



UNIVERSIDADE DE CAXIAS DO SUL
ÁREA DO CONHECIMENTO DE CIÊNCIAS DA VIDA
INSTITUTO DE BIOTECNOLOGIA
PROGRAMA DE PÓS GRADUAÇÃO EM BIOTECNOLOGIA

***ANÁLISE IN SILICO* DA EXPRESSÃO DIFERENCIAL DE GENES EM
CÂNCER DE CABEÇA E PESCOÇO**
NIKAEL SOUZA DE OLIVEIRA

CAXIAS DO SUL, 2023

NIKAEL SOUZA DE OLIVEIRA

**ANÁLISE *IN SILICO* DA EXPRESSÃO DIFERENCIAL DE GENES EM
CÂNCER DE CABEÇA E PESCOÇO**

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade de Caxias do Sul, como parte dos requisitos para a obtenção de grau de Mestre em Biotecnologia.

Orientador: Scheila de Avila e Silva

CAXIAS DO SUL, 2023

Dados Internacionais de Catalogação na Publicação (CIP)
Universidade de Caxias do Sul
Sistema de Bibliotecas UCS - Processamento Técnico

- O48a Oliveira, Nikael Souza de
Análise *in silico* da expressão diferencial de genes em câncer de cabeça e pescoço [recurso eletrônico] / Nikael Souza de Oliveira. – 2023.
Dados eletrônicos.
Dissertação (Mestrado) - Universidade de Caxias do Sul, Programa de Pós-Graduação em Biotecnologia, 2023.
Orientação: Scheila de Avila e Silva.
Modo de acesso: World Wide Web
Disponível em: <https://repositorio.ucs.br>
1. Câncer. 2. Cabeça. 3. Pescoço. 4. Genes. I. Silva, Scheila de Avila e, orient. II. Título.

CDU 2. ed.: 616-006

Catalogação na fonte elaborada pela(o) bibliotecária(o)
Carolina Machado Quadros - CRB 10/2236

**ANÁLISE *IN SILICO* DA EXPRESSÃO DIFERENCIAL DE GENES EM
CÂNCER DE CABEÇA E PESCOÇO**

NIKAEL SOUZA DE OLIVEIRA

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade de Caxias do Sul, como parte dos requisitos para a obtenção de grau de Mestre em Biotecnologia.

Aprovada em _____ de _____ de 2023.

Banca Examinadora

Orientadora: Profa. Dra. Scheila de Ávila e Silva
Universidade de Caxias do Sul

Prof. Dr. Rafael Colombo
Universidade de Caxias do Sul

Prof. Dr. Igor Araujo Vieira
Universidade do Vale do Rio dos Sinos

Profa. Dra. Mariana Recamonde Mendoza
Universidade Federal do Rio Grande do Sul

AGRADECIMENTOS

Todos os acertos e erros de minha vida me trouxeram até o presente momento, cada uma das etapas que por vezes me arrependo foram essenciais para me tornar quem sou, e por este motivo nada mudaria. Agradeço ao Nikael do passado pelas escolhas que foram feitas no processo de tornar o Nikael do presente e desejo os melhores aos caminhos ao Nikael do futuro.

Agradeço a todos que estiveram presentes comigo neste caminho, a meus amigos, minha família, meus colegas de trabalho dos três diferentes lugares em que trabalhei durante este mestrado. Agradeço também a equipe do laboratório de Biologia Computacional e Bioinformática UCS, bem como as equipes dos laboratórios o que nos cercam, principalmente a Professora Scheila de Avila e Silva que me orienta desde o início da minha primeira graduação e a meus colegas Fernanda de Abreu Pessi, Pedro Lenz Casa e Marcos Rossetto pelo apoio neste trabalho.

Por último, mas não menos importante, agradeço a CAPES pelo fomento, e a Universidade de Caxias do Sul pelo apoio estrutural.

ÍNDICE

Resumo.....	08
Abstract.....	09
1. Introdução.....	10
2. Revisão da literatura.....	12
2.2. Câncer de cabeça e pescoço.....	12
2.3. Expressão gênica e biomarcadores.....	19
2.4 Biologia computacional.....	21
3. Objetivos.....	23
3.1. Objetivo geral.....	23
3.2. Objetivos específicos.....	23
4. Resultados.....	24
5. Conclusões e perspectivas.....	37
6. Referências.....	55

LISTA DE SIGLAS/ABREVIATURAS

> – Maior que

< – Menor que

DNA – Ácido desoxirribonucleico

DEGs – Genes diferencialmente expressos

GEPIA – Gene Expression Profiling Interactive Analysis

GEO – Gene Expression omnibus

HPV – Human papillomavirus

TCGA – The Cancer Genome Atlas

TNM – Tumor, Nodo, Metástase

RNA – Ácido Ribonucleico

LISTA DE TABELAS

Tabela 1 – Classificação TNM do câncer de orofaringe.....15

Tabela 2 – Estadiamento do câncer de orofaringe.....17

LISTA DE FIGURAS

Figura 1 – Principais sítios anatômicos do câncer e cabeça e pescoço.....	12
Figura 2 – Extração do RNA, conversão em cDNA; Hibridização das sondas do microarranjo; Leitura do valor de expressão com base na coloração formada durante a hibridização.....	20

Resumo

Câncer de cabeça e pescoço é um termo utilizado como denominação a um grupo de cânceres que ocorrem acima da região da clavícula. Devido a particularidade anatômica estes tipos de cânceres tendem a ser agressivos e desfigurantes. De modo geral, cânceres são patologias de origem genética divergindo o fenótipo celular normal, onde as células se multiplicam descontroladamente, invadem tecidos vizinhos e causam prejuízos ao organismo. As alterações de fenótipos podem ser identificadas por alterações nos padrões de expressão genica do tecido normal de origem. A evolução das tecnologias moleculares tem permitido a obtenção de um maior volume de dados, os quais vêm sendo depositados em repositórios públicos como *The cancer genome atlas* (TCGA) e *Gene expression omnibus database* (GEO) para que pesquisadores realizem novas descobertas. Neste sentido, este trabalho visa a identificação de genes diferencialmente expressos (DEGs) em câncer de cabeça e pescoço utilizando dados destes repositórios públicos. Devido a particularidade amostral as análises realizadas foram divididas em 2 grupos: câncer de cavidade oral e câncer de tireoide. Os dados foram organizados e preparados utilizando o ambiente R, e após submetidos a estatística com pontos de corte $p < 0,05$ e $\log_2FC > |1|$. Em seguida a identificação dos DEGs realizou-se as curvas de sobrevivência e pesquisa na literatura. Como resultados obteve-se 4 DEGs para câncer de cavidade oral: LOXL2, FAT1, SOD3 e DPT e 12 DEGs para câncer de tireoide: CRABP1, CSGALNCT1, WSCD2, ITM2A, ID3, PAX9, SELENBP1, TNFRS11B, SDPR, LRP1B, MT1F e SGK223. A partir da curva de sobrevivência FAT1, CRABP1, ITM2A e ID3 demonstraram potencial prognóstico. Todos os DEGs identificados possuem referencial teórico demonstrando relações com câncer. CSGALNCT1 possui apenas descrições relacionadas a família gênica. Métodos experimentais são necessários para a confirmação do uso destes genes como biomarcadores. Desta forma, os resultados encontrados apresentam contribuição científica. Ao mesmo tempo em que este trabalho possui potencial a contribuição à pesquisas futuras auxiliando na descoberta de melhores métodos diagnósticos e prognósticos.

Palavras-chave: Câncer de cabeça e pescoço; Câncer de cavidade oral; Câncer de tireoide; Análise de expressão diferencial de genes; GEO; TCGA.

Abstract

Head and neck cancer is a term used as denomination for a cancer group that occurs above the clavicle. Due to the anatomy particularity those cancer types tend to be aggressive and disfiguring. Generally, cancers are genetics pathologies diverging from the normal cellular phenotype, causing an uncontrollably cell division, which invade nearby tissues and cause body damage. The phenotype alteration can be identified by changes in the gene expression patterns of normal tissues. The molecular technology evolution has allowed the obtaining a larger volume of data, which has been deposited on public repositories as the cancer genome atlas (TCGA) and Gene expression omnibus database (GEO) allowing new discoveries. In this regard, this work aims the identification of differentially expressed genes on head and neck cancer using data by those public repositories. Due to the sampling particularity the data analysis was divide at 2 groups: Oral Cavity Cancer and Thyroid Cancer. The data was prepared and organized using the R environment, and after submitted to statistics with the cut off $p < 0,05$ and $\log_2FC > |1|$. After the DEGs identification survival curves were performed and literature researched. As results it was obtained 4 DEGs for oral cavity cancer: LOXL2, FAT1, SOD3 and DPT and 12 DEGs for thyroid cancer: CRABP1, CSGALNCT1, WSCD2, ITM2A, ID3, PAX9, SELENBPQ, TNFRS11B, SDPR, LRP1B, MT1F, SGK22. With the survival curve FAT1, CRABP1, ITM2A and ID3 evidence prognostic potential. All the DEGs identified on this work had papers describing their relation with cancer. CSGALNCT1 has only descriptions about the genic family. Molecular works are needed for confirmation about use of these genes as biomarkers. This way, the results found here have scientific contribution. At the same time this work could contribute for futures research and help at new discoveries about diagnostics and prognostics.

Keywords: Head and neck cancer; Oral Cavity Cancer; Thyroid Cancer; Differential expression Gene; Analysis; GEO; TCGA.

1. Introdução

Câncer é uma patologia de origem genética, ocasionada por alterações do funcionamento de alguns genes presentes no organismo, causando uma fuga dos padrões celulares normais. Naturalmente as células se dividem, cumprem suas funções e morrem. Entretanto, devido a mutações que ocorrem durante os processos de divisão celular, os ciclos normais se alteram e uma única célula pode gerar uma massa que se multiplica descontroladamente, invade tecidos vizinhos e causa prejuízos ao portador (ALBERTS, 2017; GRIFFITHS, 2019).

Há dois grupos principais de genes relacionados ao câncer: (i) Oncogenes e (ii) genes supressores de tumor. O primeiro, antes de se tornarem oncogenes propriamente ditos, são denominados de proto-oncogenes, e atuam nas funções normais das células. O segundo, são genes que atuam na regulação do ciclo celular impedindo a formação de tumores, entretanto quando mutados perdem sua ação (PIERCE, 2017; RODRIGUES, 2016). Vale ressaltar que são necessárias um acúmulo crítico de mutações nestes genes para que uma célula se torne câncer. Sendo que, durante a vida de um ser humano, ocorrem cerca de 10^{16} divisões celulares e que mutações espontâneas ocorrem a uma taxa de 10^{-6} mutações por gene por mitose, desta forma, cada gene deve sofrer cerca de 10^{10} mutações em ocasiões independentes. Portanto, se uma única mutação fosse suficiente para originar uma neoplasia, esse processo seria incompatível com a vida (ALBERTS, 2017).

Câncer de cabeça e pescoço é um termo guarda-chuva utilizado como agrupamento aos cânceres que ocorrem acima da região da clavícula, com exceção do sistema nervoso central; pele; linfonodos; olhos; esôfago e traqueia. Este grupo de cânceres foi responsável, em 2020, por 8,4% dos novos casos de cânceres, bem como 5,2% das mortes entre todos os cânceres em todo o mundo. Ao mesmo tempo, no Brasil houve uma incidência de 10,35% entre os novos casos de cânceres e uma taxa de mortalidade de 6,03% (WHO, 2021).

De modo geral, os cânceres de cabeça e pescoço possuem como principais fatores de risco o tabagismo e o consumo de álcool, bem como a infecção pelo vírus HPV. Havendo uma correlação indiretamente proporcional entre o índice de desenvolvimento de um país e a incidência de novos casos de câncer de cabeça e pescoço.

Quando um paciente com esta patologia é diagnosticado em estágios iniciais, estima-se que a chance de cura seja em torno de 80%. Entretanto, no Brasil, a maioria dos pacientes são diagnosticados em estágios avançados, e devido a isso são submetidos a tratamentos agressivos e desfigurantes e com uma menor taxa de sobrevivência.

Com o advento das tecnologias “ômicas”, cada vez mais grandes volumes de dados são gerados, ocasionando o desafio da busca por informações biologicamente relevantes. Uma deste tipo de tecnologia é a transcriptômica, baseada na análise dos transcritos de mRNA. Há duas tecnologias principais a este tipo de análise, *microarray* e RNA-seq.

Repositórios públicos internacionais como o *Gene Expression Omnibus* (GEO)¹ e *The Cancer Genome Atlas* (TCGA)² armazenam dados provenientes das técnicas de *microarray* e RNA-seq respectivamente. Estes dados são de livre utilização, depositados por pesquisadores para que outros pesquisadores possam realizar novas pesquisas. Necessitando de análises estatísticas e computacionais complexas, como algoritmos de inteligência artificial. Para isto, é necessário que o pesquisador compreenda não só a técnica molecular, mas também possua conhecimentos dos métodos computacionais e linguagens de programação de modo a lidar apropriadamente com a informação disponível.

Diversos trabalhos tem sido descritos na literatura relacionando os conjuntos de dados depositados nestes bancos, com novas descobertas de potenciais biomarcadores moleculares para diagnóstico, prognóstico e alvos terapêuticos. As descobertas relacionadas a este campo têm grande potencial á auxilio a qualidade de vida dos pacientes portadores de neoplasias, podendo resultar em tratamentos precoces, bem como guiar novas terapias. A análise *in silico* por si só não traz grandes aplicações de mercado, entretanto podem direcionar pesquisas nessa área, seja para análises *in vitro*, seja pelo direcionamento de fármacos já utilizados *in vivo* à novos tipos de neoplasias.

Neste contexto, este trabalho visa a utilização dos dados de expressão gênica, para câncer de cabeça e pescoço, presentes nos repositórios públicos GEO e TCGA, com auxílio ambiente computacional R e objetivando a identificação de potenciais biomarcadores que possam vir a auxiliar no diagnóstico precoce. Evitando tratamentos agressivos e aumentando as chances de cura dos pacientes.

¹ <https://www.ncbi.nlm.nih.gov/geo/>

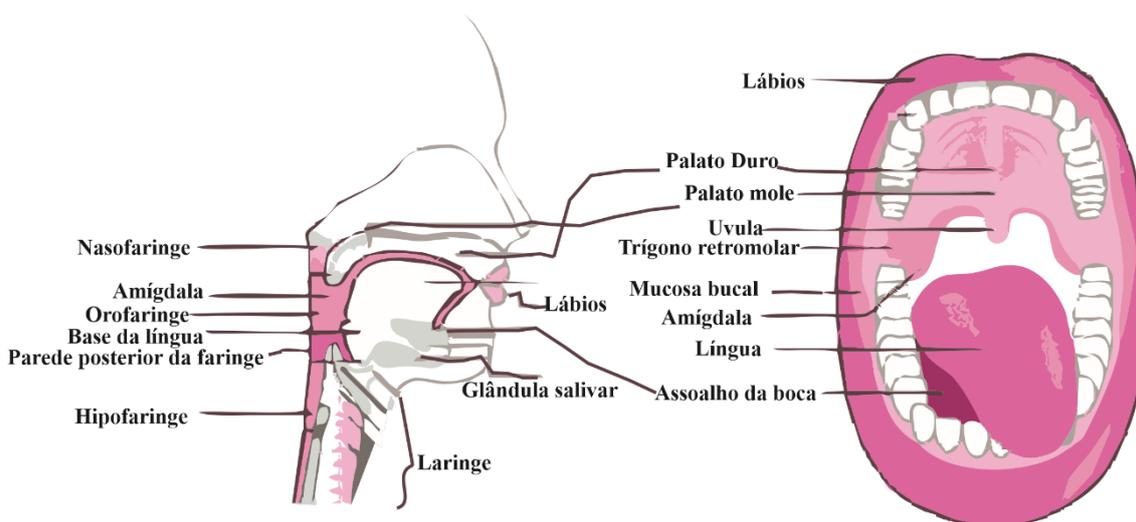
² <https://portal.gdc.cancer.gov/>

2. Revisão da literatura

2.1. Câncer de cabeça e pescoço

Câncer de cabeça e pescoço é um termo guarda-chuva utilizado como denominação a um grupo de cânceres que ocorrem acima da região da clavícula, havendo algumas divergências na literatura, mas, com exceção do cérebro, pele, linfonodos, olhos, esôfago e traqueia (CHOW, 2020; KONINGS *et al*, 2020; WHO, 2021). Os principais sítios anatômicos presentes neste tipo de câncer são: (i) cavidade oral, que inclui: lábios, mucosa bucal, língua, assoalho da boca, palato duro e gengiva; (ii) faringe; (iii) laringe; (iv) cavidade nasal; (v) seios paranasais; (vi) ouvido e (vii) tireoide. Além dos sítios anatômicos, o câncer de cabeça e pescoço também é dividido por tipo histológico, sendo o carcinoma de células escamosas o mais comum, responsável por 90% dos casos, seguido pelo adenocarcinoma (ARQUETTE, 2002; CHOW, 2020; GALBIATTI *et al*, 2013; JOHNSON *et al*, 2020). A figura 1 demonstra os principais sítios anatômicos envolvidos no câncer de cabeça e pescoço.

Figura 1 – Principais Sítios anatômicos ao Câncer de Cabeça e Pescoço



Adaptado de: Johnson *et al*, 2020.

Segundo dados da Organização Mundial da Saúde (WHO), estima-se que em 2020 o câncer de cabeça e pescoço representou em todo o planeta cerca de 1.518.133

novos casos, 8,4% entre todos os tipos de cânceres, e 5.510.771 mortes, 5,2% entre todos os tipos de cânceres. No Brasil, estima-se que a incidência deste tipo de câncer em 2020 foi de 57.633 novos casos, 10,35% entre todos os tipos de cânceres e 15.506 mortes, 6,03% entre todos os tipos de cânceres. Destes, destes 586.202 (3,23%) são de novos cânceres de tireoide, e 43.646 óbitos devido a progressão da patologia em todo o mundo. Enquanto no Brasil, surgiram 30.607 (5,49%) novos casos e 1081 (0,42%) novos óbitos (WHO, 2021).

Existe uma correlação entre a incidência deste tipo de câncer com o desenvolvimento de uma nação. Países subdesenvolvidos tendem a apresentar um número maior de casos de câncer de cabeça e pescoço do que países desenvolvidos. Associa-se tal correlação aos fatores de risco associado ao surgimento deste câncer, visto que estes se apresentam uma proporção inversa ao índice de desenvolvimento (ARQUETTE, 2002; GALBIATTI *et al*, 2013; JOHNSON *et al*, 2020).

Consumo de cigarros, álcool, má higiene bucal, deficiências nutricionais, exposição ocupacional e infecção por HPV são os principais fatores de risco associado ao câncer de cabeça e pescoço. O tabagismo é o principal fator de risco a este tipo de câncer, relacionando-se com a intensidade e a duração do hábito. Não só o ato de fumar em si, como o ato de mascar o fumo também eleva o risco de desenvolvimento deste câncer. Isso acontece pois o tabaco contém agentes carcinogênicos e a exposição local tende, mais fortemente, a causar mutações locais (ARQUETTE, 2002; GALBIATTI *et al*, 2013; JOHNSON *et al*, 2020).

É interessante a correlação entre o consumo de álcool associado ao tabagismo. O tabagismo e o alcoolismo são fatores de risco já quando ocorrem individualmente, entretanto em conjunto, provocam um aumento significativo na chance do desenvolvimento deste tipo de câncer, estima-se que em torno de 40 vezes mais. Entende-se que o álcool age como um solvente, aumentando a exposição da mucosa oral aos agentes carcinogênicos (ARQUETTE, 2002; GALBIATTI *et al*, 2013; JOHNSON *et al*, 2020).

Deficiências nutricionais estão relacionadas a diversos tipos de doenças, dependendo do nutriente. No câncer de cabeça e pescoço, a deficiência de vitamina A bem como a deficiência de ferro associada a anemia férrica se demonstram como fatores de risco. A exposição ocupacional está relacionada a inalação de asbestos, níquel, cromo e ao pó gerado no processamento do couro e de madeiras (ARQUETTE, 2002; GALBIATTI *et al*, 2013; JOHNSON *et al*, 2020).

A infecção pelo *Human papillomavirus* (HPV) é responsável por alguns tipos de câncer, como o câncer cervical; o câncer de vulva; o câncer de pênis; o câncer anal e câncer de cabeça e pescoço (HARDEN; MUNGER, 2017). Os HPVs são vírus de DNA, pertencentes a família Papillomaviridae, e possuem 9 proteínas, sendo E6 e E7 responsáveis pela alteração do fenótipo celular. Sua classificação taxinômica é realizada com base em homologia, entretanto confusa, sendo mais comum sua utilização por tipos. Sendo os tipos HPV-18 e HPV-16 os mais comuns associado aos cânceres e designados como tipos de *alto risco* (ROMANOS; MENDES, 2021).

A transmissão do HPV ocorre por via sexual, seja ela oral; vaginal ou anal, necessitando de micro-abrasões ou, no caso do câncer cervical, da junção escamo-celular. A partícula viral infecta as células das camadas basais da epiderme por meio de endocitose, e não costuma se espalhar pelo corpo. Sua infecção é local, e contínua apenas para as células vizinhas. O HPV estimula a divisão celular, ocasionando a formação dos papilomas. Nos casos em que ocorre o processo oncogênico, o DNA viral se integra ao genoma do hospedeiro, ocasionando alterações gênicas em ambos e aumentando a expressão dos genes virais E6 e E7 (ROMANOS; MENDES, 2021).

O sistema imunológico geralmente possui capacidade de combater o HPV, entretanto após infecção concretizada o vírus consegue enganar o sistema imunológico. Entretanto, há medidas profiláticas a este processo infeccioso, como a utilização de preservativos em qualquer forma de sexo e vacinação. No Brasil, o sistema único de vacinação fornece a administração de duas doses da vacina contra o HPV-6, 11, 16 e 18 para adolescentes (BRASIL, 2021). Além disso, é importante ressaltar que em países desenvolvidos com altas taxas de vacinação ocorreu uma diminuição significativa nos cânceres de genitália feminina, câncer retal e câncer de cabeça e pescoço (HARDEN; MUNGER, 2017; JOHNSON *et al*, 2020).

O estadiamento deste tipo de câncer é dado pelo sistema TNM (tumor, nodo, metástase). Sendo o T geralmente referente ao tamanho do tumor e pequenos processos invasivos; N referente ao acometimento de linfonodos adjacentes e M metástases distantes. É importante ressaltar que como câncer de cabeça e pescoço é um termo guarda-chuva para alguns tipos de cânceres, há diferenças singelas no modelo de classificação, como os tipos de invasão em tecidos adjacentes (CHOW, 2020; BRIERLEY *et al*, 2017).

De forma relativamente recente o status de HPV se tornou um fator importante na classificação TNM do câncer de orofaringe. Havendo diferenças na classificação em T e N. A tabela 1 demonstra a classificação do câncer de orofaringe, dividindo em status de

HPV positivo e negativo (CHOW, 2020; BRIERLEY *et al*, 2017). Após a classificação TNM, se realiza o estadiamento do tumor conforme tabela 2. Os estágios variam de 0 à IV, havendo subdivisões, entretanto o estágio 0 é sempre carcinoma *in situ* e a última subdivisão do IV (quando há subdivisões) já ocorre metástase distante.

Tabela 1 – Classificação TNM do câncer de orofaringe.

Classificação (TNM)	Câncer de orofaringe HPV positivo	Câncer de orofaringe HPV negativo
TX	Tumor primário não pode ser avaliado	Tumor primário não pode ser avaliado
Tis	Carcinoma in situ	Carcinoma in situ
T0	Nenhum tumor identificado	Nenhum tumor identificado
T1	Tumor menor que 2 cm	Tumor menor que 2 cm
T2	Tumor entre 2 e 4 cm	Tumor entre 2 e 4 cm
T3	Tumor maior que 4 cm ou com extensão da superfície da língua para a epiglote	Tumor maior que 4 cm ou com extensão da superfície da língua para a epiglote
T4	Avanço local moderado da patologia; Invasão da laringe, músculos da língua, musculatura acessória adjacente, palato duro ou mandíbula	-
T4a	-	Avanço local moderado da doença; O tumor invade laringe, músculos da língua, musculatura

		acessória adjacente, palato duro ou mandíbula
T4b	-	Patologia local muito avançada; Tumor invade musculatura acessória, lateral da nasofaringe, base do crânio ou parte da artéria carótida
Nx	Linfonodos locais não podem ser avaliados	Linfonodos locais não podem ser avaliados
N0	Sem metástase em linfonodos adjacentes	Sem metástase em linfonodos adjacentes
N1	Metástase para 1 ou mais linfonodos laterais menor que 6cm	Metástase para 1 único linfonodo lateral, menor ou igual a 3cm, sem extensão
N2	Metástase para linfonodo contralateral ou bilateral, menor que 6cm	-
N2a	-	Metástase para um único linfonodo lateral; Entre 3 e 6 cm, sem extensão
N2b	-	Metástase para vários linfonodos laterais; menor que 6cm, sem extensão
N2c	-	Metástase contralateral ou bilateral em linfonodos, menor que 6cm, sem extensão

N3	Metástase para 1 ou mais linfonodos, maior que 6cm	
N3a		Metástase para linfonodo, maior que 6cm, sem extensão
N3b		Metástase para 1 ou mais linfonodos, com extensão clinicamente evidenciada
M0	Sem metástase distante	Sem metástase distante
M1	Metástase distante	Metástase distante

Adaptado de: BRIERLEY, 2017; CHOW, 2020.

Tabela 2 – Estadiamento do câncer de orofaringe

Estágio	Câncer de orofaringe HPV positivo			Câncer de orofaringe HPV negativo		
	Tumor	Nodo	Metástase	Tumor	Nodo	Metástase
0	Tis	N0	M0	Tis	N0	M0
I	T0, T1, ou T2	N0 ou N1	M0	T1	N0	M0
II	T0, T1 OU T2	N2	M0	T2	N0	M0
	T3	N0, N1 ou N2	M0	-	-	-
III	Qualquer T	N3	M0	T1, T2 ou T3	N1	M0
	T4	Qualquer N	M0	-	-	-
IV	Qualquer T	Qualquer N	M1	-	-	-
IVa	-	-		T4a	N0 ou N1	M0

	-	-	T1, T2, T3 OU T4a	N2	M0
IVb	-	-	Qualquer T	N3	M0
	-	-	T4b	Qualquer T	M0
IVc	-	-	Qualquer T	Qualquer T	M1

Adaptado de: BRIERLEY, 2017; CHOW, 2020.

Estima-se que no Brasil, 70% dos pacientes com câncer de cabeça e pescoço são diagnosticados em estágios avançados, diminuindo a chance de cura e a qualidade de vida durante o tratamento. Esta situação ocorre devido à falta de políticas públicas relacionadas a prevenção deste tipo de câncer, como avaliações odontológicas periódicas ou campanhas de autocuidado. Entretanto, quando o diagnóstico ocorre de forma precoce, estima-se que a chance de cura atinja 80% (REIS *et al*, 2018).

A importância de estudos relacionados ao câncer de cabeça e pescoço vai além de taxas de mortalidade e incidência. A face é o local por onde um ser humano reconhece o outro, responsável pela fala e por praticamente todas as interações sociais diárias. Além disso, é o local por onde ocorre a entrada de todos os elementos essenciais ao funcionamento do organismo. Sendo assim, uma patologia deste nível afeta o indivíduo de forma física, fisiológica e psicossocial, acarretando baixa autoestima, comunicação afetada, dor local, secreção, odores, perda de autonomia, eliminações fisiológicas e alimentação prejudicada (ARQUETTE, 2002; GENDEN, 2020; REIS *et al*, 2018).

Dependendo do estágio em que o câncer é diagnosticado, o tratamento tende a ser agressivo e desfigurante. Quimioterapias, radioterapias, imunoterapias e cirurgias são as principais formas de tratamento, e devido a particularidade da localização deste tipo de tumor, o paciente necessita de utilização de sondas para alimentação, traqueostomias e em alguns casos perca de parte da face. Estudos que possam auxiliar na construção do conhecimento molecular e por conseguinte em métodos preventivos, diagnósticos e em tratamentos, podem vir a auxiliar este tipo de paciente e evitando sofrimento (ARQUETTE, 2002; GENDEN, 2020; REIS *et al*, 2018).

2.2. Expressão gênica e biomarcadores

O material genético é a molécula responsável pela hereditariedade. Um organismo multicelular, como um ser humano, se origina de uma mesma célula inicial, e, portanto, todas as células de um indivíduo possuem a mesma sequência gênica no material genético, com algumas exceções ocasionadas por mutações aleatórias nas células somáticas (ALBERT, 2017). Diferentes tecidos no corpo apresentam os diferentes fenótipos mesmo apresentando o mesmo genótipo, devido a diferenças na expressão gênica, bem como aos mecanismos reguladores desta expressão.

A expressão gênica em eucariontes depende de vários fatores e reguladores, os quais definem a diferenciação celular entre tecidos e, também um tecido normal de um tecido tumoral. (GRIFFITHS *et al*, 2016; PIERCE, 2016; SNUSTAD; SIMMONS, 2017). Entender estes padrões de expressão permite a compreensão dos mecanismos moleculares, e assim a identificação de biomarcadores para diagnóstico e prognóstico, bem como potenciais alvos de terapia.

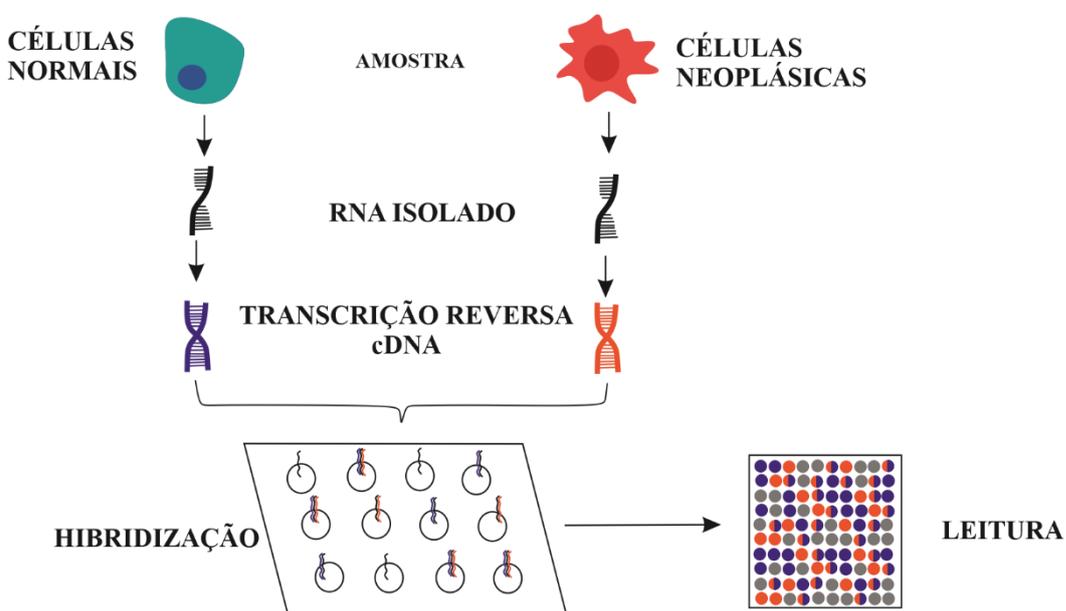
Uma das formas de análise da expressão diferencial de genes é com a transcriptômica, a qual baseia-se na análise dos transcritos de DNA e são indicativos de atividade gênica (ADAMS, 2008). Há dois métodos principais para obtenção de transcriptomas: (i) *microarrays* (BUMGARNER, 2013) e (ii) RNA-seq (WANG *et al*, 2009). Ambos se baseiam na quantidade de mRNA expresso pela célula em dado momento. Para isso, a análise é feita através de cDNA, ou seja, é utilizada uma enzima chamada de transcriptase reversa (RT), a qual converte o mRNA em DNA complementar. É este cDNA que é então quantificado para as análises de expressão. A análise global dos dados possível é dependente de ferramentas computacionais específicas para esta finalidade (SARTOR; ROSSETO, 2020).

A tecnologia de *microarray* ou microarranjo se baseia em um chip, com diversos *spots* contendo sondas de DNA. O mRNA é extraído do tecido biopsiado e convertido em cDNA pela transcriptase reversa. O material é posto sobre o chip, e o cDNA se hibridiza com as sondas de DNA gerando coloração. A partir disso, o chip é posto em equipamento especializado que mede a intensidade da cor permitindo sua transformação em um valor de expressão (BUMGARNER, 2013). A figura 2 ilustra este processo.

A tecnologia de RNA-Seq se baseia nas tecnologias de sequenciamento de DNA. Da mesma forma que no microarranjo, o mRNA é extraído e convertido em cDNA, que é sequenciado. Estas tecnologias geralmente necessitam da fragmentação do DNA, e a

leitura é feita a partir de hibridização e coloração. Cada base que se hibridiza gera um comprimento de onda diferente, e esta é lida por equipamento específico em tempo real. Ao final do processo, se tem a sequência de bases de DNA como um dado. Utilizando ferramentas computacionais, esses dados são montados, e o número de sequências presentes na amostra é quantificado em valores de expressão por gene. A tecnologia de RNA-seq é considerada mais acurada do que a tecnologia de microarranjo (WANG *et al*, 2009).

Figura 2 – Etapas envolvidas na execução da metodologia de Microarranjo.



Fonte: BUMGARNER, 2013; SARTOR & ROSSETO, 2020.

Diversos trabalhos tem sido realizados nesta área nos diferentes tipos de câncer. Sartor; Recamonde-Mendoza & Ashton-Prolla (2019) identificaram TULP3 como um potencial biomarcador ao diagnóstico do câncer colorretal, a partir de dados de expressão obtidos do bando de dados GEO e TCGA. Da mesma forma, Liu *et al* (2021) identificaram AGT, SERPINH1 e MMP7 como biomarcadores ao câncer gástrico, estudo o qual foi realizado *in silico* e confirmado experimentalmente.

XU *et al* (2019) demonstraram a identificação de CDC20 e ASPM como potenciais biomarcadores ao câncer de bexiga utilizando os dados presente no GEO e TCGA. Da mesma forma, Zheng; Min & Zhou (2021) identificaram ADCY8, ADRB2,

CALCA, GCG, GNGT1 e NPSR1 como potenciais biomarcadores e alvos terapêuticos ao adenocarcinoma de pulmão utilizando os dados presentes nos repositórios públicos.

Investigações de microRNAs circulantes também tem sido avaliado como potenciais biomarcadores a alguns tipos de cânceres, utilizando dados de transcriptômica (WANG *et al*, 2018). Ao mesmo tempo, fatores epigenéticos como metilação de DNA podem ser avaliados na identificação de potenciais biomarcadores, utilizando também dados de expressão gênica (LOCKE *et al*, 2019; TERRY *et al*, 2017). Desta forma, percebe-se que estudos de transcriptômica possuem um grande potencial, e reforçam a importância dos repositórios públicos com dados ômicos.

2.3. Biologia computacional

Metabolômica, transcriptômica, proteômica, são ciências intimamente ligadas à bioinformática, seja para análise de dados, seja para simulações (DATO *et al*, 2021). Neste processo, bancos de dados têm sido criados em cooperativas mundiais. GEO *database* (EDGAR *et al*, 2002) e *The cancer genome atlas* (Cancer Genome Atlas Research Network, 2014) são exemplos disso. Ambos são repositórios públicos de dados genômicos, publicados por pesquisadores, e permitem que outros pesquisadores realizem novas descobertas.

O TCGA é um projeto que possui como objetivo catalogar as principais alterações genéticas causadoras do câncer, descrevendo por meio de sequenciamento genômico, RNA-seq e análises multidimensionais integradas, tecidos tumorais e normais. Com isso, fornece conjuntos de dados genômicos e transcriptômicos de câncer, os quais são disponibilizados publicamente permitindo a melhoria dos métodos de diagnóstico, tratamento e prevenção (Cancer Genome Atlas Research Network *et al*, 2014).

O GEO é um repositório público internacional que armazena e disponibiliza dados de microarranjo. O GEO tem três principais objetivos; (i) fornecer um banco de dados robusto e versátil para armazenar dados de alto rendimento; (ii) oferecer procedimentos de apresentação simples e formatos que suportam depósitos de dados completos da comunidade de pesquisa; (iii) fornecer mecanismos de fácil utilização que permita que o usuário pesquise, localize, revise e faça *download* de estudos e perfis de expressões dos genes (EDGAR *et al*, 2002).

Ao mesmo tempo, ferramentas computacionais que permitam a análise de grandes volumes de dados são de extrema importância. Existem diversos programas específicos

para a análise destes dados biológicos, tais como o GEOR2 e GEPIA (TANG *et al*, 2017). Ainda assim, ferramentas para análises de dados gerais também são utilizadas como R (R development core team, 2021) e python (ROSSUM; DRAKE, 1995).

Diversos trabalhos tem sido realizados utilizando de dados presentes em bancos de dados na identificação de potenciais biomarcadores para diagnóstico, prognóstico e alvos terapêuticos. Exemplos de trabalhos relacionados são a identificação de CDC20 e ASPM (XU *et al*, 2019) como potenciais biomarcadores ao câncer de bexiga, ADCY8, ADDR2, CALCA, GCG, GNGT1 e NPSR1 (ZHENG; MIN; ZHOU, 2021) como potenciais biomarcadores para adenocarcinoma de pulmão ou TULP 3 como potencial biomarcador ao prognóstico de adenocarcinoma de ducto pancreático (SARTOR *et al*, 2014).

Da mesma forma trabalhos relacionados ao câncer de cabeça e pescoço tem sido descrito, como a identificação de PLAU, CLDN8 e CDKN2A como potenciais preditores de sobrevivência global em carcinoma oral de células escamosas (ZHAO *et al*, 2018). Estudos neste sentido tem demonstrado grande potencial científicos, demonstrando a importância da bioinformática e impulsionando novas descobertas.

3. Objetivos

3.1. Objetivo geral

O presente trabalho possui como objetivo geral identificar potenciais biomarcadores em câncer de cabeça e pescoço utilizando dados disponíveis em repositórios públicos.

3.2. Objetivos específicos

- Comparar a expressão diferencial de genes entre biópsias de câncer de tireoide e cavidade oral com seus respectivos tecidos normais;
- Relacionar a expressão gênica em câncer de cabeça e pescoço com os índices de sobrevivência dos pacientes;
- Identificar potenciais biomarcadores em câncer de cabeça e pescoço.

4. Resultados

Os resultados da presente dissertação estão organizados em um capítulo o qual possui dois artigos.

O primeiro artigo é um tutorial introdutório para obtenção e análise de genes diferencialmente expressos utilizando o repositório público GEO, intitulado: “ANÁLISE DE GENES DIFERENCIALMENTE EXPRESSOS: UMA ABORDAGEM METODOLÓGICA *IN SILICO*” e publicado na revista brasileira UNILUS ensino e pesquisa³ em 30 de junho de 2023⁴.

O artigo de número 2 apresenta os resultados da obtidos durante a dissertação, dividindo o câncer de cabeça e pescoço em câncer de cavidade oral e câncer de tireoide. Este é intitulado: “GENES DIFERENCIALMENTE EXPRESSOS EM CÂNCER DE CAVIDADE ORAL E DE TIREOIDE UMA ABORDAGEM *IN SILICO* UTILIZANDO DADOS DO TCGA E GEO”. O mesmo será traduzido ao inglês e submetido a revista: *Brazilian Archives of Biology and Technology*⁵.

³ <http://revista.unilus.edu.br/index.php/ruep>

⁴ <http://revista.unilus.edu.br/index.php/ruep/article/view/1688>

⁵ <https://www.scielo.br/j/babt/>

ANÁLISE DE GENES DIFERENCIALMENTE EXPRESSOS: UMA ABORDAGEM METODOLÓGICA *IN SILICO*

DIFFERENTIALLY EXPRESSED GENES ANALYSIS: AN *IN SILICO* METHODOLOGICAL APPROACH

Nikael Souza de Oliveira; Natalia Rech; Scheila de Avila e Silva

Universidade de Caxias do Sul

Instituto de Biotecnologia

Laboratório de Biologia Computacional e Bioinformática

RESUMO

Tecidos normais e tumorais apresentam o mesmo genótipo, entretanto diferentes fenótipos. Assim, espera-se encontrar diferentes padrões de expressão quando estes são comparados. Devido ao grande volume de dados gerados na obtenção destes dados e o desafio da união do conhecimento molecular e computacional, bancos de dados públicos têm sido criados para que novas pesquisas sejam realizadas. Desta forma, este trabalho demonstrou uma abordagem metodológica na identificação de genes diferencialmente expressos (DEGs), utilizando repositórios públicos com o auxílio do ambiente R. Para isto, a metodologia se dividiu em quatro etapas: (i) Obtenção dos dados; (ii) Preparação dos dados; (iii) Análise estatística e (iv) Análise dos resultados. Sendo identificados 12 DEGs em câncer de tireoide. Esperamos que este trabalho sirva de guia a projetos futuros na identificação de potenciais DEGs.

Palavras-chave: Câncer; Bancos de Dados; Repositórios públicos; Genes diferencialmente expressos; Abordagem metodológica

ABSTRACT

Normal and tumor tissues present the same genotype, however different phenotypes. In this way, it's expected to find different expression patterns when those comparisons. Due to the large volume of data generated in obtaining these data and the challenge of uniting molecular and computational knowledge, public databases have been created so that new research can be carried out. In this way, this work demonstrated a methodological approach in the identification of differentially expressed genes (DEGs), using public repositories with the R environment. For this, the methodology was divided into four stages: (i) Obtaining data; (ii) Data preparation; (iii) Statistical analysis and (iv) Analysis of results. Being identified 12 DEGs in thyroid cancer. We hope that this work will serve as a guide for future projects in identifying potential DEGs.

Key-words: Cancer; Data bases; Public repositories; differentially expressed genes; methodological approach.

INTRODUÇÃO

Câncer é um grupo de patologias de origem genética, ocasionada devido alterações no funcionamento de alguns genes presentes no organismo, os quais modificam o metabolismo celular normal, causando uma fuga dos padrões originais. Naturalmente as células se dividem, cumprem suas funções e morrem. Entretanto, devido a mutações que ocorrem durante os processos de divisão celular, influenciados também por fatores ambientais, os ciclos normais se alteram e uma única célula pode gerar uma massa que se multiplica descontroladamente, invade tecidos vizinhos e causa danos ao organismo, tornando-se, portanto, a patologia conhecida como câncer (BRUCE, *et al* 2017).

Todas as células de um mesmo organismo apresentam o mesmo genótipo, entretanto diferentes fenótipos, isso ocorre devido a mecanismos de regulação gênica os quais alteram os

genes expressos em cada conjunto celular, formando os diferentes tecidos. Nesse sentido, se entende o câncer como uma patologia de origem genética, a qual foge dos parâmetros celulares normais, e, portanto, apresentam fenótipos distintos do seu tecido de origem (JORDE; CAREY; BAMSHAD, 2017). Desta forma, trabalhos tem demonstrado como diferentes padrões de expressões são encontrados nestes conjuntos, podendo ser estes utilizados como biomarcadores de atividades patológicas, exemplifica-se TULP3 no câncer colorretal (SARTOR; RECAMONDE-MENDONZA; ASHTON-PROLLA, 2019), AGT, SERPINH1 e MMP7 no câncer gástrico (LIU *et al*, 2021), CDC20 e ASPM no câncer de bexiga (XU *et al*, 2019).

Há dois grupos principais de genes relacionados ao câncer: (i) oncogenes e (ii) genes supressores de tumor. O primeiro, antes de se tornarem oncogenes propriamente ditos, são denominados de proto-oncogenes, e atuam nas funções normais das células, como divisão, metabolismo, regulação do ciclo celular e senescência. O segundo grupo compreende genes que atuam na regulação do ciclo celular impedindo a formação de tumores, atuando em etapas de controle e verificação, a qual como exemplo induzem a apoptose células mutadas com potencial de divisão descontrolada, entretanto, estes genes quando mutados perdem sua ação, e com as etapas de controle de qualidade não funcionando corretamente, abre-se espaço para a formação de neoplasias (PIERCE, 2017).

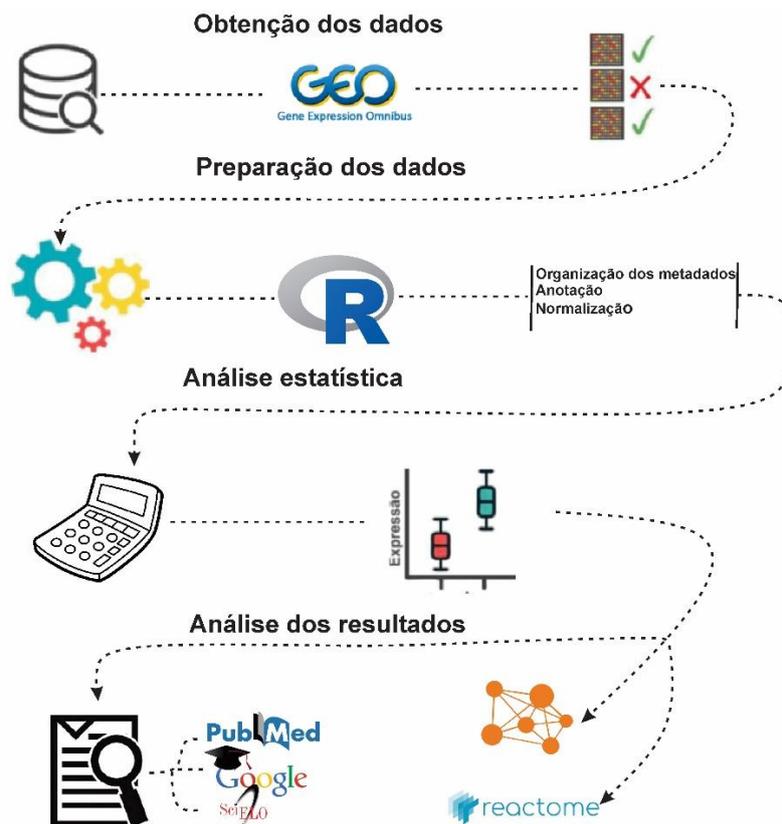
Com o advento das novas tecnologias moleculares, bem como o desenvolvimento das tecnologias computacionais, novas áreas de pesquisa vêm sendo desenvolvidas, e unindo estes dois tipos de conhecimento forma-se a bioinformática. Uma forma de estudo nesta jovem área é com análise de expressão de genes. Desta forma, visando pesquisas globais relacionadas a expressão gênica com intuítos de descobertas que possam auxiliar nos entendimentos destas patologias, bancos de dados têm sido criados permitindo o depósito de dados de expressão de genes. Como exemplo, cita-se: *Gene Expression Omnibus (GEO) database* (<https://www.ncbi.nlm.nih.gov/geo/>) (EDGAR; DOMRACHEV; LASH, 2002) e *The cancer genome atlas* (<https://portal.gdc.cancer.gov/>) (Cancer Genome Atlas Research Network, 2014). Em relação a esses repositórios, o primeiro é voltado a dados de expressão obtidos a partir de análises de *microarray* e o segundo a partir de dados de RNA-seq.

O grande volume de dados gerados a partir das técnicas moleculares que identificam a expressão gênica, o formato dos dados e análises estatísticas adequadas requerem a utilização de ferramentas computacionais de análises, sendo essencial a integração entre o conhecimento molecular e o conhecimento computacional. Está é uma dificuldade encontrada por diversos pesquisadores e programas específicos têm sido desenvolvidos para facilitar estas análises como GEOR2 e GEPIA (TANG *et al*, 2017). Ainda assim, devido a manipulação para análises específicas são utilizadas linguagens de programação como R (R Development Core Team, 2021) e python (ROSSU; DRAKE, 1995). Nesse sentido, o presente trabalho possui como objetivo demonstrar um passo a passo para realização da coleta de dados e análise de expressão diferencial de genes em câncer utilizando o banco de dados GEO, bem como sua manipulação de dados no ambiente R.

PROCEDIMENTOS PARA ANÁLISE *IN SILICO* DE EXPRESSÃO DIFERENCIAL DE GENES

Para realização da análise de genes diferencialmente expressos, este trabalho irá se dividir em 4 etapas: (i) Obtenção dos dados pelos bancos de dados GEO; (ii) Manipulação e preparação dos dados; (iii) Análise estatística para obtenção dos genes diferencialmente expressos; (iv) Análise dos resultados obtidos, identificando os genes, suas funções e interações, bem como relatos na literatura associados a neoplasias. As etapas II e III são realizadas com auxílio do ambiente R. A figura 1 demonstra um fluxograma metodológico visualmente.

FIGURA 1- fluxograma metodológico.



Fonte: Elaborado pelos autores.

Obtenção dos dados

O GEO é um repositório público internacional que armazena e disponibiliza dados de *microarray*. Este possui três principais objetivos; (i) fornecer um banco de dados robusto e versátil para armazenar dados de alto rendimento; (ii) oferecer procedimentos de apresentação simples e formatos que suportam depósitos de dados completos da comunidade de pesquisa; (iii) fornecer mecanismos de fácil utilização que permita que o usuário pesquise, localize, revise e faça *download* de estudos e perfis de expressões dos genes (EDGAR; DOMRACHEV; LASH, 2002).

O tipo de câncer escolhido para utilização neste trabalho é o câncer de tireoide. Sendo necessária esta definição para demonstrar a forma de busca no banco. A pesquisa no banco de dados GEO se dá pela interface gráfica do website, utilizando de uma *string* de busca, a qual para a presente demonstração será: “*Thyroid Cancer*”.

O banco de dados GEO pode ser encontrado pelo link <https://www.ncbi.nlm.nih.gov/geo/> e a *string* de busca deve ser digitada no campo “*search*” entre aspas (figura 2a). Em caso de busca de mais de um tipo de câncer este pode ser realizado em uma mesma pesquisa, separados por seus diferentes conjuntos de aspas e com a palavra OR no meio. Como o GEO apresenta uma vasta coleção de dados de diferentes espécies e diferentes tipos de análises filtros devem ser selecionados na plataforma. No trabalho em questão estes serão tipo de organismo: *Homo sapiens* e tipo de estudo: *Expression profiling by array* (figura 2b).

Figura 2 – Interface gráfica GEO.

NCBI Resources How To Sign in to NCBI
GEO Home Documentation Query & Browse Email GEO

a
Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Thyroid Cancer Search

Getting Started	Tools	Browse Content
Overview	Search for Studies at GEO DataSets	Repository Browser
FAQ	Search for Gene Expression at GEO Profiles	DataSets: 4348
About GEO DataSets	Search GEO Documentation	Series: 194232
About GEO Profiles	Analyze a Study with GEO2R	Platforms: 24766
About GEO2R Analysis	Studies with Genome Data Viewer Tracks	Samples: 6541742

b
National Library of Medicine
National Center for Biotechnology Information Log in

GEO DataSets GEO DataSets [("Thyroid Cancer") AND "Homo sapiens"[porgn: __bid9606]] Search Help

Entry type Summary 20 per page Sort by Default order Filters: Manage Filters

DataSets (2)
Series (52)
Samples (0)
Platforms (0)

Search results
Items: 1 to 20 of 55

Organization
Customize ...

Study type
Expression profiling by array
Methylation profiling by array
Customize ...

Author

Filters activated: Expression profiling by array. Clear all to show 1174 items.

1. [Transcriptomic landscape of TIMP3 oncosuppressor activity in thyroid carcinoma](#)
(Submitter supplied) In thyroid tumors TIMP3 methylation, and hence its downregulation, has been reported (Hoque MO et al, J Clin Endocrinol Metabolism 2005) and also found significantly associated with several aggressive features of PTC, including extrathyroidal invasion, lymph node metastasis, multifocality, advanced tumor stages and BRAF mutation (Hu S et al, Int J Cancer 2006). In a previous work, in addition to confirming

Top Organisms [Tree]
Homo sapiens (55)
synthetic construct (1)

Find related data
Database: Select
Find items

Fonte: EDGAR; DOMRACHEV; LASH, 2002

A busca apresenta diversas amostras depositadas, e estas devem ser analisadas individualmente e selecionadas para análise, para isto deve-se definir critérios de inclusão e exclusão, os quais devem ser baseados no objetivo do trabalho do pesquisador. Cada conjunto de dados apresenta um código de acesso determinado por GSE seguido por um numeral, estes devem ser anotados para posterior utilização. Ao mesmo tempo, cada GSE apresenta um GPL também seguido por um numeral, a qual é a plataforma utilizada na realização do experimento de *microarray*, e contém a marca e o modelo do chip. Este também deve ser anotado, pois será necessário durante o processo de obtenção dos dados utilizando o ambiente R.

Visando a obtenção de genes diferencialmente expressos *in vivo* que possam ser utilizados como potenciais biomarcadores os critérios de inclusão e exclusão para cada conjunto de dados devem ser definidos. Como exemplo, os critérios utilizados para este trabalho foram: (i) Cada conjunto de amostra deve possuir amostras normais e tumorais; (ii) Cada conjunto de amostras deve conter apenas amostras de tecidos biopsiados; (iii) Cada conjunto de amostras deve conter no mínimo 10 amostras. Após a seleção dos conjuntos de dados com seus respectivos GSEs anotados deve-se realizar o download das amostras com os dados de expressão e os respectivos metadados, utilizando como auxílio o pacote GEOquery (DAVIS; MELTZER, 2007), no ambiente R (Material suplementar 1).

PREPARAÇÃO DOS DADOS

Os dados baixados devem ser preparados utilizando o ambiente R (neste artigo foi utilizada a versão 4.1.1). Cada gene é representado por uma sonda, sendo um dos processos a anotação gênica a partir desta sonda. Os conjuntos de dados de expressão são organizados em arquivos diferentes do seu conjunto de metadados com as informações clínicas dos pacientes. Este conjunto de metadados carregam informações importantes, as quais podem ser utilizadas com bases em novas pesquisas relacionadas aos aspectos clínicos dos pacientes. Quando estes apresentam-se nos objetivos da pesquisa e serão utilizados, faz-se importante a associação das informações dos diferentes arquivos. Os quais podem ser assimilados, utilizando o ambiente R, em um mesmo dado para realização de análises posteriores.

Para realização da análise estatística deve-se realizar a normalização dos dados (material suplementar 2), a qual é uma prática de organização e transformação dos valores de modo que estes possam ser comparados. Existem diversos tipos de normalização de dados, não havendo um consenso claro de qual o melhor método a ser utilizado. A transformação logarítmica auxilia na modelação do erro visto que deste modo apresenta alterações proporcionais e não aditivas, da mesma forma auxilia nas comparações de pequenas alterações. Ao mesmo tempo, pode-se pensar na amplificação do cDNA, utilizado nas técnicas moleculares de análise, uma vez que há resultados exponenciais sem seguir o padrão de distribuição normal, e a transformação logarítmica na base 2 é a que mais a aproxima (AMBROISE *et al*, 2011). Além disso, a transformação em LOG2 é amplamente citada para trabalhos com microarray (SINHA *et al*, 2022; STEKEL, 2003; VALIZADEH *et al*, 2022). O pacote limma no ambiente R possui ferramentas que auxiliam neste processo (RITCHIE *et al*, 2015).

ANÁLISE ESTATÍSTICA

As diversas metodologias de comparação de médias são amplamente discutidas referente as análises relacionadas a comparação de dados de expressão gênica, não havendo um consenso claro de qual a melhor abordagem analítica. Entende-se que o fator de probabilidade associado, presente na estatística bayesiana se adapta melhor a incerteza dos dados biológicos e permite responder de forma mais apropriada as questões biológicas (JIMENEZ *et al*, 2021).

Para comparação entre amostras tumorais e normais é recomendada a estatística bayesiana com correção da falsa taxa de descoberta de Benjamini e Hochberg para o valor p. (SINHA *et al*, 2023). Os pontos de cortes definidos para este trabalho são valor $p < 0,05$ e Log_2 Fold Change $> |1|$. Sendo o valor p a significância estatística, e log_2 Fold Change o quanto a expressão do gene em câncer varia da normal, sendo estes valores negativos para os genes menos expressos e os valores positivos genes mais expressos. A definição dos pontos de corte cabe ao pesquisador, sendo que para o Log_2 Fold Change varia entre 1 e 2 na maior parte dos trabalhos encontrados na literatura (SINHA *et al*, 2023; SONG *et al*, 2023; VALIZADEH *et al*, 2022; LIU *et al* 2023; ALAM; SULTANA *et al*, 2022). Definiu-se neste trabalho o valor de 1 para obtenção de um maior número de resultados (material suplementar 3).

ANÁLISE DOS RESULTADOS

Os diferentes GSEs analisados resultam em diferentes conjuntos de genes diferencialmente expressos. Deste modo deve ser realizado a intersecção dos conjuntos de resultados utilizando o diagrama de *venn* (Material suplementar 4). A análise dos resultados se dá por uma vasta pesquisa na literatura, utilizando-se de bancos de dados de artigos como PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), e bancos de dados de exploração gênica como UNIPROT (<https://www.uniprot.org/>) ou CARDGENE (<https://www.genecards.org/>).

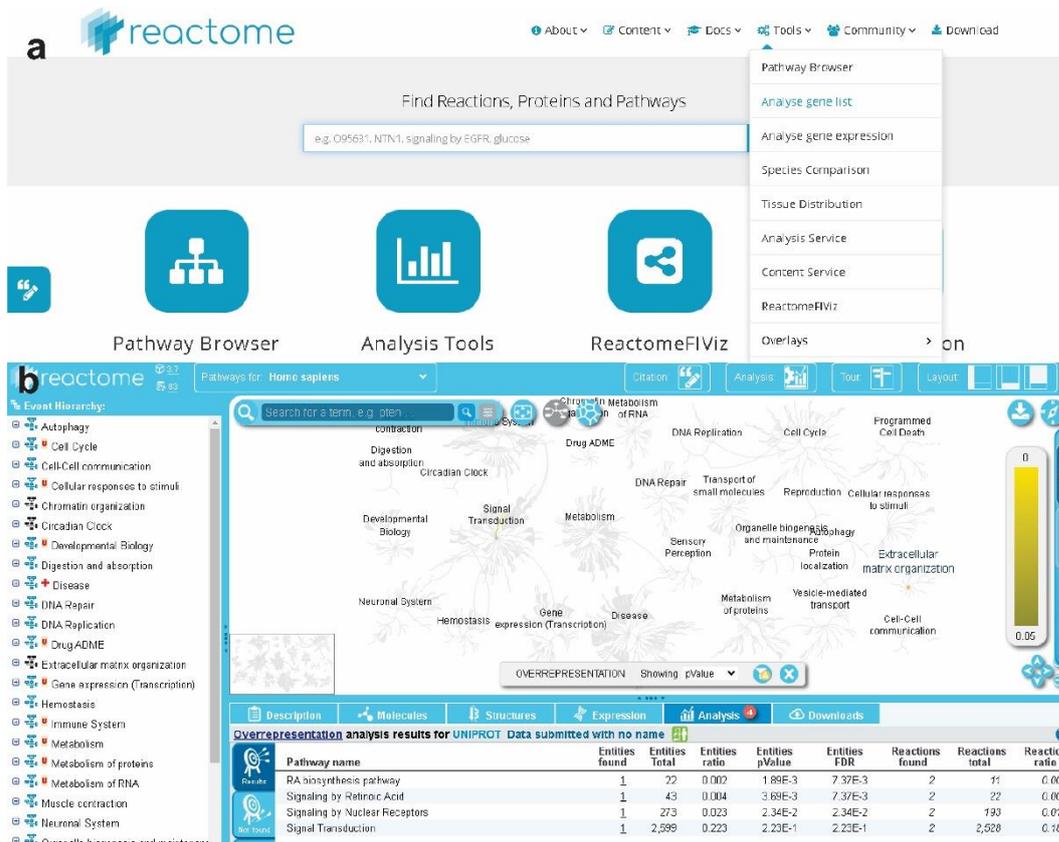
A partir disso, pode-se inferir a participação destes genes em processos biológicos normais, suas alterações relacionadas a atividades neoplásicas, bem como comparar se os resultados encontrados em um determinado tipo de câncer também já foram descritos em outros tecidos. Vale ressaltar que alguns destes genes diferencialmente expressos também são utilizados como alvos terapêuticos em determinadas neoplasias, podendo ser assimilados este tratamento já em uso a um novo tipo de câncer.

Uma forma de compreender o contexto metabólico destes genes é relacionando os processos biológicos e as vias metabólicas que estes atuam, sendo uma ferramenta disponível para análise a plataforma Reactome (<https://reactome.org/>). Esta é uma ferramenta de código aberto, curada manualmente e revisada por pares, a qual tem como objetivo fornecer caminhos em bioinformática para a visualização de dados, interpretação e análise de conhecimento. A

plataforma permite a busca por lista de genes e indica os processos biológicos que estes participam, bem como demonstram em que etapa das rotas eles influenciam (GRISS *et al*, 2020).

A busca pode ser realizada na interface gráfica do website, selecionando o item ferramentas e após o item *Analyse gene list* (figura 3a). Em seguida deve-se realizar a busca utilizando o nome dos genes, sendo demonstrados os resultados referentes as vias metabólicas e os processos biológicos em que estes genes estão envolvidos (figura 3b). Os resultados obtidos a partir desta plataforma é dado em forma de interface gráfica demonstrando os processos biológicos relacionados, e em forma de relatório final, onde apresenta uma breve descrição literária do envolvimento dos genes pesquisados nas atividades celulares.

Figura 3 – Interface gráfica do banco de dados Reactome.



Fonte: GRISS *et al*, 2020

Outra forma de verificar os conceitos biológicos dos genes identificados é pela construção de redes de interação proteína-proteína. Uma análise importante deste tipo de rede está relacionada a lista de genes, por vezes os genes diferencialmente expressos atuam em processos biológicos semelhantes, e interagem com os mesmos genes, os quais nem sempre são diferencialmente expressos, entretanto tem papéis importantes no câncer. Pode-se entender, de forma a exemplificar, a mutação de um gene de fator transcricional, o qual não apresentará alteração na expressão por si só, mas causará a diferença de expressão em algum outro gene, ou em um conjunto de genes. Sendo importante a identificação neste contexto de interação.

Desta forma, pode-se inferir importantes interações biológicas. Uma ferramenta com esta finalidade é IntAct (<https://www.ebi.ac.uk/intact/home>), que permite a busca na interface gráfica por lista de genes (figura 4a), retornando suas respectivas interações (figura 4b) (ORCHAD *et al*, 2013). IntAct é um banco de dados público, o qual apresenta um amplo repositório relacionado a interações moleculares. Possui como principal objetivo aproveitar ao máximo os dados de interação pública e facilitar a adoção de padrões em biologia molecular. Além da interface gráfica, IntAct pode ser trabalhado como extensão dentro da ferramenta cytoscape, permitindo a visualização de redes maiores e com maior nível de complexidade.

Figura 4 – Interface gráfica banco de dados Intact.

The image shows the IntAct Molecular Interaction Database interface. At the top, there is a navigation bar with links for Home, Download, About, Documentation, and Feedback. The main heading is 'IntAct Molecular Interaction Database'. Below this, there is a description of the database and a featured dataset for COVID-19. The search bar is used to search for 'CRABP1, ... SGK223'. The results are displayed as an interaction network with various filters and layout options.

Fonte: ORCHAD *et al*, 2013

Vale ressaltar que existe uma ampla gama de ferramentas e bancos de dados disponíveis na internet e descritas na literatura. Cabe ao pesquisador realizar a busca pelas plataformas que lhe interessam e atendem seus objetivos. Aqui, foi demonstrado apenas alguns exemplos de ferramentas que podem auxiliar na pesquisa e na compreensão do aspecto biológico dos genes identificados como diferencialmente expressos. Ao mesmo tempo, ressaltase a importância da busca de trabalhos relacionados já descritos em atividades neoplásicas.

ESTUDO DE CASO USANDO COMO EXEMPLO O CÂNCER DE TIREOIDE

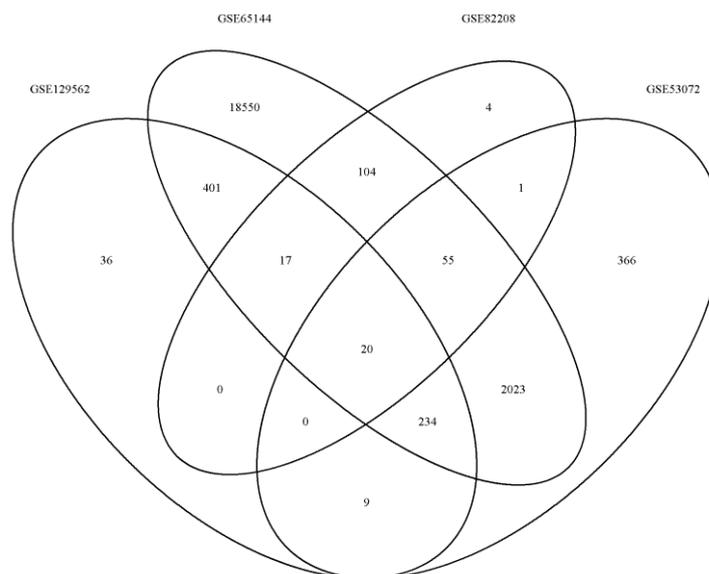
Ao realizar a busca no banco de dados GEO, a partir da *string* “*Thyroid Cancer*”, se obteve como resultado preliminar 55 conjuntos de dados. Estes foram submetidos a investigação manual e aplicação dos critérios de exclusão e inclusão, retornando em 4 conjuntos de dados, os quais são identificados pelo número de acesso: GSE82208, GSE65144, GSE53072 e GSE129562. Cada GSE foi submetido individualmente as etapas de obtenção dos dados no ambiente R, preparação dos dados e análise estatística, processos descritos na seção anterior. A partir disso, para cada GSE obteve-se diferentes conjuntos de genes diferencialmente expressos, demonstrados na tabela 1.

Tabela 1 – Número de genes diferencialmente expressos encontrado em cada conjunto de dados.

IDENTIFICAÇÃO	DEGS
GSE82208	257
GSE65144	6922
GSE53072	3206
GSE129562	884

Nota-se diferença no número de genes diferencialmente expressos encontrados nos diferentes conjuntos de dados. Esse resultado é esperado, visto que os dados de expressão são elaborados por diferentes pesquisadores em diferentes laboratórios com diferentes chips de *microarray*. Ao mesmo tempo, entende-se o organismo biológico em um estado dinâmico sobre as diferentes condições ambientais, o que pode alterar aspectos fenotípicos de expressão por si só. Desta forma, demonstra-se a necessidade dos cruzamentos dos dados para identificação dos DEGs presentes em todos os conjuntos. Para isso, com auxílio do R, foi realizado a intersecção dos resultados obtidos para cada GSE em um diagrama de *venn* (Figura 5). A intersecção dos 4 estudos resultou em 20 genes: CRABP1, GLT8D2, TNFRSF11B, FOS, SLC25A15, MT1F, LRP1B, DNALI1, WSCD2, SDPR, CSGALNACT1, CPQ, SDC2, PAX8, ITM2A, SELENBP1, SGK223, IER2, ID1, ID3.

Figura 5 – Diagrama de venn.



Fonte: Elaborado pelos autores.

Visto que o presente trabalho tem como objetivo uma abordagem metodológica referente a expressão diferencial de genes utilizando bancos de dados públicos, optou-se por 2 genes aleatórios, para a realização de uma breve discussão: ID3 e LRP1B. Ressalta-se que ambos os genes apresentaram menor valor de expressão no tecido neoplásico quando comparado ao tecido normal.

ID3 é um inibidor de proteína de ligação ao DNA. A qual atua como regulador transcricional que regula negativamente os fatores de transcrição básicos hélice-alfa-hélice (HLH), formando heterodímeros que inibem sua ligação ao DNA e atividade transcricional, implicando em processos celulares como crescimento, senescência, diferenciação e apoptose (DEED e *t al*, 1993). Sendo já relatado na literatura sua menor expressão em leucemia mieloide

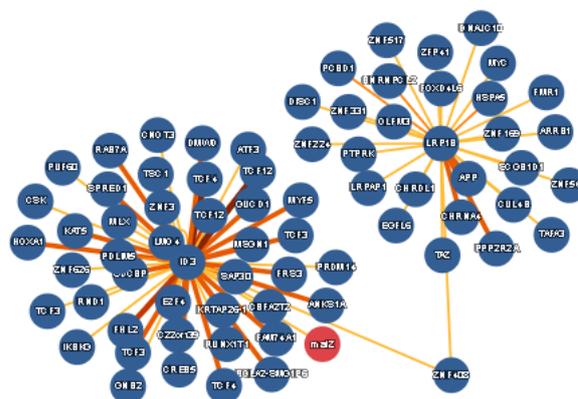
aguda (ZHAO *et al*, 2022) e relacionado a um circuito de regulação transcricional sobre a proteína dedo de zinco ZNF148 e ao proto-oncogene MYC em cânceres de mama agressivos (KIM *et al*, 2022). A busca no banco de dados reatome demonstrou ID3 envolvido em processos nucleares de transcrição e sinalização.

LRP1B é receptor de lipoproteína de baixa densidade localizado na membrana celular. Desempenhando diversos papéis na função e desenvolvimento celular normal, como transporte e sinalização celular, sendo considerado também um suposto supressor tumoral (PRINCIPLE *et al*, 2021). Trabalhos relacionaram a frequente mutação deste gene associada a câncer gástrico (KUNG *et al*, 2023) e adenocarcinoma de pulmão (RAO *et al*, 2023). Além disso, da mesma forma que este trabalho, sua menor expressão já havia sido relatada em câncer gástrico (MIAO *et al*, 2022) e câncer de bexiga (CONCONI *et al*, 2022), demonstrando seu potencial como biomarcador ao diagnóstico desta patologia.

A busca no bando de dados Reactome resultou na identificação de processos biológicos relacionados a ID3, entretanto não foram encontrados resultados relacionados ao LRP1B. ID3 foi descrito como indutor transcricional de fator de crescimento nervoso, relacionado a eventos nucleares como quinase e ativador de fatores de transcrição, sinalização a partir de neurotrofinas tirosina quinase e sinalização por receptores de tirosina quinase. De modo geral, ID3 atua em processos de sinalização essencial ao metabolismo normal da célula, geralmente em células do sistema nervoso, entretanto também relacionado a outros tipos celulares, desta forma demonstrando que sua alteração implica em nesses processos básicos, sendo possível relacionar as alterações demonstradas em câncer (GRISS *et al*, 2020).

A busca no banco de dados intAct demonstrou uma interação mutua entre ID3 e LRP1B com uma proteína dedo de zinco ZNF408, (figura 6) porém não foi encontrado trabalhos relacionando este gene e câncer, apenas com outras patologias relacionadas a retina (AVILA-FERNANDEZ *et al*, 2015; CHEN *et al*, 2022). Entretanto, entende-se ZNF408 como uma proteína pertencente à família PRDM, a qual regulam a expressão gênica por modificação de histonas ou por interação de proteínas (KARJOSUKARSO *et al*, 2020). Permitindo a especulação de que esta proteína dedo de zinco, de alguma maneira, pode afetar a expressão destes genes e estar relacionada a suas diferenças de expressão. Desta forma, se sugere a necessidade de mais estudos moleculares *in vivo* e *in vitro* para compreensão do mecanismo moleculares de ZNF408 e suas interações com ID3, LRP1B e com outros genes, identificando talvez no futuro uma associação entre estes genes e câncer de tireoide.

Figura 6 – Rede de interação gerada a partir da busca no banco de dados IntAct



Fonte: ORCHAD *et al*, 2013

CONCLUSÃO

Câncer é uma patologia que assola a humanidade causando óbitos e prejuízos a qualidade de vida dos portadores. Estudos relacionados ao câncer, se fazem de suma importância em um aspecto que busca processos terapêuticos que possibilitem a cura e preserve a dignidade dos pacientes. Diversas pesquisas têm-se voltado a esta área auxiliando na compreensão desta patologia. Atualmente, entende-se cada câncer como uma patologia ímpar, de origem genética com expressões gênicas diferentes dos tecidos normais. Desta forma, como já demonstrado em trabalhos descritos na literatura, se busca, com base nessas diferenças de expressões identificar padrões moleculares que sirvam como potenciais biomarcadores a diagnóstico e prognóstico, bem como potenciais alvos terapêuticos.

Bancos de dados públicos como o GEO e TCGA possuem conjuntos de dados, com enorme potencial de análise para novas descobertas relacionadas ao câncer. Estes permitem a deposição de dados gênicos, os quais poderão ser analisados por outros pesquisadores em novas pesquisas, permitindo maiores descobertas. Ao mesmo tempo a vasta literatura e repositórios biológicos possibilitam novas inferências moleculares, os quais podem apontar biomoléculas potenciais a realização de novos estudos *in vitro* e *in vivo*.

Uma grande dificuldade observada no cenário atual é a integração do conhecimento genômico e do conhecimento computacional, bem como das análises estatísticas complexas realizadas em dados de expressão. Nota-se o crescimento de bancos de dados e de ferramentas de análises deste material. Entretanto, em grande maioria, mantém-se a utilização da linguagem de programação em R e em alguns casos python, visto que estas ferramentas permitem uma manipulação de dados mais dinâmica e individual. Desta forma, o presente trabalho aborda um contexto metodológico que visa exemplificar a análise de expressão diferencial de genes em câncer utilizando o banco de dados GEO e o ambiente R.

As etapas descritas no corpo do texto são divididas em obtenção dos dados; preparação dos dados, análise estatística e análise dos resultados. Os códigos utilizados para a realização dos mesmos se encontram no material suplementar. Para exemplificar este trabalho foi utilizado o câncer de tireoide, resultado em 20 genes diferencialmente expressos encontrados em todos os conjuntos analisados. Foram selecionados 2 genes para uma breve descrição e demonstrado com base na interação molecular, sua relação com a proteína dedo de zinco ZNF408.

O presente trabalho apresentou uma abordagem metodológica relacionada a identificação de genes diferencialmente expressos utilizando o banco de dados público GEO, bem como uma breve discussão acerca de dois genes selecionados, demonstrando seus aspectos biológicos e identificações prévias em outros estudos sobre câncer. Espera-se que este trabalho possa servir de auxílio a pesquisas futuras, e que seja utilizado como guia a identificação de novos genes diferencialmente expressos.

REFERÊNCIAS

- ALAM, S. *et al.* Gene expression profile analysis to discover molecular signatures for early diagnosis and therapies of triple-negative breast cancer. **Frontiers**. 9. 2022.
- AMBROISE, J. *et al.* Impact of the spotted microarray preprocessing method on fold-change compression and variance stability. 12:413. 2011.
- AVILA-FERNANDEZ, A; PEREZ-CARRO, R.; CORTON, M. Whole-exome sequencing reveals ZNF408 as a new gene associated with autosomal recessive retinitis pigmentosa with vitreal alteration. **Hum Mol Genet**. 24. 4037-4048. 2015.
- BRUCE, A. *et al.* **Fundamentos de Biologia Celular**. 4.ed. Porto Alegre: Artmed, 2017.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. **Nature**. 513. 202-209. 2014.

- CHEN, C. *et al.* Long-term clinical prognosis of 335 infant single-gene positive FEVR cases. *22(1)*. 2022.
- CONCONI, D. *et al.* Analysis of copy number alteration in bladder cancer stem cells revealed a prognostic role of LRP1B. **World J Urol.** 40. 2267-2273. 2022.
- DAVIS, S.; MELTZER, P. GEOQUERY: a bridge between the Gene Expression Omnibus (GEO) and Bioconductor. **Bioinformatics.** 23. 1846-11847. 2007.
- DEED, R. W. *et al.* An immediate early human gene encodes an id-like helix-loop-helix protein and is regulated by protein kinase C activation in diverse cell types. *8(3)*: 599-607. 1993.
- EDGAR, R.; DOMRACHEV, M.; LASH, A. E. Gene expression Omnibus: NCBI gene expression and hybridization array data repository. **Nucleic Acids Research.** 30. 207-210. 2002.
- GRISS, J. *et al.* Efficient Multi-Omics Comparative Pathways Analysis. **Mol cell proteomics.** 2020.
- JIMÉNEZ, V.; G.; MARTÍ-GÓMEZ, C.; ANGEL, M. Bayesian inference of gene expression. IN: HELDER, I. N, editor. **Bioinformatics.** Brisbane(AU): Exon publications. 2021.
- JORDE, L. B.; CAREY, J.C.; BAMSHAD, M.J. **Genética Médica.** 5.ed. Rio de Janeiro: GEN. 2021.
- KARJOSULKARO, D. W. *et al.* Modeling ZNF408-associated FEVR in Zebrafish Results in abnormal retinal vasculature. **Biochemistry and molecular biology.** 2020.
- KIM, M. *et al.* A MYC-ZNF148-ID1/3 regulatory axis modulating cancer stem cell traits in aggressive breast cancer. **Oncogenesis.** 60. 2022.
- LIU, W. *et al.* Bioinformatics analysis of key biomarkers for bladder cancer. **Biomedical reports.** 18. 2023.
- MIAO, J. *et al.* Integrative analysis of the proteome and transcriptome in gastric cancer indentified LRP1b as a potential biomarker. **Biomark med.** 15. 1101-1111. 2022.
- ORCHAD, S.; AMMARI, M.; ARANDA, B. *et al.* (2013). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction database. **Nucleic Acids Res.** 42.
- PIERCE, B. A. (2017). **Genética um enfoque conceitual.** Rio de Janeiro: Guanabara Koogan.
- PRINCIPE, C. *et al.* LRP1B: A giant lost in cancer translation. **Pharmaceuticals.** 14. 836. 2021.
- R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. 2020.
- RAO, W. *et al.* Frequently mutated genes in predicting the relapse of stage I lung adenocarcinoma. **Clin transl oncol.** 2023.
- RITCHIE, M. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. **Nuclei Acids Research.** 43. 2015.
- ROSSUM, V.; DRAKE, J. F. L. **Python reference manual.** Centrum voor Wiskunde em informatica Amsterdam. 1995.
- SINHA, B. K. *et al.* (2022). Gene expression. profiling elucidates cellular responses to NCX3030 in human ovarian tumor cells: Implications in the mechanisms of action of NCX4040. **Cancers Basel.** 15. 2022.
- SONG, Y. *et al.* Comprehensive molecular analyses of notch pathway-related genes to predict prognosis and immunotherapy response in patients with gastric cancer. **Journal of oncology.** 2023.
- STEKEL, D. **Microarray bioinformatics.** New York: Cambridge. 2003.
- TANG, Z. *et al.* Epigenetic Biomarkers of Breast Cancer Risk, Across the breast cancer prevention continuum. **Advances in experimental medicine and biology.** 2017.

VALIZADEH, S. *et al.* Upregulation of miR-142 in papillary thyroid carcinoma tissues: a report based on in silico and in vitro analysis. **Molecular Biology Research Communications**. 11. 133-141. 2022.

ZHAO, Q.*et al.* Comprehensive analysis of ID genes reveals the clinical and prognostic value of ID3 expression in acute myeloid leukemia using bioinformatics identification and experimental validation. **BMC Cancer**. 22(1). 2022.

GENES DIFERENCIALMENTE EXPRESSOS EM CÂNCER DE CAVIDADE ORAL E DE TIREOIDE UMA: ABORDAGEM *IN SILICO* UTILIZANDO DADOS DO GEO E TCGA

Nikael Souza de Oliveira¹

<https://orcid.org/0000-0003-4713-4527>

Nathalia Rech¹

<https://orcid.org/0000-0001-5550-7245>

Ivaine Tais Sauthier Sartor

<https://orcid.org/0000-0002-8775-6622>

Scheila de Avila e Silva¹

<https://orcid.org/0000-0002-3472-3907>

¹Universidade de Caxias do Sul; Instituto de Biotecnologia; Laboratório de Biologia Computacional e Bioinformática; Caxias do Sul; Rio Grande do Sul; Brasil.

HIGHLIGHTS

- Utilização de dados do GEO e TCGA para identificação de genes diferencialmente expressos;
- Identificação de 4 genes diferencialmente expressos para o câncer de cavidade oral;
- Identificação de 12 genes diferencialmente expressos para o câncer de tireoide.

RESUMO: Os cânceres de cabeça e pescoço compreendem as regiões acima da clavícula, com exceção de pele, linfonodos, olhos, esôfago e traqueia. Estes tipos de cânceres além dos malefícios usuais, devido a particularidade anatômica, causam prejuízos físicos, fisiológicos e psicossociais. A evolução das tecnologias moleculares tem permitido a obtenção de um volume de dados cada vez maior, os quais vêm sendo depositados em repositórios públicos como *The cancer genome atlas* (TCGA) e *Gene expression omnibus database* (GEO *database*). O presente trabalho realizou a análise da expressão diferencial de genes em cânceres de cavidade oral e câncer de tireoide. Com o auxílio do ambiente R, os dados foram retirados do banco de dados GEO *database* e TCGA, preparados e analisados. Obteve-se 4 genes diferencialmente expressos (DEG's) para o câncer de cavidade oral: LOXL2; FAT1; SOD3 e DPT, e 12 DEG's para o câncer de tireoide: CRABP1; CSGALNCT1; WSCD2; ITM2A; ID3; PAX9; SELENBP1; TNFRS11B; SDPR; LRP1B; MT1F e SGK233. Todos os genes identificados aqui apresentam descrições anteriores relacionadas a cânceres, fundamentando os resultados obtidos neste trabalho. A análise da curva de sobrevida demonstrou, para o câncer de cavidade oral uma maior expressão de FAT1 relacionada a uma menor sobrevivência ao longo do tempo. Para o câncer de tireoide, menores valores de expressão dos genes ITM2A e ID3 apresentaram menor sobrevivência. Este trabalho identificou 4 genes como potenciais biomarcadores de diagnóstico e prognóstico. Esperamos que este trabalho possa contribuir para pesquisas futuras e

auxiliar na escolha de biomarcadores de diagnóstico e prognóstico mais específicos para tumores de cavidade oral e tireoide.

Palavras-chaves: Câncer de cabeça e pescoço; Câncer de cavidade oral; Câncer de tireoide; Bancos de dados; Genes diferencialmente expressos.

INTRODUÇÃO

Câncer é uma patologia de origem genética ocasionada por alteração em alguns genes, os quais fazem que as células alterem seu funcionamento normal, formando uma massa neoplásica que se multiplica descontroladamente invade tecidos vizinhos e causa danos ao organismo [1]. Câncer de cabeça e pescoço é um termo utilizado como denominação a um grupo de cânceres que ocorrem acima da região da clavícula, com exceção do cérebro, pele, linfonodos, olhos esôfago e traqueia [2-4]. Os principais sítios anatômicos presentes neste tipo de câncer são: cavidade oral, faringe, laringe, cavidade nasal, seios paranasais, ouvidos e tireoide. Além da divisão por sítios anatômicos, este agrupamento de cânceres também é dividido por tipo histológico, sendo o carcinoma de células escamosas o mais comum, responsável por 90% dos casos, seguido pelo adenocarcinoma [2,5-7].

Segundo dados da Organização Mundial da Saúde (WHO), estima-se que em 2020 o câncer de cabeça e pescoço representou em todo o planeta cerca de 1.518.133 novos casos, 8,4% entre todos os tipos de cânceres, e 5.510.771 mortes, 5,2% entre todos os tipos de cânceres. No Brasil, estima-se que a incidência deste tipo de câncer em 2020 foi de 57.633 novos casos, 10,35% entre todos os tipos de cânceres e 15.506 mortes, 6,03% entre todos os tipos de cânceres. Destes, destes 586.202 (3,23%) são de novos cânceres de tireoide, e 43.646 óbitos devido a progressão da patologia em todo o mundo. Enquanto no Brasil, surgiram 30.607 (5,49%) novos casos e 1081 (0,42%) novos óbitos [4].

Este grupo de cânceres afetam o indivíduo de forma física, fisiológica e psicossocial, acarretando em problemas de autoestima, comunicação, perda de autonomia e dificuldade na alimentação. Dependendo do estágio em que o câncer é diagnosticado, o tratamento tende a ser agressivo e desfigurante, sendo necessário a utilização de sondas para alimentação, traqueostomias e em alguns casos perda de parte da face [5,8,9]. Estima-se que no Brasil, 70% dos pacientes com câncer de cabeça e pescoço são diagnosticados em estágios avançados, diminuindo a chance de cura e a qualidade de vida durante o tratamento. Entretanto, quando o diagnóstico ocorre de forma precoce, a chance de cura atinge em torno de 80% [9]. Visto isto, ressalta-se a importância de estudos que possibilitem o diagnóstico precoce.

A evolução das tecnologias moleculares e computacionais tem gerado cada vez mais informações biológicas. Neste sentido bancos de dados públicos como *Gene Expression Omnibus* (GEO) database [10] e *The Cancer Genome Atlas* (TCGA) [11] apresentam uma gama de dados relacionados a câncer, depositados por pesquisadores para realização de novas pesquisas. Um dos tipos de dados encontrados nestes bancos são dados de expressão gênica biopsiados de tecidos normais e tumorais. Sendo que o GEO armazena principalmente dados obtidos a partir de *microarray* e o TCGA principalmente dados advindos de metodologias de RNA-seq.

Diversos trabalhos têm sido realizados utilizando os conjuntos de dados encontrados nestes bancos para identificação de potenciais biomarcadores para diagnóstico e prognóstico, bem como alvos terapêuticos. XU e colaboradores em 2020 demonstraram em seu trabalho a identificação de CDC20 e ASPM como potenciais biomarcadores ao câncer de bexiga [12]. Zheng, Min & Zhou em 2021 identificaram ADCY8, ADDR2, CALCA, GCG, GNGT1 e NPSR1 como potenciais biomarcadores ao adenocarcinoma de pulmão [13]. Da mesma forma, Sartor e colaboradores em 2014

descreveram TULP3 como potencial biomarcador de prognóstico do adenocarcinoma ductal pancreático [14].

Trabalhos relacionados a estes bancos de dados e câncer de cabeça e pescoço também têm sido descritos na literatura. Ye e colaboradores em 2021 utilizam 1 único conjunto de dados do banco de dados GEO e os dados no repositório do TCGA para identificação de genes diferencialmente expressos em carcinoma espinocelular de cabeça e pescoço [15]. No carcinoma oral de células escamosas, a partir de dados disponibilizados no GEO e TCGA há relatos dos genes *PLAU*, *CLDN8* e *CDKN2A* como potenciais genes preditores de sobrevida global [16]. Da mesma forma, WAN e colaboradores em 2020 utilizaram 4 conjuntos de dados presentes no banco de dados GEO e os dados presentes no TCGA para análise de expressão diferencial em câncer de tireoide obtendo 8 candidatos: *FN1*, *CCND1*, *CDH2*, *CXCL12*, *MET*, *IRS1*, *DCN* e *FMOD* [17].

Desta forma, utilizando como base trabalhos já descritos na literatura, o presente estudo possui como objetivo a identificação de genes diferencialmente expressos entre tecidos normais e tumorais de câncer de cabeça e pescoço, mais especificamente, câncer de tireoide e cânceres de cavidade oral, utilizando utilizando múltiplos presentes no banco de dados GEO e TCGA a fim de contribuir para a descoberta de potenciais novos potenciais biomarcadores moleculares.

MATERIAIS E MÉTODOS

Devido a particularidades do número amostral dos dados encontrados no banco GEO, as análises aqui realizadas foram divididas em 2 grupos: cânceres de cavidade oral e câncer de tireoide. Sendo 5 conjuntos de dados obtidos no GEO referente ao primeiro grupo e 4 do segundo. Os conjuntos de dados para câncer de cavidade oral podem ser identificados pelo número de acesso: *GSE38517*, *GSE58911*, *GSE13601*, *GSE14648*, *GSE83519*. Os conjuntos de dados para câncer de tireoide são identificados pelo número de acesso: *GSE82208*, *GSE65144*, *GSE53072* e *GSE129562*. Além disso, a partir do TCGA, foram adquiridas uma base de dados ampla/compreensiva para câncer de cavidade oral e outra para câncer de tireoide, denominados neste trabalho de TCGA-HNC e TCGA-THC.

Os dados foram preparados utilizando o ambiente R. Para os dados advindos do banco GEO, foi necessária a transformação em Log_2 [18], enquanto os dados obtidos do banco de dados TCGA foram normalizados em *transcript per million* (TPM) e em seguida transformados em LOG_2 [19,20]. Para identificação dos genes diferencialmente expressos (DEGs) os dados foram submetidos a comparação por estatística bayesiana com correção da falsa taxa de descoberta de Benjamini e Hocheberg [21]. Os pontos de corte utilizados foram valor $p.\text{adj} < 0,5$ e $\log_2\text{FC} > |1|$.

De modo a validar os resultados, os DEGs obtidos nos diferentes conjuntos de dados foram interseccionados, sendo selecionados apenas os genes que se apresentaram como diferencialmente expressos em todos os conjuntos. Resultando em 4 DEGs para câncer de cavidade oral e 12 DEGs para o câncer de tireoide. Após com os resultados obtidos, foi realizada a análise de enriquecimento funcional utilizando o banco de dados *Gene Ontology* [22]. Por último realizou-se uma extensa pesquisa na literatura relacionando cada gene a seu papel biológico, bem como descrições anteriores relacionadas a patologias.

Para estes DEGs selecionados foi desenhada a curva de sobrevivência para cada gene individualmente, utilizando os dados provenientes do banco TCGA, uma vez que esses dados não estão disponíveis no GEO. As análises foram realizadas utilizando o pacote *R-Survival*. O ponto de corte para dicotomização dos pacientes foi definido pela mediana. O método Kaplan-Meier foi utilizado para estimar as curvas de sobrevivência e o teste LogRank para comparar a sobrevivência [14].

RESULTADOS

Cada conjunto de dados analisados apresentou diferentes quantidades de genes diferencialmente expressos, (tabela 1) isso se dá devido a particularidade de cada experimento. A fim de validar estes achados, a intersecção dos resultados de todos os conjuntos de dados diferencialmente expressos contribui para diminuir o número de genes a serem analisados e complementar a relevância de uma discussão acerca destes genes. Assim, após a intersecção, para o câncer de cavidade oral foram identificados 4 DEGs: LOXL2, FAT1, SOD3 e DPT. Destes, LOXL2 e FAT1 apresentaram maior valor de expressão em tecidos de câncer quando comparados ao tecido não-tumoral, e SOD3 e DPT um menor valor de expressão. Já para o câncer de tireoide, o cruzamento dos DEGs obtidos resultaram em 12 genes: CRABP1, CSGALNCT1, WSCD2, ITM2A, ID3, PAX9, SELENBP1, TNFRS11B, SDPR, LRP1B, MT1F e SGK223.

Tabela 1- Quantidade de genes diferencialmente expressos encontrados para cada conjunto de dados.

DEG's em Câncer de Cavidade Oral		DEG's em Câncer de Tireoide	
ID	Número de genes	ID	Número de genes
GSE13601	1822	GSE53072	3206
GSE38517	113	GSE65144	6922
GSE58911	2245	GSE82208	257
GSE93519	2689	GSE129562	884
GSE146483	6636	TIREOIDE-TCGA	4100
HNC-TCGA	2164	-	-

Vale ressaltar a ausência de igualdade entre os genes identificados após a intersecção dos conjuntos para câncer de cavidade oral e câncer de tireoide. Diferentemente dos genes de cavidade oral, os genes identificados no câncer de tireoide apresentaram, todos, uma menor expressão quando comparados ao tecido não tumoral. Após a obtenção destes genes realizou-se o enriquecimento funcional, e deste foram selecionadas as 10 primeiras funções biológicas, ranqueadas com base no valor p, tais quais apresentam-se descritos na tabela 2 e 3.

Nota-se que no câncer de cavidade oral o gene LOXL2 apresenta-se em todos os principais processos biológicos identificados a partir do enriquecimento funcional, em contrapartida, FAT1 não foi identificado em nenhum destes processos. De modo geral, os processos identificados no câncer de cavidade oral são relacionados a respostas á íons, respostas a níveis de oxigênio e organização da estrutura celular, processos os quais são esperados quando se discute câncer. Em contrapartida, os processos biológicos identificados sobre o câncer de tireoide são amplos e divergentes, sendo alguns relacionados à formação de tecidos distantes da tireoide, talvez devido a particularidade glandular deste tecido. Nota-se também, uma maior diversidade de genes identificados, entretanto, da mesma forma que no câncer de cavidade oral, nem todos os DEG's são encontrados, porém há uma maior predominância de PAX8.

Tabela 2 - Enriquecimento funcional dos genes identificados nos cânceres de cavidade oral.

Função biológica	Genes envolvidos
Resposta a íons de cobre	LOXL2/SOD3
Organização das Fibras de colágeno	LOXL2/DPT
Resposta a Hipoxia	LOXL2/SOD3
Resposta a íons de metal	LOXL2/SOD3
Resposta a diminuição dos níveis de oxigênio	LOXL2/SOD3
Resposta a níveis de oxigênio	LOXL2/SOD3
Organização da matriz extracelular	LOXL2/DPT
Organização da estrutura extracelular	LOXL2/DPT
Organização da estrutura de encapsulamento	LOXL2/DPT
Oxidação de proteínas	LOXL2

Tabela 3 - Enriquecimento funcional dos genes identificados nos cânceres de tireoide.

Função biológica	Genes envolvidos
Desenvolvimento de metanefros	ID3/PAX8
Processos metabólicos hormonais	CRABP1/PAX8
Processo metabólico de isoprenóides	CRABP1
Morfogênese de túbulos metanéfricos	PAX8
Regulação de processos apoptóticos envolvidos na morfogênese	PAX8
Processos biossintético da heparina	CSGALNACT1
Desenvolvimento do ducto coletor metanéfrico	PAX8
Regulação de processos apoptóticos envolvidos no desenvolvimento	PAX8
Regulação de processos apoptóticos de células mesenquimais	PAX8
Transição mesenquimal para epitelial na morfogênese dos metanefros	PAX8

Em relação à sobrevida dos pacientes (Figura 1), para o câncer de cavidade oral apenas FAT1 demonstrou uma associação entre a alta expressão e menor sobrevida no decorrer do tempo. Ressalta-se que este gene também havia sido identificado com maior valor de expressão em tecidos tumorais quando comparados a tecidos não-tumorais. Entretanto LOXL2, SOD3 e DTP não apresentaram diferença em seu nível de expressão e porcentagem de sobrevida.

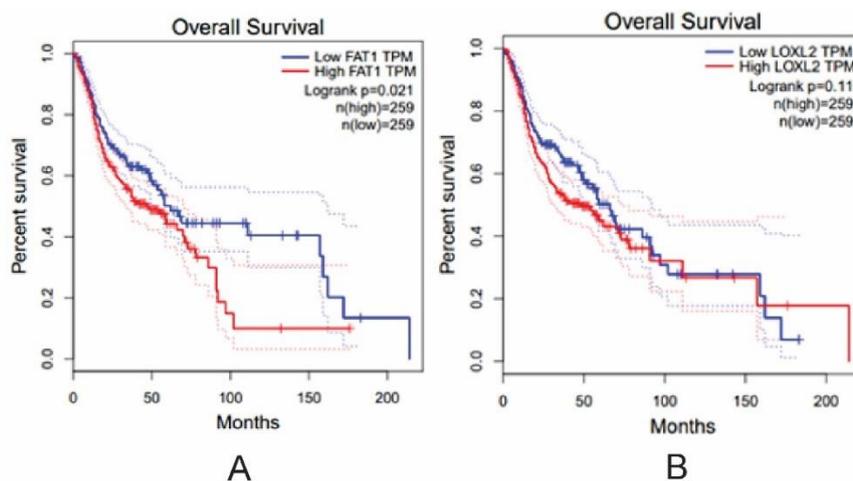
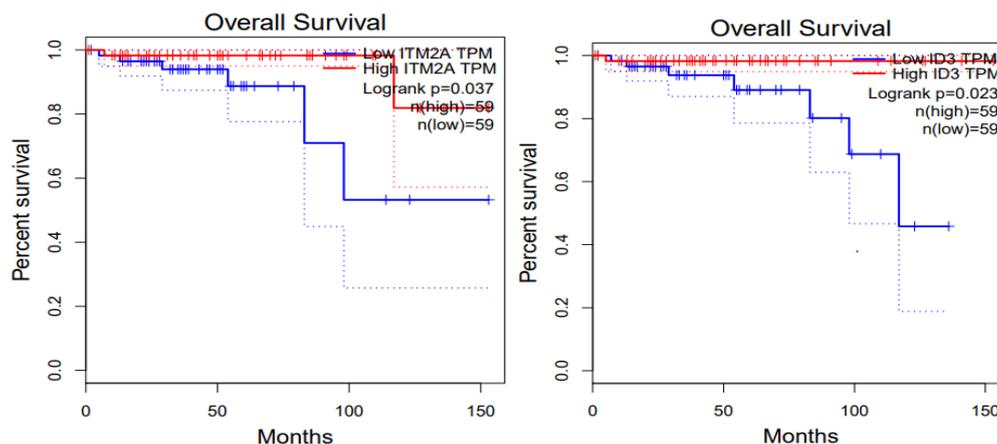


Figura 1. Curva de sobrevivência para os genes FAT1 (A) e LOXL2 (B).

Para o câncer de tireoide 2 genes apresentaram associação entre baixa expressão e aumento na porcentagem de mortalidade no decorrer do tempo: ITM2A e IDE3 (Figura 2). Nota-se uma maior porcentagem de sobrevivência no câncer de tireoide quando comparado aos cânceres de cavidade oral, o que corrobora com a baixa mortalidade relacionada ao câncer de tireoide, conforma dados da Organização Mundial da Saúde (WHO, 2021). Os genes CSGALNACT1, WSCD2, PAX8, SELENBP1, TNFRSF11B, SDPR, LRP1B, MT1F e SKG223 não apresentaram relação entre seu diferencial de expressão e a porcentagem de sobrevivência ao decorrer do tempo. Vale ressaltar que estes resultados demonstram que esses genes não possuem potencial como biomarcadores de prognóstico, entretanto foram identificados como DEGs e possuem potencial de estudo como biomarcadores moleculares para diagnóstico.

Figura 2. Curvas de sobrevivência para os genes ITM2A e ID3.



DISCUSSÃO

Em relação às funções biológicas dos cânceres de cavidade oral, duas delas foram relacionadas com o enriquecimento funcional quanto à resposta à íons de metal e resposta a íons de cobre, sendo os genes LOXL2 e SOD3 envolvidos neste processo. Percebe-se a presença de íons no corpo humano com participações em várias atividades relacionadas à vida como em aspectos do ciclo celular e composição de proteínas, sendo inclusive alvos estudados a novas terapias contra o câncer [23]. Desta forma, processos neoplásicos e metastáticos tendem a apresentar distorções na homeostase de diversos componentes iônicos, incluindo cobre [24]. Sendo o cobre um

importante elemento estudado no sentido de marcador para diagnóstico e prognóstico em câncer, bem como alvo de estratégias terapêuticas no sentido de atuação em sua homeostase [25].

LOXL2 também foi identificado em funções biológicas relacionadas à organização da estrutura celular, em conjunto com DPT, sendo estas: Organização das fibras de colágeno; Organização da matriz extracelular; organização da estrutura extracelular e organização da estrutura de encapsulamento. É de amplo conhecimento o câncer como uma patologia que altera o funcionamento celular normal e alteração do sítio anatômico onde se encontra. Sendo esta alteração anatômica e histológica como alterações no citoplasma, morfologia, organização nuclear utilizadas rotineiramente no diagnóstico desta patologia [26]. Neste processo, a anormalidade da estrutura causa um desbalanceamento no nível de oxigênio que chega às células tumorais, ocasionando menores níveis de oxigênio e até condições de hipóxia [27], desta forma, as células cancerígenas se adaptam à nova condição e passam a sobreviver a mecanismos anaeróbios [28]. Neste sentido, LOXL2 e SOD3 foram identificados como genes atuantes em processos metabólicos relacionados a respostas de níveis de oxigênio.

A tireoide é uma glândula constituinte do sistema endócrino a qual produz hormônios que atuam em todo o organismo [29], desta forma, uma modificação em seu funcionamento pode causar alterações sistêmicas. Neste trabalho, os genes diferencialmente expressos ID3 e PAX8 demonstraram, a partir do enriquecimento funcional, atuações relacionadas a metanefros, o qual no processo embrionário origina os rins, um órgão diferente e distante. Sendo este, um achado particularmente interessante, visto que há uma relação bidirecional entre câncer de tireoide e cânceres renais onde os pacientes que apresentam primariamente a patologia em um destes dois sítios tendem a ter uma recorrência no segundo sítio [30]. Da mesma forma, não só relacionada ao câncer, mas de modo geral disfunções da tireoide acarretam em alterações renais [31].

Da mesma forma, os genes CRABP1 e PAX8 foram identificados relacionados a processos metabólicos hormonais, aspecto biológico esperado devido à natureza da glândula tireoide. PAX8 também se apresentou relacionado a mecanismos de processos apoptóticos, sendo este perdido em células neoplásicas, o que permite a evolução da patologia [32]. Sendo a superexpressão de PAX8 identificada como responsável pela evolução de cânceres [33-35], resultado diferente do encontrado neste trabalho, visto que PAX8 apresentou menores valores de expressão.

Em relação à carcinogênese de cavidade oral, os DEG's identificados neste trabalho possuem papel documentado na literatura. LOXL2 é um gene que codifica a enzima lisil oxidase-like 2, possuindo papel para a progressão de tumores malignos em vários tipos de cânceres [36]. Sendo verificada uma relação entre o metabolismo de cobre e expressão de LOXL2, onde a desregulação do metabolismo de cobre encontra-se intimamente relacionada com a ocorrência e progressão de tumores [37]. Zhang e colaboradores em 2022 demonstraram em seu trabalho com câncer gástrico uma interação entre TRIM44 e LOXL2, onde o primeiro modula o nível do segundo remodelando a matriz extracelular e promovendo a imunidade tumoral [38].

Outra via metabólica descrita para LOXL2 relacionada a progressão tumoral se dá pela contribuição ao metabolismo de glicose, onde em células tumorais esofágicas este gene parece aumentar o metabolismo de glicose, facilitando a proliferação da neoplasia [39]. LOXL2 também foi associado a processos de proliferação, migração e invasão de tecidos em câncer de bexiga, sendo identificado como potencial biomarcador e alvo terapêutico a este tipo de neoplasia [40]. Da mesma forma, o aumento da expressão de LOXL2 foi identificado em adenocarcinoma pancreático [41] e carcinoma hepático [42]. Entretanto, o silenciamento total de LOXL2 também demonstrou aspectos negativos, causando a hipertrofia uterina em ratos e papel na progressão do câncer de útero [43].

No câncer cavidade oral o aumento da expressão LOXL2 já havia sido identificado em células tumorais [44] possuindo papel na proliferação, migração e

invasão de tecidos [45]. Nota-se LOXL2 como um gene importante nas funções celulares normais, sendo que suas alterações nos padrões de expressão acarretam efeitos negativos aos tecidos. Percebe-se que a alta expressão de LOXL2 é encontrada em diversos tipos de cânceres, inclusive em câncer de cabeça e pescoço, sendo, portanto, um potencial biomarcador e alvo terapêutico.

FAT1 expressa o gene FAT caderina atípica 1, o qual é um membro da superfamília caderinas, um grupo de proteínas integrais de membrana e relatado como supressor tumoral ou oncogene, dependendo dos tipos de câncer, associado a regulação da adesão celular, proliferação, motilidade e cinética de actina [46]. Sendo um gene frequentemente mutado em vários tipos de câncer e possuindo sua maior taxa de mutação em carcinomas espinocelular de cabeça e pescoço [47]. Da mesma forma que no presente trabalho, FAT1 é descrito na literatura como um gene altamente expresso na maioria dos tumores e correlacionado aos seus respectivos prognósticos [48]. Já possuindo descrições na literatura relacionando sua alta expressão com o desenvolvimento de carcinoma oral de células escamosas, bem como a um mal prognóstico [49].

SOD3 é um gene que codifica uma isoenzima pertencente à família de proteínas superóxido dismutase (SOD), estas convertem superóxidos em peróxidos de hidrogênio atuando, portanto, na proteção ao estresse oxidativo [50]. Entendendo estas espécies reativas de oxigênio como algo que afeta negativamente a célula, é fácil a compreensão de que uma menor expressão das enzimas desta superfamília vai apresentar efeitos negativos ao contexto celular, podendo evoluir para uma neoplasia, sendo SOD3 já identificado na literatura como um supressor tumoral [51], e suas alterações extracelulares associadas a fenótipos agressivos de carcinoma epidermóide oral [52]. Além disso, a superexpressão induzida de SOD3 demonstrou efeitos benéficos ao câncer de próstata [53] e uma potencial melhora da resposta quimioterápica por doxorrubicina [54].

DPT é o gene responsável pela expressão da proteína dermatopontina, uma proteína de matriz extracelular, a qual apresenta funções relacionadas à estrutura da matriz, bem como à interação célula-matriz e adesão celular [55]. Sendo já documentado na literatura, assim como neste trabalho, uma menor expressão de DPT relacionada a câncer de endométrio [56] e câncer gástrico [57]. Em câncer oral humano, semelhante aos resultados encontrados neste trabalho, a menor expressão de DPT é documentada na literatura associada ao desenvolvimento da neoplasia [58], bem como ao desenvolvimento de metástases [59].

Em relação aos cânceres de tireoide, todos os genes identificados neste trabalho possuem alguma relação descrita na literatura associada a desenvolvimento de cânceres, sendo diferentes em alguns casos a mudança no perfil de expressão. CRABP1 relaciona-se a vitamina A, está também conhecida como retinol é um nutriente essencial necessário para diversos processos fisiológicos [60], executando suas atividades por meios de ligações a receptores nucleares RA que geralmente se pareiam com receptores retinóides X [61]. Os que atuam em conjunto com fatores de transcrição resultando na regulação de processos celulares como proliferação, diferenciação, apoptose entre outros processos fisiológicos. Sendo a sinalização de ácido retinóico desregulada associada a condições de doença [62,63]. Persaud e colaboradores (2013) demonstraram em seu estudo a associação entre CRABP1 na mediação rápida de ERK1/2 associada ao ácido retinóico, com importantes implicações biológicas, onde há expansão da população celular em G1, demonstrando uma supressão no ciclo celular, e proliferação celular [64]. Da mesma forma, Park e colaboradores (2019) alcançaram os mesmos resultados, e demonstram uma relação direta entre CRABP1 e ERK1/2, confirmando a relação no controle mitótico [65].

Estudos têm demonstrado a associação entre uma menor expressão de CRABP1 e câncer. Celestino e colaboradores em 2018, da mesma forma que no presente trabalho, demonstram em seu estudo a presença de uma menor expressão de CRABP1 comparada em tecidos normais de câncer de tireoide [66]. Resultados

semelhantes, são encontrados para o câncer de esôfago [67]. Em contrapartida, Sakata e colaboradores (2022) demonstraram em seu estudo que o aumento da expressão de CRABP1 indica fenótipos malignos e prognósticos ruins relacionados ao câncer gástrico [68]. Demonstrando a necessidade de estudos mais específicos relacionados a tecidos e tipos de câncer. Atualmente, entende-se CRABP1 como um importante gene relacionado a câncer, e com potencial para estudos específicos como biomarcador.

O gene CSGALNACT1 sintetiza a proteína sulfato de condroitina N-acetilgalactosaminiltransferase, possuindo papel importante na biossíntese da cadeia de condroitina na formação da cartilagem e subsequente ossificação endocondral. Apresentando-se também relacionado a regulação, manutenção da proliferação celular, citodiferenciação e morfogênese de tecido [69]. Não foram encontrados trabalhos relacionando câncer com o gene CSGALNACT1, entretanto, alguns estudos com a classe de genes GALNTS, que são tipos de N-acetilgalactosaminiltransferases são descritos. Como o aumento da expressão de GALNT14 relacionado a prognósticos negativos em câncer de cabeça e pescoço [70]. Da mesma forma, uma maior expressão de GALNT5 demonstrou-se relacionada a prognósticos negativos em colangiocarcinomas [71] e GALNT2 para câncer colorretal [72].

Diferentemente dos trabalhos encontrados onde o aumento da expressão de N-acetilgalactosaminiltransferases indica prognósticos negativos, no presente estudo CSGALNACT1 foi identificado como sub expresso. Entretanto esta foi apenas uma discussão especulativa baseada em funções semelhantes, sendo necessários mais estudos caracterizar a real relação entre o câncer de tireoide e CSGALNACT1.

WSC Domain containing 2 é descrito na literatura como um componente integral de membrana. Sendo correlacionado a sensibilidade da insulina às células [73]. Estudos recentes demonstraram uma correlação positiva entre a expressão de WSCD2 em ilhotas de Langerhans com a secreção de insulina, sugerindo uma hipermetilação deste gene em células diabéticas e um papel importante no metabolismo de glicose [74]. Há poucos trabalhos relacionando WSCD2 a cânceres, sendo demonstrado uma correlação entre a baixa expressão de WSCD2 e câncer de mama de pacientes femininas, sendo que quanto menor a expressão maior o risco de mortalidade [30]. Resultados semelhantes ao encontrado em nosso trabalho, visto que células de câncer de tireoide apresentam uma menor expressão deste gene quando comparado a células normais, havendo um maior risco de morte relacionado a esta menor expressão.

Considerando que as células tumorais alteram o metabolismo da glicose, é de interesse geral estudos sobre a plasticidade tumoral relacionada ao metabolismo de glicose [75]. Desta forma, sendo WSCD2 um gene relacionado ao metabolismo de glicose, pode-se esperar uma relação em sua alteração de expressão como um importante fator no desenvolvimento de neoplasias. Entretanto cabe ressaltar a necessidade de estudos específicos relacionados a este gene nos diferentes tecidos tumorais, bem como análises *in vitro*.

ITM2A codifica uma proteína de membrana tipo II, pertencendo a família ITM2. Sendo encontrada na membrana plasmática e no complexo de Golgi [76]. ITM2 está relacionada à diferenciação celular, regulação de função de células T e processos autofágicos. Além disso, estudos têm demonstrado o papel deste gene como um supressor de tumor em diversos cânceres epiteliais [77]. O presente trabalho avaliou uma menor expressão de ITM2A em câncer de tireoide quando relacionado a tecidos normais. Resultados estes semelhantes aos encontrados em outros tipos de câncer como o de mama [78], leucemia mieloide aguda [79], câncer de cabeça e pescoço [80] e câncer de ovário [81]. Sendo que a superexpressão de ITM2A demonstrou-se como um indicativo de diminuição da proliferação de câncer de mama e estimulante a apoptose, por mecanismos de resposta imune induzida [78].

Zhou e colaboradores (2019) em demonstraram seu trabalho, ITM2A como um inibidor de crescimento celular em células tumorais de câncer de mama, por mecanismos de reações imunes induzindo a autofagia [82]. Da mesma forma, Abdulwahab e colaboradores (2020), encontraram resultados semelhantes também no

câncer de mama, indicando este gene como potencial biomarcador e alvo terapêutico [83]. Não foram encontrados trabalhos relacionando ITM2A e câncer de tireoide. Entretanto, o processo de autofagia e resposta imune mediada por ITMA2A demonstra-se bem descrita no câncer de mama, possuindo potencial de ser encontrada em outros tipos de cânceres, inclusive de tireoide.

O inibidor de ligação/diferenciação de DNA-3 (ID3) é uma proteína contendo hélice-alça-hélice (HLH) que pertence à família de proteínas ID. ID3 carece de um domínio de ligação ao DNA, mas pode se ligar a outras proteínas HLH e direcionar fatores de transcrição para inibir ainda mais sua ligação ao DNA [84]. O presente trabalho identificou ID3 como um gene menos expresso em tecidos de câncer de tireoide quando comparados com tecidos normais, resultados semelhantes aos encontrados na literatura para câncer de pulmão [85] e pancreático [86]. Da mesma forma, menores valores de ID3 já foram identificados relacionados ao câncer de tireoide e correlacionado com metástases linfonodais [87].

PAX8 é um fator de transcrição envolvido na regulação do desenvolvimento de uma variedade de tecidos em diferentes espécies, principalmente na morfogênese da glândula tireóide, sistema renal e sistema mulleriano [88,89], se apresentando como um gene que se comporta em paralelo ao desenvolvimento embrionário e o processo tumorigênico [90]. Sendo alterações neste gene relatadas associadas ao desenvolvimento de diversos tipos de tumores como de ovário [91], renais [92] e de tireoide [93]. Entretanto, diferente dos resultados encontrados neste trabalho, PAX8 tem sido referido como um oncogene com níveis de expressão maiores em tecidos de câncer do que em tecidos normais [94].

SELENBP1 é um gene que expressa a proteína de ligação ao selênio 1, a qual catalisa a oxidação do metanotiol, composto organossulfurado produzido pela microbiota intestinal, e gera como produto peróxido de hidrogênio e sulfeto de hidrogênio, moléculas de sinalização para diferenciação celular [95]. Este gene aparece descrito como um gene supressor de tumor, o qual sua diminuição de expressão está correlacionada com diversos tipos de cânceres [96], resultado semelhante ao encontrado neste trabalho. Sendo este gene descrito também como um provável preditor de prognóstico ao câncer colorretal [97].

TNFRSF11B é um membro da superfamília dos receptores do fator de necrose tumoral 11B, sendo estudado por estar envolvido no desenvolvimento de diversas patologias humanas [98]. Este gene aparece descrito como superexpresso em câncer colorretal, bem como associado a um mal prognóstico [99]. Em contrapartida, TNFRSF11B aparece como um gene regulado negativamente em câncer de mama [100], resultados semelhantes ao encontrado neste trabalho. Ressalta-se também a importância deste gene relacionada ao alvo terapêutico para câncer gástrico [98].

O gene SDPR expressa a proteína conhecida por resposta a privação de soro ou proteína de ligação à fosfatidilserina, atua principalmente na formação de invaginações da membrana plasmática e substrato da proteína quinase C [101]. Este gene também foi descrito recentemente como um gene supressor de tumor em determinados tipos de cânceres [102], como câncer de mama [103] e carcinoma hepatocelular [104] associado à promoção de apoptose. No câncer de tireoide, resultado semelhante ao encontrado neste trabalho, uma menor expressão de SDPR já foi relatada associada a este tipo de câncer, bem como a um pobre prognóstico [105].

LRP1B é receptor de lipoproteína de baixa densidade localizado na membrana celular. Desempenhando diversos papéis na função e desenvolvimento celular normal, como transporte e sinalização celular, sendo considerado também um suposto supressor tumoral [106]. Trabalhos relacionaram a frequente mutação deste gene associada a câncer gástrico [107] e adenocarcinoma de pulmão [108]. Além disso, da mesma forma que este trabalho, sua menor expressão já havia sido relatada em câncer gástrico [109] e câncer de bexiga [110], demonstrando seu potencial como biomarcador ao diagnóstico desta patologia.

O gene MT1F pertence à família de genes MT os quais expressam proteínas de ligação com metais de baixo peso molecular, metalotioneína, sendo MT1F uma das 18 isoformas do gene MT1, envolvido especificamente no estado estacionário dos metais e na proteção contra o estresse oxidativo [111]. A desregulação da expressão dos genes MT é discutida em diversos tipos de cânceres, podendo haver um aumento da expressão associada ao mau prognóstico em câncer de mama, pulmão, rim e pâncreas ou uma diminuição da expressão em cânceres hepatocelulares gástricos, sistema nervoso e tireoide [112]. Da mesma forma que neste trabalho uma menor expressão de MT1F já foi relatada na literatura e sugerida como potencial biomarcador ao câncer de tireoide [113].

SGK223 expressa uma proteína pseudoquinase, a qual atua como um regulador negativo da família SFK, regulando sua localização subcelular [114]. Há poucos estudos referentes a este gene e câncer, entretanto SGK223 parece possuir um importante papel de sinalização em vias oncogênicas [115]. SGK223 já foi relatado como um gene superexpresso em adenocarcinoma ductal pancreático e identificado como responsável pela aquisição de um fenótipo migratório invasivo [116]. Em contrapartida, semelhantes aos achados deste trabalho, SGK223 foi relatado como um gene com menor valor de expressão em cânceres de tireoide [117].

CONCLUSÃO

Cada câncer parece ser uma patologia ímpar com seus aspectos moleculares individuais, entretanto apresentam características semelhantes que permitem agrupamentos específicos, possibilitando diagnósticos, prognósticos e estratégias terapêuticas. As evoluções das técnicas de biologia molecular têm elevado o padrão de estudos e possibilitado novas descobertas, ao mesmo tempo em que formam uma enxurrada de dados os quais têm sido depositados em repositórios públicos. TCGA e GEO *database* são exemplos destes repositórios, os quais apresentam dados de expressão gênica possibilitando que novas pesquisas sejam realizadas por novos pesquisadores.

O presente trabalho utilizou estes dados para a realização da análise de expressão diferencial de genes em cânceres de cabeça e pescoço, dividindo-os, devido à disponibilidade dos dados, em dois grandes grupos: cânceres de cavidade oral e câncer de tireoide. Como resultado obteve-se 4 genes diferencialmente expressos para o câncer de cavidade oral: LOXL2; FAT1; SOD3 e DPT, e 12 genes para o câncer de tireoide: CRABP1, CSGALNCT1, WSCD2, ITM2A, ID3, PAX9, SELENBP1, TNFRS11B, SDPR, LRP1B, MT1F e SGK223, sendo todos estes potenciais biomarcadores de diagnóstico. Ressalta-se que, com exceção de CSGALNCT1, todos estes genes possuem referencial teórico demonstrando relações com câncer, entretanto CSGALNCT1 possui referencial relacionado à família gênica. Fundamentando os resultados encontrados no presente estudo.

A partir da análise da curva de sobrevida identificou-se associação entre maiores valores de expressão de FAT1 e uma menor sobrevida ao passar dos anos, podendo ser FAT1 um potencial biomarcador de mau prognóstico do câncer de cavidade oral. Em contrapartida, a curva de sobrevida de CRABP 1, ITM2 e ID3 demonstrou uma associação entre menores valores de expressão e uma menor taxa de sobrevivência ao passar dos anos, podendo ser estes potenciais biomarcadores de prognóstico do câncer de tireoide. Trabalhos complementares laboratoriais são necessários para a confirmação do uso destes genes como biomarcadores de diagnóstico e prognóstico. Desta forma, acreditamos que este trabalho apresenta contribuição científica devido aos genes identificados, bem como sua discussão na literatura já relacionada. Ao mesmo tempo, esperamos que este trabalho possa contribuir para pesquisas futuras e auxiliar em melhores métodos diagnósticos e prognósticos.

Conflitos de interesse: Os autores não declaram qualquer conflito de interesses.

REFERÊNCIAS

1. Snustad DP, Simmons MJ. [Fundamentals of genetics]. 7.ed. Rio de Janeiro: Guanabara Koogan. 2020. Portuguese version.
2. Chow LQM. Head and Neck Cancer. *N Engl J Med.* 2020; 382:60-72.
3. Konkigs H, Stappers S, Geens M, Winter BYD, Lamote K, Meerbeeck JPV, et al. A literature review of the potential diagnostic biomarkers of Head and Neck Neoplasms. *Front oncol.* 2020 Jun 26;10:1020.
4. World Health Organization (WHO) [Internet]. Cancer facts Sheets. [cited 2021 oct 26]. Available from: <https://gco.iarc.fr/today/fact-sheets-cancers>
5. Arquette M A. [Head and Neck Cancer]. IN: ORG: Ramaswamy, G. (2002). Washington – Manual de Oncologia. 1.ed. Rio de Janeiro: Guanabara Koogan. 2002. Portuguese version.
6. Galbatti ALS, Padovani-Junior, JA, Maniglia JV, Rodrigues CDS, Pavarino EC, Goloni-Bertollo EM. Head and neck cancer: causes, prevention and treatment. *Braz J Otorhinolaryngol.* 2013 Mar-Apr;79(2) 239-47.
7. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers* 2020, 6, 92.
8. Genden EM. Head and neck cancer: management and reconstruction. New York: Thieme. 2.ed. 2020.
9. Reis JB, Oliveira JM, Nascimento VF, Cabral JF, Lucietto GC, Silva RA. [Head and neck cancer: communication and meaning]. *Revista de enfermagem.* 2018, 12(12):3263-70. Portuguese.
10. Edgar R, Domrachev M, Lash AE. Gene expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002 jan 1;30(1):207-10.
11. The Cancer Gemone Atlas Reesearch Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature.* 2014 513, 202-209.
12. Xu Y, Wu G, Li J, Li J, Ruan N, Ma L. Screening and identification of key biomarkers for bladder cancer: A study based on TCGA and GEO data. *Biomed res int.* 2020 Jan 23;2020:8283401.
13. Zheng Q, Min S, Zhou Q. Identification of potential diagnostic and prognostic biomarkers for LUAD based on TCGA and GEO databases. *Biosci Rep.* 2021 jun 25;41(6):BSR20204370.
14. Sartor ITS, Zeidán-Chuliá F, Albanus RD, Dalmolin RJS, Moreira JCF. Computational analyses reveal a prognostic impact of TULP3 as a transcriptional master regulator in pancreatic ductal adenocarcinoma. *Mol BioSyst.* 2014 jun;10(6):1461-8.
15. Ye Y, Wang J, Liang F, Song P, Yan X, Wu S. Identification of key genes for HNSCC from public databases using bioinformatics analysis. *Cancer cell international.* 2021 21, 549.
16. Zhao X, Sun S, Zeng X, Cui L. Expression profiles analysis identifies a novel three-mRNA signature to predict overall survival in oral squamous cell carcinoma. *Am J Cancer Res.* 2018 Mar 1;8(3):450-461.
17. Wan Y, Zhang X, Leng H, Yin W, Zeng W, Zhang C. Identifying hub genes of papillary thyroid carcinoma in the TCGA and GEO database using bioinformatics analysis. *PeerJ.* 2020 Jul 9;8:e9120.
18. Sketel, D. Microarray bioinformatics. New York: Cambridge. 2003.
19. Korpelainen E, Tuimala J, Somervuo P, Huss M, Wong G. RNA-seq data analysis: A practical approach. New York: CRC Press. 2005.
20. Song Y, Gao N, Yang Z, Zhang S, Fan T, Zhang B. Comprehensive Molecular Analyses of Notch Pathway-Related Genes to Predict Prognosis and Immunotherapy Response in Patients with Gastric Cancer. *J Oncol.* 2023 Jan 24;2023:2205083.
21. Sinha BK, Tokar EJ, Li J, Bushel PR. Gene Expression Profiling Elucidates Cellular Responses to NCX4040 in Human Ovarian Tumor Cells: Implications in the Mechanisms of Action of NCX4040. *Cancers.* 2023; 15(1):285.
22. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D419-D426.
23. Chi Y, Sun P, Gao Y, Zhang J, Wang L. Ion Interference Therapy of Tumors Based on Inorganic Nanoparticles. *Biosensors (Basel).* 2022 Feb 6;12(2):100.

24. Fnu G, Weber GF. Alterations of Ion Homeostasis in Cancer Metastasis: Implications for Treatment. *Front Oncol.* 2021 Dec 20;11:765329.
25. Lelièvre P, Sancey L, Coll JL, Deniaud A, Busser B. The Multifaceted Roles of Copper in Cancer: A Trace Metal Element with Dysregulated Metabolism, but Also a Target or a Bullet for Therapy. *Cancers (Basel).* 2020 Dec 1;12(12):3594.
26. Muthuswamy SK. Self-organization in cancer: Implications for histopathology, cancer cell biology, and metastasis. *Cancer Cell.* 2021 Apr 12;39(4):443-446.
27. Dunn T. Oxygen and cancer. *N C Med J.* 1997 Mar-Apr;58(2):140-3.
28. Orang AV, Petersen J, McKinnon RA, Michael MZ. Micromanaging aerobic respiration and glycolysis in cancer cells. *Mol Metab.* 2019 May;23:98-126.
29. Tortora, GJ, Derrickson B. [Principles of anatomy and physiology]. 14. ed. Rio de Janeiro: Guanabara Koogan. Portuguese Version.
30. Busel PR, Ward J, Burkholder A, Li J, Anchang B. Mitochondrial-nuclear epistasis underlying phenotypic variation in breast cancer pathology. *Scientific Reports.* 2022 12 1393.
31. Basu G, Mohapatra A. Interactions between thyroid disorders and kidney disease. *Indian J Endocrinol Metab.* 2012 Mar;16(2):204-13.
32. Pfeffer CM, Singh ATK. Apoptosis: A Target for Anticancer Therapy. *Int J Mol Sci.* 2018 Feb 2;19(2):448.
33. Di Palma T, Filippone MG, Pierantoni GM, Fusco A, Soddu S, Zannini M. Pax8 has a critical role in epithelial cell survival and proliferation. *Cell Death Dis.* 2013 Jul 18;4(7):e729.
34. Di Palma T, Lucci V, Cristofaro T, Filippone MG, Zannini M. A role for PAX8 in the tumorigenic phenotype of ovarian cancer cells. *BMC cancer.* 2014 14 292.
35. Song X, Chen Y, Peng Y, Wang X, Zheng S, Shi F, Lan J. LncRNA-PAX8-AS1 Silencing Decreases Cell Viability, Enhances Apoptosis, and Suppresses Doxorubicin Resistance in Myeloid Leukemia via the miR-378g/ERBB2 Axis. *Evid Based Complement Alternat Med.* 2022 Oct 6;2022:2295044.
36. Qiu Z, Qiu S, Mao W, Lin W, Peng Q, Chang H. LOXL2 reduces 5-FU sensitivity through the Hedgehog/BCL2 signaling pathway in colorectal cancer. *Exp Biol Med (Maywood).* 2022 Dec 27:15353702221139203.
37. Chang W, Li H, Zhong L, Zhu T, Chang Z, Ou W, et al. Development of a copper metabolism-related gene signature in lung adenocarcinoma. *Front Immunol.* 2022 Nov 29;13:1040668.
38. Zhang X, Wu X, Sun Y, Chu Y, Liu F, Chen C. TRIM44 regulates tumor immunity in gastric cancer through LOXL2-dependent extracellular matrix remodeling. *Cell Oncol (Dordr).* 2023 Apr;46(2):423-435.
39. Jiao JW, Zhan XH, Wang JJ, He LX, Guo ZC, Xu XE, et al. LOXL2-dependent deacetylation of aldolase A induces metabolic reprogramming and tumor progression. *Redox Biol.* 2022 Oct 3;57:102496.
40. Chen H, Yang W, Xue X, Li Y, Jin Z, Ji Z. Integrated Analysis Revealed an Inflammatory Cancer-Associated Fibroblast-Based Subtypes with Promising Implications in Predicting the Prognosis and Immunotherapeutic Response of Bladder Cancer Patients. *Int J Mol Sci.* 2022 Dec 15;23(24):15970.
41. Li L, Xiao Z, He P, Zou W, Deng Z, Zhang G, et al. Molecular subtyping based on TRP family and prognostic assessment for TRP-associated LncRNAs in pancreatic adenocarcinoma. *BMC Gastroenterology.* 2022 22 454.
42. Zhao N, Chen C, Guo Y, Liu T, Che N, Zhang D, et al. LOXL2 serves as a prognostic biomarker for hepatocellular carcinoma by mediating immune infiltration and vasculogenic mimicry. *Dig Liver Dis.* 2023 May;55(5):661-672.
43. Lu X, Xin DE, Du JK, Zou QC, Wu Q, Zhang YS, et al. Loss of LOXL2 Promotes Uterine Hypertrophy and Tumor Progression by Enhancing H3K36ac-Dependent Gene Expression. *Cancer Res.* 2022 Dec 2;82(23):4400-4413.
44. Zhu G, Wang L, Meng W, Lu S, Cao B, Liang X, et al. LOXL2-enriched small extracellular vesicles mediate hypoxia-induced premetastatic niche and indicates poor outcome of head and neck cancer. *Theranostics.* 2021 Sep 3;11(19):9198-9216.
45. Liu C, Guo T, Sakai A, Ren S, Fukusumi T, Ando M, et al. A novel splice variant of LOXL2 promotes progression of human papillomavirus-negative head and neck squamous cell carcinoma. *Cancer.* 2020 Feb 15;126(4):737-748.
46. Huang ZL, Zhang PB, Zhang JT, Li F, Li TT, Huang XY. Comprehensive Genomic Profiling Identifies *FAT1* as a Negative Regulator of EMT, CTCs, and Metastasis of Hepatocellular Carcinoma. *J Hepatocell Carcinoma.* 2023 Mar 7;10:369-382.

47. Chen ZG, Saba NF, Teng Y. The diverse functions of FAT1 in cancer progression: good, bad, or ugly? *J Exp Clin Cancer Res.* 2022 Aug 15;41(1):248.
48. Wang Z, Lin K, Xiao H. A pan-cancer analysis of the FAT1 in human tumors. *Sci Rep.* 2022 Dec 14;12(1):21598. doi: 10.1038/s41598-022-26008-1.
49. Lan T, Ge Q, Zheng K, Huang L, Yan Y, Zheng L, et al. FAT1 Upregulates in Oral Squamous Cell Carcinoma and Promotes Cell Proliferation *via* Cell Cycle and DNA Repair. *Front Oncol.* 2022 May 11;12:870055.
50. Zelko IN, Mariani TJ, Folz RJ. Superoxide dismutase multigene family: a comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. *Free Radic Biol Med.* 2002 Aug 1;33(3):337-49.
51. Griess B, Tom E, Domann F, Teoh-Fitzgerald M. Extracellular superoxide dismutase and its role in cancer. *Free Radic Biol Med.* 2017 Nov;112:464-479.
52. Yokoe H, Nomura H, Yamano Y, Fushimi K, Sakamoto Y, Ogawara K, et al. Alteration of extracellular superoxide dismutase expression is associated with an aggressive phenotype of oral squamous-cell carcinoma. *Exp Ther Med.* 2010 Jul;1(4):585-590.
53. Kim J, Mizokami A, Shin M, Izumi K, Konaka H, Kadono Y, et al. SOD3 acts as a tumor suppressor in PC-3 prostate cancer cells via hydrogen peroxide accumulation. *Anticancer Res.* 2014 Jun;34(6):2821-31.
54. Mira E, Carmona-Rodríguez L, Pérez-Villamil B, Casas J, Fernández-Aceñero MJ, Martínez-Rey D, et al. SOD3 improves the tumor response to chemotherapy by stabilizing endothelial HIF-2 α . *Nat Commun.* 2018 Feb 8;9(1):575.
55. Kuroda K, Okamoto O, Shinkai H. Dermato-pontin expression is decreased in hypertrophic scar and systemic sclerosis skin fibroblasts and is regulated by transforming growth factor-beta1, interleukin-4, and matrix collagen. *J Invest Dermatol.* 1999 May;112(5):706-10.
56. Huang H, Hao Z, Long L, Yin Z, Wu C, Zhou X, Zhang B. Dermato-pontin as a potential pathogenic factor in endometrial cancer. *Oncol Lett.* 2021 May;21(5):408.
57. Huang S, Ma L, Lan B, Liu N, Nong W, Huang Z. Comprehensive analysis of prognostic genes in gastric cancer. *Aging (Albany NY).* 2021 Oct 22;13(20):23637-23651.
58. Koike K, Uzawa K, Miyamoto I, Kimura Y, Kasama H, Ishige S, et al. Inactivation of dermatopontin via histone deacetylation in human oral cancer. *Journal of Oral and Maxillofacial Surgery.* 2017 29, 5, 400-404.
59. Yamatoji M, Kasamatsu A, Kouzu Y, Koike H, Sakamoto Y, Ogawara K, et al. Dermato-pontin: a potential predictor for metastasis of human oral cancer. *Int J Cancer.* 2012 Jun 15;130(12):2903-11.
60. Tanumihardjo SA, Russell RM, Stephensen CB, Gannon BM, Craft NE, Haskell MJ, et al. Biomarkers of Nutrition for Development (BOND)-Vitamin A Review. *J Nutr.* 2016 Sep;146(9):1816S-48S. doi: 10.3945/jn.115.229708. Epub 2016 Aug 10.
61. Cunningham TJ, Duester G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat Rev Mol Cell Biol.* 2015 Feb;16(2):110-23.
62. Wolf G. Retinoic acid as cause of cell proliferation or cell growth inhibition depending on activation of one of two different nuclear receptors. *Nutr Rev.* 2008 Jan;66(1):55-9.
63. Nhieu J, Lin YL, Wei LN. CRABP1 in Non-Canonical Activities of Retinoic Acid in Health and Diseases. *Nutrients.* 2022 Apr 6;14(7):1528.
64. Persaud SD, Lin YW, Wu CY, Kagechika H, Wei LN. Cellular retinoic acid binding protein I mediates rapid non-canonical activation of ERK1/2 by all-trans retinoic acid. *Cell Signal.* 2013 Jan;25(1):19-25.
65. Park SW, Nhieu J, Persaud SD, Miller MC, Xia Y, Lin YW, et al. A new regulatory mechanism for Raf kinase activation, retinoic acid-bound Crabp1. *Sci Rep.* 2019 Nov 14;9(1):17042.
66. Celestino R, Nome T, Pestana A, Hoff AM, Gonçalves AP, Pereira L, et al. CRABP1, C1QL1 and LCN2 are biomarkers of differentiated thyroid carcinoma, and predict extrathyroidal extension. *BMC Cancer.* 2018 Jan 10;18(1):68.
67. Tanaka K, Imoto I, Inoue J, Kozaki K, Tsuda H, Shimada Y, et al. Frequent methylation-associated silencing of a candidate tumor-suppressor, CRABP1, in esophageal squamous-cell carcinoma. *Oncogene.* 2007 Sep 27;26(44):6456-68.
68. Sakata K, Kanda M, Shimizu D, Nakamura S, Inokawa Y, Hattori N, et al. Expression of cellular retinoic acid binding protein 1 predicts peritoneal recurrence of gastric cancer. *Int J Oncol.* 2022 Jun;60(6):63.
69. Uyama T, Kitagawa H, Tamura Ji J, Sugahara K. Molecular cloning and expression of human chondroitin N-acetylgalactosaminyltransferase: the key enzyme for chain initiation and

- elongation of chondroitin/dermatan sulfate on the protein linkage region tetrasaccharide shared by heparin/heparan sulfate. *J Biol Chem.* 2002 Mar 15;277(11):8841-6.
70. Lin NC, Shih YH, Chiu KC, Li PJ, Yang HW, Lan WC, et al. Association of rs9679162 Genetic Polymorphism and Aberrant Expression of Polypeptide *N-Acetylgalactosaminyltransferase 14 (GALNT14)* in Head and Neck Cancer. *Cancers (Basel).* 2022 Aug 30;14(17):4217.
71. Detarya M, Lert-Itthiporn W, Mahalapbutr P, Klaewkla M, Sorin S, Sawanyawisuth K, et al. Emerging roles of GALNT5 on promoting EGFR activation in cholangiocarcinoma: a mechanistic insight. *Am J Cancer Res.* 2022 Sep 15;12(9):4140-4159.
72. Liao YY, Chuang YT, Lin HY, Lin NY, Hsu TW, Hsieh SC, et al. GALNT2 promotes invasiveness of colorectal cancer cells partly through AXL. *Mol Oncol.* 2023 Jan;17(1):119-133.
73. Gao C, Hsu FC, Dimitrov LM, Okut H, Chen YI, Taylor KD, et al. A genome-wide linkage and association analysis of imputed insertions and deletions with cardiometabolic phenotypes in Mexican Americans: The Insulin Resistance Atherosclerosis Family Study. *Genet Epidemiol.* 2017 May;41(4):353-362.
74. Wang WJ, Huang R, Zheng T, Du Q, Yang MN, Xu YJ, et al. Genome-Wide Placental Gene Methylations in Gestational Diabetes Mellitus, Fetal Growth and Metabolic Health Biomarkers in Cord Blood. *Front Endocrinol (Lausanne).* 2022 May 26;13:875180.
75. Lin X, Xiao Z, Chen T, Liang SH, Guo H. Glucose Metabolism on Tumor Plasticity, Diagnosis, and Treatment. *Front Oncol.* 2020 Mar 6;10:317.
76. Cegarra C, Chaves C, Déon C, Do TM, Dumas B, Frenzel A, Kuhn P, Roudieres V, Guillemot JC, Lesuisse D. Exploring ITM2A as a new potential target for brain delivery. *Fluids Barriers CNS.* 2022 Mar 21;19(1):25.
77. Li Y, Wang J, Gao C, Hu Q, Mao X. Integral membrane protein 2A enhances sensitivity to chemotherapy via notch signaling pathway in cervical cancer. *Bioengineered.* 2021 Dec;12(2):10183-10193.
78. Zhang R, Xu T, Xia Y, Wang Z, Li X, Chen W. ITM2A as a Tumor Suppressor and Its Correlation With PD-L1 in Breast Cancer. *Front Oncol.* 2021 Feb 12;10:581733.
79. Nagy Á, Ósz Á, Budczies J, Krizsán S, Szombath G, Demeter J, Bödör C, Györffy B. Elevated HOX gene expression in acute myeloid leukemia is associated with NPM1 mutations and poor survival. *J Adv Res.* 2019 Jun 11;20:105-116.
80. Cao Z, Ao Y, Guo Y, Zhou S. Comprehensive Analysis of mRNA Expression Profiles in Head and Neck Cancer by Using Robust Rank Aggregation and Weighted Gene Coexpression Network Analysis. *Biomed Res Int.* 2020 Dec 7;2020:4908427.
81. Nguyen TM, Shin IW, Lee TJ, Park J, Kim JH, Park MS, et al. Loss of ITM2A, a novel tumor suppressor of ovarian cancer through G2/M cell cycle arrest, is a poor prognostic factor of epithelial ovarian cancer. *Gynecol Oncol.* 2016 Mar;140(3):545-53.
82. Zhou C, Wang M, Yang J, Xiong H, Wang Y, Tang J. Integral membrane protein 2A inhibits cell growth in human breast cancer via enhancing autophagy induction. *Cell Commun Signal.* 2019 Aug 22;17(1):105.
83. Abuderman AA, Harb OA, Gertallah LM. Prognostic and clinicopathological values of tissue expression of MFAP5 and ITM2A in triple-negative breast cancer: an immunohistochemical study. *Contemp Oncol (Pozn).* 2020;24(2):87-95.
84. Deed RW, Bianchi SM, Atherton GT, Johnston D, Santibanez-Koref M, Murphy JJ, et al. An immediate early human gene encodes an Id-like helix-loop-helix protein and is regulated by protein kinase C activation in diverse cell types. *Oncogene.* 1993 Mar;8(3):599-607.
85. Chen F, Zhao Q, Wang S, Wang H, Li X. Upregulation of Id3 inhibits cell proliferation and induces apoptosis in A549/DDP human lung cancer cells in vitro. *Mol Med Rep.* 2016 Jul;14(1):313-8.
86. Xu J, Palestino Dominguez M, Alewine C. Loss of ID3 in pancreatic cancer cells increases DNA damage without impairing MDC1 recruitment to the nuclear foci. *Cancer Commun (Lond).* 2022 Mar;42(3):269-272.
87. Xu S, Mo C, Lin J, Yan Y, Liu X, Wu K, Zhang H, Zhu Y, Chen L, Chen X. Loss of ID3 drives papillary thyroid cancer metastasis by targeting E47-mediated epithelial to mesenchymal transition. *Cell Death Discov.* 2021 Aug 30;7(1):226.
88. Cristofaro T, Mascia A, Pappalardo A, D'Andrea B, Nitsch L, Zannini M. Pax8 protein stability is controlled by sumoylation. *J Mol Endocrinol.* 2009 Jan;42(1):35-46.
89. Khizer K, Padda J, Khedr A, Tasnim F, Al-Ewaidat OA, Patel V, et al. Paired-Box Gene 8 (PAX8) and Its Association With Epithelial Carcinomas. *Cureus.* 2021 Aug 16;13(8):e17208.

90. Kakun RR, Melamed Z, Perets R. PAX8 in the Junction between Development and Tumorigenesis. *Int J Mol Sci.* 2022 Jul 3;23(13):7410.
91. Xiang L, Kong B. PAX8 is a novel marker for differentiating between various types of tumor, particularly ovarian epithelial carcinomas. *Oncol Lett.* 2013 Mar;5(3):735-738.
92. Patel SA, Hirose S, Rodrigues P, Vojtasova E, Richardson EK, Ge J, et al. The renal lineage factor PAX8 controls oncogenic signalling in kidney cancer. *Nature.* 2022 Jun;606(7916):999-1006.
93. Rosignolo F, Sponziello M, Durante C, Puppini C, Mio C, Baldan F, et al. Expression of PAX8 Target Genes in Papillary Thyroid Carcinoma. *PLoS One.* 2016 Jun 1;11(6):e0156658.
94. Fan T, Zhu W, Kong M, Yang X, Wang C, Wang M, et al. The Significance of PAX8-PPAR γ Expression in Thyroid Cancer and the Application of a PAX8-PPAR γ -Targeted Ultrasound Contrast Agent in the Early Diagnosis of Thyroid Cancer. *Contrast Media Mol Imaging.* 2022 Jun 23;2022:3265342.
95. Steinbrenner H, Micoogullari M, Hoang NA, Bergheim I, Klotz LO, Sies H. Selenium-binding protein 1 (SELENBP1) is a marker of mature adipocytes. *Redox Biol.* 2019 Jan;20:489-495.
96. Zhang Y, He Q. The role of SELENBP1 and its epigenetic regulation in carcinogenic progression. *Front Genet.* 2022 Nov 1;13:1027726.
97. Zhang X, Hong R, Bei L, Hu Z, Yang X, Song T, et al. SELENBP1 inhibits progression of colorectal cancer by suppressing epithelial-mesenchymal transition. *Open Med (Wars).* 2022 Sep 1;17(1):1390-1404.
98. Luan F, Li X, Cheng X, Huangfu L, Han J, Guo T, et al. TNFRSF11B activates Wnt/ β -catenin signaling and promotes gastric cancer progression. *Int J Biol Sci.* 2020 Apr 25;16(11):1956-1971.
99. Zhang JR, Hou P, Wang XJ, Weng ZQ, Shang-Guan XC, Wang H, et al. TNFRSF11B Suppresses Memory CD4 $^{+}$ T Cell Infiltration in the Colon Cancer Microenvironment: A Multiomics Integrative Analysis. *Front Immunol.* 2021 Dec 6;12:742358.
100. Luo P, Lu G, Fan LL, Zhong X, Yang H, Xie R, et al. Dysregulation of Tmprss3 and TNFRSF11B correlates with tumorigenesis and poor prognosis in patients with breast cancer. *Oncol Rep.* 2017 Apr;37(4):2057-2062.
101. Gustincich S, Vatta P, Goruppi S, Wolf M, Saccone S, Della Valle G, et al. The human serum deprivation response gene (SDPR) maps to 2q32-q33 and codes for a phosphatidylserine-binding protein. *Genomics.* 1999 Apr 1;57(1):120-9.
102. Wang Y, Song Z, Leng P, Liu Y. A systematic analysis reveals gene expression alteration of serum deprivation response (SDPR) gene is significantly associated with the survival of patients with cancer. *Oncol Rep.* 2019 Sep;42(3):1161-1172.
103. Ozturk S, Papageorgis P, Wong CK, Lambert AW, Abdolmaleky HM, Thiagalingam A, et al. SDPR functions as a metastasis suppressor in breast cancer by promoting apoptosis. *Proc Natl Acad Sci U S A.* 2016 Jan 19;113(3):638-43.
104. Chen X, Ma W, Yao Y, Zhang Q, Li J, Wu X, et al. Serum deprivation-response protein induces apoptosis in hepatocellular carcinoma through ASK1-JNK/p38 MAPK pathways. *Cell Death Dis.* 2021 Apr 30;12(5):425.
105. Wang QX, Chen ED, Cai YF, Zhou YL, Dong SY, Zhang XH, et al. Serum deprivation response functions as a tumor suppressor gene in papillary thyroid cancer. *Clin Genet.* 2019 Nov;96(5):418-428.
106. Príncipe C, Dionísio de Sousa IJ, Prazeres H, Soares P, Lima RT. LRP1B: A Giant Lost in Cancer Translation. *Pharmaceuticals (Basel).* 2021 Aug 24;14(9):836.
107. Kung CY, Fang WL, Hung YP, Huang KH, Chen MH, Chao Y, et al. Comparison of the mutation patterns between tumor tissue and cell-free DNA in stage IV gastric cancer. *Aging (Albany NY).* 2023 Feb 10;15(3):777-790.
108. Rao W, Yang L, Dai N, Zhang L, Liu J, Yang B, et al. Frequently mutated genes in predicting the relapse of stage I lung adenocarcinoma. *Clin Transl Oncol.* 2023 Feb 5.
109. Miao J, Yang Z, Guo W, Liu L, Song P, Ding C, et al. Integrative analysis of the proteome and transcriptome in gastric cancer identified LRP1B as a potential biomarker. *Biomark Med.* 2022 Oct;16(15):1101-1111.
110. Conconi D, Jemima A, Giambra M, Redaelli S, Croci GA, Dalprà L, et al. Analysis of copy number alterations in bladder cancer stem cells revealed a prognostic role of LRP1B. *World J Urol.* 2022 Sep;40(9):2267-2273.
111. Dai H, Wang L, Li L, Huang Z, Ye L. Metallothionein 1: A New Spotlight on Inflammatory Diseases. *Front Immunol.* 2021 Nov 5;12:739918.

112. Yan DW, Fan JW, Yu ZH, Li MX, Wen YG, Li DW, et al. Downregulation of metallothionein 1F, a putative oncosuppressor, by loss of heterozygosity in colon cancer tissue. *Biochim Biophys Acta*. 2012 Jun;1822(6):918-26.
113. Wojtczak B, Pula B, Gomulkiewicz A, Olbromski M, Podhorska-Okolow M, Domoslawski P, et al. Metallothionein Isoform Expression in Benign and Malignant Thyroid Lesions. *Anticancer Res*. 2017 Sep;37(9):5179-5185.
114. Patel O, Griffin MDW, Panjekar S, Dai W, Ma X, Chan H, et al. Structure of SgK223 pseudokinase reveals novel mechanisms of homotypic and heterotypic association. *Nature communications* 2017 8 1157.
115. Patel O, Roy MJ, Murphy JM, Lucet IS. The PEAK family of pseudokinases, their role in cell signalling and cancer. *FEBS J*. 2020 Oct;287(19):4183-4197.
116. Tactacan CM, Phua YW, Liu L, Zhang L, Humphrey ES, Cowley M, et al. The pseudokinase SgK223 promotes invasion of pancreatic ductal epithelial cells through JAK1/Stat3 signaling. *Mol Cancer*. 2015 Jul 29;14:139.
117. Shen Y, Dong S, Liu J, Zhang L, Zhang J, Zhou H, et al. Identification of potential Biomarkers for Thyroid Cancer Using Bioinformatics Strategy: A study Based on GEO datasets. *Biomed Res Int*. 2020.

5. Conclusões e Perspectivas

O presente trabalho possibilitou uma breve discussão acerca dos cânceres de cavidade oral e câncer de tireoide. Bem como a elaboração de um artigo tutorial em português para análise de expressão diferencial de genes utilizando dados presentes no GEO e um artigo com resultados obtidos para análise de expressão diferencial em câncer de cavidade oral e câncer de tireoide. Neste trabalho, utilizando de 6 conjuntos de dados diferentes foram identificados 4 genes potenciais biomarcadores ao diagnóstico de câncer de cavidade oral: LOXL2, FAT1, SOD3 e DPT, sendo FAT1 potencial biomarcador ao prognóstico. Ao mesmo tempo foram identificados 12 genes potenciais biomarcadores ao diagnóstico de câncer de tireoide: CRABP1, CSGALNCT1, WSCD2, ITM2A, ID3, PAX9, SELENBP1, TNFRS11B, SDPR, LRP1B, MT1F e SGK223, sendo CRABP1, ITM2A e ID3 potenciais biomarcadores ao prognóstico.

Todos os genes identificados neste trabalho possuem descrições na literatura relacionadas ao câncer, em sua maioria, relacionadas aos cânceres aqui estudados, fundamentando os resultados encontrados neste trabalho. A exceção se no gene CSGALNCT1, o qual não possui descrições próprias, entretanto há descrições relacionando a família gênica. Ressalta-se que determinados genes apresentam valores aumentados ou diminuídos de expressão em determinados tipos de cânceres, demonstrando a necessidade de maiores trabalhos laboratoriais.

Como perspectivas futuras, espera-se a possibilidade de trabalhos de bancada que possibilitem a fundamentação dos resultados encontrados neste trabalho, bem como a obtenção de outros potenciais biomarcadores que possam ser utilizados como produtos. Ao mesmo tempo, espera-se uma maior coleta de dados que permita uma melhor divisão por sítio anatômico e também por tipo histológico. Vale ressaltar que um maior número de dados poderia possibilitar comparações mais específicas como sexo, idade, etnia, fatores ambientais possibilitando estudos mais precisos e protocolos padronizados.

6. Referências

- Adams, J. U. (2008). Transcriptome: Connecting the Genome to Gene Function. Nature education. *Disponível (online)* 2021, <https://www.nature.com/scitable/topicpage/transcriptome-connecting-the-genome-to-gene-function-605/> (acessado em 20 de setembro).
- Alberts, B. Fundamentos da biologia celular. Tradução: Andrade, A. E. B. *et al.* 2017. 4.ed. Porto Alegre: Artmed.
- Arquette, M.A. Câncer de Cabeça e Pescoço. IN: ORG: Ramaswamy, G. Washington – Manual de Oncologia. 2002, 1.ed. Rio de Janeiro: Editora Guanabara Koogan S.A.
- Brasil. Calendário Nacional de Vacinação. *Disponível (online)* 2021, <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/calendario-de-vacinacao> (acessado em 07 de dezembro).
- Brierley, J., Gospodarowicz, M.K., Wittekind, C. TNM Classification of Malignant Tumors. 2017.8.ed.
- Bumgarner, R. 2013. DNA microarrays: Types, Applications and their future. *Curr Protoc Mol Biol*.
- Cancer Genome Atlas Research Network. 2014. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 513, 202-209.
- Chow, L. Q. M. 2020. Head and Neck Cancer. *The New England Journal of Medicine*. 382, 60-82.
- Dato, S., Crocco, P., Migliore, N.R., Lescai, F. 2021. Omics in a digital world: The role of Bioinformatics in Providing New Insights Into Human Aging. *Frontiers in Genetics*.10.
- Edgar, R., Domrachev, M., Lash, A.E. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*. 30. 207-210.

Galbiatti, A. L. S., Padovani-Junior, J.A., Maníglia, J.V., Rodrigues, C.D.S., Pavarino, E.C., Goloni-Bertollo, E.M. 2013. Head and Neck Cancer: Causes, prevention and treatment. *Brazilian Journal of Otorhinolaryngology*. v. 79. p. 239-247.

Genden, E.M. Head and nec cancer: management and reconstruction. New York: Thieme. 2020.2.ed.

Griffiths, A.J.F., Doebley, J., Peichel, C., Wassarman, D.A. Introdução a genética. Tradução: Sylvia Werdmüller von Elgg Roberto. 2016. 2.ed. Rio de Janeiro: Guanabarra Koogan.

Harden, M.E., Munger, K. 2017. Human Papillomavirus molecular biology. Mutation research. *Reviews in mutation research*. v. 772.

Johnson, D.E., Burtneess, B., Leemans, C.R., Lui, V.W.Y., Bauman, J.E., Grandis, J.R. 2020. Head and Neck squamous cell carcinoma. *Nature reviews*. 92.

Konings, H., Stappers, S, Geens, M, Winter, B.Y., Lamote, K., Meerbeck, J.P., Spencenier, P., Vanderveken, O.M., Ledeganck, K.J. 2020. A literature review of the potential diagnostic biomarkers of Head and Neck Neoplasms. *Frontiers in oncology*. v.10.

Liu, L., Pang, H., He, Q., Pan, B., Sun, X., Shan, J., Wu, L., Wu, K., Yao, X., Guo, Y. 2021. A novel strategy to identify candidate diagnostic and prognostic biomarkers for gastric cancer. *Cancer cell international*.2, 21.

Locke, W. J., Guanzon, D., Ma, C., Liew, Y.J., Duesing, K.R., Fung, K.Y.C., Ross, J.P. DNA Methylation Cancer Biomarkers: Translation to the Clinic. *Frontiers in Genetics*. v. 10.

Pierce, B. A. Genética: um enfoque conceitual. Tradução: Rosário, B. A. 2017. 5.ed. Rio de Janeiro: Guanabara Koogan.

R Core Team (2020). R: A language and enviroment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Rodrigues, A. B. Oncogênese. IN: ORG: Rodrigues, A. B., Martin, L. G. R., Moraes, M. W. Oncologia Multiprofissional: Bases para assistência. 2016. Barueri, SP: Manoele.

- Romanos, M.T.V., Mendes, G.S. 2021. Viroses Oncogênicas. IN: ORG: Santos, N. S. O. *et al.* Virologia Humana.
- Rossum, V., Drake Jr.F,L, 1995. Python reference manual. Centrum voor Wiskunde em informática Amsterdam.
- Reis, J.B., Oliveira, J.M., Nascimento, V.F., Cabral, J.F., Lucietto, G.C., Silva, R.A. 2018. Câncer de Cabeça e pescoço: A comunicação e seus significados. Revista de Enfermagem UFPE. 12, 3263-3270.
- Sartor, I. T. S., Recamonde-Mendonza, M., Ashton-Prolla, P. 2019. TULP3: A potential biomarker in colorretal cancer? Plos one. 14, 1.
- Sartor, I.T.S., Rossetto, M.V. 2020. Análise de dados de expressão gênica. IN: ORG: Silva, S. A.; Notari, D. L.; Dall’Alba, G. Bioinformática: Contexto computacional e aplicações.
- Snustad, D.P., Simmons, M.J. Fundamentos de genética. 2017. 7.ed. Rio de Janeiro: Guanabara Koogan.
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., Zhang, Z. 2017. GEPIA: a web server for cancer and normal gene expression. profiling and interactive analyses. Nucleic Acids Res. 45, w98-w102.
- Terry, M.B., McDonald, J.A., Wu, H.C, Eng, S., Santella, R.M. 2017. Epigenetic Biomarkers of Breast Cancer Risk: Across the Breast Cancer Prevention Continuun. Advances in experimental medicine and biology. 882, 33-68.
- Xu, Y., Wu, G., Li, J., Li, J., Ruan, N., Ma, L., Han, X., Wei, Y., Li, L., Zhang, H., Chen, Y., Xia, Q. 2019. Screening and Identification of Key Biomarkers for Bladder Cancer: A study Based on TCGA and GEO Data. BioMed Research International.2020, 20.
- Wang, Z, Gerstein, M., Snyder, M. 2009. RNA-seq: a revolutionary tool for transcriptomics. Nature reviews genetics. 10, 57-63.
- Wang, H., Peng, R., Wang, J., Qin, Z., Xue, L. 2018. Circulating microRNAs as potential cancer biomarkers: the advatange and disadvantage. Clinical epigenetics. 23. 10.

World Health Organization (WHO). Cancer facts Sheets. *Disponível (online)* 2021, <https://gco.iarc.fr/today/fact-sheets-cancers> (acessado 26 de outubro).

Zhao, X., Sun, S., Zeng, X., Cui, L. 2018. Expression profiles analysis identifies a novel three-mRNA signature to predict overall survival in oral squamous cell carcinoma. *American Journal of Cancer Research*. 8. 450-461.

Zheng, Q., Min, S., Zhou, Q. 2021. Identification of potential diagnostic and prognostic biomarkers for LUAD based on TCGA and GEO databases. *Bioscience Reports*. v.41.