UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

Trabalho de Conclusão de Curso

Glioblastoma Multiforme: análise de dados ômicos para a caracterização do processo tumoral

João Gabriel Fragoso Paulo

Natal/RN
2024

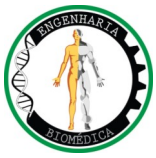
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

Glioblastoma Multiforme: análise de dados ômicos para a caracterização do processo tumoral

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

Graduando: João Gabriel Fragoso Paulo
Orientador: Professora Doutora Beatriz Stransky Ferreira
Co-orientador: Professor Doutor Patrick Cesar Alves Terrematte

Natal/RN
2024



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

Glioblastoma Multiforme: análise de dados ômicos para a caracterização do processo tumoral

Banca Examinadora do Trabalho de Conclusão de Curso:

Profª. Drª. Beatriz Stransky Ferreira _____

UFRN - Orientadora

Prof. Dr. Jorge Estefano de Santana Souza _____

UFRN – Avaliador Interno

Prof. Dr. Leonardo Capistrano Ferreira _____

UFRN – Avaliador Interno

Natal/RN

2024

Dedico este trabalho à estimada professora-orientadora Beatriz Stransky que muito me ensinou ao longo do curso, e abriu novos horizontes para minha formação como pesquisador e engenheiro biomédico.

AGRADECIMENTOS

Primeiramente a'*O D'us único* e em seguida, aos irmãos espirituais por terem me inspirado durante o desempenho deste trabalho, bem como, pelo encorajamento interior para prosseguir pujante na jornada de 6 anos de Engenharia Biomédica.

Ao incentivo e suporte de todas as naturezas por parte de meus familiares e meus pais, em especial à minha mãe. À professora-orientadora Beatriz Stransky, a quem dedico este trabalho, por ter me instruído com tenacidade e munida dos melhores conhecimentos e técnicas. Ao co-orientador Patrick Terrematte por ter auxiliado nos *scripts*. Aos professores e avaliadores Jorge Estefano e Leonardo Capistrano por terem aceito o convite de avaliar este trabalho. Ao corpo docente do Departamento de Engenharia Biomédica pelo zelo e esmero com que mantêm o curso, bem como pelos ensinamentos únicos e valiosos. À toda UFRN por ter ofertado a estrutura e recursos de qualidade no decurso da graduação. Aos colegas que tornaram esta jornada fluida e amigável. Aos amigos Nayane, Paulo, Antônio e à minha namorada Lavínnya que trouxeram alegria e descontração, de maneira especial, e que foram fundamentais para manter o equilíbrio psíquico.

SUMÁRIO

1. INTRODUÇÃO
2. OBJETIVOS
- **2.1. Geral**
- **2.2. Específicos**

3. REVISÃO BIBLIOGRÁFICA
- **3.1. Câncer: Conceitos Gerais**
- **3.2. O programa TCGA - The Cancer Genome Atlas**
- **3.3. Natureza dos dados ômicos**

4. METODOLOGIA
- **4.1 - Desenvolvimento**
- **4.2 - Obtenção de dados**
- **4.3 - Análise dos dados clínicos**
- **4.4 - Análise de sobrevivência e taxa de risco**
- **4.5 - Análise Genômica**
- **4.6 - Análise de Metilação**
- **4.7 - Análise de Expressão Diferencial**
- **4.8 - Análise de Enriquecimento Funcional**

5. RESULTADOS
- **5.1. Descrição da Coorte**
- **5.2. Análise de sobrevivência**
- **5.3. Análise da taxa de risco de sobrevivência**
- **5.4. Análise Genômica**
- **5.5. Análise de Metilação**
- **5.6. Análise de Expressão Diferencial**
- **5.7. Análise de Enriquecimento Funcional**

6. DISCUSSÃO
- **6.1. Coorte e dados clínicos**
- **6.2. Taxa de risco**
- **6.3. Análise Genômica**
- **6.4. Análise de Metilação**
- **6.5. Análise de Expressão Diferencial**
- **6.6. Análise de Enriquecimento Funcional**

7. CONCLUSÕES
REFERÊNCIAS

LISTA DE FIGURAS

Figura 1 - Conjunto de células gliais

Figura 2 - Classificação de cânceres gliais.

Figura 3 - Glioblastoma Multiforme visto no cérebro em corte frontal.

Figura 4 – Variantes de número de cópias em genes.

Figura 5 - Mecanismo geral da expressão gênica.

Figura 6 - Níveis de compactação do DNA ao longo da cromatina.

Figura 7 - Características moleculares mais informativas para a predição dos resultados dos pacientes nos projetos TCGA.

Figura 8 - Estrutura do objeto 'RangedSummarizedExperiment'.

Figura 9 - Curva de sobrevivência da coorte TCGA-GBM sem estratificação.

Figura 10 - Estimativa de Kaplan-Meier para sobrevivência dos pacientes, estratificados por raça.

Figura 11 - Razão de risco para pacientes da coorte TCGA-GBM.

Figura 12 - Panorama geral de alterações gênicas na coorte TCGA-GBM .

Figura 13 - Oncoplot dos 10 genes mais frequentemente mutados.

Figura 14 - Curva de Kaplan-Meyer para a sobrevivência dos pacientes, em relação ao status do gene TP53.

Figura 15 - Mutações mutuamente exclusivas ou co-ocorrentes.

Figura 16 - Vias oncogênicas alteradas e proporção de pacientes afetados.

Figura 17 - Genes alterados na via metabólica PI3K.

Figura 18 - Amplificações nos cromossomos em função do q-valor.

Figura 19 - Deleções nos cromossomos dos pacientes em função do q-valor.

Figura 20 - Volcano plot das regiões metiladas.

Figura 21 - Gráfico MA segundo o status vital.

Figura 22 - Níveis de expressão dos genes EPYC e SFRP5, respectivamente.

Figura 23 - Gráfico MA segundo o status do gene IDH.

Figura 24 - Níveis de expressão dos genes C6ORF15 e SLC25A15P2, respectivamente.

Figura 25 - Análise de enriquecimento funcional (processos biológicos) dos genes diferencialmente expressos, utilizando o banco de dados KEGG.

Figura 26 - Processos biológicos segundo NES e com dados do KEGG.

Figura 27 - Vias metabólicas no Glioma (KEGG).

LISTA DE TABELAS

Tabela 1 – Distribuição das variáveis clínicas de acordo com o status vital.

Tabela 2 – Descrição dos dez genes mais mutados no glioblastoma multiforme.

Tabela 3 – Genes nas citobandas mais amplificadas.

Tabela 4 – Genes nas citobandas mais depletadas.

LISTA DE ABREVIAÇÕES, SIGLAS E SÍMBOLOS

BHE - barreira hematoencefálica

CID - Código Internacional de Doença

CNV - *Copy Number Variation*

D2-HG - D-2-hidroxiglutarato

DEA - *Differentiation Expression Analysis*

GBM - Glioblastoma multiforme

GDC - *Genomic Data Commons*

GISTIC - *Genomic Identification of Significant Targets In Cancer*

INCA - Instituto Nacional do Câncer

MEC - Matriz extracelular

KEGG - *Kyoto Enciclopedia of Genes and Genomes*

MAF - *Mutation Annotation Format*

NCI – Nacional Institute of Cancer

NES - *Normalized Enrichment Score*

SNC - Sistema Nervoso Central

SEER - *Surveillance, Epidemology, and End Results Program*

SNP - *Single Nucleotide Polymorphism*

SNV - *Single Nucleotide Variation*

TCGA - *The Cancer Genome Atlas Program*

WT - *Wildtype*

PAULO, João Gabriel Fragoso. **Glioblastoma Multiforme: análise de dados ômicos para a caracterização do processo tumoral.** Trabalho de Conclusão de Curso, Graduação em Engenharia Biomédica, Universidade Federal do Rio Grande do Norte, 62p., 2024.

RESUMO

O glioblastoma multiforme é um câncer do sistema nervoso central incidente em astrócitos. Apesar de sua baixa incidência na população mundial (297.000 casos em 2018), é uma afecção de difícil tratamento em razão do seu alto grau de estadiamento (grau IV). Além disso, como todos os cânceres, as causas da manifestação são diversas e de relação não-linear. Para tanto, este trabalho propôs-se a investigar os fatores de risco e os diversos mecanismos moleculares que contribuem para sua gênese e manutenção. Tal tarefa foi realizada por meio de ferramentas de bioinformática e teve como fonte de dados a coorte de 600 pacientes do projeto TCGA (*The Cancer Genome Atlas*). A análise baseou-se em curvas de sobrevivência, razão de risco, análise genômica, de metilação, de expressão diferencial de genes, e do enriquecimento funcional das vias, utilizando a linguagem R e o ambiente de desenvolvimento R Studio. Foram identificados como fatores de risco a raça e o sexo; a metilação como principal fator genômico para o desenvolvimento da patologia; o gene *IDH* como biomarcador de prognóstico favorável; 359 dias como tempo médio de sobrevivência da coorte; os genes frequentemente alterados, como o *PTEN, TP53* e *EGFR*; o nível de expressão de genes diferencialmente expressos (*EPYC, SFRP5, C6ORF15* e *SLC25A15P2*); as vias metabólicas mais afetadas como a TP53, TGF-Beta, RTK-RAS. Em suma, o entendimento preciso dos mecanismos moleculares do glioma, pode auxiliar o desenvolvimento de tratamentos baseados na *medicina de precisão* para oferecer uma alternativa aos métodos tradicionais, ainda muito debilitantes ao organismo.

Palavras-chave: Glioblastoma multiforme. Sobrevivência. Genômica. Expressão Diferencial. Enriquecimento funcional.

PAULO, João Gabriel Fragoso. **Glioblastoma Multiforme: data analysis for tumoral process characterization**, Biomedical Enginnering Bachelor Degree, Federal University of Rio Grande do Norte, 62p., 2024.

ABSTRACT

Glioblastoma multiforme is a cancer of the central nervous system that affects astrocytes. Despite its low incidence in the world population (297,000 cases in 2018), it is difficult to treat due to its high degree of staging (grade IV). In addition, like all cancers, the causes of the manifestation are diverse and have a non-linear relationship. To this end, this study set out to investigate the risk factors and the various molecular mechanisms that contribute to its genesis and maintenance. This task was carried out using bioinformatics tools and had as its data source the cohort of 600 patients from the TCGA (The Cancer Genome Atlas) project. The analysis was based on survival curves, risk ratios, genomic analysis, methylation, differential gene expression and functional enrichment of pathways, using R language and R Studio development environment. Were identified race and sex as significant risk factors; methylation as main genomic factor for pathology development; *IDH* gene as a biomarker of favorable prognosis; 359 days as cohort‘s average survival time; the frequently altered genes, such as *PTEN, TP53* and *EGFR*; the expression level of differentially expressed genes (*EPYC, SFRP5, C6ORF15* and *SLC25A15P2*); the most affected metabolic pathways such as TP53, TGF-Beta, RTK-RAS. To sum up, the precise understanding of the molecular mechanisms of glioma can help to develop treatments based on *precision medicine* to offer an alternative to traditional methods, which are still very debilitating for organism.

Keywords: Glioblastoma multiforme. Genomics. Survival. Differential Expression. Functional Enrichment.

1. INTRODUÇÃO

De modo geral, os cânceres são uma doença genética causada pelo crescimento desordenado de células que ao dividirem-se rapidamente, originam tumores que se instalam em tecidos, até mesmo distante do local de origem (metástase) (MINISTÉRIO DA SAÚDE, [s.d]). As mortes pela enfermidade vêm aumentando mundialmente. Os principais fatores são: envelhecimento, aumento da população, aumento da expectativa de vida - na literatura, é usual a alta incidência da doença em idosos[1]. Todavia, é aceita no meio científico, a noção de que a maioria dos tumores malignos têm como maior contributo a carga ambiental (WU *et al.*, 2016) - apenas 30% com predominância por fatores genéticos.

O câncer a que se dedica este trabalho, o glioblastoma multiforme (GBM), é de baixa incidência mundial - o décimo sétimo mais frequente. Apesar de raro, na ampla maioria dos casos, conduz o indivíduo ao óbito (WILD *et al.*, 2020). Urge, portanto, a análise acurada em bancos de dados confiáveis objetivando extração de conhecimento tais como: as classes de pacientes mais suscetíveis a desenvolver o astrocitoma; o nível de efetividade dos tratamentos; genes tipicamente disfuncionais e a intensidade dessa disfunção. Questões como estas que constituíram o propósito de pesquisa associado a este trabalho de conclusão de curso.

Nos Estados Unidos da América, conforme dados do SEER (*Surveillance, Epidemiology, and End Results Program* - Programa de Vigilância, Epidemiologia e Resultados Finais) no período entre 2000 e 2021, foram registrados uma taxa média de incidência de glioblastoma multiforme no cérebro e outras partes do sistema nervoso, na qual 400.000 pacientes eram homens e 250.000 eram mulheres. Analisando a prevalência étnica-racial, dentre hispânicos, índios americanos e nativos do Alasca, asiáticos, pretos e brancos, houve um destaque para a raça branca, diferindo em média em 100.000 casos em relação a todas as demais raças e etnias. No período de 2004 a 2019, quanto ao estadiamento (estágio) da doença, dentre a detecção do câncer de forma distante (além do órgão de origem), localizada (pontual), por região, não estadiada, protagonizou, de modo importante, a forma localizada - 250.000 casos versus 50.000 correspondente ao subsequente estadiamento mais frequente, o regional. De todas as categorias disponibilizadas, o fator etário é o mais preponderante: é substancialmente maior a taxa de incidência entre os idosos (65 anos ou mais) em relação às demais idades. Uma média de 120.000 casos versus 64.000, essa sendo referente a segunda faixa mais prevalente (entre 50 e 64 anos). Ao longo dos anos, essa taxa tenderá a aumentar em 0,2%.

No contexto nacional, há aspectos que sinalizam a necessidade de maior observação e estudo de cânceres envolvendo o sistema nervoso em geral, e, por consequência, o glioblastoma

1 Disponível em: <https://ourworldindata.org/cancer>. Acesso em: 25 de out. 2022.

multiforme (GBM), que é um subtipo de astrocitoma, o câncer mais comum nesse sistema. O GBM se configura como o astrocitoma mais agressivo por apresentar maior estadiamento, grau IV, e é o câncer mais nocivo do sistema nervoso (ROSS *et al.*, 2011). Segundo o Atlas de Mortalidade do Instituto Nacional do Câncer (INCA, 2022), em 2020, os cânceres de sistema nervoso central são o oitavo tipo de câncer mais prevalente em mortalidade na população brasileira, sendo a taxa de óbito, entre dez tipos de cânceres, de 4,1% para homens, e, 4,2% para mulheres. Dentre todos os tipos de tumores malignos existentes, a taxa de óbito incidente na população mundial varia anualmente entre 1,4% a 1,8%. No período compreendido entre 2000 e 2020, houve um aumento de 83,86% da mortalidade do primeiro ano registrado em relação ao último, bem como aumentos entre um ano e outro. Em 2000, houve 5.008 óbitos. Já em 2020, 9.208 óbitos. No Brasil, assim como ocorre nos Estados Unidos, há a prevalência do sexo masculino - 81.413 mortes de homens contra 73.924 mortes de mulheres.

Além dos dados epidemiológicos, é essencial avaliar também os processos moleculares. Os dados clínicos, genômicos e moleculares deste trabalho foram extraídos do projeto TCGA-GBM. O TCGA (*The Cancer Genome Atlas*) é um programa dedicado a fornecer estes tipos de dados para variados cânceres. Para isso, emprega tecnologias de rápido sequenciamento gênico e é apoiado pelos Institutos Nacionais de Saúde dos Estados Unidos (NIH - *National Institutes of Health*). Em virtude disso, possui grande impacto nas pesquisas empreendidas para a compreensão e tratamento de cânceres. Por isso, é um dos projetos com maior aceitação no portal GDC (*Genomic Data Commons*) - repositório que reúne dados advindos de vários programas.

2. OBJETIVOS

2.1. Geral

A partir da análise e revisão de dados clínicos, moleculares e de publicações relativas ao câncer em questão, buscou-se encontrar os padrões associados à gênese tumoral do glioblastoma bem como fatores que sinalizassem risco ao seu desenvolvimento.

2.2. Específicos

As metas para se atingir o objetivo geral serão a caracterização e análise de:

- Dados clínicos;
- Sobrevivência e de risco em relação à atributos clínicos;
- Alterações no genoma, em relação a nucleotídeos únicos (SNVs) e número de cópias (CNVs);
- Metilação;
- Expressão dos genes diferencialmente expressos (DEA);
- Enriquecimento funcional das vias metabólicas e processos contendo os genes diferencialmente expressos.

3. REVISÃO BIBLIOGRÁFICA

3.1. Câncer: Conceitos Gerais

O câncer pode ser definido como uma desordem celular na qual há uma multiplicação irrefreada, e, com possibilidade de espalhamento em direção a células normais do organismo, competindo por recursos energéticos do sistema, para o crescimento tumoral segundo um ganho funcional. O tumor que sofreu o espalhamento além do tecido de origem, o câncer propriamente dito, é considerado de causa genética: são mutações em determinadas classes de genes (proto-oncogenes, supressores de tumor e reparadores de DNA) que terão sua expressão fenotípica em um período de tempo variável em função do tecido, órgão e diversos parâmetros do indivíduo portador, como idade, herança genética, exposição a fatores de risco, tais como, radiação ultravioleta proveniente do Sol, substâncias com alta concentração de tóxicos, estresse sistemático (NCI, 2007).

Ressalta-se que a causa de origem é necessária, porém não suficiente para o surgimento efetivo do câncer: no estágio de iniciação, os fatores cancerígenos ou mutações aleatórias alteraram o material genético da célula sem haver a manifestação tumoral. Apenas no estágio de promoção, havendo o contato continuado com carcinógenos, um tumor benigno é instalado no tecido. A sua mudança para o aspecto maligno se dá na passagem para o estágio de progressão (INCA, 2022). Nele, os chamados oncogenes estão plenamente ativos, gerando e fomentando o comportamento de retroalimentação positiva tumoral e metástase - cria-se um microambiente de variados tipos celulares (NCI, 2007) a fim de alimentar as células cancerígenas por meio dos aparatos sadios do corpo. No contexto metastático encontram-se as maiores incidências de óbito.

Devido a qualquer tipo de célula ser passível de se transformar em uma célula cancerígena, existem tantos tipos da doença quanto há tipos teciduais. Ainda que o tumor possa atingir outros órgãos além do originário, sua nomenclatura parte do local do tumor primário. Logo, como exemplo, células cancerígenas que primeiramente manifestaram-se no pulmão e dispersaram-se para as mamas (metástase), diz-se, ainda, que o indivíduo desenvolveu "câncer de pulmão".

Um dado tipo de tumor maligno possui indicativos qualitativos relevantes na decisão do enfoque clínico: o grau e o estadiamento (*stage*). O grau avalia a proximidade em aspectos morfológicos (diferenciação) do tumor em relação a uma célula normal (NCI, 2013). Tem-se no espectro, os graus 1, 2, 3, 4 e grau X (desconhecido), tal que há uma ordem crescente de indiferenciação celular. Graus maiores implicam em taxa de metástase mais acelerada. Paralelamente, a medida de estadiamento (*stage*), refere-se a três condições para a maioria dos cânceres (sistema TNM): tamanho tumoral, T, linfonodos afetados, N (em número e em

localização), e existência de metástase, M (NCI, 2015). Os estadiamentos são classificados em ordem crescente de severidade (I, II, III e IV) em função dos parâmetros referidos (TNM). Tal definição confere maior elucidação prognóstica, entretanto, um grau tumoral elevado tende a igualmente possuir estadiamento avançado.

Os modos de se tratar o câncer são diversos e sua escolha parte da análise médica considerando a relação beneficência-maleficência mais favorável, uma vez que efeitos colaterais são inerentes aos tratamentos. Dentre os convencionais, estão a cirurgia, quimioterapia, radioterapia e transplante de medula óssea, as quais, por vezes, são aplicadas em conjunto (INCA, 2022). A cirurgia reside na remoção do tumor. Se especialmente este for benigno, é um método resolutivo. Já para tumores que sofreram metástase, a quimioterapia atua no sentido de inserir uma substância ativa na corrente sanguínea que irá aniquilar, desacelerar ou parar o crescimento tumoral. A radioterapia segue-se analogamente: aplica-se um fármaco radioativo em dada região demarcada do paciente para suscitar a morte ou a inibição da atividade das células cancerosas. É comum que a cirurgia proceda de sessões de radioterapia ou quimioterapia. As primeiras refreando o avanço metastático de modo emergencial, e a última, pondo fim aos tumores. Para os casos em que o tecido alvo do câncer é o sanguíneo, a abordagem de tratamento passa a protagonizar o transplante de medula óssea - há a recepção de células saudáveis de medula óssea visando a reposição dos elementos figurados do sangue (plaquetas, leucócitos, hemácias). Por isso, o transplante é empregado também após sessões de quimioterapia ou radioterapia, onde há perda desses componentes, como efeito colateral.

Na atualidade, os tratamentos da patologia vêm adquirindo um viés de medicina de precisão, a qual é realizada a partir de análises estatísticas de dados sobre alterações moleculares depositadas em diversos bancos de dados, e estudos populacionais. Esta abordagem genômica tem auxiliado sobremaneira o entendimento dos mecanismos que promovem o câncer, além de auxiliar o desenvolvimento de drogas 'customizadas' para tipos ou subtipos específicos de câncer - “terapia alvo” (NCI, 2007), minimizando, portanto, a agressividade das intervenções terapêuticas tradicionais como a quimioterapia, a qual mobiliza um significativo contingente celular, ainda que, em sua maioria, não englobe a população tumoral.

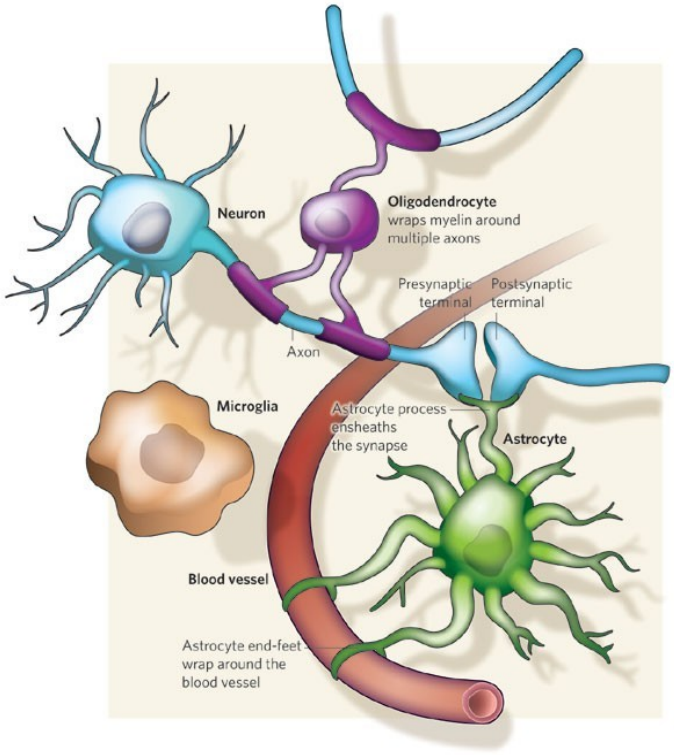
Aliado aos tratamentos convencionais, órgãos como o Instituto Nacional do Câncer (INCA), provê cuidados paliativos (INCA, 2022), isto é, a manutenção da saúde considerando aspectos gerais e individuais do paciente. Na prática, esses cuidados traduzem-se pelo acompanhamento nutricional, vide o desequilíbrio orgânico causado pelo transtorno e seu tratamento; assistência psicológica atuando no alívio dos impactos emocionais provocados pelo significado de portar

câncer, principalmente nas fases terminais; interações de quaisquer naturezas que fomentem o bem-estar e conforto.

3.1.1. Câncer de Sistema Nervoso e Glioblastoma Multiforme

O câncer de análise deste trabalho, glioblastoma multiforme (GBM), é incidente em astrócitos, tipos de células gliais. Estas são responsáveis por fornecerem aprovisionamento e suporte ao neurônio (Figura 1). Para cada tipo de célula glial, há um câncer associado (Figura 2). No entanto, é mais comum observar neoplasias em astrócitos, uma vez que há maior quantidade dos mesmos em relação aos demais subtipos celulares. Particularmente, a atuação astrocítica refere-se a modulação da propagação elétrica nos axônios e regulação metabólica de neurônios (Figura 1): remove excesso de neurotransmissores nas regiões intercalares das células nervosas (ROSS *et al.*, 2011). Além disso, é notável na constituição da barreira hematoencefálica (BHE), estrutura na qual dificulta-se a difusão de grande número de substâncias do sangue para o sistema nervoso central. Os astrócitos transportam os compostos aos endotélios integrantes da BHE (KIERSZENBAUM, 2008), bem como conferem seletividade a esta, alterando-se o tônus vascular dos capilares (DEUS, 2013).

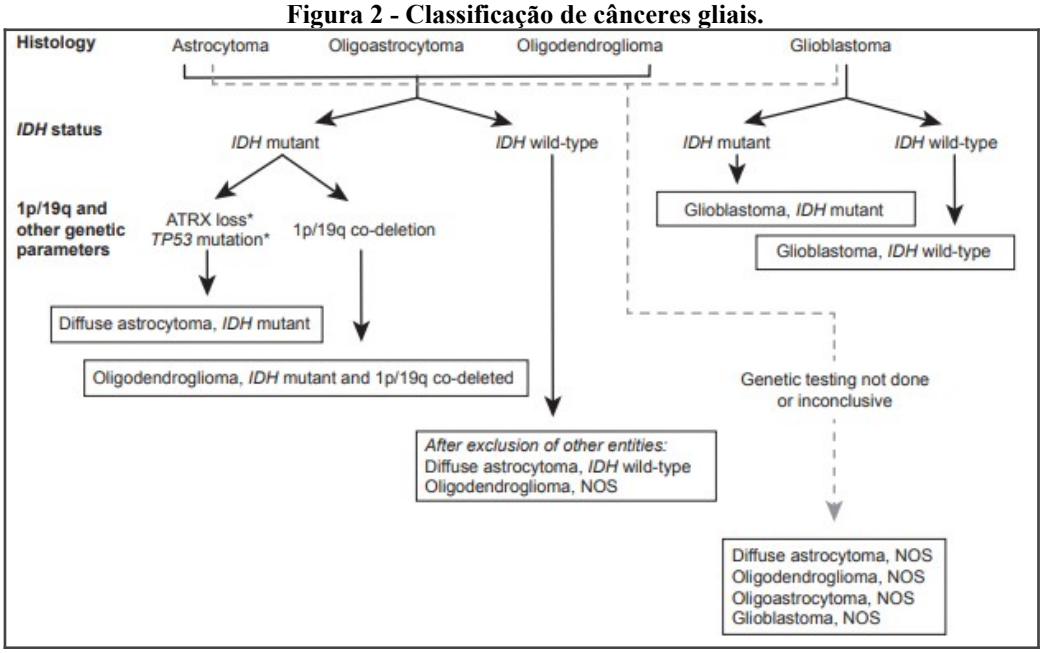
Figura 1 - Conjunto de células gliais.



Neurônio (azul), oligodendrócito (roxo) e micróglia (marrom). O astrócito (verde), célula onde incide o glioblastoma multiforme, é mostrado conectando vasos sanguíneos e sinapses nervosas, a fim de difundir substâncias entre estas partes. Fonte: Allen *et al.*, 2009.

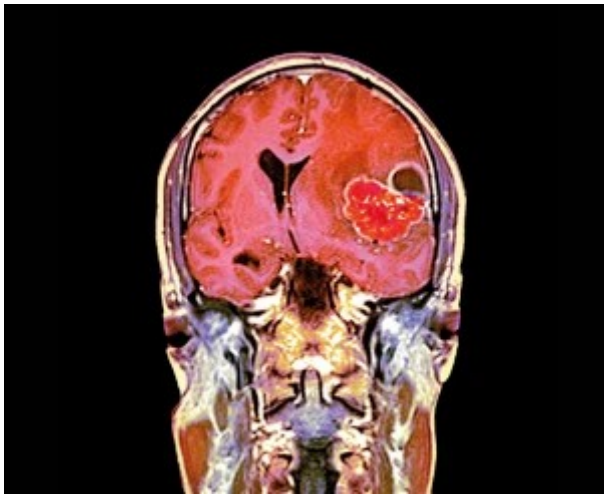
A maioria dos tumores cerebrais são relativos às células da glia, dado que estas superam em número os neurônios, e conforme supracitado, nesta classe, os astrócitos são o tipo mais abundante, por isso, 80% dos tumores primários no cérebro em adultos são astrocitomas fibrosos (ROSS *et al.*, 2011). Em crianças, "a sobrevivência global é de apenas 20 a 30% no prazo de três anos após o tratamento" (MANUAL MSD, 2021). Os astrocitomas, em geral, são detectados por ressonância magnética, tomografia computadorizada, ou biópsia (Figura 3), sendo esta última, o meio para a classificação do estadiamento.

Figura 2 - Classificação de cânceres gliais.



Os cânceres gliais são classificados de acordo com sua histologia, *status* do gene *IDH* e outros parâmetros genéticos. O glioblastoma é indicado como um subtipo de astrocitoma. Fonte: World Health Organization, 2020.

Figura 3 - Glioblastoma Multiforme visto no cérebro em corte frontal.



Fonte: Sidaway, 2017.

Em razão da natureza agressiva correspondente ao GBM (grau IV), na prática, seus danos a nível sistêmico se dão pelo acúmulo descontrolado de massa tecidual oriundo do tumor em metástase. Dessa forma, o volume anômalo resultante gera pressões, fazendo com que a área cerebral submetida perca progressivamente sua função. Há também uma série de possíveis sintomas associados, como convulsões, perda de equilíbrio, problemas visuais, auditivos ou de fala, e aumento no tamanho da cabeça em crianças. O tecido-alvo, com menos frequência, pode ser a medula espinhal, havendo variação dos sintomas e ausência da perda funcional por pressão, como ocorre nos astrocitomas cerebrais (NCI, 2022).

3.1.2. Panorama global e nacional da incidência de GBM

Sendo o glioblastoma multiforme um câncer de Sistema Nervoso Central (SNC), tem-se que o cérebro é o órgão no qual mais surgem os tumores, dada a maior concentração neuronal. Em âmbito mundial, os cânceres de cérebro são de baixa incidência, o décimo sétimo mais frequente. Em 2018, apenas 297.000 indivíduos manifestaram GBM. A diferença de casos entre países se deve a diferenças metodológicas (modo como se aplica o diagnóstico, classificação) (IARC, 2020). No cenário nacional, os cânceres relativos ao Sistema Nervoso Central (SNC) não constam entre os mais incidentes: no período de 1979 até 2020, segundo dados do Atlas On-line de Mortalidade do Instituto Nacional do Câncer (INCA), as 5 localizações primárias mais frequentes por 100.000 habitantes foram nos brônquios e pulmões; mama; próstata; estômago; cólon. Ademais, observa-se uma baixa distribuição populacional (proporcional) dos cânceres de SNC (CIDS C71 - Encéfalo; e C72 - Medula Espinhal e outras partes do SNC). Nos períodos de 2011 a 2015, 3,84% para tumores encefálicos, 0,43% para tumores na medula espinhal e outras partes do SNC; de 2016 a 2020, 3,62% para tumores encefálicos, e 0,49% para tumores na medula espinhal e outras partes do SNC.

Quanto ao número de anos potenciais de vida perdidos por cânceres do SNC, no Brasil, os dados do INCA são pouco díspares com o observatório *Our World in Data* comparando os óbitos por qualquer tipo de câncer. Este registrou que os indivíduos que possuíam 70 anos ou mais, concentraram, em termos percentuais, quase metade do número de óbitos por cânceres de qualquer tipo, do ano de 1990 até 2019. Enquanto que, percentualmente, a “taxa de anos potenciais de vida perdidos” (TAPVP) em cânceres do SNC (INCA), são maioria para o grupo etário de 60 à 69 anos - 1,6 anos potenciais de vida perdidos. Qualitativamente, os dados coadunam com o fator atribuído à

mortalidade crescente de cânceres no mundo apresentado no observatório *Our World in Data*: *o aumento da expectativa de vida*[2].

3.2. O programa TCGA - The Cancer Genome Atlas

Dentre os Institutos Nacionais de Saúde dos Estados Unidos (NIH - *National Institutes of Health*), o Instituto Nacional do Câncer (NCI - *National Cancer Institute*) possui um centro dedicado ao estudo genômico do câncer (*NCI Center for Cancer Genomics*). Sob seu comando, é disponibilizado ao público um repositório de diversos bancos de dados de cânceres (GDC - *Genomic Data Commons*). Ao todo, 23 programas produzem segundo suas técnicas próprias os dados e muitos deles recebem apoio do NCI. O programa TCGA *(The Cancer Genome Atlas*) é um dos principais, com um reconhecido impacto no avanço de pesquisas e geração de dados relativos aos cânceres.

Com o avanço das tecnologias de alto rendimento, aliado ao aumento da escalabilidade e diminuição do custo (NIH, 2019), o TCGA ao longo dos anos, gerou, organizou e analisou as alterações moleculares e dados clínicos de 33 tipos de câncer. Para cada tipo de câncer é possível resgatar o ID (sequência numérica) e informações clínicas do paciente, o tecido primário de ocorrência, lâminas para a visualização da amostra tecidual, e uma série de arquivos contendo dados brutos e harmonizados (processados) de sequenciamento do genoma, perfil do transcriptoma e proteoma, variação de nucleotídeo único (SNVs), variação do número de cópias (CNVs), metilação do DNA, variação estrutural (cromossomal), e espécime (informações do tecido para amostragem histológica). Porém, nem todos os arquivos são de acesso aberto. A maior concentração de dados no GDC para o glioblastoma multiforme é do TCGA (599 casos), fator que contribuiu com a escolha do projeto.

Em relação às intervenções terapêuticas, o TCGA vem atuando de forma importante na medicina de precisão aplicada ao câncer (HUTTER e ZENKLUSSEN, 2018), isto é, aplica-se um método de tratamento com as características genômicas individuais. Assim, descobertos pontualmente os alvos (*targets*) - podendo ser gênicos, domínios proteicos, e outras macromoléculas (HOPKINS e GROOM, 2002), criam-se novos fármacos, e em paralelo, tornam direcionadas às terapias já existentes.

2 Disponível em: <https://ourworldindata.org/life-expectancy >. Acesso em: 25 de out. 2022.

3.3. Natureza dos dados ômicos

O conceito de dados ômicos (*omic data*) é usado para designar o conjunto de moléculas de uma determinada natureza (proteínas - proteômica, RNAs - transcriptômica, etc). Nesta seção será detalhada sumariamente a fonte biológica de *componentes* que foram usados na caracterização do glioblastoma multiforme. Os fenômenos abaixo são comuns de ocorrerem nas células, portanto não acarretam necessariamente na manifestação de câncer. Efetivamente, é a conjunção desses eventos somada às suas intensidades que torna maior o risco de desenvolver a doença.

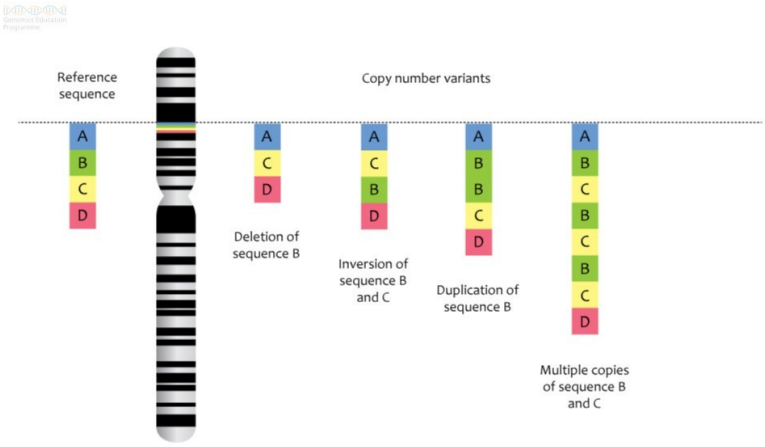
3.3.1. Alterações gênicas

As alterações em nucleotídeos únicos podem ser caracterizadas como mutações, ou de forma mais abrangente, como polimorfismos. Tais mudanças são classificadas como: (i) sinônima, quando as alterações no nucleotídeo não alteram o códom de identificação do aminoácido; (ii) sem sentido (*nonsense*), quando ocorre a criação de um códon de parada, ou (iii) não-sinônima (*missense*), quando há a modificação do aminoácido codificado. Há também outros tipos de alteração, como a mudança na janela de leitura devido à inserção ou deleção de nucleotídeos (*indels*), e, a mudança no sítio de ligação (*splice site*), ou seja, nas delimitações do processo de maturação do RNA (*splicing*), todas com um impacto significativo no produto gênico.

3.3.2. Variação do número de cópias (CNV)

Ao nível do cromossomo, durante seu processo de replicação, podem ser produzidas cópias gênicas com duplicações (amplificações) ou deleções de uma região cromossômica (HAMAD, [s.d]). Dentro de uma mesma citobanda pode haver diferentes níveis de amplificação ou deleção do número de cópias. A variação das cópias gênicas também pode ser oriunda da recombinação, replicação e reparo de DNA (PÖS et al., 2021).

Figura 4 – Variantes de número de cópias em genes.

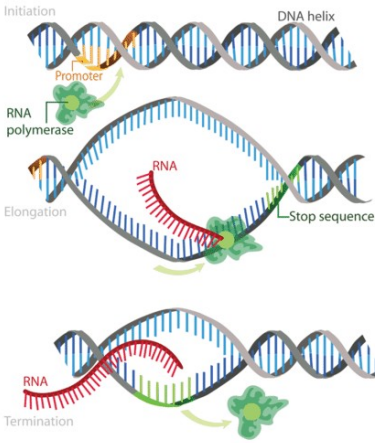


As cópias podem apresentar deleção, inversão, duplicação ou mesmo repetição de sequências nucleotídicas.
Fonte: Hamad, [s.d].

3.3.3. Expressão gênica diferencial

A expressão de um gene é entendida como a quantidade dos RNAs transcritos, sejam eles um RNA não codante ou um RNA-mensageiro (Figura 5). Uma série de fatores influencia a expressão gênica, como mutações gênicas, alterações do número de cópias, rearranjos cromossômicos, regulação por RNAs não codantes e fenômenos epigenéticos como a metilação. E logicamente, a expressão gênica está intimamente relaciona à linhagem celular.

Figura 5 - Mecanismo geral da expressão gênica.



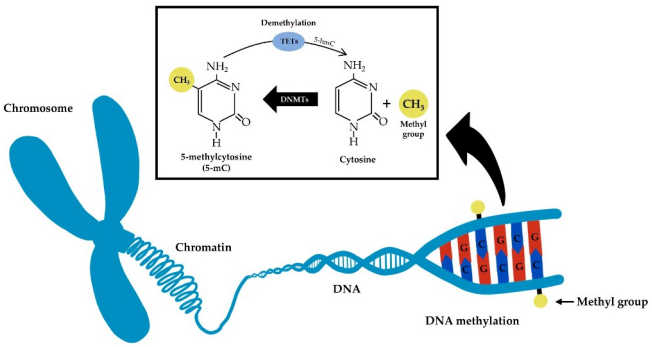
A partir de uma fita de DNA e da complementariedade de bases, a enzima RNA polimerase codifica uma molécula de RNA mensageiro correspondente. Os diferentes fenótipos observados são, em parte, devidos à variação quantitativa e qualitativa de transcritos, isto é, à expressão diferencial. Fonte: LibreTexts Biology[3].

3.3.4. Metilação

A metilação é um fenômeno epigenético reversível que consiste na adição (hipermetilação) ou retirada (hipometilação) de grupos metil, principalmente em ilhas CpGs (5'-citosina-p-guanina-3'). Em todos os tipos de cânceres, há um aumento ou diminuição significativa dessa atividade, conferindo ganhos ou perdas de expressão em determinados genes. Do ponto de vista molecular, esse mecanismo é mediado pelo nível de compactação do DNA. A acetilação, adição do grupo acetil, leva à descompactação de DNA, facilitando a transcrição gênica. Em contrapartida, a metilação promove o enrolamento de DNA, tornando a região metilada pouco acessível para transcrição (MASCARO, 2020) (Figura 6).

3 Disponível em: <https://bio.libretexts.org/Courses/Lumen_Learning/Biology_for_Non-Majors_I_%28Lumen%29/09%3A_DNA_Transcription_and_Translation/9.02%3A_Transcription>. Acesso em: 16 ago. 2024.

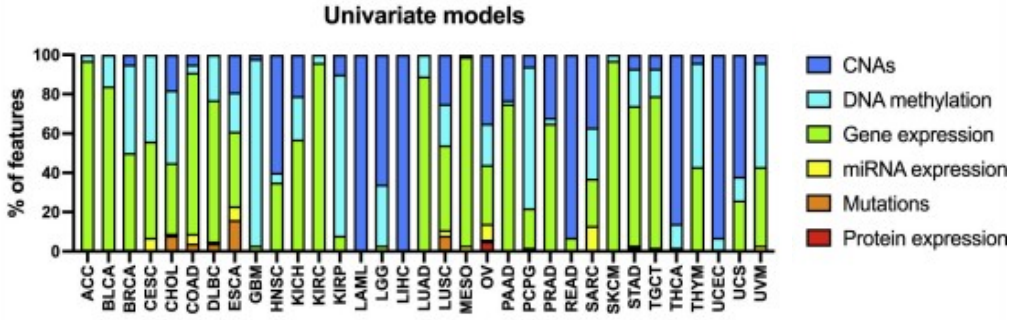
Figura 6 - Metilação do DNA ao nível da cromatina.



A adição de grupos metis (CH_3) provenientes de enzimas metiltransferases (DNMTs), diminui a expressão gênica de uma ilha CpG (dinucleotídeos de citosina e guanina). Já a extração de grupos metis pode ocorrer passivamente durante a divisão celular, ou, pela ação das enzimas TETs. Fonte: MDPI[4].

Dos 33 tipos de cânceres que contemplam o projeto TCGA, o glioblastoma multiforme é o que apresenta a metilação como o fator mais relevante para a qualidade do prognóstico em modelo univariado (SMITH *et al.*, 2022), ou seja, estatisticamente dependente apenas de características genômicas do paciente (Figura 7). Isso é ocasionado pela metilação diferencial em pelo menos um dos genes, como *MGMT, SOX10, TBX3* ou *FNDC3B*. Adicionalmente, a análise do padrão de metilação no GBM permite identificar oncogenes como possíveis alvos do desenvolvimento de fármacos "metiladores", os quais já são empregados em algumas neoplasias com esse propósito (ETCHEVERRY et al., 2010).

Figura 7 - Características moleculares mais informativas para a predição dos resultados dos pacientes nos projetos TCGA.



De acordo com o modelo de Cox univariado, a metilação do DNA é o principal preditor de sobrevivência dos pacientes com GBM. Fonte: Smith *et al.*, 2022.

4 Disponível em: <https://www.mdpi.com/2079-4991/13/12/1880/>. Acesso em: 16 ago. 2024

4. METODOLOGIA

4.1 - Desenvolvimento

Este trabalho foi desenvolvido utilizando a linguagem R versão 4.3.1, e a interface gráfica associada, RStudio Server versão 2023.09.0 Build 463 sediado pelo Centro Multiusuário de Bioinformática do Instituto Metrópole Digital - UFRN (BioME - IMD/UFRN). Todos os pacotes utilizados são de código aberto e foram obtidos nos repositórios Bioconductor[5], CRAN[6] e Tidyverse[7]. Os códigos estão disponíveis no GitHub[8].

4.2 - Obtenção de dados

Os dados de 599 pacientes do projeto TCGA-GBM (The Cancer Genome Atlas - Glioblastoma Multiforme) foram extraídos do GDC Data Portal (Genome Data Commons[9]) utilizando o pacote TCGA Biolinks, versão 2.24.3 (COLAPRICO et al., 2015). Os dados clínicos foram adquiridos em centros especializados. Os dados moleculares foram gerados a partir do sequenciamento do exoma de amostras tumorais de alta qualidade (identificação de mutações), medição baseada em microarrays (número de cópias - CNV), sequenciamento de RNA e geração de microarrays para análise transcriptômica e de expressão gênica[10]. A coluna TCGA barcode é comum ao conjunto dos dados clínicos e moleculares, isso permitiu referenciá-los de forma mútua.

Para a análise do número de cópias gênicas, utilizou-se o website cBioPortal[11] para realizar o download dos dados clínicos e genômicos da coorte TCGA-GBM. Dentre os estudos disponíveis, foi escolhido como fonte a *Firehouse Legacy* visto que o número de pacientes selecionados, 606, foram os que mais se equiparavam à mesma coorte obtida pelo GDC.

4.3 - Análise dos dados clínicos

Para a depuração e processamento dos dados clínicos foi utilizada a suíte de pacotes *tidyverse* versão 1.3.2 (WICKHAM *et al.*, 2019) e o pacote *skimr* versão 2.1.4 (WARNING *et al.*, 2022). A maioria das variáveis clínicas foram excluídas da análise, em face ao grande número de

5 Disponível em: <https://bioconductor.org/>.
6 Disponível em: <https://cran.r-project.org/>.
7 Disponível em: <https://tidyverse.tidyverse.org/>.
8 Disponível em: <https://github.com/jgabrielbio/tcga-gbm>.
9 Disponível em: <https://gdc.cancer.gov/>.
10 Disponível em: <https://portal.gdc.cancer.gov/projects/TCGA-GBM>.
11 Disponível em: <http://www.cbioportal.org/>.

valores não identificados, bem como, a natureza de algumas serem irrelevantes para o desenvolvimento de GBM. A análise exploratória foi realizada utilizando comandos nativos do R e funções do *Tidyverse* para a geração de tabelas customizadas, histogramas, testes qui-quadrado, boxplot, tabelas de correlação (pacote *corrplot* versão 0.92 - WEI *et al*., 2021). A disposição final das informações clínicas foi feita com o pacote f*inalfit* versão 1.0.6 (HARISSON *et al*., 2023). Empregou-se o *teste de Wilcoxon* para a análise estatística da idade. Já para todas as demais, foi aplicado o *teste exato de Fisher*.

4.4 - Análise de sobrevivência e taxa de risco

Utilizando o pacote *survminer* versão 0.4.9 (KASSAMBARA *et al*., 2021), obtiveram-se curvas de *Kaplan-Meier*, que calcula a probabilidade de indivíduos permanecerem vivos até o fim de determinado tempo (probabilidade condicional). A sobrevivência global (OS) é o período desde a data do diagnóstico (dia 0) até a data da morte por qualquer causa. O tempo censurado é contabilizado a partir da data do diagnóstico inicial até a data do último contato. A derivação de resultados de OS não foi problemática, pois dependeu somente do status do paciente - vivo ou morto, podendo incluir mortes não relacionadas ao câncer. O nível de significância entre curvas de sobrevivência mensurado pelo valor-p é deduzido pela medida log-rank - exclusiva para aplicação em dados censurados (símbolo "+" visto nas figuras 9 e 10).

Uma análise de taxa de risco (*hazard ratio*) foi realizada a fim de confirmar a maior ou menor propensão do desenvolvimento do glioblastoma de determinado subgrupo para cada fator contemplado anteriormente nas curvas de Kaplan-Meier. Para tanto, utilizou-se funções do pacote *survival*, versão 3.5-0 (THERNEAU, 2023). O cálculo da taxa de risco é uma razão entre razões, eq. (1):

$$HR = \frac{\left(\frac{O_a}{E_a}\right)}{\left(\frac{O_b}{E_b}\right)} \quad (1)$$

O_a e O_b são a quantidade de observações com risco dos subgrupos das variáveis binárias A e B. E_a e E_b são o espaço amostral. Tem-se um risco aumentado em certo número de vezes para determinado subgrupo quando a taxa de risco é maior que 1 (HR > 1), e um risco diminuído quando a taxa de risco é menor que 1 (HR < 1). O teste estatístico empregado foi o log-rank. A partir dele, foram consideradas significativas as variáveis categóricas cujo p-valor foi menor que 0,05.

4.5 - Análise Genômica

A partir de funções do pacote *Maftools* versão 2.12.0 , foi possível obter-se o arquivo MAF (*Mutation Annotation Format*), que caracteriza um "arquivo de texto delimitado por tabulação padrão introduzido pelo TCGA para armazenar e distribuir variantes somáticas, contendo a paisagem somática completa da coorte" (MAYAKONDA *et al*., 2018). O *Maftools* permite verificar graficamente a relevância de determinados genes tanto na formação tumoral quanto na variação dos efeitos do glioblastoma multiforme, em particular, na sobrevivência do paciente.

A análise do número de cópias gênicas (CNV - *Copy Number Variation*) a partir de arquivos GISTIC (*Genomic Identification of Significant Targets In Cancer*) foi realizada pela plataforma online GenePatterns[12], selecionando-se o genoma de referência (hg19) e anexando-se o arquivo de segmentação (.seg).

4.6 - Análise de Metilação

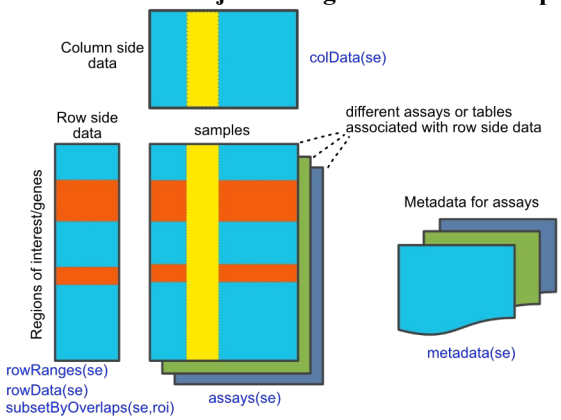
O grau de metilação de um determinado segmento de DNA pode ser mensurado pelo valor β que varia entre 0 e 1, e se baseia na diferença entre a condição metilada e não-metilada. Assim, valores próximos à 1 são mais significativamente metilados, considerando p-valor < 0,05. Para a análise e visualização de regiões com grau relevante de metilação (β = 0,5), utilizou-se funções do pacote *TCGA Biolinks*, versão 2.24.3 (COLAPRICO *et al.*, 2015).

4.7 - Análise de Expressão Diferencial

Para a análise dos genes diferencialmente expressos foi utilizado o pacote *DESeq2,* versão 1.41.10 (LOVE; HUBER; ANDERS, 2014). Os dados brutos de contagem de RNAs (*raw counts*) foram transformados em objeto da classe *RangedSummarizedExperiment* do pacote *Summarized Experiment,* versão 1.30.2 (MORGAN *et al.*, 2023), que cria uma matriz onde as linhas se associam aos genes, e as colunas às amostras de cada paciente, sendo possível selecionar diversas variáveis experimentais por meio dos metadados (Figura 8).

12 Disponível em: <https://www.genepattern.org/>.

Figura 8 - Estrutura do objeto 'RangedSummarizedExperiment'.



Na matriz de ensaio (*assay(se)*) as linhas em laranja correspondem aos genes e as colunas em amarelo correspondem às amostras. Os metadados com informações de cada experimento se encontram na matriz (metadata(se)). Fonte: Computational Genomics with R[13].

A matriz DESeq estimou para cada gene o grau de alteração de expressão (o número de dobras, *log fold-change*) em relação à expressão no grupo controle (a média das contagens normalizadas para todas as amostras). Para os dois parâmetros experimentais adotados, o status vital e status do gene *IDH*, assumiu-se um p-valor de 0,01. Os genes considerados superexpressos (*log fold-change* ≥ 1,5*),* ou subexpressos (*log fold-change* ≤ -1,5), foram selecionados para posterior análise de enriquecimento funcional. A visualização da expressão dos genes foi feita em *volcano plots*. Para tornar acessível a investigação dos genes diferencialmente expressos, converteu-se seus IDs de *ENSEMBL* para *HUGO* através da ferramenta online BioTools[14].

4.8 - Análise de Enriquecimento Funcional

A análise das vias enriquecidas no GBM foi realizada utilizando o pacote *ClusterProfiler*, versão 4.3.1.900 (WU, 2023), tendo como dados de entrada a matriz de genes diferencialmente expressos (seção 4.6) e como parâmetro experimental, o *status* do gene *IDH*. Optou-se pelo tipo de análise GSEA (análise de enriquecimento de grupo gênico) em detrimento da análise ORA (análise de sobre-representação), uma vez que o primeiro método classifica o grau de enriquecimento baseando-se na distribuição de expressão diferencial ao longo do conjunto de dados, ao invés de realizar sobreposições ao acaso entre dois genes, como ocorre no segundo método.

A partir dos resultados da análise GSEA, obteve-se a visualização de processos biológicos mais alterados, em função do tamanho dos conjuntos gênicos, contra o banco de dados KEGG[15],

13 Disponível em <https://compgenomr.github.io/book/genomic-intervals-with-more-information-summarizedexperiment-class.html>. Acesso em: 20 fev. 2024.

14 Disponível em <https://www.biotools.fr/human/ensembl_symbol_converter>.

15 Disponível em: <https://www.genome.jp/pathway/map05214>.

utilizando como a medida o escore de enriquecimento normalizado (NES). A montagem dos gráficos foi feita utilizando o pacote ggplot2, versão 3.5.0 (WICKHAM, 2016). O *ClusterProfiler* permite acessar o mapa das vias metabólicas envolvidas no GBM, segundo o KEGG. Agregando os dados GSEA no mapa, verificou-se os genes mais enriquecidos em determinadas rotas.

5. RESULTADOS

5.1. Descrição da Coorte

De 141 variáveis de dados clínicos, epidemiológicos e dados sobre as amostras, apenas 7 foram tomadas como relevantes para a análise (Tabela 1). Com o pacote *finalfit* foi possível extrair o valor e o percentual dos grupos nas categorias principais a partir do total dos pacientes (599). Considerando apenas as variáveis cujo resultado é significativo (p-valor < 0,05), os portadores de glioblastoma multiforme, apresentaram uma média de idade de 57,8 anos, representando 14,4% da amostra; a maioria é de brancos (88,8%), homens (366 - 61,4%), com tratamento farmacêutico (473 - 87,4%), com tratamento radioativo (497 - 87,3%).

Tabela 1 – Distribuição das variáveis clínicas da coorte TCGA-GBM, de acordo com o status vital. (continua)

Categoria	Níveis	Vivo	Morto	p-valor
Idade	Média (SD – Desvio Padrão)	54,7 (15,9%)	58,4 (14%)	0,017
Raça	Não-branca	24 (25%)	40 (8,5%)	<0,001
	Branca	72 (75%)	432 (91,5%)	
Sexo	Feminino	46 (45,5%)	184 (37,4%)	0,156
	Masculino	55 (54,5%)	308 (62,6%)	
Etnia	Hispânica ou latina	3 (3,6%)	10 (2,4%)	0,788
Tratamento prévio	Não	100 (99%)	471 (95,9%)	0,219
	Sim	1 (1%)	20 (4,1%)	
Tratamento farmacêutico	Não	6 (6,4%)	62 (13,9%)	0,067
	Sim	88 (93,6%)	383 (86,1%)	

Tabela 1 – Distribuição das variáveis clínicas da coorte TCGA-GBM, de acordo com o status vital. (conclusão)

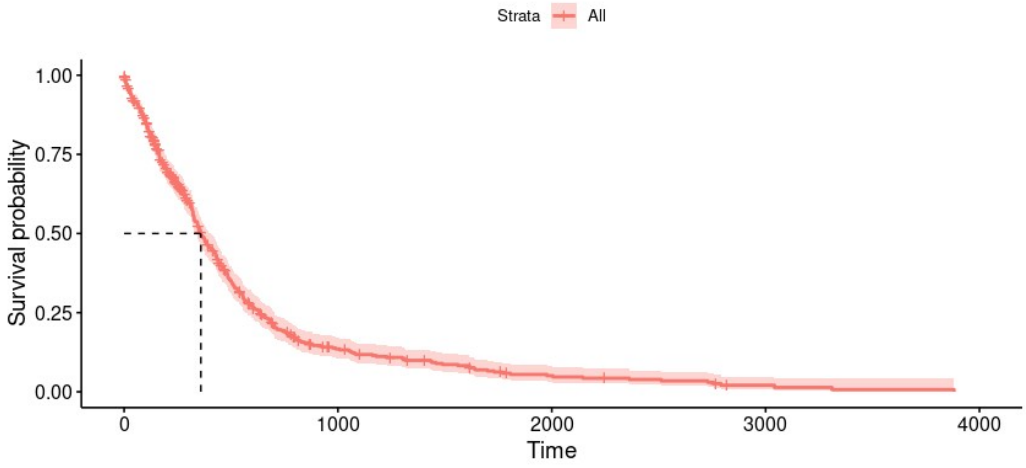
Categoria	Níveis	Vivo	Morto	p-valor
Tratamento radioativo	Não	3 (3,1%)	69 (14,7%)	0,003
	Sim	94 (96,9%)	40 (85,3%)	

Fonte: Autoral.

5.2. Análise de sobrevivência

Faz-se fundamental identificar quantitativamente os fatores que predispõe a óbitos por glioblastoma multiforme, ressaltando seus níveis de significância, para assim mensurar a gravidade da doença e por conseguinte alertar os grupos de risco, bem como, priorizá-los enquanto pacientes afetados. A análise de sobrevivência sem estratificação para os dados do TCGA-GBM apresentou uma mediana de sobrevivência de cerca de 1 ano (Figura 9).

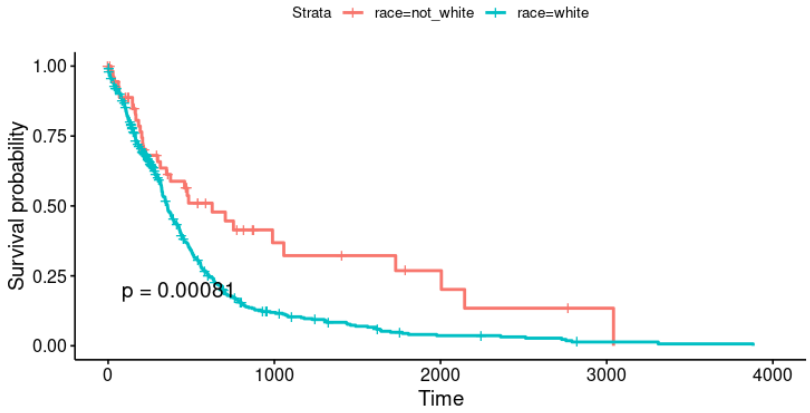
Figura 9 - Curva de sobrevivência da coorte TCGA-GBM sem estratificação.



A probabilidade de sobrevivência está em função do tempo em dias. Fonte: Autoral.

Para o fator racial, a diferença mostrou-se bastante relevante (p = 0,00081) com uma melhor sobrevivência para os não-brancos, por possuir decaimento de probabilidade menos íngreme. A mediana da sobrevivência para não-brancos é de aproximadamente 1,38 anos, enquanto para brancos é de 10 meses.

Figura 10 - Estimativa de Kaplan-Meier para sobrevivência dos pacientes, estratificados por raça.



Curvas de sobrevivência de acordo com a raça: não-branca (azul) e branca (vermelho). P-valor determinado pelo teste de Log-rank. Fonte: Autoral.

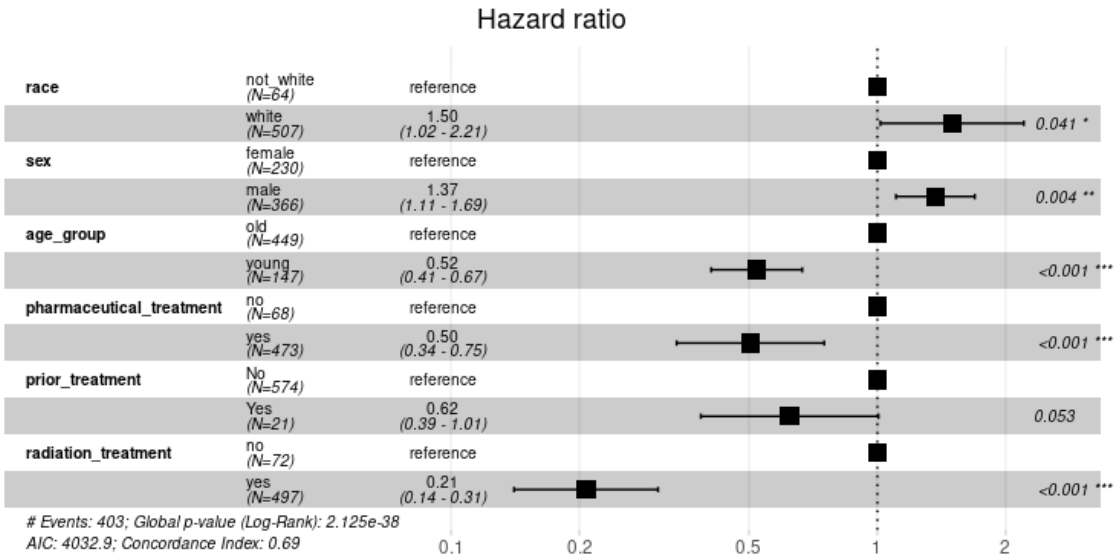
Outros fatores onde a diferença entres as categorias mostrou-se significativa foram faixas etárias ($p < 0{,}0001$); tratamento farmacêutico ($p < 0{,}001$); radioterapia ($p << 0{,}05$); tratamento prévio ($p << 0{,}05$). Nessas variáveis, os subgrupos com melhor prognóstico foram, respectivamente, jovens, pacientes que não ingeriram fármacos, que não receberam radioterapia e que não foram acompanhados previamente.

5.3. Análise da taxa de risco de sobrevivência

Enquanto a análise de sobrevivência fornece uma noção temporal dos riscos, a análise da taxa de risco infere a probabilidade de dada variável ocasionar ou não uma queda de sobrevivência. Desse modo, essa medida é um meio de informar, de forma isolada, a grandeza do risco de algum fator predisponente ao paciente a um avanço mais ou menos severo do tumor, para então sinalizar um enfoque diferenciado no tratamento (SOARES et al., [s. d]).

Os resultados apresentados na figura 11 mostram que para o fator raça, ser branco representa um risco aumentado (1,50 vezes maior de mortalidade que a classe de referência - não-branca). Analogamente, para o fator sexo, os homens possuem 1,37 vezes mais susceptibilidade em relação às mulheres; para o fator etário, ser jovem implica em risco diminuído em 0,52 vezes; analisando-se o peso do risco em tratamento farmacológico, sua ausência acarreta em risco diminuído em 0,5 vezes. A ausência de radioterapia como tratamento confere redução de risco de 0,21 vezes em relação a referência (o emprego da radioterapia).

Figura 11 - Razão de risco para pacientes da coorte TCGA-GBM.



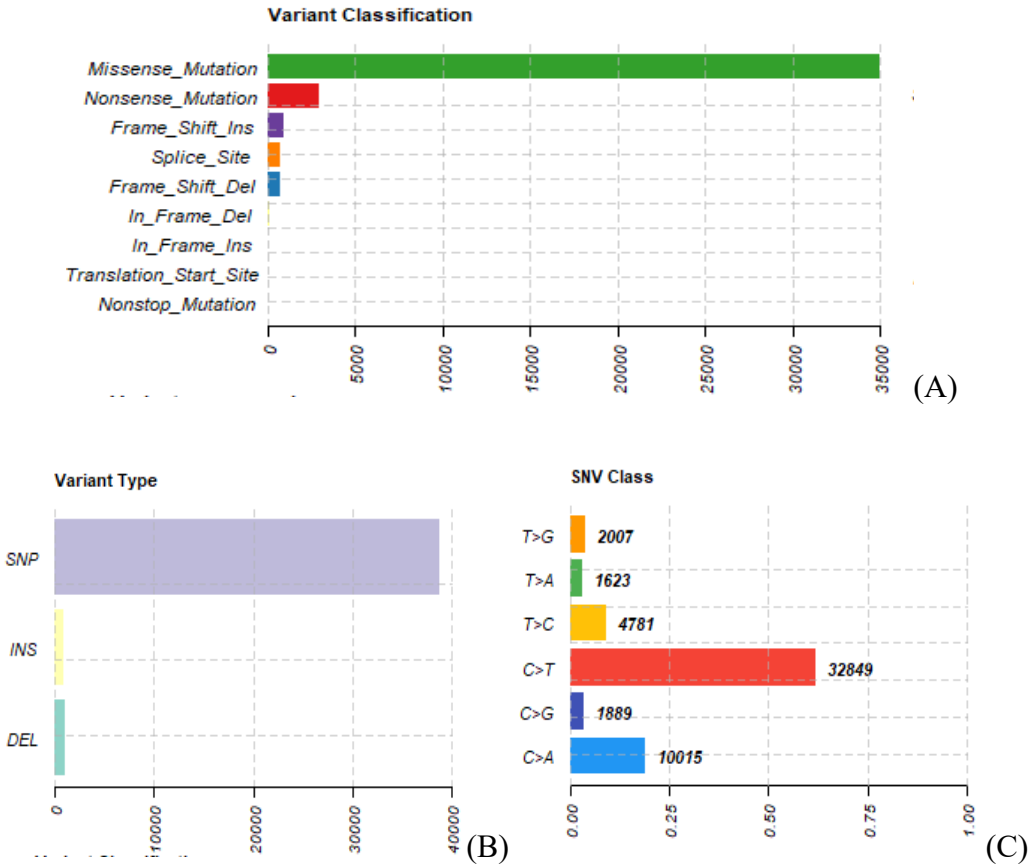
Razão de risco medida em relação à raça (race), sexo (sex), grupo etário (age_group), tratamento farmacêutico (pharmaceutical_treatment), tratamento prévio (prior_treatment), tratamento radioativo (radiation_treatment). Na segunda coluna, estão as classes e o número de observações N. Na terceira coluna, o valor de HR em relação à referência (reference) e o intervalo de confiança. Asteriscos denotam p-valores significativos medidos pelo teste de Log-Rank. Fonte: Autoral.

5.4. Análise Genômica

5.4.1. Alterações gênicas

A fim de avaliar os processos moleculares subjacentes à manifestação do glioblastoma multiforme, procedeu-se à análise das alterações gênicas. Inicialmente, verificou-se os aspectos gerais das variações gênicas - respectivamente, a classificação (Figura 12.A), o tipo (Figura 12.B), as classes de variações de nucleotídeo único (SNV - *Single Nucleotide Variation*) (Figura 12.C). Como esperado, a alteração majoritária foi o de mutações não-sinônimas, liderando a de sentido trocado (*Missense Mutation*). Nos tipos de variações, tem-se que as preponderantes são os polimorfismos de nucleotídeo único (SNP), e em seguida, INDELS (inserção/deleção de pares de nucleotídeos). Das substituições de bases nitrogenadas, as mais expressivas foram as de citosina para timina (transição). Contabilizou-se uma mediana de 44 variações gênicas por amostra.

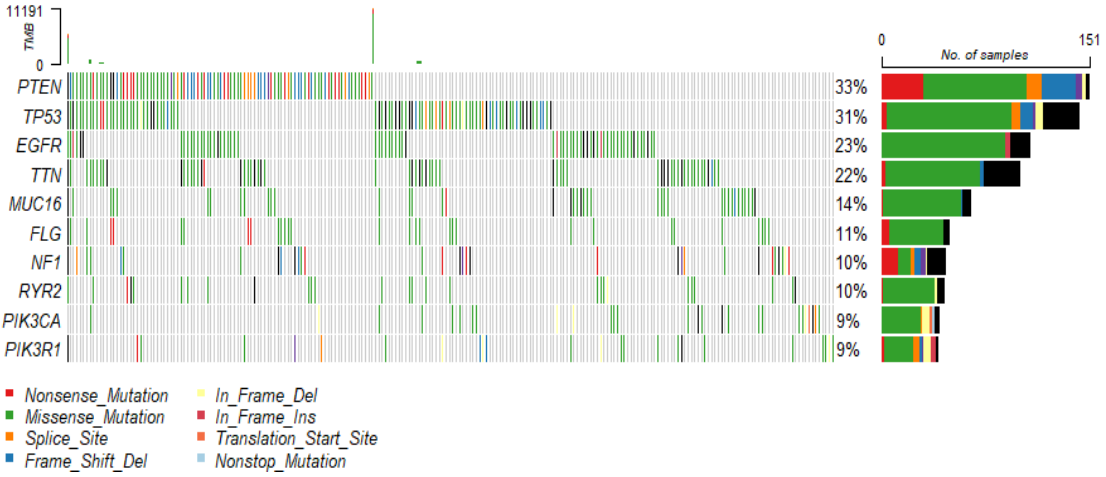
Figura 12 - Panorama geral de alterações gênicas na coorte TCGA-GBM.



Quantidade total de alterações gênicas. (A) Classificação de variações: mutações de sentido trocado (verde), mutações sem sentido (vermelho), inserção dentro da janela de leitura (roxo), alteração do sítio de *splicing* (alaranjado) e deleção alterando a janela de leitura (azul claro). (B) Tipos de variações: polimorfismos de nucleotídeo único (SNP), inserção (INS) e deleção (DEL). (C) Transições e transversões em SNPs. Fonte: Autoral.

O gráfico abaixo, *oncoplot*, dispõe de forma ordenada os genes mais frequentemente mutados e a frequencia de cada tipo de mutação, representada pela barra colorida ao lado direito. Incluindo todos os genes sequenciados, 380 amostras de um total de 461, apresentaram mutações. A seção à direita do *oncoplot* apresenta a quantidade amostral e o tipo de alteração gênica; e a seção superior exibe o número de variações gênicas por amostra. As informações sobre as funções e as doenças associadas aos genes da figura 13 são apresentadas na tabela 2.

Figura 13 - Oncoplot dos 10 genes mais frequentemente mutados.



A secção à direita, a frequência de mutações de cada gene na coorte e a classificação da alteração. Na parte superior, a carga de mutação tumoral (TMB - Tumoral Mutation Burden) em megabases, por amostra. Mutações sem sentido (vermelho), mutações de sentido trocado (verde), alteração do sítio de splicing (alaranjado), deleção alterando a janela de leitura (azul escuro), deleção dentro da janela de leitura (amarelo claro), inserção dentro da janela de leitura (vermelho), alteração do sítio de início da transcrição (alaranjado) e mutação no códon de parada (azul claro). Fonte: Autoral.

Tabela 2 – Processos biológicos relativos aos 10 genes mais mutados (continua).

Gene	Função / Produto	Doenças associadas
TP53	Processos genéticos que levam ao crescimento celular ou apoptose.	Cânceres em geral. Síndrome de Li-Fraumeni. Papiloma do Plexo Coróide.
EGFR	Receptor do fator de crescimento epidérmico; fosforilação de certas proteínas; regulação da migração celular.	Promotor de tumor; câncer de pulmão, de mama (tipo basal). Doença inflamatória da pele e intestino.
TTN	Regulação da contração excêntrica nos músculos esqueléticos. Estruturação cromossômica. Atuação como enzima quinase.	Alteração da mioelasticidade (incluem cardiomiopatias).
MUC16	Produção de muco (barreira epitelial contra antígenos). Marcador de cânceres com alta taxa de prognóstico pobre.	Câncer de pulmão de células pequenas.

Tabela 2 – Processos biológicos relativos aos 10 genes mais mutados (conclusão).

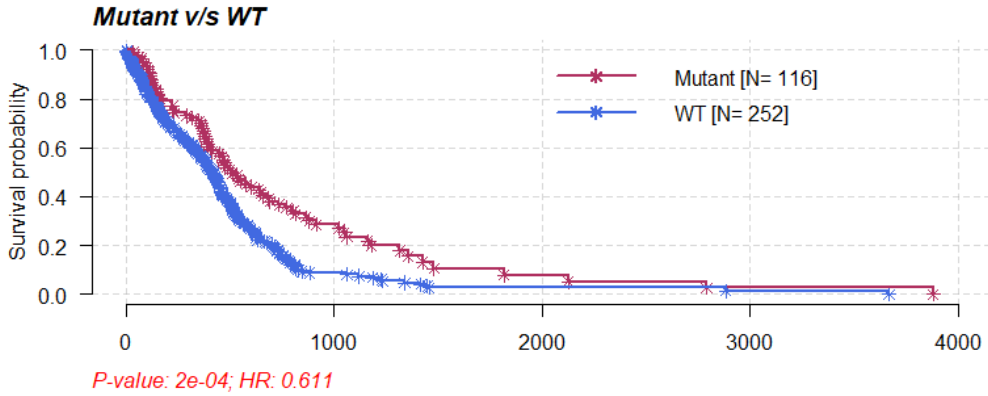
Gene	Função / Produto	Doenças associadas
FLG	Manter coeso os filamentos de queratina da epiderme.	Doenças de pele, como ictiose e dermatite.
PIK3CA	Fosforilar lipídeos e sinalizar vários processos celulares como crescimento, proliferação, fagocitose e pinocitose.	Oncogene - Câncer de mama (comumente o gene mais enriquecido), cânceres cervicais, síndromes envolvendo malformações vasculares, megalencefalia.
PIK3R1	Metabolizar insulina e administrar sobrecargas metabólicas, em especial, no fígado.	Resistência à insulina, Síndrome de Short, agamaglobulinemia.

Fonte: Elaborada pelo autor baseado em Genecards[16] (2023).

5.4.2. Análise de Sobrevivência

Para avaliar o impacto das mutações, traçaram-se as curvas de sobrevivência dos pacientes segundo o estado dos 10 genes mais mutados na coorte. Somente o gene TP53 teve uma influência signifcativa na sobrevivência ($p < 0,05$), a qual foi positiva para os casos em que houve mutação ($HR < 1$) (Figura 14). De fato, este grupo (116 indivíduos) alcançou um tempo maior de sobrevivência em relação ao tipo selvagem (252 indivíduos).

Figura 14 - Curva de Kaplan-Meyer para a sobrevivência dos pacientes, em relação ao status do gene *TP53*.



Curvas de sobrevivência dos pacientes da coorte segundo o estado do gene *TP53*, em selvagem (WT) e mutante (*Mutant*). A diferença foi significativa (p-valor $< 0,05$) e a taxa de risco diminuída ($HR < 1$) para os pacientes com o gene *p53* mutado.

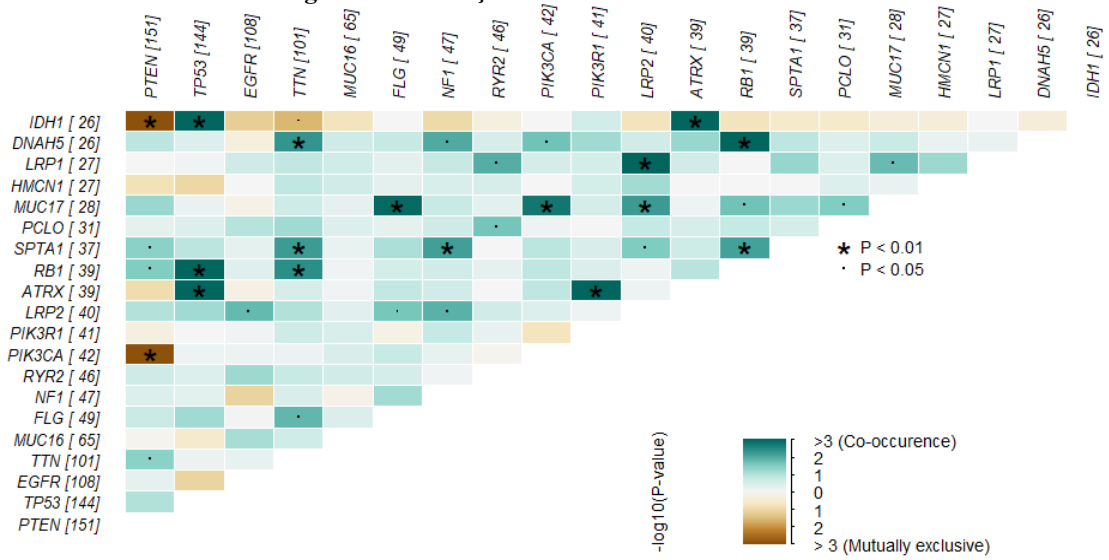
16 Disponível em: <https://www.genecards.org/cgi-bin/carddisp.pl>. Acesso em: 2 dez. 2023.

5.4.3. Análise funcional

A visualização das interações somáticas entre genes torna possível a identificação dos principais genes responsáveis pela ocorrência de ganho funcional ainda que haja algum grau de incerteza. Segundo Mayakonda e colaboradores (2018), tal grupo gênico "pode apresentar rotas metabólicas do tipo *de novo* e mecanismos subjacentes da formação tumoral".

As mutações são distinguidas segundo a relação entre pares de genes, por mutuamente exclusiva e co-ocorrente (Figura 15). A primeira diz respeito a mutações que são observadas em um **ou** outro gene do par. Inversamente, mutações co-ocorrentes são aquelas onde são observadas mutações em **ambos** os genes. Para o caso da mutação mutuamente exclusiva, tem-se como exemplo, na coorte, o gene *IDH1*. Quando este é mutado, não se observa a ocorrência de mutação no gene *PTEN,* com uma diferença bastante significativa ($p < 0,01$). Em contrapartida, se o gene *IDH1* está mutado, o gene *TP53* também apresentará mutação, com alta probabilidade ($p < 0,01$) - mutação co-ocorrente. Como outros resultados marcantes ($\log10(\text{p-valor}) > 3$) tem-se a mutação mutuamente exclusiva entre os genes *PIK3CA* e *PTEN*. Já como mutações co-ocorrentes destacam-se as que contêm pelo menos um gene frequentemente mutado, componente das vias canônicas e com alta significância ($p < 0,01$), como as interações envolvendo *TP53 e IDH1, TTN e RB1, LRP2 e LRP1, PIK3CA e PCLO, PIKR1 e LRP2* (Figura 13).

Figura 15 - Mutações mutuamente exclusivas ou co-ocorrentes.

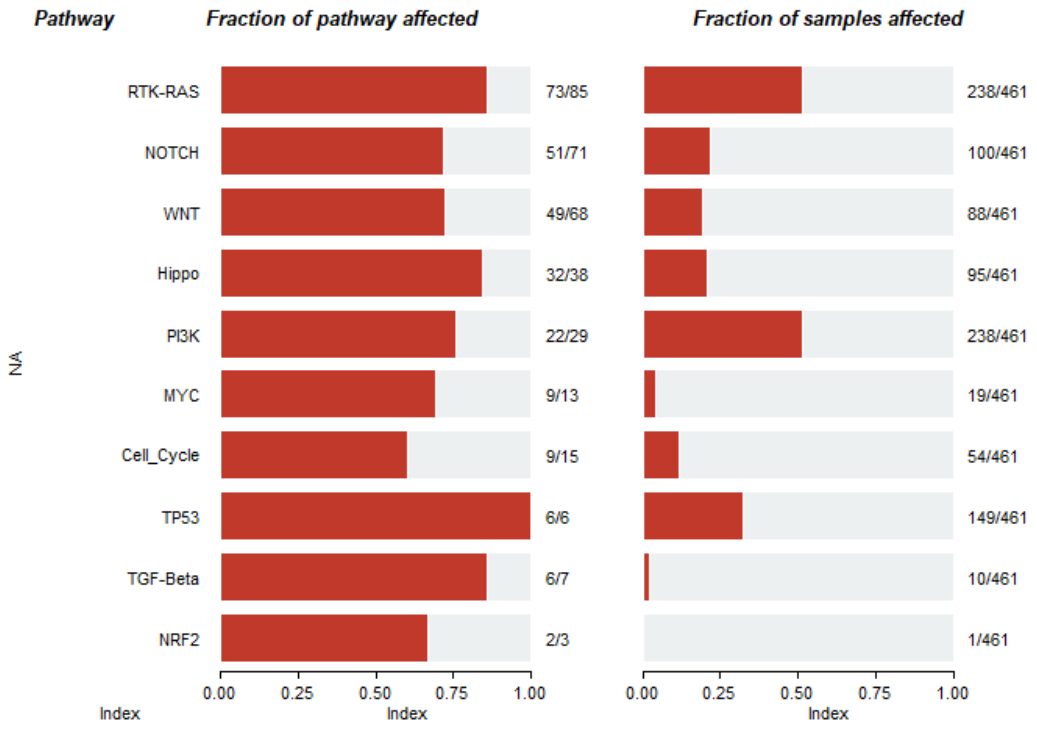


O verde mais escuro indica maior tendência de co-ocorrência enquanto o marrom mais escuro, maior tendência de exclusividade, avaliada pelo teste exato de Fisher. Fonte: Autoral.

Uma outra análise relevante para se entender os mecanismos na gênese tumoral do GBM, é a verificação de enriquecimentos de genes alterados em vias metabólicas associadas ao câncer. A figura 16 exibe as vias que apresentam genes mutados e a fração de amostras afetadas. Observa-se

que todas elas são vias canônicas do câncer (SANCHEZ-VEGA *et al*., 2018), sendo as com maior porcentagem de alteração, TP53 (100%), TGF-Beta (85,7%), RTK-RAS (85,6%) e Hippo (84,2%). Todas elas envolvidas com a regulação da proliferação celular.

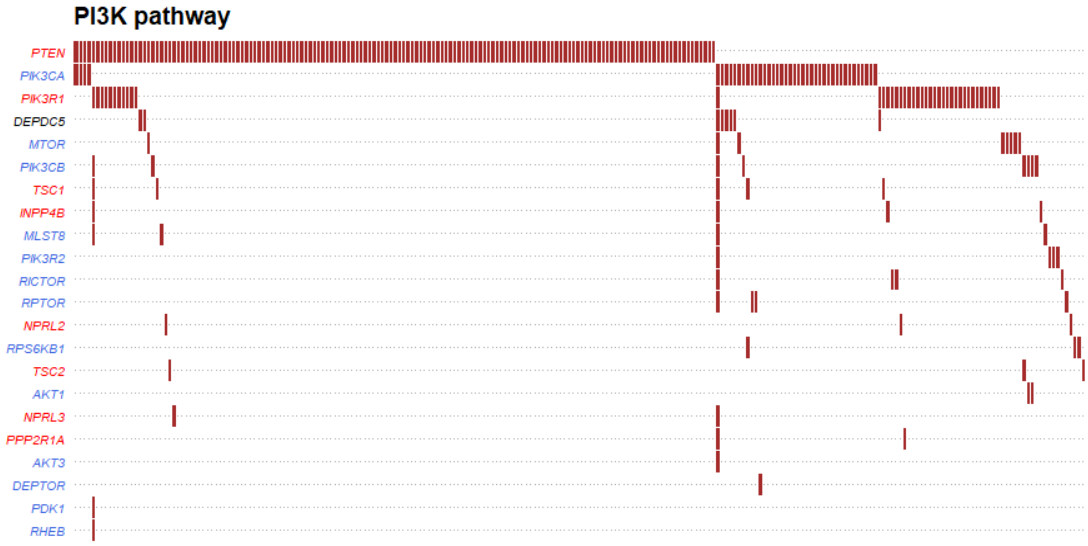
Figura 16 - Vias oncogênicas alteradas e proporção de pacientes afetados.



À esquerda, a proporção de genes alterados em cada via metabólica. À direita, a fração das amostras afetadas correspondente a mesma via. Fonte: Autoral.

A figura 17 apresenta a visualização expandida da quantidade e tipo dos genes envolvidos em cada amostra na via PI3K, a qual está afetada em 75,86% das amostras (Figura 16). Esta via promove o crescimento celular baseando-se na ativação de oncogenes ora pela ativação de outros oncogenes ora pela inibição funcional de supressores de tumor. Os genes estão listados de forma decrescente de quantidade de mutação, sendo o *PTEN* (supressor de tumor) o mais mutado, seguido do *PIK3CA* (supressor de tumor) e *PIK3R1* (oncogene).

Figura 17 - Genes alterados na via metabólica PI3K.

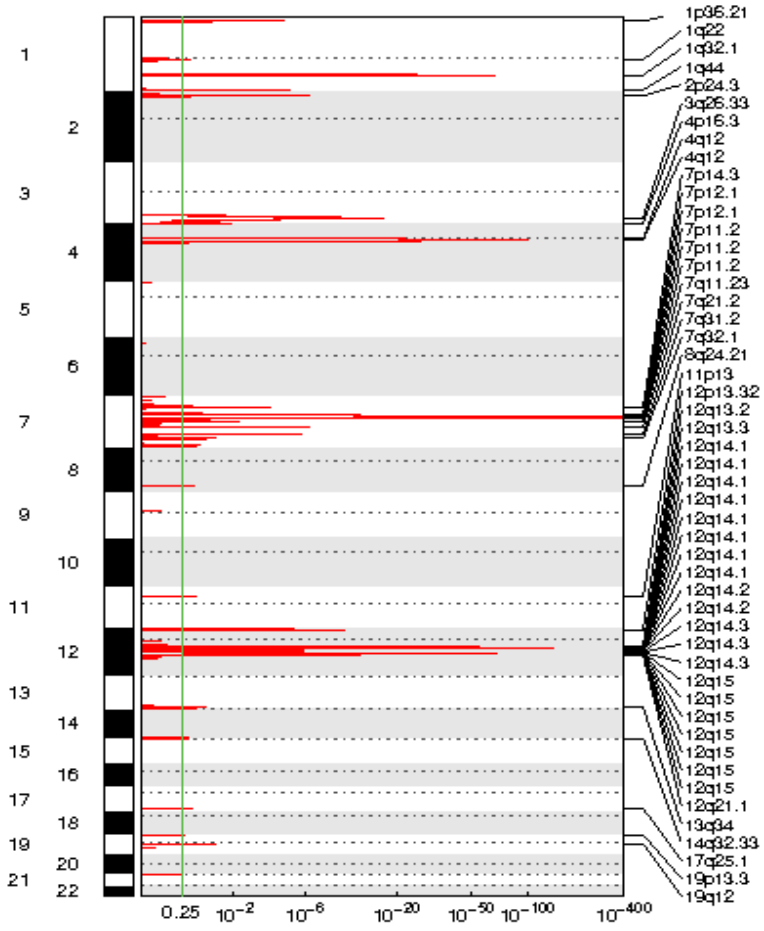


A barra representa um paciente onde foi detectado pelo menos uma mutação no respectivo gene. Oncogenes são grafados em azul e supressores de tumor em vermelho. O gene grafado em preto não se encaixa em nenhuma das categorias anteriores. Fonte: Autoral.

5.4.4. Análise do número de cópias

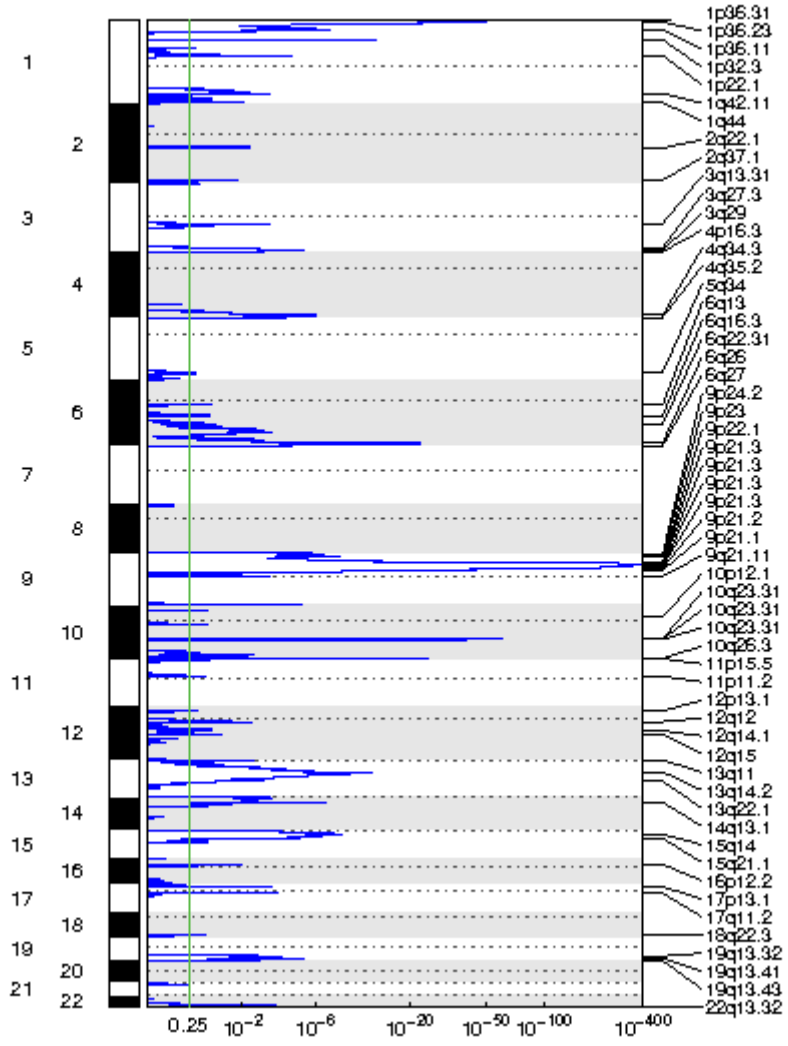
Além das alterações do tipo SNV, avaliou-se também a alteração do número de cópias (CNV) uma vez que a amplificação ou deleção de regiões cromossômicas alteram simultaneamente a atividade de vários genes. Na coorte TCGA-GBM, as *amplificações* (secção 3.3.2) concentraram-se de forma significativa nas citobandas 4q12, 7p12.1, 7p11.2, 12q13.2, 12q13.3 e 12q14.1 (Figura 18). Já as deleções nas regiões 1p36.31, 6q27, 9p22.1, 9p21.3, 10q23.31 (Figura 19). Vale salientar que existem genes sem envolvimento com o glioblastoma nas regiões de altas alterações de número de cópias. Nas tabelas 3 e 4, encontram-se os genes presentes nas amplificações e deleções mencionadas.

Figura 18 - Amplificações nos cromossomos em função do q-valor.



Os cromossomos autossômicos são ordenados verticalmente à esquerda, e à direita encontram-se as citobandas amplificadas. Os q-valores GISTIC são mostrados abaixo em escala log e a linha verde indica o limiar de significância ($q \geq 0{,}25$). Fonte: Autoral.

Figura 19 - Deleções nos cromossomos dos pacientes em função do q-valor.



Os cromossomos autossômicos são ordenados verticalmente à esquerda, e à direita encontram-se as citobandas deletadas. Os q-valores GISTIC são mostrados abaixo em escala log e a linha verde indica o limiar de significância ($q \geq 0{,}25$). Fonte: Autoral.

Tabela 3 – Genes nas citobandas mais amplificadas. (continua)

Citobanda	Gene	Função
4q12	PDGFRA	Promotor de tumor na via RTK-RAS (constituinte do GBM), gene mutado frequentemente na coorte.
7p11.2	EGFR	Promotor de tumor na via RTK-RAS (constituinte do GBM), gene mutado frequentemente na coorte.

Tabela 3 – Genes nas citobandas mais amplificadas. (conclusão)

Citobanda	Gene	Função
12q13.3	MARS G	Promotor de tumor na via RTK-RAS (constituinte do GBM), gene mutado frequentemente na coorte.
	ARHGAP9	Ativação do fator de crescimento neural na via PI3RK (constituinte do GBM).
12q13.2	PDE1B	Participação nos processos cognitivos.
7p12.1	POM121L12	Composição do nucleoporo.
12q14.1	CDK4	Controle de proliferação celular em G1. Presente na via metabólica "Ciclo celular" (constituinte do GBM).
	TSPAN31	Proteína transmembranar.
	MARCH9	Empacotamento de proteínas no endossomo.
	AGAP2	Ativação do fator de crescimento neural na via PI3RK (constituinte do GBM).

Fonte: Elaborada pelo autor a partir de dados do GenePatterns e Genecards (2023)[7].

Tabela 4 – Genes nas citobandas mais depletadas. (continua)

Citobanda	Gene	Função
9p21.3	SLC24A2	Transmissão nervosa na retina.
	CDKN2A	Supressão de tumor na via TP53 e RB1 (vias do GBM)
	CDKN2B	Supressão de tumor na via RB1 (constituinte do GBM).
	MTAP	Metabolização da poliamina.
	IFNE	Inibição de proliferação celular.
	MIR31	Regulação gênica pós-transcricional.

Tabela 4 – Genes nas citobandas mais depletadas. (conclusão)

Citobanda	Gene	Função
1p36.31	RPL22	Composição ribossomal.
	KCNAB2	Determinação de propriedades elétricas de células excitáveis.
	ACOT7	Produção de Acyl-CoA (percussor do acetil-CoA).
	ICMT	Endereçamento de proteínas.
	NPH4	Sinalização celular no sistema nervoso.
	GPR153	Formação de receptor de proteína G, principalmente, no sistema nervoso.
	RNF207	Regulação da atividade nervosa no coração.
10q23.31	PTEN	Supressão de tumor na via PI3K (constituinte do GBM).

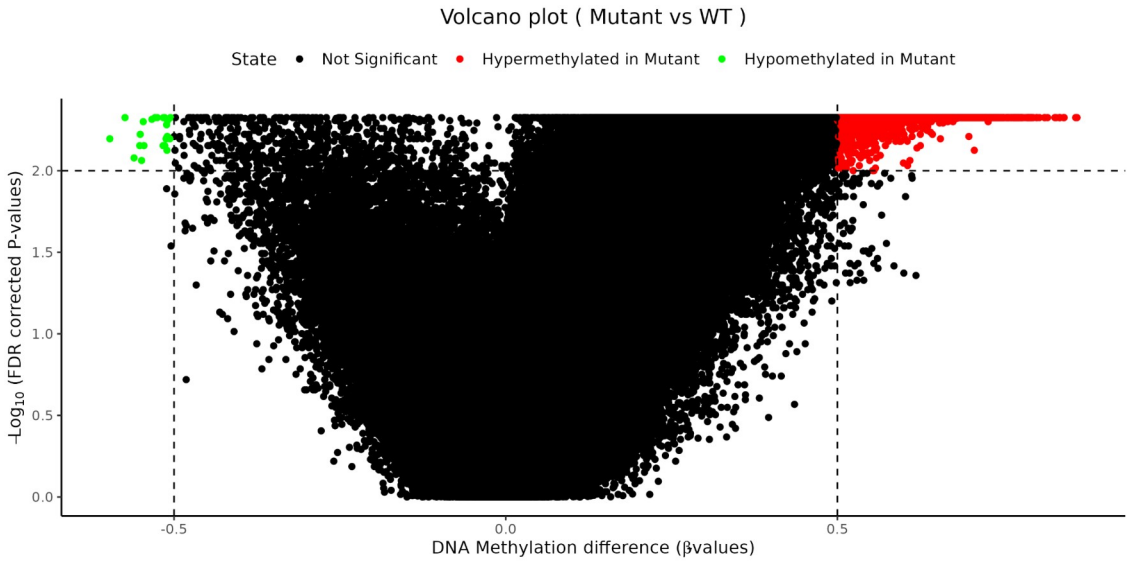
Fonte: Elaborada pelo autor a partir de dados do GenePatterns e Genecard (2023).

5.5. Análise de Metilação

Sabe-se que há subtipos de gliomas oriundos de alterações no gene *IDH,* tais como, o 'L6m6-GBM', 'tipo clássico', 'tipo mesenquimal', que são presentes também no GBM (MAO *et al.*, 2022; CECARELLI *et al.*, 2016). Essa variedade de subtipos se associa a padrões precisos de metilação, os quais por sua vez implicam na qualidade do resultado do paciente (NOUSHMEHR *et al.*, 2010; STURM *et al*., 2012; VERHAAK *et al.*, 2010). A hipermetilação do DNA vista em alguns desses subtipos é o padrão de maior interesse uma vez que está correlacionada a um melhor prognóstico (CECCARELLI *et al*., 2016). Para avaliar o nível de metilação nos sítios CpGs e não CpGs no DNA (seção 3.3.4) dentre os subtipos mencionados, dividiu-se a coorte entre os pacientes que portavam o gene *IDH* no estado mutado ou selvagem. Os resultados mostraram uma diferença significativa ($p < 0,01$) do nível de metilação ($\beta = 50\%$). Nos pacientes que portam o gene *IDH*

mutado, a maioria dos sítios estão hipermetiladas (3765 regiões), apenas 23 estavam na forma hipometilada (Figura 20).

Figura 20 - Volcano plot das regiões metiladas.



As regiões hipermetiladas, em vermelho, e hipometiladas, em verde. A linha horizontal delimita os dados em termos do inverso dos falsos positivos em escala logarítimica (-log *false discovery rate*). As linhas verticais delimitam os níveis de metilação em relação ao controle, medido em um valor β. Fonte: Autoral.

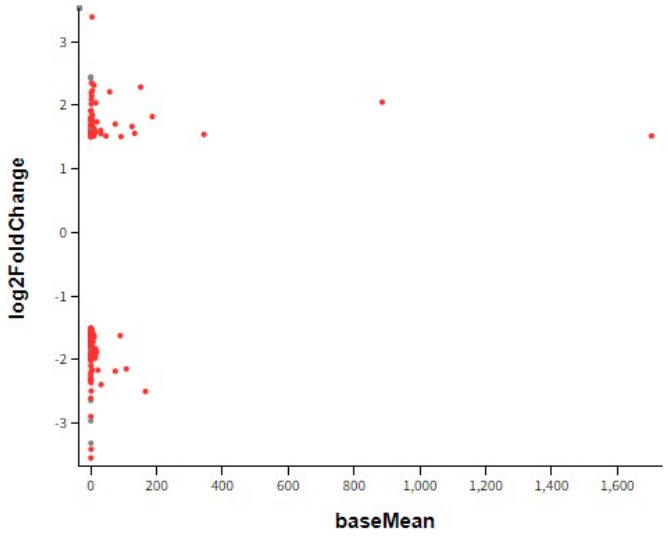
5.6. Análise de Expressão Diferencial

Nesta análise, os genes diferencialmente expressos correspondem aos que apresentam alteração aumentada (*upregulated*) ou alteração diminuída (*downregulated*) em relação ao nível de expressão no grupo controle. A análise de expressão diferencial foi feita sobre os parâmetros - o *status* do gene *IDH* e o *status* vital do paciente. A primeira variável de modelo se relaciona com a qualidade do prognóstico - quando o *IDH* é detectado mutado é um indicador de bom prognóstico, e inversamente quando do tipo selvagem (IDBAIH; MEHDI, [s.d.]). Apesar desta aparente correlação, os padrões de expressão gênica diferiram nitidamente em ambos os designs de estudo.

Tomando-se o *status* vital como referência, aplicando-se um limiar de expressão um aumento de 2,83 vezes em relação ao controle, e um nível de significância de 1% (Figura 21), obtiveram-se 174 genes diferencialmente expressos, sendo o gene com maior nível de expressão o *EPYC,* com expressão aumentada em cerca de 13 vezes em relação ao controle, enquanto o de menor nível de expressão foi o *SFRP5,* com expressão cerca de 7,36 vezes menor em relação ao controle (Figura 22). Os gráficos da figura 22 apresentam a expressão dos genes EPYC e SFRP5 em cada paciente da coorte, subdividindo-os em grupos de pacientes vivos e mortos. É interessante

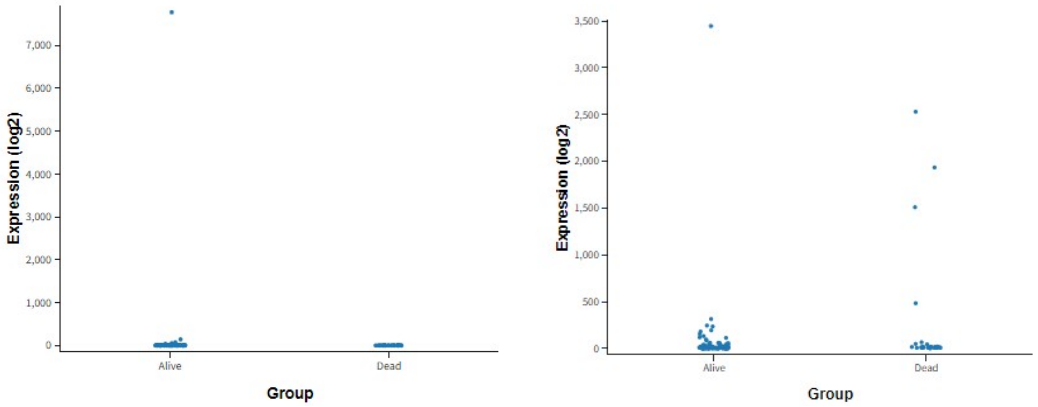
notar que, em termos visuais, a categoria ‘vivo’ suplantou a ‘morto’ em relação ao maior módulo de nível de expressão observado.

Figura 21 - Gráfico MA segundo o status vital.



Na ordenada, a variação da expressão em escala logarítmica (log_2 *Fold Change*). Na abscissa, a média normalizada da quantidade de amostras dos genes. Os genes diferencialmente expressos estão corados em vermelho, sendo o com maior expressão, o *EPYC,* e, com menor, o *SFRP5*. Considerou-se um p-valor ≤ 0,001. Fonte: Autoral.

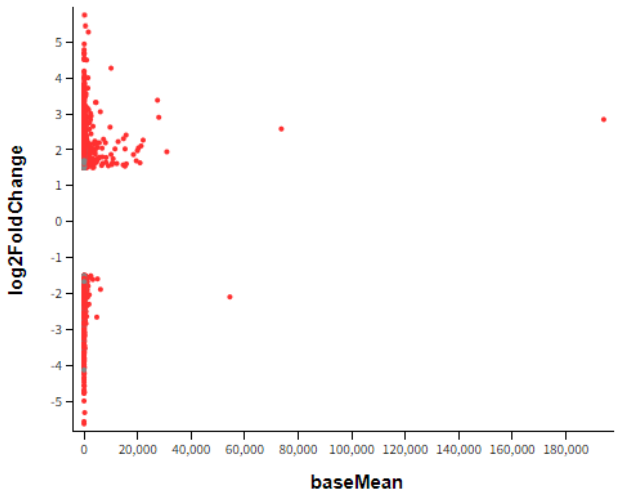
Figura 22 - Níveis de expressão dos genes *EPYC* e *SFRP5*, respectivamente.



Expressão em log 2 dos referidos genes segundo o status vital, vivo (*Alive*) e morto (*Dead)*. Fonte: Autoral.

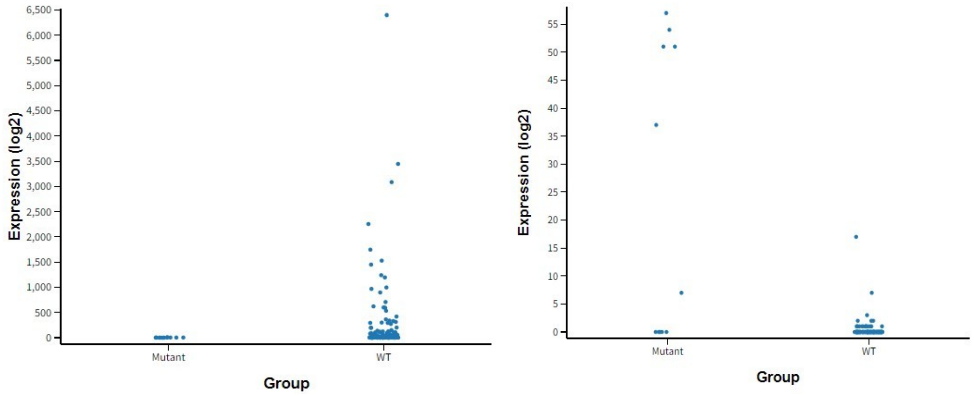
Tomando-se os dados de expressão gênica diferencial segundo o status do gene *IDH* - mutado ou não mutado (WT), e aplicando um valor de significância de 1%, foram obtidos 2481 genes, sendo os de maior e menor níveis de expressão, respectivamente, *C6ORF15* e *SLC25A15P2* (Figura 23). Avaliando o gene mais subexpresso na figura 24 (*SLC25A15P2)*, o grupo mutante foi o que visualmente mais apresentou um maior número de dobras. Apesar disso, os pacientes com a versão selvagem foram numericamente maiores.

Figura 23 - Gráfico MA segundo o status do gene *IDH*.



Na ordenada, a variação da expressão em escala logarítmica (log_2 *Fold Change*). Na abscissa, a média normalizada da quantidade de amostras dos genes. Os genes diferencialmente expressos estão corados em vermelho, sendo o com maior expressão, o *C6ORF15,* e, com menor, o *SLC25A15P2*. Considerou-se um $p \leq 0{,}01$. Fonte: Autoral.

Figura 24 - Níveis de expressão dos genes *C6ORF15* e *SLC25A15P2,* respectivamente.



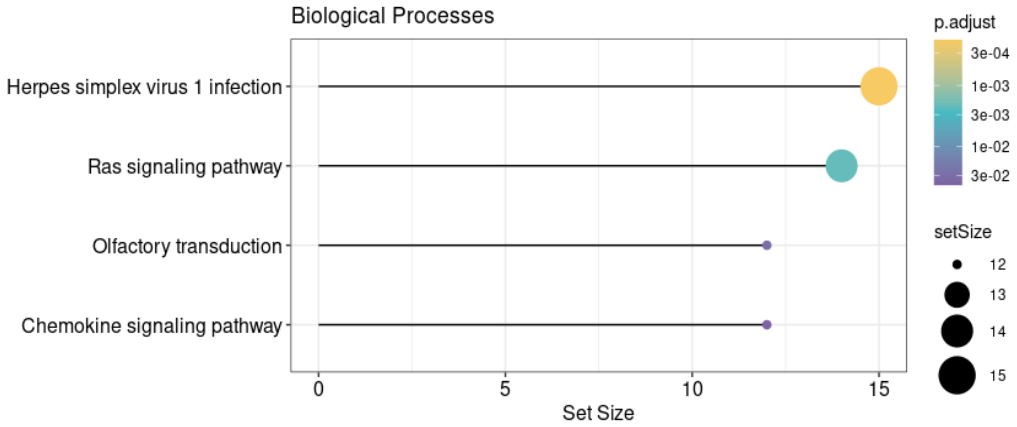
Expressão em log 2 dos referidos genes segundo o status do gene *IDH* (mutante e selvagem - WT). Fonte: Autoral.

5.7. Análise de Enriquecimento Funcional

A partir do conjunto de dados gênicos obtidos na análise GSEA, identificamos os processos biológicos desempenhados por esses genes (YU, 2022), com base no banco de dados *Kyoto Encyclopedia of Genes and Genomes* (KEGG), bastante utilizado para análise funcional devido à sua constante curadoria. Dentre os processos biológicos com maior conjunto de genes e melhores níveis de significância estão: infecção pelo vírus herpes simples 1; sinalização da via Ras; transdução olfativa; sinalização da via de quimiocina (Figura 25). Todos esses eventos se correlacionam ao glioblastoma. Segundo o escore de enriquecimento normalizado (NES) na figura 26, os processos de *sinalização da via Ras* e *sinalização da via de quimiocina* foram positivos, isto é, seu conjunto gênico apresentando expressão aumentada do número de dobras (*log fold-change*).

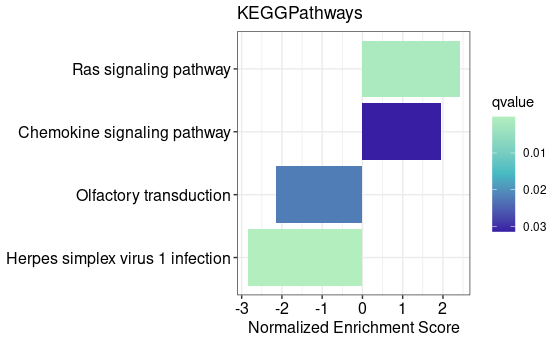
Inversamente ocorreu com escores negativos (KIM, 2016) - os processos de infecção pelo vírus *herpes simples 1* e *transdução olfativa* (Figura 26). Além dessa análise, o banco KEGG forneceu um mapa com as rotas metabólicas envolvidas na manutenção do GBM e sinalizou em cores os genes mais alterados na coorte TCGA-GBM (Figura 27).

Figura 25 - Análise de enriquecimento funcional (processos biológicos) dos genes diferencialmente expressos, utilizando o banco de dados KEGG.



Fonte: Autoral.

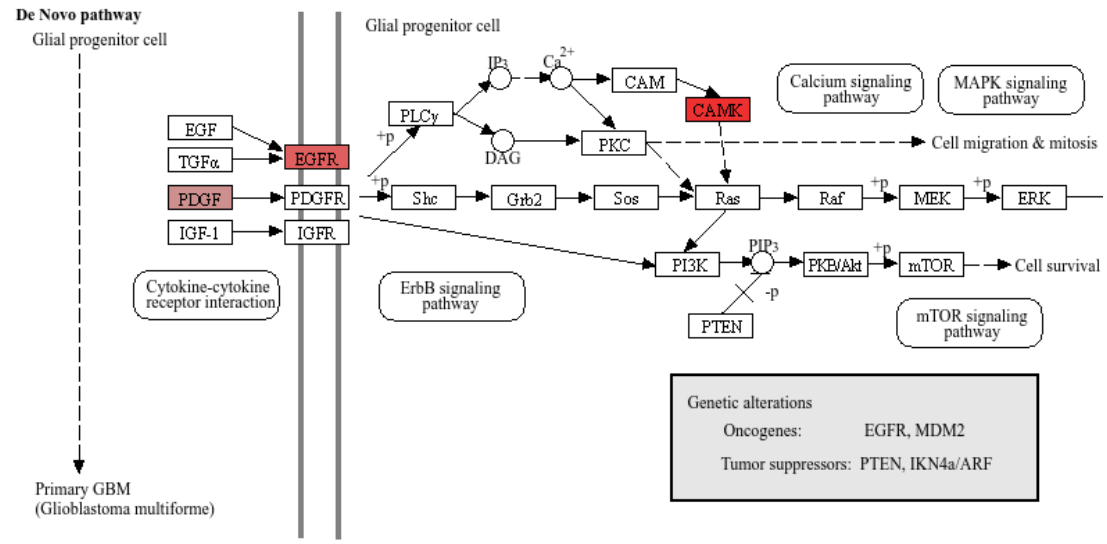
Figura 26 - Processos biológicos segundo NES e com dados do KEGG.



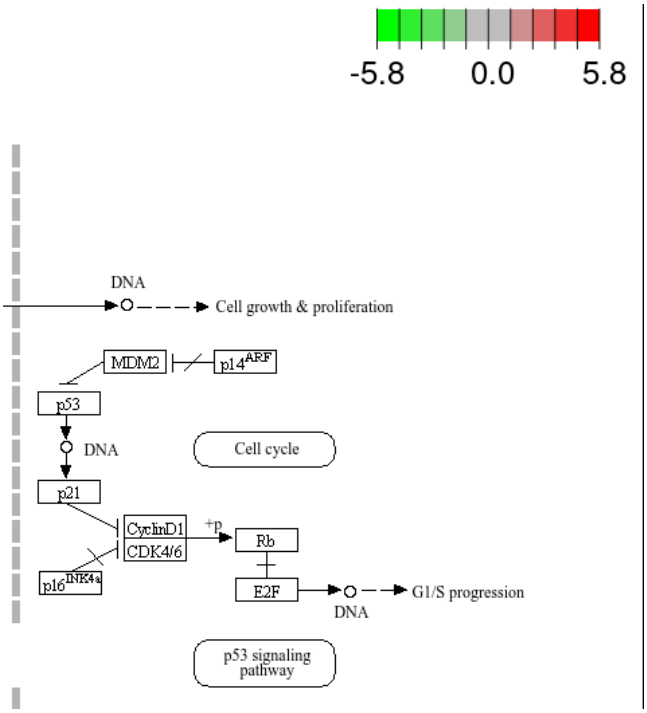
Fonte: Autoral.

Figura 27 - Vias metabólicas no Glioma (KEGG).

A)



B)



A) Vias envolvidas fora da célula, membrana plasmática (linhas paralelas) e citoplasma. B) Vias envolvidas no núcleo celular. Os genes coloridos estão em função do grau de expressão na coorte TCGA-GBM, de acordo com o gradiente no lado superior direito da figura 27.B (vermelho indica superexpressão, e verde, sub-expressão). Fonte: Autoral.

6. DISCUSSÃO

Neste trabalho, efetuou-se a caracterização clínica e molecular do astrocitoma glioblastoma multiforme. Através de análises de bioinformática, pôde-se identificar e avaliar as alterações gênicas relacionadas a SNVs e CNVs, genes envolvidos em vias metabólicas afetadas pelo tumor, curvas de sobrevivência, alterações relacionadas a metilação, e, por fim, a expressão dos genes diferencialmente expressos (DEA) e seu enriquecimento funcional.

6.1. Coorte e dados clínicos

Analisando-se os atributos clínicos da coorte em função do status vital (Tabela 1), tem-se que somente algumas categorias possuem significância estatística ($p < 0,05$), a saber, a idade, raça e tratamento radioativo. A partir de uma meta-análise de dados de sobrevivência de bancos de dados do TCGA, SEER, CNS Disease e MSBase, um estudo concluiu que a idade média de incidência do GBM foi de 65 anos e uma sobrevivência média de 10 meses, quando há o tratamento com radiação, temozolomida e a ressecção tumoral (KIM *et al.*, 2021). Quanto ao fator racial, a raça branca contabilizou 91,5% das mortes por GBM, coadunando assim com diversos dados da literatura que apontam os caucasianos como um fator de risco importante (OSTROM *et al.*, 2018; BOHN *et al*., 2018; GROCHANS *et al*., 2022). Já no tratamento radioativo, apesar de dentre os pacientes mortos, 85,3% deles terem recebido esse tipo de terapia, 96,9% dentre os pacientes vivos também foram tratados com radiação. Desde o início da década de 60, notou-se que a radiação aprimorou o prognóstico dos pacientes de GBM em relação à época em que se realizava apenas a ressecção do tumor. Hoje em dia, o tratamento radioativo aplicado de forma isolada, aumenta a sobrevivência em aproximadamente 1 ano (SULMAN; ISMAILA; CHANG, 2017).

Atendo-se aos fatores que mais se mostraram relevantes para um bom prognóstico, tem-se a idade como líder. Pela taxa de risco, observa-se que, dentre os dados de maior significância, o risco mais reduzido (0,52) é associado a estar na faixa-etária jovem. Ademais, a diferença entre a curva de sobrevivência entre jovens e idosos se mostrou bastante significativa ($p < 0,0001$). De um modo geral, a mediana de sobrevivência sem estratificação da coorte foi em torno de 12 meses, se aproximando à mediana de 10 meses do referido estudo de Kim e colaboradores (2021). Esse parâmetro mostrou ser demasiadamente mais baixo em relação a outros tumores de sistema nervoso central - apenas 6,9% dos pacientes ultrapassaram pelo menos 5 anos após o diagnóstico entre 2001 e 2018 (CBTRUS - *Central Brain Tumor Registry of the United States)*. Ressalta-se que na coorte

TCGA-GBM, houve um grupo seleto de pacientes que sobreviveram também por pelo menos 5 anos após o diagnóstico. Os resultados precedentes são compatíveis qualitativamente com artigos científicos (DOLECEK *et al*., 2013; OSTROM *et al*., 2022; OSTROM *et al*., 2019; OSTROM *et al*., 2014). Entretanto, quantitativamente sofreram alguma variação, a qual pode ser justificada por ser tomada, nesses estudos, o censo populacional de um país ao invés de uma coorte reduzida. Outra razão se deve a diferença de período escolhido nos dados estatísticos - enquanto alguns estudos capturaram como ponto inicial da curva o diagnóstico de GBM entre 2001 e 2018, outros realizaram analisaram entre 2012 e 2016, outros entre 2015 e 2019. Já a coorte TCGA-GBM, alguns pacientes apresentaram sobrevivência capturada no período de 2013 até 2023, e, outra parcela da coorte, de 2008 até 2018. Essa variação mínima da mediana de sobrevivência também foi observado em fontes nacionais como o INCA, levando a crer que os dados clínicos são similares quantitativamente entre os países. As fontes de pesquisa carecem desses números em razão da baixa porcentagem mundial de casos e mortes de câncer no sistema nervoso: 1,6% e 2,5%, respectivamente (SUNG e *et al*., 2023).

Quanto à implicação da mutação de genes específicos na taxa de sobrevivência, sabe-se que uma única mutação não é determinante para a sobrevivência do paciente, embora sirva como indicativo da qualidade do prognóstico, o que é uma correlação esperada em doenças multifatoriais, tal como é o câncer.

6.2. Taxa de risco

Considerando a prevalência de certos grupos de pacientes como representantes de maior risco de manifestação do glioblastoma, todos apresentaram concordância com dados já publicados. Entretanto, há categorias que destoam da proporção observada em outros tumores do SNC. O risco foi aumentado para o sexo masculino em 37% (HR = 1.37), visto na figura 11. Tal medida é oposta à proporção entre homens e mulheres para tumores do SNC em geral, tanto no cenário nacional (INCA, 2022) quanto nos Estados Unidos (SEER, 2023).

Para cada fator adotado nesta análise, possíveis explicações do seu risco são embasadas pela literatura. O grupo etário mais avançado é mais sujeito a desenvolver cânceres devido a um maior tempo do organismo sofrendo divisão celular. Nesse processo, a possibilidade de ocorrer uma mutação gênica *de novo* é natural para todos os indivíduos, sendo eles, portadores ou não de susceptibilidades. Tratando-se sobre o fator racial, houve uma sobrerepresentação dos brancos na coorte, ocasionando em um risco muito maior do que o previsto na literatura. Este fator de risco pode se relacionar com o câncer em questão através de práticas culturais como a endogamia

(MISHRA *et al*., 2022; AMAYIRI *et al*., 2016), a qual pode reforçar o "efeito fundador" em populações isoladas de modo a restringir a variabilidade genética e assim elevar a taxa de risco em certos grupos para cânceres de cérebro ou outros (RUDAN *et al.*, 1999).

Há evidências fortes de que a causa imunológica é o fator interno mais preponderante no desenvolvimento de gliomas malignos, tal como é o GBM. Em gliomas de modo geral, duas causas em particular se mostraram claramente associadas a um melhor prognóstico: o histórico de infecção por varicela-zoster traz uma redução de risco de 20% para manifestar gliomas, principalmente os de maior estadiamento, como o GBM. Apresentar histórico de certas alergias - asma, eczema ou/e alergias respiratórias também representa uma proteção contra gliomas - redução de risco de, respectivamente, 23%, 30% e 30%. Até então, o único fator externo reconhecido como agentes cancerígeno em gliomas são radiações ionizantes (WILD, 2020).

6.3. Análise Genômica

De acordo com a figura 5.A, há uma predominância forte de mutações do tipo *missense* (de sentido trocado) em relação às demais classificações, padrão já descrito na literatura (BRENNAN *et al*., 2013). Devido ao fato de frequentemente mudar a conformação e, portanto, funcionamento da proteína, a mutação *missense* é um fator frequentemente associado ao desenvolvimento de patologias. Já as mutações sem sentido (*nonsense*) que introduzem um *stop-codon*; e inserções ou deleções (INDELS) que alteram a mudança de leitura e de sítio de *splice*, apresentam um impacto claramente deletério. Todos estes fenômenos podem acarretar na susceptibilidade à doenças gênicas.

Em relação aos genes frequentemente mutados, a proporção da classificação da variante, usualmente, se mantém para cada gene. Além disso, como era esperado, as posições elevadas do *ranking* são pertencentes às vias metabólicas mais afetadas, sendo mais nítida a correlação nas vias RTK-RAS, PI3K e TP53 (Figura 6). Esse comportamento se repetiu em outros estudos, contudo, cada trabalho apresentou *ranking* próprio: além de posições diferentes para um mesmo gene, alguns genes eram inéditos (BEROUKIN *et al.*, 2007; BRENNAN *et al.*, 2013; BAILEY *et al*., 2018; FRATINNI *et al*., 2013).

A mutação no gene *TP53* esteve relacionada à uma elevação na sobrevivência (Figura 14). Apesar de contraintuitiva, essa relação é explicada por esta mutação está constantemente acompanhada da mutação do gene *IDH (*LIU *et al.*, 2022*),* o qual é um biomarcador de prognóstico positivo. Entretanto, as mutações de *TP53* que acarretam em ganho funcional são fatores de queda

na sobrevivência por promoverem inflamação. Esta é causada pela infiltração de células do sistema imune que levam ao crescimento tumoral e a resistência aos tratamentos (HAM *et al.*, 2019).

Os cânceres, de modo geral, possuem vias metabólicas afetadas em comum, são as chamadas vias canônicas do câncer. Para sumarizar a análise das alterações nas vias do GBM, o enfoque foi tomado para as vias canônicas com um alto percentual de amostras afetadas dentro de um espaço amostral considerável (RTK-RAS, PI3K, TP53, vide figura 16). Vale salientar que as alterações em componentes de uma via metabólica podem afetar outras, devido a regulações cruzadas, como ocorre entre a rota "ciclo celular" e "p53" (Figura 27).

No segmento RTK da via RTK-RAS, destaca-se a ação do gene *EGFR*, o qual, foi predominante mutado no glioblastoma multiforme (50%) (SANCHEZ-VEGA *et al*., 2018). Como visto na tabela 2, o *EGFR* é um promotor de tumor e está ligado à regulação da migração celular. Tal gene foi o maior alvo de alterações mutuamente exclusivas na via RTK. Os eventos frequentes e co-ocorrentes no gene são de amplificação do número de cópias e fusão. Nesta em particular, a mais frequente no glioblastoma multiforme é a *EGFR* com *SEPT14* (*EGFR-SEPT14*), cujo fenótipo é a independência mitótica, isto é, a proliferação sem haver um sinal externo de estimulação. Tal mecanismo torna desafiador tratamentos farmacêuticos para o GBM. No entanto, outra consequência dessa fusão é o de sensibilizar a célula à inibição do *EGFR*, levando ao retardo de mais formação tumoral. Na prática, apenas algumas amostras exibiram isto *in vitro* ao serem submetidas à ação de dois fármacos, *erptinib* e *lapatinib* (FRATINNI *et al*., 2013).

Com relação à via PI3K, o gene *PIK3CA* foi o maior alvo de ativação, potencializando a sua função sinalizadora de crescimento celular. Por outro lado, *PTEN* foi o alvo de maior inativação, levando por conseguinte, na diminuição do seu papel supressor de tumor (Tabela 2). Estes genes agem antagonicamente na via de sinalização mTOR, ligada à sobrevivência celular.

A via TP53 (também referida como p53) com apenas 6 genes contribui diretamente para o desenvolvimento do GBM quando está inativa, uma vez que, durante seu funcionamento normal a mesma conduz a célula à apoptose (morte), além de regular o reparo de DNA e parada de crescimento (ALBERTS, 2015). O gene *TP53*, fundamental para tais processos (Figura 27), é o mais afetado em sua via metabólica (Figura 17). Pesquisas atuais tentam tratar com fármacos as alterações no *TP53* para atenuar de forma significativa o que tais danos acarretam na via p53 (JIANG et. al, 2022).

A desregulação das vias metabólicas envolvidas na manifestação de glioblastoma multiforme possui maior relação com genes de mutação mutuamente exclusiva (YEANG *et al*., 2008). As vias mais afetadas na coorte TCGA-GBM, em sua maioria, seguem este padrão de ocorrência, tais como, RTK-RAS, TP53, PI3K, embora nesta última tenham sido observadas também alterações co-ocorrentes (SANCHEZ-VEGA *et al*., 2018). Nos resultados deste trabalho, apenas um gene não incluído nas vias canônicas do câncer, o *IDH1*, apresentou uma predominância de interação mutuamente exclusiva com outros genes. O gene *IDH1* codifica a isoforma da enzima isocitrato desidrogenase encontrada tanto no citoplasma como nos peroxissomos. Esta proteína converte NADPH e alfa-cetoglutrato em D-2HG (D-2-hidroxiglutarato) que terá sua produção aumentada no IDH1 mutado. Este oncometabólito D-2HG irá bloquear a participação de alfa-cetogluturato em diversos processos celulares, incluindo a desmetilação, causando então um padrão de hipermetilação. É isto que reduzirá a expressão gênica de oncogenes (IDBAIH; MEHDI, [s.d.]).

Tratando-se do padrão co-ocorrente de interação gênica, foram trazidas nos resultados como relevantes: *TP53* e *IDH1, TTN* e *RB1, LRP2* e *LRP1, PIK3CA* e *PCLO, PIKR1 e LRP2.* Em contraste, a maioria destes estão em vias mutuamente exclusivas (TP53, PI3K e RTK-RAS). Essa relação pode indicar regulação do produto gênico a partir da ação de outro, ou mesmo, a retroalimentação positiva ou negativa entre vias distintas.

Quanto à variação do número de cópias, nas regiões de maior pico nas figuras 18 e 19, encontramos também alguns genes mutados, pertencentes às vias mais afetadas (Figura 16). Nas citobandas com elevadas ocorrências de amplificação (seção 5.4.4), observam-se genes que são concomitantemente os mais afetados nas vias canônicas e estão entre os dez mais mutados na coorte, são eles, *EGFR* (na região 7p11.2), *PDGFRA* (na região 4q12), ambos promotores de tumor e afetados na via RTK-RAS (Tabela 3). Nos genes presentes nas regiões restantes não houve identificação com as vias canônicas nem com alta frequência na coorte (dados não mostados). Todavia, nestes demais genes, foi comum detectar papéis funcionais no neurônio e na regulação das vias canônicas relativas ao GBM, como os genes *ARHGAP9* e *AGAP2* (oncogene) na via PI3K e CDK4 no ciclo celular (Figura 27).

Detendo-se às regiões com os maiores picos de deleções (seção 5.4.4), apenas a região 10q23.31 apresentou um gene *driver* supressor de tumor para o glioblastoma. O gene supressor de tumor *PTEN* possuiu o maior número de amostras mutadas (Figura 6) e é o mais afetado na sua via, PI3KA (Figura 11). Nos demais genes presentes nos maiores picos de deleção, poucos estavam diretamente ligados às vias metabólicas do glioma. Com efeito, se relacionaram a diversos tipos

celulares, ou com supressão de tumores em geral. Em contrapartida, os genes *CDKN2A e CDKN2B,* supressores de tumor, estão contidos em regiões de deleções significativas na porção 9p21.3. Essas alterações acarretam na perda da funcionalidade supressora da "sinalização *Rb",* responsável por manter o DNA na fase G1 do ciclo celular.

Nas regiões amplificadas, há promotores de tumor (oncogenes) e ausência de supressores de tumor (Tabela 3). Ademais, a presença desses últimos tendeu a ocorrer nas regiões de deleção, intensificando ainda mais a redução do efeito benéfico de contenção do câncer. Esse padrão sugere que o processo tumoral se assemelha a um mecanismo de seleção natural, onde a diversidade e adaptações celulares são geradas por meio de mutações aleatórias, e os organismos (células) são selecionados através de pressões ambientais (imunidade, terapias, drogas), fazendo com que apenas os mais aptos sobrevivam (MERLO *et al*., 2006).

6.4. Análise de Metilação

Um biomarcador importante de bom prognóstico no GBM é a mutação no gene *IDH*. Na figura 20, observa-se que a maioria dos genes dos pacientes da coorte estão hipermetilados segundo o padrão de mutação do *IDH*, como observado também em outros estudos (HASSAN *et al*., 2023; YEN *et al*., 2010; WELLER *et al*., 2024). É este fenótipo de maior grau de metilação em ilhas CpGs (CIMP) que contribui para uma maior sobrevivência do paciente. Do ponto de vista biológico, isso é devido, em parte, pela produção aumentada de um oncometabólito produzido pelo *IDH* mutante, o D-2-hidroxiglutarato (D-2-HG). Essa molécula inibe outras envolvidas em processos celulares que manteriam o tumor, tais como angiogênese, maturação de colágeno e regulação epigenética. Essa regulação causa a inibição de enzimas que iriam desmetilar o DNA, diminuindo assim a expressão gênica de oncogenes e consequentemente a proliferação de células cancerígenas. O aumento do tempo de sobrevivência pode ser aproximadamente três vezes maior em relação aos pacientes sem a mutação no *IDH* (LIU *et al*., 2024). Contudo, um outro estudo não apresentou diferença significativa no tempo de sobrevivência de pacientes menores de 65 anos com ou sem a referida mutação (CHRISTIANS *et al*., 2019).

6.5. Análise de Expressão Diferencial

Dentre os parâmetros experimentais disponíveis, optou-se por analisar a variação de expressão gênica em relação ao status do gene *IDH* e o status vital, em razão de ambos se relacionarem à malignidade do tumor. No tocante ao último parâmetro, os genes enriquecidos em pacientes mortos não necessariamente indicam que são causadores dessa condição, é preciso analisar a via metabólica em que estão envolvidos. O status do *IDH* mutado, apesar de estar associado ao aparecimento mais precoce do GBM, é um indicativo de maior tempo de sobrevivência do paciente. Alternativamente, o *IDH* em estado selvagem é tipicamente um marcador de pior prognóstico. Vale salientar que apenas 6% com glioblastoma multiforme primário possuem a vantagem do *IDH* mutado (IDBAIH; MEHDI, [s.d.]).

O conjunto de genes com expressão mais alta (*upregulated*) ou mais reduzida (*downregulated*) diferiram em função dos parâmetros experimentais adotados. Contudo, os genes com expressão mais aumentada em cada parâmetro, *C6ORF15 (*status do *IDH)* e *EPYC (*status vital*)*, são codificadores de proteínas ligadas a componentes que atuam na manutenção e organização da matriz extracelular[7]. Em cânceres, há uma produção diminuída de componentes da matriz extracelular (MEC), o que pode facilitar a migração celular, aumentando assim o potencial metastático (RUOSLAHTI, 1984). Entretanto, os resultados mostraram um aumento na expressão do *C6ORF15* (Figura 25) em pacientes com *IDH*-WT, o grupo com pior prognóstico, necessitando de maior elucidação por outros estudos. Já a expressão de *EPYC* é aumentada em pacientes vivos, o que corrobora com o raciocínio de que uma maior organização MEC está relacionada a um estado, no mínimo, menos agressivo do câncer.

Os genes com menor expressão, *SLC25A15P2 (*status do *IDH) e SFRP5* (*status vital*), desempenham diferentes papéis. O primeiro é um pseudogene que está relacionado ao transporte de mitocôndrias e ornitina (ação no ciclo da uréia)[7]. Como há a diminuição do transporte mitocondrial, haverá uma desestruturação neuronal e perda da distribuição da função dessa organela (SAXTON; HOLLENBECK, 2012). Esse fator causa a ativação de diversas vias retrógradas mitocondriais, isto é, vias que codificam genes no núcleo da célula para produzir proteínas mitocondriais (TORELLI, 2014) responsáveis por promover a malignidade celular (WANG; TSENG; LEE, 2023). O *SFRP5* é um gene codificante que atua no controle de crescimento e diferenciação celular[7]. Devido à sua expressão estar inibida, a consequência será a perda desse controle, levando a multiplicação desenfreada.

6.6. Análise de Enriquecimento Funcional

Após a análise de expressão diferencial, verificou-se as vias onde estes genes estão envolvidos. Dos processos biológicos obtidos pelo banco KEGG na seção 5.7 (Figura 25), todos se correlacionaram ao glioblastoma. A "*infecção por herpes do tipo 1*" se dá no sistema nervoso central. O vírus pode se instalar devido a uma imunodeficiência, a IMD83, causada pela mutação do gene *TLR3*[17]. Contudo, a fragilização dos sistemas imune e nervoso pelo GBM também pode propiciar a doença viral. Já sobre a "*transdução olfativa*", foi mostrado que gliomas surgem preferencialmente no bulbo olfatório. Em estudo com camundongos, o nocaute do gene *Igf1* (*insulin-like growth factor 1*) envolvido na regulação de sistemas olfatórios em neurônios locais, foi capaz de suprimir o desenvolvimento de gliomas em geral. (CHEN *et al.*, 2022). Já o processo biológico, "*via de sinalização de quimiocinas*", está associado à manutenção tumoral de cânceres em geral, uma vez que a produção dessas moléculas pelo próprio tumor ativam diretamente as vias de proliferação, como PI3K e MAPK. Além disso, as quimiocinas nesse estado alterado, inibem a supressão da oncogênese ao prevenir a apoptose celular, e promovem a migração das células cancerígenas para "locais metastáticos". Em cânceres cerebrais, as quimiocinas envolvidas nesses mecanismos são as CXCR4 (receptor de quimiciona tipo 4 CXC) e ACKR3 (receptor 3 de quimiocina atípica, também conhecido como CXCR7). Estudos buscam desenvolver fármacos que inibem os receptores das mesmas, como uma forma de terapia ao câncer (MOLLICA *et al.*, 2019).

O banco de dados KEGG apresenta uma série de rotas metabólicas que podem ser acessadas por funções do pacote *ClusterProfiler*. Desta forma, foi possível obter as vias envolvidas na gênese tumoral do glioblastoma multiforme, bem como o grau de enriquecimento de vias ligadas aos genes diferencialmente expressos (Figura 27). Os principais processos biológicos relacionados à tumorigênese do GBM envolvem as vias metabólicas: ciclo celular, RTK-RAS e PI3K (SANCHEZ-VEGA *et al*., 2018).

Por fim, é importante sintetizar todas estas informações, descrevendo as principais vias envolvidas no GBM. Na figura 27, moléculas presentes no meio extracelular como citocinas e fatores de crescimento se acoplam à receptores próprios na membrana e irão ativar vias metabólicas específicas. Os resultados mostram que o fator de crescimento derivado de plaqueta (PDGF) está com expressão aumentada. Já na membrana, os EGFRs, apresentam um aumento em torno de 13 vezes em relação ao tipo selvagem. A ligação dos EGFRs ativa uma cascata de sinalização intracelular. Ao nível do citoplasma, a fosfolipase C (PLC) ativará um ligante ativado de cálcio (IP3) que por sua vez ativará a quinase dependente de cálcio (CAMK). Esta macromolécula é o

17 Disponível em: <https://omim.org/entry/613002>. Acesso em: 20 jun. 2024.

produto mais enriquecido na coorte, com um aumento de expressão em cerca de 5 vezes em relação ao controle. CAMK junto a outros produtos gênicos serão os estímulos iniciais da via canônica RTK-RAS[7]. Apesar de não termos observado enriquecimentos significativos de outras moléculas nos nossos resultados, no citoplasma também haverá a ativação de vias promotoras de migração celular e mitose, e ativação da via PI3K. Nela, PTEN apresenta perda funcional, causando o aumento da atividade de MTOR, levando assim à sobrevivência da célula. Paralelamente, há o crescimento e proliferação celulares pela ativação consecutiva de genes como *Ras*, *Raf, MEK* e *ERK*. No interior do núcleo o ciclo celular se encontra alterado uma vez que o supressor de tumor $p14^{ARF}$ está com perda desta função, levando ao aumento da expressão do *MDM2*. Sua ativação inibe a função supressora de tumor própria do *TP53.* Alterações no número de cópias, mutações e/ou rearranjos genômicos neste gene levam a parada da replicação celular, e, frequentemente, a apoptose (SANCHEZ-VEGA *et al*., 2018). Neste processo o gene *p21* também promove a supressão do ciclo celular da fase G1 para S ao inibir o complexo de ciclina-CDK, atuando como um freio celular para manutenção em G1. Devido à disfunção de outro supressor tumoral (gene *p16)*, há a desassociação deste complexo que culmina na desativação de outro complexo gênico, o *Rb-E2F* que também atua como mantenedor do DNA na fase G1. Feito isto, o fator de transcrição *E2F,* finalmente permite a continuidade do ciclo celular para a fase S, onde ocorre a replicação do material genético, e com isso a proliferação celular, mantendo o processo tumoral em curso.

7. CONCLUSÕES

Este trabalho apresentou uma caracterização abrangente do glioblastoma multiforme em relação aos fatores de risco clínicos e epidemiológicos, assim como características moleculares associadas ao genoma, epigenoma, expressão diferencial e enriquecimento de vias funcionais associados ao processo tumorigênico, utilizando dados da coorte TCGA-GBM. Como esperado, foram identificadas alterações genômicas importantes de SNV e CNA em genes *drivers e* supressores de tumor, representando serem fortes candidatos para terapias gênicas, ou mesmo, para a supressão dos mecanismos que dificultam o tratamento, tal como demonstrou o projeto brasileiro que silenciou a inibição do *TP53*, removendo completamente a resistência à radioterapia[18]. Por outro lado, necessitam-se de melhores elucidações acerca de quais genes são mais ou menos metilados uma vez que este fenômeno regulatório é o fator genômico mais relevante no prognóstico

18 Disponível em: <https://jornal.usp.br/ciencias/glioblastoma-descoberto-mecanismo-capaz-de-reverter-resistencia-do-tumor-cerebral-a-radioterapia/>. Acesso em: 24 jun. 2024.

do paciente. Além da perspectiva molecular, o conhecimento de padrões clínicos, tal como a existência de fatores de risco, podem favorecer o diagnóstico precoce do câncer, e, trazer um enfoque mais urgente no tratamento de alguns grupos. Em suma, as diferentes frentes de análise e caracterização, efetivam a possibilidade de combate ao glioblastoma por diferentes vias, contribuindo para o aumento do tempo e qualidade de vida do paciente.

REFERÊNCIAS

AKALIN, A. **Computational Genomics with R**. [s.l: s.n.].

ALBERTS, B. **Molecular biology of the cell**. Sixth edition ed. New York, NY: Garland Science, Taylor and Francis Group, 2015.

ALLEN, N. J.; BARRES, B. A. Glia — more than just brain glue. **Nature**, v. 457, n. 7230, p. 675–677, fev. 2009.

AMAYIRI, N. *et al.* High frequency of mismatch repair deficiency among pediatric high grade gliomas in Jordan. International Journal of Cancer, v. 138, n. 2, p. 380–385, 15 jan. 2016.

BAILEY, M. H. *et al*. Comprehensive Characterization of Cancer Driver Genes and Mutations. **Cell**, v. 173, n. 2, p. 371- 385.e18, abr. 2018.

BEROUKHIM, R. *et al*. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. **Proceedings of the National Academy of Sciences**, v. 104, n. 50, p. 20007–20012, 11 dez. 2007

BOHN, A. *et al*. The association between race and survival in glioblastoma patients in the US: A retrospective cohort study. **PLOS ONE**, v. 13, n. 6, p. e0198581, 21 jun. 2018.

BONIFACE, Rachel; GUENOT, Frederique; SAIKALI, Stephan; HAMLAT, Abderrahmane; RIFFAUD, Laurent; MENEI, Philippe. **DNA methylation in glioblastoma: impact on gene expression and clinical outcome**. Bmc Genomics, [S.L.], v. 11, n. 1, dez. 2010. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/1471-2164-11-701. Disponível em: https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-701. Acesso em: 24 set. 2023.

BRENNAN, C. W. *et al*. The Somatic Genomic Landscape of Glioblastoma. **Cell**, v. 155, n. 2, p. 462–477, out. 2013.

Câncer do sistema nervoso central. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/sistema-nervoso-central>. Acesso em: 2 jun. 2024.

Câncer. **Ministério da Saúde**, [s.d]. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer/cancer>. Acesso em: 6 ago. 2024.

CECCARELLI, M. *et al*. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma**. Cell**, v. 164, n. 3, p. 550–563, jan. 2016.

COLAPRICO, Antonio; SILVA, Tiago C.; OLSEN, Catharina; GAROFANO, Luciano; CAVA, Claudia; GAROLINI, Davide; SABEDOT, Thais S.; MALTA, Tathiane M.; PAGNOTTA, Stefano M.; CASTIGLIONI, Isabella. TCGAbiolinks: an r/bioconductor package for integrative analysis of tcga data. **Nucleic Acids Research**, [S.L.], v. 44, n. 8, p. 71-71, 23 dez. 2015. Oxford University Press (OUP). http://dx.doi.org/10.1093/nar/gkv1507.

CHEN, P. et al. Olfactory sensory experience regulates gliomagenesis via neuronal IGF1. **Nature**, v. 606, n. 7914, p. 550–556, 16 jun. 2022.

CHRISTIANS, A. *et al*. The prognostic role of IDH mutations in homogeneously treated patients with anaplastic astrocytomas and glioblastomas. **Acta Neuropathologica Communications**, v. 7, n. 1, p. 156, dez. 2019.

Ch. 11 – Gene Expression. **ABC Science**, 25 set. 2018. Disponível em: <https://astrobiochem.wordpress.com/grade-10-biology/ch-11-gene-expression/>. Acesso em: 2 jun. 2024.

DEUS VIEIRA, G. DE; SOUSA, C. M. DE. Aspectos celulares e fisiológicos da Barreira Hematoencefálica. **Journal of Health & Biological Sciences**, 19 dez. 2013. v. 1, n. 4, p. 166. . Acesso em: 4 nov. 2022.

DOLECEK, T. A. *et al*. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2005-2009. **Neuro-Oncology**, v. 14, n. suppl 5, p. v1–v49, 1 nov. 2012.

ENSEMBL ID to Gene Symbol Converter - Genomics Biotools. Disponível em: <https://www.biotools.fr/human/ensembl_symbol_converter>. Acesso em: 13 maio. 2024.

Erratum. **Neuro-Oncology**, v. 16, n. 5, p. 760–760, 1 maio 2014.

ETCHEVERRY, Amandine; AUBRY, Marc; TAYRAC, Marie de; VAULEON, Elodie; KIM, M. *et al*. Glioblastoma as an age-related neurological disorder in adults. **Neuro-Oncology Advances**, v. 3, n. 1, p. vdab125, 1 jan. 2021.

Estatísticas de câncer. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estatisticas-de-cancer>. Acesso em: 2 jun. 2024.

FERRARA, R. *et al*. Clinical and Translational Implications of RET Rearrangements in Non–Small Cell Lung Cancer. **Journal of Thoracic Oncology**, v. 13, n. 1, p. 27–45, jan. 2018.

FERREIRA, J. C.; PATINO, C. M. What is survival analysis, and when should I use it? **Jornal Brasileiro de Pneumologia**, v. 42, p. 77–77, fev. 2016.

FRATTINI, V. *et al*. The integrated landscape of driver genomic alterations in glioblastoma. **Nature genetics**, v. 45, n. 10, p. 1141–1149, out. 2013.

GENECARDS. Disponível em : <https://www.genecards.org/cgi-bin/carddisp.pl>. Acesso em: 2 dez. 2023.

GROCHANS, S. *et al*. Epidemiology of Glioblastoma Multiforme–Literature Review. **Cancers**, v. 14, n. 10, p. 2412, 13 maio 2022.

HAM, S. W. et al. TP53 gain-of-function mutation promotes inflammation in glioblastoma. **Cell Death & Differentiation**, v. 26, n. 3, p. 409–425, mar. 2019.

HAMAD, D. A. **Copy number variants — Knowledge Hub**. Disponível em: <https://www.genomicseducation.hee.nhs.uk/genotes/knowledge-hub/copy-number-variants/>. Acesso em: 19 ago. 2024.

HASSAN, U. *et al*. Gemistocytic Differentiation in Isocitrate Dehydrogenase Mutant Astrocytomas: A Histopathological and Survival Analysis. **Cureus**, 13 abr. 2023.

HOPKINS, A. L.; GROOM, C. R. The druggable genome. **Nature Reviews** Drug Discovery, v. 1, n. 9, p. 727–730, set. 2002.

HUTTER, C.; ZENKLUSEN, J. C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. **Cell**, v. 173, n. 2, p. 283–285, abr. 2018.

IDBAIH, A.; MEHDI, T. **Factsheets on Biomarkers | OncologyPRO**. Disponível em: <https://oncologypro.esmo.org/education-library/factsheets-on-biomarkers/idh1-2-mutations-in-glioma>. Acesso em: 23 abr. 2024.

INSTITUTO NACIONAL DO CÂNCER - INCA. **Atlas On-line de Mortalidade**, 2022. Disponível em:<https://www.inca.gov.br/MortalidadeWeb/pages/Modelo02/consultar.xhtml#panelResultado>. Acesso em: 8 nov. 2022.

INSTITUTO NACIONAL DO CÂNCER -INCA. **Como surge o câncer?**. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/como-surge-o-cancer/como-surge-o-cancer>. Acesso em: 4 nov. 2022.

INSTITUTO NACIONAL DO CÂNCER -INCA. **Tratamento do câncer**. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/tratamento/tratamento-do-cancer>. Acesso em: 4 nov. 2022.

INSTITUTO NACIONAL DO CÂNCER - INCA. **Cuidados Paliativos**. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/tratamento/cuidados-paliativos/cuidados-paliativos>. Acesso em: 4 nov. 2022.

JIANG, J. *et al*. Systematic illumination of druggable genes in cancer genomes. **Cell Reports**, v. 38, n. 8, p. 110400, fev. 2022.

KEGG PATHWAY: Glioma - Homo sapiens (human). Disponível em: <https://www.genome.jp/pathway/hsa05214>. Acesso em: 2 jun. 2024.

KIERSZENBAUM, Abraham L. **Histologia e biologia celular: uma introdução à patologia**; [revisão científica Marcelo Sampaio Narciso; tradução Adriana Paulino do Nascimento et al - Rio de Janeiro: EIsevier, 2008.

KIM, K. *et al*. PIF1 Regulates Plastid Development by Repressing Photosynthetic Genes in the Endodermis. **Molecular Plant**, v. 9, n. 10, p. 1415–1427, out. 2016.

LIU, Y. *et al*. IDH-mutant grade 4 astrocytoma: a comparison integrating the clinical, pathological, and survival features between primary and secondary patients. **Journal of Neurosurgery**, v. 140, n. 1, p. 94–103, 1 jan. 2024.

LIU, H.-Q. et al. Integrated analysis of the genomic and transcriptional profile of gliomas with isocitrate dehydrogenase-1 and tumor protein 53 mutations. **International Journal of Immunopathology and Pharmacology**, v. 36, p. 039463202211392, jan. 2022.

LOVE, Michael I; HUBER, Wolfgang; ANDERS, Simon. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, [S.L.], v. 15, n. 12, p. 0-550, dez. 2014. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/s13059-014-0550-8.

maftools : Summarize, Analyze and Visualize MAF Files. Disponível em: <https://www.bioconductor.org/packages/devel/bioc/vignettes/maftools/inst/doc/maftools.html>. Acesso em: 2 jun. 2024.

MANUAL MSD**. Astrócitos,** junho. 2021. Disponível em: <https://www.msdmanuals.com/pt-br/casa/problemas-de-sa%C3%BAde-infantil/c%C3%A2nceres-na-inf%C3%A2ncia/astrocitomas>. Acesso em: 8 nov. 2022.

MAO, X. *et al*. Deep learning identified glioblastoma subtypes based on internal genomic expression ranks. **BMC Cancer**, v. 22, n. 1, p. 86, dez. 2022.

MASCARO, B. Epigenética: O que é e como ocorre a regulação. **Varsomics**, 30 jul. 2020. Disponível em: <https://blog.varsomics.com/epigenetica/>. Acesso em: 16 nov. 2023.

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.

MEDEIROS, Baptista de Oliveira. Propriedades mecânicas musculares: compreendendo o papel da proteína titina na produção de força ativa no músculo esquelético, 2021. **Repositório Institucional UFSC**. Disponível em: <https://repositorio.ufsc.br/handle/123456789/225973>. Acesso em: 23 abr. 2024.

MERLO, L. M. F. *et al.*. Cancer as an evolutionary and ecological process. **Nature Reviews Cancer**, v. 6, n. 12, p. 924–935, dez. 2006.

MIOT, H. A. Análise de sobrevivência em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, v. 16, p. 267–269, dez. 2017.

MORGAN, Martin; OBENCHAIN, Valerie; HESTER, Jim; PAGÈS, Hervé (2023). ***SummarizedExperiment: SummarizedExperiment container*. R package version 1.30.2**. Disponível em: <https://bioconductor.org/packages/SummarizedExperiment.>

MISHRA, A. *et al*. Germline Biallelic Mismatch Repair Deficiency in Childhood Glioblastoma and Implications for Clinical Management. **Neurology India**, v. 70, n. 2, p. 772, 2022.

MOLLICA, V. P. *et al*. Chemokines and Chemokine Receptors: New Targets for Cancer Immunotherapy. **Frontiers in Immunology**, v. 10, p. 379, 6 mar. 2019.

NATIONAL CANCER INSTITUTE. **Adult Central Nervous System Tumors Treatment (PDQ®)–Patient Version**. Disponível em: <https://www.cancer.gov/types/brain/patient/adult-brain-treatment-pdq>. Acesso em: 7 nov. 2022.

NATIONAL CANCER INSTITUTE. **What Is Cancer?**, 17 set. 2007. Disponível em: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>. Acesso em: 4 nov. 2022.

NATIONAL CANCER INSTITUTE. **Cancer Staging**, 2022. Disponível em: <https://www.cancer.gov/about-cancer/diagnosis-staging/staging>. Acesso em: 4 nov. 2022.

NATIONAL CANCER INSTITUTE. **Tumor Grade**, 9 maio. 2013. Disponível em: <https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-grade>. Acesso em: 4 nov. 2022.

NATIONAL CANCER INSTITUTE. **TCGA-GBM**. Disponível em: <https://portal.gdc.cancer.gov/projects/TCGA-GBM>. Acesso em: 30 ago. 2022.

NCI Dictionaries - NCI. cgvMiniLanding. Disponível em: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/hazard-ratio>. Acesso em: 2 jun. 2024.

OSTROM, Q. T. *et al*. Adult Glioma Incidence and Survival by Race or Ethnicity in the United States From 2000 to 2014. **JAMA Oncology**, v. 4, n. 9, p. 1254, 1 set. 2018.

OSTROM, Q. T. *et al*. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016. **Neuro-Oncology**, v. 21, n. Supplement_5, p. v1–v100, 1 nov. 2019.

OSTROM, Q. T. *et al*. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019. **Neuro-Oncology**, v. 24, n. Supplement_5, p. v1–v95, 5 out. 2022.

PICO, A. *et al*. Glioblastoma signaling, 22 maio 2024. Disponível: <https://www.wikipathways.org/pathways/WP2261.html>. Acesso em: 1 dez. 2023.

PÖS, O. et al. DNA copy number variation: Main characteristics, evolutionary significance,

and pathological aspects. **Biomedical Journal**, v. 44, n. 5, p. 548–559, out. 2021.

P. WILD, C.; STEWART W., B. World Cancer Report - Cancer Research for cancer prevention. **World Health Organization**, 2020.

ROSER, Max; RITCHIE, Hannah. **Cancer**, 2019. Disponível em: https://ourworldindata.org/cancer. Acesso em: 8 nov. 2022.

ROSS, Michael H. *et al*. **Supporting cells of the nervous system: The Neuroglia. *In*: HISTOLOGY: A Text and Atlas**. Sixth Edition. ed. U. S: Wolters Kluwer, 2011. cap. Nerve Tissue, ISBN 978-0-7817-7200-6.

RUDAN, I. Inbreeding and Cancer Incidence in Human Isolates. **Human Biology**, v. 71, n. 2, p. 173–187, 1999.

RUOSLAHTI, E. Fibronectin in cell adhesion and invasion. **Cancer and Metastasis Review**, v. 3, n. 1, p. 43–51, 1984.

SANCHEZ-VEGA, F. *et al*. Oncogenic Signaling Pathways in The Cancer Genome Atlas. **Cell**, [S.L.], v. 173, n. 2, p. 321-337, abr. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.cell.2018.03.035.

SAXTON, W. M.; HOLLENBECK, P. J. The axonal transport of mitochondria. **Journal of Cell Science**, p. jcs.053850, 1 jan. 2012.

SOARES, A. N. *et al*. Estudos sobre Experimento Clínico. **Sociedade Brasileira de Oncologia Clínica**. Disponível em: <https://www.sboc.org.br/app/webroot/leitura-critica/LEITURA-CRITICA_C9.pdf>. Acesso em: 2 ago. 2024.

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

SEER. **Glioblastoma of the Brain and ONS Recent Trends in SEER Age-Adjusted Incidence Rates, 2000-2021**, 19 abr. 2023. Disponível em: <https://seer.cancer.gov/statistics-network/explorer/application.html?site=661&data_type=1&graph_type=2&compareBy=sex&chk_sex_3=3&chk_sex_2=2&hdn_rate_type=1&race=1&age_range=1&stage=101&advopt_precision=1&advopt_show_ci=on&hdn_view=0&advopt_show_apc=on&advopt_display=1#resultsRegion0>. Acesso em: 2 ago. 2024.

SIDAWAY, P. Glioblastoma subtypes revisited. **Nature Reviews Clinical Oncology**, v. 14, n. 10, p. 587–587, out. 2017.

SULMAN, E. P.; ISMAILA, N.; CHANG, S. M. Radiation Therapy for Glioblastoma: American Society of Clinical Oncology Clinical Practice Guideline Endorsement of the American Society for Radiation Oncology Guideline. **Journal of Oncology Practice**, v. 13, n. 2, p. 123–127, fev. 2017.

SUNG, H. *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA: A Cancer Journal for Clinicians**, v. 71, n. 3, p. 209–249, maio 2021.

Second Cancers Related to Treatment. Disponível em: <https://www.cancer.org/cancer/survivorship/long-term-health-concerns/second-cancers-in-adults/treatment-risks.html>. Acesso em: 2 jun. 2024.

TCGA Molecular Characterization Platforms - NCI. cgvArticle. Disponível em: <https://www.cancer.gov/ccg/research/genome-sequencing/tcga/using-tcga-data/technology>. Acesso em: 29 maio. 2024.

TORELLI, N. Q. **Sinalização retrógrada RTG-dependente controla a atividade mitocondrial e resistência a estresse em Saccharomyces cerevisiae**. Mestrado em Bioquímica— São Paulo: Universidade de São Paulo, 12 dez. 2014.

t-Test, Chi-Square, ANOVA, Regression, Correlation... Disponível em: <https://datatab.net/tutorial/log-rank-test>. Acesso em: 2 jun. 2024.

WANG, E. Understanding genomic alterations in cancer genomes using an integrative network approach. **Cancer Letters**, v. 340, n. 2, p. 261–269, nov. 2013.

WANG, S.-F.; TSENG, L.-M.; LEE, H.-C. Role of mitochondrial alterations in human cancer progression and cancer immunity. **Journal of Biomedical Science**, v. 30, n. 1, p. 61, 31 jul. 2023.

What is the difference between Over Representation Analysis and gene set enrichment analysis? | 3 Answers from Research papers. Disponível em: <https://typeset.io/questions/what-is-the-difference-between-over-representation-analysis-hkdkl28zkt>. Acesso em: 2 jun. 2024.

WELLER, M. *et al*. Improved prognostic stratification of patients with isocitrate dehydrogenase-mutant astrocytoma. **Acta Neuropathologica**, v. 147, n. 1, p. 11, jun. 2024.

WIKIPÉDIA. **Retinoblastoma protein**, 29 mar. 2024. (Nota técnica). Disponível em: <https://en.wikipedia.org/wiki/Retinoblastoma_protein>. Acesso em: 23 maio 2024.

WIKIPÉDIA. **Teste exato de fisher**, 20 maio 2024. (Nota técnica). Disponível em: <https://pt.wikipedia.org/wiki/Teste_exato_de_Fisher>. Acesso em: 10 nov. 2022.

WILD, C.P; WEIDERPASS E.; STEWART B.W. World Cancer Report: Cancer Research for Cancer Prevention. Lyon, França: **International Agency for Research on Cancer**, 2020. Disponível em: <http://publications.iarc.fr/586>. Licença: CC BY-NC-ND 3.0 IGO.

YEN, K. E. *et al*. Cancer-associated IDH mutations: biomarker and therapeutic opportunities. **Oncogene**, v. 29, n. 49, p. 6409–6417, 9 dez. 2010.

YU, G. **Introduction | Biomedical Knowledge Mining using GOSemSim and clusterProfiler**, 22 abr. 2022. Disponível em: <https://yulab-smu.top/biomedical-knowledge-mining-book>. Acesso em: 15 abr. 2024.