



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
FACULDADE DE ENGENHARIA ELÉTRICA E BIOMÉDICA

CAMILA SINIMBÚ FORTE

**INVESTIGAÇÃO DO PERFIL DE EXPRESSÃO DO *CFTR* EM
SUBTIPOS DE CÂNCER GÁSTRICO: ANÁLISE DE EXPRESSÃO
DIFERENCIAL, COEXPRESSÃO E VIAS BIOLÓGICAS**

BELÉM/PA

2024

CAMILA SINIMBÚ FORTE

**INVESTIGAÇÃO DO PERFIL DE EXPRESSÃO DO *CFTR* EM
SUBTIPOS DE CÂNCER GÁSTRICO: ANÁLISE DE EXPRESSÃO
DIFERENCIAL, COEXPRESSÃO E VIAS BIOLÓGICAS**

Trabalho de Conclusão de Curso, apresentado como
requisito parcial para a obtenção de grau de Bacharel
em Engenharia Biomédica, pela Universidade Federal do
Pará.

Orientador: Prof. Dr. Gilderlanio Santana de
Araújo
Instituto de Ciências Biológicas, Universi-
dade Federal do Pará

BELEM/PA
2024

CAMILA SINIMBÚ FORTE

**INVESTIGAÇÃO DO PERFIL DE EXPRESSÃO DO *CFTR* EM
SUBTIPOS DE CÂNCER GÁSTRICO: ANÁLISE DE EXPRESSÃO
DIFERENCIAL, COEXPRESSÃO E VIAS BIOLÓGICAS**

Trabalho de Conclusão de Curso, apresentado como
requisito parcial para a obtenção de grau de Bacharel
em Engenharia Biomédica, pela Universidade Federal do
Pará.

DATA DE APROVAÇÃO: 28/11/2024

CONCEITO: Excelente

Prof. Dr. Gilderlanio Santana de Araújo
Orientador - PPGBM/ICB/UFPA

Prof. Dr. Antônio Pereira Junior
Membro - FEEB/ITEC/UFPA

Prof. Dr. Fabiano Cordeiro Moreira
Membro - NPO/HUJBB/UFPA

BELÉM/PA
2024

Dedico aos meus pais e a toda a minha família.

"Existem muitas hipóteses na ciência que estão erradas. Isso é perfeitamente aceitável, elas são a abertura para achar as que estão certas."
(Carl sagan)

RESUMO

Para investigar o potencial do gene Regulador de Condutância Transmembrana da Fibrose Cística (*CFTR*) nos subtipos de câncer de estômago realizamos a análise de expressão gênica diferencial em seis subtipos de câncer de estômago usando dados do The Cancer Genome Atlas Stomach Adenocarcinoma (TCGA-STAD). Além disso, para cada subtipo de câncer, exploramos módulos de genes coexpressos utilizando o CEMiTool e análises de super-representação de vias funcionais. Para aprimorar nossa análise, integramos redes de interação gene-gene e análises de ontologia gênica, a fim de identificar o papel do *CFTR* e seu interatoma. A análise de coexpressão demonstrou que o *CFTR* interage com 58 genes no subtipo de câncer de células em anel de sinete, sugerindo que ele desempenha um papel específico em redes regulatórias mais amplas, em vez de exibir mudanças fortes e independentes nos níveis de expressão. Apesar de seu papel hipotetizado, não observamos expressão diferencial significativa do gene *CFTR* em uma análise abrangente de 15 comparações pareadas entre seis subtipos moleculares de adenocarcinoma gástrico para o STAD-TCGA. Embora o *CFTR* não tenha atendido individualmente aos critérios para expressão diferencial, uma investigação mais aprofundada revelou percepções sistêmicas sobre seu envolvimento no câncer gástrico. O *CFTR* pode desempenhar um papel crítico na biologia tumoral, influenciando o comportamento de genes coexpressos e suas vias relacionadas. Esses achados sugerem que o *CFTR* pode influenciar a progressão do câncer por meio de interações complexas gene-gene e modulação de vias.

Palavras-chave: *CFTR*. Cancer Gástrico. Redes de coexpressão. Cancer heterogeneidade.

ABSTRACT

To investigate the potential role of Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*) gene in stomach cancer subtypes we performed differential gene expression analysis across six stomach cancer subtypes using data from the The Cancer Genome Atlas Stomach Adenocarcinoma (TCGA-STAD). Further for each cancer subtype, we explored modules of co-expressed genes using CEMiTool and over-representation analysis of functional pathways. To enhance our analysis, we integrated experimental gene-gene interaction networks and gene ontology analyses, to identify the potential role of *CFTR* and its interactome. Co-expression analysis demonstrated that *CFTR* interacts with 58 genes in the signet ring cell cancer subtype, suggesting it plays a specific role in broader regulatory networks rather than exhibiting strong, independent changes in expression levels. Despite its hypothesized role, we did not observe significant differential expression of the *CFTR* gene in a comprehensive analysis of 15 pairwise comparisons across six molecular subtypes of stomach adenocarcinoma for STAD-TCGA. While *CFTR* did not individually meet the criteria for differential expression, further investigation revealed systemic insights into its involvement in gastric cancer. *CFTR* may play a critical role in tumor biology by influencing the behavior of co-expressed genes and their related pathways. These findings suggest that *CFTR* may influence cancer progression through complex gene-gene interactions by pathway modulation.

Keywords: *CFTR*. Gastric cancer. Co-expression networks. Cancer heterogeneity.

LISTA DE FIGURAS

Figura 1.1 – Localização do gene <i>CFTR</i> situado no braço longo do cromossomo 7 (7q31.2).	12
Figura 1.2 – Arquitetura do <i>CFTR</i>	13
Figura 1.3 – Pipeline de inferência da rede de coexpressão.	20
Figura 4.1 – Pipeline para análise de expressão diferencial, coexpressão diferencial e ontologia gênica	27
Figura 5.1 – Distribuição de contagens de genes hipo e hiper-expressos da análise de expressão diferencial pareada entre subtipos de câncer gástrico (TCGA-STAD).	29
Figura 5.2 – Resultado da expressão diferencial entre os subtipos de câncer gástrico (STAD-TCGA)	30
Figura 5.3 – Expressão do <i>CFTR</i> para cada subtipo de câncer gástrico (TCGA-STAD)	32
Figura 5.4 – Resultados da detecção de módulos de coexpressão gênica para cada subtipo de câncer gástrico (TCGA-STAD), conforme identificado pelo Cemitool.	33
Figura 5.5 – Visualização do Módulo 1 do subtipo de carcinoma de células em anel de sinete (SRCC), destacando <i>CFTR</i> como hub gene.	33
Figura 5.6 – Análise de sobre-representação mostrando as vias biológicas significativas associadas aos genes no módulo 1 do subtipo de carcinoma de células em anel de sinete (SRCC).	34

LISTA DE TABELAS

Tabela 1.1 – Visão geral de alguns bancos de dados de bioinformática	17
Tabela 1.2 – Exemplo de tabela de <i>reads counts</i> para um experimento de RNA-Seq	18
Tabela 4.1 – Distribuição de amostras e dados clínicos para as amostras de subtipos de adenocarcinoma gástrico do Programa The Cancer Genome Atlas (STAD-TCGA).	24
Tabela 5.1 – Contagem de genes hipo e hiper-expressos para cada comparação entre os subtipos de câncer gástrico (TCGA-STAD).	28
Tabela 5.2 – Expressão diferencial do gene <i>CFTR</i> para todas as comparações entre os subtipos de câncer gástrico (TCGA-STAD).	31
Tabela 5.3 – Análise de enriquecimento funcional por Ontologia Gênica (Go) dos genes coexpressos em conjunto do <i>CFTR</i> no módulo 1 do subtipo de carcinoma de células em anel de sinete (SRCC).	35

LISTA DE SIGLAS

AC-NOS	Adenocarcinoma Não Especificado
cDNA	DNA Complementar
CFTR	Regulador da Condutância Transmembrana na Fibrose Cística
CCR	Câncer colorretal
DC	Carcinoma do Tipo Difuso
DEGs	Genes diferencialmente expressos
DNA	Ácido Desoxirribonucleico
ENaC	Canal de Sódio Epitelial
FC	Fibrose Cística
FDR	Taxas de Descobertas Falsas (False Discovery Rate)
IA	Adenocarcinoma Intestinal
logFC	Log Fold Change
MA	Adenocarcinoma Mucinoso
mRNA	RNA Mensageiro
NBD	Domínio de Ligação de Nucleotídeos Cistólicos
NGS	Sequenciamento de nova geração
ORA	Análise de Super-Representação de Vias Funcionais (Over-representation analysis)
PKA	Proteína Kinase A
PKC	Proteína Kinase C
R-domain	Domínio Regulatório
RNA	Ácido Ribonucleico
RNA-Seq	Sequenciamento de RNA (RNA Sequencing)
SRCC	Carcinoma de Células em Anel de Sinete
STAD-TCGA	The Genome Cancer Atlas Stomach Adenocarcinoma
TA	Adenocarcinoma Tubular
TGI	Trato gastrointestinal
TMD	Domínio Transmembranar

LISTA DE SÍMBOLOS

Cl^-	Cloreto
NO_3^-	Nitrato
HCO_3^-	Bicarbonato
Na^+	Sal
Ca^+	Cálcio

Sumário

1	CONTEXTUALIZAÇÃO	12
1.1	O gene e a proteína CFTR	12
1.1.1	Classificação Funcional das mutações do CFTR	14
1.2	CFTR e o trato gastrointestinal	15
1.3	Biologia de Sistemas	16
1.3.1	Bancos de dados	16
1.3.2	Análise transcriptômica	17
1.3.3	Análise de expressão diferencial	18
1.3.4	Análise de coexpressão	19
1.3.5	Ontologia Gênica	20
2	JUSTIFICATIVA	22
3	OBJETIVOS	23
3.1	Objetivo geral	23
3.2	Objetivos específicos	23
4	METODOLOGIA	24
4.1	RNA-seq para seis subtipos de câncer gástrico	24
4.2	Análise de expressão diferencial	25
4.3	Análise de Rede de Coexpressão e Identificação de Genes Hub	25
4.4	Análise de Ontologia Gênica	26
4.5	Fluxograma da metodologia	27
5	RESULTADOS E DISCUSSÃO	28
5.1	Heterogeneidade molecular no câncer gástrico	28
5.2	CFTR não apresenta expressão diferencial nos subtipos de câncer gástrico	31
5.3	O interatoma do <i>CFTR</i> é coexpresso exclusivamente no SRCC	32
5.4	Discussão	36
6	CONCLUSÃO	38
6.1	Trabalhos Futuros	38
	REFERÊNCIAS	39

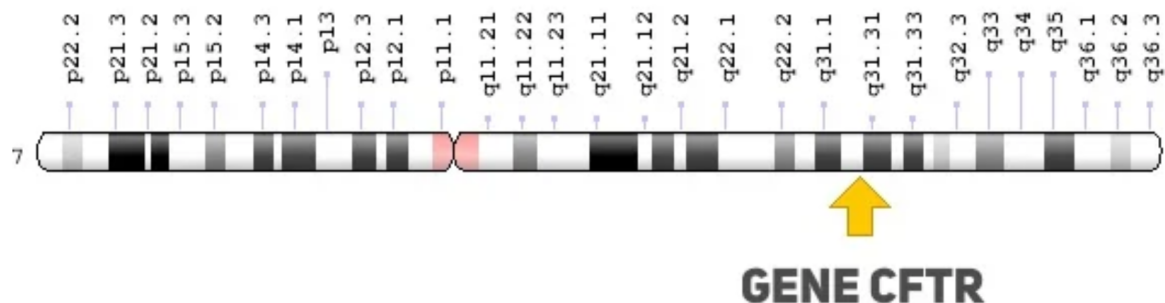
1 CONTEXTUALIZAÇÃO

Devido o avanço da triagem neonatal, tratamentos como nutrição, antibióticos e transplantes de órgãos, a expectativa de vida de pacientes acometidos por Fibrose cística (FC) aumentou. Apesar disso, síndromes novas relacionadas à FC começaram a ser diagnosticadas. O câncer gastrointestinal vem sendo associado como comorbidade da idade adulta em pacientes FC (SLAE; WILSCHANSKI, 2018). Estudos recentes apontam o gene Regulador de Condutância Transmembrana da Fibrose Cística (CFTR), tipicamente associado à fibrose cística, como tendo um papel potencial em vários tipos de câncer (LIU et al., 2014; BHATTACHARYA et al., 2022; HOU et al., 2016; MAKAROVA et al., 2023; PARISI et al., 2023).

1.1 O gene e a proteína CFTR

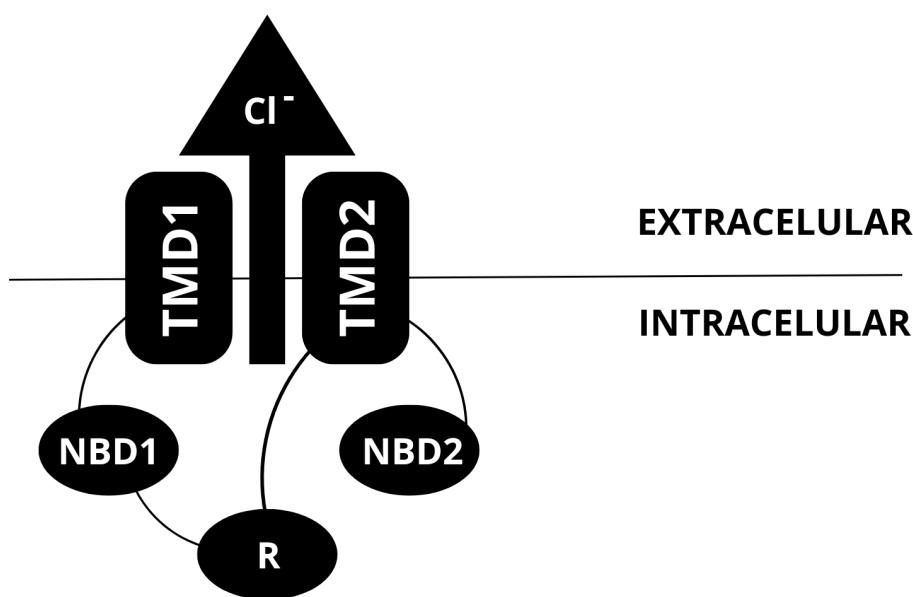
O gene Regulador da Condutância Transmembrana na Fibrose Cística (*CFTR*, também conhecido como *ABCC7*), localizado no cromossomo 7 (7q31.2) contém 27 exons e 180.000 pares de bases (Figura 1.1), é responsável por codificar a proteína de mesmo nome (BAREIL; BERGOUGNOUX, 2020). A proteína CFTR pertence a superfamília transportadora ABC (ATP-binding-cassette), encarregada de transportar uma grande variedade de substratos de maneira dependente de energia; sendo a proteína CFTR a única que atua como canal iônico (LUKASIAK; ZAJAC, 2021). A proteína CFTR regula e conduz Cl⁻ através das membranas apicais de células epiteliais, é essencial para o equilíbrio osmótico e viscosidade do muco, além de secretar e absorver eletrólitos e fluídos (HANSSENS; DUCHATEAU; CASIMIR, 2021).

Figura 1.1 – Localização do gene CFTR situado no braço longo do cromossomo 7 (7q31.2).



A arquitetura da proteína contém dois domínios transmembranares (TMDs), 2 domínios de ligação de nucleotídeos cistólicos (NBDs) e um domínio regulatório (R-domain), que contém sites de fosforilação; a abertura do canal acontece pela fosforilação do domínio regulatório pela proteína kinase A (PKA) e/ou proteína kinase C (PKC), esse processo muda a afinidade do CFTR para o ATP, para que este se ligue nos NBDs (dimerização); enquanto que o fechamento ocorre com a hidrólise do ATP, resultando na separação do dímero do NBD e a restauração da conformação do TMD (LUKASIAK; ZAJAC, 2021; HANSSENS; DUCHATEAU; CASIMIR, 2021) (Figura 1.2).

Figura 1.2 – Arquitetura do CFTR.



Legenda: TMDs = domínios transmembranares; NBDs = domínios de ligação de nucleotídeos cistólicos; R = domínio regulatório; Cl^- = Cloreto.

Fonte: Autoria própria

O *CFTR* é comumente conhecido como um canal de Cl^- , no entanto não é exclusivo para este anion; é permeável para anions como NO_3^- , e HCO_3^- , e também regula o Canal de Sódio Epitelial (ENaC) (LUKASIAK; ZAJAC, 2021). O canal CFTR secreta o Cl^- através da membrana apical e cria uma força motriz que secreta o Na^+ pelas vias paracelulares, e por osmose, a água segue pela mesma via ou pelas aquaporinas. Em condições saudáveis, a secreção e absorção de Cl^- e Na^+ é essencial para a regulação do volume e líquido na superfície das vias aéreas, prevenindo a desidratação e garantindo a limpeza mucociliar adequada. Uma proteína CFTR desfuncional compromete a secreção e absorção do Cl^- , causando uma despolarização do potencial da membrana apical, prevenindo a difusão passiva de Na^+ e resultando em uma alta concentração de Na^+ e Cl^- no lumen, desidratando a superfície das vias aéreas, contribuindo para uma alta viscosidade do muco e restringindo a função de limpeza dos cílios, predispõe consequentemente o tecido a inflamações e infecções (ANDERSON; CORMIER; SCOTT, 2019).

A proteína CFTR também secreta HCO_3^- em diferentes tecidos, este é importante para o controle do PH de fluídos na superfície de células epiteliais e para a viscosidade do muco, dependente das mucinas que necessitam do HCO_3^- . O momento mais importante na

geração do muco saudável fisiologicamente é a secreção para fora do meio intracelular; para que isso seja possível precisa haver uma rápida remoção de íons cálcio da mucina, que se dá pela ação do bicarbonato que aumenta o PH intracelular e precipita os íons cálcio para fora, permitindo a expansão da mucina e a formação de uma rede de muco extracelular. Uma secreção desfuncional do HCO_3^- prejudica a remoção do Ca^{+} , impossibilitando a expansão normal da mucina, assim estas tendem a se manter extremamente aglomeradas, pouco solubilizadas, resultando em um aumento da viscosidade. Um muco anormal permite o contato de bactérias com o epitélio, além da obstrução de órgãos (BAREIL; BERGOUGNOUX, 2020).

Atualmente há mais de 2000 variantes do gene CFTR catalogadas, que podem ser divididas em dois tipos de classificações (BAREIL; BERGOUGNOUX, 2020):

1. Clínica: de acordo com o fenótipo associado a Fibrose Cística (FC), as variantes são classificadas como causadoras da FC ou como doenças relacionadas ao CFTR;
2. Funcional: de acordo com o impacto da mutação na proteína, dividido em 6 classes.

1.1.1 Classificação Funcional das mutações do CFTR

A classificação funcional se divide da seguinte forma (BAREIL; BERGOUGNOUX, 2020; SLAE; WILSCHANSKI, 2021; BOECK; AMARAL, 2016):

- **Classe I** - Síntese defeituosa da proteína, resulta na ausência de proteína funcional da superfície da célula uma vez que a proteína não sai do núcleo. Inclui mutações nonsense, missense e frameshift;
- **Classe II** - Processamento e maturação defeituosa da proteína (misfolding), fica retida no retículo endoplasmático, impedida de trafegar até a superfície apical da célula. Inclui a mutação mais frequente entre a população caucasiana, a F508del, além de missenses, em que sua patogenicidade depende das propriedades do aminoácido atingido e sua localização;
- **Classe III** - Proteína sem função na superfície da célula, regulação do canal anormal, defeituoso. Inclui mutações missense, localizadas principalmente nas sequências que codificam os domínios de ligação a nucleotídeos, impedindo a ligação e hidrólise do ATP em um dos NBDs;
- **Classe IV** - Condutância do canal defeituosa, função reduzida. Inclui mutações missense, localizadas sobretudo nas sequências que codificam os domínios transmembranares envolvidos na formação do poro do canal;
- **Classe V** - Quantidade reduzida da proteína na superfície apical da célula. Inclui variantes que causam um splicing anormal e intrônicas, as quais sua patogenicidade depende da quantidade residual de transcritos normais;
- **Classe VI** - Proteína com baixa estabilidade (maior endocitose ou menor reciclagem das moléculas de CFTR na superfície celular).

As mutações das classes I - III são dadas como severas (tanto para heterozigotos como homozigotos) em razão de conferirem nenhuma proteína funcional na membrana apical; as mutações das classes IV-VI tem consequências brandas uma vez que ainda têm função residual do canal iônico.

1.2 CFTR e o trato gastrointestinal

O trato gastrointestinal (TGI) é um sistema complexo que digere e absorve nutrientes e excreta resíduos, sendo essencial para a nutrição, imunidade e bem-estar. Distúrbios no TGI podem causar várias condições de saúde. Recentemente, descobriu-se que a fibrose cística afeta significativamente o TGI, gerando complicações que comprometem a nutrição e a saúde geral dos pacientes (SLAE; WILSCHANSKI, 2021).

O *CFTR* é expresso elevadamente na superfície apical do epitélio intestinal, nas células-tronco nas criptas do intestino delgado, na região proximal e ceco no colon, além das células ductais do pâncreas (ANDERSON; CORMIER; SCOTT, 2019). As células-tronco na base das criptas do intestino são responsáveis pela renovação do epitélio intestinal e é sugerido que estas podem iniciar o processo de oncogênese do câncer colorretal (CCR), sendo assim, o *CFTR* sendo altamente expresso nas criptas pode influenciar no processo de renovação epitelial e consequentemente na oncogênese do CCR quando em disfunção (ANDERSON; CORMIER; SCOTT, 2019). A disfunção da proteína resulta em secreções viscosas, obstruindo dutos e o intestino, em razão disso, pacientes de FC apresentam sintomas gastrointestinais desde a infância como o Íleo Meconial (ocorre em 15% dos pacientes), que se manifesta nos primeiros dias de vida, o infante apresenta distensão abdominal e vômito bilioso; na idade adulta é conhecido como Síndrome da Obstrução Intestinal Distal, dá-se em um bloqueio crônico e recorrente no íleo terminal e no colon proximal a partir de um conteúdo intestinal denso e aderente (SLAE; WILSCHANSKI, 2021).

As mutações do *CFTR*, particularmente a mutação $\Delta F508$, têm sido associadas a um risco aumentado de desenvolver várias malignidades, incluindo câncer gástrico, é destacado que pacientes com fibrose cística exibem um risco elevado ao longo da vida para cânceres do trato digestivo, incluindo o câncer gástrico (SUGUNARAJ et al., 2016). O estudo de Liu et al. demonstrou que a expressão do *CFTR* aumenta com a idade e está associado com o estágio clínico do câncer gástrico, além de ser um potencial biomarcador deste câncer em conjunto do gene *CA199* (LIU et al., 2014). Essa associação sugere que o papel do *CFTR* vai além da função de transporte iônico, influenciando a tumorigênese por meio de predisposição genética. Esse comportamento também corrobora um estudo experimental realizado por Than et al., que catalogou o *CFTR* como um gene supressor de tumor no trato intestinal, mas também mostrou que a deleção do *CFTR* causou altas taxas de tumores intestinais (THAN et al., 2016).

O envolvimento do *CFTR* no câncer gástrico é ainda mais apoiado por seu papel regulador em vias inflamatórias. O *CFTR* demonstrou suprimir a sinalização inflamatória mediada pelo NF- κ B, que muitas vezes está implicada na progressão do câncer (HUNTER et al., 2010). A inflamação crônica é um fator de risco bem estabelecido para o câncer gástrico, e a desregulação do *CFTR* pode contribuir para o microambiente inflamatório que promove o desenvolvimento de tumores (HOU et al., 2016). Inflamações crônicas provocadas por úlceras, por exemplo, são um fator de risco para alterações no tecido gástrico e potencialmente o desenvolvimento de câncer (SOGAARD et al., 2016). O *CFTR* é uma das proteínas transportadoras de HCO_3^- essenciais nas células epiteliais e tem um papel fundamental na regulação da secreção de HCO_3^- na mucosa duodenal (WEN et

al., 2016). Segundo o estudo de Wen et al., a infecção por *Helicobacter pylori* diminuiu a expressão e a atividade funcional do CFTR na mucosa duodenal pela via de sinalização do $TGF\beta$, o que contribui para o desenvolvimento de úlcera duodenal (WEN et al., 2018).

1.3 Biologia de Sistemas

A biologia de sistemas é uma área emergente da biologia computacional dedicada a entender a complexidade dos sistemas biológicos de maneira integrada. A área combina dados e modelos de diferentes áreas, como genômica e transcriptômica, e utiliza métodos computacionais, estatísticos e matemáticos para compreender a regulação e a maquinaria celular. Seus objetivos principais incluem a integração de dados "ômicos" (como genômica e transcriptômica), a modelagem computacional, para prever o comportamento de sistemas complexos, e a análise de interações moleculares, que ajuda a entender a homeostase celular e as causas de doenças. Esse campo é essencial para a medicina personalizada, pois permite a descoberta de biomarcadores e o desenvolvimento de terapias específicas para doenças complexas como o câncer (VOIT, 2022).

A pesquisa frequentemente inicia com a coleta e análise de informações disponíveis publicamente. A seguir apresentamos algumas bases de dados.

1.3.1 Bancos de dados

Bancos de dados públicos como , *BioGRID*, *Reactome* e *TCGA* oferecem acesso a grandes volumes de dados biológicos e são essenciais para a pesquisa na bioinformática. (Tabela 1.1).

- **BioGRID (Biological General Repository for Interaction Datasets)** é um banco de dados público que reúne informações sobre interações biológicas, incluindo interações proteína-proteína, gene-gene e interações químicas, em diversos organismos. Compilado a partir de publicações científicas e experimentos, o BioGRID oferece dados atualizados, sendo uma ferramenta essencial para pesquisas em biologia molecular, é ideal para o mapeamento de redes de interação e para a compreensão das funções de proteínas e genes em processos biológicos (STARK et al., 2006).
- **Reactome** é um banco de dados gratuito que fornece informações detalhadas sobre vias e reações biológicas, especialmente em humanos. O reactome mapeia diversas vias metabólicas e de sinalização, documentando reações químicas e interações celulares. Mantido por especialistas, é constantemente atualizado com dados científicos confiáveis e se integra a outros bancos e ferramentas bioinformáticas, como Ensembl e KEGG, facilitando a pesquisa interdisciplinar (MILACIC et al., 2023).
- **TCGA (The Cancer Genome Atlas)** é um catálogo com alterações genômicas associadas ao câncer, seu objetivo é melhorar o diagnóstico, tratamento e prevenção do câncer a partir da caracterização das mudanças moleculares em vários tipos de câncer. Atualmente o banco de dados conta com sequências genômicas, transcriptômicas e dados clínicos de mais de 20 mil tumores em 33 diferentes tipos de câncer que podem ser acessados livremente pela comunidade científica (CAROLYN; ZENKLUSEN, 2018).

Tabela 1.1 – Visão geral de alguns bancos de dados de bioinformática

Nome	Conteúdo	Números	URL	Referência
The Biological General Repository for Interaction Datasets (BioGRID)	Arquiva e dissemina dados de interação genética e proteica de organismos modelo e humanos.	1.740.000 interações curadas a partir de conjuntos de dados derivados de mais de 70.000 publicações na literatura	< https://thebiogrid.org/ >	(STARK et al., 2006)
Reactome	Banco de dados open-source de vias de sinalização de acesso livre, curado manualmente e revisado por pares	Compõe-se de 15.492 reações bioquímicas organizadas em 2.742 vias, envolvendo 30.892 proteínas codificadas por 11.289 genes humanos diferentes, com suporte de 39.318 referências bibliográficas	< https://reactome.org/ >/	(MILACIC et al., 2023)
The Cancer Genome Atlas Program (TCGA)	Catálogo com alterações genômicas associadas ao câncer	Mais de 2,5 petabytes de dados genômicos, epigenômicos, transcriptômicos e proteômicos, caracterizando molecularmente mais de 20.000 amostras primárias de câncer e amostras normais correspondentes, abrangendo 33 tipos de câncer	< https://www.cancer.gov/ccg/research/genome-sequencing/tcga >	(CAROLYN; ZENKLUSEN, 2018)

Fonte: Autoria própria

1.3.2 Análise transcriptômica

O ácido ribonucleico (RNA) é uma molécula polimérica essencial para funções biológicas, incluindo o transporte de instruções genéticas, o carregamento de aminoácidos e a composição de estruturas celulares. O ácido desoxirribonucleico (DNA) codifica a informação genética de um organismo, que se manifesta por meio da transcrição. As moléculas de RNA produzidas a partir da transcrição do DNA são chamadas de transcritos. O conjunto total de transcritos de RNA em um organismo é denominado transcriptoma (LOWE et al., 2017).

A análise transcriptômica permite medir a expressão gênica em diferentes tecidos e condições, revelando detalhes sobre a regulação dos genes e processos biológicos. A expressão gênica é o processo pelo qual a informação contida em um gene é convertida em um produto funcional, geralmente uma proteína. O aumento da expressão gênica (hiperexpressão) resulta em uma maior produção do produto gênico, enquanto a diminuição da expressão (hipo-expressão) reduz essa produção. A expressão gênica pode variar entre diferentes tipos celulares e tecidos, ser influenciada pelo estágio de desenvolvimento e mudar

ao longo do tempo. Além disso, essa expressão pode ser modulada por sinais externos ou pela ação de repressores que inibem a transcrição do gene. Estudar a expressão gênica é crucial para compreender o desenvolvimento de doenças, como o câncer, pois permite identificar padrões genéticos específicos em amostras tumorais e entender os mecanismos de regulação gênica (LANG et al., 2023; ERGIN; KHERAD; ALAGOZ, 2022).

Tecnologias como o Sequenciamento de Nova Geração (RNA-Sequencing, RNA-Seq) possibilitam a quantificação precisa de transcritos e a identificação de novos genes e variantes, sendo capazes de detectar transcritos em baixas quantidades, ao contrário de outros métodos, como microarrays (LOWE et al., 2017). A partir do sequenciamento, são obtidas sequências de nucleotídeos denominadas de “*reads*”, que representam fragmentos do DNA complementar (cDNA) derivados do RNA mensageiro (mRNA), e então são contadas para estimar quantas vezes cada gene foi lido, gerando uma matriz de contagens (*counts*) que representa a abundância de transcritos (genes) em diferentes amostras (Tabela 1.2).

1.3.3 Análise de expressão diferencial

Para uma matriz de *counts*, refere-se um conjunto de contagens de *reads* (*reads counts*) para uma amostra como uma *biblioteca*, e o número total de *reads* em uma biblioteca como o *tamanho da biblioteca* (CHEN; LUN; SMYTH, 2018).

Tabela 1.2 – Exemplo de tabela de *reads counts* para um experimento de RNA-Seq

	Caso		Controle	
	Amostra 1	Amostra 2	Amostra 3	Amostra 4
Gene 1	24	31	74	59
Gene 2	0	13	7	2
...	...	...	...	...
Total	22.341.344	20.132.435	12.543.543	34.875.342

Fonte: Autoria própria

Na tabela 1.2 cada coluna corresponde a uma amostra caso ou controle, cada célula é o número de *reads* mapeadas para um gene específico em uma amostra específica. A soma das contagens em cada coluna (última linha) é o tamanho da biblioteca para cada amostra correspondente. O propósito da análise de expressão diferencial é verificar se a proporção de um mesmo gene difere entre as condições experimentais, caso ou controle, segundo o exemplo da tabela (CHEN; LUN; SMYTH, 2018).

A contagem de *reads* são números inteiros discretos que apresentam forte relação de média-variância. Um gene pouco replicado pode acarretar em uma falsa aproximação da sua verdadeira variabilidade de expressão na amostra. Essa variabilidade pode ser dividida em duas: técnica e biológica. A variabilidade técnica é referente aos aspectos técnicos do sequenciamento que incluem a contaminação das amostras, qualidade do sequenciamento e o processamento de dados. De outro modo, a variabilidade biológica se refere às diferenças naturais da expressão gênica nas condições experimentais como diferenças genéticas entre indivíduos, condições ambientais, estágio do desenvolvimento da amostra, entre outros (CHEN; LUN; SMYTH, 2018). Para genes com poucas *reads*, a variabilidade técnica é dominante, a medida que aumentam, as diferenças reais biológicas entre as amostras superam as variações técnicas. É importante que ocorra a distinção entre as variabilidades no momento da análise do RNA-Seq para que a interpretação dos resultados seja confiável

e acurado. Para tal, pacotes como *EdgeR* utilizam Modelos lineares Generalizados e Métodos Empíricos Bayesianos para estimar a relação e evitar o viés de média-variância mesmo quando o número de réplicas é muito pequeno (CHEN et al., 2024).

Modelos Lineares Generalizados (GLM) permitem a análise de amostras que não seguem uma distribuição normal a partir da distribuição da família exponencial. A distribuição binomial negativa (NB) é frequentemente utilizada para a modelagem da contagem de *reads* pois permite que a variância seja maior que a média (superdispersão), característica comum dos dados de RNA-Seq (CHEN et al., 2024). A estimação dos parâmetros da distribuição Binominal Negativa (NB) é feita pela estratégia empírica de verossimilhança ponderada bayesiana. A abordagem bayesiana empírica estima a distribuição diretamente dos dados disponíveis observados, ao invés de especificá-la antes da coleta de dados (CHEN; LUN; SMYTH, 2018).

Na análise de expressão diferencial em dados de RNA-Seq, métricas estatísticas ajudam a interpretar mudanças na expressão gênica entre condições. O log fold change (logFC) quantifica a alteração na expressão de um gene: valores positivos indicam maior expressão na condição experimental. O log Counts Per Million (logCPM) normaliza as contagens para comparações entre amostras de diferentes profundidades e fornece uma ideia da abundância relativa do gene em relação ao total de reads na amostra. O Likelihood Ratio (LR) avalia a consistência dos dados com a hipótese de diferença na expressão, um valor alto sugere que os dados são mais consistentes com a hipótese alternativa. O valor de p (p-value) indica a probabilidade de observar os dados sob a hipótese nula, um valor baixo (geralmente $< 0,05$) sugere que há evidências suficientes para rejeitar a hipótese nula, indicando que o gene é considerado diferencialmente expresso entre as condições. A taxa de descobertas falsas (FDR) estima a proporção de falsos positivos, é geralmente utilizado como critério para considerar um gene como diferencialmente expresso ($FDR < 0,05$), indicando baixa probabilidade de que os genes identificados como significativos sejam falsos positivos.

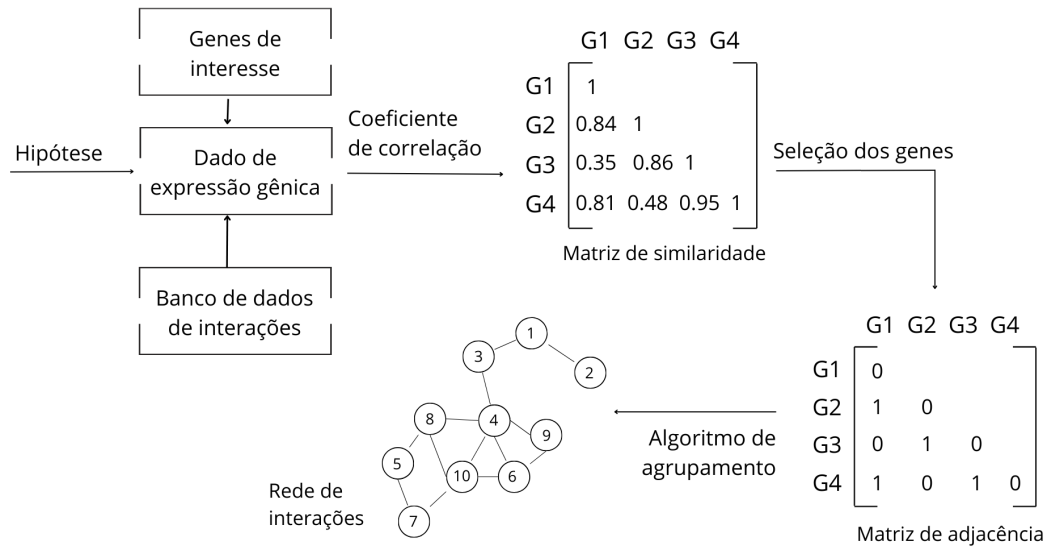
1.3.4 Análise de coexpressão

A análise de coexpressão identifica padrões de expressão entre genes em diferentes condições, a abordagem infere redes regulatórias e preve funções gênicas a partir do cálculo da correlação dos níveis de expressão gênicas (DAM et al., 2017). As matrizes de *counts* são utilizadas para calcular um escore de similaridade entre pares de genes, com base nos padrões de expressão e então são formadas redes de coexpressão, em que os genes são agrupados em módulos de acordo com sua função biológica ou participação em vias metabólicas; é extremamente útil em situações em que a função de um gene em um organismo ainda é desconhecida, e pode ser inferida com base nas funcionalidades conhecidas de outros genes com os quais está coexpresso na rede (SERIN et al., 2016). Além disso, existem genes que desempenham papéis centrais em processos biológicos, os *hub* genes, são aqueles com alto grau de conectividade (quantos genes estão ligados a um específico) com funções regulatórias ou estruturais significativas.

A hipótese direciona a estratégia para a análise da coexpressão diferencial, com o conhecimento prévio da investigação são escolhidos os genes de interesse e bancos de dados de interações podem ser consultados para analisar padrões, a Figura 1.3 ilustra o processo da análise de coexpressão (CHOWDHURY; BHATTACHARYYA; KALITA, 2020). Inúmeros bancos de dados públicos fornecem conjuntos de dados úteis para a montagem da rede de coexpressão, como o reactome e o biogrid, estes podem consultados

de forma global (sem especificação) ou direcionada, guiado pelos genes de interesse (SERIN et al., 2016).

Figura 1.3 – Pipeline de inferência da rede de coexpressão.



Legenda: Após a hipótese guiar a identificação dos genes de interesse e a consulta do banco de dados, a similaridade entre os padrões de expressão gênica são avaliados por coeficientes de correlação como Pearson e Spearman. Um limiar é definido (0,8 no exemplo) que seleciona os genes e cria a matriz de adjacência binária, os genes que apresentam coexpressão significativa são marcados com o valor 1. Por fim aplica-se um algoritmo de agrupamento que identifica as redes de genes significativamente coexpressos. No exemplo, os genes estão representados por nós numerados (vértices) que se conectam através de arestas (links), facilitando a interpretação das relações entre os diferentes genes.

Fonte: Autoria própria

1.3.5 Ontologia Gênica

A Ontologia gênica (Gene Ontology) é uma iniciativa que busca unificar a representação de genes e seus produtos em diferentes organismos, estabelecida em 1998, seu desenvolvimento acompanha o ritmo das descobertas científicas, com a adição de novos termos, ajuste ou descontinuação de outros, refletindo o estado atual do conhecimento científico (CONSORTIUM, 2018). Em 2023, a base de dados já contemplava mais de 45.000 termos interconectados, curada por uma equipe internacional de especialistas que analisam minuciosamente a literatura científica para manter a precisão e atualidade das informações (CONSORTIUM et al., 2023).

A estrutura da Gene Ontology se organiza em três aspectos que proporcionam uma linguagem padronizada para a análise comparativa e compreensão das funções gênicas em diversos contextos biológicos (ASHBURNER et al., 2000; CONSORTIUM, 2018; CONSORTIUM et al., 2023):

1. **Função Molecular:** caracteriza a atividade bioquímica específica dos produtos gênicos no nível molecular. Esta categoria descreve ações precisas como catálise e transporte, seja por produtos gênicos individuais ou por complexos macromoleculares, estas descrições focam exclusivamente na ação realizada, independentemente do contexto

temporal ou espacial, permitindo sua aplicação universal em diferentes organismos e sistemas biológicos.

2. **Componente Celular:** mapeia a distribuição espacial dos produtos gênicos dentro da estrutura celular, identificando a localização. Esta categorização abrange desde estruturas universalmente presentes em células (como núcleo e mitocôndrias) até componentes específicos de determinados tipos celulares ou organismos. Este conhecimento é fundamental para compreender as interações e redes moleculares no contexto celular.
3. **Processo Biológico:** se refere a sequências de eventos ou atividades geneticamente programadas que visam alcançar um resultado específico no organismo. Cada processo é definido por seu resultado, como a divisão celular, que resulta na formação de duas células-filhas a partir de uma célula-mãe. Estes processos representam a integração de múltiplas funções moleculares e componentes celulares, operando em conjunto para alcançar objetivos biológicos definidos.

2 JUSTIFICATIVA

As estimativas globais de incidência de câncer, baseadas nas últimas atualizações da Agência Internacional para Pesquisa em Câncer (IARC) reportadas em 2022, indicam que o câncer de estômago ocupa o quinto lugar em termos de incidência (novos casos = 968.350) e mortalidade ($n = 659.853$) em todo o mundo (BRAY et al., 2024). Em comparação com outros países, a China lidera no diagnóstico em estágio inicial desse câncer. Apesar dos avanços no diagnóstico e tratamento, os mecanismos moleculares subjacentes aos vários subtipos de câncer gástrico ainda são pouco compreendidos (PUCCINI et al., 2022; ZHAO et al., 2023).

Com base na literatura mencionada, poucos estudos associaram a mutação e a expressão do gene *CFTR* com o câncer gástrico (BHATTACHARYA et al., 2022; HOU et al., 2016; LIU et al., 2014; PARISI et al., 2023). Assim, o papel biológico do *CFTR* ainda não é completamente compreendido. Além disso, esses estudos não exploraram a especificidade tecidual ou tumoral do *CFTR* no câncer gástrico, apresentando, portanto, resultados mais amplos. Isso ressalta a importância de realizar análises sistêmicas para descrever melhor a relação entre *CFTR* e os subtipos de câncer gástrico.

Neste estudo, investigamos a expressão gênica diferencial do gene *CFTR* em tecidos, sua coexpressão e sua rede, para compreender melhor as interações gene-gene e as vias em seis subtipos de câncer de estômago. Nosso objetivo foi identificar possíveis interações gênicas que pudessem prever os subtipos gástricos. Nossas análises investigaram módulos distintos de genes que incluem o *CFTR* como interator.

3 OBJETIVOS

3.1 Objetivo geral

Investigar a expressão diferencial, coexpressão e ontologia gênica do gene CFTR em subtipos de câncer gástrico.

3.2 Objetivos específicos

- Pré-processar os dados de sequenciamento RNA-Seq em amostras de câncer gástrico do conjunto de dados The Genome Cancer Atlas Stomach Adenocarcinoma (STAD-TCGA);
- Analisar a expressão gênica diferencial entre os subtipos de câncer gástrico;
- Analisar os padrões de coexpressão gênica para cada subtipo de câncer gástrico;
- Modelar redes de interações a partir da análise de coexpressão de amostras de subtipos de câncer gástrico;
- Análise de vias, processos e funções moleculares de redes que envolvem o CFTR como modulador da expressão gênica.

4 METODOLOGIA

4.1 RNA-seq para seis subtipos de câncer gástrico

Utilizando a ferramenta TCGABiolinkGUI (COLAPRICO et al., 2015), acessamos remotamente os dados clínicos e moleculares de cada amostra da coleção The Cancer Genome Atlas Stomach Adenocarcinoma (TCGA-STAD) (NETWORK, 2014). Os dados de expressão gênica, gerados por RNA-seq, foram extraídos para seis subtipos de câncer gástrico. O TCGA-STAD é composto de 154 amostras de adenocarcinoma não especificado (AC-NOS), referente a tumores gástricos que não atendem aos critérios diagnósticos de outros tipos conhecidos de câncer, 79 de adenocarcinoma intestinal (IA), 73 de adenocarcinoma tubular (TA), 66 de carcinoma do tipo difuso (DC), 17 de adenocarcinoma mucinoso (MA) e 13 de carcinoma de células em anel de sinete (SRCC). As amostras de tecidos são referentes a 146 mulheres e 264 homens, com idade média de 66 anos submetidos a terapia farmacêutica, ou radioterápica, conforme descrito na Tabela 4.1.

Tabela 4.1 – Distribuição de amostras e dados clínicos para as amostras de subtipos de adenocarcinoma gástrico do Programa The Cancer Genome Atlas (STAD-TCGA).

		Amostras	Média	SD
Subtipo de Adenocarcinoma do estômago	AC-NOS	154		
	IA	79		
	TA	73		
	DC	66		
	MA	17		
	SRCC	13		
Gênero	Feminino	146		
	Masculino	264		
Idade		410	66	11
Tratamento terapêutico				
	Não	180		
	Não reportado	43		
	Sim	187		
Tratamento radioterápico				
	Não	293		
	Não reportado	39		
	Sim	78		

Legenda: AC-NOS: adenocarcinoma não especificado. IA: adenocarcinoma intestinal. TA: adenocarcinoma tubular. DC: carcinoma do tipo difuso. MA: adenocarcinoma mucinoso. SRCC: carcinoma de células em anel de sinete. SD: desvio padrão.

4.2 Análise de expressão diferencial

A análise de expressão diferencial foi realizada utilizando o pacote *EdgeR* no R 4.4.1 e RStudio (v4.1) (CHEN et al., 2024). Um total de 15 comparações pareadas foram conduzidas entre seis subtipos de câncer gástrico usando o *EdgeR*. Inicialmente os genes com poucas contagens de leituras são filtradas antes da análise principal. A função *filterByExpr* é utilizada para manter somente os genes que têm contagens suficientes para atingir significância estatística e biológica (*reads* ≥ 10). Em seguida a normalização aplicada foi a contagem por milhão (counts-per-million, CPM).

Posteriormente, considera-se uma matriz de *design* que comporta o fenótipo e as covariáveis do experimento tal como o subtipo, gênero, idade, e tratamento (farmacêutico e radioterapia) a partir dos dados clínicos disponibilizados. Além de um objeto *DGEList*, o “container” de dados do *EdgeR* com a matriz de counts pareados entre os 6 subtipos de adenocarcinoma do estômago, especificando a covariável de comparação que são os subtipos.

A função *calcNormFactors* normaliza a contagem de *reads* do objeto *DGEList* a partir das diferenças na profundidade de sequenciamento entre as amostras, os fatores calculados pela função são as diferenças das distribuições de contagens, um número excessivo de reads atribuídos a genes altamente expressos reduzem a profundidade disponível para outros, sendo assim, a normalização garante que as comparações entre as amostras não sejam enviesadas, minimizando o impacto das variações técnicas em análises futuras.

A estimativa de dispersão é calculada com a função *estimateDisp* que implementa a abordagem bayesiana empírica de verossimilhança ponderada, descrita anteriormente, e recebe o objeto *DGEList* e a matriz de design. Os genes diferencialmente expressos (DEGs) na comparação pareada são então calculados pelo modelo linear generalizado binomial negativo pela função *glmFit*. Por fim é aplicado um teste de razão de verossimilhança para calcular os p-valores da expressão diferencial dado pela função *glmLRT*.

O resultado estatístico de cada gene é individualmente composto por: “logFC”, “logCPM”, “LR”, “PValue” e “FDR”. Como considerações estatísticas utiliza-se o log fold-change (logFC) e a Taxa de falsas descobertas (False Discovery Rate, FDR) para a escolha dos DEGs. O logFC é a razão entre a expressão de um gene e as condições (subtipos) e informa se a expressão do gene aumenta ou diminui sob determinadas condições por meio do sinal (negativo ou positivo); o FDR representa a proporção de falsos positivos entre as descobertas. Para a análise geral, os DEGs foram filtrados para o $|\logFC| > 0$ e uma taxa de descobertas falsas (FDR) $\leq 0,05$ e os resultados para o gene *CFTR* foram isolados para uma investigação sistêmica.

A visualização dos resultados foi realizada por meio de um volcano plot com a biblioteca *EnhancedVolcano* em R, que mostra a distribuição dos genes em função da diferença de expressão entre os grupos (logFC) e de sua significância estatística, geralmente em termos de $-\log_{10}$ do p-valor, de forma que os genes mais significativos ficam posicionados mais altos no gráfico (BLIGHE et al., 2024).

4.3 Análise de Rede de Coexpressão e Identificação de Genes Hub

A análise de coexpressão foi realizada utilizando o CEMiTool (RUSSO et al., 2018), para identificar módulos de coexpressão por meio de uma abordagem não supervisionada. O CEMiTool aplica uma distribuição gama inversa para a filtragem de genes, otimiza a seleção de parâmetros para identificação de módulos, realiza análise de enriquecimento funcional usando o banco de dados de vias Reactome e cria redes de interação (MILACIC et al., 2023).

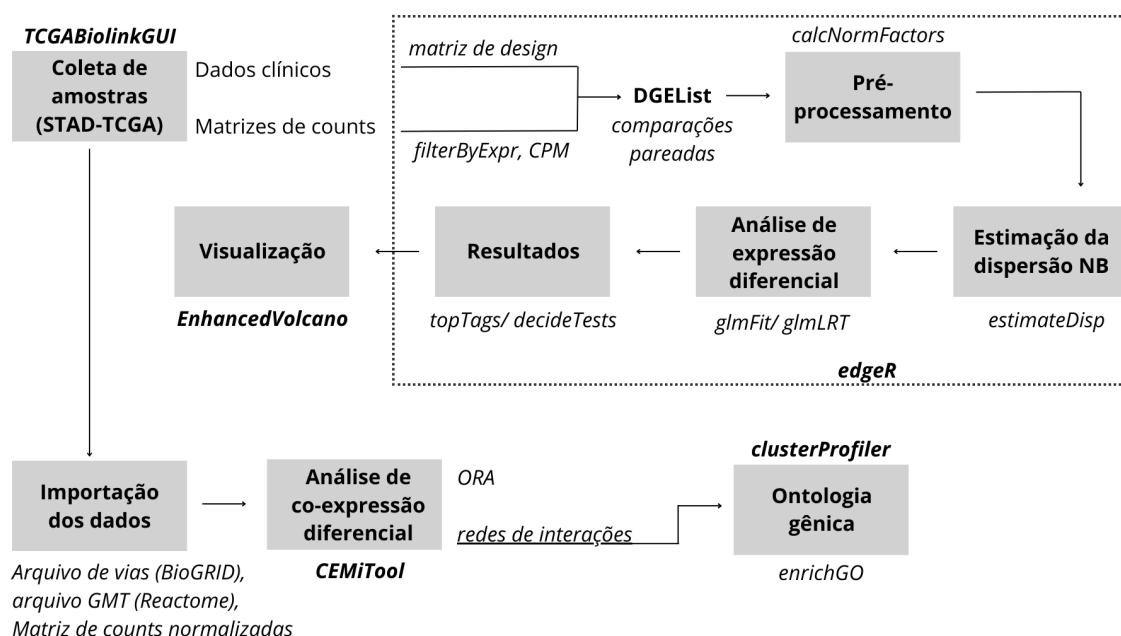
Embora o CEMiTool não tenha sido originalmente projetado para dados de RNA-seq, ele pode processar contagens de leituras utilizando sua transformação interna estabilizadora de variância (VST). Essa ferramenta se baseia no pacote R WGCNA, um método amplamente utilizado para análise de coexpressão gênica (LANGFELDER; HORVATH, 2008). Os procedimentos do WGCNA incorporados ao CEMiTool incluem funções para o agrupamento hierárquico de genes em módulos e uma versão modificada da função de seleção automática de potência de limiar suave. O CEMiTool determina automaticamente o parâmetro β ideal, este é o valor de suavização utilizado para transformar a matriz de correlação dos genes em uma matriz de adjacência. Para dados de RNA-seq, o CEMiTool recomenda o uso da função VST; os dados de entrada foram mantidos em uma forma discreta normalizada. Os resultados incluem um resumo das contagens de genes, análise de super-representação de vias funcionais (ORA), redes de coexpressão gene-gene, redes de interação gene-gene e hub genes. O CEMiTool foi realizado a partir da função de mesmo nome *cemitool* com o coeficiente de correlação de Spearman e VST, utilizando o banco de dados Reactome (versão de 25-09-2024) para análise ORA (FDR $\leq 0,01$), e o banco de dados BioGRID (v.4.4.238) para investigar a sobreposição de links de coexpressão e interações físicas (STARK et al., 2006).

4.4 Análise de Ontologia Gênica

Para capturar informações biológicas, aplicamos o pacote R *clusterProfiler* (YU et al., 2012) nas interações do *CFTR* em SRCC, que automatiza a classificação de termos biológicos e a análise de enriquecimento. O *clusterProfiler* fornece análise de enriquecimento de ontologia gênica dado um vetor de genes (aqueles encontrados na rede de interações) e os anota em processos biológicos, funções moleculares e componentes celulares com a função *enrichGO*, além de realizar um controle rigoroso do (FDR) por meio da estimativa dos q-valores.

4.5 Fluxograma da metodologia

Figura 4.1 – Pipeline para análise de expressão diferencial, coexpressão diferencial e ontologia gênica



Legenda: O experimento começa com a coleta da matrizes de counts e dados clínicos do STAD-TCGA por meio do TCGABiolinkGUI. Inicia-se com a análise de expressão diferencial utilizando o pacote *EdgeR* e a visualização é feita com o pacote *EnhancedVolcano*. A análise de coexpressão diferencial é feita com a importação das matrizes de counts já normalizadas anteriormente pelo *EdgeR*, o arquivo GMT disponibilizado pelo *Reactome* e as vias de interações do *BioGRID* como entrada para o *CEMiTool*. Por fim, com as redes de interações resultantes do processo anterior, é feita a ontologia gênica com o pacote *clusterProfiler*. Os códigos podem ser acessados no repositório: <<https://github.com/camiliane/TCC>>

Fonte: Autoria própria

5 RESULTADOS E DISCUSSÃO

5.1 Heterogeneidade molecular no câncer gástrico

Após o pré-processamento dos dados de expressão do conjunto TCGA-STAD, avaliamos a expressão diferencial de 19.932 transcritos comparando os diferentes tipos de câncer gástrico. Foram encontrados conjuntos e quantidades distintas de genes diferencialmente expressos nos seis subtipos de câncer de estômago do conjunto de dados de RNA-seq avaliado, destacando a alta heterogeneidade molecular dentro do câncer gástrico. Entretanto, relativamente muito baixa entre alguns dos subtipos comparados.

Observamos que a comparação entre o Carcinoma do tipo difuso e adenocarcinoma intestinal obteve a maior quantidade de DEGs, com 5.023 genes significativos, destes 2449 são hipo-expressos e 2754 hiper-expressos. Ao contrário da comparação entre o carcinoma de células em anel de sinete e o adenocarcinoma mucinoso que não apresentou nenhum DEG (Tabela 5.1).

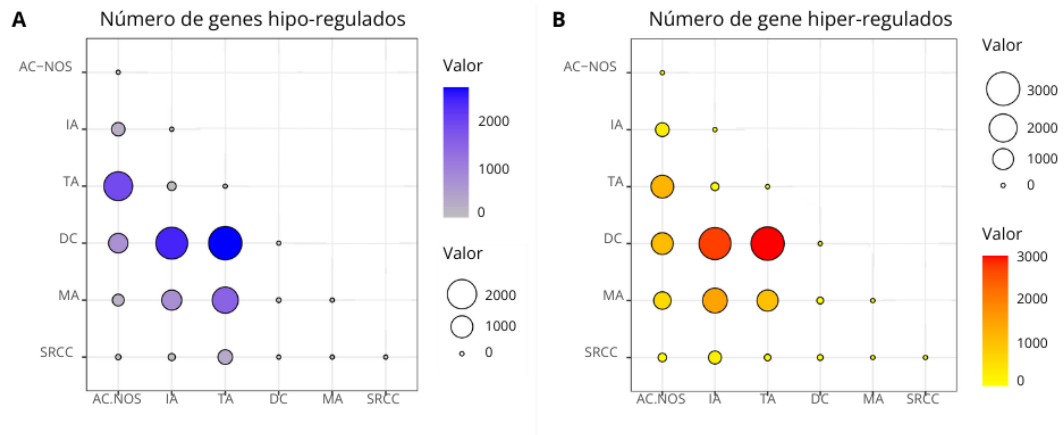
Tabela 5.1 – Contagem de genes hipo e hiper-expressos para cada comparação entre os subtipos de câncer gástrico (TCGA-STAD).

Comparações	hipo-regulado	hiper-regulado
IA x AC-NOS	269	323
IA x MA	803	1503
IA x SRCC	28	274
IA x TA	66	52
AC-NOS x MA	181	621
AC-NOS x SRCC	5	65
AC-NOS x TA	1941	1221
DC x IA	2449	2754
DC x AC-NOS	742	1097
DC x MA	1	23
DC x SRCC	0	9
DC x TA	2715	3040
SRCC x MA	0	0
SRCC x TA	347	20
TA x MA	1535	1035

Legenda: AC-NOS: adenocarcinoma sem outra especificação. IA: adenocarcinoma intestinal. AT: adenocarcinoma tubular. CD: carcinoma tipo difuso. MA: adenocarcinoma mucinoso. SRCC: carcinoma de células em anel de sinete.

O fator de heterogeneidade molecular foi avaliado em função do número de genes e fatores de transcrição regulados (hipo e hiper-expressos) entre os tecidos de câncer gástrico (Figura 5.1).

Figura 5.1 – Distribuição de contagens de genes hipo e hiper-expressos da análise de expressão diferencial pareada entre subtipos de câncer gástrico (TCGA-STAD).

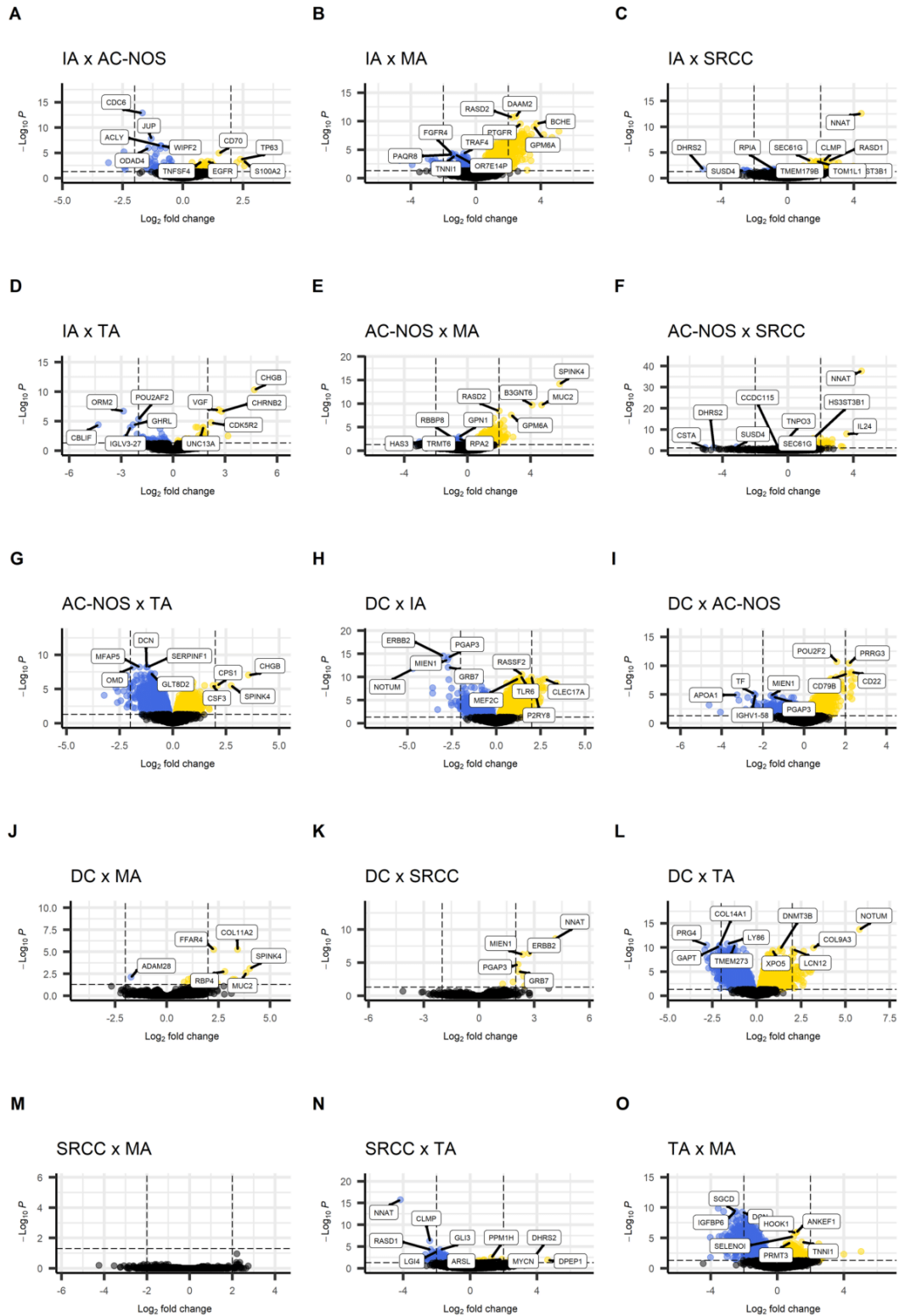


Legenda: AC-NOS: adenocarcinoma não especificado; IA: adenocarcinoma intestinal; TA: adenocarcinoma tubular. DC: carcinoma do tipo difuso; MA: adenocarcinoma mucinoso; SRCC: carcinoma de células em anel de sinete.

No volcano plot, cada ponto representa um gene, distribuído de acordo com sua significância estatística e a magnitude de alteração na expressão entre grupos. Destacamos os Top-5 genes hipo e hiper-expressos (Figura 5.2), que se situam nas extremidades esquerda (em azul) e direita (em amarelo) do gráfico, respectivamente, enquanto aqueles sem significância residem no meio (em preto).

Figura 5.2 – Resultado da expressão diferencial entre os subtipos de câncer gástrico (STAD-TCGA)

LEGENDA: ● HIPO-REGULADOS ● SEM SIGNIFICÂNCIA ● HIPER-REGULADOS



Legenda: Volcano plot destacando os top 5 genes hipo e hiper-expressos. AC-NOS: adenocarcinoma não especificado; IA: adenocarcinoma intestinal; TA: adenocarcinoma tubular. DC: carcinoma do tipo difuso; MA: adenocarcinoma mucinoso; SRCC: carcinoma de células em anel de sinete.

5.2 CFTR não apresenta expressão diferencial nos subtipos de câncer gástrico

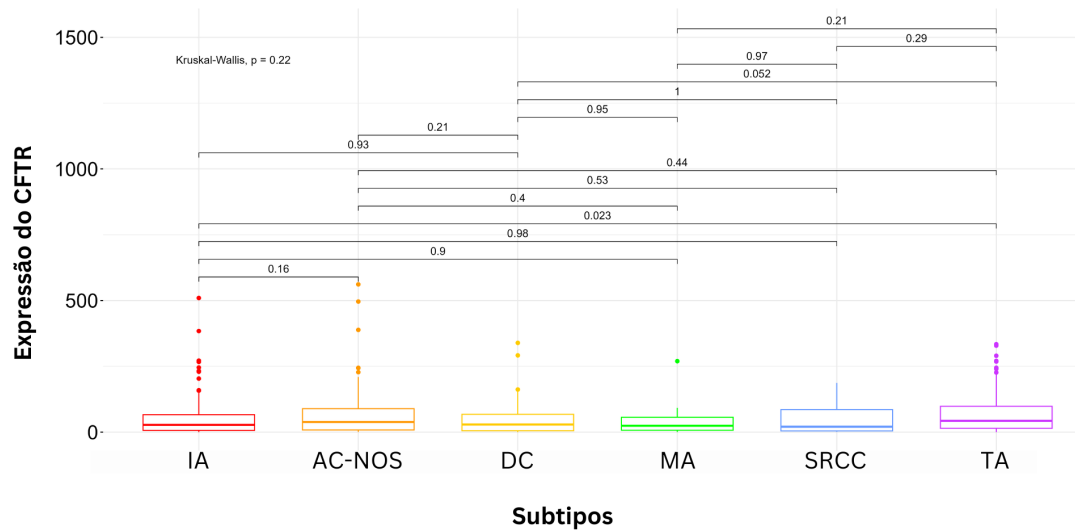
Observamos que o *CFTR* não é diferencialmente expresso entre os tecidos, com base na Tabela 5.2. Portanto, seu papel não é preditivo para a diferenciação entre os mecanismos dos subtipos de câncer com base nos dados do TCGA-STAD. A expressão do *CFTR* é relativamente baixa e consistente nesses tecidos (Figura 5.3), sugerindo que sua atividade não varia significativamente para distinguir um tipo de tecido ou subtipo de câncer gástrico de outro.

Tabela 5.2 – Expressão diferencial do gene *CFTR* para todas as comparações entre os subtipos de câncer gástrico (TCGA-STAD).

Comparações	logFC	FDR	gene
IA x AC-NOS	-0,4093	0,3953	CFTR
IA x MA	-0,6302	0,4483	CFTR
IA x SRCC	-0,8396	0,5367	CFTR
IA x TA	0,1574	0,9106	CFTR
AC-NOS x MA	-0,0471	0,9824	CFTR
AC-NOS x SRCC	-0,5330	0,8669	CFTR
AC-NOS x TA	0,6783	0,0526	CFTR
DC x IA	-0,3698	0,4130	CFTR
DC x AC-NOS	0,1762	0,7515	CFTR
DC x MA	-0,1138	0,9991	CFTR
DC x SRCC	-0,3885	1,0000	CFTR
DC x TA	0,5412	0,1862	CFTR
SRCC x MA	-0,2481	0,9999	CFTR
SRCC x TA	0,6817	0,6980	CFTR
TA x MA	0,7805	0,3536	CFTR

Legenda: os DEGs significativos foram filtrados para o $|logFC| > 0$ e uma taxa de descobertas falsas (FDR) $\leq 0,05$. Nenhuma comparação foi significativa em relação ao gene *CFTR*. AC-NOS: adenocarcinoma não especificado. IA: adenocarcinoma intestinal. TA: adenocarcinoma tubular. DC: carcinoma do tipo difuso. MA: adenocarcinoma mucinoso. SRCC: carcinoma de células em anel de sinete.

Figura 5.3 – Expressão do *CFTR* para cada subtipo de câncer gástrico (TCGA-STAD)



Legenda: AC-NOS: adenocarcinoma não especificado; IA: adenocarcinoma intestinal; TA: adenocarcinoma tubular. DC: carcinoma do tipo difuso; MA: adenocarcinoma mucinoso; SRCC: carcinoma de células em anel de sinete.

5.3 O interatoma do *CFTR* é coexpresso exclusivamente no SRCC

Partindo do princípio de que a expressão singular do *CFTR* não prediz, hipotizamos que o *CFTR* e suas interações podem afetar o desenvolvimento dos subtipos de cancer gastrointestinal. Dessa forma investigamos módulos de genes coexpressos para cada subtipo de câncer de forma individual.

O adenocarcinoma não especificado (NOS) exibiu três módulos, o adenocarcinoma intestinal apresentou um módulo de genes coexpressos, no carcinoma do tipo difuso encontramos dois módulos coexpressos, no adenocarcinoma mucinoso encontramos dois módulos, no carcinoma de células em anel de sinete (SRCC) dois módulos, e no adenocarcinoma tubular um módulo. Cada módulo abrange um conjunto diverso de genes, variando de 35 a 134 genes (Figura 5.4).

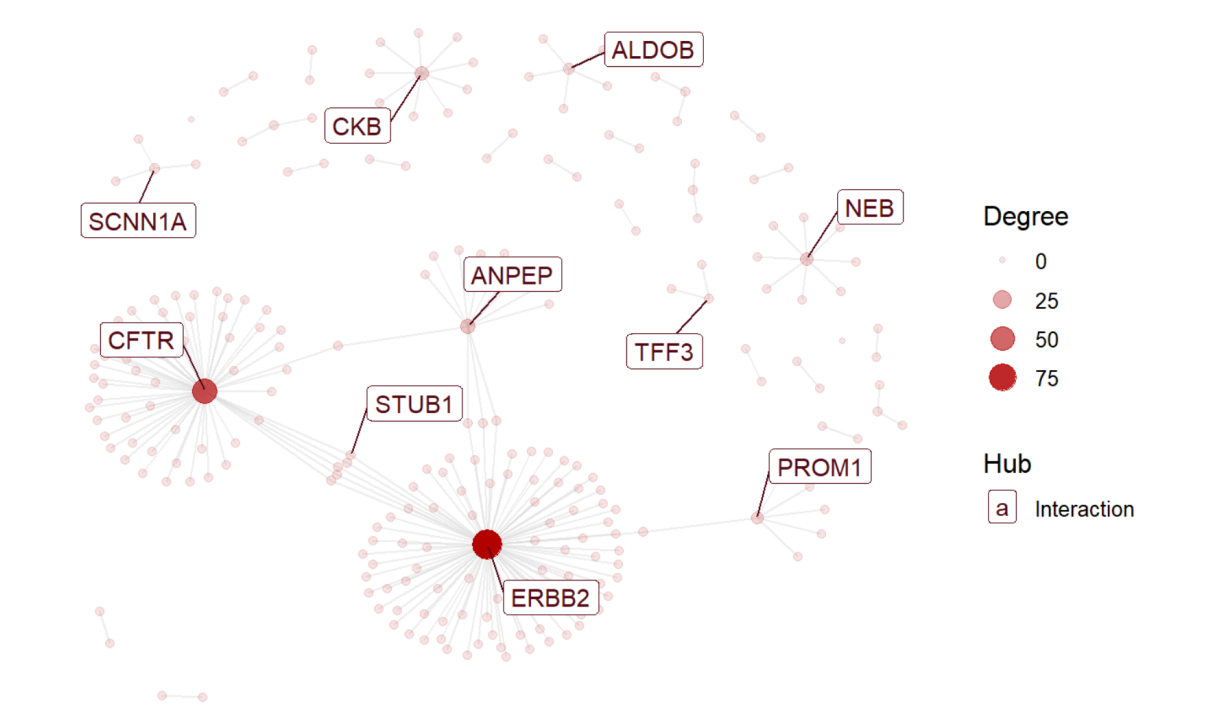
Figura 5.4 – Resultados da detecção de módulos de coexpressão gênica para cada subtipo de câncer gástrico (TCGA-STAD), conforme identificado pelo Cemitoool.

	Adenocarcinoma NOS	Adenocarcinoma Intestinal	Carcinoma tipo Difuso	Adenocarcinoma Mucinoso	Carcinoma de células em anel de sinete	Adenocarcinoma Tubular
MÓDULO 1 (M1)	92	134	101	90	123	126
MÓDULO 2 (M2)	77		63	88	94	
MÓDULO 3 (M3)	35					
NÃO CORRELACIONADO	77	132	92	70	10	107

Legenda: Os círculos cinzas representam os genes não correlacionados em nenhum módulo, os círculos rosa-claro são os módulos em que o CFTR não foi encontrado, em contrapartida o círculo vermelho é o módulo em que o CFTR foi encontrado.

Notavelmente, o *CFTR* foi identificado como coexpresso exclusivamente no Módulo 1 do subtipo SRCC (SRCC-M1), que contém 123 genes (Figura 5.4). O *CFTR* também foi destacado como um dos genes centrais na rede de interação, apoiado por estudos experimentais catalogados para a construção da rede de interação gene-gene, conforme representado na Figura 5.5. O *CFTR* está coexpresso com 58 genes no módulo identificado pelo CEMiTool, e essas interações também foram corroboradas pelos dados do BioGRID, que armazena evidências experimentais.

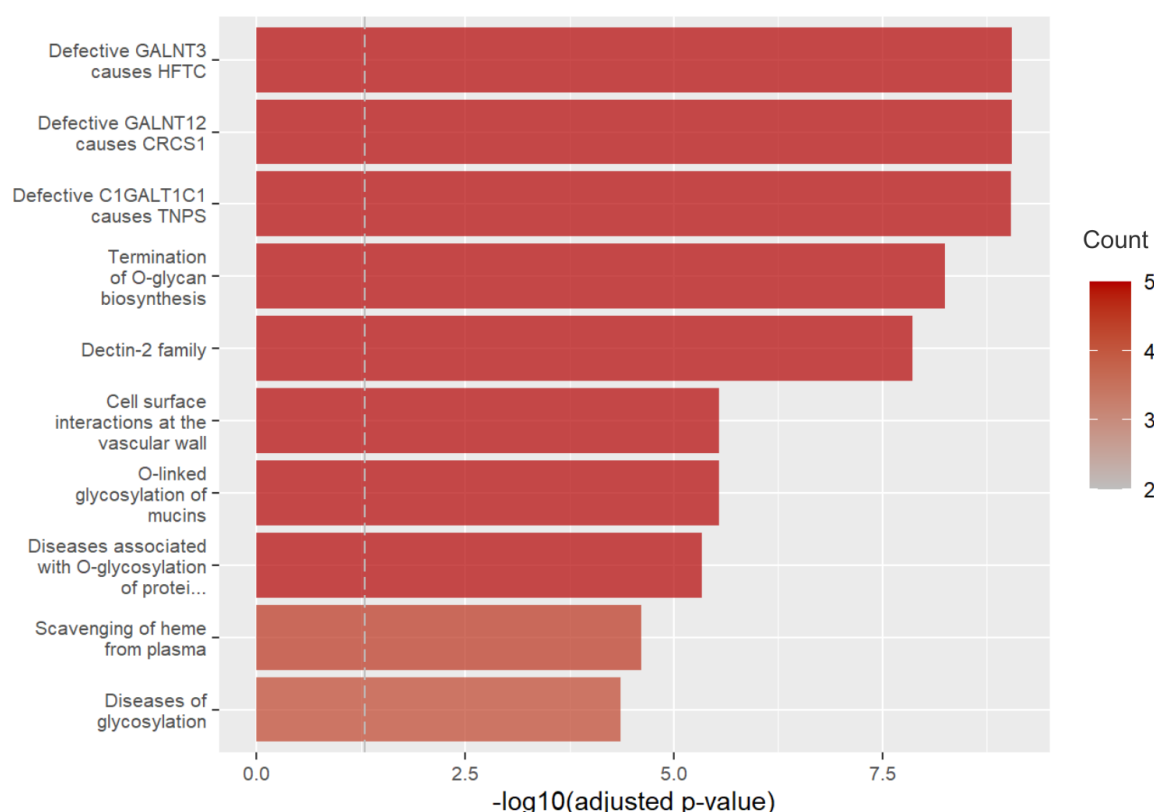
Figura 5.5 – Visualização do Módulo 1 do subtipo de carcinoma de células em anel de sinete (SRCC), destacando CFTR como hub gene.



A análise de super-representação de vias para o módulo SRCC-M1 revelou um enriquecimento significativo em vias associadas a defeitos de glicosilação e modificação

de proteínas, com as principais vias sendo GALNT3 defeituoso, GALNT12 defeituoso e C1GALT1C1 defeituoso. Outras vias notáveis incluem a terminação da biossíntese de O-glicanos, sinalização da família dectina-2, O-glicosilação de mucinas e doenças associadas à glicosilação de proteínas. Isso destaca a importância dos processos de glicosilação e seus defeitos nos mecanismos das doenças. Além disso, vias relacionadas a interações vasculares e captação de heme também foram significativamente enriquecidas (Figura 5.6).

Figura 5.6 – Análise de sobre-representação mostrando as vias biológicas significativas associadas aos genes no módulo 1 do subtipo de carcinoma de células em anel de sinete (SRCC).



Com base na Tabela 5.3 é possível observar os resultados da análise de enriquecimento funcional por Ontologia Gênica para os genes associados ao *CFTR* do módulo 1 do subtipo SRCC, com os 10 primeiros resultados de cada ontologia a partir do número de counts. Os processos biológicos enriquecidos estão fortemente ligados ao controle de conformação de proteínas, incluindo respostas a proteínas mal dobradas e topologicamente incorretas, dobramento de proteínas, *refolding* e montagem de complexos mediados por chaperonas, especialmente sob condições de estresse, como estresse térmico e do retículo endoplasmático. A análise de componentes celulares identificou enriquecimento em estruturas envolvidas no tráfego vesicular, incluindo vesículas revestidas, vesículas revestidas por clatrina, adesões focais e junções célula-substrato, indicando papéis significativos na adesão celular e transporte. Por fim, As funções moleculares enriquecidas neste módulo incluem ligação a chaperonas de dobramento de proteínas, ligação a proteínas mal dobradas, inibição de canais iônicos e transportadores, e ligação a caderinas, sugerindo papéis críticos na homeostase celular e proteostase.

Tabela 5.3 – Análise de enriquecimento funcional por Ontologia Gênica (Go) dos genes coexpressos em conjunto do *CFTR* no módulo 1 do subtipo de carcinoma de células em anel de sinete (SRCC).

Ontologia Gênica	Descrição	p.adjust	Counts
Processo Biológico			
	Resposta a proteína desnaturada	4,63E-12	12
	Resposta a proteína topologicamente incorreta	7,52E-12	12
	Dobramento de proteína	1,39E-08	11
	Redobramento de proteína	4,35E-06	8
	Resposta ao estímulo de temperatura	7,92E-06	8
	Resposta celular a proteína desnaturada	4,98E-05	8
	Resposta celular a proteína topologicamente incorreta	8,23E-05	8
	Resposta ao calor	8,23E-05	8
	Montagem de complexos proteicos mediada por chaperonas	8,23E-05	8
	Resposta ao estresse do retículo endoplasmático	1,22E-04	8
Componente Celular			
	Vesícula revestida	2,52E-05	9
	Adesão focal	4,35E-05	9
	Junção célula-substrato	4,35E-05	9
	Membrana lisossomal	3,29E-04	8
	Membrana do vacúolo lítico	3,29E-04	8
	Parte apical da célula	4,29E-04	8
	Membrana da vesícula revestida	4,35E-05	7
	Vesícula revestida por clatrina	4,35E-05	7
	Membrana plasmática apical	8,88E-04	7
	Endossomo precoce	1,26E-03	7
Função Molecular			
	Ligação ao chaperona de dobramento de proteína	1,07E-06	8
	Ligação à ligase de proteína ubiquitina	1,29E-04	8
	Ligação à ligase de proteína similar a ubiquitina	1,29E-04	8
	Ligação a cadherina	1,29E-04	8
	Ligação a proteína desnaturada	4,90E-06	7
	Atividade de hidrólise de ATP	3,13E-03	7
	Chaperona de dobramento de proteína	4,90E-06	6
	Ligação a proteína de choque térmico	6,90E-04	5
	Ligação a actina	3,59E-02	5
	Atividade inibidora de canal iônico	1,29E-04	4

5.4 Discussão

Em nosso estudo, analisamos os perfis diferenciais e de coexpressão de *CFTR* em seis subtipos de câncer gástrico e identificamos uma rede genética única associada ao *CFTR* em amostras de SRCC. Essa rede inclui 58 genes coexpressos com pelo menos duas interações experimentalmente validadas, revelando uma super-representação de processos de glicosilação.

Pesquisas recentes destacaram o gene *CFTR*, associado à fibrose cística, pelo seu papel potencial em diversos tipos de câncer (BHATTACHARYA et al., 2022; HOU et al., 2016; LIU et al., 2014; PARISI et al., 2023). Estudos sugerem que o *CFTR* atua não apenas como um canal iônico, mas também como um supressor tumoral. Sua expressão tem sido associada ao desenvolvimento de várias malignidades, incluindo cânceres gastrointestinais, pancreáticos e colorretais. Variantes genéticas específicas do *CFTR* foram associadas a um risco aumentado de câncer (LIU et al., 2014). Liu et al. relatam que pacientes com fibrose cística, especialmente aqueles que possuem a mutação $\Delta F508$, apresentam um risco elevado de desenvolver cânceres do trato digestivo, incluindo o câncer pancreático (LIU et al., 2014). Isso sugere que o papel do *CFTR* se estende além de sua função tradicional de transporte iônico, influenciando potencialmente a tumorigênese através da predisposição genética (LIU et al., 2014).

A análise de componentes celulares revelou um enriquecimento significativo de estruturas envolvidas no tráfego vesicular, incluindo vesículas revestidas, vesículas revestidas de clatrina, adesões focais e junções célula-substrato. Essas estruturas desempenham papéis essenciais nos processos de adesão e transporte celular. Esse achado é consistente com a análise de *single-cell* de Zhao et al., que mostrou que células de SRCC exibem uma adesão celular diminuída, o que pode facilitar a metástase (ZHAO et al., 2023). Além disso, similar aos nossos resultados, o *CFTR* foi encontrado em uma via de componente celular associada à membrana plasmática apical no estudo de Pang et al (PANG; YAN; SU, 2021). A baixa expressão do *CFTR* e outros genes foram relacionados ao baixo prognóstico em pacientes com câncer gástrico (PANG; YAN; SU, 2021).

Os níveis de expressão de *CFTR* podem variar significativamente em diferentes tipos de câncer. Por exemplo, enquanto o *CFTR* é frequentemente regulado negativamente em cânceres como o de mama e o colorretal, ele é regulado positivamente no câncer gástrico (SHIN et al., 2020; HOU et al., 2016). Essa expressão diferencial pode contribuir para o papel complexo do *CFTR* na progressão do câncer. Observa-se que modificações epigenéticas, como a hipermetilação do promotor do *CFTR*, podem levar à diminuição da expressão de *CFTR*, o que frequentemente está correlacionado com prognósticos desfavoráveis em vários tipos de câncer, incluindo o câncer gástrico (MAKAROVA et al., 2023).

A participação do *CFTR* no câncer gástrico é ainda respaldada por seu papel regulador em vias inflamatórias. Foi demonstrado que o *CFTR* suprime a sinalização inflamatória mediada pelo NF- κ B, frequentemente implicada na progressão do câncer (HUNTER et al., 2010). A inflamação crônica é um fator de risco bem estabelecido para o câncer gástrico, e a desregulação do *CFTR* pode contribuir para o microambiente inflamatório que favorece o desenvolvimento tumoral (HOU et al., 2016). O estudo de Wen et al. demonstrou que a infecção por *Helicobacter pylori* reduz a expressão e a atividade funcional do *CFTR* na mucosa duodenal pela via de sinalização do TGF β , contribuindo para o desenvolvimento de úlceras duodenais (WEN et al., 2018). As inflamações crônicas associadas a essas úlceras são um fator de risco para alterações no tecido gástrico, o que pode aumentar a probabilidade de desenvolvimento de câncer (SOGAARD et al., 2016).

Os mecanismos regulatórios que governam sua expressão ilustram ainda mais a complexidade da participação do *CFTR* no câncer. Collobert et al. enfatizam a importância de compreender os elementos cis-reguladores que controlam a expressão do gene *CFTR*, pois esse conhecimento pode levar a novas estratégias terapêuticas destinadas a modular a atividade do *CFTR* em pacientes com câncer (COLLOBERT et al., 2021). A interação entre a expressão do *CFTR*, variantes genéticas e modificações epigenéticas apresenta um sistema complexo que merece maior investigação. O papel do *CFTR* na metástase do câncer foi documentado por Zhang et al., que demonstraram que o *CFTR* serve como alvo direto do miR-125b, um microRNA que promove a migração e invasão de células de câncer colorretal ao regular negativamente a expressão de *CFTR* (ZHANG et al., 2021). Isso sugere que o *CFTR* está envolvido tanto na iniciação tumoral quanto na progressão metastática das células cancerígenas. A função supressora tumoral do *CFTR* na metástase também foi observada em outros tipos de câncer, incluindo o câncer de pulmão não pequenas células e o carcinoma nasofaríngeo, destacando seu papel protetor mais amplo em várias malignidades (ZHANG et al., 2021). No estudo de Pang et al., o *CFTR* é destacado como hub gene na investigação de potenciais biomarcadores associados a metaplasia gastrointestinal que progride para o câncer gástrico (PANG; YAN; SU, 2021).

6 CONCLUSÃO

Em conclusão, nosso estudo investigou a expressão diferencial, coexpressão e a ontologia gênica do gene *CFTR* em seis subtipos de câncer de estômago para revelar interações gene-gene e vias associadas ao câncer gástrico. Ao analisar módulos genéticos distintos nos quais o *CFTR* atua como um interator, demonstramos que o *CFTR* pode desempenhar um papel fundamental na biologia tumoral ao influenciar a atividade de genes coexpressos e suas vias associadas no câncer gástrico de células em anel de sinete. Esses achados sugerem que o *CFTR* contribui para a progressão do câncer por meio de interações gene-gene complexas, destacando seu potencial como um modulador-chave da dinâmica de vias no câncer gástrico de células em anel de sinete.

6.1 Trabalhos Futuros

- Investigar redes de coexpressão entre o subtipo Carcinoma de células em anel de sinete e controle;
- Validar os achados em amostras independentes;
- Investigar processos celulares pro câncer gástrico via experimentos de Single Cell.

REFERÊNCIAS

- ANDERSON, K. J.; CORMIER, R. T.; SCOTT, P. M. Role of ion channels in gastrointestinal cancer. **World Journal of Gastroenterol.**, 2019. 13, 15
- ASHBURNER, M.; BALL, C. A.; BLAKE, J. A.; BOTSTEIN, D.; BUTLER, H.; CHERRY, J. M.; DAVIS, A. P.; DOLINSKI, K.; DWIGHT, S. S.; EPPIG, J. T.; HARRIS, M. A.; HILL, D. P.; ISSEL-TARVER, L.; KASARSKIS, A.; LEWIS, S.; MATESE, J. C.; RICHARDSON, J. E.; RINGWALD, M.; SHERLOCK, G. M. R. . G. Gene Ontology: tool for the unification of biology. **Nature Genetics**, 2000. 20
- BAREIL, C.; BERGOUGNOUX, A. Cftr gene variants, epidemiology and molecular pathology. **Archives de Pédiatrie**, 2020. 12, 14
- BHATTACHARYA, R.; BLANKENHEIM, Z.; SCOTT, P.; CORMIER, R. Cftr and gastrointestinal cancers: an update. **Journal of Personalized Medicine**, v. 12, p. 868, 2022. 12, 22, 36
- BLIGHE, K.; RANA, S.; TURKES, E.; OSTENDORF, B.; GRIONI, A.; LEWIS, M. **EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling**. 2024. <<https://github.com/kevinblighe/EnhancedVolcano>>. 25
- BOECK, K. D.; AMARAL, M. D. Progress in therapies for cystic fibrosis. **The Lancet Respiratory Medicine**, 2016. 14
- BRAY, F.; LAVERSANNE, M.; SUNG, H.; FERLAY, J.; SIEGEL, R. L.; SOERJOMATARAM, I.; JEMAL, A. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: a cancer journal for clinicians**, Wiley Online Library, v. 74, n. 3, p. 229–263, 2024. 22
- CAROLYN, C. H.; ZENKLUSEN, J. C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. **Cell**, 2018. 16, 17
- CHEN, Y.; CHEN, L.; LUN, A. T. L.; BALDONI, P. L.; SMYTH, G. K. edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Cold Spring Harbor Laboratory, 2024. 19, 25
- CHEN, Y.; LUN, A. T. L.; SMYTH, G. K. Differential expression analysis of complex rna-seq experiments using edgeR. **Frontiers in Probability and the Statistical Sciences**, 2018. 18, 19
- CHOWDHURY, H. A.; BHATTACHARYYA, D. K.; KALITA, J. K. (differential) co-expression analysis of gene expression: A survey of best practices. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, 2020. 19
- COLAPRICO, A.; SILVA, T. C.; OLSEN, C.; GAROFANO, L.; CAVA, C.; GAROLINI, D.; SABEDOT, T. S.; MALTA, T. M.; PAGNOTTA, S. M.; CASTIGLIONI, I.; CECCARELLI, M.; BONTEMPI, G.; NOUSHMEHR, H. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. **Nucleic Acids Research**, 2015. 24

COLLOBERT, M.; BOCHER, O.; NABEC, A. L.; G  NIN, E.; F  REC, C.; MOISAN, S. Cftr cooperative cis-regulatory elements in intestinal cells. **International Journal of Molecular Sciences**, 2021. 37

CONSORTIUM, T. G. O. The Gene Ontology Resource: 20 years and still GOing strong. **Nucleic Acids Research**, 2018. 20

CONSORTIUM, T. G. O.; ALEKSANDER, S. A.; BALHOFF, J.; CARBON, S.; CHERRY, J. M.; DRABKIN, H. J.; EBERT, D.; FEUERMAN, M.; GAUDET, P.; HARRIS, N. L.; HILL, D. P.; LEE, R.; MI, H.; MOXON, S.; MUNGALL, C. J.; MURUGANUGAN, A.; MUSHAYAHAMA, T.; STERNBERG, P. W.; THOMAS, P. D.; AUKEN, K. V.; RAMSEY, J.; SIEGELE, D. A.; CHISHOLM, R. L.; FEY, P.; ASPROMONTE, M. C.; NUGNES, M. V.; QUAGLIA, F.; TOSATTO, S.; GIGLIO, M.; NADENDLA, S.; ANTONAZZO, G.; ATTRILL, H.; SANTOS, G. dos; MARYGOLD, S.; STRELETS, V.; TABONE, C. J.; THURMOND, J.; ZHOU, P.; AHMED, S. H.; ASANITTHONG, P.; BUITRAGO, D. L.; ERDOL, M. N.; GAGE, M. C.; KADHUM, M. A.; LI, K. Y. C.; LONG, M.; MICHALAK, A.; PESALA, A.; PRITAZAHRA, A.; SAVERIMUTTU, S. C. C.; SU, R.; THURLOW, K. E.; LOVERING, R. C.; LOGIE, C.; OLIFERENKO, S.; BLAKE, J.; CHRISTIE, K.; CORBANI, L.; DOLAN, M. E.; DRABKIN, H. J.; HILL, D. P.; NI, L.; SITNIKOV, D.; SMITH, C.; CUZICK, A.; SEAGER, J.; COOPER, L.; ELSE, J.; JAISWAL, P.; GUPTA, P.; JAISWAL, P.; NAITHANI, S.; LERA-RAMIREZ, M.; RUTHERFORD, K.; WOOD, V.; PONS, J. L. D.; DWINELL, M. R.; HAYMAN, G. T.; KALDUNSKI, M. L.; KWITEK, A. E.; LAULEDERKIND, S. J. F.; TUTAJ, M. A.; VEDI, M.; WANG, S.-J.; D'EUSTACHIO, P.; AIMO, L.; AXELSEN, K.; BRIDGE, A.; HYKANOUSPIKEL, N.; MORGAT, A.; ALEKSANDER, S. A.; CHERRY, J. M.; ENGEL, S. R.; KARRA, K.; MIYASATO, S. R.; NASH, R. S.; SKRZYPEK, M. S.; WENG, S.; WONG, E. D.; BAKKER, E.; BERARDINI, T. Z.; REISER, L.; AUCHINCLOSS, A.; AXELSEN, K.; ARGOUDE-PUY, G.; BLATTER, M.-C.; BOUTET, E.; BREUZA, L.; BRIDGE, A.; CASALS-CASAS, C.; COUDERT, E.; ESTREICHER, A.; FAMIGLIETTI, M. L.; FEUERMAN, M.; GOS, A.; GRUAZ-GUMOWSKI, N.; HULO, C.; HYKANOUSPIKEL, N.; JUNGO, F.; MERCIER, P. L.; LIEBERHERR, D.; MASSON, P.; MORGAT, A.; PEDRUZZI, I.; POURCEL, L.; POUX, S.; RIVOIRE, C.; SUNDARAM, S.; BATEMAN, A.; BOWLER-BARNETT, E.; BYE-A-JEE, H.; DENNY, P.; IGNATCHENKO, A.; ISHTIAQ, R.; LOCK, A.; LUSSI, Y.; MAGRANE, M.; MARTIN, M. J.; ORCHARD, S.; RAPOSO, P.; SPERETTA, E.; TYAGI, N.; WARNER, K.; ZARU, R.; DIEHL, A. D.; LEE, R.; CHAN, J.; DIAMANTAKIS, S.; RACITI, D.; ZAROWIECKI, M.; FISHER, M.; JAMES-ZORN, C.; PONFERRADA, V.; ZORN, A.; RAMACHANDRAN, S.; RUZICKA, L.; WESTERFIELD, M. The Gene Ontology knowledgebase in 2023. **Genetics**, 2023. 20

DAM, S. van; V  SA, U.; GRAAF, A. van der; FRANKE, L.; MAGALH  ES, J. P. de. Gene co-expression analysis for functional classification and gene-disease predictions. **Briefings in Bioinformatics**, 2017. 19

ERGIN, S.; KHERAD, N.; ALAGOZ, M. Rna sequencing and its applications in cancer and rare diseases. **Mol Biol Rep**, 2022. 18

HANSSSENS, L. S.; DUCHATEAU, J.; CASIMIR, G. J. Cftr protein: Not just a chloride channel? **Cells**, 2021. 12, 13

- HOU, Y.; GUAN, X.; YANG, Z.; LI, C. Emerging role of cystic fibrosis transmembrane conductance regulator - an epithelial chloride channel in gastrointestinal cancers. **World Journal of Gastrointestinal Oncology**, 2016. 12, 15, 22, 36
- HUNTER, M. J.; TREHARNE, K. J.; WINTER, A. K.; CASSIDY, D.; LAND, S. C.; MEHTA, A. Expression of wild-type cftr suppresses nf-b-driven inflammatory signalling. **Plos One**, 2010. 15, 36
- LANG, J.; CHO, W. C.; HUANG, T.; XU, T. W. and Junlin. Applications of rna-seq in cancer and tumor research. **Frontiers in Genetics**, 2023. 18
- LANGFELDER, P.; HORVATH, S. Wgcna: an r package for weighted correlation network analysis. **BMC bioinformatics**, BioMed Central, v. 9, n. 1, p. 1–13, 2008. 26
- LIU, H.; WU, W.; LIU, Y.; ZHANG, C.; ZHOU, Z. Predictive value of cystic fibrosis transmembrane conductance regulator (cftr) in the diagnosis of gastric cancer. **Clinical & Investigative Medicine**, 2014. 12, 15, 22, 36
- LOWE, R.; SHIRLEY, N.; BLEACKLEY, M.; DOLAN, S.; SHAFEE, T. Transcriptomics technologies. **PLOS Computational Biology**, 2017. 17, 18
- LUKASIAK, A.; ZAJAC, . M. The distribution and role of the cftr protein in the intracellular compartments. **Membranes**, 2021. 12, 13
- MAKAROVA, M. V.; NEMTSOVA, M. V.; DANISHEVICH, A.; CHERNEVSKIY, D.; BELENIKIN, M. S.; KRINITINA, A.; , ; SAGAYDAK, O. V.; VORONTSOVA, M. S.; KHATKOV, I.; , ; BODUNOVA, N.; NIKOLAEV, S.; BYAKHOVA, M.; SEMENOVA, A. A.; , ; GADZHIEVA, S. The cftr gene germline heterozygous pathogenic variants in russian patients with malignant neoplasms and healthy carriers: 11,800 wgs results. **International Journal of Molecular Sciences**, 2023. 12, 36
- MILACIC, M.; BEAVERS, D.; CONLEY, P.; GONG, C.; GILLESPIE, M.; GRISS, J.; HAW, R.; JASSAL, B.; MATTHEWS, L.; MAY, B.; PETRYSZAK, R.; RAGUENEAU, E.; ROTHFELS, K.; SEVILLA, C.; SHAMOVSKY, V.; STEPHAN, R.; TIWARI, K.; VARUSAI, T.; WEISER, J.; WRIGHT, A.; WU, G.; STEIN, L.; HERMJAKOB, H.; D'EUSTACHIO, P. The Reactome Pathway Knowledgebase 2024. **Nucleic Acids Research**, 2023. 16, 17, 25
- NETWORK, T. C. G. A. R. Comprehensive molecular characterization of gastric adenocarcinoma. **Nature**, 2014. 24
- PANG, L.; YAN, X.; SU, D. Screening key genes and biomarkers in gastrointestinal metaplasia that progress to gastric cancer by integrated bioinformatics analysis. **Journal of physiology and pharmacology**, 2021. 36, 37
- PARISI, G. F.; PAPALE, M.; PECORA, G.; ROTOLO, N.; MANTI, S.; RUSSO, G.; LEONARDI, S. Cystic fibrosis and cancer: Unraveling the complex role of cftr gene in cancer susceptibility. **Cancers**, 2023. 12, 22, 36
- PUCCINI, A.; POORMAN, K.; CATALANO, F.; SEEGER, A.; GOLDBERG, R. M.; SALEM, M. E.; SHIELDS, A. F.; BERGER, M. D.; BATTAGLIN, F.; TOKUNAGA, R. et al. Molecular profiling of signet-ring-cell carcinoma (srcc) from the stomach and colon reveals potential new therapeutic targets. **Oncogene**, Nature Publishing Group UK London, v. 41, n. 26, p. 3455–3460, 2022. 22

- RUSSO, P. S.; FERREIRA, G. R.; CARDOZO, L. E.; BÜRGER, M. C.; ARIAS-CARRASCO, R.; MARUYAMA, S. R.; HIRATA, T. D.; LIMA, D. S.; PASSOS, F. M.; FUKUTANI, K. F. et al. Cemitool: a bioconductor package for performing comprehensive modular co-expression analyses. **Bmc Bioinformatics**, BioMed Central, v. 19, n. 1, p. 1–13, 2018. 25
- SERIN, E. A. R.; NIJVEEN, H.; HILHORST, H. W. M.; LIGTERINK, W. Learning from co-expression networks: Possibilities and challenges. **Frontiers in Plant Science**, 2016. 19, 20
- SHIN, Y.; KIM, M.; WON, J.; KIM, J. C.; OH, S.; LEE, J. H.; PARK, K. Epigenetic modification of cftr in head and neck cancer. **Journal of Clinical Medicine**, 2020. 36
- SLAE, M.; WILSCHANSKI, M. Cystic fibrosis: a gastrointestinal cancer syndrome. **The Lancet Oncology**, 2018. 12
- SLAE, M.; WILSCHANSKI, M. Cystic fibrosis and the gut. **Frontline Gastroenterology**, 2021. 14, 15
- SOGAARD, K. K.; FARKAS, D. K.; PEDERSEN, L.; LUND, J. L.; THOMSEN, R. W.; SØRENSEN, H. T. Long-term risk of gastrointestinal cancers in persons with gastric or duodenal ulcers. **cancer Medicine**, 2016. 15, 36
- STARK, C.; BREITKREUTZ, B.-J.; REGULY, T.; BOUCHER, L.; BREITKREUTZ, A.; TYERS, M. BioGRID: a general repository for interaction datasets. **Nucleic Acids Research**, 2006. 16, 17, 26
- SUGUNARAJ, J. P.; MIRSHAHI, U.; WARDEH, A.; MANNEY, C.; MANICKAM, K.; MURRAY, M.; CAREY, D.; STAMM, J. Cancer risks in heterozygous cystic fibrosis transmembrane conductance regulator (cftr) delf508 carriers. **Chest**, 2016. 15
- THAN, B. L. N.; LINNEKAMP, J. F.; STARR, T. K.; LARGAESPADA, D. A.; ROD, A.; ZHANG, Y.; BRUNER, V.; ABRAHANTE, J.; SCHUMANN, A.; LUCZAK, T.; NIEMCZYK, A.; O'SULLIVAN, M. G.; MEDEMA, J. P.; FIJNEMAN, R. J. A.; MEIJER, G. A.; BROEK, E. Van den; HODGES, C. A.; SCOTT, P. M.; VERMEULEN, L.; CORMIER, R. T. Cftr is a tumor suppressor gene in murine and human intestinal cancer. **Oncogene**, 2016. 15
- VOIT, E. O. Perspective: Systems biology beyond biology. **Frontiers in Systems Biology**, 2022. 16
- WEN, G.; DENG, S.; SONG, W.; JIN, H.; XU, J.; LIU, X.; XIE, R.; SONG, P.; TUO, B. Helicobacter pylori infection downregulates duodenal cftr and slc26a6 expressions through tgf signaling pathway. **BMC Microbiology**, 2018. 16, 36
- WEN, G.; JIN, H.; DENG, S.; XU, J.; LIU, X.; XIE, R.; TUO, B. Seffects of helicobacter pylori infection on the expressions and functional activities of human duodenal mucosal bicarbonate transport proteins. **Helicobacter**, 2016. 16
- YU, G.; WANG, L.-G.; HAN, Y.; HE, Q.-Y. clusterprofiler: an r package for comparing biological themes among gene clusters. **OMICS: A Journal of Integrative Biology**, 2012. 26

ZHANG, X.; LI, T.; HAN, Y.; GE, M.; WANG, P.; SUN, L.; LIU, H.; CAO, T.; NIE, Y.; FAN, D.; GUO, H.; WU, K.; ZHAO, X.; LU, Y. mir-125b promotes colorectal cancer migration and invasion by dual-targeting cftr and cgn. **Cancers**, 2021. 37

ZHAO, W.; JIA, Y.; SUN, G.; YANG, H.; LIU, L.; QU, X.; DING, J.; YU, H.; XU, B.; ZHAO, S. et al. Single-cell analysis of gastric signet ring cell carcinoma reveals cytological and immune microenvironment features. **Nature Communications**, Nature Publishing Group UK London, v. 14, n. 1, p. 2985, 2023. 22, 36