UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

Trabalho de Conclusão de Curso

ANÁLISE DO TRANSCRIPTOMA DE CARCINOMA RENAL DE CÉLULAS CLARAS BASEADA EM RNAs NÃO CODIFICANTES

Epitácio Dantas de Farias Filho

Natal/RN

2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

ANÁLISE DO TRANSCRIPTOMA DE CARCINOMA RENAL DE CÉLULAS CLARAS BASEADA EM RNAs NÃO CODIFICANTES

Trabalho de conclusão de curso apresentado ao Departamento de Engenharia Biomédica da Universidade Federal do Rio Grande do Norte para obtenção do título de Bacharel em Engenharia Biomédica.

Graduando: Epitácio Dantas de Farias Filho

Orientadora: Prof.ª Dra. Beatriz Stransky Ferreira

Coorientador: Prof. Dr. Patrick Cesar Alves Terrematte

Natal/RN

2022

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Farias Filho, Epitácio Dantas de.
Análise do transcriptoma de carcinoma renal de células claras baseada em rnas não codificantes / Epitácio Dantas de Farias Filho. - 2022.
70f.: il.

Monografia (Graduação) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Departamento de Engenharia Biomédica, Natal, 2022.
Orientadora: Drª. Beatriz Stransky Ferreira.
Coorientador: Dr. Patrick Cesar Alves Terrematte.

1. Carcinoma Renal - Monografia. 2. RNAs não codificantes - Monografia. 3. Análise de Expressão Diferencial - Monografia. 4. Rede de Competição Endógena - Monografia. 5. Análise de Sobrevida - Monografia. I. Ferreira, Beatriz Stransky. II. Terrematte, Patrick Cesar Alves. III. Título.

RN/UF/BCZM CDU 61:6

Elaborado por Raimundo Muniz de Oliveira - CRB-15/429

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

ANÁLISE DO TRANSCRIPTOMA DE CARCINOMA RENAL DE CÉLULAS CLARAS BASEADA EM RNAs NÃO CODIFICANTES

Banca Examinadora do Trabalho de Conclusão de Curso:

Prof.ª Drª. Beatriz Stransky Ferreira _____

UFRN - Orientadora

Prof. Dr. Patrick Cesar Alves Terrematte _____

UFRN – Coorientador

Prof. Dr. Taffarel Melo Torres _____

UFERSA – Avaliador Externo

Drª. Iara Dantas de Souza _____

Avaliador Externo

Natal/RN

2022

DEDICATÓRIA

Dedico este trabalho a minha mãe e minha irmã, grandes apoiadoras e incentivadoras dos meus sonhos. Dedico aos meus amigos que fizeram parte da minha jornada durante o curso.

AGRADECIMENTOS

Agradeço primeiramente a Deus, aos orixás e a todas as entidades espirituais, que durante todos esses anos, tem me fortalecido, protegido, guiado e ajudado a conquistar tudo que conquistei e ainda conquistarei. Agradeço à minha família, em especial, à minha mãe Aida, à minha irmã Bárbara, à minha tia Adriana, e à minha avó Maria da Luz, que são as mulheres mais fortes que conheço, por terem sido, e serem, meus exemplos de dedicação, lutadoras, fortes e dedicadas, que sempre acreditaram em mim, e me apoiaram em todos os momentos de dificuldade e me incentivaram a estudar e seguir os meus sonhos.Sem vocês tudo isso e o que está por vir não teria sido possível.

Aos meus professores que estiveram me guiando, ensinando e incentivando a seguir minhas metas, em especial a minha orientadora Beatriz Stransky e meu coorientador Patrick Terrematte, como também as minhas professoras Luciana Mendes e a Heliana Bezerra.

Aos meus amigos que conheci no início da graduação, quando ainda estava na UFERSA em Mossoró cursando o primeiro ciclo, por todas as conversas e brincadeiras que faziam com que as disciplinas cursadas se tornassem mais fáceis e divertidas, em especial a Renan Crisóstomo, Matheus Diógenes, Mariana Duarte, Myllena Lopes e José Geraldo. Não tenho nem palavras para descrever o quanto sou grato aos meus amigos do Instituto de Química da UFRN, por terem me acolhido durante o período que não conhecia ninguém e foram tudo que precisei, em especial a João Vitor, Victor Eduardo, Mike, Beatriz, Katherine e Amanda.

Aos amigos da Engenharia Biomédica, pessoas que mesmo em meio a uma pandemia, mesmo estando longe pareciam estar perto, mostrando uma unicidade e companheirismo entre os alunos do curso, ajudando uns aos outros. Muito obrigado ao grupinho que construí nesse período, o “DEB Que Surte”, Jackson, Jade e João Pedro, chegar até aqui não teria sido tão fácil o quanto foi junto de vocês, meu muito obrigado!!

A minha grande amiga e parceira Eulália Ribeiro, que além de colega de curso se tornou uma pessoa especial em minha vida, com quem dividi conquistas e tristezas durante todo esse período de Engenharia Biomédica, que pretendo levar para o resto da minha vida.

EPÍGRAFE

A hereditariedade pertence à totalidade do genoma,
não a um único gene.
(Steven Pinker)

SUMÁRIO

DEDICATÓRIA IV
AGRADECIMENTOS V
EPÍGRAFE VI
LISTA DE FIGURAS IX
LISTA DE TABELAS XII
LISTA DE ABREVIAÇÕES, SIGLAS E SÍMBOLOS XIII
RESUMO XV
ABSTRACT XVI
1. INTRODUÇÃO 1
2. OBJETIVOS 3
2.1. OBJETIVO GERAL 3
2.2. OBJETIVOS ESPECÍFICOS 3
3. REFERENCIAL TEÓRICO 4
3.1. EPIDEMIOLOGIA DO CÂNCER RENAL 4
3.2. FISIOPATOLOGIA DO CÂNCER DE RIM 6
3.3. MECANISMOS MOLECULARES E INFLUÊNCIA DOS RNAs NÃO CODANTES NO CÂNCER 7
3.4. RNAs E SEUS PAPÉIS NA ONCOGÊNESE 8
3.4.1. REDE DE RNAs ENDÓGENOS COMPETIDORES (ceRNAS) 10
3.5. REPOSITÓRIO DE DADOS E FERRAMENTAS PARA ANÁLISE EM BIOINFORMÁTICA 10
3.5.1. GENOMIC DATA COMMONS - GDC 11
3.5.2. XENA BROWSER 11
3.5.3. BIOCONDUCTOR 12
3.5.4. RSTUDIO 12
4. METODOLOGIA 13
4.1. DADOS DOS PACIENTES 13
4.2. AQUISIÇÃO E ANÁLISE EXPLORATÓRIA DE DADOS 13
4.3. ANÁLISE DE ALTERAÇÕES GENÔMICAS 14
4.4. ANÁLISE DE ALTERAÇÕES DO TRANSCRIPTOMA 14
4.4.1. 15
4.4.2. ANÁLISE DE ENRIQUECIMENTO FUNCIONAL 15

4.4.3. REDE DE ceRNAs 15
4.5. ANÁLISE DE SOBREVIDA 16
4.6. DESENVOLVIMENTO 16
5. RESULTADOS 17
5.1. ANÁLISE EXPLORATÓRIA 17
5.2. ANÁLISE DE ALTERAÇÕES GENÔMICAS 18
5.3. ANÁLISE DE ALTERAÇÕES DO TRANSCRIPTOMA 22
5.3.1. ANÁLISE DE EXPRESSÃO DIFERENCIAL 22
5.3.2. ANÁLISE DE ENRIQUECIMENTO FUNCIONAL 23
5.3.3. REDE DE ceRNAs 24
5.4. ANÁLISE DE SOBREVIDA 25
6. DISCUSSÃO 27
7. CONCLUSÃO 30
REFERÊNCIAS BIBLIOGRÁFICAS 31
APÊNDICES 36
APÊNDICE A 37
APÊNDICE B 38
APÊNDICE C 39
APÊNDICE D 44
APÊNDICE E 45
APÊNDICE F 46

LISTA DE FIGURAS

Figura 1: Taxa de sobrevida de 5 anos (1969-2019), para pacientes com carcinoma renal e pelve renal a partir do estadiamento da doença 5
Figura 2:(a) Posicionamento anatômico, (b) corte coronal do rim. 6
Figura 3: RNAs codificantes e não codificantes do genoma humano e suas classificações em relação a quantidade de nucleotídeos. 9
Figura 4:Subconjuntos dos dados do projeto KIRC-TCGA. 13
Figura 5: Organização da análise transcriptômica. 15
Figura 6: Sumário das alterações genéticas para TCGA-KIRC: (a) classificação e contagem das mutações somáticas; (b) os tipos e contagem das variações; (c) contagens de transições e transversões dos SNVs; (d) a quantidade de mutações por amostra; (e) o b 19
Figura 7: Oncoplot com os 10 genes mais frequentemente mutados. O histograma à direita do gráfico representa a porcentagem de alteração do gene em toda a coorte e os tipos de mutações sofridas, o histograma acima representa o quantitativo de mutações por paciente.20
Figura 8: Principais vias oncogênicas alteradas. A esquerda a fração de alteração dos genes na via e a direita a fração de amostras com a via alterada. 21
Figura 9: Volcano plot de miRNAs diferencialmente expressos entre amostras tumorais e normais. Os pontos em vermelho indicam genes diferencialmente expressos a uma taxa de falsa descoberta de 1% quando o ajuste de teste múltiplo de Benjamini-Hochberg é utilizado e logFC = 2. 22
Figura 10: Volcano plot de lncRNAs diferencialmente expressos entre amostras tumorais e normais. Os pontos em vermelho indicam genes diferencialmente expressos a uma taxa de falsa descoberta de 1% quando o ajuste de teste múltiplo de Benjamini-Hochberg é utilizado e logFC = 2. 23
Figura 11: Principais vias enriquecidas pelo Disease Ontology para os mRNAs diferencialmente expressos. 23
Figura 12: (a) Enriquecimento de mRNAs diferencialmente expressos contra base de dados; (b) enriquecimento funcional para as vias: HP:0012126, KEGG: hsa05200, KEGG: hsa05230, WP:WP4585, WP:WP4018 e hsa-miR-335-5p. 24
Figura 13: (a) e (b)Rede de ceRNAs da coorte KIRC-TCGA. Em laranja os miRNAs, em verde os lncRNAs e em vermelho os mRNAs. 25

Figura 14: Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do SNHG15, em relação ao tempo. p-valor determinado pelo teste de log-rank. 26
Figura 15: Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do EPB41L4A-AS1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 26
Figura 16:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AC005154.1, em relação ao tempo. p-valor determinado pelo teste de log-rank 48
Figura 17:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AC015813.1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 48
Figura 18:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AC016876.2, em relação ao tempo. p-valor determinado pelo teste de log-rank. 49
Figura 19:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AC021078.1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 49
Figura 20:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AF117829.1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 50
Figura 21:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do C1RL-AS1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 50
Figura 22:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do GAS5, em relação ao tempo. p-valor determinado pelo teste de log-rank. 51
Figura 23:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do H19, em relação ao tempo. p-valor determinado pelo teste de log-rank. 51
Figura 24:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do MAGI2-AS3, em relação ao tempo. p-valor determinado pelo teste de log-rank. 52

Figura 25:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do MALAT1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 52
Figura 26:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do MIR497HG, em relação ao tempo. p-valor determinado pelo teste de log-rank. 53
Figura 27:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do NEAT1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 53
Figura 28:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do PVT1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 54
Figura 29:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do SNHG1, em relação ao tempo. p-valor determinado pelo teste de log-rank. 54
Figura 30:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do SNHG5, em relação ao tempo. p-valor determinado pelo teste de log-rank. 55
Figura 31:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do SNHG20, em relação ao tempo. p-valor determinado pelo teste de log-rank. 55

LISTA DE TABELAS

Tabela 1:Fatores de Risco Relacionados ao Carcinoma Renal 5
Tabela 2:Classificação histológica do Carcinoma Renal. 7
Tabela 3:Distribuição das variáveis clínicas em relação ao status vital na coorte TCGA-KIRC. 18
Tabela 4: Função e associações a doenças dos 10 genes mais frequentemente mutados. 20
Tabela 5: Informações associadas aos dados clínicos – TCGA-KIRC 37
Tabela 6: lncRNAs da ceRNAs 38
Tabela 7: mRNAs diferencialmente expressos participantes da ceRNA. 39
Tabela 8: 18 miRNAs diferencialmente participantes da rede de ceRNA. 44
Tabela 9: Bancos de dados de vias funcionais 47

LISTA DE ABREVIAÇÕES, SIGLAS E SÍMBOLOS

ccRCC – Carcinoma Renal de Células Claras.

ceRNAs – RNAs endógenos concorrentes.

DATASUS - Departamento de Informática do Sistema Único de Saúde.

DEA – Análise de expressão diferencial.

DElncRNAs – lncRNAs diferencialmente expressos.

DEmiRNAs – miRNAs diferencialmente expressos.

DEmRNAs – mRNAs diferencialmente expressos.

DNA – Ácido Desoxirribonucleico.

DO – Disease Ontology.

ENCODE – Enciclopédia Elementos do DNA.

FDR – False Rate Discovery.

GCO – Global Cancer Observatory.

GDC – Genomic Data Commons.

GO – Gene Ontology.

INCA – Instituto Nacional do Câncer.

KEGG – Enciclopédia de Genes e Genomas de Kyoto.

lncRNAs – RNAs longos não codificantes.

miRNAs – microRNAs.

NCI – Instituto Nacional do Câncer.

ncRNAs – RNAs não codificantes.

RNA – Ácido Ribonucleico.

SEER – Programa de Vigilância, Epidemiologia e Resultados Finais.

snoRNAs – RNAs nucleolares pequenos.

SNP – Polimorfismo de nucleotídeo único.

snRNAs – RNAs nucleares curtos.

TCGA – The Cancer Genome Atlas.

WXS - Sequenciamento completo do exoma.

FARIAS FILHO, Epitácio Dantas de. **Análise do Transcriptoma de Carcinoma Renal de Células Claras Baseada em RNAs Não Codificantes.** Trabalho de Conclusão de Curso, Graduação em Engenharia Biomédica, Universidade Federal do Rio Grande do Norte, 70 p., 2022.

RESUMO

O carcinoma renal de células claras é o subtipo mais comum entre os pacientes com câncer no sistema urinário e quando em metástase, apresenta uma baixa taxa de sobrevida. As alterações em nível de genoma e transcriptoma, especificamente dos mRNAs, já foram extensivamente analisadas e apenas nos últimos anos começaram a aparecer estudos mostrando a importância dos RNAs não-codificantes na regulação da expressão gênica. Desta forma, o objetivo deste estudo é investigar a atividade dos RNAs, com um foco nos não codificantes, no carcinoma renal de células claras, através da análise e integração de dados de pacientes do The Cancer Genome Atlas utilizando técnicas de bioinformática, analisando padrões dos pacientes da coorte, alterações genômicas e de transcriptoma. A análise exploratória mostrou que o estado de sobrevida dos pacientes está atrelado significativamente com o estadiamento da doença. A análise de expressão diferencial entre tecidos normais e tumorais, identificou 2.999 RNAs codificantes, 271 lncRNAs e 132 miRNAs, com um Log2FC = |2| e FDR < 0.01. A análise funcional dos genes diferencialmente expressos mostrou um enriquecimento de termos relacionados a vias oncológicas renais e doenças relacionadas ao rim. Por fim, a construção da rede de competição endógena, permitiu a identificação e visualização de alvos comuns entre 18 lncRNAs e 75 miRNAs. A partir de curvas de sobrevida com p-valor significativo associadas à expressão dos 18 lncRNAs pertencentes à rede, foi observado o comportamento inverso de dois lncRNAs, EPB41L4A-AS1 e SNHG15, que quando subexpressos se relacionam com pior e melhor prognóstico, respectivamente. Por fim, estudos na literatura mostram que estes lncRNAs apresentam características para prováveis biomarcadores.

Palavras-chave: Carcinoma Renal. RNAs não codificantes. Análise de Expressão Diferencial. Rede de Competição Endógena. Análise de Sobrevida.

FARIAS FILHO, Epitácio Dantas de. **Transcriptome Analysis of Renal Clear Cell Carcinoma Based on Non-Coding RNAs.** Conclusion Work Project, Biomedical Engineering Bachelor Degree, Federal University of Rio Grande do Norte, 70p., 2022.

ABSTRACT

Clear cell renal carcinoma is the most common subtype among patients with cancer of the urinary system, presenting a poor prognosis when metastasized. Alterations at the genome and transcriptome level, specifically of mRNAs, have already been extensively analyzed, however recent studies reveal the importance of non-coding RNAs in the regulation of gene expression. This study aims to investigate the activity of RNAs, with a focus on non-coding genes in clear cell renal cell carcinoma by analyzing patient data from The Cancer Genome Atlas using bioinformatics techniques, analyzing cohort patient patterns, genomic and transcriptome alterations. Exploratory analysis showed that patient survival status is significantly correlated with disease staging. Differential expression analysis between normal and tumor tissues, identified 2,999 coding RNAs, 271 lncRNAs, and 132 miRNAs, with a Log2FC = 2 and p-value of 0.01. Functional analysis of the differentially expressed genes showed an enrichment of terms related to renal oncology pathways and kidney-related diseases. Finally, the construction of the endogenous competition network allowed the identification and visualization of common targets among 18 lncRNAs and 75 miRNAs. From survival curves with a significant p-value associated with the expression of the 18 lncRNAs belonging to the network, the inverse behavior of two lncRNAs, EPB41L4A-AS1 and SNHG15 when underexpressed are relate to a poor and better prognosis, respectively. Finally, studies have shown that these lncRNAs are cancer-related and potential biomarkers.

Keywords: Renal Carcinoma. non-coding RNAs. Differential Expression Analysis. Endogenous Competition Network. Survival Analysis.

1. INTRODUÇÃO

O câncer de rim, ou câncer renal, é uma doença que tem como característica base a proliferação descontrolada de células neoplásicas, que podem ocorrer em ambos os rins ou em somente um (MOTZER et al., 1996). A partir dos registros disponibilizados pelo *Global Cancer Observatory* (GCO) em 2020, foram confirmados mundialmente cerca 432 mil casos de cânceres renais e uma média de 179 mil mortes também causadas por essa patologia (IARC, 2022). Segundo dados da American Cancer Society de 2017, os pacientes de câncer renal apresentam uma taxa de sobrevida por cinco anos de 93% quando o tumor não entra em metástase e 15,3% quando em metástase. Estima-se que em 2022 sejam registrados cerca de 79.000 casos de câncer renal e aproximadamente 13.920 mortes, nos Estados Unidos da América (EUA), segundo dados do Programa de Vigilância, Epidemiologia e Resultados Finais (SEER) do Instituto Nacional do Câncer (NCI) dos Estados Unidos (SEER, 2022). Em território nacional, segundo dados disponibilizados pelo Instituto Nacional de Câncer, o câncer renal representou cerca de 0,23%, aproximadamente 3.630 mortes, dos óbitos relacionados a câncer no país em 2020 em ambos os sexos (INCA, 2022).

Apesar de apresentar baixos valores de incidência e mortalidade no mundo atualmente, o GCO projeta um grande crescimento nos registros de casos e mortes relacionados ao câncer de rim (PADALA *et al.*, 2020). As razões por esses baixos índices nos dias de hoje podem ser relacionadas a inúmeros fatores, tais como a forma de diagnóstico, dado que este carcinoma é encontrado muitas vezes de forma acidental, bem como as características ambientais ao qual o indivíduo acometido com a patologia está inserido, o estilo de vida que o paciente levava antes do diagnóstico e entre outros.

Somado aos fatores mencionados anteriormente, as alterações a nível do genoma e na regulação de processos celulares também estão ligadas aos números de incidência e mortalidade. O projeto Atlas do Genoma do Câncer (TCGA), entre 2006 e 2018 realizou a caracterização molecular de amostras de câncer primário, gerando um alto quantitativo de dados genômicos, proteômicos, transcriptômicos e epigenômicos. Segundo o *The Cancer Genome Atlas Research Network* (2013) a atividade dos ácidos ribonucleicos (RNAs) durante o desenvolvimento e nas diferenças de sobrevida dos pacientes acometidos com o carcinoma nas células claras do rim foram associadas significativamente com a progressão/remissão da doença.

O reconhecimento dos RNAs como estruturas de caráter regulatório no desenvolvimento do câncer, como em diversas outras patologias, tem sido estudado desde a década de 50 de

maneira transversal ao estudo do código genético e da estrutura helicoidal do ácido desoxirribonucleico (DNA) (MORRIS e MATTICK, 2014). Por ser uma área de estudo que segue em constante evolução, o conhecimento das atividades desempenhadas pelos RNAs tido como não-codificantes, que são moléculas de RNA que não serão traduzidas em proteína, dentro da biologia molecular ainda é inicial e necessita de mais estudos exploratórios. A realização de uma análise dos RNAs não-codificantes, visando compreender quais deles estão expressos no contexto da doença a ser estudada, permite realizar inferências sobre o comportamento e a evolução dela.

Diante desse contexto, é possível perceber que a atividade dos RNAs não codificantes está relacionada ao carcinoma renal de células claras e que a identificação dessas ações está atrelada geralmente ao enriquecimento de várias vias imunológicas específicas. Logo, compreender tais diferenças pode contribuir para uma melhor avaliação de risco e para o desenvolvimento de potenciais tratamentos que poderiam reduzir as mortes por carcinoma renal de células claras.

O trabalho será organizado seguindo a formatação: no Capítulo 2 são apresentados os objetivos, tanto gerais como específicos, do trabalho; no Capítulo 3 o referencial teórico é apresentado; no Capítulo 4 é feita a apresentação dos materiais e métodos aplicados; o Capítulo 5 traz os resultados; O capítulo 6 traz as discussões e por último, o Capítulo 7 apresenta as conclusões do estudo e trabalhos futuros.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Este trabalho tem como objetivo identificar e avaliar a função dos RNAs não codificantes no carcinoma renal de células claras e relacionar o impacto desses genes com o desenvolvimento e agravamento da doença**.**

2.2. OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste estudo consistem em:

- Analisar de forma exploratória os dados clínicos da coorte TCGA-KIRC;
- Identificar as principais alterações genômicas na coorte;
- Identificar os RNAs codificantes e não codificantes diferencialmente expressos;
- Avaliar o enriquecimento funcional dos RNAs diferencialmente expressos;
- Construir uma rede de RNAs endógenos competidores (ceRNAs) a partir dos RNAs codificantes e não codificantes diferencialmente expressos;
- Avaliar curva de sobrevida a partir dos RNAs participantes da rede de ceRNAs.

3. REFERENCIAL TEÓRICO

Neste capítulo será feita uma revisão bibliográfica sobre os principais temas relacionados a esse trabalho. Os assuntos abordados a seguir irão contribuir para um melhor entendimento sobre o assunto que será discutido.

3.1. EPIDEMIOLOGIA DO CÂNCER RENAL

Segundo Padala *et al.* (2020) o câncer renal, em suas posições anatômicas tida como renal e pélvica renal, tem se tornado o subtipo patológico com um grande aumento de notificações, tornando-o mais comum no "mundo desenvolvido" nas últimas décadas. Por se tratar de uma doença onde ocorre uma multiplicação desordenada das células renais, existem variações do câncer de rim relacionadas ao tipo celular atingido (MOTZER *et al.*, 1996), que será abordado na seção 3.2.

De acordo com os dados do Programa de Vigilância, Epidemiologia e Resultados Finais (SEER) do Instituto Nacional do Câncer (NCI) dos Estados Unidos da América (EUA), estima-se para 2022 aproximadamente 79.000 novos casos de câncer de rim e pelve renal, representando cerca de 4,1% de todos novos casos de câncer. Atrelado à este dado, calcula-se que haverão a de 2,3% de mortes registradas por câncer no território estado-unidense equivalente a aproximadamente 13.920 mortes, o que faz com que ele fique entre os 10 cânceres mais comuns no país (PADALA *et al.*,2020).

Ao ser comparado com outros tipos de câncer, o carcinoma renal apresenta um índice de sobrevida de 5 anos de aproximadamente 93%. Partindo para uma observação baseada no estadiamento da doença, quando o câncer se espalha para os linfonodos regionais a taxa sobrevida de 5 anos é de aproximadamente 72,3%, e quando se espalha para regiões distantes, caracterizando um câncer metastático, a taxa de sobrevida passa a ser de 15,3%, como podemos observar na Figura 1 (SEER, 2022). O câncer renal apesar de ser mais comumente diagnosticado em homens na faixa etária dos 65 aos 74 anos de idade, é também diagnosticado em mulheres com idade média em torno de 65 anos (HOWLANDER *et al.*, 2020).

No Brasil, dados do Departamento de Informática do Sistema Único de Saúde (DATASUS) mostram que foram diagnosticados aproximadamente 1.941 casos de carcinoma renal, exceto pelve renal, dentre os 259.542 casos de câncer registrados. Este quantitativo de casos de carcinoma renal representa uma incidência de cerca de 0,75% dos casos de câncer no último ano. Em 2020, as mortes por carcinoma renal representaram cerca de 0,23% das mortes

relacionadas a câncer no país, o que em números brutos representa aproximadamente 3.630 mortes por carcinoma renal (INCA, 2020).

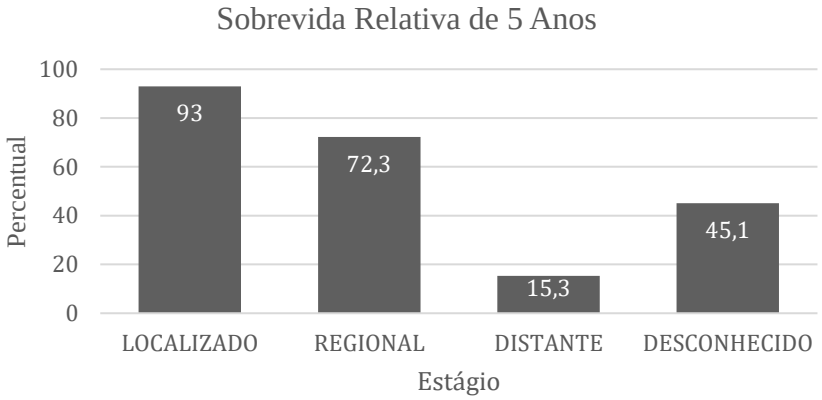


Figura 1: Taxa de sobrevida de 5 anos (1969-2019), para pacientes com carcinoma renal e pelve renal a partir do estadiamento da doença.
Fonte: Adaptado SEER, 2022.

Por se tratar de uma patologia multifatorial, existem diversos fatores (Tabela 1) que podem influenciar na sua progressão, relacionados às características ambientais, o estilo de vida, doenças pré-existentes relacionadas a síndromes genéticas ou neoplasias hereditárias, como por exemplo Síndrome de von Hippel-Lindau, Esclerose Tuberosa, Síndrome papilar hereditária, Síndrome de Birt-Hogg-Dubé, e entre outros fatores (NABI *et al.*, 2018; MOTZER *et al.*, 1996; PADALA *et al.*,2020).

Tabela 1:Fatores de Risco Relacionados ao Carcinoma Renal

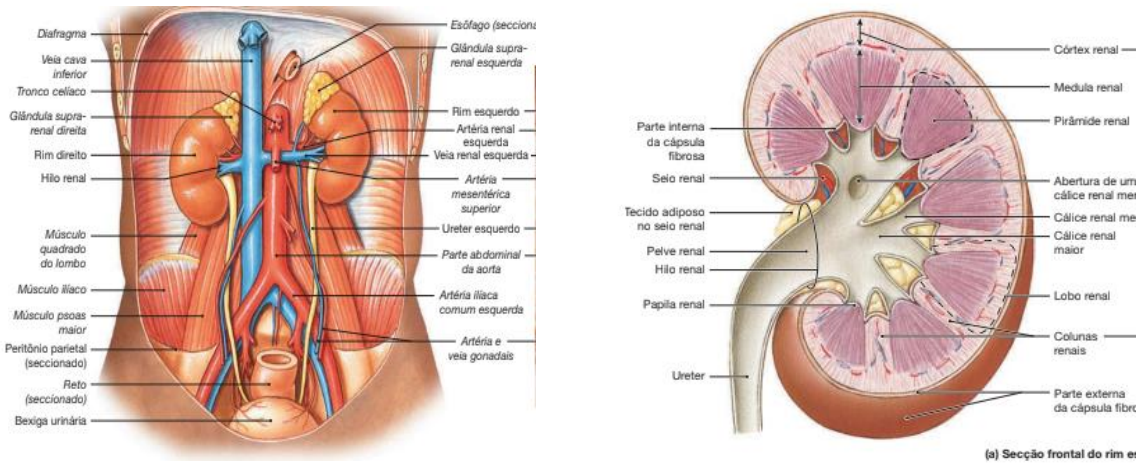
Fatores de Risco	
Modificáveis	**Não-modificáveis**
Tabagismo	Idade
Obesidade	Sexo
Dieta	Raça
Hipertensão Exposição Ocupacional	Doenças pré-existentes

Fonte: Nabi *et al.*(2018), Motzer *et al.* (1996), Padala *et al.* (2020).

3.2. FISIOPATOLOGIA DO CÂNCER DE RIM

Os rins são órgãos bilaterais, em conjunto das outras estruturas que compõem o sistema urinário, localizam-se em região retroperitoneal, entre os músculos do dorso e o peritônio

parietal (Figura 2a). Ao observar superficialmente, o rim tem um formato estrutural equivalente ao de um feijão, possuindo uma coloração castanha-avermelhada, em adultos saudáveis. Quando se realiza um corte coronal em sua estrutura é possível visualizar internamente as seguintes estruturas: córtex renal, medula renal e o seio renal (Figura 2b) (MARTINI *et al.*,2009).



(a) (b)

Figura 2:(a) Posicionamento anatômico, (b) corte coronal do rim.
Fonte: Martini et al., 2009.

Sabe-se que o câncer renal está fortemente relacionado aos fatores de risco ao qual o paciente pode estar exposto, como os listados anteriormente na seção 3.1 e extensamente discutidos por Nabi *et al.* (2018), Motzer *et al.* (1996) e Padala *et al.* (2020). A origem do câncer de rim ocorre no córtex renal, que tem em sua estrutura o glomérulo, os aparatos tubulares e o ducto coletor (Figura 2b) e se classifica histologicamente em três tipos (Tabela 2).

Como apresentado na Tabela 2, a partir de sua classificação histológica o câncer renal apresenta uma taxa de incidência, relacionando o subtipo e sua ocorrência perante os casos registrados. O carcinoma renal de células claras é o mais comum entre os três tipos, com uma incidência variando entre 70% e 80%, já os papilares e os cromófobos são menos comuns com taxas de incidência que variam de 10% a 15%, e de 5%, respectivamente. Independentemente do tipo célula que lhe deu origem, o carcinoma renal pode manifestar-se tanto unilateral quanto bilateralmente, ou seja, tanto somente em um ou nos dois rins (DALL'OGLIO *et al.*,2006).

Tabela 2:Classificação histológica do Carcinoma Renal.

Tipo	Características	Incidência (%)
Carcinoma Renal de Células Claras	Caracteriza-se pela coloração clara, "cristalina", que as células neoplásicas assumem depois de passar pelo tratamento histológico e ao ser visualizada pelo equipamento de microscopia óptica.	70 - 80
Carcinoma Renal Papilar	As células neoplásicas apresentam um padrão de crescimento papilar e o epitélio tumoral ainda apresenta resquícios do tecido epitelial proximal dos ductos contorcidos, onde iniciou-se o processo de oncogênese.	10 - 15
Carcinoma Renal Cromófobo	É caracterizado pela presença de células neoplásicas em formatos poligonais e com as membranas celulares bem proeminentes, e diferentemente do carcinoma de células claras, os vasos sanguíneos tumorais são bastante espessos e hialinizados de forma excêntrica.	5

Fonte: Prasad *et al.* (2006) e Dall'Oglio *et al*. (2006).

Por se tratar de uma doença silenciosa, devido ao tamanho da estrutura de origem, o diagnóstico do carcinoma renal ocorre frequentemente de maneira inesperada durante exames clínicos de rotina, como exames de urina, imagens (ultrassonografia, tomografia computadorizada ou ressonância magnética), ou em metástases, quando o paciente apresenta quadro clínico da clássica tríade de hematúria, dores lombares e presença de massas palpáveis na lombar, chegando também a ter sinais de febre, perda de peso e leucocitose (NKF, 2017).

3.3. MECANISMOS MOLECULARES E INFLUÊNCIA DOS RNAs NÃO CODANTES NO CÂNCER

Os tumores renais, mesmo de uma mesma origem celular, apresentam uma heterogeneidade entre si, ou seja, podem apresentar características histológicas e clínicas distintas fazendo com que diferentes prognósticos e tratamentos sejam identificados e aplicados. Essas diferenças podem, de certa forma, serem justificadas pela existência de diversas classificações do carcinoma renal (Tabela 2) a partir da sua origem celular, exigindo uma terapia específica (DALL'OGLIO *et al*.,2006).

Modificações a nível do genoma, transcriptoma e proteoma são comuns dentro do processo de oncogênese e estadiamento da doença, e diversos estudos têm sido realizados para conhecer essas variações, como eles influenciam no desenvolvimento, no prognóstico, na intervenção terapêutica utilizada e afins (GUIMARÃES, 2020).

Com este propósito, o programa Atlas do Genoma do Câncer (TCGA, 2013) realizou diversos estudos a partir de uma base de dados composta por 537 pacientes diagnosticados com carcinoma renal de células claras (KIRC-TCGA). Estudos voltados para as características clínicas e patológicas, alterações genômicas, como a alteração no número de cópias, perfis de metilação de DNA, expressão de RNAs e assinaturas proteicas foram realizados, permitindo assim uma análise global dos mecanismos moleculares desta patologia.

Neste estudo (TCGA, 2013) foi identificada uma gama de alterações a nível do genoma, como a perda do braço cromossômico 3p, abrangendo todos os quatro genes mais comumente mutados (VHL, PBRM1, BAP1 e SETD2). Foram identificados também 19 genes significativamente mutados, como o gene supressor de tumor VHL, envolvido no processo de oxigenação celular, e o gene PBRM1 e o complexo SWI/SNF, que atuam na manutenção dos estados das cromatinas. Além disso, foi observado que várias destas podem estar relacionadas com o estadiamento e severidade da doença. Adicionalmente, a partir da análise do sequenciamento de mRNAs e miRNAs, foram identificados clusters distintos, indicando as interações destes subtipos de RNAs como um importante componente da regulação do ccRCC (TCGA, 2013). Os subtipos moleculares do ccRCC e os eventos que promovem a letalidade do tumor, bem como a interação entre as células tumorais e seu microambiente, foram revistos e também foram elucidados os mecanismos essenciais no desenvolvimento do processo tumoral, e os estudos apontam suas possíveis implicações terapêuticas (JONASCH *et al.*, 2021).

3.4. RNAs E SEUS PAPÉIS NA ONCOGÊNESE

O papel dos RNAs para a biologia molecular foi estabelecido desde a década de 50, após o conhecimento da estrutura de dupla hélice do DNA, e foi inicialmente relacionado à maquinaria envolvida na síntese de proteínas, incluindo os RNAs mensageiros (mRNA), os RNAs ribossômicos (rRNA) e os RNAs transportadores (tRNA) (MORRIS e MATTICK, 2014). Entretanto, apenas no fim do século 20 começou a se perceber a real complexidade do transcriptoma. O consórcio Enciclopédia Elementos do DNA (ENCODE) através do projeto GENCODE, analisou o genoma humano e dos 85% que é transcrito, apenas cerca de 1% (Figura 3) é composto por RNA codificador de proteína, como apresentado por Djebali *et al.* (2012)

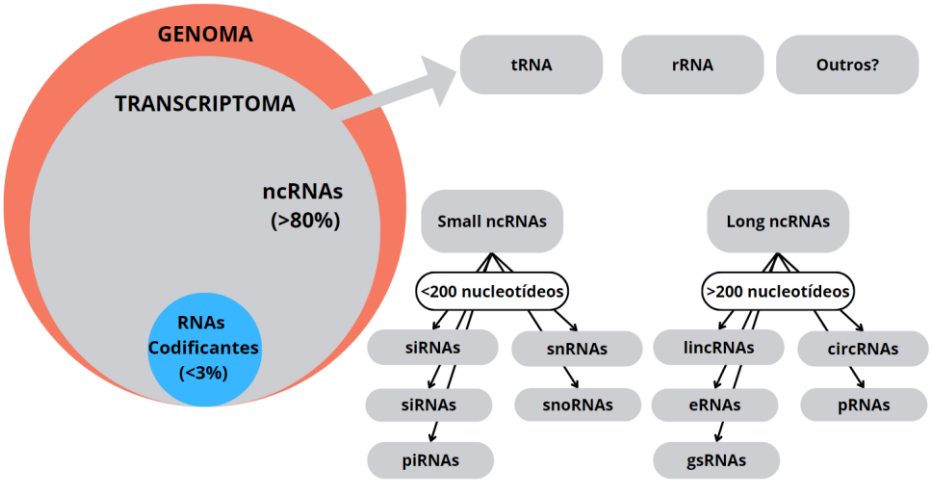


Figura 3: RNAs codificantes e não codificantes do genoma humano e suas classificações em relação a quantidade de nucleotídeos.
Fonte: Adaptado de Gomes *et al.* (2019).

Os RNAs não codificantes (ncRNAs) são aqueles que não são traduzidos em uma proteína funcional. Estruturalmente, os RNAs não codificantes podem ser classificados pela quantidade de nucleotídeos que os compõem - caso o RNA possua menos de 200 nucleotídeos ele é tido como um RNA pequeno não-codificante (sncRNA) e caso possua mais de 200 nucleotídeos, é chamado de RNA longo não-codificante (lncRNA) (KLINGE, 2018). Funcionalmente, os ncRNAs desempenham diversos papéis no ambiente celular, dependendo de sua localização e interações específicas com DNA, proteínas ou outros RNAs. Várias evidências mostram que os ncRNA estão envolvidos com processos de regulação do ciclo celular, diferenciação e desenvolvimento, como apoptose e metástases, além de processos de regulação epigenética, guiando as enzimas modificadoras de cromatina para os sítios de ação (MORRIS e MATTICK;2014).

Quando desregulados, os lncRNAs agem em conjunto, afetando os genes relacionados ao câncer, modificando a atividade dos fatores de transcrição (TF), das proteínas ligantes de RNA (RBPs) e os efetores de miRNAs ou reguladores canônicos. A partir da construção e análise de uma rede de lncRNAs, Chiu *et al.* (2018) sugere que inúmeros lncRNAs são especialistas em transcrição quanto em pós-transcrição, agindo de diversas formas: (i) como "iscas" para modular os efetores de seus alvos; (ii) como cofatores ou guias para alterar as interações promotor-TF; (iii) e como um "interruptor molecular" para alterar atividade de TF, ou RBPs, através de múltiplos alvos.

As evidências do envolvimento dos lncRNAs na oncogênese permitiram que eles fossem reconhecidos como marcadores de prognósticos tumorais, relacionado-os aos dados clínicos dos pacientes, como na progressão tumoral, sobrevida do paciente e infiltração imune (STAELLO *et al.*, 2020).

3.4.1. REDE DE RNAs ENDÓGENOS COMPETIDORES (ceRNAS)

Com o avanço dos estudos a respeito dos RNAs não codantes, como apresentado anteriormente, têm-se aumentado o entendimento sobre a associação dos lncRNAs com atividades de regulação, mas esse entendimento ainda é raso, quando comparado com o entendimento das funções desempenhadas pelos RNAs codantes. Em um ensaio publicado em 2011, Salmena e colaboradores apresentam a hipótese do "RNA endógeno competidor" (ceRNA), tido por ele como a "Pedra Rosetta da comunicação escondida do RNA", por apresentar uma "linguagem/comunicação" mediada pelos sítios de ligações dos miRNAs, ou elementos reconhecedores de miRNAs (MREs) (SALMENA *et al.*, 2011). Essa comunicação se torna uma forma de regulação desempenhada pelos miRNAs, uma vez que um único miRNAs pode regular mRNAs e também são regulados por outros ncRNAs, como os lncRNAs, formando uma rede de regulação/competição endógena. (ALA, 2020).

Como os lncRNAs podem desempenhar a função de modular os efetores dos seus alvos, diversos estudos mostram que ao realizar essa função com os miRNAs eles modulam a expressão gênica e caso exista um descontrole em algum componente dessa rede isso poderia afetar o complexo regulatório e culminar no desenvolvimento e na progressão do câncer (CHAN E TAY, 2018).

Com isso, a partir dessa necessidade de entender o papel desempenhado pelos RNAs não codificantes na oncogênese e no prognóstico de pacientes com câncer, o intuito deste trabalho é analisar os dados de carcinoma renal de células claras, partindo da perspectiva da rede de RNAs endógenos competidores composta pela tríade lncRNA-miRNA-mRNA.

3.5. REPOSITÓRIO DE DADOS E FERRAMENTAS PARA ANÁLISE EM BIOINFORMÁTICA

Como mencionado anteriormente, os estudos a respeito das atividades a nível de genoma e de transcriptoma vêm sendo desenvolvidos para compor os estudos relacionados a diversas patologias, o que resulta na existência de um grande quantitativo de dados biológicos. Apesar da existência deste montante de dados, as informações que eles trazem não se apresentam de

forma sucinta e direta, fazendo com que necessite de uma aplicação computacional para a extração das informações mais relevantes. Agindo como um ponto de intersecção entre a biologia e a ciência da computação, a Bioinformática é o campo de estudo que busca além de extrair as informações, realizando análises e tratamento dos dados que muitas vezes se apresentam em um elevado grau de complexidade para o entendimento, realizar modelagem e simulações de sistemas biológicos e afins.

Os resultados obtidos por análises de bioinformática podem contribuir para um melhor entendimento da patologia, ou outro tópico que esteja sendo estudado, demonstrando as causas e efeitos da doença, a incidência/prevalência, a influência do tratamento e/ou técnicas utilizadas, como também auxiliar na criação e identificação de novos modelos e hipóteses de estudo. Alguns dos repositórios de dados e o ambiente de desenvolvimento utilizados no desenvolvimento deste estudo são apresentados a seguir.

3.5.1. GENOMIC DATA COMMONS - GDC

O repositório *Genomic Data Commons* (GDC) (https://portal.gdc.cancer.gov) desenvolvido pelo NCI apresenta os maiores e mais abrangentes conjuntos de dados genômicos do câncer, dentre eles o TCGA e Pesquisa Terapeuticamente Aplicável para Gerar Tratamentos Eficazes (TARGET) (ZHANG,2021). Cada um desses projetos possui uma variedade de tipos de dados moleculares processados e não processados, incluindo genômica, epigenômica, proteômica, imagem, clínica e outros.

3.5.2. XENA BROWSER

O *Xena Browser* (https://xenabrowser.net) é uma plataforma que supre a necessidade que os bioinformatas têm de visualizar os dados de câncer a partir de diferentes fontes, públicas ou privadas, permitindo ao usuário uma forma segura e possível de integração com outros conjuntos de dados ômicos (GOLDMAN *et al.*; 2020). O Xena complementa algumas outras ferramentas de análise de dados existentes, tais como cBioPortal, Portal ICGC, Portal GDC, IGV e se destaca em mostrar tendências em coortes de mostras, células ou linhagens celulares, sendo de extrema importância para os pesquisadores, auxiliando-os a descobrir e validar seus resultados. Apesar de apresentar um foco para a genômica do câncer, a plataforma é completa o suficiente para hospedar quaisquer outros dados genômicos funcionais.

3.5.3. BIOCONDUCTOR

O Bioconductor (https://www.bioconductor.org) é uma iniciativa de criação colaborativa de *software* para biologia computacional e bioinformática baseada em linguagem R (GENTLEMAN, 2004). Desde 2001, o projeto tem como meta possibilitar o acesso a métodos estatísticos e gráficos para a análise de dados genômicos, fornecer uma plataforma para rápido desenvolvimento e implementação de softwares e facilitar a inclusão de metadados biológicos na análise de dados genômicos. Os objetivos do Bioconductor são: (i) simplificar os processos; (ii) promover análises estatísticas e melhor compreensão de ensaios biológicos; (ii) habilitar os usuários e desenvolvedores a explorar e interagir com dados e algoritmos para obter inferências precisas. Para tanto, este projeto apresenta uma documentação acessível e escalável, além de ser *open-source* e *open-development*.

3.5.4. RSTUDIO

O RStudio (https://www.rstudio.com) é um ambiente de desenvolvimento *open-source* para a linguagem de programação estatística R, disponível para diferentes tipos de sistemas operacionais, o qual, através de uma interface dinâmica e intuitiva, torna possível o estudo estatístico de dados e sua visualização em gráficos. Visualmente a ferramenta é dividida em quatro janelas que permitem uma maior exploração dos seus recursos: o console onde o código é executado, o ambiente de desenvolvimento do código, o ambiente de exibição de variáveis e o ambiente de plotagem de gráficos e manuseio dos arquivos, pacotes e ajuda. Neste *software* existem algumas funções pré-definidas, mas por meio de rotinas de instalação e ativação de pacotes desenvolvidos para fins específicos é possível ter um maior aproveitamento de uma infinidade de outras funções para complementar o seu estudo.

4. METODOLOGIA

4.1. DADOS DOS PACIENTES

Os dados dos pacientes com carcinoma renal de células claras do projeto KIRC-TCGA foram consultados a partir do portal *Genomic Data Commons* (GDC) e do Xena Browser, sendo obtidos os dados na versão de *Data Release* v18.0 (07-19-2019). Foram analisados 3 conjuntos de dados (Figura 4): (1) dados clínicos, composto por informações sobre idade ao ser diagnosticado, tempo de remissão do paciente sem sintomas da doença, tempo de sobrevida, ocorrência de diagnóstico de câncer anterior, entre outros; (2) dados genômicos, obtidos a partir do sequenciamento completo do exoma (WXS) de cada paciente e apresentado em arquivos do Formato de Anotação de Mutação (MAF); (3) dados de contagem transcriptoma, obtidos a partir do sequenciamento dos transcritos (RNASeq).

A mesma estrutura de código de barras TCGA é usada para dados clínicos e moleculares, permitindo a análise integrada de dados clínicos de pacientes e dados moleculares baseados em amostras. Amostras de DNA normais foram fornecidas como alíquotas de rim normal não envolvido no processo tumoral ou alíquotas de sangue de pacientes.

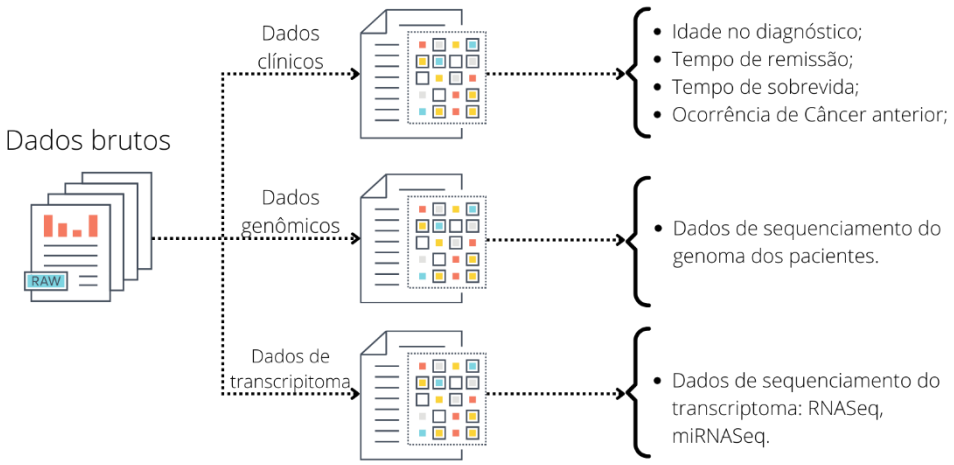


Figura 4:Subconjuntos dos dados do projeto KIRC-TCGA.
Fonte: Autor.

4.2. AQUISIÇÃO E ANÁLISE EXPLORATÓRIA DE DADOS

Para a aquisição dos dados junto ao GDC utilizou-se dos pacotes TCGAWorkflow, versão 1.19.1 (SILVA *et al.*, 2016) e TCGAbiolinks, versão 2.24.3 (COLAPRICO *et al.*, 2015) e para a etapa de pré-processamento e análise dos dados clínicos foram utilizados os pacotes

Tidyverse, versão 1.3.0 (WICKHAM et al., 2019), SkimR, versão 2.1.4 (WARNING, *et al.*, 2022), Finalfit, versão 1.0.4 (HARISSON *et al.*, 2021), Tableone versão 0.13.2 (YOSHIDA *et al*, 2022). Inicialmente, para a limpeza dos dados clínicos, todas as variáveis que apresentavam menos de 50% de observações foram removidas e mantidas apenas as que apresentavam registros únicos dos pacientes. A aplicação de boas práticas no tratamento dessas variáveis, como a checagem das frequências, alterações no nome, e adequação do tipo e níveis das variáveis foram realizadas.

A análise exploratória da distribuição das variáveis numéricas, feita a partir da inspeção visual de gráficos univariados, de histogramas e tabelas para dados categóricos, possibilitou o reconhecimento de 24 variáveis significativas, tomando um p-valor <0.01 como nível de significância, para este estudo (APÊNDICE A).

4.3. ANÁLISE DE ALTERAÇÕES GENÔMICAS

As análises genômicas foram realizadas sobre arquivos do tipo *Mutation Annotation Format* (MAF) utilizando o pacote Maftools, versão 2.6.05 (Mayakonda *et al.*, 2018).

Foi realizado um panorama do conjunto de dados e algumas importantes alterações a nível de genoma, como: (i) os tipos de variações existes; (ii) a classificação dessas variações; (iii) as classes das variações de nucleotídeo único; (iv) a quantidade de variantes por amostra e (v) os 10 genes mais mutados na coorte. Por fim, foi realizada a checagem das vias de sinalização oncogênicas, analisando por meio de gráficos de barras, as frações das vias afetadas e a quantidade de amostras do conjunto de dados que apresentaram alterações nestas vias.

4.4. ANÁLISE DE ALTERAÇÕES DO TRANSCRIPTOMA

As análises relacionadas às alterações do transcriptoma (Figura 5) foram realizadas utilizando dados de sequenciamento de RNAs (RNASeq) e os pacotes GDCRNATools, versão 1.16.2 (Li *et al.*, 2022) e gprofiler2, versão 0.2.1 (KOLBERG *et al.*, 2020).

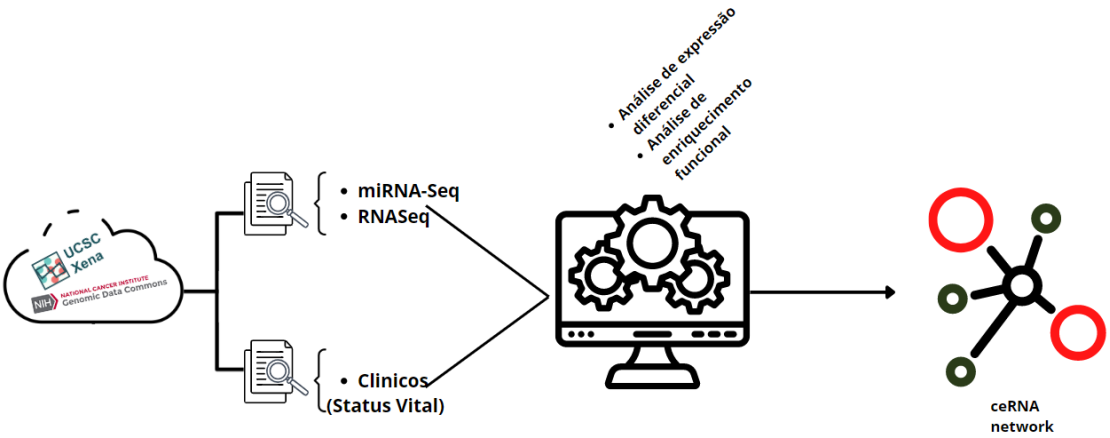


Figura 5: Organização da análise transcriptômica.
Fonte: Autoral.

4.4.1. ANÁLISE DE EXPRESSÃO DIFERENCIAL

Os dados de expressão dos RNAs foram obtidos a partir dos dados de contagem de RNASeq e de miRNA-Seq, seguindo o método da média comparada do componente M (TMM), que consiste numa forma simples e efetiva de estimar os níveis de RNA (Robinson e Oshlack, 2010), e da transformação dos dados de contagem para o formato de log2-contagens por milhão (logCPM), estimando a relação média-variância (Ritchie *et al.*, 2015), para os dados de sequenciamento.

Com os dados de expressão já normalizados e transformados, foi realizada a análise de expressão diferencial entre os grupos de tecidos normais (72 amostras) e tumorais (530 amostras), baseando-se na metodologia aplicada pelo DESeq2, versão 1.36.0 (LOVE *et al.*, 2014), utilizando log2FC = |2| e FDR < 0.01 , incorporado ao pacote *GDCRNATools*. Este método foi escolhido por conseguir detectar um maior número de genes diferencialmente expressos (MONTI *et al.*, 2022).

4.4.2. ANÁLISE DE ENRIQUECIMENTO FUNCIONAL

A anotação funcional dos genes diferencialmente expressos foi realizada contra bancos de dados especialistas (APÊNDICE E) Para a análise de enriquecimento funcional foi utilizado o pacote gprofiler2, versão 0.2.1 (KOLBERG *et al.*, 2020) com FDR < 0,01.

4.4.3. REDE DE ceRNAs

A rede de ceRNAs foi construída a partir dos dados de expressão diferencial de mRNAs e ncRNAs, visando estabelecer as interações existentes entre os lncRNAs-miRNAs-mRNAs. Foi utilizado o pacote GDCRNATools, versão 1.16.2 (LI *et al.*,2022) e a base de dados

desenvolvida por Li et *al.* (2013), *starBase* (https://starbase.sysu.edu.cn/), que tem como objetivo decodificar as redes de interações celulares através de inúmeras RBPs e RNAs. A seleção dos pares lncRNAs-mRNA e miRNA-mRNA foi feita a partir de 3 análises: (i) teste hipergeométrico, que analisa a quantidade de miRNAs que um lncRNA e um mRNA compartilham de forma significativa; (ii) da correlação de Pearson, que avalia a força da associação entre os lncRNAs e os miRNAS; e por fim (iii) dos padrões de regulação que avaliam a similaridade de regulação, a partir da correlação de expressão entre lncRNA-miRNA e mRNA-miRNA, e avaliam a sensibilidade de correlação, para medir a correlação entre os mRNAs e lncRNAs e se ela é mediada por um miRNA (LI *et al.*,2022). Os pontos de corte desses resultados foram: p-valor do teste hipergeométrico e p-valor da correlação de Pearson < 0,01 e os padrões de regulação diferentes de 0, por fim foi utilizado o software *Cytoscape*, versão 3.9.1 (SHANNON et al., 2003) para visualização da rede.

4.5. ANÁLISE DE SOBREVIDA

Para as análises de sobrevida, se considerou o *overall survival status* e o *survival time* dos pacientes com alta ou baixa expressão dos lncRNAs que compõem a rede de ceRNAs. A probabilidade de sobrevida foi estimada usando o método de Kaplan-Meier e o teste de log-rank p < 0,01, seguindo as funções disponibilizadas no pacote GDCRNATools.

4.6. DESENVOLVIMENTO

Os *scripts* desenvolvidos neste estudo foram baseados na linguagem de programação estatística R, versões 4.0.4 e 4.2.0, utilizando o ambiente de desenvolvimento *RStudio* versão 2022.02.3+492. As implementações computacionais foram realizadas no servidor do Centro Multiusuário Bioinformática (BioME) do Instituto Metrópole Digital. Os códigos e *datasets* foram disponibilizados em um repositório no *GitHub* (https://github.com/epfarias/lncRNAsKIRC_analysis).

5. RESULTADOS

5.1. ANÁLISE EXPLORATÓRIA

Inicialmente, foi realizada uma análise exploratória dos dados clínicos para descrever as características básicas da coorte do TCGA-KIRC e após uma verificação cautelosa de cada uma das variáveis, foram selecionadas as 23 variáveis que traziam informações importantes relacionadas a aspectos demográficos e clínicos da coorte (APÊNDICE A). Este grupo apresentou uma idade mediana ao diagnóstico de 61 anos, sendo que 65% dos pacientes eram do sexo masculino, 35% do sexo feminino e em relação à raça 93% se declararam caucasianos, 3,4% afro-americanos/negros e 1,6% asiáticos. Uma representação considerável foi fornecida por sites europeus, alterando adicionalmente o equilíbrio de amostras de tumores em favor de doadores caucasianos. Os tumores apresentaram uma distribuição de estágio tumoral e grau típico da doença. Especificamente, 372 tumores representam doença localizada (fase I-estágio III), com 74 ocorrendo com doença metastática sincrônica (estágio IV).

Entre as variáveis selecionadas, o status vital (*vital_status*) é uma variável importante para explorar quais aspectos clínicos são mais correlacionados com a sobrevida, com a vantagem de haver ambiguidade mínima na definição do evento - o paciente está vivo ou morto (LIU *et al.*, 2018). Esta relação é mostrada na Tabela 3, onde podemos observar uma relação significativa entre o status vital e as variáveis estágio da doença segundo o Comitê Conjunto Americano de Câncer (*ajcc_pathologic_stage*), etnia (*ethnicity*), tratamento ou terapia farmacêutica (*pharmaceutical_treatment_or_theraypy*), terapia ou tratamento à base de radiação (*radiation_treatment_or_therapy*), dias até o último acompanhamento (*days_to_last_follow_up*), idade (*age*). Por outro lado, não foi observada uma correlação entre doença síncrona (*synchronous_malignancy*), doença anterior (*prior_malignancy*), gênero (*gender*).

Tabela 3:Distribuição das variáveis clínicas em relação ao status vital na coorte TCGA-KIRC.

Variável dependente: Status Vital		**Vivo**	**Morto**	**p-valor**
Variável independente				
synchronous_malignancy	No	308 (85.6)	153 (86.4)	0.371
	No reported	48 (13.3)	24 (13.6)	
	Yes	4 (1.1)	0 (0.0)	
ajcc_pathologic_stage	Stage I	224 (62.6)	45 (25.6)	<0.001
	Stage II	44 (12.3)	13 (7.4)	
	Stage III	74 (20.7)	51 (29.0)	
	Stage IV	16 (4.5)	67 (38.1)	
primary_diagnosis	Clear cell adenocarcinoma	346 (96.1)	177 (100.0)	0.018
	Renal cell carcinoma	14 (3.9)	0 (0.0)	
prior_malignancy	no	310 (86.1)	153 (86.4)	1.000
	yes	50 (13.9)	24 (13.6)	
prior_treatment	No	352 (97.8)	167 (94.4)	0.069
	Yes	8 (2.2)	10 (5.6)	
race	asian	7 (1.9)	1 (0.6)	0.074
	black or african american	45 (12.5)	11 (6.2)	
	not reported	5 (1.4)	2 (1.1)	
	white	303 (84.2)	163 (92.1)	
gender	female	126 (35.0)	65 (36.7)	0.767
	male	234 (65.0)	112 (63.3)	
ethnicity	hispanic or latino	23 (6.4)	3 (1.7)	0.008
	not hispanic or latino	227 (63.1)	132 (74.6)	
pharmaceutical_treatment_or_theraypy	no	164 (45.6)	29 (16.4)	<0.001
	not reported	171 (47.5)	89 (50.3)	
	yes	25 (6.9)	59 (33.3)	
radiation_treatment_or_therapy	no	175 (48.6)	49 (27.7)	<0.001
	not reported	170 (47.2)	86 (48.6)	
	yes	15 (4.2)	42 (23.7)	
days_to_last_follow_up		1534.2 (1020.8)	890.4 (702.6)	0.003
Age		58.8 (11.9)	64.2 (12.0)	<0.001

*As variáveis quantitativas são expressas como média e (desvio-padrão)
** As variáveis qualitativas são expressas como contagem e (porcentagem).
Fonte: Autoral.

5.2. ANÁLISE DE ALTERAÇÕES GENÔMICAS

A análise das alterações a nível do genoma possibilita avaliar o impacto destas sobre seu produto, podendo ser a interrupção da produção de proteínas, alteração da própria proteína,

ou até mesmo, uma produção de proteínas de forma desordenada e em alta quantidade. Utilizando o pacote Maftools, avaliamos os polimorfismos de nucleotídeo único (SNPs) no genoma dos pacientes da coorte.

Dentre as mutações observadas em todo o conjunto de dados, Figura 6, a mutação *missense* foi a mais presente na coorte, apesar da maioria das mutações terem sido do tipo SNPs, com uma mediana de 44 mutações por amostra, também foram identificadas mutações baseadas em inserções e deleções. Ao observar as alterações a nível de nucleotídeo, foi possível notar diversas transições e transversões, onde a transversão entre as bases nitrogenadas citosina e timina foi a mais presente com 6.477 registros.

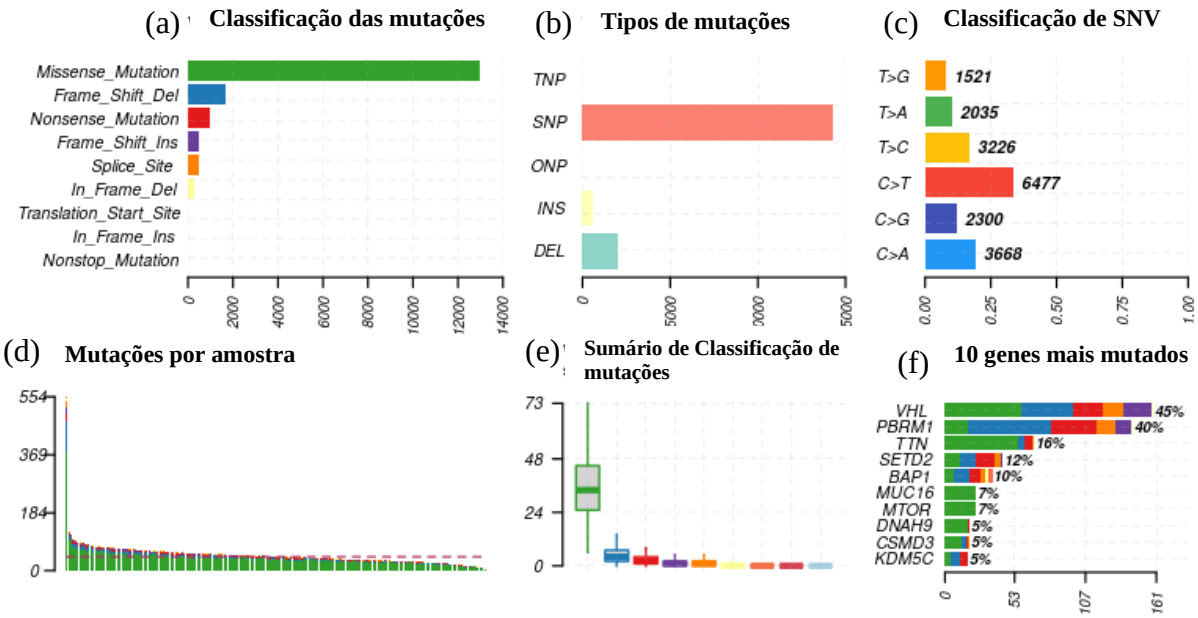


Figura 6: Sumário das alterações genéticas para TCGA-KIRC: (a) classificação e contagem das mutações somáticas; (b) os tipos e contagem das variações; (c) contagens de transições e transversões dos SNVs; (d) a quantidade de mutações por amostra; (e) o b
Fonte: Autoral.

A Figura 7 apresenta os 10 genes mais mutados VHL, PBRM1, TTN, SETD2, BAP1, MUC16, MTOR, DNAH9, CSMD3, KDM5C os tipos de mutações ocorridas para cada paciente do estudo, a fração do tipo de mutação em cada um dos 10 genes alterados e o quanto alterado o gene está dentro da coorte. Pode-se observar que cerca de 77,52% da coorte apresenta alguma mutação gênica e que dentre estas possíveis mutações, a *missense* e a *fram_shitf_del*, que representa as deleções nas janelas de leitura, são as mais comuns. O gene mais alterado foi o VHL, apresentando alteração em 45% dos pacientes do estudo.

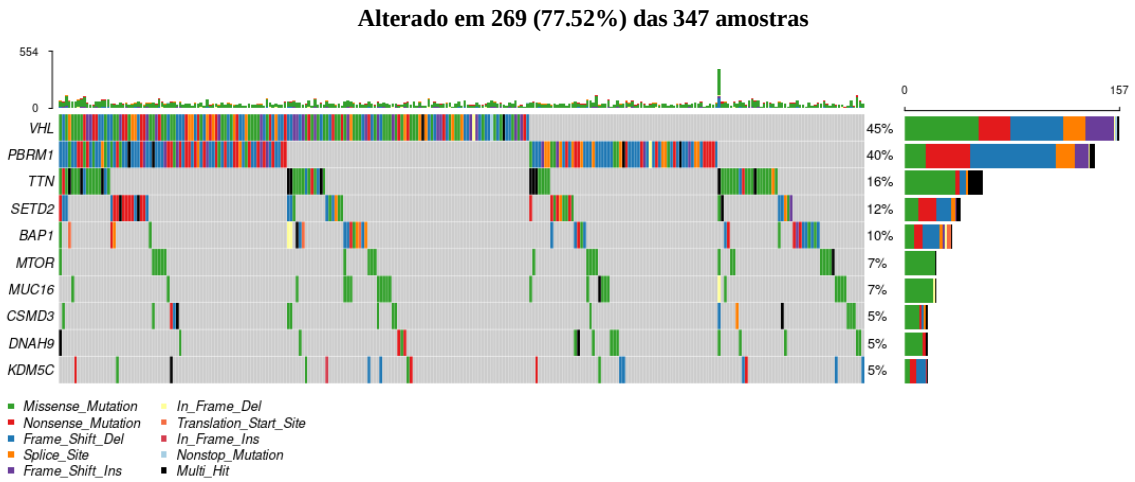


Figura 7: Oncoplot com os 10 genes mais frequentemente mutados. O histograma à direita do gráfico representa a porcentagem de alteração do gene em toda a coorte e os tipos de mutações sofridas, o histograma acima representa o quantitativo de mutações por paciente.
Fonte: Autoral.

A Tabela 4 apresenta a função e associação dos 10 genes mais frequentemente mutados. Os genes VHL e BAP1 apresentam papel de supressor de tumor, diversos deles apresentam alterações em cânceres, como: CSMD3 aparece alterado no Carcinoma do Endométrio do Corpo Uterino e Carcinoma de células escamosas de Cabeça e Pescoço; o VHL, PBRM1, MUC16 e KDM5C estão relacionados com o ccRCC. (SAFRAN et al., 2021)

A análise das vias de sinalização oncogênicas alteradas (SANCHEZ-VEGA et al., 2018) possibilita observar quais estão enriquecidas, tomando como base o número de genes alterados em cada via na coluna à esquerda, e a fração de amostras alteradas para cada via, na coluna à direita (Figura 8). Na coorte analisada as vias mais alteradas foram a RTK-RAS, NOTCH, WNT, Hippo e PI3K.

Tabela 4: Função e associações a doenças dos 10 genes mais frequentemente mutados.

(continua)

Gene	Função	Associação
VHL	Codifica um componente do complexo de ubiquitinação	Síndrome de Von Hippel-Lindau
PBRM1	Codifica complexos de remodelação da cromatina dependentes de ATP	Carcinomas renais: de células claras, não papilar e papilar de células claras
TTN	Codifica uma proteína grande e abundante de músculo estriado.	Miopatia de Salih e Miopatia Miofibrilar, 9
SETD2	Codifica uma proteína que interage com a huntingtina.	Síndrome de Luscan-Lumish e Síndrome de Sotos

Tabela 4: Função e associação a doenças dos 10 genes frequentemente mutados

(conclusão)

GENE	FUNÇÃO	ASSOCIAÇÃO
BAP1	Está envolvido na remoção de ubiquitina de proteínas	Síndrome de Kury-Isidor e Síndrome de Predisposição Tumoral
MTOR	Codifica genes da família de quinases relacionadas à fosfatidilinositol quinase	Displasia Cortical Focal, Tipo II e Síndrome de Smith-Kingsmore
MUC16	Codifica uma proteína que é membro da família das mucinas	Cistadenocarcinoma Seroso e Adenocarcinoma de Célula Clara
CSMD3	Regulação do desenvolvimento dendrítico	Síndrome Tricorrinofalangeana, Tipo II e Carcinoma do Endométrio do Corpo Uterino
DNAH9	Codifica a subunidade de cadeia pesada da dineína axonemal.	Discinesia Celular Primária e Dextrocardia com *Situs Inversus*.
KDM5C	Envolvidos na regulação da transcrição e remodelação da cromatina	Transtorno do Desenvolvimento Intelectual e Deficiência Intelectual, ambas do Tipo Claes-Jensen

Fonte: Genecards (https://www.genecards.org/).

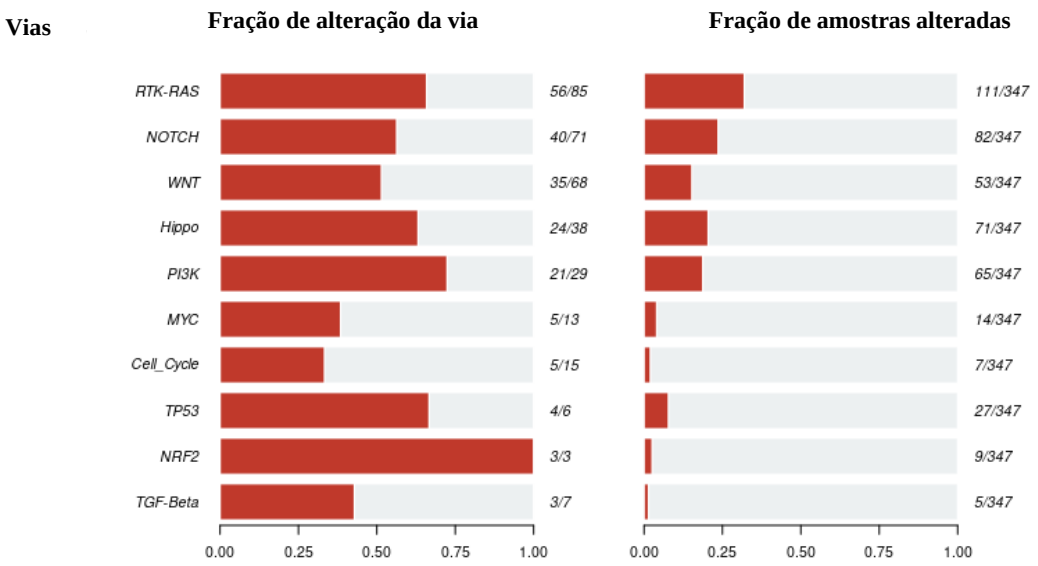


Figura 8: Principais vias oncogênicas alteradas. A esquerda a fração de alteração dos genes na via e a direita a fração de amostras com a via alterada.
Fonte: Autoral

5.3. ANÁLISE DE ALTERAÇÕES DO TRANSCRIPTOMA

5.3.1. ANÁLISE DE EXPRESSÃO DIFERENCIAL

Após a análise das alterações genômicas, realizamos a análise de expressão diferencial de genes codantes e não-codantes, entre os tecidos tumorais (530) e normais (72), utilizando um log2FC = |2| e FDR < 0,01. Dos 13.266 RNAs codificantes, somente 2.842 estão diferencialmente expressos, sendo 1872 superexpressos e 970 sub expressos.

Em relação aos RNAs não-codificantes, foram identificados 132 miRNAS diferencialmente expressos, de um total de 418 miRNAs, sendo 60 superexpressos e 72 sub expressos (Figura 09). Pode-se observar alguns miRNAs com uma grande variação, como por exemplo hsa-miR-508-3p e hsa-miR-514a-3p extremamente subexpressos, como também os hsa-miR-155-5p e hsa-miR-122-5p extremamente superexpressos, com um logFC > |2.5|

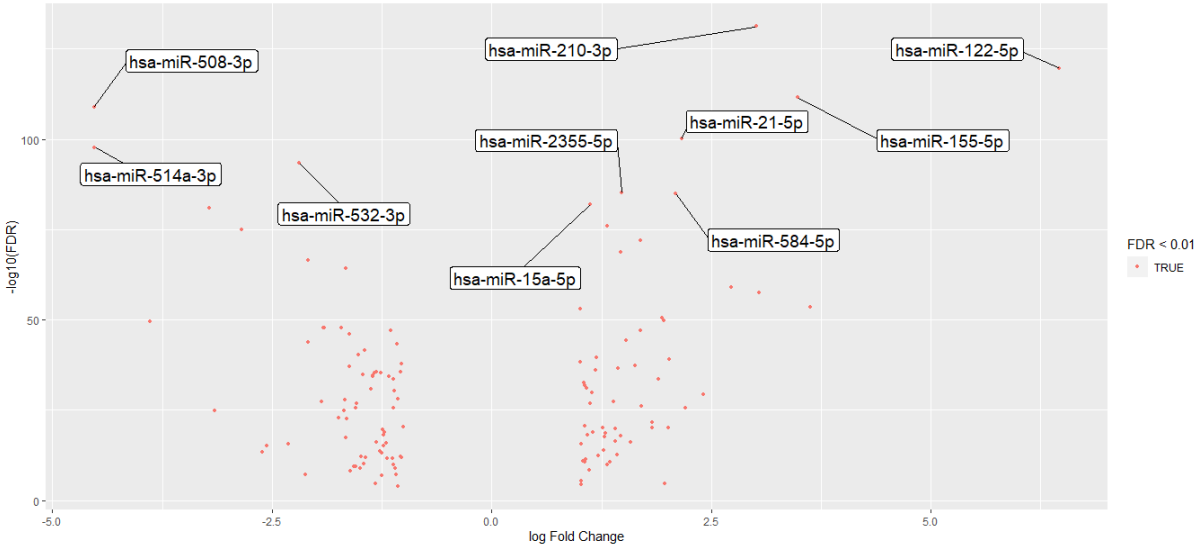


Figura 9: Volcano plot de miRNAs diferencialmente expressos entre amostras tumorais e normais. Os pontos em vermelho indicam genes diferencialmente expressos a uma taxa de falsa descoberta de 1% quando o ajuste de teste múltiplo de Benjamini-Hochberg é utilizado e logFC = 2.
Fonte: Autoral.

Em relação aos lncRNAs, foram identificados 271 diferencialmente expressos, de um total de 816 lncRNAs, sendo 221 superexpressos e 49 sub expressos (Figura 9). Novamente podemos observar lncRNAs com uma grande variação, como o AC079760.2, LINC00887, LINC00462, LINC02348 e TTC21B-AS1, todos apresentando um aumento de expressão, com logFC > 5.

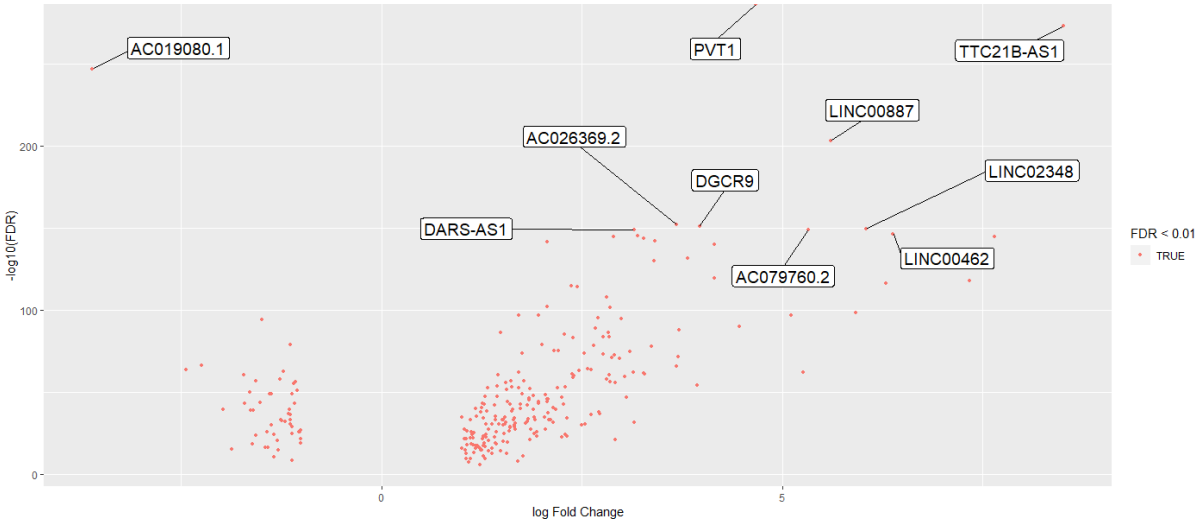


Figura 10: Volcano plot de lncRNAs diferencialmente expressos entre amostras tumorais e normais. Os pontos em vermelho indicam genes diferencialmente expressos a uma taxa de falsa descoberta de 1% quando o ajuste de teste múltiplo de Benjamini-Hochberg é utilizado e logFC = 2.
Fonte: Autoral

5.3.2. ANÁLISE DE ENRIQUECIMENTO FUNCIONAL

Para avaliar os processos ou condições em que os genes diferencialmente expressos identificados neste estudo podem estar envolvidos, foi realizada a análise de enriquecimento funcional contra as bases de dados listadas na metodologia. Na anotação destes dados contra os termos do banco de dados Disease Ontology (DO) observamos que os genes diferencialmente expressos estão relacionados a diversas doenças como câncer no sistema urinário, câncer renal, câncer pulmonar e doenças virais e bacterianas, com valores estatísticos significativos (Figura 12).

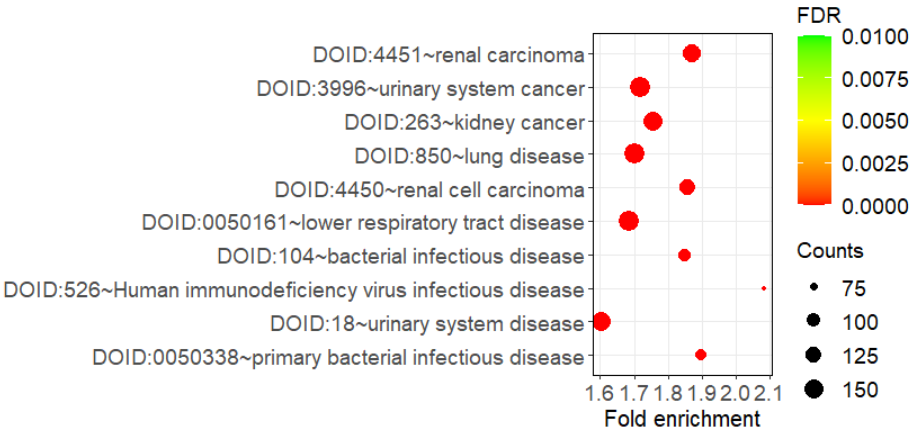
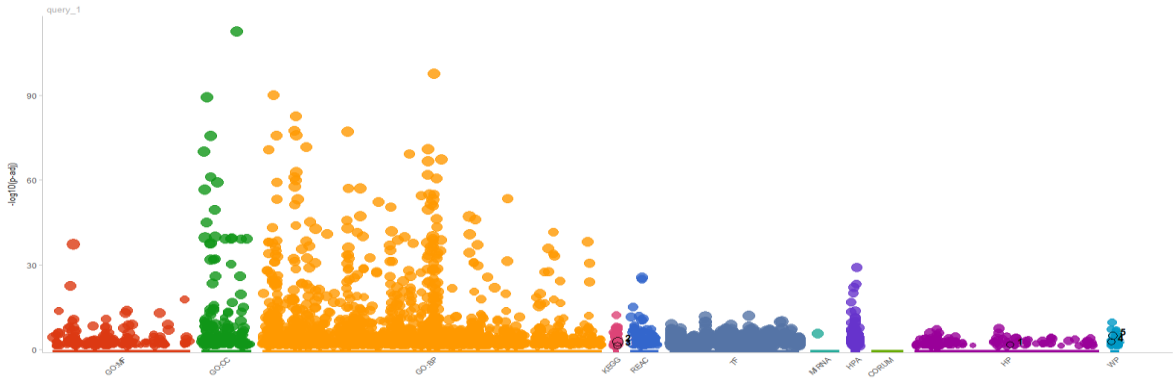


Figura 11: Principais vias enriquecidas pelo Disease Ontology para os mRNAs diferencialmente expressos.
Fonte: Autoral.

Quando feito o enriquecimento funcional contra as outras bases de dados funcionais (Figura 12), foi observado uma via de miRNA enriquecida e também foi realizada uma filtragem para as vias que estavam relacionadas diretamente com o câncer, independente do seu sítio de origem e também com foco no carcinoma renal. A partir dessa filtragem foram encontradas 6 vias metabólicas enriquecidas relacionadas a câncer tanto renal quanto de outros sítios de origem: HP:0012126, KEGG: hsa05200, KEGG: hsa05230, WP:WP4585, WP:WP4018 e a via hsa-miR-335-5p relaciona com os miRNAs.



(a)

id	source	term_id	term_name	term_size	intersection_size		p_value
1	HP	HP:0012126	Stomach cancer	28	13		2.3e-02
2	KEGG	KEGG:05200	Pathways in cancer	517	120		1.3e-03
3	KEGG	KEGG:05230	Central carbon metabolism in cancer	69	20		4.8e-02
4	WP	WP:WP4585	Cancer immunotherapy by PD-1 blockade	23	13		1.5e-03
5	WP	WP:WP4018	Clear cell renal cell carcinoma pathways	85	37		9.2e-06

g:Profiler (biit.cs.ut.ee/gprofiler)

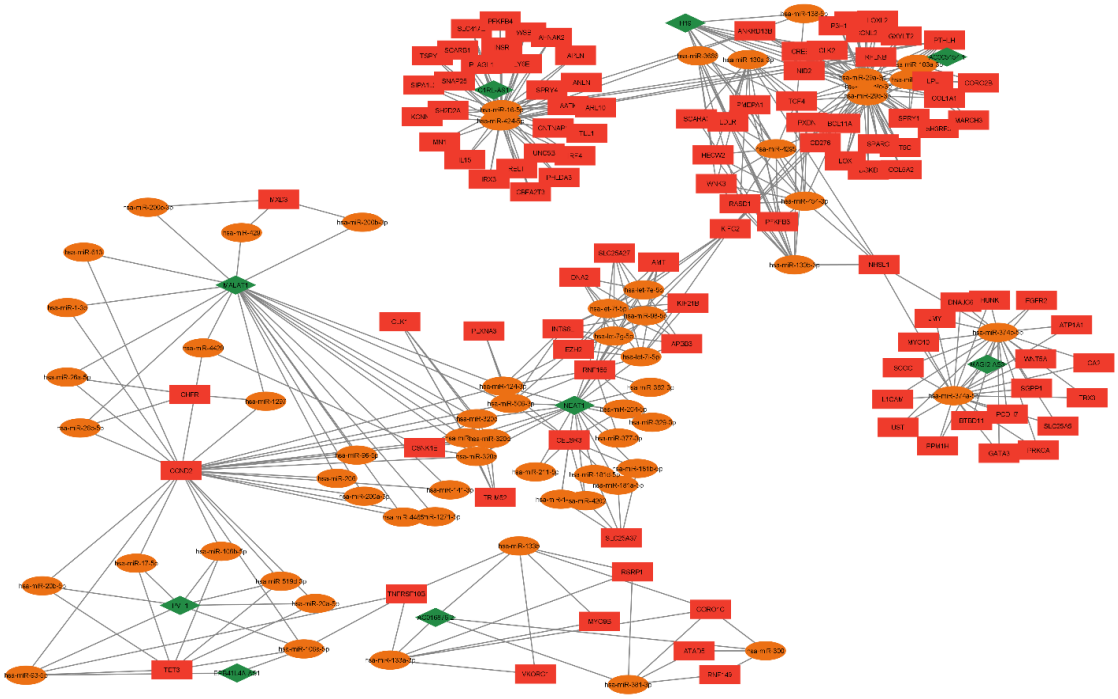
(b)

Figura 12: (a) Enriquecimento de mRNAs diferencialmente expressos contra base de dados; (b) enriquecimento funcional para as vias: HP:0012126, KEGG: hsa05200, KEGG: hsa05230, WP:WP4585, WP:WP4018 e hsa-miR-335-5p.
Fonte: Autoral.

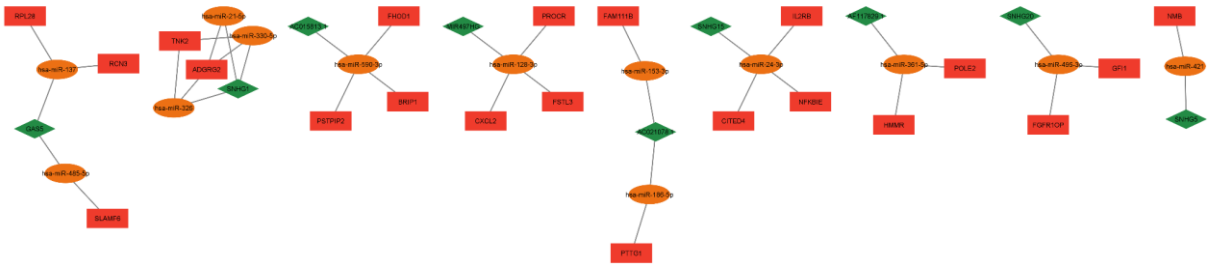
5.3.3. REDE DE ceRNAs

A partir dos lncRNAs diferencialmente expressos (DElncRNAs) e dos mRNAs diferencialmente expressos (DEmRNAs) foi construída a rede de ceRNAs (Figura 13(a) e (b)), sendo composta por 18 lncRNAs (APÊNDICE B), 128 mRNAs (APÊNDICE C) e 75 miRNAs, onde destes 75, 18 apresentaram-se como diferencialmente expressos na coorte (APÊNDICE D). Podemos observar a existência de um subgrafo maior com uma grande quantidade de ligações entre os miRNAs, os lncRNAs e os mRNAS. Neste subgrafo, observamos tanto regiões de onde muitos mRNAs se conectam por um ou dois miRNAs, com a presença de 1 lncRNAs, quanto regiões que apresentam uma grande quantidade de miRNAs para poucos mRNAs. Além

deste subgrafo maior, outros 9 subgrafos pequenos e independentes, compostos de 3 a 6 elementos, formam a rede de competição endógena.



(a)



(b)

Figura 13: (a) e (b)Rede de ceRNAs da coorte KIRC-TCGA. Em laranja os miRNAs, em verde os lncRNAs e em vermelho os mRNAs.
Fonte: Autoral.

5.4. ANÁLISE DE SOBREVIDA

Com as informações de status vital (*vital_status*) dos pacientes, disponibilizada pelos dados clínicos e os 18 lncRNAs da rede ceRNA, foram construídas e analisadas as curvas de sobrevivência para cada um deles, para avaliar a associação entre a sobrevida alta ou baixa expressão destes RNAs (APÊNDICE F). Como exemplo, foram destacados pela distinção significativa das curvas de sobrevida dois lncRNAs: *Small Nucleolar RNA Host Gene 15* (SNHG15) e o *EPB41L4A Antisense RNA 1* (EPB41L4A-AS1). Como observado, a baixa

expressão de SNHG15 está significativamente associada a uma maior sobrevida dos pacientes (Figura 14), enquanto a baixa expressão de EPB41L4A-AS1 está significativamente associada a uma menor sobrevida (Figura 15), indicando diferentes processos de regulação dos lncRNAS no câncer renal.

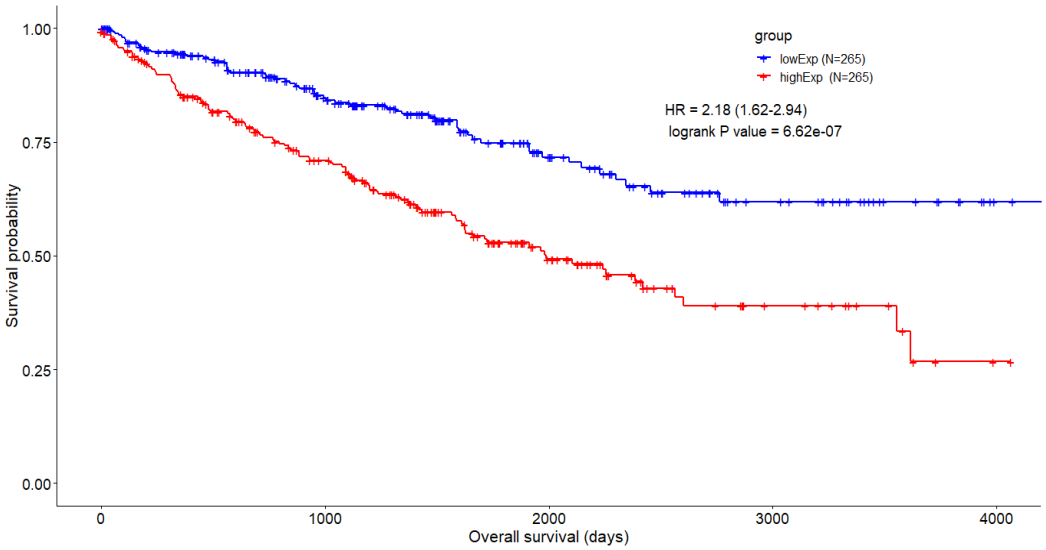


Figura 14: Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do SNHG15, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte: Autoral.

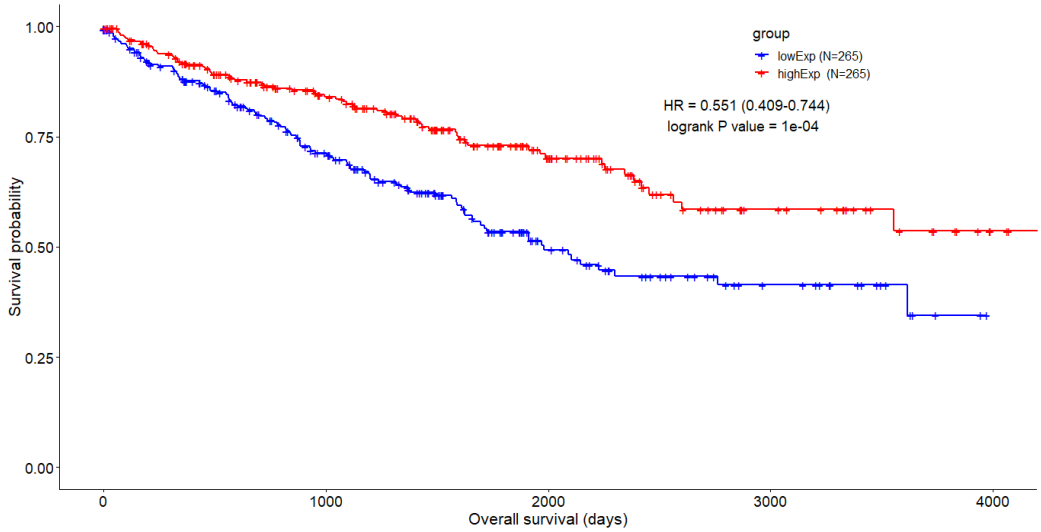


Figura 15: Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do EPB41L4A-AS1, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte: Autoral.

6. DISCUSSÃO

No presente estudo, foi realizada uma análise utilizando técnicas de bioinformática dos dados de sequenciamento completo de exoma do banco de dados do TCGA para identificar as alterações genômicas e transcriptômicas existentes na coorte, associadas ao estado de sobrevida dos pacientes com carcinoma renal de células claras.

Os resultados das alterações genômicas, corroborando com a literatura, indicam uma alta taxa de mutação no gene VHL (Figura 7). Como se sabe, o VHL é um gene um supressor de tumor que codifica um componente do complexo de ubiquitinação, e apresenta um papel chave no sensoriamento de oxigênio celular, por meio da regulação do fator induzível de hipóxia (HIF), sendo um dos principais mecanismos relacionados ao ccRCC (GOSSAGE *et al*.; 2015).

Outros genes identificados como frequentemente mutados foram PBRM1, SETD2 e BAP1, que apresentam importantes funções na biologia da cromatina (Figura 7). Estas alterações podem estar associadas a mutações ou com a perda do braço cromossômico 3p. A perda ou alteração do PBRM1, está relacionada a uma potencialização do HIF1-alpha, a perda do BAP1 ao crescimento celular e uma assinatura transcricional de HIF (JONASCH *et al*.; 2021), e a perdo do SETD2 ao aumento da perda de metilação de DNA nas regiões não promotoras (TCGA; 2013) e a um alto grau de Fuhrman, uma classificação patológica e preditora de prognóstico de ccRCC (BRUGAROLAS, 2013; GOSSAGE *et al*., 2015).

As modificações que ocorrem a nível de genoma podem gerar alterações em vias oncogênicas, e foram identificadas 5 vias alteradas: RTK-RAS, WTN, Hippo, NOTCH, PI3K (Figura 8). Todas estas vias atuam em processos celulares básicos, como proliferação celular, diferenciação, apoptose, e quando desreguladas no câncer podem estar associadas a metástase e sobrevida dos pacientes. Além disso, algumas apresentam funções específicas como as vias PI3K e WNT estão associadas a resposta e resistência terapêutica, a via Hippo com o crescimento de órgãos e a via NOTCH relaciona-se com a homeostase (ZYGULSKA *et al*., 2017; BUGTER *et al*.,2021; REGAD, 2015; VARA et al., 2004; YUAN et al., 2015).

Para avaliar as alterações a nível de transcriptoma, inicialmente foi feita uma análise de expressão diferencial entre o grupo de pacientes com ccRCC e o grupo controle foram identificados 271 lncRNAs, 132 miRNAs e 2.999 mRNAs diferencialmente expressos, sendo que alguns exibindo ampla e significativa variação - miRINAs com log2FC > |2.5| e lncRNAs apresentando log2FC > 5.

Entre os miRNAS subexpressos, podemos citar o hsa-miR-508-3p e o hsa-miR-514a-3p, que quando subexpressos, apresentam relação com câncer de mama, de ovário e germinativo de células testiculares, afetando a definição dos subtipos mesenquimais no ovário (ZHAO et al., 2019) e desregulando o gene PEG3, que regula a via NF-kappa-B no testiculo (ÖZATA et al., 2017). Entre os miRNAs identificados como superexpressos, hsa-miR-155-5p e hsa-miR-122-5p apresentam funções relacionadas ao carcinoma renal. O miR-155-5p regulando o papel do lncTCL6 que quando superexpresso induz apoptose as células do câncer renal (KULKARNI et al., 2021), e o miR-122-5p induzindo um aumento da proliferação celular e invasão das células tumorais promovendo uma sub expressão do FOXO3 e ativando a via de sinalização PI3K/Akt (LIAN et al., 2013; NIE et al., 2018). Em relação aos lncRNAs superexpressos, alguns se mostraram relacionados à carcinogênese (i) AC079760.2; (ii) LINC00887: lncRNA associado a diversos cânceres, como no carcinoma renal, e influenciando negativamente os níveis de sobrevida quando superexpressos no câncer das células não pequenas pulmonares (TIAN et al., 2019); (iii) LINC00462: promove progressão no câncer hepatocelular, por meio da via PI3K/Akt; (iv) LINC02348 e (v) TTC21B-AS1.

A partir dos resultados da análise de expressão diferencial dos mRNAs, foi realizada uma análise de enriquecimento funcional contra vários bancos de dados especialistas. As vias enriquecidas contra o DO confirmam a relação dos genes diferencialmente expressos com carcinoma renal e diversas doenças do sistema geniturinário, de outros sistemas e também relacionadas a doenças infecciosas (Figura 11). Quando feito contra outras bases de dados funcionais (Figura 12) foi observado que existia um enriquecimento de vias de câncer bem discutidas na literatura, como por exemplo a via WP:WP4585 que envolve a regulação das células T imunológicas e se encontra alterada e inibida em processos cancerígenos Sadreddini et al. (2019) propõem que a utilização de imunoterápicos nesta via pode fornecer uma resposta positiva para o tratamento do câncer.

Para analisar como os ncRNAs participam da regulação do ccRCC, foi construída uma rede de ceRNAs mediada pelos lncRNAs, composta por 18 lncRNAs, 75 miRNAs e 128 mRNAs, identificados como diferencialmente expressos, e levando em consideração os testes hipergeométricos, a correlação de Pearson e os padrões de regulação. Além disso, foram analisadas as curvas de sobrevida dos pacientes da coorte, associadas a cada um dos 18 lncRNAs. A partir das curvas de sobrevida, do p-valor e do log-rank associado a elas, os lncRNAs EPB41L4A-AS1 e SNGH15 foram destacados por apresentarem padrões de baixa e alta expressões distintas e possuírem diferentes curvas sobrevida significativas.

O EPB41L4A-AS1 é um lncRNA *anti-sense* localizado na região 5q22.1 do cromossomo. Nossos resultados mostram que, em baixa expressão, o EPB41L4A-AS1 apresenta uma associação negativa e significativa (p-valor = 1e-04) na probabilidade de sobrevida dos pacientes com ccRCC, e resultado que é corroborado por estudos já publicados. Yang e Lv (2022) realizaram testes laboratoriais e confirmaram que para o câncer de mama EPB41L4A-AS1 atua como um gene supressor de tumor e que sua baixa expressão implica no desenvolvimento da doença. Já Liao *et al.* (2019) também confirma que sua baixa expressão, ou deleção, está ligada ao baixo prognóstico em pacientes com câncer e complementa que alterações nele estão relacionadas com o aumento da glicólise aeróbica e a glutaminólise, que ficam mais acelerada devido a presença do HDAC2 nas regiões promotoras do VHL.

De forma contrária, a baixa expressão do RNA longo intergênico SNHG15, por sua vez, apresenta uma associação significativa (p-valor = 6.7e-07) com uma melhor sobrevida destes pacientes. O SNHG15 tem sua localização no cromossomo 7p13, e a família SNHG tem sido associada a diversas funções celulares, como também na oncogênese e na metástase de diversos cânceres. Du *et al*. (2018) demonstram que em estado de super expressão, como em nosso estudo, o SNHG15 pode ser um onco-lncRNA que está relacionado com a proliferação, invasão e migração de células cancerosas do ccRCC, por meio da via de sinalização NF-kappa -B, influenciando negativamente no prognóstico de pacientes com ccRCC. Shuai *et al*. (2019), corroborando também com estes resultados, mostra que em um estado de sub expressão as fases G1/G0 do ciclo celular são afetadas,induzem a apoptose e quando em *knockdown*, ocorre a regulação dos processo de transição epitélio-mesênquima.

Apesar da literatura apresentar descrições testadas e validadas, o estudo sobre os lncRNAs e suas atividades como participante de uma rede de ceRNAs no contexto da progressão dos cânceres e de outras patologias, ainda continua em desenvolvimento. O estudo atual apresenta diversas contribuições, mas ainda sim há limitações. O TCGA fornece uma grande quantidade de dados, possibilitando as análises utilizando técnicas de bioinformática, possibilitando a exploração da ceRNA construída, entretanto existe uma lacuna no estudo dos lncRNAs o que dificulta o entendimento das alterações e os mecanismos moleculares ao qual eles estão relacionados.

7. CONCLUSÃO

Este trabalho teve como objetivos, explorar os dados clínicos do projeto TCGA-KIRC, identificar as principais alterações genômicas na coorte, identificar os RNAs codificantes e não codificantes diferencialmente expressos, avaliar o enriquecimento funcional dos RNAs diferencialmente expressos, construir uma rede de RNAs endógenos competidores a partir dos RNAs codificantes e não codificantes e por fim avaliar a curva de sobrevida a partir da expressão dos RNAs participantes da rede de ceRNAs. Os resultados encontrados mostram que as (i)alterações genômicas e do transcriptoma indicam importantes alterações genéticas e funcionais no ccRCC corroborando com dados da literatura; (ii) a incorporação de informações relacionadas aos ncRNAs possibilita analisar os mecanismos relativos ao ccRCC sob uma nova perspectiva; (iii) a construção e análise de redes de ceRNAs auxilia o entendimento da regulação do RCC a nível transicional e pode sugerir possíveis novos alvos terapêuticos e biomarcadores para o ccRCC; (iv) pôr fim a análise dos 18 lncRNAs diferencialmente expressos participantes da rede, mostrou uma relação significativa da expressão deles com a sobrevida dos pacientes.

Em relação aos trabalhos futuros, pretende-se ampliar o estudo sobre os outros lncRNAs componentes da rede, para entender melhor como os processos biológicos aos quais eles participam se relacionam ao carcinoma renal ou a outros cânceres. Além disso, pretende-se aplicar técnicas de aprendizagem de máquina para realizar a seleção de variáveis (features) para a construção de assinaturas transcricionais de genes não codificantes relacionadas ao prognóstico dos pacientes.

REFERÊNCIAS BIBLIOGRÁFICAS

ALA, U. Competing endogenous RNAs and cancer: How coding and non-coding molecules cross-talk can impinge on disease. **The International Journal of Biochemistry & Cell Biology**, v. 130, p. 105874, jan. 2021.

Brasil, Ministério da Saúde. Banco de dados do Sistema Único de Saúde-DATASUS. Disponível em <http://www.datasus.gov.br>. Acesso em: 18 maio 2022.

BRUGAROLAS, J. PBRM1 and BAP1 as Novel Targets for Renal Cell Carcinoma. **The Cancer Journal**, v. 19, n. 4, p. 324–332, jul. 2013.

CHAN, J.; TAY, Y. Noncoding RNA:RNA Regulatory Networks in Cancer. **International Journal of Molecular Sciences**, v. 19, n. 5, p. 1310, 27 abr. 2018.

CHIU, H.-S. et al. Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. **Cell Reports**, v. 23, n. 1, p. 297- 312.e12, abr. 2018.

COLAPRICO, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. **Nucleic Acids Research**, v. 44, n. 8, p. e71–e71, 5 maio 2016.

DALL'OGLIO, Marcos et al. Câncer de Rim. In: LOPES, Antônio Carlos. Tratado de Clínica Médica. 2. Ed. Rio de Janeiro: **Roca**, 2006. Cap. 305. P. 3242-3248.

Ewen Harrison, Tom Drake and Riinu Ots (2021). Finalfit: Quickly Create Elegant Regression Results Tables and Plots hit Modelling. R package version 1.0.4. Disponível em:<https://CRAN.R-project.org/package=finalfit> Acesso em: 20 de Abril de 2022;

GENTLEMAN, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biology**, v. 5, n. 10, p. R80, 2004.

GDC Data Portal - Data Release v18.0. 2019. Disponível em: <https://docs.gdc.cancer.gov/Data/Release_Notes/Data_Release_Notes/#data-release-180>. Acesso em: 1 jun.2022.

GILLESPIE, Marc et al. The reactome pathway knowledgebase 2022. **Nucleic Acids Research**, [S.L.], v. 50, n. 1, p. 687-692, 12 nov. 2021. Oxford University Press (OUP).

GIURGIU, Madalina et al. CORUM: the comprehensive resource of mammalian protein complexes**.:**2019. **Nucleic Acids Research**,v. 47, n. 1, p. 559-563, 24 out. 2018. Oxford University Press (OUP).

GOLDMAN, M. J. et al. Visualizing and interpreting cancer genomics data via the Xena platform. **Nature Biotechnology**, v. 38, n. 6, p. 675–678, jun. 2020.

GOMES, Clarissa et al. Catalyzing Transcriptomics Research in Cardiovascular Disease: the cardiorna cost action ca17129. Non-Coding Rna. **MDPI AG**. v. 5, n. 2, p. 31, 29 mar. 2019.

GOSSAGE, L.; EISEN, T.; MAHER, E. R. VHL, the story of a tumour suppressor gene. **Nature Reviews Cancer**, v. 15, n. 1, p. 55–64, jan. 2015.

GUIMARÃES, B. D. Associação de Polimorfismos de Nucleotídeo Único (SNPS) dos genes PSCA, TP53 e NQO1 e sua relação com o desenvolvimento de carcinoma mamário em mulheres no Estado da Paraíba. Dissertação. Programa de Pós-Graduação em Biologia Celular e Molecular. Universidade Federal da Paraíba. 2020.

Howlader N. *et al.* SEER Cancer Statistics Review, 1975-2017, **National Cancer Institute**. Bethesda, MD, https://seer.cancer.gov/csr/1975_2017/, based on November 2019 SEER data submission, posted to the SEER web site, April 2020.

HUANG, Hsi-Yuan et al. MiRTarBase 2020: updates to the experimentally validated microrna∶target interaction database. **Nucleic Acids Research**, v. 48, n. 1, p. 148-154, 24 out. 2019. Oxford University Press (OUP).

INCA – Instituto Nacional do Câncer. Câncer de Mama. Disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>. Acesso em 18 de abril de 2022.

JONASCH, E.; WALKER, C. L.; RATHMELL, W. K. Clear cell renal cell carcinoma ontogeny and mechanisms of lethality. **Nature Reviews Nephrology**, v. 17, n. 4, p. 245–261, abr. 2021.

KANEHISA, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27–30, 1 jan. 2000.

KANEHISA, M. Toward understanding the origin and evolution of cellular organisms. **Protein Science**, v. 28, n. 11, p. 1947–1951, nov. 2019.

KANEHISA, M. et al. KEGG: integrating viruses and cellular organisms. **Nucleic Acids Research**, v. 49, n. D1, p. D545–D551, 8 jan. 2021.

KARTHA, R. V.; SUBRAMANIAN, S. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. **Frontiers in Genetics**, v. 5, 2014.

Kazuki Yoshida and Alexander Bartel (2022). Tableone: Create ‘Table 1’ to Describe Baseline Characteristics with or without Propensity Score Weights. R package version 0.13.2. Disponível em:<https://CRAN.R-project.org/package=tableone>. Acesso em: 20 de Abril de 2022;

KLINGE, C. M. Non-coding RNAs: long non-coding RNAs and microRNAs in endocrine-related cancers. **Endocrine-Related Cancer**, v. 25, n. 4, p. R259–R282, abr. 2018.

KOLBERG, Liis et al. Gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g: profiler. **F1000Research**, v. 9, p. 709-736, 17 nov. 2020. F1000 Research Ltd.

KÖHLER, S. et al. The Human Phenotype Ontology in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D1207–D1217, 8 jan. 2021.

LI, J.-H. et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. **Nucleic Acids Research**, v. 42, n. D1, p. D92–D97, jan. 2014.

LI, R. et al. GDCRNATools: an R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. **Bioinformatics**, v. 34, n. 14, p. 2515–2517, 15 jul. 2018.

LIAO, M. et al. LncRNA EPB41L4A-AS1 regulates glycolysis and glutaminolysis by mediating nucleolar translocation of HDAC2. **EBioMedicine**, v. 41, p. 200–213, mar. 2019.

LIU, J. et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. **Cell**, v. 173, n. 2, p. 400- 416.e11, abr. 2018.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 550, dez. 2014.

MARTENS, Marvin et al. WikiPathways: connecting communities. **Nucleic Acids Research**, [S.L.], v. 49, n. 1, p. 613-621, 19 nov. 2020. Oxford University Press (OUP).

MARTINI, Frederic H. et al. O Sistema Urinário. In: MARTINI, Frederic H. et al. Anatomia Humana. 6. Ed. Porto Alegre: Artmed, 2009. Cap. 26. P. 693-713.

MAYAKONDA, A. et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. **Genome Research**, v. 28, n. 11, p. 1747–1756, nov. 2018.

MORRIS, K. V.; MATTICK, J. S. The rise of regulatory RNA. **Nature Reviews Genetics**, v. 15, n. 6, p. 423–437, jun. 2014.

MOTZER, R. J.; BANDER, N. H.; NANUS, D. M. Renal-Cell Carcinoma. **New England Journal of Medicine**, v. 335, n. 12, p. 865–875, 19 set. 1996.

NABI, S. et al. Renal cell carcinoma: a review of biology and pathophysiology. **F1000Research**, v. 7, p. 307, 12 mar. 2018.

NKF - NATIONAL KIDNEY FOUNDATION. RENAL CARCINOMA GUIDELINES. LONDRES – UK, 2017

PADALA, S. A. et al. Epidemiology of Renal Cell Carcinoma. **World Journal of Oncology**, v. 11, n. 3, p. 79–87, 2020.

PRASAD, S. R. et al. Common and Uncommon Histologic Subtypes of Renal Cell Carcinoma: Imaging Spectrum with Pathologic Correlation. **RadioGraphics**, v. 26, n. 6, p. 1795–1806, nov. 2006.

PUJATO, M. et al. Prediction of DNA binding motifs from 3D models of transcription factors; identifying TLX3 regulated genes. **Nucleic Acids Research**, v. 42, n. 22, p. 13500–13512, 16 dez. 2014.

REGAD, T. Targeting RTK Signaling Pathways in Cancer. **Cancers**, v. 7, n. 3, p. 1758–1784, 3 set. 2015.

RITCHIE, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. **Nucleic Acids Research**, v. 43, n. 7, p. e47–e47, 20 abr. 2015.

ROBINSON, M. D.; OSHLACK, A. A scaling normalization method for differential expression analysis of RNA-seq data. **Genome Biology**, v. 11, n. 3, p. R25, 2010.

SADREDDINI, S. et al. Immune checkpoint blockade opens a new way to cancer immunotherapy. **Journal of Cellular Physiology**, v. 234, n. 6, p. 8541–8549, jun. 2019.

SAFRAN, M. et al. The GeneCards Suite. Em: ABUGESSAISA, I.; KASUKAWA, T. (Eds.). **Practical Guide to Life Science Databases**. Singapore: Springer Nature Singapore, 2021. p. 27–56.

SALMENA, L. et al. A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? **Cell**, v. 146, n. 3, p. 353–358, ago. 2011.

SANCHEZ-VEGA, F. et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. **Cell**, v. 173, n. 2, p. 321- 337.e10, abr. 2018.

SCHRIML, L. M. et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. **Nucleic Acids Research**, v. 47, n. D1, p. D955–D962, 8 jan. 2019.

SILVA, T. C. et al. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. **F1000Research**, v. 5, p. 1542, 28 dez. 2016.

STATELLO, L. et al. Gene regulation by long non-coding RNAs and its biological functions. **Nature Reviews Molecular Cell Biology**, v. 22, n. 2, p. 96–118, fev. 2021.

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Mortality - All COD, Aggregated With State, Total U.S. (1969-2019) <Katrina/Rita Population Adjustment>, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021. Underlying mortality data provided by NCHS (www.cdc.gov/nchs)

THE CANCER GENOME ATLAS RESEARCH NETWORK. Comprehensive molecular characterization of clear cell renal cell carcinoma. **Nature**, v. 499, n. 7456, p. 43–49, jul. 2013.

THE GENE ONTOLOGY CONSORTIUM et al. The Gene Ontology resource: enriching a GOld mine. **Nucleic Acids Research**, v. 49, n. D1, p. D325–D334, 8 jan. 2021.

VARA, J. Á. F. et al. PI3K/Akt signalling pathway and cancer. **Cancer Treatment Reviews**, v. 30, n. 2, p. 193–204, abr. 2004.

WARING, E.; QUINN, M.; MCNAMARA, A.; RUBIA, E. A. L.; ZHU, H.; O AND ELLIS, WICKHAM, H. et al. Welcome to the Tidyverse. Journal of Open Source Software, v. 4, n. 43, p. 1686, 21 nov. 2019.

World Health Organization. IARC – International Agency for Research on Cancer. Disponível em: <https://www.iarc.who.int/>. Acesso em 18 de abril de 2022.

YANG, F.; LV, S. LncRNA EPB41L4A-AS1 Regulates Cell Proliferation, Apoptosis and Metastasis in Breast Cancer. **Annals of Clinical and Laboratory Science**, v. 52, n. 1, p. 3–11, jan. 2022.

YUAN, X. et al. Notch signaling: An emerging therapeutic target for cancer treatment. **Cancer Letters**, v. 369, n. 1, p. 20–27, dez. 2015.

ZHANG, Z. et al. Uniform genomic data analysis in the NCI Genomic Data Commons. **Nature Communications**, v. 12, n. 1, p. 1226, dez. 2021.

ZYGULSKA, A. L.; KRZEMIENIECKI, K.; PIERZCHALSKI, P. Hippo pathway - brief overview of its relevance in cancer. **Journal of Physiology and Pharmacology: An Official Journal of the Polish Physiological Society**, v. 68, n. 3, p. 311–335, jun. 2017.

APÊNDICES

APÊNDICE A

Tabela 5: Informações associadas aos dados clínicos – TCGA-KIRC

Variável	Tipo	Descrição
patient_id	chr*	Identificação de cada paciente
synchronous_malignancy	fct**	Presença de outra câncer durante participação do projeto
ajcc_pathologic_stage	fct	Estadiamento da Doença
days_to_last_follow_up	num**	Dias até o último acompanhamento
prior_malignancy	fct	Diagnóstico anterior de câncer
year_of_diagnosis	num	Ano de Diagnóstico
prior_treatment	fct	Tratamento prévio
ajcc_pathologic_t	fct	Classificação do tamanho tumoral
ajcc_pathologic_n	fct	Grau de espalhamento tumoral
ajcc_pathologic_m	fct	Classificação de distância metastática
race	fct	Raça
gender	fct	Gênero
ethnicity	fct	Etnia
vital_status	fct	Status de sobrevivência
age	num	Idade
year_of_birth	num	Ano de nascimento
pharmaceutical_treatment_or_therapy	fct	Utilização de terapia ou tratamento farmacêutico
radiation_treatment_or_therapy	fct	Utilização de terapia ou tratamento radioterápico

*Caracter
** Fator
**Numérico

Fonte: Autoral.

APÊNDICE B

Tabela 6: lncRNAs da ceRNAs

lncRNA	LogFC	p-valor	FDR
PVT1	4,6699E+14	0	0
AC016876.2	2,00427E+13	1,90139E-67	5,94132E-66
SNHG1	1,62174E+14	8,20628E-45	1,20199E-43
SNHG15	1,77383E+14	1,47247E-45	2,12462E-43
GAS5	1,71819E+14	1,2035E-40	1,53394E-39
MAGI2-AS3	-1,63937E+14	1,20123E-37	1,39631E-36
AC021078.1	1,76055E+14	9,29266E-37	1,0524E-35
EPB41L4A-AS1	1,40784E+14	1,10275E-29	9,65241E-30
SNHG20	1,17689E+14	7,39169E-29	6,08506E-27
AC005154.1	1,63469E+14	3,92023E-27	3,14701E-26
C1RL-AS1	1,52538E+13	6,84728E-21	4,45549E-20
MIR497HG	1,46701E+14	6,0323E-18	3,4562E-17
AC015813.1	1,65621E+14	2,06128E-17	1,15952E-16
SNHG5	1,12029E+13	3,91405E-13	1,86097E-12
AF117829.1	1,11369E+14	1,13561E-09	4,62585E-09
NEAT1	1,5715E+13	9,78378E-07	3,53452E-06
MALAT1	1,24428E+14	0,01029378	0,030934564
H19	1,22691E+14	38887736,13	70145300,72

Fonte: Autoral

APÊNDICE C

Tabela 7: mRNAs diferencialmente expressos participantes da ceRNA.

mRNAs	Log2FC	p-valor	FDR
SCARB1	4,08183E+14	3,0338E-200	4,0074E-197
ATP1A1	-2,78382E+14	6,3981E-176	4,4269E-173
AHNAK2	4,08349E+14	1,5221E-118	2,3038E-116
PTHLH	6,42007E+14	2,9298E-104	3,1302E-102
DGKD	2,1064E+14	2,9092E-103	2,956E-101
IL2RB	3,01186E+14	1,4902E-101	1,4532E-99
RELT	2,09074E+14	1,51826E-93	1,19893E-91
PFKFB4	2,42103E+14	3,00341E-92	2,26112E-90
RNF149	1,31407E+14	5,89672E-87	3,74146E-85
TNFRSF10B	1,50585E+14	1,06938E-85	6,42069E-84
FAM111B	2,52046E+14	3,57607E-85	2,07013E-83
EZH2	1,85936E+14	1,3694E-78	6,19855E-77
LOX	4,05092E+14	4,00939E-76	1,69844E-74
APLN	2,46288E+14	7,38053E-65	2,13624E-63
MXD3	2,58312E+14	1,72045E-63	4,7525E-62
SPARC	1,82467E+13	3,43032E-63	9,35132E-62
KIF21B	2,12804E+14	3,59783E-62	9,48757E-61
MYO9B	1,16665E+14	6,18216E-61	1,5704E-59
INTS6L	1,75746E+14	1,70855E-57	3,82515E-56
LOXL2	2,69089E+14	1,62265E-54	3,32072E-53
CSNK1E	1,11511E+14	6,86921E-53	1,32373E-51
SH2D2A	2,75421E+14	7,12822E-53	1,37183E-51
CD276	1,21264E+14	7,25124E-53	1,39366E-51
CHFR	1,12043E+14	9,21742E-53	1,76921E-51
ZNF469	2,07153E+14	3,25696E-52	5,89335E-51
FHOD1	1,26884E+14	7,72672E-51	1,31003E-48
GFI1	2,48888E+13	4,07458E-48	6,60018E-47
PTTG1	2,39149E+14	5,23039E-47	8,23376E-46
SLAMF6	2,5291E+14	4,31092E-46	6,38508E-44

SGPP1	-1,27863E+14	1,3849E-42	1,88414E-41
CORO1C	1,06197E+14	1,04086E-41	1,37755E-40
UNC5B	1,75291E+14	4,84431E-41	6,26784E-40
GATA3	-3,08467E+14	9,9163E-41	1,27508E-39
MARCH3	1,44201E+14	7,1554E-40	8,89375E-39
NMB	2,58873E+14	1,00095E-39	1,242E-38
P3H1	1,33451E+14	2,48442E-39	3,04373E-38
COL5A2	2,06604E+14	3,63179E-40	4,4233E-38
TNK2	1,39562E+14	4,30774E-39	5,23778E-38
TET3	1,01269E+14	5,95359E-39	7,19681E-38
PROCR	1,60024E+14	6,50163E-40	7,85275E-38
RASD1	-2,53109E+14	4,08934E-38	4,8229E-37
BRIP1	1,80406E+14	9,27197E-36	1,00689E-34
SLC41A2	1,20905E+14	2,59455E-35	2,78631E-34
ANLN	2,31616E+14	7,95164E-34	8,09652E-34
NHSL1	-1,25008E+14	1,17131E-33	1,1827E-32
PHLDA3	1,66751E+14	5,15567E-33	5,08567E-32
PLXNA3	1,35504E+14	1,03865E-31	9,76806E-31
POLE2	1,33917E+14	1,19038E-31	1,11661E-30
TBX3	-1,84524E+14	1,25806E-31	1,17933E-30
CNTNAP1	1,6477E+13	6,32727E-30	5,59557E-29
COL1A1	2,82095E+14	1,76233E-29	1,53425E-28
PXDN	1,68253E+14	2,50284E-28	2,10089E-27
CREB5	1,48735E+14	6,35217E-28	5,25011E-27
FGFR1OP	1,35648E+14	2,09615E-27	1,70151E-26
INSR	1,25311E+14	2,8576E-27	2,31185E-26
NFKBIE	1,21977E+14	2,884E-27	2,33062E-28
CCNL2	1,8094E+14	6,77997E-28	5,4128E-26
SIPA1L2	1,13631E+14	3,83014E-26	2,98882E-26
PMEPA1	1,6071E+14	7,93845E-26	6,14849E-25
APBB3	1,48491E+14	1,30284E-25	1,0016E-24
JMY	-1,21216E+14	1,64824E-25	1,2658E-24

ATAD5	1,24491E+14	3,1497E-24	2,31487E-24
HUNK	-1,73276E+14	4,10831E-25	3,00119E-23
AATK	1,38529E+14	1,50713E-23	1,0813E-22
PPM1H	-1,91676E+14	2,07144E-23	1,47179E-22
SPRY4	1,27172E+14	2,04341E-22	1,41452E-21
RFLNB	1,44534E+13	6,18085E-23	4,19466E-21
TRIM52	1,08986E+14	2,75915E-21	1,83062E-20
SNAP25	2,01169E+14	4,06116E-20	2,54787E-19
KMT5C	1,2872E+14	4,55998E-20	2,85342E-19
ANKRD13B	1,46476E+14	4,69832E-20	2,93494E-19
ARL10	1,30789E+14	1,02642E-19	6,30339E-19
CLK2	1,05415E+14	1,60273E-19	9,76424E-19
CORO2B	-2,01168E+14	4,46456E-19	2,67285E-18
VKORC1	1,17703E+14	5,1376E-19	3,06946E-18
DNA2	1,4762E+14	1,08829E-18	6,43847E-18
FSTL3	1,76809E+14	1,85823E-18	1,08696E-17
CA2	-1,50101E+14	2,371E-18	1,38189E-17
PRKCA	-1,22496E+14	2,82048E-18	1,6373E-18
LY6E	1,69497E+14	2,91135E-18	1,68937E-17
IL15	1,00849E+13	5,14695E-19	2,9571E-17
RPL28	1,12731E+14	8,2691E-18	4,72102E-18
SH3RF3	1,15905E+13	3,68335E-17	2,04976E-16
SCOC	-1,00719E+14	5,15114E-17	2,85454E-16
SPRY1	1,33455E+14	1,40296E-17	7,63769E-16
IRX3	1,51018E+14	1,51443E-16	8,22912E-16
CLK1	1,15724E+14	6,88707E-17	3,66017E-15
CCND2	1,19241E+14	8,72806E-16	4,62167E-16
DNAJC6	-1,4626E+14	3,01368E-15	1,55776E-14
PSTPIP2	1,09616E+14	7,00955E-15	3,57741E-14
PFKFB3	-1,25291E+14	1,14003E-14	5,73964E-14
PLAGL1	1,2306E+14	4,06963E-14	2,01197E-13
TCF4	1,17186E+14	1,10571E-13	5,38402E-13

BCL11A	1,54115E+14	2,9349E-14	1,40462E-12
RCN3	1,52461E+14	4,16665E-13	1,97719E-13
ADGRG2	2,54108E+14	5,69376E-13	2,68779E-13
L1CAM	-3,19645E+14	7,16055E-13	3,3573E-12
AMT	-1,10507E+14	4,12197E-12	1,88045E-11
HECW2	1,39887E+14	4,15901E-12	1,89556E-11
TSPYL2	1,17383E+14	4,88867E-12	2,22393E-11
FGFR2	-1,27885E+14	6,55928E-12	2,9635E-11
IRF4	1,98921E+14	1,89628E-11	8,40285E-11
TLL1	1,51031E+14	5,31236E-11	2,29865E-10
WNK3	-1,72098E+14	6,31741E-11	2,72461E-10
SLC25A37	1,01696E+14	1,64131E-10	6,9266E-10
MYO10	-1,14001E+14	7,4271E-10	3,04675E-09
GXYLT2	1,95459E+14	1,29933E-09	5,26619E-09
KCNN4	1,58516E+14	6,52147E-09	2,56933E-08
SLC25A5	-1,15851E+14	8,42436E-09	3,30381E-09
CBFA2T3	1,04998E+14	9,24838E-10	3,6211E-08
HMMR	1,37698E+14	7,0274E-08	2,66184E-07
CELSR3	1,49734E+14	8,42325E-08	3,18226E-07
KIFC2	1,22996E+14	1,38336E-06	4,96057E-06
RNF165	1,38501E+14	4,50895E-06	1,57678E-05
WSB1	1,00528E+13	2,42344E-06	8,22341E-05
LPL	-1,23124E+13	9,26748E-05	0,000306595
WNT5A	-1,27922E+14	0,04016336	0,011737201
CXCL2	1,65503E+14	0,059259536	0,171487961
RSRP1	1,03267E+14	0,168736737	0,478295902
CITED4	1,05182E+14	0,606452994	1,674285008
LDLR	-1,27286E+14	2,868511382	7,685685115
NID2	1,00438E+14	0,998011306	26,00162144
PCDH7	-1,36979E+14	69,49792917	173,2975649
MN1	1,00207E+14	24,82989182	601,8991126
SCARA3	-1,19418E+14	5056,049575	11351,11253

UST	-1,27048E+14	5552,048186	1242,242996
BTBD11	-1,43442E+14	5940,874848	13271,97287
SLC25A27	1,05083E+14	15746,6542	34318,11692

Fonte: Autoral.

APÊNDICE D

Tabela 8: 18 miRNAs diferencialmente participantes da rede de ceRNA.

miRNA	Log2FC	p-valor	FDR
hsa-miR-21-5p	2,16097E+14	1,02741E-88	8,58918E-87
hsa-miR-16-5p	1,00895E+14	5,56881E-41	1,05807E-39
hsa-miR-429	-1,91674E+14	7,84777E-36	1,26168E-35
hsa-miR-181b-5p	1,1947E+14	2,8082E-27	3,00982E-27
hsa-miR-93-5p	1,0119E+14	4,98106E-26	4,95734E-26
hsa-let-7g-5p	-1,02642E+14	2,01385E-25	1,95765E-24
hsa-miR-200b-3p	-1,62143E+14	9,71807E-25	9,02701E-24
hsa-let-7e-5p	-1,25654E+14	5,68936E-23	4,75631E-22
hsa-miR-200a-3p	-1,36956E+14	2,34997E-18	1,51121E-17
hsa-miR-590-3p	1,14282E+14	3,08842E-17	1,9268E-16
hsa-miR-200c-3p	-3,152E+14	2,30498E-12	1,16082E-11
hsa-miR-1-3p	-1,74508E+14	3,74086E-10	1,77691E-09
hsa-miR-1271-5p	1,26318E+14	1,80996E-07	7,79961E-07
hsa-miR-20b-5p	-1,23432E+14	2,3269E-05	8,92333E-05
hsa-miR-204-5p	-1,66042E+14	0,000159463	0,000584699
hsa-miR-130b-3p	1,02041E+14	0,005712187	0,018949955
hsa-miR-141-3p	-2,61534E+14	1,318655439	3,881675869
hsa-miR-133a-3p	-1,11811E+14	4330,990048	10525,31302

APÊNDICE E

Tabela 9: Bancos de dados de vias funcionais

Banco de dados	Característica	Site	Referência
Enciclopédia de Genes e Genomas de Kyoto (KEGG)	vias envolvidas no carcinoma renal	www.kegg.jp/pathway/hsa05211	(KANEHISA, 2000, 2019; KANEHISA et al., 2021)
Disease Ontology (DO)	descrições consistentes de termos de doenças humanas, características fenotípicas e características das doenças relacionadas ao vocabulário médico	disease-ontology.org	(SCHRIML *et al.*, 2018)
Gene Ontology (GO)	identifica os processos biológicos, os componentes celulares e as funções moleculares envolvidas	geneontology.org	THE GENE ONTOLOGY CONSORTIUM et al., 2021
Reactome (REAC)	base de dados de moléculas metabólicas e sinalizadoras e suas relações, organizadas em processos e vias biológicas	reactome.org	GILLESPIE *et al.*; 2022
WikiPathways (WP)	plataforma colaborativa para a curadoria de vias biológicas	wikipathways.org	MARTENS *et al.*; 2020
Atlas de Proteína Humana (HPA)	mapeamento de todas as proteínas humanas nas células, tecidos e órgãos	www.proteinatlas.org	UHLÉN *et al.*; 2015
Comprehensive Resource of Mammalian Protein Complexe (CORUM)	coleção de complexos proteicos de mamíferos validados experimentalmente	mips.helmholtz-muenchen.de/corum	GIURGIU *et al.*; 2019
Transcription Factor database (TF)	informações abrangentes sobre os *motifs* de ligação do fator de transcrição e seus genes regulados para cinco organismos modelo e humanos	www.fiserlab.org/tf2dna_db/index.html	PUJATO et al., 2014
mirTARbase (MIRNA)	banco de dados composto pelas interações miRNA-alvo	mirtarbase.cuhk.edu.cn	HUANG *et al.*; 2020
Fenótipo Humano (HP)	banco de dados com vocabulário de anormalidades fenotípicas encontradas em doenças humanas.	hpo.jax.org	KÖHLER et al., 2021

APÊNDICE F

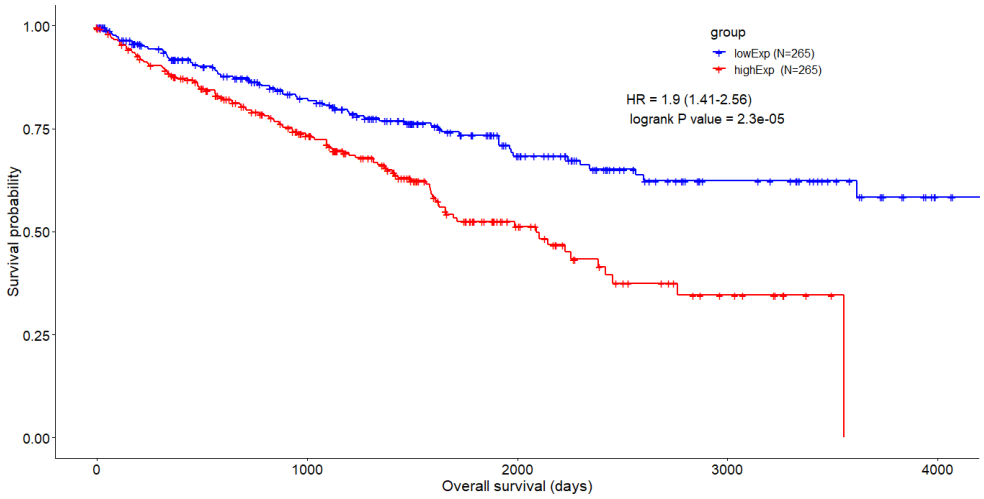


Figura 16:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AC005154.1, em relação ao tempo. p-valor determinado pelo teste de log-rank

Fonte Autoral

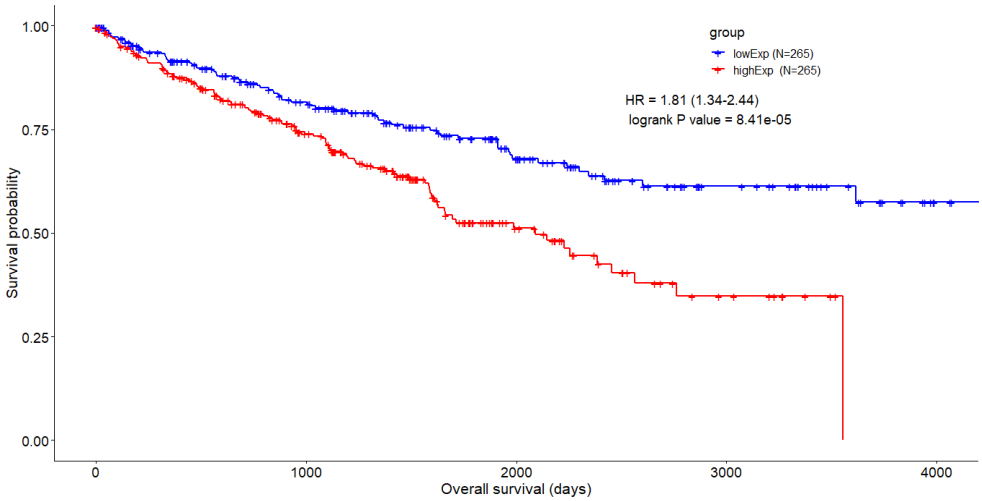


Figura 17:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AC015813.1, em relação ao tempo. p-valor determinado pelo teste de log-rank.

Fonte Autoral

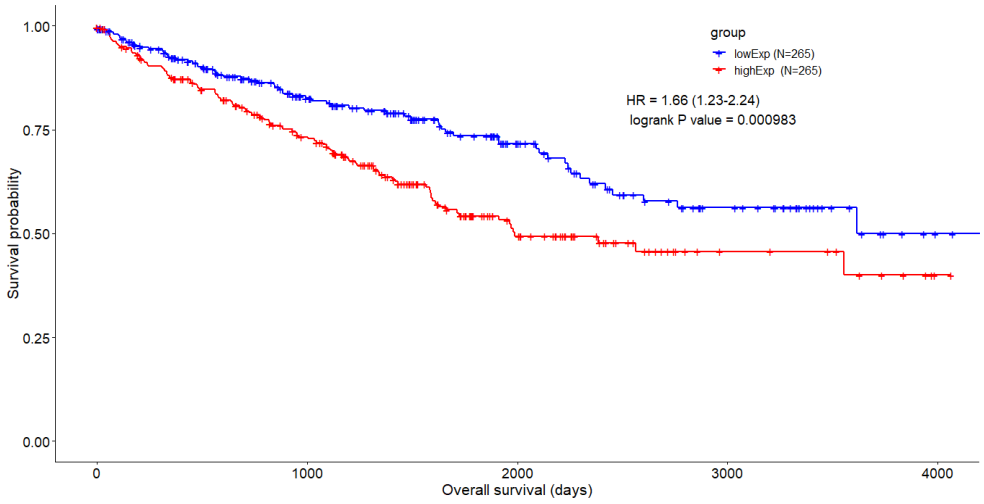


Figura 18:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AC016876.2, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

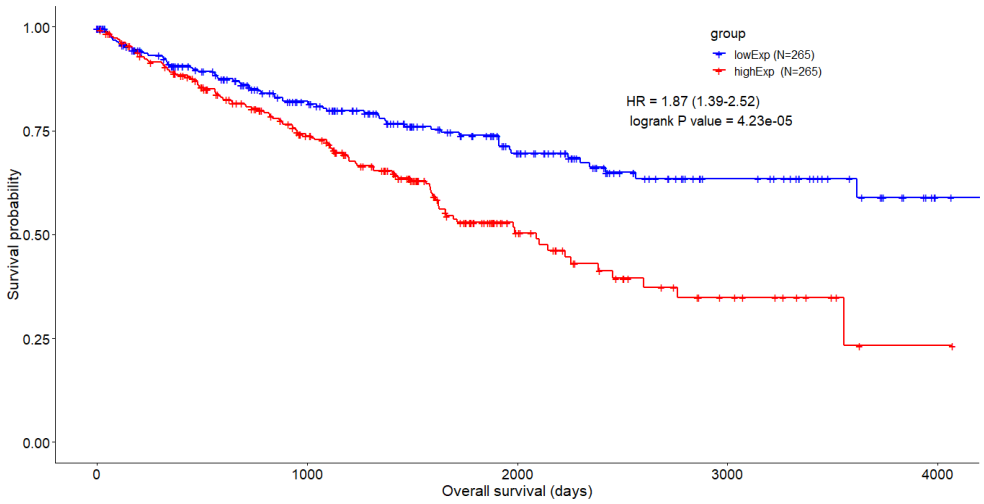


Figura 19:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AC021078.1, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

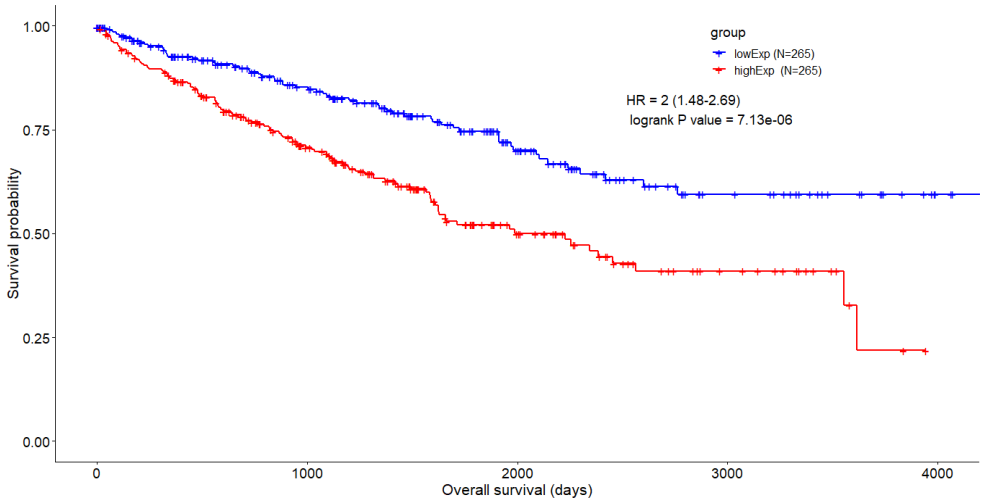


Figura 20:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do AF117829.1, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

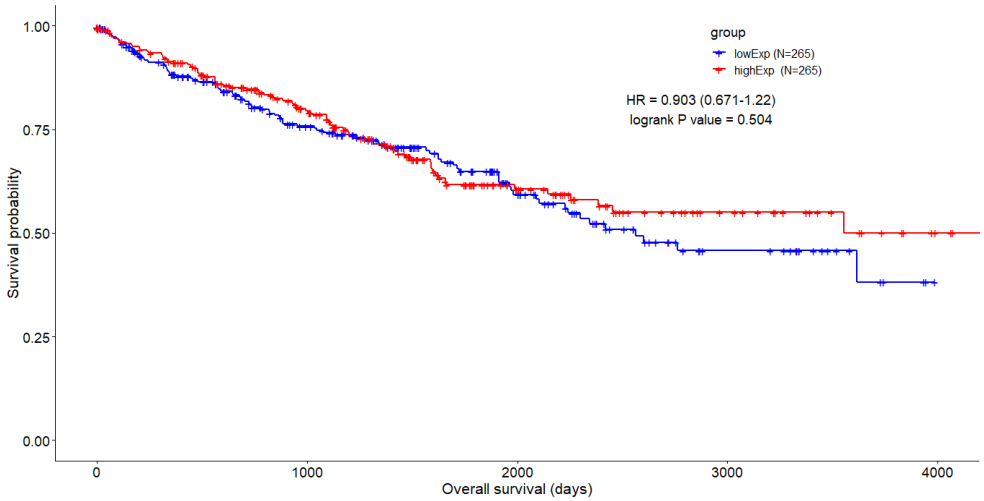


Figura 21:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do C1RL-AS1, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

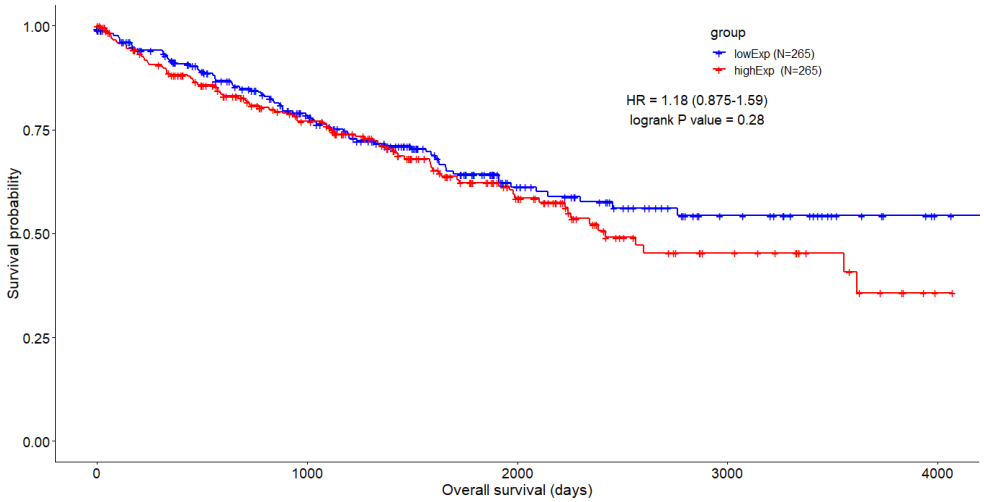


Figura 22:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do GAS5, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

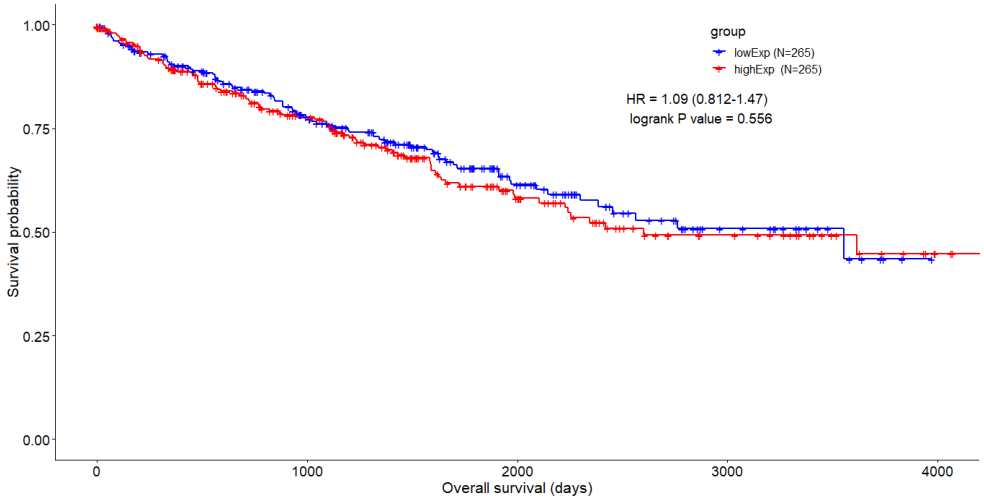


Figura 23:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do H19, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

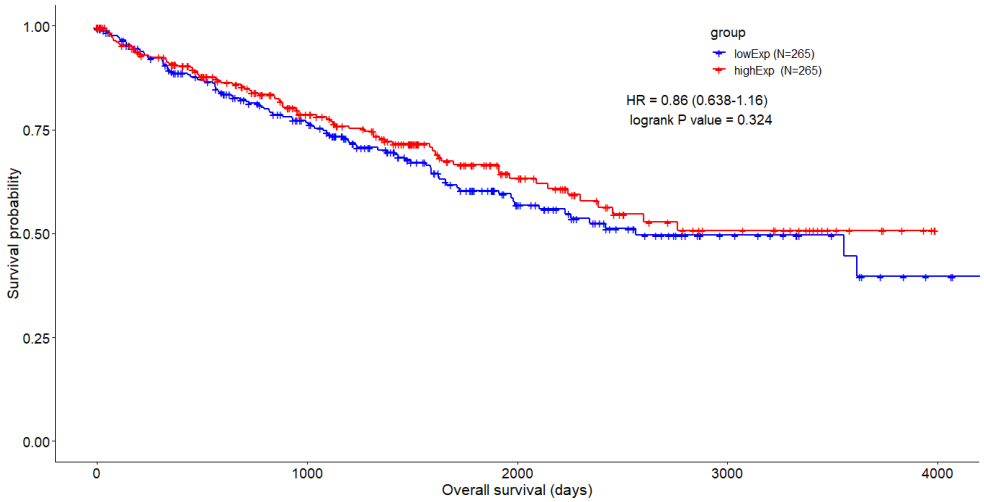


Figura 24:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do MAGI2-AS3, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

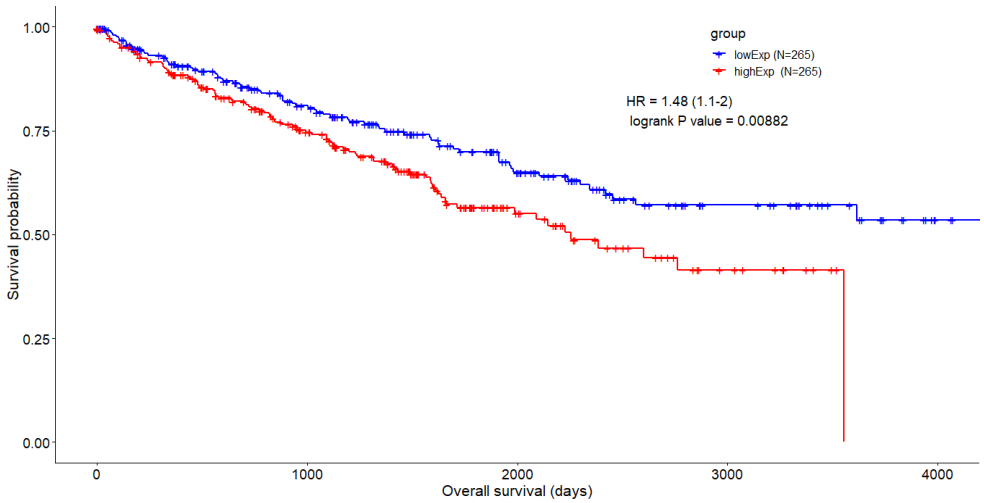


Figura 25:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do MALAT1, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

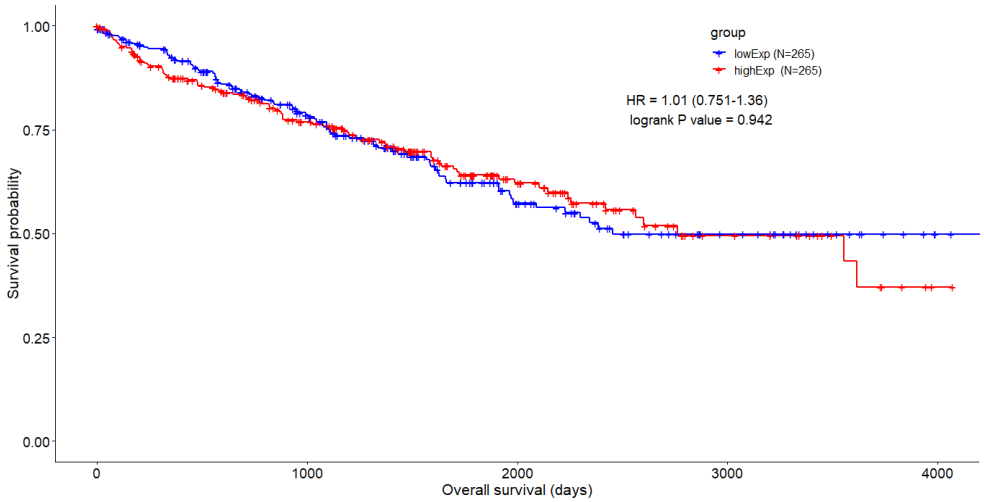


Figura 26:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do MIR497HG, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

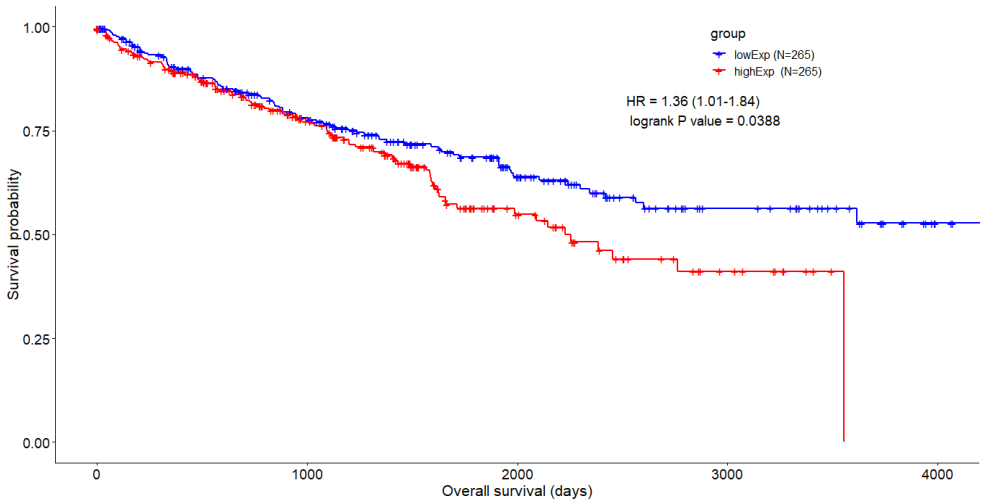


Figura 27:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do NEAT1, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

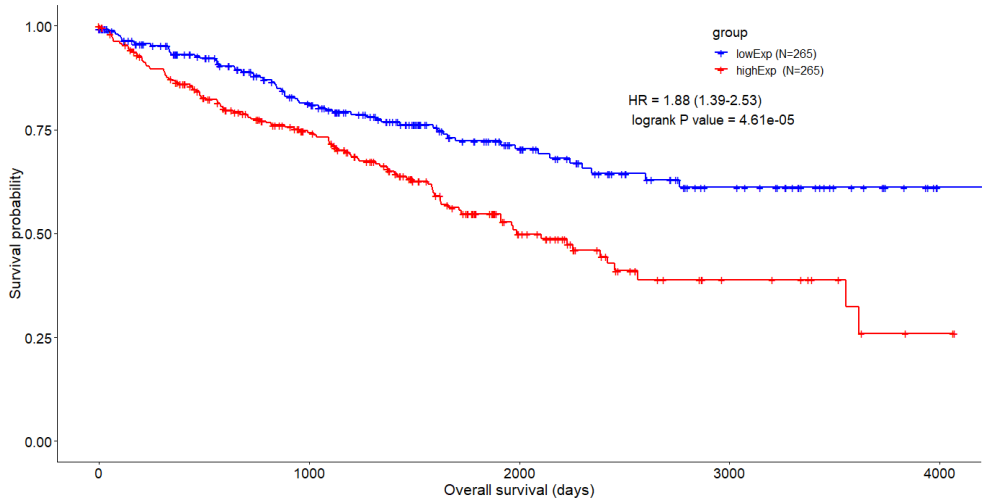


Figura 28:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do PVT1, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

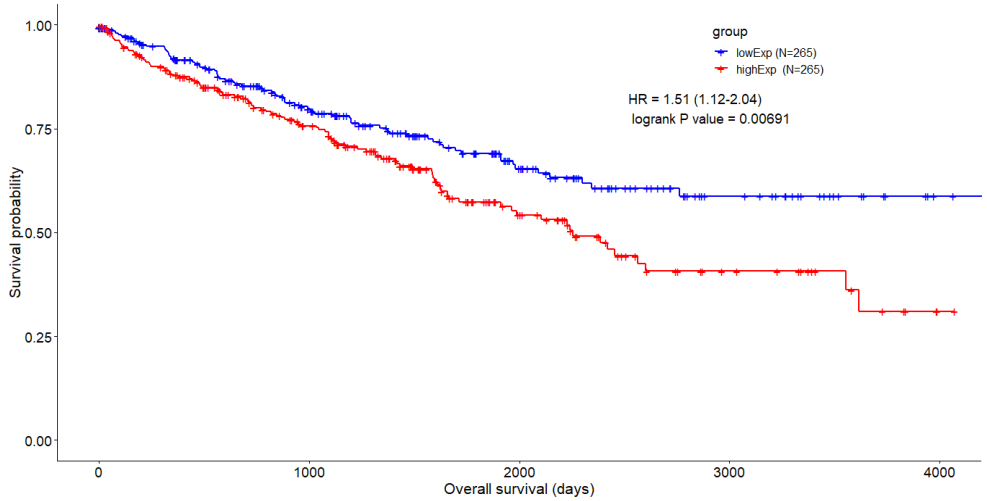


Figura 29:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do SNHG1, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

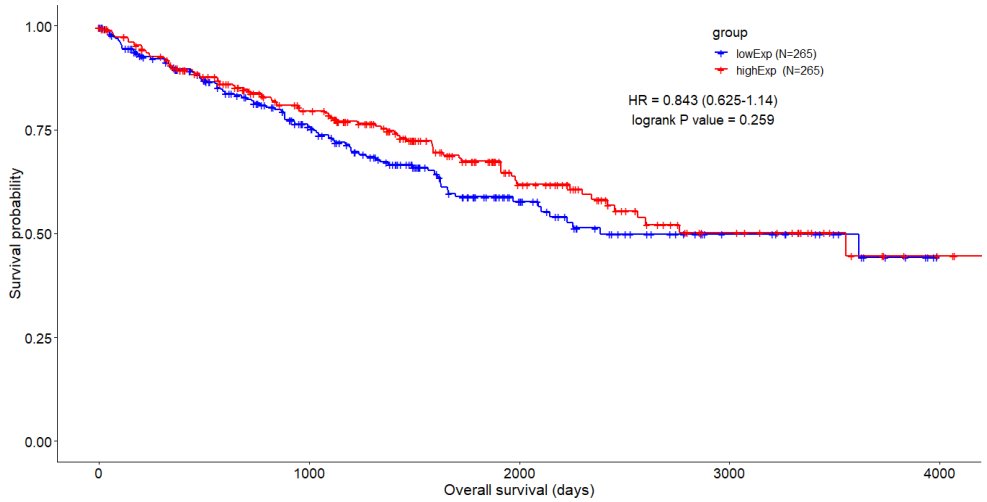


Figura 30:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do SNHG5, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral

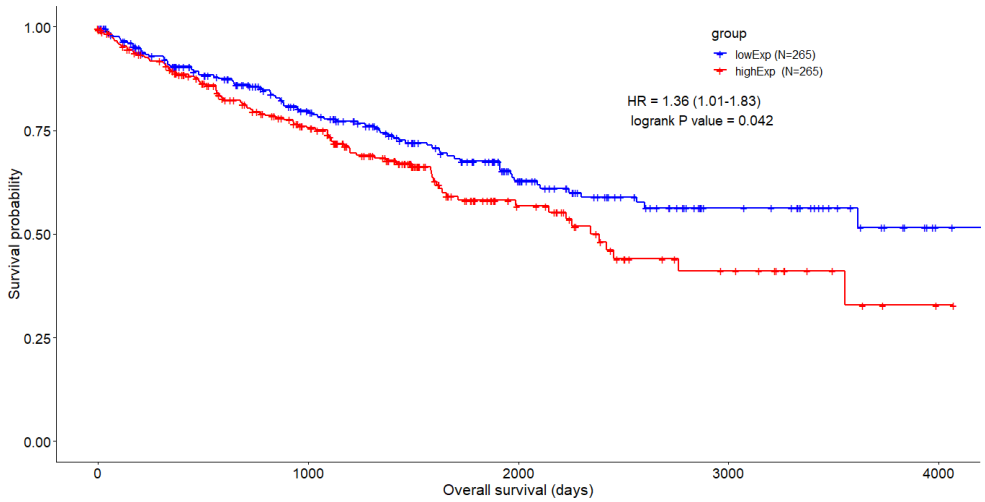


Figura 31:Curva de Kaplan-Meier. Probabilidade de sobrevida de pacientes de ccRCC com alta (curva em vermelho) e baixa (azul) expressão do SNHG20, em relação ao tempo. p-valor determinado pelo teste de log-rank.
Fonte Autoral.