC (Shen et al., 2021). Deste modo, os resultados apresentados neste projeto, estão de acordo com os apresentados nesse artigo recente, onde o circ GLI3 (hsa_circ_08687) está subexpresso na linhagem SW480,

talvez como um mecanismo de tentativa de não progressão tumoral, corroborando com os demais circRNAs subexpressos nesta linhagem.

Ying *et al*, mostraram que o circ TSPAN4 foi significativamente regulado positivamente em tecidos LUAD de linhagens celulares. O aumento do circ TSPAN4 foi associado ao estágio avançado de tumor, nódulos, metástases, nódulo linfático e metástase à distância. O *knockdown* do circ TSPAN4 mediado por lentivírus atenuou dramaticamente a capacidade metastática das células LUAD tanto *in vitro* quanto *in vivo*. Mecanicamente, circ TSPAN4 interagiu diretamente com miR-665, aumentando a expressão de ZEB1, um iniciador de metástase bem conhecido. Foi ressaltado que a capacidade metastática reduzida causada pela depleção circ TSPAN4 foi parcialmente resgatada pelo silenciamento de miR-665 ou superexpressão de ZEB1 (Ying *et al.*, 2019). Nas análises apresentadas, o circ TSPAN4 (hsa_circ_18995) se mostrou *down-regulated* na SW480, tumor primário, indicativo que o gene TSPAN4 esteja subexpresso, assim como descrito no artigo, que está superexpresso apenas nas linhagens metastáticas.

Para a linhagem metastática SW620, os circRNAs diferencialmente expressos com fortes evidências quando se foi analisada a correlação das redes PPI, KEGG e termos GO foram DNMT1 (hsa_circ_28013), RUNX1, UBE3C (hsa_circ_12017), SETD3 (hsa_circ_13006), KMT2C, ATRX (hsa_circ_0140609), FOXP1 (hsa_circ_11346) e ST18 (hsa_circ_13413).

Em um estudo, foram analisados os painéis de 12 amostras normais de tecidos mamários e tecidos tumorais correspondentes. Como resultado, os

pesquisadores verificaram que a expressão de ST18 estava desregulada no câncer de mama. A regulação negativa de ST18 parecia dever-se à hipermetilação das ilhas CpG (Jandrig et al., 2004). Em outro estudo alguns anos depois, através da técnica de PCR quantitativo, os resultados indicaram que ST18 em rato era abundantemente expresso na tireóide, medula espinhal, aorta e cérebro, e mais modestamente no estômago, fígado e pâncreas, mas não no intestino. A análise por PCR e imunofluorescência do pâncreas dos ratos detectou alta expressão de ST18 no pâncreas alfa e beta endócrino, mas não nas células acinares ou ductais. Nas células secretoras de insulina INS832 / 13 de rato e nas células ilhotas humanas normais, ST18 foi expresso no citoplasma (Henry et al., 2014). Mais recentemente, Shukla *et al.* identificaram 14 miRNAs diferencialmente expressos em tecido de câncer cervical e amostras de soro por meio da realização de análise de sequenciamento e PCR em tempo real. Entre eles, miR-17-5p, miR-32-5p e miR-454-3p foram superexpressos, enquanto miR-409-3p foi regulado para baixo em amostras de soro e biópsias tumorais de pacientes com câncer cervical. Foi observada uma correlação inversa entre os níveis de miR-409-3p e miR-454-3p e a expressão de seus genes alvo MTF2 e ST18, respectivamente, em biópsias de câncer cervical (Li et al., 2017). Os genes ST18 mostraram inibir a formação de colônias de células cancerosas em ágar mole e regular a expressão gênica pró-apoptótica e pró-inflamatória em fibroblastos e em células B pancreáticas (Jandrig et al., 2004; Yang et al., 2008). É interessante observar que circ ST18 (hsa_circ_13413) está *down-regulated* na SW620, o que pode levar a hipótese de que ele subexpresso corrobora com a progressão tumoral e metástase.

A expressão de RUNX1 é regulada em excesso nos tecidos de CRC. O RUNX1 regulado positivamente promove a metástase celular e a transição epitelial para mesenquimal (EMT) do CRC, tanto *in vitro* quanto *in vivo*. RUNX1 pode ativar a sinalização Wnt/ β -catenina nas células CRC interagindo diretamente com a β -catenina e visando as regiões promotora e potenciadora do KIT para promover a transcrição do KIT. Essas observações demonstram que a regulação positiva do RUNX1 é um evento comum em amostras de CRC e está intimamente correlacionada com metástases do câncer e que RUNX1 promove EMT de células CRC ativando a sinalização de Wnt/ β -catenina. Além disso, o RUNX1 é regulado pela Wnt/ β -catenina (Li et al., 2019). Observa-se ele positivamente regulado na linhagem celular SW620 associado a DNMT1 na rede PPI colaborando com os resultados da literatura, também está associado na via de câncer do KEGG onde outros fatores como WNT5A estão presentes como um todo nos genes que fazem parte dessa via.

Em células e tecidos CRC, a expressão de circ_0040809 e DNMT1 está acentuadamente aumentada, enquanto a expressão de miR-515-5p está diminuída. Além disso, a expressão alta do circ_0040809 está significativamente ligada a uma sobrevida global mais curta. Os experimentos de compensação da função celular revelam que o silenciamento do circ_0040809 inibe a proliferação e migração das células CRC e promove a apoptose, enquanto a superexpressão circ_0040809 tem os efeitos opostos. Mecanicamente, circ_0040809 se liga competitivamente a miR-515-5p para elevar a expressão de DNMT1. O ensaio de resgate revela que o miR-515-5p superexpresso neutraliza parcialmente o impacto de facilitação do tumor circ_0040809 (Mao et al., 2021). Nas análises apresentadas,

circ DNMT1 (hsa_circ_28013) está *up-regulated* na linhagem SW620, podendo aumentar as chances de proliferação e migração das células tumorais. Ele encontra-se associado em diversos termos do GO com RUNX1 também superexpresso na linhagem metastática.

Por análise de uma matriz de expressão de tumor / tecido normal, foi descrito que a expressão de FOXP1 era menor nos tumores do cólon e maior nos tumores do estômago e da próstata em comparação com os tecidos normais correspondentes (Banham et al., 2001). Em outro estudo, pela análise imuno-histoquímica de um microarray de tecido difuso de linfoma de células B grandes (BCL), pacientes não tratados com uma alta porcentagem de núcleos positivos para FOXP1 tiveram redução significativa da sobrevida e progressão precoce em comparação com pacientes negativos para FOXP1 (Banham et al., 2005) (OMIM 605515). FOXP1 é um supressor de tumor no câncer de cólon. No entanto, circ FOXP1 derivado de FOXP1 é um oncogene. Em um artigo, os autores investigaram o papel de circ FOXP1 no câncer de cólon e o mecanismo regulatório entre circ FOXP1 e Foxp1. A expressão de circ FOXP1 foi aumentada e Foxp1 foi reduzida em tecidos de câncer de cólon, que foram associados a uma baixa taxa de sobrevida geral dos pacientes com câncer de cólon. O circ FOXP1 recrutou DNMT1 para o promotor de Foxp1, levando à hipermetilação do promotor, inibindo assim a transcrição de Foxp1. O circ FOXP1 pode regular a metilação e desmetilação do DNA para coordenar a proliferação das células cancerígenas do cólon e participar nas respostas aos medicamentos de quimioterapia. Assim, circ FOXP1 pode ser um alvo terapêutico potencial para o câncer de cólon (Luo et al, 2020). Na linhagem SW620, circ FOXP1 (hsa_circ_11346)

está subexpresso, podendo estar associado a uma tentativa de não progressão tumoral na linhagem metastática e melhor sobrevida de acordo com a literatura.

O carcinoma epidermóide oral (CECO) é um dos tipos mais comuns de câncer de cabeça e pescoço. No entanto, os mecanismos subjacentes à patogênese do CECO permanecem obscuros. Em um estudo de 2021, verificou-se que circ MTO1 foi regulado positivamente em células e tecidos tumorais CECO. Verificou-se que o knockdown de circ MTO1 inibiu a proliferação, migração e invasão de células CECO. Além disso, os autores conduziram uma série de experimentos para elucidar os mecanismos subjacentes. Um novo eixo circ MTO1 / miR-320a / ATRX foi identificado. Os resultados sugeriram que circ MTO1 modulou a expressão de ATRX para acelerar a progressão de CECO por esponja miR-320a (Zou et al, 2021). Em outro estudo, os autores encontraram mudanças extensas na acessibilidade da cromatina em regiões repetitivas do DNA e regiões regulatórias não repetitivas após a perda de ATRX. Essas mudanças estavam altamente correlacionadas com as mudanças na transcrição, que levavam a alterações nas vias de sinalização relacionadas ao câncer, como a regulação positiva da via do TGF- β e a regulação negativa da família de proteínas da caderina. Esses achados indicaram que a deficiência de ATRX induziu mudanças epigenéticas e promoveu a tumorigênese por meio da instabilidade do genoma e mudanças na transcrição (Junbo et al., 2019). Na análise apresentada, o circ ATRX (hsa_circ_0140609) encontra-se sub regulado na SW620, podendo estar associado a uma maior taxa de progressão tumoral no CRC.

Em um estudo de 2020, os autores discutem sobre a histona 3 lisina 4 (H3K4) monometilase KMT2C que sofre mutação em vários tipos de câncer, no entanto, os efeitos das mutações na organização do epigenoma, na expressão gênica e no crescimento celular não são claros. Uma mutação frequentemente recorrente no câncer colorretal (CRC) com instabilidade de microssatélites é uma única deleção de nucleotídeo dentro da repetição do exon 38 poli-A (9) (c.8390delA). Para estudar os efeitos da expressão de KMT2C em células CRC, eles restauraram um alelo para KMT2C de tipo selvagem nas duas linhas de células de CRC RKO e HCT116, que são mutantes c.8390delA homozigotos. A restauração de KMT2C reduziu o crescimento de células CRC e reforçou a deposição de H3K4me1 em todo o genoma em potenciadores, no entanto, os efeitos variaram dependendo do status de H3K4me1 das células deficientes em KMT2C. Os resultados indicam que a inativação de KMT2C pode promover o desenvolvimento de câncer colorretal por meio da desregulação da transcrição em várias vias com relevância conhecida para o câncer (Larsson et al., 2020). As análises apresentam circ KMT2C subexpresso na linhagem celular SW620, podendo levar a ideia de uma progressão tumoral facilitada.

Os perfis de expressão e funções de circRNAs no carcinoma hepatocelular (HCC) são amplamente obscuros. Neste artigo publicado, circ SETD3 (hsa_circRNA_0000567/ hsa_circRNA_101436) foi significativamente regulado para baixo em tecidos e linhas celulares de HCC. A baixa expressão de circ SETD3 em tecidos de HCC previu significativamente um prognóstico desfavorável e foi correlacionada com tamanho maior do tumor e diferenciação pobre de HCC em pacientes. Experimentos *in vitro* mostraram que circ SETD3 inibiu a proliferação

de células HCC e induziu a parada G1 / S em células HCC. Estudos *in vivo* revelaram que circ SETD3 foi superexpresso de forma estável em um modelo de xenoenxerto de camundongo e inibiu o crescimento de HCC. Além disso, foi demonstrado que circ SETD3 atua como uma esponja para miR-421 e verificou-se que a proteína quinase ativada por mitogênio (MAPK) 14 é um novo alvo de miR-421 (Xu et al., 2019). Em uma revisão, circ SETD3 (hsa_circ_0000567), foi encontrado para ser significativamente regulado para baixo em tecidos CRC em comparação com aqueles em tecidos não tumorais, esta regulação negativa foi intimamente correlacionada com o tamanho do tumor, metástases linfáticas e metástases distais no CRC (Hao et al., 2019). Estando em similaridade encontrada na literatura, o circ SETD3 (hsa_circ_13006) apresentado nas análises, está subexpresso na linhagem SW620, colaborando com a progressão tumoral assim como os demais circRNAs encontrados associados a ele nas redes PPI como ATRX, FOXP1 e KMT2C.

O circ UBE3C já citado anteriormente para a linhagem SW480, também está expresso na linhagem SW620, mas *down-regulated*. Em um estudo sobre o carcinoma de células renais (CCR) o qual apresenta um prognóstico extremamente ruim devido ao alto potencial de crescimento tumoral, invasão vascular, metástase e recorrência, os autores demonstraram que a proteína UBE3C pode ser um fator determinante para o crescimento e metástase de CCR. A expressão de UBE3C foi aumentada em tecidos CCR em comparação com tecidos normais adjacentes. Pacientes com CCR com alta expressão da proteína UBE3C em tumores foram associados com sobrevida pós-operatória significativamente pior. A destruição da expressão de UBE3C em células ACHN

inibiu a proliferação celular, migrações e invasividade *in vitro*, enquanto a superexpressão de UBE3C em células 786-O exerceu os efeitos opostos. UBE3C regulou positivamente os níveis de proteína β -catenina e promoveu o acúmulo nuclear de β -catenina, levando à ativação da via de sinalização Wnt/ β -catenina em células CCR. Essas observações sugerem que o UBE3C desempenha um papel importante no desenvolvimento e progressão do CCR (Wen et al., 2015). Na linhagem SW480, ele parece estar associado a progressão tumoral, porém na linhagem SW620, ele encontra-se *down-regulated*, como circ FOXP1, podendo estar associado com uma não progressão do tumor metastático.

Conclusão

Na comparação entre metodologias de preparo de bibliotecas de *RNA-sequencing*, o preparo específico utilizando RNase R para detecção de circRNAs foi mais eficiente na chamada dos circRNAs do que o preparo padrão da biblioteca, comprovando a nossa hipótese.

Existem diversos algoritmos de detecção para circRNAs provenientes de experimentos de *RNA-Seq* comum ou específico com tratamento de RNase R, mas poucos mostraram eficiência na detecção correta dos RNAs circulares e isso continua sendo uma área em desenvolvimento e melhoramento contínuo. A escolha do circExplorer mostrou-se com mais resultados confiáveis de acordo com os dados obtidos das análises que colaboram com o descrito na literatura e as revisões disponíveis de comparação publicadas de diversos autores comparando-o a outros algoritmos de detecção. A desvantagem de se usar o circExplorer se mostrou apenas no tempo de processamento e capacidade da máquina que deve-se ter.

Os algoritmos de estimativa de transcritos escolhidos para essa análise, já bem descritos na literatura, Kallisto e RSEM, embora utilizem metodologias diferentes para contagem, apresentaram resultados similares para análise da expressão dos circRNAs.

Nas análises de expressão diferencial realizadas, pode-se perceber um papel importante dos circRNAs TGM2, TSPAN4 (*hsa_circ_18995*), WNT5A, GLI3

(hsa_circ_08687), SMURF2 e UBE3C (hsa_circ_12017) na linhagem SW480 do tumor primário.

Alguns circRNAs parecem estar envolvidos em processos biológicos na tentativa de influenciar na inibição da progressão tumoral.

A regulação negativa de circ TGM2 pode inibir a angiogênese e a expressão de β -catenina influenciando positivamente na não progressão do CRC. Pode-se hipotetizar, que o circ TGM2 pouco expresso também pode estar influenciando na via Wnt do KEGG, que se encontra *down* regulada, com circ WNT5A e circ NFAT, suprimindo essa via, e colaborando na supressão da progressão tumoral. O circ WNT5A quando pouco expresso, também pode inibir a tumorigênese. A menor expressão do circ GLI3 pode colaborar nesse processo de supressão de tumor, já que quando este apresenta uma expressão elevada, está associado a pior prognóstico dos pacientes com CRC. O *knockdown* do circ TSPAN4 pode atenuar dramaticamente a capacidade metastática das células, ele se mostrou *down-regulated* na SW480.

Em outra análise, alguns circRNAs podem estar envolvidos na progressão tumoral regulando ela positivamente.

O circ UBE3C (hsa_circ_12017) na linhagem SW480 encontra-se *up-regulated*, assim ele pode ser um oncogene candidato envolvido no desenvolvimento e progressão do tumor. Já o circ SMURF2 está *up* e *down-regulated* na SW480, assim, ele pode desempenhar papéis distintos dependendo da via que se correlaciona. Ele pode agir como um supressor de tumor, porém, quando está *up-regulated*, embora as evidências estejam fracas,

ele pode atuar junto na via da ubiquitinação com circ UBE3C (hsa_circ_12017), facilitando a progressão tumoral. Quando *down-regulated*, ele pode atuar na via Wnt seguindo a interação PPI com WNT5A, como um supressor de tumor, podendo se relacionar com os demais circRNAs como TGM2 e GLI3.

Nas análises de expressão diferencial realizadas, pode-se perceber um papel importante dos circRNAs DNMT1 (hsa_circ_28013), RUNX1, UBE3C (hsa_circ_12017), SETD3 (hsa_circ_13006), KMT2C, ATRX (hsa_circ_0140609), FOXP1 (hsa_circ_11346) e ST18 (hsa_circ_13413) na linhagem SW620, em um possível papel epigenético dos circRNAs na progressão tumoral do CRC.

É possível observar uma via geral com circRNAs os quais estão sub expressos e podem influenciar no avanço do CRC, como o circ ST18, sub expresso pode corroborar com a progressão tumoral e metástase, o circ ATRX, onde sua deficiência pode induzir mudanças epigenéticas e promover a tumorigênese por meio da instabilidade do genoma e mudanças na transcrição, circ KMT2C *down-regulated* pode promover o desenvolvimento de câncer colorretal por meio da desregulação da transcrição em várias vias com relevância conhecida para o câncer e o circ SETD3 encontrado para ser significativamente regulado para baixo em tecidos CRC em comparação com aqueles em tecidos não tumorais, pode estar correlacionado com o avanço no tamanho do tumor, metástases linfáticas e metástases distais no CRC.

Há ainda os circRNAs *up-regulated*, como o circ DNMT1, que quando superexpresso colabora na proliferação e migração das células tumorais, e circ RUNX1, *up-regulated*, ajuda no processo de metástase.

Embora encontremos diversos papéis dos circRNAs relacionados à manutenção e avanço da progressão tumoral, há aqueles circRNAs *down-regulated* que podem estar associados a uma resposta de *break* da progressão, como circ UBE3C e circ FOXP1. Com a destruição da expressão de UBE3C, pode haver uma inibição da proliferação celular, da migração e invasividade das células tumorais. O circ FOXP1 quando *down-regulated* pode aumentar a sobrevida dos pacientes e colaborar para uma progressão tumoral mais lenta.

Importantes vias e perfis de expressão para esclarecimento do envolvimento dos circRNAs no câncer colorretal foram detectadas. Uma análise mais profunda sobre correlação com fatores de transcrição e miRNAs precisa ser feita para um melhor entendimento do mecanismo de ação desses circRNAs. A literatura disponível sobre os circRNAs e seu papel na regulação e manutenção tumoral é escassa e com poucas evidências. Pretendemos usar dados de sequenciamento dessas mesmas linhagens disponíveis em nosso laboratório para correlacionar as ações e implicações da expressão dos circRNAs na progressão tumoral do CRC, além de validar os alvos descritos em experimentos subjacentes como qRT-PCR e verificar a eficiência do *pipeline* aplicado nas análises apresentadas.

Anexo 1: Relatório FASTQC



Thu 15 Mar 2018
SW480-Circ-BC4_S2_L001_R1_001.fastq.gz

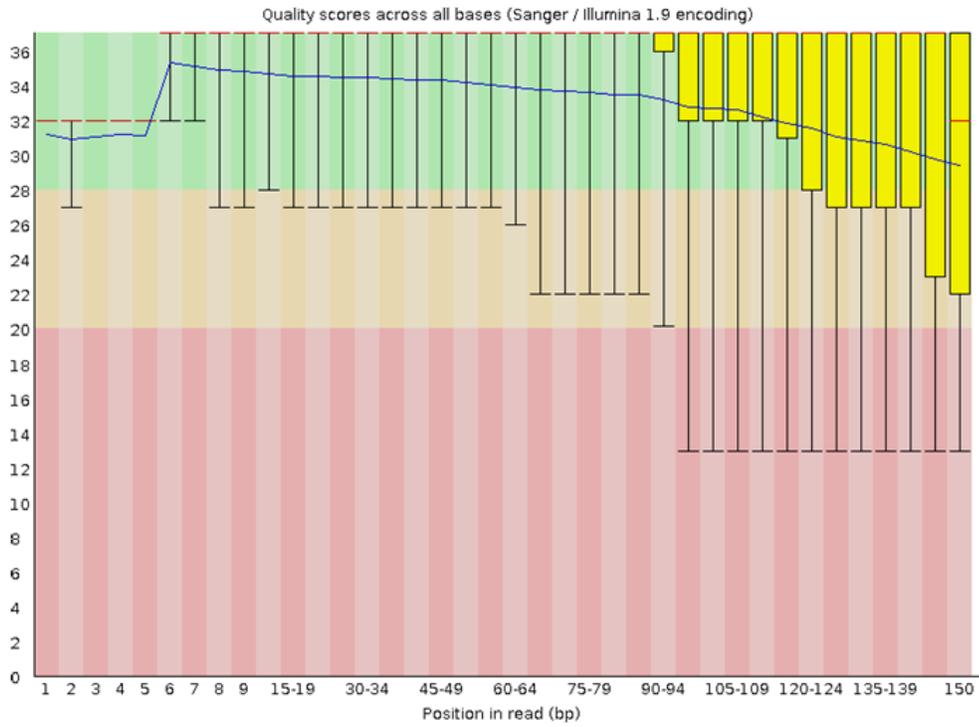
Summary

-  Basic Statistics
-  Per base sequence quality
-  Per tile sequence quality
-  Per sequence quality scores
-  Per base sequence content
-  Per sequence GC content
-  Per base N content
-  Sequence Length Distribution
-  Sequence Duplication Levels
-  Overrepresented sequences
-  Adapter Content
-  Kmer Content

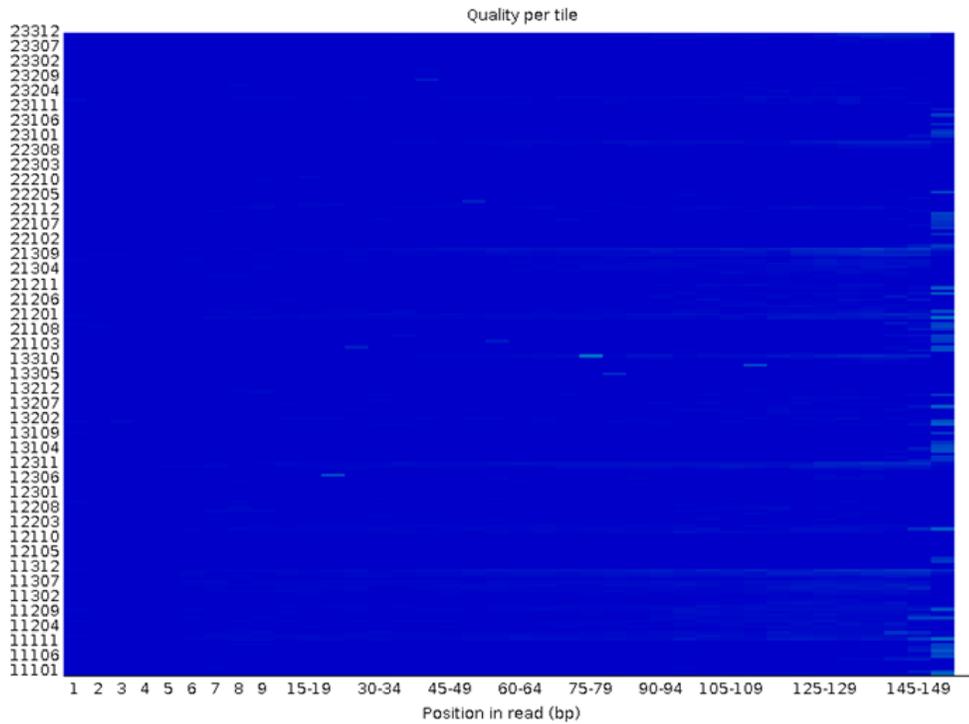
Basic Statistics

Measure	Value
Filename	SW480-Circ-BC4_S2_L001_R1_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	22815504
Sequences flagged as poor quality	0
Sequence length	35-150
%GC	57

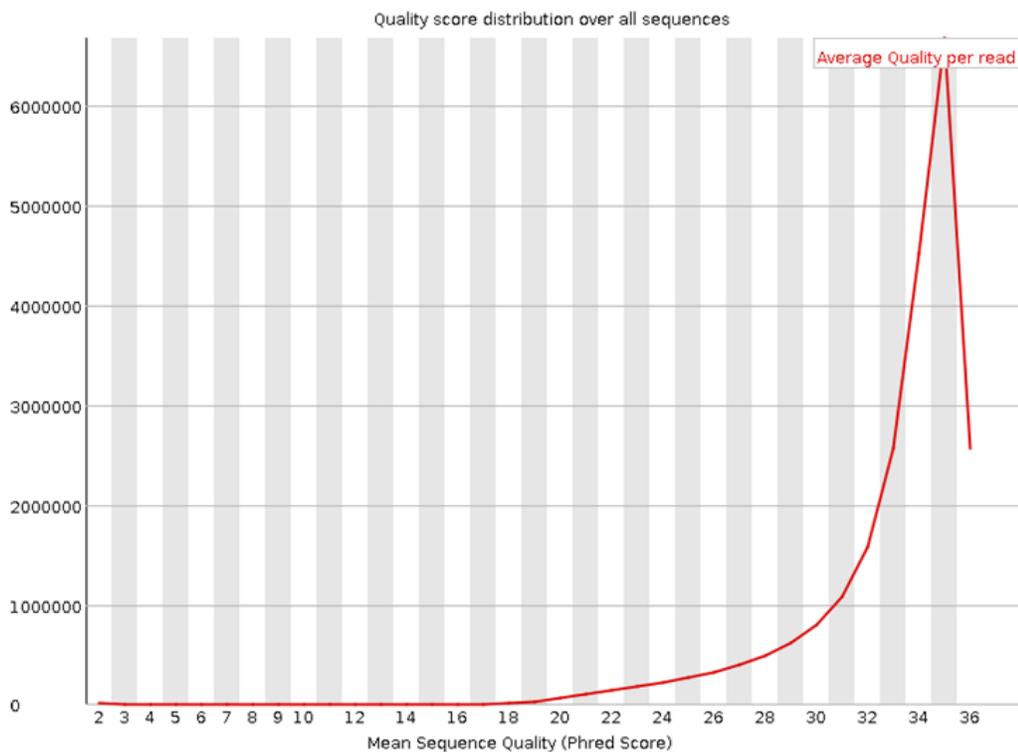
✔ Per base sequence quality



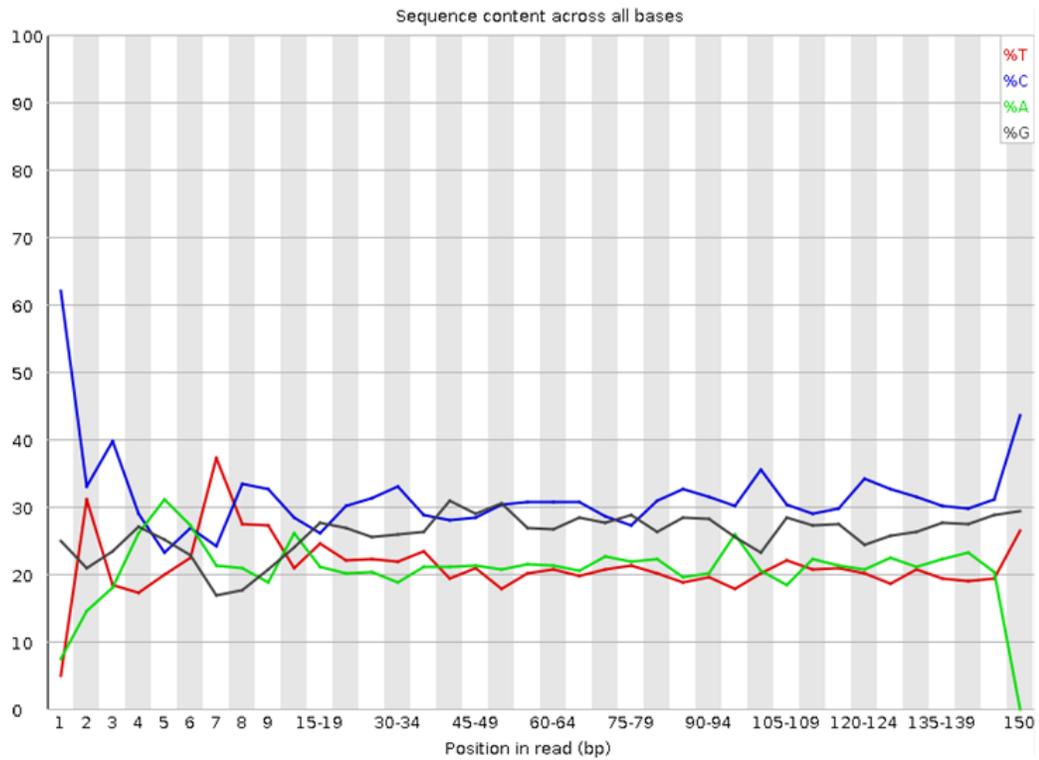
✔ Per tile sequence quality



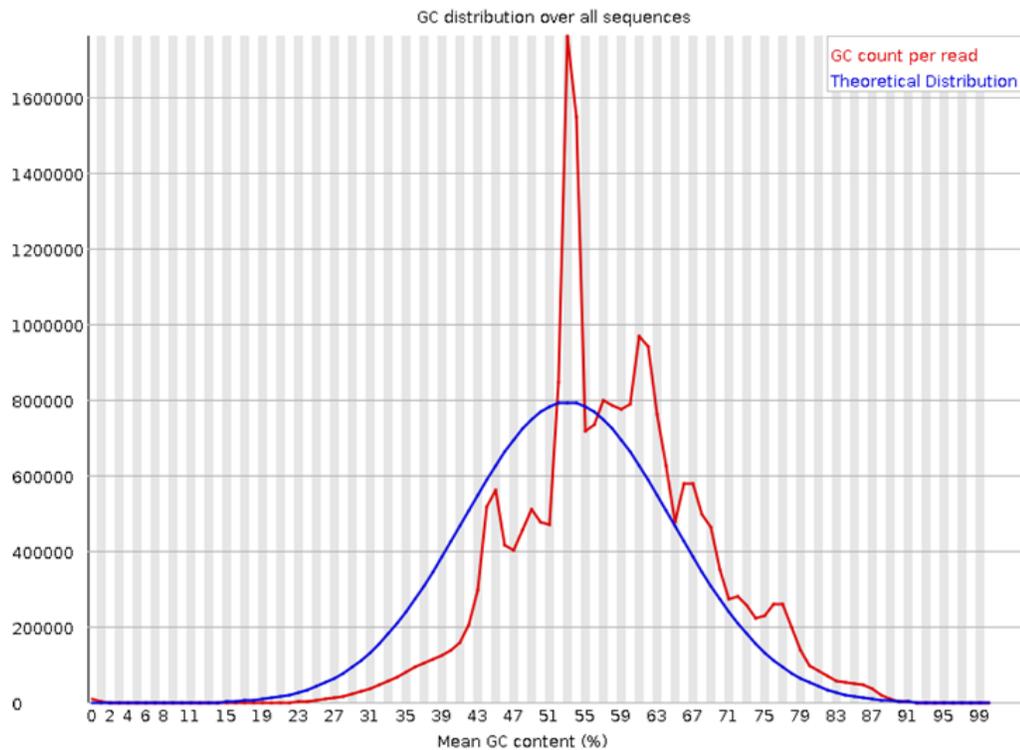
✔ Per sequence quality scores



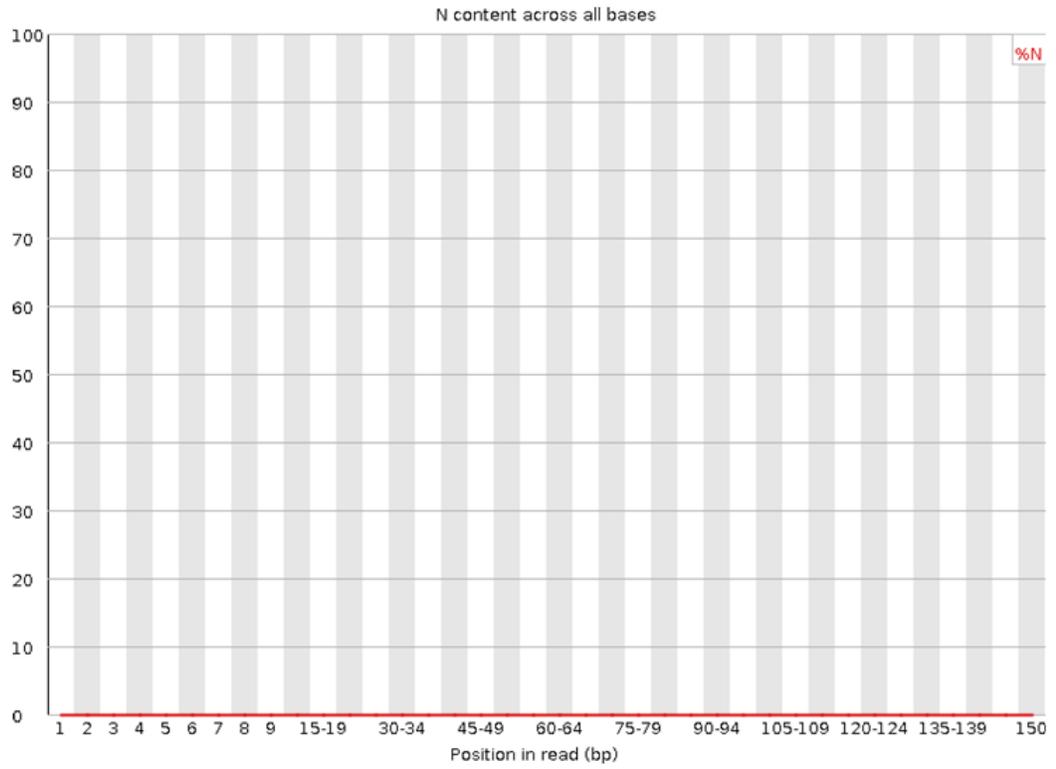
✖ Per base sequence content



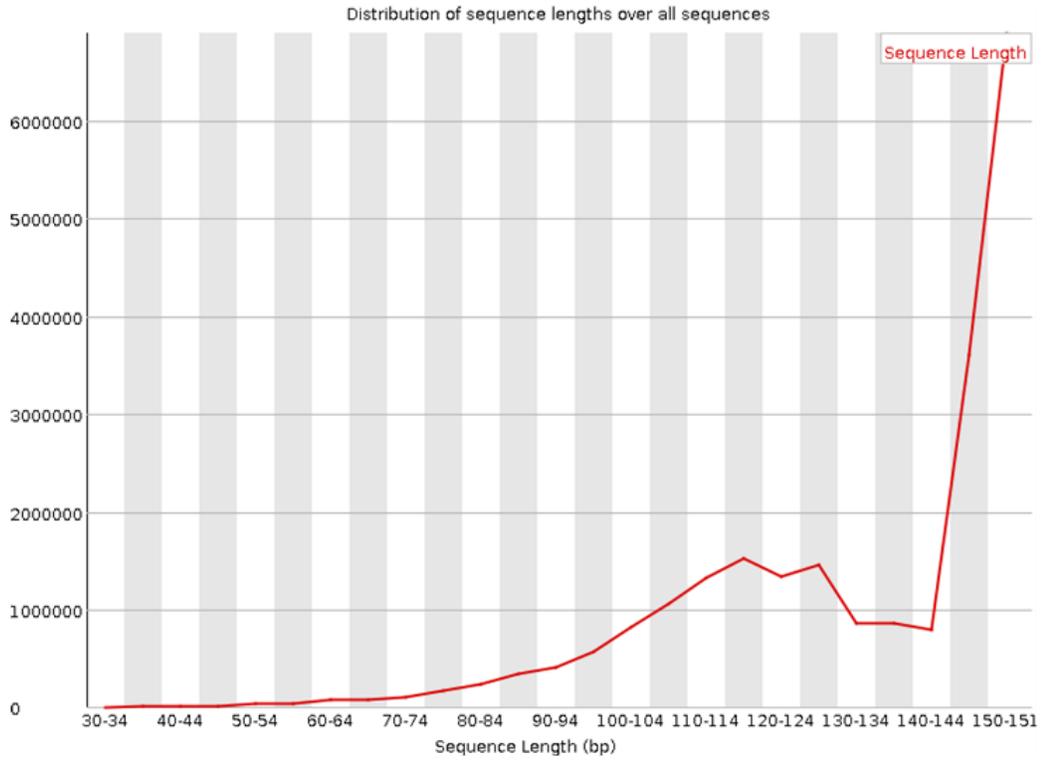
✖ Per sequence GC content



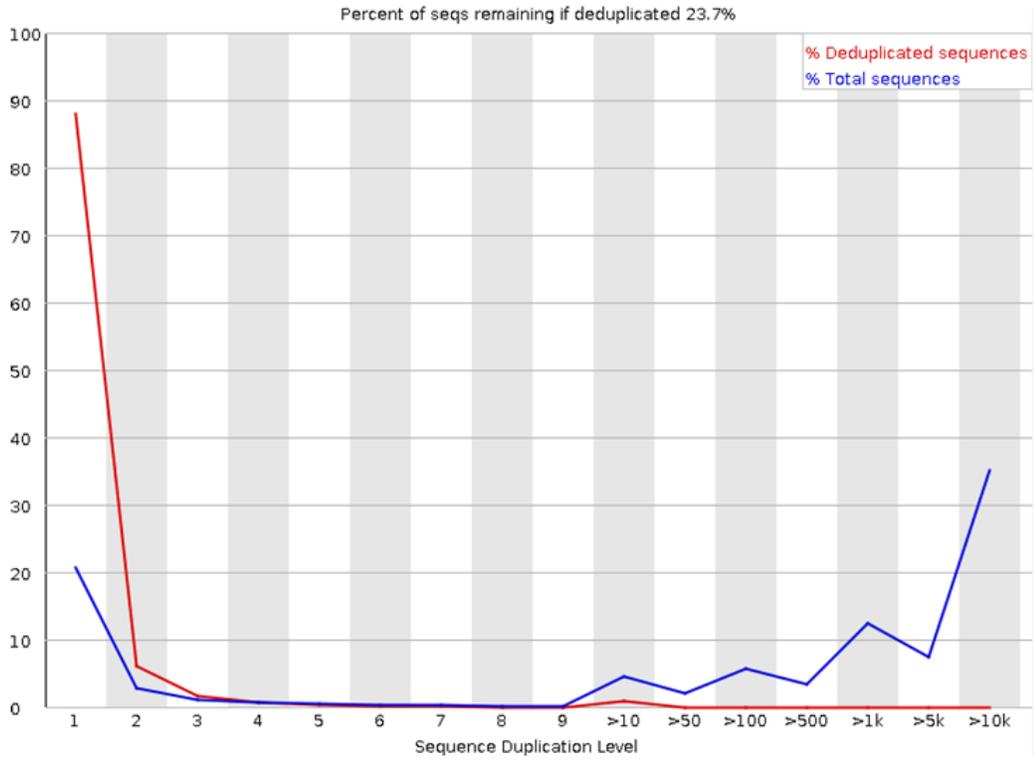
✔ Per base N content



Sequence Length Distribution



Sequence Duplication Levels



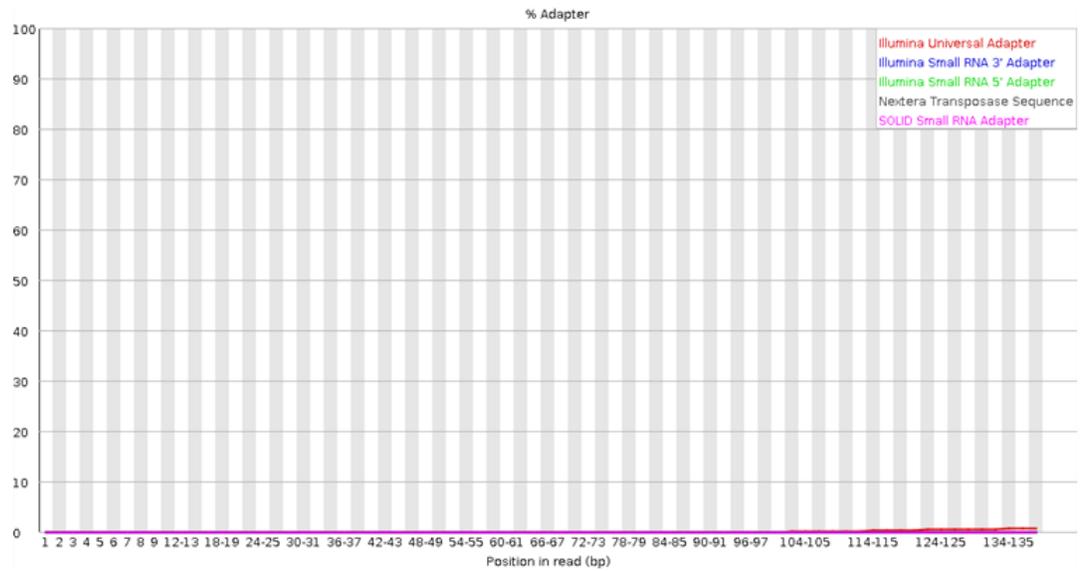
Overrepresented sequences

Sequence	Count	Percentage	Possible Source
CCCCTACCACAAATTATGCAGTCGAGTTTCCACATTTGGGAAATCGC	599474	26.274.852.398.614.500	No Hit
CCCCACTACCACAAATTATGCAGTCGAGTTTCCACATTTGGGAAATCG	493632	2.163.581.396.229.510	No Hit
CACTACCACAAATTATGCAGTCGAGTTTCCACATTTGGGAAATCGCAG	286182	12.543.312.652.659.300	No Hit
CCACAAATTATGCAGTCGAGTTTCCACATTTGGGAAATCGCAGGGGTC	258926	11.348.686.402.018.500	No Hit
CCTTAGGCAACCTGTGGTCCCCGCTCCCGGGAGGTCACCATATTGATG	235425	1.031.864.121.870.810	No Hit
CCACTACCACAAATTATGCAGTCGAGTTTCCACATTTGGGAAATCGCA	205985	0.9028290586962269	No Hit
CACAAATTATGCAGTCGAGTTTCCACATTTGGGAAATCGCAGGGTCA	186391	0.8169488607396094	No Hit
CTTAGCAACCTGGTGTCCCCGCTCCCGGGAGGTCACCATATTGATGC	153776	0.6739978218320314	No Hit
CTTGAACCTCTCTTCAAAGTTCTTTTCAACTTTCCCTACGGTACTTGT	132256	0.5796759957614788	No Hit
COGGTATTAGCCTTAGAGTGGAGTTTACCACCGCTTGGGCTGCATTCC	124613	0.5461768453591909	No Hit
CTACCACAAATTATGCAGTCGAGTTTCCACATTTGGGAAATCGCAGGG	108846	0.4770703290183728	No Hit
GTGCGCATGTATTAGCTCTAGAATTACCACAGTTATCCAAGTAGGAGAGG	101569	0.4451753509368016	No Hit
CTTCCGTACGCCACATGTCCCGCGCCCGCGCGGGCGGGATTCCGCG	97844	0.42884873373825094	No Hit
CGCCACTCCGGATTCCGGGATCTGAACCCGACTCCCTTCGATCGGCCGA	96166	0.42149408577605824	No Hit
GCCCTCTTGAACCTCTCTTCAAAGTTCTTTTCAACTTTCCCTACGGTA	81959	0.35922502522845867	No Hit
CTCGTTCGACCTGGGCCGTTTCAACCCCTCTTAGGCAACCTGGTGGT	79092	0.3466590087161782	No Hit
GTCCCCCACTACCACAAATTATGCAGTCGAGTTTCCACATTTGGGAAA	73708	0.32306102026060873	No Hit
CTGGA TAGTAGG TAGGGACAGTGGGAATCTCGTTCA TCCATTCATGCGCG	66022	0.2893734015255591	No Hit
CTAACACGTGCGCTCGTCTCCACCTCCCGCGCGCGGGGCGAGACGGG	64534	0.28285152061510455	No Hit
CCCGAAGTTACG GATCCGGCTTGCCGACTTCCCTTACCTACATTGTTCCA	63187	0.2769476405167293	No Hit
GTGAGTTTCCACATTTGGGAAATCGCAGGGGTCAGCACATCCGGAGT	62239	0.27279257122700423	No Hit
CCCTCCTTAGGCAACCTGGTGTCCCCGCTCCCGGAGGTCACCATAT	56448	0.24741070808692195	No Hit
CGGGCTTCTTACCATTTAAAGTTTGAGAA TAGGTTGAGATCGTTTCG GC	54736	0.23990703865231292	No Hit
GGCAGACGTTGAA TGGGTGTCGCGCCACGGGGGGCGTGCATCGGCC	54123	0.23722026916433667	No Hit
CTTAGAGCCAATCCTTATCCCGAAGTTACGGATCCGGCTTGCCGACTTCC	52233	0.2289364284917835	No Hit
CTCTGAACTCTCTTCAAAGTTCTTTTCAACTTTCCCTACGGTACTT	51168	0.224268550017567	No Hit
GTCTGTTCCAAGCTCCGGCAAAGGAGGCATCCGCGGGCCCCTCCCGAA	50493	0.22131003549165515	No Hit
ACTACCACAAATTATGCAGTCGAGTTTCCACATTTGGGAAATCGCAGG	47000	0.20600027069312166	No Hit
CTGCTGCCCTTCTGGATGTGGTAGCGTTTCTCAGGCTCCCTCTCCGGA	46176	0.20238869147926777	No Hit
GGCATGTATTAGCTCTAGAATTACCACAGTTATCCAAGTAGGAGAGGAGC	46004	0.20163481814822062	No Hit

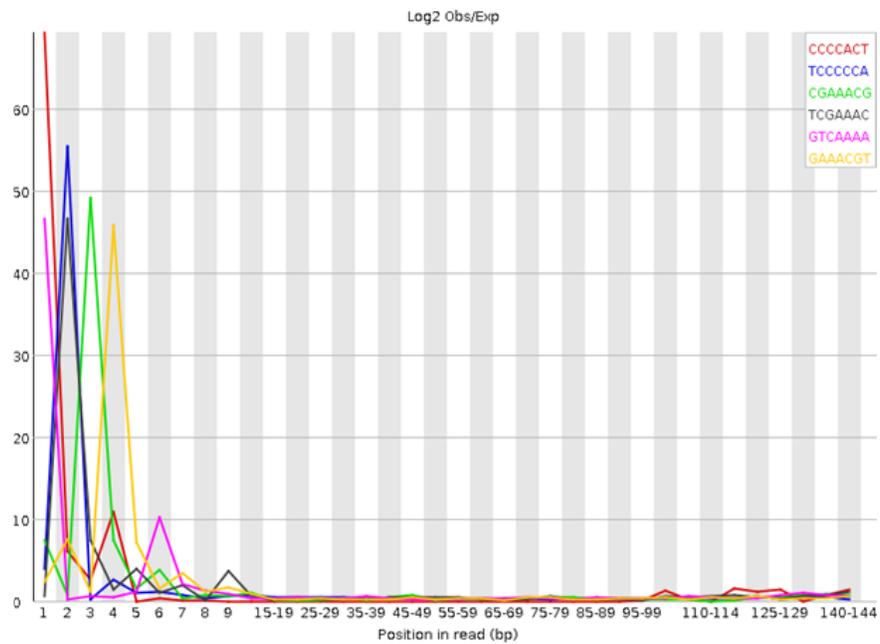
ACCACAAATTATG CAGTCGAGTTTCCACATTTGGGAAATCGCAGGGGT	45846	0.20094230659993306	No Hit
CCTCCTTAGGCAACCTGGTGTCCCCGCTCCCGGA GGTCACCATATTG	45479	0.1993337512947336	No Hit
CCCTCCTTAGGCAACCTGGTGTCCCCGCTCCCGGAGGTCACCATATT	44638	0.19564766134467157	No Hit
CGGTATTTAGCCTTAGATGGAGTTTACCACCCGCTTTGGCTGCATTCCC	43956	0.1926584659273799	No Hit
CTCCTTAGGCAACCTGGTGTCCCCGCTCCCGGAGGTCACCATATTGA	43064	0.18874884376869344	No Hit
CCGTTTCCGACCTGGGCCGTTCAACCCTCCTTAGGCAACCTGGTGTCC	42166	0.18481292370310995	No Hit
CCCCACTACCACAAATTATG CAGTCGAGTTTCCACATTTGGGAAATC	41456	0.181701004720299	No Hit
CTGCTGTCTATATCAACCAACACCTTTTCTGGGTCTGATGAGCGTCGGC	39992	0.17528431543743236	No Hit
CCACTCTGACTGCCGGCGACGGCCGGTATGGCCCGACGCTCCAGCGC	39282	0.17217239645462137	No Hit
CGTACGCCACATGTCCCGCGCCCGCGCGGGCGGGGATTCCGGCGCTG	38923	0.17059890502528458	No Hit
CTGACTGCCGGCGACGGCCGGTATGGGCCGACGCTCCAGCGCCATCC	38272	0.16774558212696067	No Hit
CCCGGGGCTCCCGCCGCTTCCCGGATCGGTCGCTTACCGCACTGGA	38112	0.1670443046097075	No Hit
CGTGAGTCTGTTCCAAGCTCCGGCAAAGGAGGCATCCGCGGGCCCTCC	36859	0.16155242505271852	No Hit
GGCAACCTGGTGTCCCCGCTCCCGGAGGTCACCATATTGATGCCGAA	35730	0.15660403557160077	No Hit
CTGATGAGCGTCGGCATCGGGCCCTTAACCCGCGGTTCCGTTTATCCCG	35132	0.15398301085086702	No Hit
CCTAGATGGAGTTTACCACCCGCTTTGGGCTGCATTCCAAGCAACCCGA	34399	0.1507702832249509	No Hit
CTGGAACTCACCTCCCGAAGCTCAGGGAGAGCCCTGTTAGG GCCGCCT	32938	0.14436674289553278	No Hit
GTTGGTTTTGATCTGATAAATGCACGCATCCCCCGCGAAGGGGGTCAG	32290	0.14152656895065743	No Hit
GTTTCGACCTGGGCCGGTTCAACCCTCCTTAGGCAACCTGGTGTCCCC	32289	0.14152218596617458	No Hit
CCGGGGCTCCCGCGCTTCTCCGGGATCGGTCGCGTTACCGCACTGGAC	31869	0.139681332483385	No Hit
CTTACCTACATTGTTCCAACATGCCAGAGGCTGTTACCTTGGAGACTG	31765	0.1392255020971704	No Hit
CGGATAAACTGCGTGCGGGGGTGCCTCGGTCTGCGAGAGCGCCAGCT	30607	0.13415000606605054	No Hit
CTCACTCGCTGGCGTGAGTCTGTTCCAAGCTCCGGCAAAGGAGGCATC	30590	0.13407549532984236	No Hit
CCCAGGCATAGTTACCATCTTTCGGGTCTAACACGTGCGCTCGTGCTC	30245	0.13256336568326518	No Hit
CAGGCTGGAGTGCAGTGGCTATTCACAGGCGGATCCCACTACTGATCAG	30037	0.13165170491083608	No Hit
TGAGCGTCGGCATCGGGCGCCTTAACCCGCGGTTCCGTTTATCCCGCAGC	29890	0.13100740619185972	No Hit
GTTTTGATCTGATAAATGCACGCATCCCCCGCGAAGGGGGTCAAGCGCC	29353	0.12865374352457873	No Hit
CTGGGCTCTCCCTGTTCACTCGCCGTTACTGAGGGAATCCTGTTAGTT	29328	0.12854416891250792	No Hit
CTCGCATCCACGCCCGCTCCACGCCAGCGAGCCGGGCTTCTTACCCAT	29287	0.12836446654871178	No Hit
CCGAGGTTATCTAGAGTCAACAAAGCCGCGGGCGCCCGCCCGGCCGG	29226	0.12809710449525902	No Hit
CCGAAAGTTACGGATCCGGCTTCCGACTTCCCTTACCTACATTGTTCAA	27733	0.12155330866239027	No Hit
CTCGATCAGAAGGACTTGGCCCGCCACGAGCGCGCGGGAGCGGGTC	27675	0.121299095562386	No Hit
GATAAACTGCGTGGCGGGGTGCGTGGGCTGCGAGAGCGCCAGCTAT	27656	0.12121581885721218	No Hit

COGGGATCGGTGCGTTACCGCACTGGACGCCTCGCGGCGCCCATCTCCG	27495	0.12051015835547617	No Hit
AGACGTTGGAATGGGTCGTGCGCCACGGGGGGCGTGCGATCGGCCCGA	27155	0.11901994363131316	No Hit
CTGGAGTGCAGTGGCTATTACAGGCGCGATCCCACTACTGATCAGCAAG	26814	0.11752534592266732	No Hit
CCAGGCTGGAGTGCAGTGGCTATTACAGGCGCGATCCCACTACTGATCA	26275	0.11516291728642068	No Hit
CAGTTATCCAAGTAGGAGAGGAGCGAGCGACCAAAGGAACCATAACTGAT	25980	0.11386993686398512	No Hit
GGATAAACTGCGTGGCGGGGGTGCCTGCGGTCTGCGAGAGCGCCAGCTA	25804	0.11309853159500662	No Hit
GTCTGACCTCGCGCGGAGCCCGTTCTCTGGAACTCACCTCCCGAAGC	25665	0.11248929675189293	No Hit
CACAGTTATCCAAGTAGGAGAGGAGCGAGCGACCAAAGGAACCATAACTG	25648	0.11241478601568478	No Hit
CGGGGCTCCCGCGGCTTCTCGGGATCGGTGCGTTACCGCACTGGACG	25599	0.11220001977602599	No Hit
COGCCACTCCGATTCGGGGATCGAACCCGACTCCCTTTCGATCGGCCG	25524	0.11187129593981356	No Hit
CCCTCTGAACTCTCTCTCAAAGTCTTTTCAACTTCCCTTACGGTAC	25414	0.11138916764670198	No Hit
CCCGCTTTGGGCTGCATTCCCAAGCAACCCGACTCCGGGAAGACCCGGGC	25307	0.11092018830703894	No Hit
CGCGTAACTAGTTAGCATGCCAGAGTCTCGTTGTTATCGGAATTAACCA	25285	0.11082376264841662	No Hit
CTGAATTGGGTTATGAGGTCCTCGCGGGTACCTCACTCAGCCATTGA	25247	0.11065720923806897	No Hit
TGATGAGCGTCGGCATCGGGCGCCTTAACCCGGCGTTGCGTTCATCCCGC	24536	0.10754090727077517	No Hit
COGGATTCGGGGATCTGAACCCGACTCCCTTTCGATCGGCCGAGGGCAAC	24462	0.10721656641904558	No Hit
CTCGTGCCGGTATTTAGCCTTAGATGGAGTTTACCACCCGCTTTGGGCTG	24342	0.10669060828110569	No Hit
GTGATGCGTGGAGTGGACGGAGCAAGCTCCTATTCCATCTCCTATTTCC	23967	0.10504698910004356	No Hit
CGGATTTCAAGGGCCAGCGAGAGCTACCCGACGCCGCGGAAACCGCGA	23849	0.10452979693106933	No Hit
CGGGTCTTCGTA CGCCACATGTCCCGCGCCCGCGCGGGGCGGGGATT	23727	0.10399507282416377	No Hit
COGGATTTCAAGGGCCAGCGAGAGCTACCGGACGCCGCGGAAACCGCG	23005	0.10083055802755882	No Hit
CCCCCCTCACCCGGTCAAGTGA AAAACGATCAGAGTAGTGATTTTAC	22980	0.10072098341548799	No Hit

✓ Adapter Content



✗ Kmer Content



Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CCCCACT	101015	0.0	69.23689	1
TCCCCCA	19170	0.0	55.463863	2
CGAAACG	2710	0.0	49.21502	3
TCGAAAC	2750	0.0	46.686108	2
GTCAAAA	8720	0.0	46.671295	1
GAAACGT	3020	0.0	45.814205	4
CGTGCAA	3060	0.0	43.989624	8
CCACTAC	211660	0.0	40.875816	2
AATTGCC	11700	0.0	40.697105	5
AGGTCGA	10560	0.0	40.604828	2
CACTACC	218690	0.0	40.39683	3
AAATTGC	12070	0.0	39.9141	4
ATTGCCA	12210	0.0	39.150616	6
GGTCGAT	11735	0.0	37.92	3
ACCACAA	258720	0.0	36.74508	7
AAACGTG	3830	0.0	36.45065	5
CCCGCTA	9050	0.0	35.741432	1
CTACCAC	260050	0.0	35.54667	5
ACGTGCA	3855	0.0	35.40273	7
CCCACTA	244475	0.0	35.373905	1

Produced by FastQC (version 0.11.5)

Anexo 2: Relatório RNASEQC

Total Reads

Sample	Note	Total Purity Filtered Reads Sequenced	Alternative Alignments	Failed Vendor QC Check	Read Length	Estimated Library Size
SW480circlane1	sw480	39,560,458	149,622,909	NA	150	NA
SW480circlane2	sw480	37,987,089	147,278,579	NA	150	NA
SW480circlane3	sw480	35,079,903	132,645,636	NA	150	NA
SW480circlane4	sw480	34,749,886	142,885,837	NA	150	NA
SW480ribolane1	sw480	47,344,410	94,151,353	NA	150	NA
SW480ribolane2	sw480	45,852,108	92,440,362	NA	150	NA
SW480ribolane3	sw480	42,225,524	83,799,483	NA	150	NA
SW480ribolane4	sw480	43,585,602	86,617,825	NA	150	NA
SW620circlane1	sw620	41,986,237	150,284,579	NA	150	NA
SW620circlane2	sw620	41,876,597	146,215,403	NA	150	NA
SW620circlane3	sw620	38,358,688	124,875,515	NA	150	NA
SW620circlane4	sw620	39,613,179	140,258,862	NA	150	NA
SW620ribolane1	sw620	41,123,340	92,251,975	NA	150	NA
SW620ribolane2	sw620	39,632,144	88,294,270	NA	150	NA
SW620ribolane3	sw620	36,842,650	81,668,252	NA	150	NA
SW620ribolane4	sw620	37,770,646	84,064,121	NA	150	NA

Total Purity Filtered Reads Sequenced are filtered for vendor fail flags and exclude alternative alignment reads. **Alternative Alignments** are duplicate read entries providing alternative coordinates. **Failed Vendor QC Check** are reads which have been designated as failed by the sequencer. **Read Length** is the maximum length found for all reads. **Estimated Library Size** is the number of expected fragments based upon the total number of reads and duplication rate assuming a Poisson distribution.

Mapped Reads

Sample	Note	Mapped	Mapping Rate	Mapped Unique	Mapped Unique Rate of Total	Unique Rate of Mapped	Duplication Rate of Mapped	Base Mismatch Rate	rRNA	rRNA rate
SW480circlane1	sw480	39,560,458	1.000	39,560,458	1.000	1.000	NA	0.007	2,346,127	0.059
SW480circlane2	sw480	37,987,089	1.000	37,987,089	1.000	1.000	NA	0.008	2,256,868	0.059
SW480circlane3	sw480	35,079,903	1.000	35,079,903	1.000	1.000	NA	0.008	1,852,492	0.053
SW480circlane4	sw480	34,749,886	1.000	34,749,886	1.000	1.000	NA	0.008	370,855	0.011
SW480ribolane1	sw480	47,344,410	1.000	47,344,410	1.000	1.000	NA	0.007	4,477,839	0.095
SW480ribolane2	sw480	45,852,108	1.000	45,852,108	1.000	1.000	NA	0.007	4,161,410	0.091
SW480ribolane3	sw480	42,225,524	1.000	42,225,524	1.000	1.000	NA	0.008	3,931,920	0.093
SW480ribolane4	sw480	43,585,602	1.000	43,585,602	1.000	1.000	NA	0.008	4,062,991	0.093
SW620circlane1	sw620	41,986,237	1.000	41,986,237	1.000	1.000	NA	0.007	1,340,043	0.032
SW620circlane2	sw620	41,876,597	1.000	41,876,597	1.000	1.000	NA	0.008	3,198,873	0.076
SW620circlane3	sw620	38,358,688	1.000	38,358,688	1.000	1.000	NA	0.009	2,725,174	0.071
SW620circlane4	sw620	39,613,179	1.000	39,613,179	1.000	1.000	NA	0.008	3,001,312	0.076
SW620ribolane1	sw620	41,123,340	1.000	41,123,340	1.000	1.000	NA	0.007	4,710,730	0.115
SW620ribolane2	sw620	39,632,144	1.000	39,632,144	1.000	1.000	NA	0.007	4,597,796	0.116
SW620ribolane3	sw620	36,842,650	1.000	36,842,650	1.000	1.000	NA	0.008	4,275,027	0.116

Mapped reads are those that were aligned. **Mapping Rate** is per total reads. **Mapped Unique** are both aligned as well as non-duplicate reads. **Mapped Unique Rate of Total** is per total reads. **Unique Rate of Mapped** are unique reads divided by all mapped reads. **Duplication Rate of Mapped** is the duplicate read divided by total mapped reads. **Base Mismatch Rate** is the number of bases not matching the reference divided by the total number of aligned bases. **rRNA** reads are non-duplicate and duplicate reads aligning to rRNA regions as defined in the transcript model definition. **rRNA Rate** is per total reads.

Mate Pairs

Sample	Note	Mapped Pairs	Unpaired Reads	End 1 Mapping Rate	End 2 Mapping Rate	End 1 Mismatch Rate	End 2 Mismatch Rate	Fragment Length Mean	Fragment Length StdDev	Chimeric Pairs
SW480circlane1	sw480	18,554,805	NA	1.000	1.000	0.006	0.008	126	74	2,241,858
SW480circlane2	sw480	18,123,068	NA	1.000	1.000	0.007	0.009	126	72	2,077,450
SW480circlane3	sw480	16,803,872	NA	1.000	1.000	0.008	0.009	128	75	2,219,844
SW480circlane4	sw480	17,340,750	NA	1.000	1.000	0.007	0.010	125	71	2,127,427
SW480ribolane1	sw480	23,659,203	NA	1.000	1.000	0.006	0.008	136	138	7,206,625
SW480ribolane2	sw480	22,911,330	NA	1.000	1.000	0.006	0.008	137	140	6,958,894
SW480ribolane3	sw480	21,098,333	NA	1.000	1.000	0.007	0.009	137	139	6,474,946
SW480ribolane4	sw480	21,781,541	NA	1.000	1.000	0.006	0.009	137	140	6,609,745
SW620circlane1	sw620	20,743,969	NA	1.000	1.000	0.006	0.008	126	86	2,937,375
SW620circlane2	sw620	19,925,338	NA	1.000	1.000	0.007	0.009	126	84	3,027,832
SW620circlane3	sw620	18,321,063	NA	1.000	1.000	0.008	0.009	127	91	2,846,761
SW620circlane4	sw620	18,746,784	NA	1.000	1.000	0.006	0.010	127	79	2,794,649
SW620ribolane1	sw620	20,550,648	NA	1.000	1.000	0.006	0.007	133	130	7,096,639
SW620ribolane2	sw620	19,805,301	NA	1.000	1.000	0.006	0.008	133	132	6,983,949
SW620ribolane3	sw620	18,412,396	NA	1.000	1.000	0.007	0.009	134	134	6,370,038
SW620ribolane4	sw620	18,875,889	NA	1.000	1.000	0.006	0.009	133	131	6,551,873

Mapped Pairs is the total number of pairs for which both ends map. **Unpaired Reads** are the number of reads that are lacking a mate. **End 1/2 Mapping Rate** is the number of mapped divided by the total number of End1/End2 reads. **End 1/2 Mismatch Rate** is the number of End 1 and 2 bases not matching the reference divided by the total number of mapped End 1 and 2 bases. **Fragment Length Mean/StdDev** is the mean distance, standard deviation between the start of an upstream read and the end of the downstream one. Only fragments contained within single exons are used. **Chimeric Pairs** are pairs whose mates map to different genes.

Transcript-associated Reads

Sample	Note	Intragenic Rate	Exonic Rate	Intronic Rate	Intergenic Rate	Split Reads	Expression Profiling Efficiency	Transcripts Detected	Genes Detected
SW480circlane1	sw480	0.654	0.329	0.325	0.284	1,644,164	0.329	105,992	17,058
SW480circlane2	sw480	0.662	0.343	0.319	0.292	1,564,855	0.343	105,441	16,976
SW480circlane3	sw480	0.659	0.331	0.328	0.299	1,455,506	0.331	104,025	16,773
SW480circlane4	sw480	0.703	0.363	0.341	0.295	1,486,515	0.363	104,438	16,872
SW480ribolane1	sw480	0.803	0.501	0.302	0.197	6,582,567	0.501	123,211	21,356
SW480ribolane2	sw480	0.798	0.499	0.299	0.201	6,256,447	0.499	123,019	21,448
SW480ribolane3	sw480	0.801	0.501	0.300	0.198	5,803,961	0.501	122,070	21,100
SW480ribolane4	sw480	0.806	0.497	0.309	0.194	5,956,986	0.497	122,400	21,214
SW620circlane1	sw620	0.729	0.354	0.375	0.259	2,001,220	0.354	111,091	18,073
SW620circlane2	sw620	0.694	0.349	0.345	0.257	1,909,989	0.349	110,698	18,027
SW620circlane3	sw620	0.685	0.324	0.360	0.270	1,766,190	0.324	109,691	17,770
SW620circlane4	sw620	0.690	0.389	0.301	0.256	1,814,739	0.389	110,120	17,931
SW620ribolane1	sw620	0.785	0.473	0.311	0.215	4,661,365	0.473	118,921	20,390
SW620ribolane2	sw620	0.785	0.468	0.317	0.215	4,431,756	0.468	118,762	20,378
SW620ribolane3	sw620	0.776	0.464	0.311	0.224	4,123,847	0.464	117,620	20,088
SW620ribolane4	sw620	0.779	0.466	0.313	0.221	4,226,931	0.466	117,927	20,168

All of the above rates are per mapped read. **Intragenic Rate** refers to the fraction of reads that map within genes (within introns or exons). **Exonic Rate** is the fraction mapping within exons. **Intronic Rate** is the fraction mapping within introns. **Intergenic Rate** is the fraction mapping in the genomic space between genes. **Split Reads** is the number of reads spanning an exon exon junction. **Expression Profile Efficiency** is the ratio of exon reads to total reads. **Transcripts/Genes Detected** is the number of transcripts/Genes with at least 5 reads.

Strand Specificity

Sample	Note	End 1 Sense	End 1 Antisense	End 2 Sense	End 2 Antisense	End 1 % Sense	End 2 % Sense
SW480circlane1	sw480	545,152	10,990,569	10,976,987	543,385	4.726	95.283
SW480circlane2	sw480	533,659	10,532,867	10,512,903	531,547	4.822	95.187
SW480circlane3	sw480	477,128	9,563,169	9,534,499	475,439	4.752	95.250
SW480circlane4	sw480	507,437	10,279,705	10,254,734	505,245	4.704	95.304
SW480ribolane1	sw480	708,734	16,366,876	16,369,808	705,996	4.151	95.866
SW480ribolane2	sw480	721,721	15,785,336	15,808,411	718,855	4.372	95.650
SW480ribolane3	sw480	714,525	14,576,854	14,583,651	712,076	4.673	95.345
SW480ribolane4	sw480	728,645	15,169,553	15,180,649	725,733	4.583	95.437
SW620circlane1	sw620	750,543	12,488,809	12,452,818	746,950	5.669	94.341
SW620circlane2	sw620	723,689	12,075,668	12,045,562	720,389	5.654	94.357
SW620circlane3	sw620	652,945	10,324,309	10,312,334	649,911	5.948	94.071
SW620circlane4	sw620	689,631	11,001,170	10,949,876	686,557	5.899	94.100
SW620ribolane1	sw620	713,782	13,913,947	13,918,583	710,124	4.880	95.146
SW620ribolane2	sw620	755,462	13,407,433	13,415,127	752,900	5.334	94.686
SW620ribolane3	sw620	615,738	12,279,663	12,285,822	610,916	4.775	95.263
SW620ribolane4	sw620	686,293	12,725,358	12,733,388	682,816	5.117	94.911

End 1/2 Sense are the number of End 1 or 2 reads that were sequenced in the sense direction. Similarly, **End 1/2 Antisense** are the number of End 1 or 2 reads that were sequenced in the antisense direction. **End 1/2 Sense %** are percentages of intragenic End 1/2 reads that were sequenced in the sense direction.

Referências

Adler AS, McClelland ML, Yee S, Yaylaoglu M, Hussain S, Cosino E. An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. *Genes Dev*, 2014. 28:1068–84. doi:10.1101/gad.237206.113.

American Cancer Society. Colorectal Cancer Facts & Figures, 2014-2016. Available from: URL: <http://www.cancer/research/cancerfactsstatistics/colorectal-cancer-facts-figures>.

American Cancer Society. What is Colorectal cancer? *Am Cancer Soc*, 2015. 74.

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.

Annie R, Chloé C, Christa K, Thomas F. Comparative Analysis of the Circular Transcriptome in Muscle, Liver, and Testis in Three Livestock Species. *Frontiers in Genetics*, 2021. 12; 660. doi=10.3389/fgene.2021.665153

Arvelo F, Sojo F, Cotte C. Biology of colorectal cancer. *E cancer medical science*, 2015. 9: 520. PMID: 25932044. doi: 10.3332/ecancer.2015.520.

Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, et al. CircRNA Biogenesis competes with Pre-mRNA splicing. *Mol Cell*, 2014. 56:55–66. doi:10.1016/j.molcel.2014.08.019.

Bachmayr-Heyda A, Reiner AT, Auer K, Sukhbaatar N, Aust S, Bachleitner-Hofmann T, Mesteri I, Grunt TW, Zeillinger R, Pils D. Correlation of circular RNA abundance with proliferation--exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci Rep*, 2015. 5: 8057. PMID:25624062. doi:10.1038/srep08057.

Bachmayr-Heyda A, Reiner AT, Auer K, Sukhbaatar N, Aust S, Bachleitner-Hofmann T, et al. Correlation of circular RNA abundance with proliferation – exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci Rep*, 2015.5:8057. doi:10.1038/srep08057.

Banham AH, Beasley N, Campo E, Fernandez PL, Fidler C, Gatter K, et al. The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p. *Cancer Res*, 2001. 61:8820–9.

Banham AH, Connors JM, Brown PJ, Cordell JL, Ott G, Sreenivasan G, et al. Expression of the FOXP1 transcription factor is strongly associated with inferior survival in patients with diffuse large B-cell lymphoma. *Clin Cancer Res*, 2005. 11:1065–72.

Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. 2004. 116, 281–297.

Bergmann-Leitner ES, Abrams SI, Gamen S, Monle n I, Lasierra P, Larrad L, et al. Differential role of Fas/Fas ligand interactions in cytolysis of primary and metastatic colon carcinoma cell lines by human antigen-specific CD8+ CTL. *J Immunol*, 2000.164:4941–54. doi:10.4049/jimmunol.164.9.4941.

Best MG. RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics, 2015. 1–11. doi:10.1016/j.ccell.2015.09.018.

Blank M, Tang Y, Yamashita M, Burkett SS, Cheng SY, Zhang YE. A tumor suppressor function of Smurf2 associated with controlling chromatin landscape and genome stability through RNF20. *Nat Med*, 2012. 18:227–34. doi:10.1038/nm.2596.

Bosman FT, Hamilton SR, Lambert R. World Cancer Report 2014. 5.5 Colorectal Cancer. In: Stewart BW, Wild CP. World Cancer Report IARC, 2014. 560-576.

Brawand D, Harrigan P, Soumillon M, Necsulea A, Julien P, Aximu-petri A, Gru F. The evolution of gene expression levels in mammalian organs. 2011. doi:10.1038/nature10532.

Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science*, 2011. 331:1559–64. doi:10.1126/science.1203543.

Chakrabarti A, Jha B, Silverman R. New insights into the role of RNase L in innate immunity. *J Interf Cytokine Res: off J Int Soc Interf Cytokine Res*, 2011. 31(1):49–57.

Chen J, Xiao H, Huang Z, Hu Z, Qi T, Zhang B, et al. MicroRNA124 regulate cell growth of prostate cancer cells by targeting iASPP. *Int J Clin Exp Pathol* 2014. 7:2283–90.

Chen J, Xiao H, Huang Z, Hu Z, Qi T, Zhang B, Liu SH. MicroRNA124 regulate cell growth of prostate cancer cells by targeting iASPP. *International Journal of Clinical and Experimental Pathology*, 2014. 7(5), 2283–2290.

Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Scientific Reports*, 2016. 6(1), 34985. doi:10.1038/srep34985.

Chen X, Han P, Zhou T. circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep*, 2016. 6, 34985.

Chou YT, Lin HH, Lien YC, Wang YH, Hong CF, Kao YR, Wu CW. EGFR promotes lung tumorigenesis by activating miR-7 through a Ras/ERK/Myc pathway that targets the Ets2 transcriptional repressor ERF. *Cancer Research*, 2010. 70(21), 8822– 8831. doi:10.1158/0008-5472.CAN-10-0638.

Cocquerelle C, Daubersies P, Majerus M, Kerckaert J. Splicing with inverted.

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire M-D, Williams C, et al. RNA- SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics*, 2012. 28:1530–2. doi:10.1093/bioinformatics/bts196.

Dobin A, Gingeras TR, Spring C. Mapping RNA-seq Reads with STAR 2016. doi:10.1002/0471250953.bi1114s51.Mapping.

Dong R, Zhang XO, Zhang Y, Ma XK, Chen LL, Yang L. CircRNA-derived pseudogenes. *Cell Res*, 2016. 26, 747–750. doi: 10.1038/cr.2016.42]

Dong Y, He D, Peng Z, Peng W, Shi W, Wang J, Duan C. Circular RNAs in cancer: an emerging key player. *Journal of Hematology & Oncology*, 2017. 10(1), 2. doi:10.1186/s13045-016-0370-2.

Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biology*, 2016. 13(1), 34–42. doi:10.1080/15476286.2015.1128065.

Duursma AM, Kedde M, Schrier M, Sage CLE, Agami R. miR-148 targets human DNMT3b protein coding region. *Ence SCI. Viroids: Structure and Funct*, 2008. 872–7. 1979;205:859–66. doi:10.1261/rna.972008.2004.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 2009. 10:48. doi: 10.1186/1471-2105-10-48. PMID: 19192299; PMCID: PMC2644678.

Enuka Y, Lauriola M, Feldman ME, Sas-Chen A, Ulitsky I, Yarden Y. Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res*, 2016. 18;44(3):1370-83. doi: 10.1093/nar/gkv1367. PMID: 26657629; PMCID: PMC4756822.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, 2015. 136: E359-E386. PMID: 25220842. doi: 10.1002/ijc.29210.

Gaffo E, Boldrin E, Dal Molin A. Circular RNA differential expression in blood cell populations and exploration of circRNA deregulation in pediatric acute lymphoblastic leukemia. *Sci Rep*, 2019. 9, 14670. <https://doi.org/10.1038/s41598-019-50864-z>

Gaffo E, Bonizzato A, Kronnie GT, Bortoluzzi S. CirComPara: A Multi-Method Comparative Bioinformatics Pipeline to Detect and Study circRNAs from RNA-seq Data. *Non-coding RNA*, 2017. 3(1), 8. <https://doi.org/10.3390/ncrna3010008>

Gall JG. Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. 2014. 1476–1487. doi:10.1261/rna.045781.114.

Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 2020. 36(8):2628-2629. doi:10.1093/bioinformatics/btz931. PMID: 31882993; PMCID: PMC7178415.

Ghosal S, Das S, Sen R, Basak P, Chakrabarti J. Circ2Traits: A comprehensive database for circular RNA potentially associated with disease and traits. *Frontiers in Genetics*, 2013. 4(DEC), 1–9. doi:10.3389/fgene.2013.00283.

Glazar P, Papavasileiou P, Rajewsky N, Ar PGLaŽ, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs circBase: a database for circular RNAs. 2014. 0–5. doi:10.1261/rna.043687.113.overview.

Granados-riveron JT, Aquino-jarquín G. The complexity of the translation ability of circRNAs. *BBA - Gene Regulatory Mechanisms*, 2016. 1859(10), 1245–1251. doi:10.1016/j.bbagr. 2016.07.009.

Gruber R, Keller V, Gruber R, Keller V. Cluster optimization. *HPC@Green IT*, 2010.121–60. doi:10.1007/978-3-642-01789-6_7.

Gu CC, Jianqun X, Zhijun Z, Shiming Y, Liangying Y, Qun Z, Yue F, Yuxin L, Yongfeng T, Chenge W, Xinke H, Juan L, Qingyuan H, Lu L, Xin L, Aimin L. MiR-532-3p suppresses colorectal cancer progression by disrupting the ETS1/TGM2 axis-mediated Wnt/ β -catenin signaling. *Cell Death & Disease*, 2019. 10.1038/s41419-019-1962-x.

Guarnerio J, Bezzi M, Jeong JC, Paffenholz S V., Berry K, Naldini MM, et al. Erratum: Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations. *Cell*, 2016. 166:1055–6. doi:10.1016/j.cell.2016.07.035.

Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. 2014. 1–14. doi:10.1186/s13059-014-0409-z.

Han G, Li AG, Liang Y-Y, Owens P, He W, Lu S, et al. Smad7-Induced β -Catenin Degradation Alters Epidermal Appendage Development. *Dev. Cell*, 2006. 11:301–12. doi:10.1016/j.devcel.2006.06.014.

Han Y, Donovan J, Rath S, Whitney G, Chitrakar A, Korennykh A. Structure of human RNase L reveals the basis for regulated RNA decay in the IFN response. *Science (new York, NY)*, 2014. 343(6176):1244–8.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. *Nature*, 2013. 495(7441), 384–388. doi:10.1038/nature11993.

Hansen TB, Wiklund ED, Bramsen JB, Villadsen SB, Statham AL, Clark SJ. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. 2011. 30(21), 4414–4422. doi:10.1038/emboj.2011.359

Hao S, Cong L, Qu R, Liu R, Zhang G, Li Y. Emerging roles of circular RNAs in colorectal cancer. *OncoTargets and therapy*, 2019. 12, 4765–4777. <https://doi.org/10.2147/OTT.S208235>

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*, 2012. 22:1760–74. doi:10.1101/gr.135350.111.

Henry C, Close A-F, Buteau J. A critical role for the neural zinc factor ST18 in pancreatic β -cell apoptosis. *J Biol Chem*, 2014. 289:8413–9. doi:10.1074/jbc.M114.554915.

Hewitt RE, McMarlin A, Kleiner D, Wersto R, Martin P, Tsokos M, et al. Validation of a model of colon cancer progression. *J Pathol*, 2000. 192:446–54. doi:10.1002/1096-9896(2000)9999:9999<::AID-PATH775>3.0.CO;2-K.

Hsiao KY, Lin YC, Gupta SK, Chang N, Yen L, Sun HS, et al. Noncoding effects of circular RNA CCDC66 promote colon cancer growth and metastasis. *Cancer Res*, 2017. 77:2339–50. doi:10.1158/0008-5472.CAN-16-1883.

Hsu MT, Coca-Prados M. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature*, 1979. 280:339–40.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 2009. 4:44–57. doi:10.1038/nprot.2008.211.

Huang G, Zhu H, Shi Y, et al. cir-ITCH plays an inhibitory role in colorectal cancer by regulating the Wnt/beta-catenin pathway. *PLoS One*, 2015. 10(6): e0131225.

Huo LW, Wang YF, Bai XB, Zheng HL, Wang MD. circKIF4A promotes tumorigenesis of glioma by targeting miR-139-3p to activate Wnt5a signaling. *Molecular medicine (Cambridge, Mass.)*, 2020. 26(1), 29. <https://doi.org/10.1186/s10020-020-00159-1>

Illumina NextSeq Specification Parameters. Disponível em: www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwje64vK_5HIAhXTHbkGHem9A40QFjAAegQIAhAB&url=https%3A%2F%2Fwww.illumina.com%2Fsystems%2Fsequencing.platforms%2Fnextseq%2Fspecifications.html&usq

Illumina. bcl2fastq Conversion User Guide. Energy, 2013.

Illumina. Sequencing Analysis Viewer Software User Guide, 2014.

Invitrogen. TRIzol Reagent User Guide. User Guid, 2016. 15596018:1–6.

Jandrig B, Seitz S, Hinzmann B, Arnold W, Micheel B, Koelble K, et al. ST18 is a breast cancer tumor suppressor gene at human chromosome 8q11.2. *Oncogene*, 2004. 23:9295–302. doi:10.1038/sj.onc.1208131.

Jandrig B, Seitz S, Hinzmann B, Arnold W, Micheel B, Koelble K. ST18 é um gene supressor de tumor de câncer de mama no cromossomo humano 8q11.2. *Oncogene*, 2004. 23: 9295–302. doi:10.1038 / sj.onc.1208131

Jeck WR, Sorrentino JA, Wang K. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, 2013. 19:141-157.

Jeck WR, Sorrentino JA, Wang KAI, Slevin MK, Burd CE, Liu J, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. 2013. 141–157. doi:10.1261/rna.035667.112.8.

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, 2009. 37:D412–6. doi:10.1093/nar/gkn760.

Ji P, Wu W, Chen S, Zheng Y, Zhou L, Zhang J, Cheng H, Yan J, Zhang S, Yang P, Zhao F. Expanded Expression Landscape and Prioritization of Circular RNAs in Mammals. *Cell Rep*, 2019. 26(12):3444-3460.e5. doi:10.1016/j.celrep.2019.02.078. PMID: 30893614.

Ji Y, Lv J, Sun D, Huang Y. Therapeutic strategies targeting Wnt/ β catenin signaling for colorectal cancer (Review). *International Journal of Molecular Medicine*, 2021. 49, 1. <https://doi.org/10.3892/ijmm.5056>

Jia WC, Xu B, Wu J. Circular RNA expression profiles of mouse ovaries during postnatal development and the function of circular RNA epidermal growth factor receptor in granulosa cells. *Metab. Clin. Exp.*, 2018. 85, 192–204.

Jiang W, Zhang X, Chu Q, Lu S, Zhou L, Lu X, et al. The circular RNA profiles of colorectal tumor metastatic cells. *Front Genet*, 2018. 9:1–11. doi:10.3389/fgene.2018.00034.

Jin K, Li T, van Dam H, Zhou F, Zhang L. Molecular insights into tumour metastasis: tracing the dominant events. *J Pathol*, 2017. 241:567–77. doi:10.1002/path.4871.

Junbo L, Hongchao L, Guangyu L, Jun Q, Ran G, Yanchi Z, Xiaoyue W. Global changes in chromatin accessibility and transcription after ATRX inactivation in human cancer cells. 2019. <https://doi.org/10.1002/1873-3468.13549>

Katzenellenbogen RA, Vliet-Gregg P, Xu M, Galloway DA. Cytoplasmic Poly(A) Binding Proteins Regulate Telomerase Activity and Cell Growth in Human Papillomavirus Type 16 E6-Expressing Keratinocytes. *J Virol*, 2010. 84:12934–44. doi:10.1128/JVI.01377-10.

Kim KM, Abdelmohsen K, Mustapic M, Kapogiannis D, Gorospe M. RNA in extracellular vesicles. *Wiley Interdiscipl Ver*, 2017. 8(4):e1413.

Kleaveland B, Shi C, Stefano J, Bartel D. A network of noncoding regulatory RNAs acts in the mammalian brain. *Cell*, 2018. 174(2):350-362.e317.

Kolakofsky D. Isolation and characterization of Sendai virus DI-RNAs. *Cell*, 1976. 8:547–55.

Kong Y, Cui H, Zhang H. Smurf2-mediated ubiquitination and degradation of Id1 regulates p16 expression during senescence. *Aging Cell*, 2011. 10:1038–46. doi:10.1111/j.1474-9726.2011.00746.x.

Laclef C, Anselme I, Besse L, Catala M, Palmyre A, Baas D, et al. The role of primary cilia in corpus callosum formation is mediated by production of the Gli3 repressor. *Hum Mol Genet*, 2015. 24:4997–5014. doi:10.1093/hmg/ddv221.

Larsson C, Cordeddu L, Siggins L. Restoration of KMT2C / MLL3 in human colorectal cancer cells enhances genome-wide H3K4me1 profiles and influences cell growth and gene expression. *Clin Epigenet*, 2020. 12, 74. <https://doi.org/10.1186/s13148-020-00863-z>

Lasda E, Parke R. Circular RNAs Co-Precipitate with Extracellular Vesicles: A Possible Mechanism for circRNA Clearance. 2016. 1–11. doi:10.1371/journal.pone.0148407.

Lee Y, Choe J, Park O, Kim Y. Molecular mechanisms driving mRNA degradation by mA modification. *Trends Genet: TIG*, 2020. 36(3):177–88.

Leibovitz A, Stinson JC, McCombs WB, McCoy CE, Mazur KC, Mabry ND. Classification of Human Colorectal Adenocarcinoma Cell Lines. *Cancer Res*, 1976. 36:4562–9.

Li A, Wang WC, McAlister V, Zhou Q, Zheng X. Circular RNA in colorectal cancer. *Journal of Cellular and Molecular Medicine*, 2021. <https://doi.org/10.1111/jcmm.16380>

Li A, Wang WC, McAlister V, Zhou Q, Zheng X. Circular RNA in colorectal cancer. *Journal of cellular and molecular medicine*, 2021. 25(8), 3667–3679. <https://doi.org/10.1111/jcmm.16380>

Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 2011. 12:323. doi:10.1186/1471-2105-12-323.

Li H, Liefke R, Jiang J, Kurland JV, Tian W, Deng P. Polycomb-like proteins bind the PRC2 complex to the CpG islands. *Nature*, 2017. 549: 287–91. doi: 10.1038 / nature23881

Li P, Chen S, Chen H, Mo X, Li T, Shao Y, et al. Using circular RNA as a novel type of biomarker in the screening of gastric cancer. *Clin Chim Acta*, 2015. 444:132–6. doi:10.1016/j.cca.2015.02.018.

Li Q, Lai Q, He C, Fang Y, Yan Q, Zhang Y, et al. RUNX1 promotes tumour metastasis by activating the Wnt/ β -catenin signalling pathway and EMT in colorectal cancer. *J Exp Clin Cancer Res*, 2019. 38:334. doi:10.1186/s13046-019-1330-9.

Li QH, Liu Y, Chen S, Zong ZH, Du YP, Sheng XJ, Zhao Y. circ-CSPP1 promotes proliferation, invasion and migration of ovarian cancer cells by acting as a miR-1236-3p sponge. *Biomed. Pharmacother.*, 2019. 114, 108832.

Li X, Wu Y. Detecting circular RNA from high-throughput sequence data with de Bruijn graph. *BMC Genomics*, 2020. 21, 749. <https://doi.org/10.1186/s12864-019-6154-7>

Li Y, Zheng Q, Bao C, Li S, Guo W, Zhao J. Circular RNA is enriched and stable in exosomes: A promising biomarker for cancer diagnosis. *Cell Res*, 2015; 25:981–4. doi:10.1038/cr.2015.82.

Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Shan G. Exon-intron circular RNAs regulate transcription in the nucleus. 2016. 22(3). doi:10.1038/nsmb.2959.

Li, JH, Liu S, Zhou H, Qu LH, Yang JH. StarBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, 2014. 42(D1), 92–97. doi:10.1093/nar/gkt1248.

Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res*, 2019. 47:W199–205. doi:10.1093/nar/gkz401.

Liu C, Li X, Nan F, Jiang S, Gao X, Guo S, et al. Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell*, 2019. 177(4):865-880.e821.

Liu D, Yin B, Wang Q, Ju W, Chen Y, Qiu H, et al. Cytoplasmic Poly(A) Binding Protein 4 Is Highly Expressed in Human Colorectal Cancer and Correlates with Better Prognosis. *J Genet Genomics*, 2012.39:369–74. doi:10.1016/j.jgg.2012.05.007.

Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 2013. 29:1830–1. doi:10.1093/bioinformatics/btt285.

Luo Y, Liu F, Ma J. A novel epigenetic regulation of circFoxp1 on Foxp1 in colon cancer cells. *Cell Death Dis*, 2020. 11, 782. <https://doi.org/10.1038/s41419-020-03007-6>

Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. 2007. 104.

Ma S, Kong S, Wang F, Ju S. CircRNAs: biogenesis, functions, and role in drug-resistant Tumours. *Mol Cancer*, 2020. 19(1):1–19.

Malumbres M. Control of cell proliferation pathways by microRNAs. 2008. 4101. doi:10.4161/cc.7.20.6833.

Mao G, Zhou B, Xu W, Jiao N, Wu Z, Li J, Liu Y. Hsa_circ_0040809 regulates colorectal cancer development by upregulating methyltransferase DNMT1 via targeting miR-515-5p. 2021. <https://doi.org/10.1002/jgm.3388>

Medina PP, Slack FJ. MicroRNAs and cancer: An overview. 2008. 4101. doi:10.4161/cc.7.16.6453.

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Rajewsky N. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 2013. 495(7441), 333–338. doi:10.1038/nature11928.

Memczak S, Papavasileiou P, Peters O, Rajewsky N. Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS ONE*, 2015. 10(10), 1–13. doi:10.1371/journal.pone.0141214.

Mikels AJ, Nusse R. Purified Wnt5a Protein Activates or Inhibits β -Catenin–TCF Signaling Depending on Receptor Context. *PLoS Biol*, 2006. 4:e115. doi:10.1371/journal.pbio.0040115.

Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*, 2003. 34:267–73. doi:10.1038/ng1180.

Nair AA, Niu N, Tang X, Thompson KJ, Wang L, Kocher JP, Kalari KR. Circular RNAs and their associations with breast cancer subtypes. *Oncotarget*, 2016. 7(49), 80967–80979. doi:10.18632/oncotarget.13134.

Narimatsu M, Bose R, Pye M, Zhang L, Miller B, Ching P, et al. Regulation of Planar Cell Polarity by Smurf Ubiquitin Ligases. *Cell*, 2009. 137:295–307. doi:10.1016/j.cell.2009.02.025.

Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), {date}. World Wide Web URL: omim.org.

Papatsirou M, Artemaki PI, Scorilas A, Kontos CK. The role of circular RNAs in therapy resistance of patients with solid tumors. *Pers Med*, 2020. 17(06):469–90.

Paschall AV, Yang D, Lu C, Redd PS, Choi J-H, Heaton CM. CD13 CD24 defines a 5-Fluorouracil-resistant colon cancer stem cell-like phenotype. *Oncotarget*, 2016 7:78698–712. doi:10.18632/oncotarget.12168.

Peng L, Yuan XQ, Li GC. The emerging landscape of circular RNA ciRS-7 in cancer (Review). *Oncology Reports*, 2015. 33(6), 2669–2674. doi:10.3892/or.2015.3904.

Perk J, Iavarone A, Benezra R. Id family of helix-loop-helix proteins in cancer. *Nat Rev Cancer*, 2005. 5:603–14. doi:10.1038/nrc1673.

Perriman R, Ares M. Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. 1998. 1047–1054.

Provenzani A, Fronza R, Loreni F, Pascale A, Amadio M, Quattrone A. Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer progression to metastasis. *Carcinogenesis*, 2006. 27:1323–33. doi:10.1093/carcin/bgi377.

Qin M, Liu G, Huo X, et al. Hsa_circ_0001649: A circular RNA and potential novel biomarker for hepatocellular carcinoma. *Cancer Biomark*, 2016. 16(1): 161-9.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2014. URL <http://www.R-project/>.

Rabin A, Zaffagni M, Ashwal-Fluss R. SRCP: a comprehensive pipeline for accurate annotation and quantification of circRNAs. *Genome Biol*, 2021. 22, 277. <https://doi.org/10.1186/s13059-021-02497-7>

Ragusa M, Barbagallo C, Statello L, Condorelli AG, Battaglia R, Tamburello L, Barbagallo D, Di Pietro C, Purrello M. Non-coding landscapes of colorectal cancer. *World J Gastroenterol*, 2015. 21: 11709-11739. PMID: 26556998. doi: 10.3748/wjg.v21.i41.11709.

Ramkumar C, Kong Y, Cui H, Hao S, Jones SN, Gerstein RM, et al. Smurf2 Regulates the Senescence Response and Suppresses Tumorigenesis in Mice. *Cancer Res*, 2012. 72:2714–9. doi:10.1158/0008-5472.CAN-11-3773.

Rasmussen MH, Lyskjær I, Jersie-Christensen RR, Tarpgaard LS, Primdal-Bengtson B, Nielsen MM, et al. miR-625-3p regulates oxaliplatin resistance by targeting MAP2K6- p38 signalling in human colorectal adenocarcinoma cells. *Nat Commun*, 2016. 7:12436. doi:10.1038/ncomms12436.

Rigoutsos I. New Tricks for Animal MicroRNAs: Targeting of Amino Acid Coding Regions at Conserved and Nonconserved Sites. 2009. 3245–9. doi:10.1158/0008-5472.CAN-09-0352.

Rishi AK, Zhang L, Boyanapalli M, Wali A, Mohammad RM, Yu Y, et al. Identification and characterization of a cell cycle and apoptosis regulatory protein-1 as a novel mediator of apoptosis signaling by retinoid CD437. *J Biol Chem*, 2003. 278:33422–35. doi:10.1074/jbc.M303173200.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 2015. 43(7), e47.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, 2015. 43:e47–e47. doi:10.1093/nar/gkv007.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 2010. 26:139–40. doi:10.1093/bioinformatics/btp616.

Ruppert, JM, Kinzler W. Scrambled Exons. 1991. 64, 607–613.

Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-Type Specific Features of Circular RNA Expression. 2013. doi:10.1371/journal.pgen.1003777

Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Correction: Cell-Type Specific Features of Circular RNA Expression. *PLOS Genetics*, 2012. 9(12): 10.1371/annotation/f782282b-eefa-4c8d-985c-b1484e845855.

<https://doi.org/10.1371/annotation/f782282b-eefa-4c8d-985c-1484e845855>

Sanger HL, Hans J, Albrecht K. highly base-paired .1976. 73:3852–6.

Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen H-C, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *bioRxiv*, 2016. 2116. doi:10.1101/072116.

Shen M, Zhang Z, Wang P. GLI3 Promotes Invasion and Predicts Poor Prognosis in Colorectal Cancer. *Biomed Res Int.*, 2021. 8889986. doi: 10.1155/2021/8889986. PMID: 33506047; PMCID: PMC7814942.

Shi H, Wang X, Lu Z, Zhao B, Ma H, Hsu P, et al. YTHDF3 facilitates translation and decay of N-methyladenosine-modified RNA. *Cell Res*, 2017. a;27(3):315–28.

Song J, Shu H, Zhang L and Xiong J: Long noncoding RNA GAS5 inhibits angiogenesis and metastasis of colorectal cancer through the Wnt/ β -catenin signaling pathway. *J Cell Biochem*, 2019.

Song X, Zhang N, Han P. Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic acids research*, 2016.

Su M, Xiao Y, Ma J, Tang Y, Tian B, Zhang Y, et al. Circular RNAs in Cancer: emerging functions in hallmarks, stemness, resistance and roles as potential biomarkers. *Mol Cancer*, 2019. 18(1):1-17.

Sung JJ, Ng SC, Chan FK, Chiu HM, Kim HS, Matsuda T, Ng SS, Lau JY, Zheng S, Adler S, Reddy N, Yeoh KG, Tsoi KK, Ching JY, Kuipers EJ, Rabeneck L, Young GP, Steele RJ, Lieberman D, Goh KL. An updated Asia Pacific Consensus Recommendations on colorectal cancer screening. *Gut*, 2015. 64: 121-132. PMID: 24647008. doi: 10.1136/gutjnl-2013-306503.

Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One*, 2011. 6:e21800. doi:10.1371/journal.pone.0021800.

Szabo LSJ. Detecting circular RNAs: bioinformatic and experimental challenges. *Nat Rev Genet*, 2016. 17(11), 679-692. doi:10.1002/wrna.1178.

Tao M, Zheng M, Xu Y. CircRNAs e suas funções regulatórias em cânceres. *Mol Med*, 2021. 27, 94. <https://doi.org/10.1186/s10020-021-00359-3>

Vakhshiteh F, Atyabi F, Ostad SN. Mesenchymal stem cell exosomes: a two-edged sword in cancer therapy. *Int J Nanomed*, 2019. 14:2847.

Vicens Q, Westhof E. Biogenesis of Circular RNAs. *Cell*, 2014. 159(1), 13–14. doi:10.1016/j.cell.2014.09.005.

Wang PL, Bao Y, Yee M, Barrett SP, Hogan GJ, Brown PO, et al. Circular RNA Is Expressed across the Eukaryotic Tree of Life. 2014. 9. doi:10.1371/journal.pone.0090859.

Wang Y, Mo Y, Gong Z, Yang X, Yang M, Zhang S, Xiong W. Circular RNAs in human cancer. *Molecular Cancer*, 2017. 16(1), 25. doi:10.1186/s12943-017-0598-7.

Wang Z, CAS-MP. Not just a sponge: new functions of circular RNAs discovered. 2015. 58(4), 407–408. doi:10.1038/nsmb.2959

Wen JL, Wen XF, Li RB, Jin YC, Wang XL, Zhou L, Chen HX. UBE3C promotes growth and metastasis of renal cell carcinoma via activating Wnt/ β -catenin pathway. *PLoS One*, 2015.10(2):e0115622. doi: 10.1371/journal.pone.0115622. PMID: 25658088; PMCID: PMC4319909.

Weng W, Wei Q, Toden S, et al. Circular RNA ciRS-7 - A promising prognostic biomarker and a potential therapeutic target in colorectal cancer. *Clin Cancer Res*, 2017.

William R. Jeck, N. E. S. NIH Public Access, 2008. 86(12), 3279–3288. doi:10.1007/s11103-011-9767-z.Plastid.

World Health Organization. Estimated Incidence, Mortality and Prevalence Worldwide in 2012. Available from: URL: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.

Wu Q, Huang JH, Sampson ER, Kim K-O, Zuscik MJ, O'Keefe RJ, et al. Smurf2 induces degradation of GSK-3 β and upregulates β -catenin in chondrocytes: A potential mechanism for Smurf2-induced degeneration of articular cartilage. *Exp Cell Res*, 2009. 315:2386–98. doi:10.1016/j.yexcr.2009.05.019.

Wu Q, Li P, Wu M, Liu, Q. Deregulation of circular RNAs in cancer from the perspectives of aberrant biogenesis, transport and removal. *Front. Genet*, 2019. 10:16. doi: 10.3389/fgene.2019.00016

Xia S, Feng J, Lei L, Hu J, Xia L, Wang J, Xiang Y, Liu L, Zhong S, Han L, He C. Comprehensive characterization of tissue specific circular RNAs in the human and mouse genomes. *Brief Bioinform*, 2016. doi:10.1093/bib/bbw081.

Xu L, Feng X, Hao X. CircSETD3 (Hsa_circ_0000567) acts as a sponge for microRNA-421 inhibiting hepatocellular carcinoma growth. *J Exp Clin Cancer*, 2019. Res 38, 98. <https://doi.org/10.1186/s13046-019-1041-2>

Xu, H, Guo S, Li W, Yu P. The circular RNA Cdr1as, via miR-7 and its targets, regulates insulin transcription and secretion in islet cells. *Scientific Reports*, 2015. 5(1), 12453. doi:10.1038/srep12453

Yang J, Siqueira MF, Behl Y, Alikhani M, Graves DT. The ST18 transcription factor regulates pro-apoptotic and pro-inflammatory gene expression in fibroblasts. *FASEB J*, 2008. 22: 3956–67. doi: 10.1096 / fj.08-111013

Yang P, Yu D, Zhou J, Zhuang S and Jiang T: TGM2 interference regulates the angiogenesis and apoptosis of colorectal cancer via Wnt/ β -catenin pathway. *Cell Cycle*, 2019. 18:1122–1134.

Yang Y, Du Y, Liu X. Non-coding RNAs in Colorectal Cancer. 2016. Available from: <http://link.springer.com/10.1007/978-3-319-42059-2>. doi: 10.1007/978-3-319-42059-2.

Yang Z, Xie L, Han L, Qu X, Yang Y, Zhang Y. Circular RNAs: regulators of cancer-related signaling pathways and potential diagnostic biomarkers for human cancers. *Theranostics*, 2017. 7, 3106–3117. doi: 10.7150/thno.19016

Ying X, Zhu J, Zhang Y. Circular RNA circ-TSPAN4 promotes lung adenocarcinoma metastasis by upregulating ZEB1 via sponging miR-665. *Molecular genetics & genomic medicine*, 2019. 7(12), e991. <https://doi.org/10.1002/mgg3.991>

Yoon YJ, Kim OY, Gho YS. Extracellular vesicles as emerging intercellular comunicasomes. *BMB Rep*, 2014. 47(10):531.

Zelin Liu, Huiru Ding, Jianqi She, Chunhua Chen, Weiguang Zhang, Ence Yang. DEBKS: A Tool to Detect Differentially Expressed Circular RNA, *Genomics, Proteomics & Bioinformatics*. 2021.

Zhang J, Chen S, Yang J. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun*, 2020. 11, 90. <https://doi.org/10.1038/s41467-019-13840-9>

Zhang S, Zhu D, Li H, Li H, Feng C, Zhang W. Characterization of circRNA-Associated-ceRNA Networks in a Senescence-Accelerated Mouse Prone 8 Brain. *Mol Ther*, 2017. 25:6–14. doi:10.1016/j.jymthe.2017.06.009.

Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, Yang L. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Research*, 2016. 26(9), 1277–1287. doi:10.1101/gr.202895.115

Zhang XO, Wang HBin, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*, 2014. 159(1), 134–147. doi:10.1016/j.cell.2014.09.001

Zhang Y, Zhang X, Chen T, Xiang J, Yin Q, Xing Y, Chen L. Circular Intronic Long Noncoding RNAs. *Molecular Cell*, 2013. 51(6), 792–806. doi:10.1016/j.molcel.2013.08.017.

Zhang Z, Yang T, Xiao J. Circular RNAs: promising biomarkers for human diseases. *EBioMedicine*, 2018. 34, 267–274.

Zheng LL, Li JH, Wu J, Sun WJ, Liu S, Wang ZL, Qu LH. deepBase v2.0: Identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Research*, 2016. 44(D1), D196– D202. doi:10.1093/nar/gkv1273.

Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat Commun*, 2016. 7:1–13. doi:10.1038/ncomms11215.

Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, Huang S. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nature Communications*, 2016. 7, 1– 13. doi:10.1038/ncomms11215.

Zhou J, Wan J, Gao X, Zhang X, Jaffrey S, Qian S. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature*, 2015. 526(7574):591–4.

Zhu M, Xu Y, Chen Y, Yan F. Circular BANP, an upregulated circular RNA that modulates cell proliferation in colorectal cancer. *Biomed Pharmacother*, 2017. 88:138–44. doi:10.1016/j.biopha.2016.12.097.

Zou C, Li X, Lv X, Wu S, Song J, Tang Z, Luo H, Wei H, Ai Y. Circular RNA mitochondrial translation optimization 1 homologue (CircMTO1) induced by zinc finger protein 460 (ZNF460) promotes oral squamous cell carcinoma progression through the microRNA miR-320a / alpha thalassemia/mental retardation, X-linked (ATRX) axis. *Bioengineered*, 2021. (2):9585-9597. doi: 10.1080/21655979.2021.19