

ANÁLISE DE MUTAÇÕES E EXPRESSÃO DIFERENCIAL DE GENES EM PACIENTES COM CÂNCER DE MAMA

¹ Thyanne Thysysanne de Souza Soares Costa; ² Jackson Mateus Alencar de Oliveira; ³ Taffarel Melo Torres.

¹ Graduanda em Biotecnologia pela Universidade Federal Rural do Semi-Árido – UFERSA
² Pós-graduando em m Bioquímica e Biologia molecular pela Universidade do Estado do Rio Grande do Norte - UERN; ³ Biólogo, Docente da Universidade Federal Rural do Semi-Árido – UFERSA;

Área temática: Biotecnologia e Inovação em Saúde

Modalidade: Comunicação Oral Presencial

E-mail dos autores: oncolthyanne@gmail.com¹ ; jmateusalencar@hotmail.com²;
taffareltorres@ufersa.edu.br³

RESUMO

INTRODUÇÃO: O câncer, a segunda principal causa de mortes no Brasil, inclui mais de 100 doenças com crescimento celular desordenado e potencial para metástases. Entre 2023 e 2025, são esperados mais de 704 mil novos casos, sendo o câncer de mama o mais frequente entre as mulheres. **OBJETIVO:** Analisar mutações e processos biológicos com genes diferencialmente expressos em pacientes com câncer de mama. **MÉTODOS:** Foram utilizados dados do projeto TCGA-BRCA do banco de dados do TCGA-GDC. A análise incluiu a quantificação das variações somáticas, metilação, expressão diferencial de genes e análise da sobrevida para pacientes com PGR negativo e estadiamento nível 3. **RESULTADOS:** A análise revelou elevada frequência de mutações nos genes "STAR", "APOC3", "GC", "APOA2", "NP1H4" para a via de ligação a esteroides. Além disso, a análise de expressão gênica identificou um grupo de genes diferentemente expresso para o processo de ligação a esteroides. Neste agrupamento, os genes STAR e GC possuem correlação estatística, com a sobrevida das pacientes que possuem característica PGR negativo (Progesterone Receptor) e estadiamento nível 3. **DISCUSSÃO:** O câncer de mama, como evidenciado, continua a ser um desafio significativo na saúde pública, e a compreensão das variações genéticas e expressivas é crucial para aprimorar as abordagens terapêuticas e diagnósticas. **CONCLUSÃO:** A análise genômica e transcriptômica do câncer de mama com dados do TCGA e ferramentas bioinformáticas forneceu valiosos insights sobre mutações somáticas e expressão gênica.

Palavras-chave: (Banco de dados), (Oncologia), (Bioinformática).

1 INTRODUÇÃO

O câncer é uma das doenças que mais causam mortes no mundo, sendo a segunda principal causa de óbitos no Brasil. O termo "câncer" refere-se a um conglomerado de doenças que têm em

comum o crescimento desordenado de células mutadas. Essas células podem, com o tempo, desenvolver metástases e invadir tecidos adjacentes. Segundo dados do Instituto Nacional do Câncer (INCA), são esperados mais de 704 mil novos casos de câncer no Brasil entre 2023 e 2025. Os tipos mais comuns incluem o câncer de mama e o câncer de próstata. Especificamente, o câncer de mama é a neoplasia mais frequente entre as mulheres no Brasil. Para o período de 2023 a 2025, estimam-se 73.610 novos casos de câncer de mama, o que representa um risco de 66,54 casos para cada 100 mil mulheres (Instituto Nacional do Câncer, 2023).

Os bancos de dados desempenham um papel fundamental nos avanços das pesquisas oncológicas no Brasil e no mundo. No Brasil, destaca-se o Integrador INCA - Registros Hospitalares de Câncer, que coleta e armazena dados de pacientes de diversos hospitais. Internacionalmente, ganham destaque o cBioPortal e o The Cancer Genome Atlas (TCGA), que além de dados clínicos, armazena informações genômicas, histológicas, RNA-seq, entre outras. Essa estrutura é amplamente utilizada em estudos de bioinformática para a descoberta de novas interpretações associadas ao comportamento das neoplasias.

2 MÉTODO

Os dados do TCGA sobre o projeto TCGA-BRCA para câncer de mama (biblioteca TCGA-GDC), que inclui amostras de tecido tumoral, foram recuperados com o auxílio das bibliotecas TCGA-Biolinks e TCGA-workflow via API. Utilizaram-se ferramentas como pacotes do Bioconductor, além das bibliotecas TCGA-Biolinks e TCGA-workflow, para análise (Silva, 2016).

Outras ferramentas, como RTCGA, MAFTOOLS, OPENXLSX e PLYSA, também foram empregadas para análise adicional. Em seguida, foram realizadas análises para identificar mutações (com maftools) e expressão diferencial de genes (com DESeq2), seguidas do enriquecimento de processos e vias (com DOSE e clusterProfiler), utilizando as bibliotecas já mencionadas. Além disso, foi analisada a sobrevida das pacientes com PGR negativo (Progesterone Receptor) e estadiamento nível 3 (câncer de mama), filtrando para 226 pacientes, focando na via de ligação a esteroides e identificando genes relacionados à sobrevida desses pacientes. As análises foram feitas com o auxílio da linguagem de programação R e utilizando a *Kaplan-Meier Plotter*.

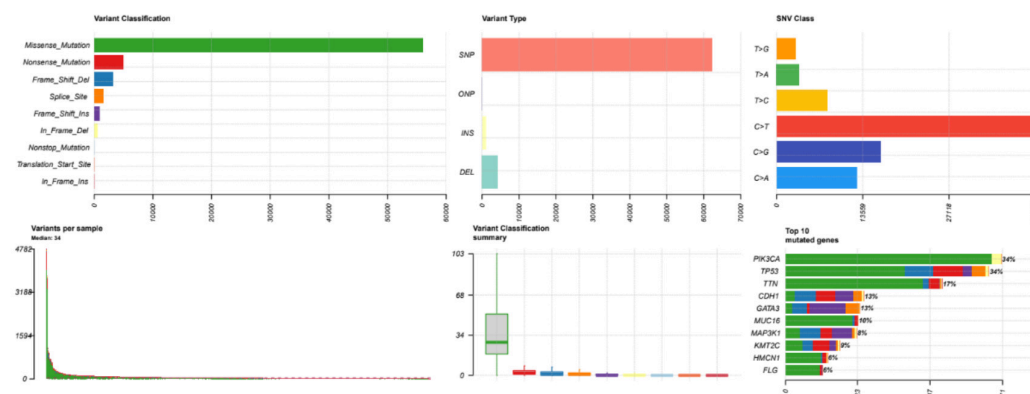
3 RESULTADOS

O projeto TCGA-BRCA do banco de dados The Cancer Genome Atlas inclui 1.098 amostras de câncer de mama, esse número representa as amostras de tecido tumoral coletadas e analisadas no estudo. As amostras foram processadas utilizando pacotes do Bioconductor, bem como as bibliotecas TCGA-Biolinks e TCGA-workflow. Na análise das mutações somáticas, foram realizadas as seguintes avaliações e visualizações: distribuição das mutações, tipos de mutação, genótipos mais frequentes e assinaturas mutacionais.

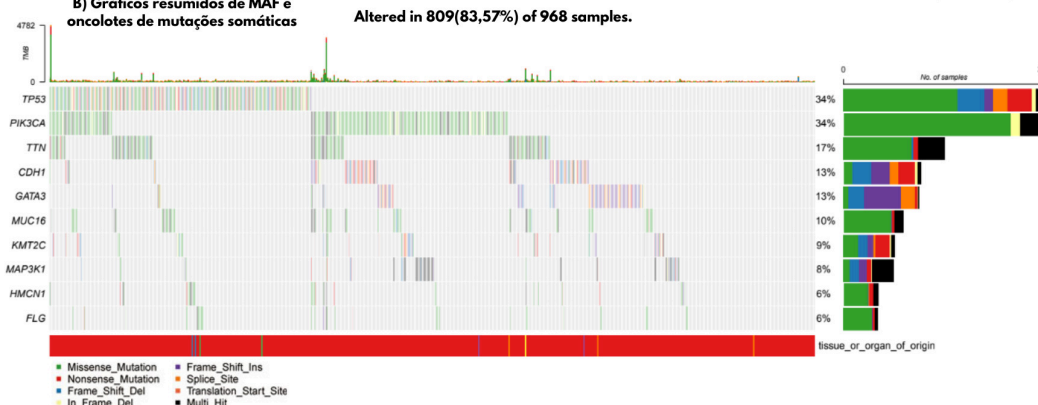
As mutações ocorrem devido a erros que causam danos ao DNA, que podem ou não ser reparados. As taxas de processos mutacionais podem variar em função de diversos fatores, como diferentes tipos de câncer e tumores. Segundo a análise (Figura 01 A), os genes PIK3CA (34%), que codifica uma subunidade da fosfatidilinositol-3-quinase; TP53 (34%), que codifica a proteína p53, essencial para a resposta ao estresse celular; TTN (17%), que codifica a titina, uma proteína estrutural do sarcômero; e CDH1 (13%), importante para a adesão celular e integridade dos tecidos, apresentaram as quatro principais frequências de mutação em pacientes com câncer de mama.

Figura 01: Análise de mutação somática utilizando TCGA-BRCA (1.098 amostras) (A) Anotação de Mutações (MAF) e (B) Gráficos resumidos de MAF e *oncoplots* de mutações somáticas.

A) Anotação de Mutações (MAF)



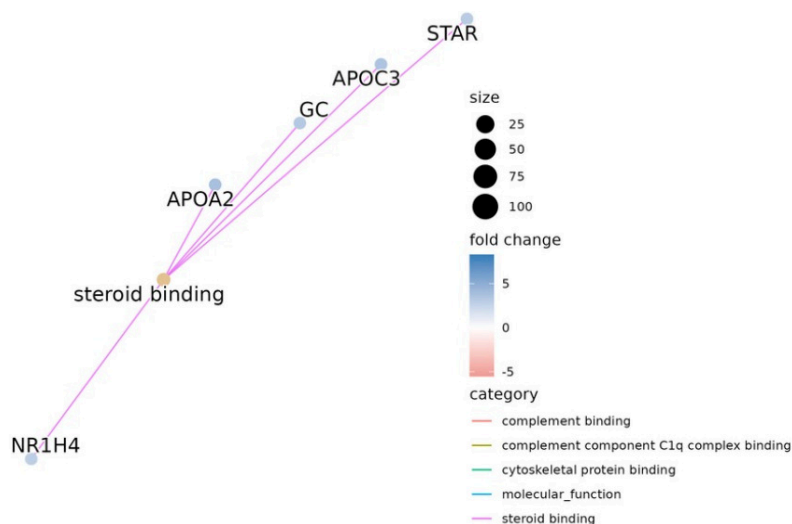
B) Gráficos resumidos de MAF e oncoplots de mutações somáticas



Fonte: Autoria própria.

O enriquecimento de processos e vias é uma análise comum na bioinformática para entender quais processos biológicos ou vias metabólicas estão associados a um conjunto de genes ou mutações de interesse. Na análise de expressão diferencial, identificam-se genes cuja expressão varia significativamente entre as condições comparadas. Esses genes diferencialmente expressos podem estar envolvidos em processos biológicos relevantes para a condição estudada, como o câncer de mama. Para o estudo BRCA, utilizam-se dados disponíveis no TCGA, baixados usando o pacote TCGAbiolinks. TCGA-Biolinks e TCGA-workflow a RTCGA, MAFTOOLS, OPENXLSX e PLYSA. Foi possível obter uma lista de 1251 genes diferencialmente expressos para o câncer em questão. A partir desses dados foi possível obter vias envolvidas nos processos e foi utilizada a via *steroid binding* para análise.

Figura 02: Genes diferencialmente expressos na via *steroid binding*.

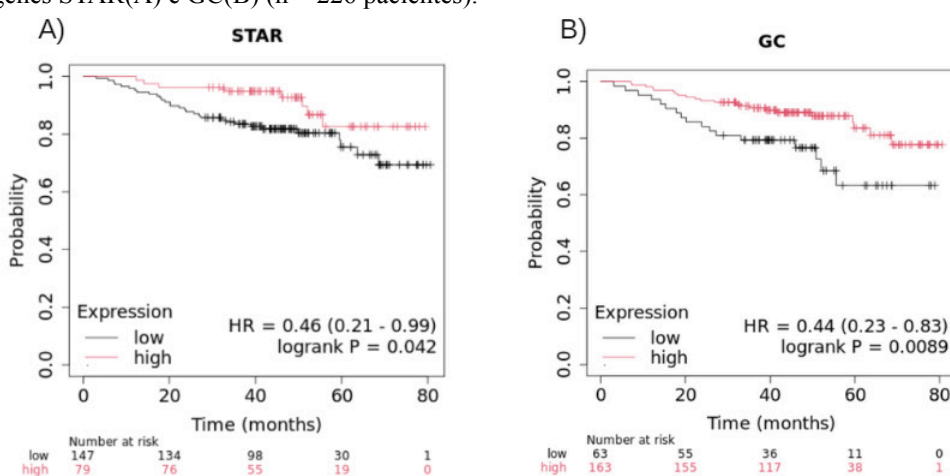


Fonte: Autoria própria.

O estudo de expressão gênica identificou um agrupamento de cinco genes relacionados ao processo de ligação a esteroides (*steroid binding*), todos regulados para cima. Os genes STAR (Steroidogenic Acute Regulatory Protein) é uma proteína crucial para a biossíntese de esteroides, APOC3 (*Apolipoprotein C3*) regula a lipoproteína lipase e o metabolismo de triglicerídeos, APOA2 (*Apolipoprotein A2*) é uma apolipoproteína que participa da formação de lipoproteínas HDL e na

regulação do metabolismo lipídico, GC (Vitamin D Binding Protein) é a proteína codificada pelo gene GC e está envolvida no transporte de vitamina D, e NR1H4 (*Nuclear Receptor Subfamily 1 Group H Member 4*) é um receptor nuclear que regula a homeostase dos ácidos biliares e o metabolismo lipídico. Os genes STAR e GC também apresentaram influência na sobrevida de pacientes com PGR negativo (*Progesterone Receptor*) e estadiamento nível 3.

Figura 03: Sobrevida de pacientes com PGR negativo (*Progesterone Receptor*) e estadiamento nível 3. Análise para os genes STAR(A) e GC(B) (n = 226 pacientes).



Fonte: Autoria própria.

4 DISCUSSÃO

Os resultados desta análise utilizando dados do TCGA-BRCA oferecem contribuições valiosas sobre o câncer de mama, especialmente em relação às mutações somáticas e aos processos biológicos associados. O câncer de mama, como evidenciado, continua a ser um desafio significativo na saúde pública, e a compreensão das variações genéticas e expressivas é crucial para aprimorar as abordagens terapêuticas e diagnósticas. Em relação à frequência de mutações, os dados indicam uma alta frequência de mutações em genes como PIK3CA e TP53. O estudo de expressão diferencial de genes revelou um grupo significativo de genes regulados para cima na via de ligação a esteroides. Os genes STAR, APOC3, APOA2, GC e NR1H4, dos quais os genes STAR e GC apresentam influência na sobrevida dos pacientes com PGR negativo e estadiamento nível 3, são notáveis.

5 CONCLUSÃO

O presente estudo concluiu que a análise de mutações e o estudo da expressão diferencial de genes, em pacientes com câncer de mama, permitiu a identificação de genes associados a via ligada a esteroides como associados à sobrevida dos pacientes com PGR negativo.

REFERÊNCIAS

CASOTTI, Matheus Correia et al. ANÁLISE DE GENES DIFERENCIALMENTE EXPRESSOS EM AMOSTRAS DE CÂNCER DE MAMA DO SEQUENCE READ ARCHIVE (SRA).

RECIMA21-Revista Científica Multidisciplinar-ISSN 2675-6218, v. 5, n. 3, p. e534955-e534955, 2024.

DE OLIVEIRA SANTOS, Marcell. Estimativa 2018: incidência de câncer no Brasil. **Revista Brasileira de Cancerologia**, v. 64, n. 1, p. 119-120, 2018.

OBEID, Elias et al. The role of tumor-associated macrophages in breast cancer progression. **International journal of oncology**, v. 43, n. 1, p. 5-12, 2013.

REIS, Vânia Isabel Gonçalves Pereira dos. **Identificação de genes diferencialmente expressos no cancro da mama**. 2012. Tese (Mestrado em Ciências Biomédicas) - Ciências da Saúde, Universidade da Beira Interior, Covilhã, 2012.

REVEL, Margot et al. Complement system: promoter or suppressor of cancer progression?. **Antibodies**, v. 9, n. 4, p. 57, 2020.

SILVA, Tiago C. et al. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. **F1000Research**, v. 5, 2016.