

UNIVERSIDADE ESTADUAL PAULISTA
“Júlio de Mesquita Filho”
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

Determinação de Biomarcadores para tratamento personalizado a
partir do Interactoma do Câncer Colorretal.

GABRIEL HENRIQUE CAXALI

Dissertação apresentada ao Instituto de
Biociências de Botucatu (IBB), para a
obtenção do título de Mestre no Programa de
Pós Graduação em Biologia Geral e Aplicada.

Orientadora: Prof^a Dr^a Flávia Karina Delella

*Coorientador: Prof. Dr. Robson Francisco
Carvalho*

Botucatu - SP

2023

UNIVERSIDADE ESTADUAL PAULISTA

“Júlio de Mesquita Filho”

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

Determinação de Biomarcadores para tratamento personalizado a
partir do Interactoma do Câncer Colorretal.

GABRIEL HENRIQUE CAXALI

FLÁVIA KARINA DELELLA

Dissertação apresentada ao Instituto de
Biociências de Botucatu (IBB), para a
obtenção do título de Mestre no Programa
de Pós Graduação em Biologia Geral e
Aplicada.

Orientadora: Profª Drª Flávia Karina Delella

*Coorientador: Prof. Dr. Robson Francisco
Carvalho*

Botucatu - SP

2023

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: MARIA CAROLINA ANDRADE CRUZ E SANTOS-CRB

Caxali, Gabriel Henrique.

Determinação de biomarcadores para tratamento
personalizado a partir do interactoma do câncer colorretal
/ Gabriel Henrique Caxali. - Botucatu, 2023

Dissertação (mestrado) - Universidade Estadual Paulista
"Júlio de Mesquita Filho", Instituto de Biociências de
Botucatu

Orientador: Flávia Karina Dellela

Coorientador: Robson Francisco Carvalho

Capes: 20601000

1. Bioinformática. 2. Cólon (Anatomia) - Câncer. 3.
Aspectos moleculares. 4. Terapia. 5. Câncer - Tratamento.

Palavras-chave: Bioinformática; Câncer colorretal;
Interactoma; Terapia personalizada.

Dedicatória

**Para Ada Donini Caxali,
Que sua presença se torne eterna no trabalho que segue.**

Agradecimentos

Primeiramente à minha família, meus pais *Elenice e Nivaldo*, que acreditaram em mim, mesmo quando eu deixei de acreditar, e confiaram sempre em meus estudos, me dando apoio nos mais diferentes aspectos. Serei sempre eternamente grato por tudo o que fizeram e fazem por mim, sem vocês isso não seria possível. Também gostaria de agradecer a minha tia/madrinha/segunda mãe *Elza*, que sempre esteve presente durante toda minha aventura nesse meio acadêmico, e sempre se fez presente. E a minha irmã *Giovana*, por tornar muitos momentos leves, com nossas brincadeiras, bobearias e risadas.

A minha namorada *Beatriz*, que assumiu em muitos dias o papel de confidente, mas sempre se fez presente e sempre foi um dos maiores motivos da minha alegria. Obrigado por deixar minha vida mais leve, por me apoiar e ser meu porto seguro. Obrigado por compartilhar a vida comigo, nem eu acredito a sorte que tenho em te ter ao meu lado.

Às minhas companheiras de laboratório. *Mirian*, muito obrigado por percorrer essa trajetória comigo, e me fazer crescer muito nesses 2 anos. Obrigado por cada risada, por cada conselho, por me ensinar tanto, principalmente sobre cultura celular. Se eu sei fazer algo na sala de cultura, devo a ti. Você é um espelho para mim. Já *Laíza*, o que tenho a dizer sobre meu Patrick? Obrigado por cada meme, cada risada, cada jogo, por compartilhar o mesmo neurônio que eu. Ser seu coorientador me fez crescer muito como cientista e vou levar todos esses ensinamentos para toda a vida. Obrigado por ser minha amiga e me ouvir quando precisei, me fazer rir e estar presente.

Ao prof. *Robson* e suas orientadas *Jakeline, Vitória e Sarah*, obrigado pelo auxílio nas análises e interpretação de dados, bem como por ceder o laboratório para rodar as análises de single cell.

A minha orientadora, *Flávia*, por acreditar em mim em todos os momentos possíveis desde 2016. Serei sempre eternamente grato por todos os conselhos e pela confiança. Obrigado pela oportunidade de crescer e estar mais próximo do meu sonho cada dia mais. Nunca esquecerei do que fez por mim.

Por fim ao Programa de Pós-graduação em Ciências Biológicas (Biologia Geral e Aplicada) e ao CNPQ pela infraestrutura e bolsa concedidos para a realização deste trabalho.

“Ninguém ignora tudo. Ninguém sabe tudo. Todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa. Por isso aprendemos sempre”

Paulo Freire

Lista de Ilustrações

Figura 1. Diferentes estágios do desenvolvimento do CCR, adaptada de SIMON, 2016.....	13
Figura 1(Artigo). Volcano Plots dos dados de expressão de COAD (A) e de READ (B).....	30
Figura 2. Diagrama de Venn dos genes up regulados detectados nas amostras tumorais de cólon e reto do TCGA	31
Figura 3. Caracterização dos 30 pares detectados no secretoma das amostras de tumor Colorretal.....	32
Figura 4. Enriquecimento funcional do interactoma gerado, contendo os ligantes e receptores identificados.....	33
Figura 5: <i>UMAP plot</i> , representando a distribuição dos genes identificados como potenciais biomarcadores nos diferentes tipos celulares identificados nos dados de scRNA-Seq.....	35
Figura 6. Interação detectada entre genes e drogas através da assinatura genética de COAD e READ.....	37

Lista de Tabelas

Tabela 1. Eficácia das linhas farmacológicas no decorrer do tratamento do câncer colorretal. Adaptada de (MODEST; PANT; SARTORE-BIANCHI, 2019)	15
Tabela 2. Marcadores Canônicos utilizados na identificação dos diferentes tipos celulares presentes nos dados de sc-RNA-Seq.....	28

Lista de Abreviações

5 FU	5- Flououracil
COAD	<i>Colon Adenocarcinoma</i>
CPE	<i>Consensus measurement of Purity Estimations</i> (Consenso de Mensuração da pureza estimada)
CRC	Colorectal Cancer
DEGs	<i>Differentially Expressed Genes</i> (Genes Diferencialmente Expressos)
DGIdb	<i>Drug Gene Interaction Database</i>
EGFR	<i>Epidermal growth factor receptor</i> (Receptor do Fator de Crescimento Epidérmico)
FC	<i>Fold Change</i>
FDR	<i>False Discovery Rate</i>
GTEX	<i>Genotype Tissue Expression Database</i>
INCA	Instituto Nacional do Câncer, José de Alencar Gomes da Silva
NGS	<i>Next Generation Sequencing</i> (Sequenciamento de Nova Geração)
OCTAD	<i>Open Cancer Therapeutic Discovery</i>
PCA	<i>Principal Component Analysis</i> (Análise de Componente Principal)
PPI	<i>Protein Protein Interaction</i> (Interação Proteína Proteína)
READ	<i>Rectum Adenocarcinoma</i>
ScRNA-SEQ	<i>Single Cell RNA-SEQ</i> (Sequenciamento de RNA de célula única)
STRING	<i>Search Tool for the Retrieval of Interacting Genes/Proteins</i>
TCGA	<i>The Cancer Genome Atlas Database</i>
TMN	Classificação de malignidade tumoral, baseada em 3 critérios: Borda do Tumor(T), Migração para LinfoNodos Próximos (N) e Metástases (M)
VEGF	<i>Vascular Endothelial Growth Factor</i> (Fator de crescimento vascular endotelial).

Resumo

O câncer se trata de uma doença complexa, quanto à sua fisiologia, progressão e alta mortalidade. Neste contexto, o tumor colorretal apresenta grande destaque, afinal se trata do segundo mais comum entre ambos os sexos, atrás apenas de tumores sexo específicos. Um dos grandes agravantes do câncer em questão, e a sua progressão lenta e silenciosa, o que se relaciona com um diagnóstico tardio e pior prognóstico. Como tratamentos possíveis há a ressecção cirúrgica, mas a quimioterapia assume papel fundamental pois em muitos momentos a ressecção é impraticável. Contudo tal tumor, apresenta resistência ao tratamento quimioterápico, e assim a busca de biomarcadores para este câncer é de suma importância, pois fornece indicadores prognósticos e terapêuticos que possibilitem tratamento personalizado e precoce. Desta forma, nosso estudo verificou os genes diferencialmente expressos que são secretados pelo CRC e atuam como ligantes e receptores das células tumorais, para assim compreender vias relacionadas com a carcinogênese deste tumor, bem como alvos passíveis de intervenção. Para melhor compreensão e elucidação de como tais marcadores interagem, realizamos o interatoma deste tumor. Para elencar os marcadores mais relevantes no interatoma formado, realizamos a clusterização dos 10 genes que mais apresentaram relações no interatoma formado, para assim verificar quais tinham interações descritas com fármacos. Logo, identificamos MMP9, EFGR e PLAU como possíveis alvos terapêuticos e marcadores diagnósticos, e para verificar o comportamento destes no microambiente tumoral, realizamos estudos de single-cell, verificando em qual tipo celular estes ocorriam. Desta forma, ficou evidente a importância destes no ambiente tumoral do CRC, bem como a possível possibilidade destes serem fatores terapêuticos, pois interagem com fármacos que não são comumente utilizados no tratamento deste tumor, sendo eles Sirolimus, Curcumina, Geldanamicina, Celecoxib.

Palavras chave: Câncer colorretal, Interactoma, terapia personalizada, bioinformática.

Abstract

Cancer is a complex disease in terms of its physiology, progression, and high mortality. In this context, the colorectal tumor stands out, after all, it is the second most common among both sexes, behind only gender-specific tumors. One of the major aggravating factors of cancer in question is its slow and silent progression, which is related to a late diagnosis and a worse prognosis. As a possible treatment, there is surgical resection, but chemotherapy plays a fundamental role because in many moments resection is impracticable. However, such a tumor is resistant to chemotherapy treatment, and thus the search for biomarkers for this cancer is of paramount importance, as it provides prognostic and therapeutic indicators that enable personalized and early treatment. In this way, our study verified the differentially expressed genes that are secreted by the CRC and act as ligands and receptors of tumor cells, in order to understand pathways related to the carcinogenesis of this tumor, as well as possible targets for intervention. For a better understanding and elucidation of how such markers interact, we performed an interactome of this tumor. To list the most relevant markers in the formed interactome, we performed the clustering of the 10 genes that presented the most relationships in the formed interactome, in order to verify which ones had interactions described with drugs. Therefore, we identified MMP9, EGFR, and PLAU as possible therapeutic targets and diagnostic markers, and to verify their behavior in the tumor microenvironment, we performed single-cell studies, verifying in which cell type they occurred. Thus, their importance in the CRC tumor environment was evident, as well as the possible possibility that they are therapeutic factors, as they interact with drugs that are not commonly used in the treatment of this tumor, namely Sirolimus, Curcumin, Geldanamycin, Celecoxib.

Keywords: Colorectal cancer, Interactome, personalized therapy , bioinformatics.

Sumário

Resumo	9
Abstract	10
1. Introdução	12
1.1. O Câncer de Cólon e Reto	12
1.2. Tratamentos para o Câncer de Colón e Reto	13
1.3. Interactoma e sua aplicação na medicina de precisão oncológica	16
1.4. Bioinformática no estudo do câncer	17
2. Justificativa e Hipótese	17
3. Objetivos	18
3.1. Objetivo Geral	18
3.2. Objetivos Específicos	18
Capítulo I	20
Referências Bibliográficas	42
Apêndices	51
Figuras Suplementares	51

1. Introdução

1.1. O Câncer de Cólon e Reto

O câncer se trata do principal problema de saúde pública mundial, sendo a segunda maior causa de morte nos Estados Unidos. Estima-se que em 2022 tenham surgido aproximadamente 1.918.030 novos casos desta patologia neste país, ocasionando aproximadamente 609.030 mortes (SIEGEL *et al.*, 2022). No Brasil, a situação é bem parecida, onde estudos do INCA para o triênio 2020-2022 estimaram que surgissem aproximadamente 685.960 novos casos desta patologia em todo o país durante o período (INCA, 2019).

Dentre os tipos tumorais mais incidentes em ambos os sexos, o câncer de cólon e reto (CRC) ganha destaque, sendo o terceiro mais incidente tanto em homens quanto em mulheres, ficando atrás apenas de tumores específicos para cada um, próstata e mama. No triênio entre 2020-2022 foram estimados aproximadamente, 41 mil novos casos deste tumor no Brasil. Esta prevalência em ambos os sexos, se deve ao fato da progressão deste tumor estar associada a comportamentos comuns em ambos e que são considerados de risco, tais como fumo, consumo de bebidas alcoólicas, uso de antibióticos que afetam a microbiota intestinal, e o comportamento mais prevalente, que é a adesão a chamada dieta ocidental, composta por alto consumo de alimentos rico em gorduras, altamente processados e pouco consumo de frutas e legumes (DEKKER *et al.*, 2019; LI, J. *et al.*, 2021).

Um dos grandes fatores de risco deste câncer é a sua lenta progressão, que ocasiona uma detecção tardia, geralmente em estágio de metástase. Desta forma, a taxa de sobrevivência dos pacientes nos próximos 5 anos, é de aproximadamente 60%, mas metade destes acaba falecendo antes mesmo da ressecção cirúrgica do tumor. Analisando a progressão tumoral, nota-se que ela ocorre seguindo uma sequência de passos conhecida como progressão adenoma-carcinoma, onde ocorrem modificações de vias específicas levando ao desenvolvimento tumoral. Esta progressão se divide em 4 estágios gerais (Figura 1), que apresentam subdivisões específicas de acordo com a classificação TMN, evoluindo de um pólipó que posteriormente se torna um carcinoma *in situ*, até o desenvolvimento do câncer (LA VECCHIA; SEBASTIÁN, 2020).

No estágio 1, há uma evolução de um carcinoma *in situ*, para as camadas submucosa e muscular própria do intestino grosso, mas ainda não se observa a presença de tumores em linfonodos próximos e metástase. Já no estágio 2, o tumor se torna mais invasivo, atingindo a chamada parede do intestino grosso e alguns órgãos próximos. No estágio 3, ocorre a migração para linfonodos próximos, mas ainda não há metástase. Por fim, no estágio 4 é possível observar

metástases para demais órgãos da cavidade abdominal, podendo haver ou não tumores em órgãos mais distantes (AMERICAN CANCER SOCIETY, 2020; LA VECCHIA; SEBASTIÁN, 2020; SIMON, 2016).

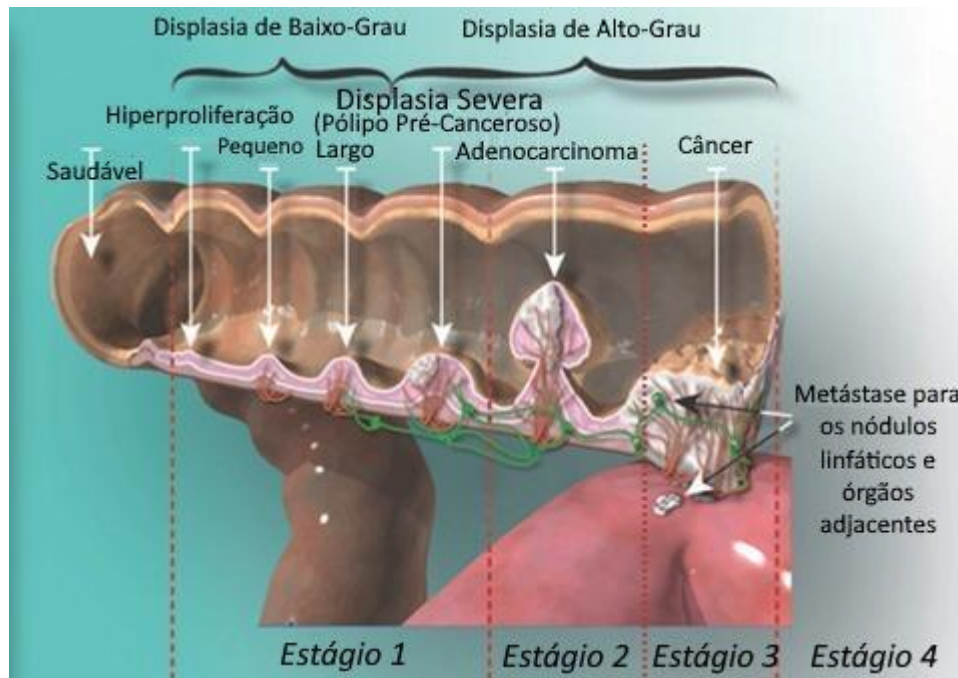


Figura 1: Diferentes estágios do desenvolvimento do CCR, adaptada de SIMON, 2016.

1.2. Tratamentos para o Câncer de Colón e Reto

Os tratamentos para o CRC se baseiam na ressecção cirúrgica combinada com a quimioterapia. No caso de tumores classificados como pólipos, ou seja, no estágio 1, é recomendada a chamada ressecção endoscópica, que pode se dar de diferentes maneiras, utilizando a chamada polipectomia, responsável pela remoção dos pólipos de diferentes tamanhos. Para pólipos menores, até 9mm, utiliza-se a técnica de “Cold Snare”, na qual ocorre cauterização por congelamento deste. Para lesões maiores são utilizadas técnicas de corte, que acabam abrangendo regiões mais profundas da mucosa intestinal (DUMOULIN; HILDENBRAND, 2019).

Porém, embora prático, o procedimento acima se limita a tumores locais, em estágio inicial da doença. Ao se abordar a ressecção cirúrgica de metástase, nota-se que esta se trata de um procedimento altamente invasivo. Observando o comportamento do CRC, um dos locais mais comuns para que se ocorra a metástase é o fígado, e a ressecção cirúrgica deste exige uma recuperação contínua que pode levar de 5 a 10 dias. A ressecção, de fato apresenta uma melhora do prognóstico do paciente metastático, onde a taxa e sobrevida nos primeiros 5 anos acaba

sendo de aproximadamente 71%, e quando combinada com quimioterapia, esta chega a apresentar até 86% de eficácia. Contudo, ainda há a possibilidade de recorrência metastática mesmo após a ressecção, em taxas que podem chegar a 70% nos próximos 2 anos, podendo ocorrer nos ossos, pulmão, cérebro e próprio fígado, além de haver metástases que são inoperáveis, logo também há utilização da quimioterapia associada (AKGÜL *et al.*, 2014).

Associada a ressecção há a utilização da quimioterapia. Essa é inicialmente recomendada em situações de metástases que são inoperáveis, buscando diminuir o tamanho do tumor para que assim seja possível uma ressecção posterior com técnicas adequadas. Tal tratamento auxilia muito nessa situação fazendo com que a taxa de sobrevida de pacientes inicialmente inoperáveis saia de 12% para 33% (AKGÜL *et al.*, 2014).

Passando agora para os tumores na qual a ressecção é possível, nota-se que com o aumento do grau de desenvolvimento tumoral, a taxa de sobrevida nos próximos 5 anos diminui gradativamente, passando de 90% no estágio I, para até 40% no estágio III. Além disso, é muito comum o desenvolvimento de metástases, e assim a combinação da ressecção com a quimioterapia pode levar a um melhor prognóstico, e diminui as chances de recorrência (POLASTRO; EL HACHEM; HENDLISZ, 2018). Os tratamentos quimioterápicos para o CRC se baseiam em três linhas de tratamento, onde são combinados fármacos citotóxicos com anticorpos específicos para VEGF e EGFR. Além disso, alguns marcadores moleculares tumorais, são fundamentais para verificar a eficácia de cada combinação de tratamento, como mutações em *RAS* e *BRAF*. A resposta do paciente a primeira linha de tratamento é fundamental para determinar o sucesso da quimioterapia no CRC, onde o principal fármaco de todo o tratamento é o 5-Fluororacil (5-FU), com combinações deste (POLASTRO; EL HACHEM; HENDLISZ, 2018).

Na primeira linha, há a combinação deste fármaco com o ácido folínico, oxiplatina e irinotecano (FOLFOXIRI), enquanto na segunda linha há a combinação com ácido folínico e irinotecano ou oxiplatina, FOLFIRI e FOLFOXI respectivamente (BENSON *et al.*, 2021; VAN CUTSEM *et al.*, 2016). Por fim, na terceira linha, há a mudança das combinações de anti-VEGF e anti-EGFR, mas neste caso, a taxa de sobrevida é amplamente reduzida (Tabela 1) (MODEST; PANT; SARTORE-BIANCHI, 2019).

Tabela 1: Eficácia das linhas farmacológicas no decorrer do tratamento do câncer colorretal. Adaptada de (MODEST; PANT; SARTORE-BIANCHI, 2019)

Parâmetros	Primeira Linha	Segunda Linha	Terceira Linha
Taxas de Resposta	38-65%	5-36%	1-31%
Taxa de Controle	81-90%	59-86%	41-75%
Sobrevida Livre de Progressão	9-12 meses	4-7 meses	2-5 meses

Observando a tabela acima, conclui-se que a presença de 3 linhas de tratamento, além de variadas combinações, indica que o tumor pode desenvolver quimioresistência. Dentre as causas conhecidas para este fenômeno há inativação da droga, efluxo acelerado da droga e modificações celulares (MARIN *et al.*, 2012). Isso se trata de algo muito comum, que ocorre com diferentes fármacos amplamente utilizados. Wei e colaboradores, 2019, demonstraram a quimioresistência em diferentes fármacos utilizados no tratamento do CRC, tais como 5-FU, oxiplatina, irinotecan, entre outros, adicionando papel para os RNAs não codificantes na modulação deste fenômeno. Ficou demonstrado que, tais RNAs acabaram por gerar a quimioresistência ao atuarem em genes alvos para os fármacos. Além disso, o próprio sistema imune contribui para a modulação da quimioresistência, através de macrófagos associados ao tumor, maior secreção de IL-6 e da inibição do miR-204-5p, que originalmente, torna as células mais sensíveis ao 5-FU (YIN *et al.*, 2017). Por fim, existem estudos que demonstram que até mesmo a própria microbiota pode auxiliar na promoção da quimioresistência tumoral no CRC, como no caso da *Fusobacterium nucleatum*, que modula a autofagia em linhagens celulares HT-29 e HCT-116, e assim garante resistência ao 5-FU e oxiplatina (YU *et al.*, 2017).

A toxicidade atrelada ao tratamento quimioterápico também deve ser lembrada, visto que muitos dos fármacos utilizados para combater o CRC apresentam efeitos adversos. Entre eles, cita-se o Irinotecan, que induz alopecia e toxicidade gastrointestinal, e a Oxiplatina, que promove neurotoxicidade cumulativa (EISTERER; PRAGER, 2019). No caso dos fármacos baseados no 5-FU, é descrito toxicidade no fígado além de toxicidade cardíaca (ALESSANDRINO *et al.*, 2019; DEPETRIS *et al.*, 2018).

Contudo, o avanço científico atrelado a grande quantidade de dados disponíveis, permitiu que fossem desenvolvidos tratamentos baseados em aspectos individuais dos pacientes, sejam estes comportamentais, mas principalmente genéticos, através da chamada medicina de precisão, que ganhou destaque a partir de 2006, com o surgimento das tecnologias

de sequenciamento de próxima geração (NGS). Na oncologia, a medicina de precisão envolve a identificação de mutações gênicas, que sirvam de biomarcadores preditivos da doença e permitam assim prever eficácia ou resistência a terapias (TSIMBERIDOU *et al.*, 2020). Tudo isso, pode ser realizado de maneira muito mais prática, afinal a busca por biomarcadores circulantes se vale da chamada biópsia líquida, na qual estes podem ser detectados através de uma amostra de sangue, sem a necessidade de procedimentos invasivos no paciente (TSIMBERIDOU *et al.*, 2020). Um dos métodos de detecção de biomarcadores, se dá através do chamado interactoma, na qual ao listar os ligantes e receptores consegue-se mensurar a ligação entre estes e se obter uma lista de interação entre proteínas, para assim mensurar locais passíveis de intervenção médica.

1.3. Interactoma e sua aplicação na medicina de precisão oncológica

Como citado anteriormente, desde que surgiram as NGS, os estudos genéticos e ômicos ampliaram muito mais as possibilidades de abordagem. Dentre as abordagens que foram desenvolvidas, uma é a compreensão das mudanças da biologia celular durante uma determinada doença (LEMBERGER, 2007). Um dos métodos mais eficazes de se realizar esta análise, é através do chamado Interactoma, que é definido como: “Todo o complemento de associações moleculares funcionais que podem ocorrer em uma célula”, permitindo que haja uma compreensão funcional da biologia celular em um determinado espaço, tempo e/ou condição (JAMES; WIPAT; COCKELL, 2022).

No câncer, a análise do interactoma permite inferir mecanismos precisos de comportamento tumoral, desde vias relacionadas com a carcinogênese e mecanismos de imunoevasão, fornecendo assim informações sobre as interações estabelecidas entre as células tumorais e o microambiente tumoral (KALKAT *et al.*, 2018; LEE, J. J. *et al.*, 2021). Contudo, um dos grandes avanços que este tipo de análise permite, é mensurar a ação de fármacos sobre as células tumorais, avaliando genes chaves da sinalização celular tumoral, e verificando se estes sofrem mutações que possam ocasionar quimioresistência ou se tornarem alvos de possíveis tratamentos (NUSSINOV; TSAI; JANG, 2021).

No CRC, alguns estudos com interactoma revelaram achados importantes sobre a tumorigênese deste. Tal análise revelou uma lista de MOTIFS curtos lineares, que estão relacionados com a tumorigênese. Desta forma, é possível desenvolver moléculas curtas, que se liguem nesses MOTIFS e assim inibam suas ações, revelando assim valor terapêutico (FASANO *et al.*, 2022). Contudo, um dos grandes enfoques foi o uso dessa abordagem para

mensurar fármacos que inibam a tumorigênese deste tumor. A análise de interactoma do CRC, demonstrou que drogas como Abacavir, Exemestano, Nortripilina, Hidrocloreto e a Tolcapona como potenciais fármacos com eficácia terapêutica no tratamento do CRC, mesmo não estando diretamente associados ao tratamento do câncer (BEKLEN *et al.*, 2021). Além disso, essa abordagem permite avaliar também a eficácia e ação de moléculas naturais no tratamento, como o caso do Cacau (ZAMANIAN-AZODI; REZAEI-TAVIRANI, 2019). Tudo isso só é possível através da integração de ferramentas de bioinformática, integradas que permitem que sejam analisadas uma grande quantidade de dados disponíveis. Desta forma, podemos se valer destes dados para estudos acerca da terapêutica de diversas doenças, principalmente o câncer.

1.4. Bioinformática no estudo do câncer

O termo *in silico* foi criado em conjunto com os termos *in vivo* e *in vitro*, e se refere a estudos que utilizam computadores ou através de simulação computacional. Essa abordagem, envolve técnicas que permitem a compreensão de sistemas biológicos como um todo, através da integração de ferramentas que permitem validação, classificação, inferência, e até mesmo predição biológica (JEAN-QUARTIER *et al.*, 2018).

No câncer, inúmeras são as ferramentas que auxiliam no desenvolvimento de pesquisas, fornecendo informações sobre prognóstico, respostas a drogas, além de permitirem uma maior compreensão do comportamento tumoral, tudo isso atrelado a um grande aumento do número de dados disponíveis e do refinamento das análises empregadas (LI, K. *et al.*, 2019; ZHENG *et al.*, 2020). Contudo, ainda não foi realizado um estudo integrativo, entre a análise de interação das moléculas secretadas pelo CRC fundamentais na sinalização das vias tumorigênicas deste, passíveis de serem biomarcadores diagnósticos, prognósticos e terapêuticos, através do uso de ferramentas de bioinformática.

2. Justificativa e Hipótese

Como demonstrado anteriormente, o CRC está intimamente associado ao padrão alimentar e comportamental da sociedade atual, afetando de maneira similarmente agressiva ambos os sexos. Contudo, mesmo conhecendo muito bem os fatores de risco, a detecção deste tipo tumoral ocorre geralmente de maneira tardia, em condição de metástase, o que acaba gerando um pior prognóstico, associado a reincidência e resistência ao tratamento.

Sobre o tratamento, como já foi dito, o principal fármaco utilizado é o 5-fluoruracil, e suas combinações com anti-corpos anti VEGF e EGFR, além de outros fármacos como Irinotecan (FOLFIRI), e Oxiplatina (FOLFOX). Contudo, é amplamente descrito para a quimioresistência para o CRC, de modo que já são listadas 3 linhas de tratamento, mas com o passar das linhas, a eficácia é reduzida e consequentemente a taxa de sobrevida também.

Desta forma, é de suma importância que se consiga traçar padrões moleculares, que permitam compreender melhor como ocorre a progressão deste tipo tumoral, bem como a interação que suas células estabelecem com as demais do microambiente ao redor, através de proteínas chaves que são secretadas pelas células tumorais, para a execução de processos chaves na progressão tumoral. Mapear e compreender tais proteínas, permite a execução de processos fundamentais no combate ao CRC, detecção precoce destes através de biopsia líquida, bem como a realização de terapia personalizada, tendo como alvo estas.

Para realizar a detecção das proteínas citadas, é fundamental que se analise uma grande quantidade de amostras, e para isso os estudos *in silico* são amplamente utilizados. Isso se deve ao fato de existir uma grande quantidade de dados já descritos e disponíveis para análise, que permitem a realização do chamado interactoma, na qual são listados os ligantes e receptores que estão com a sua expressão alterada no tumor, e como ocorre a interação destes, para assim compreender melhor vias e processos que estão envolvidos na carcinogênese do CRC e assim permitir a intervenção precoce e precisa.

A partir disso, nossa hipótese é de que a análise do interactoma dos ligantes e receptores diferencialmente expressos no CRC, permite a detecção de moléculas chave para a sinalização celular da progressão deste tumor. Essa detecção, se faz fundamental para que ocorra o diagnóstico precoce da doença, bem como para a descoberta de novas estratégias terapêuticas, personalizadas, que se baseiam nas moléculas detectadas.

3. Objetivos

3.1. Objetivo Geral

Caracterizar o conjunto de ligantes e receptores envolvidos na progressão do Câncer Colorretal, para identificar possíveis biomarcadores diagnósticos e terapêuticos e as interações estabelecidas por estes.

3.2. Objetivos Específicos

- Verificar a expressão dos genes listados no secretoma humano, em amostras normais do banco de dados *Genotype Tissue Expression* (GTEx) e amostras tumorais de cólon e reto do banco de dados *The Cancer Genome Atlas* (TCGA);
- Selecionar as amostras do TCGA com alta pureza tumoral;
- Realizar a expressão diferencial nos diferentes grupos;
- Identificar ligantes e receptores que apresentaram expressão aumentada no CRC;
- Analisar a interação (interactoma) dos pares de ligantes e receptores detectados;
- Verificar a expressão dos pares mais significativos em dados de Sequenciamento de RNA de célula única (*Single Cell RNA-SEQ/scRNA-seq*);
- Realizar a rede de Interação Proteína Proteína (PPI) do conjunto dos pares identificados na análise de interação;
- Elencar os elementos mais relevantes da rede obtida, através de clusterização;
- Realizar análise de predição de drogas nos alvos identificados após clusterização.

Capítulo I

Colorectal cancer interactome/secretome integration reveals a new approach for determining biomarkers and therapeutic targets for this tumor.

Gabriel Henrique Caxali¹, Mirian Carolini Esgoti Aal¹, Laíza Brugnerotto¹, Jakeline Santos Oliveira¹, Robson Francisco Carvalho¹, Flávia Karina Delella¹

¹Universidade Estadual Paulista (Unesp), Instituto de Biociências, Departamento de Biologia Estrutural e Funcional, Botucatu, São Paulo, Brasil

Corresponding Author:

Flávia Karina Delella

250 Professor Doutor Antonio Celso Wagner Zanin St, Botucatu, SP, Brazil. 18618-689

Phone: +55 14 3880 0500

e-mail: flavia.delella@unesp.br

Resumo

O câncer colorretal, se trata de do terceiro tipo tumoral mais incidente entre ambos os sexos, ficando atrás apenas dos tumores específicos para cada um, mama e próstata. Sua progressão é lenta, se tratando da evolução de um adenoma, para um carcinoma, em um processo que passa por 4 estágios. Por possuir uma lenta progressão, em grande parte dos casos a detecção desse tumor ocorre de maneira tardia, geralmente quando já há metástases peritoneais, direcionadas principalmente para fígado e ossos. Como principais tratamentos para este tipo tumoral, há a ressecção cirúrgica, associada a quimioterapia. Ainda assim, uma das principais problemáticas é que este tumor ainda apresenta resistência ao tratamento quimioterápico. Uma possibilidade através da grande quantidade de dados disponíveis, é a busca de marcadores através de análise integrada de banco de dados, permitindo assim que se busquem terapias personalizadas e novas, que auxiliem no combate a diversos tumores, como o CRC. Logo nosso estudo buscou analisar o conjunto de proteínas secretadas no CRC, destacando as que atuavam como ligante e receptores, a fim de caracterizar a interação entre eles, o chamado Interactoma, e assim observar novos alvos terapêuticos e drogas que interagem com estes. Ao final, conseguimos identificar um conjunto de 4 genes (FGF, EGFR, MMP9 e PLAU), que possuem expressão aumentada em regiões distintas do CRC, de modo que estes são passíveis de serem detectados através de biópsia líquida, atuam em vias fundamentais de progressão tumoral e interagem possuem interações compartilhadas com 6 fármacos. Dessa forma, acreditamos que nossos resultados oferecem uma abordagem para realizar a detecção precoce do tumor, bem como fornecer alternativas para novos tratamentos baseados na interação dos genes detectados e fármacos.

Palavras-chave: Câncer colorretal, Interactoma, terapia personalizada, bioinformática.

Introdução

O câncer se trata do maior problema de saúde pública mundial, responsável pela segunda maior causa de morte nos Estados Unidos, ficando atrás apenas das doenças cardíacas (SIEGEL *et al.*, 2022). No Brasil, temos um padrão similar, onde há considerável aumento do número de tumores diagnosticados nos últimos anos, exceto durante o período pandêmico, na qual a quarentena afetou drasticamente a diagnose do câncer (INCA, 2019; MARQUES *et al.*, 2021)

Dentre os variados tipos de tumores que afetam a população mundial, o câncer colorretal apresenta alta prevalência entre ambos os sexos, sendo o segundo mais comuns em mulheres adultas e o terceiro mais comuns em homens, representando aproximadamente 9,2% de todos as mortes por tumor ao redor do mundo (LI, J. *et al.*, 2021). A origem deste tumor varia por toda a extensão do intestino grosso, onde nos Estados Unidos, aproximadamente 41% se origina no cólon proximal, 22% no cólon distal e 28% no reto (THANIKACHALAM; KHAN, 2019). Contudo mesmo apresentando origens em regiões distintas, a tumorigênese deste segue um padrão bem descrito, onde evolui de um pólipó, oriundo de uma cripta com proliferação exacerbada, e passa por 4 estágios até se tornar metastático (DEKKER *et al.*, 2019; LA VECCHIA; SEBASTIÁN, 2020).

Este tumor se faz altamente incidente em ambos os sexos, pois sua patogênese depende de fatores comuns a ambos. Um dos principais fatores relacionados com a progressão deste tumor é alimentação inadequada, chamada dieta ocidental, rica em gordura e pobre em nutrientes, representada principalmente pelo alto consumo de carne e processados acaba contribuindo para tal progressão (MASDOR *et al.*, 2022).

Dentre os fatores que contribuem para isso, está a inflamação local causada por estes alimentos, o que acaba gerando um desbalanço na microbiota local (NIEDERREITER; ADOLPH; TILG, 2018). Outro fator que se encontra intimamente relacionado com a má alimentação, é o sedentarismo, sendo que a prática de atividade física pode contribuir para uma redução de até 27% na chance do desenvolvimento do CRC (FRIEDENREICH; RYDERBURBIDGE; MCNEIL, 2021). Além disso, estudos demonstram que fatores como fumo e consumo exacerbado de bebidas alcoólicas também são grandes agravantes no desenvolvimento deste tumor (RAWLA; SUNKARA; BARSOUK, 2019).

A ressecção cirúrgica dos pólipos é um tipo de tratamento deste tumor, porém trata-se de um procedimento invasivo, visto que é realizada de forma semelhante a uma colonoscopia. Além disso, em certas regiões do intestino grosso, a ressecção cirúrgica é impraticável, e mesmo

quando realizada existe a possibilidade de recorrência tumoral (DUMOULIN; HILDENBRAND, 2019). Logo, atrelado a ressecção cirúrgica, é muito comum o uso da quimioterapia associada.

A quimioterapia no CRC, se baseia no uso do quimioterápico 5-Fluoruracil (5-FU), que possui atividade genotóxica. Contudo, já é descrito o processo de quimioresistência tumoral, assim foram elaboradas 3 linhas de tratamento baseadas na combinação do 5-FU, com outras drogas, como a Oxiplatina, chamado de FOLFOX, com o Irinotecano, chamado de FOLFIRI. Tais combinações ainda são associadas com inibidores específicos de EGFR e VEGF, mas ainda assim ainda existe resistência e com o passar das linhas de tratamento, a chance de sobrevida diminui (BILLER; SCHRAG, 2021; JOHDI; SUKOR, 2020; MODEST; PANT; SARTORE-BIANCHI, 2019).

Desta forma, uma alternativa é a busca de substâncias ou atividades que possam ser objetivamente avaliada, para mensurar um processo biológico, normal ou tumoral, os chamados biomarcadores (WU; QU, 2015). No CRC, alguns genes que apresentam expressão alterada já são definidos como biomarcadores, tais como VEGF, BRAF e PTEN (LECH *et al.*, 2016). Um dos métodos mais eficazes para a busca destes biomarcadores, é se valer da abundância de dados disponíveis presentes em variados bancos de dados online, as chamadas análises *in silico*.

No caso do estudo do câncer, estes tipos de análise permitem a compreensão do metabolismo tumoral, bem como as vias que são alteradas para que esse processo ocorra, e assim são possíveis de detectar alvos terapêuticos e prognósticos passíveis de intervenção por drogas (LI *et al.*, 2019). Para conseguir uma informação mais precisa, pode-se inferir, como ocorre a interação destes possíveis alvos, através do chamado Interactoma, que é definido como “Todo o complemento de associações moleculares funcionais que podem ocorrer em uma célula”, permitindo que haja uma compreensão funcional da biologia celular em um determinado espaço, tempo e/ou condição (JAMES; WIPAT; COCKELL, 2022).

Portanto, a integração de dados presentes nas demais plataforma disponíveis possibilita que sejam elencados novos possíveis alvos para o tratamento do CRC. Tal processo é de suma importância, visto que este tumor apresenta resistência ao tratamento e prognóstico tardio. Logo o objetivo deste trabalho é utilizar da integração entre o conjunto de proteínas secretadas pelo organismo que atuam como ligantes e receptores, e a interação entre elas, afim de encontrar novos alvos farmacológicos, passíveis de intervenção no tratamento do CRC.

2.1. Material e Métodos

2.1.1 Escolha do Secretoma de Referência

A lista com o secretoma utilizado como referência, foi obtida através do banco de dados *The Human Protein Atlas* (<https://www.proteinatlas.org/>) (UHLÉN *et al.*, 2015). Foi utilizada a lista que contém 2793 genes, contemplando a secreção no sangue, cérebro, sistema digestivo, sistema reprodutivo masculino e feminino, imunoglobulinas, compartimento intracelular e membrana, locais desconhecidos, outros tecidos e secretado da matriz extracelular (UHLÉN *et al.*, 2019).

2.1.2 Download dos dados de expressão de amostras Normais e Tumorais.

Os genes listados no secretoma, foram submetidos a plataforma *Xena Browser* (<https://xenabrowser.net/>) (GOLDMAN *et al.*, 2020). Nesta, utilizamos o estudo *TCGA TARGET GTEx*, para obter as amostras dos tecidos normais disponíveis na plataforma *GTEx* (*Genotype Tissue Expression*) e tumorais oriundos do *TCGA* (*The Cancer Genome Atlas*). Foram selecionadas as amostras de tumor primário de cólon (*COAD- Colon Adenocarcinoma*), e de tumor primário de reto (*READ- Rectum Adenocarcinoma*). Os dados de expressão foram obtidos em $\text{Log}_2 \text{ expected count} + 1$.

2.1.3 Filtro das Amostras de Alta Pureza tumoral

Após download dos dados de expressão, foi realizada a filtragem das amostras que apresentavam uma maior proporção de células tumorais, ou seja, alta pureza tumoral. Este processo foi realizado de acordo com o método desenvolvido por Aran; Sirota; Butte, 2015, que se baseia na utilização de 4 parâmetros para estimar a pureza de uma amostra tumoral oriunda do *TCGA* (*ESTIMATE*, *ABSOLUTE*, *LUMP* E *IHC*), e elaboram o *consensus measurement of purity estimations* (*CPE*). Foram consideradas amostras com alta pureza tumoral, as amostras cujo $\text{CPE} > 0.8$.

2.1.4 Expressão Diferencial das Amostras normais e de alta pureza tumoral

Após download dos valores de expressão, foi utilizada a ferramenta *Appyters* (<https://appyters.maayanlab.cloud/#/>) (CLARKE *et al.*, 2021), para realizar a análise de expressão diferencial das listas obtidas na plataforma *XENA Browser*. Por se tratarem de dados de *RSEM expected counts*, foi utilizado do método de análise *edgeR*, onde foram considerados significativos genes em que $p\text{-value} < 0.05$ e $\text{Log}_2 \text{ Fold Change (Fc)} > 1.5$.

2.1.5 Busca de Ligantes e Receptores nos DEGs up regulados identificados

Os genes up regulados detectados nas listas de expressão diferencial de cólon e reto foram comparados com a lista de ligantes e receptores descrita por Ramilowski *et al.*, 2015. Tal lista contém 2422 pares de genes ligante/receptor descritos, através de consulta em bancos de dados diversos (ARMINGOL *et al.*, 2021). Foram considerados significativos os pares na qual ambos componentes (ligante e receptor) apresentassem $\text{Log}_2FC > 1.5$. Foi criada uma lista única, contendo os ligantes e receptores de cólon e reto, seus respectivos Log_2 de FC e *False Discovery Rate* (FDR).

2.1.4 Construção de Redes de Interação Proteína Proteína (PPI) dos genes identificados

Os genes detectados foram submetidos a plataforma *Search Tool for the Retrieval of Interacting Genes/Proteins* (STRING) (<https://string-db.org>) (SZKLARCZYK *et al.*, 2021). Foram utilizados os parâmetros definidos OLIVEIRA *et al.*, 2020: *Meaning of network edges: confidence; Active interaction sources: Experiments, database, co-expression, neighborhood, co-occurrence; Minimum required interaction score: high confidence of 0.900; Display simplifications: hide disconnected nodes in the network*. As redes geradas foram importadas para o software *Cytoscape 3.9.1* (SHANNON *et al.*, 2003), através da extensão *StringApp* (DONCHEVA *et al.*, 2019). Por fim, foi realizada a clusterização da rede importada, utilizando a extensão *cytoHubba* (CHIN *et al.*, 2014), onde foram elencados os 10 genes que mais apresentavam interações, através do método *degree*.

2.1.5 Verificação de alvos terapêuticos através das assinaturas genéticas do CRC

Para verificar os possíveis alvos terapêuticos dos datasets COAD/READ, foi utilizada a plataforma *Open Cancer Therapeutic Discovery* (OCTAD) (ZENG *et al.*, 2021). Nesta, foram selecionadas para análise apenas as amostras do TGCA, que apresentaram alta pureza tumoral, e amostras normais do GTEx. Para realizar a assinatura genética de ambos os *datasets*, foi utilizado o método EdgeR, presente na própria plataforma, considerando significativos os genes que apresentavam *Fold Change* > 1.5 e *Adjusted p-value* < 0.001 . Após isso, na mesma plataforma, foi realizada a análise de possíveis alvos terapêuticos, baseados nos resultados da assinatura genética de cada tipo de tecido. Para isso a plataforma comparou a assinatura genética obtida em cada análise com o seu banco de dados de compostos relacionados com expressão de determinado gene, que conta com 19.127 amostras de tecidos de pacientes, além de 50 tipos tumorais e 12.442 diferentes compostos.

Após isso, foram selecionados apenas os genes detectados na lista do interactoma que apresentaram o maior número de interações detectadas pelo *cytoHubba*. Para complementar a lista encontrada na plataforma OCTAD, os mesmos genes foram submetidos a plataforma *Drug Gene Interaction Database* (DGIdb)(<https://www.dgldb.org>) (COTTO *et al.*, 2018) , que contém informações de 30 bancos de dados variados sobre a sensibilidade de genes a determinado fármaco, apresentando em seu banco de dados 41.100 genes, 9.495 drogas e com 29.783 interações descritas.

2.1.6 Análise de Enriquecimento Funcional

A lista de genes presentes no interactoma foi submetida a análise de enriquecimento funcional, através da plataforma EnrichR (XIE *et al.*, 2021). Dentro da plataforma foram utilizados os bancos de dados *GO: Biological Process* e *GO: Molecular Function*. Posteriormente, foram selecionados apenas os processos comuns a ambos os componentes do par (ligante e receptor). Após isso, a mesma lista foi submetida na plataforma STRING, seguindo os mesmos parâmetros descritos anteriormente, mas foram selecionados os termos enriquecidos em na plataforma EnrichR para que o enriquecimento fosse demonstrado na rede gerada.

2.1.7 Análise de dados de sc-RNA-Seq

Para ter maior precisão da expressão dos genes detectados em cada tipo celular do microambiente tumoral, foram utilizados dados de *single cell RNA-Seq*. Utilizamos para isso os dados de câncer colorretal de Lee. *et al.*, 2020, que conta com 15 amostras de carcinoma colorretal, oriundas do Samsung Medical Center, em Seul. Os dados dessas amostras foram obtidos a partir do *Curated Cell Cancer Atlas* (3CA; <https://www.weizmann.ac.il/sites/3CA/>). Os dados de *counts* foram processados utilizando o pacote Seurat (v.4.0.6), no software RStudio (HAO *et al.*, 2021). Para realizar a distinção das células foi realizada a expressão diferencial dentro do pacote, e posterior Análise de Componente Principal (*Principal Component Analysis* – PCA). As PCs de 1 a 10 foram usados para clusterização gráfica na resolução de 0.5 e na visualização dos clusters foi utilizado o gráfico de redução dimensional UMAP. Os tipos celulares distintos foram identificados com base na expressão diferencial entre os diferentes clusters, onde foram detectados variados marcadores celulares. Para a identificação do tipo celular correspondente aos marcadores canônicos (**Tabela 2**), foi utilizada a plataforma CellMarker 2.0 (<http://117.50.127.228/CellMarker/>) (HU *et al.*, 2023). Para avaliação do perfil de expressão gênica entre os diferentes tipos celulares foram utilizadas as funções *dotplot* e *vlnplot*, presentes no pacote Seurat (v.4.0.6).

Tabela 2: Marcadores Canônicos utilizados na identificação dos diferentes tipos celulares presentes nos dados de sc-RNA-Seq

Tipo Celular	Marcadores Canônicos
Célula B	IGLC2; IGHA; IGLC3; JCHAIN; CD79a; BANK1; MS4A1; MZB1; IGLC2; IGLC3; IGLC7; IGHG3; IGHG4;
Cancer Stem Cell	OLFM4.
DCLK1+Progenitor Cell	COL1A1; COL1A2; LUM; DCN
Enterocyte	FABP1; PHGR1; CKB; PRAP1
Célula Caliciforme	SPINK4; TFF1.
Célula Tumoral	UBE2C; CDKN3; RAMP2; PLVAP; GNG11; VWF
Monócito	S100P; PLAC8.
Natural Killer	CD3D; TRAC; KLRB1; CCL5; CD7; IFNG; GZMA; CD8A; GZMB; IRF7; NR3C1
Célula de Paneth	IL1B; CCL3; CXCL8; IL1RN; SOD2; C1QA; C1QB; C1QC; HLA-DRA/DPB1
Músculo Liso	ACTA2; CALD1

2.1.8 Construção dos Volcano Plots e Diagrama de Venn

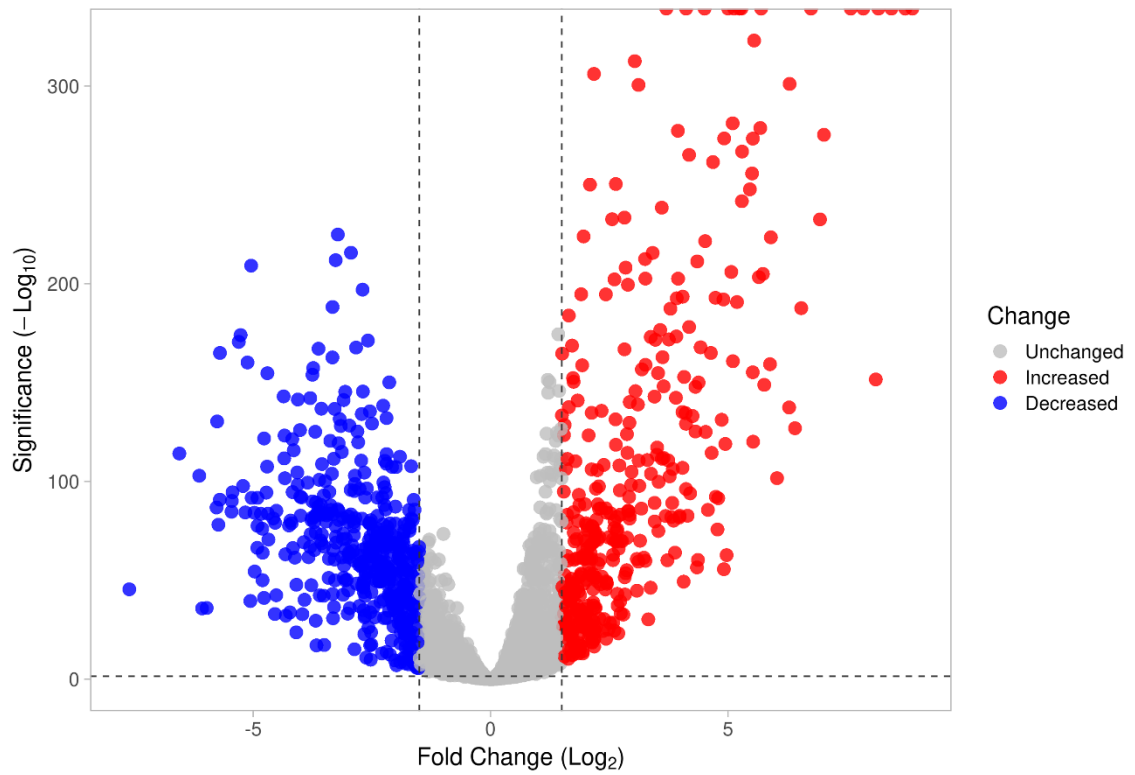
Os *Volcano Plots* com os dados de expressão foram feitos a partir da plataforma online VolcanoseR (<https://huygens.science.uva.nl/VolcanoseR/>) (GOEDHART, 2020). Os dados de expressão foram representados em Log2FC e com Log10 de *p-value*. Foram criados 2 gráficos, um para a análise de expressão em *COAD* e outro para *READ*. Além disso, foi criado um diagrama de Venn para verificar os genes *up* regulados comuns e exclusivos entre *COAD* e *READ*, através da ferramenta *Venn Diagram* (<https://bioinformatics.psb.ugent.be/webtools/Venn/>) .

2.2 Resultados

2.2.1 O perfil de expressão varia entre as amostras de alta pureza de *COAD* e *READ*

Após submeter a lista do secretoma na plataforma Xena, foram obtidos os dados de expressão de 2696 genes, nas diferentes amostras disponíveis nos bancos de dados GTEx e TCGA. No caso de *COAD*, foram obtidas no total, 594 amostras, sendo 307 amostras de tecido normal de intestino grosso, e 287 de tecido tumoral. Contudo, destas 287 tumorais, apenas 114 apresentavam alta pureza segundo a lista de Aran; Sirota; Butte, 2015. No caso de *READ*, não haviam amostras de tecido normal presentes no banco de dados GTEx, logo optamos por utilizar as amostras selecionadas para *COAD*. Desta forma, foram obtidas no total, 399 amostras, sendo 307 normais e 92 tumorais. Destas 92 amostras tumorais, apenas 39 apresentavam alta pureza (**Tabela 3**). Com a análise de expressão diferencial, foi possível notar que a quantidade de genes diferencialmente expressos varia entre os tecidos, onde em *COAD* foram encontrados 377 genes *up* regulados e 460 *down* regulados e em *READ* foram encontrados 405 genes *up* regulados e 474 genes *down* regulados (**Figura 1**) Os genes estão detalhados na Tabela Suplementar 1.

A Differentially Expressed Genes in COAD High Purity Samples



B Differentially Expressed Genes in READHigh Purity Samples

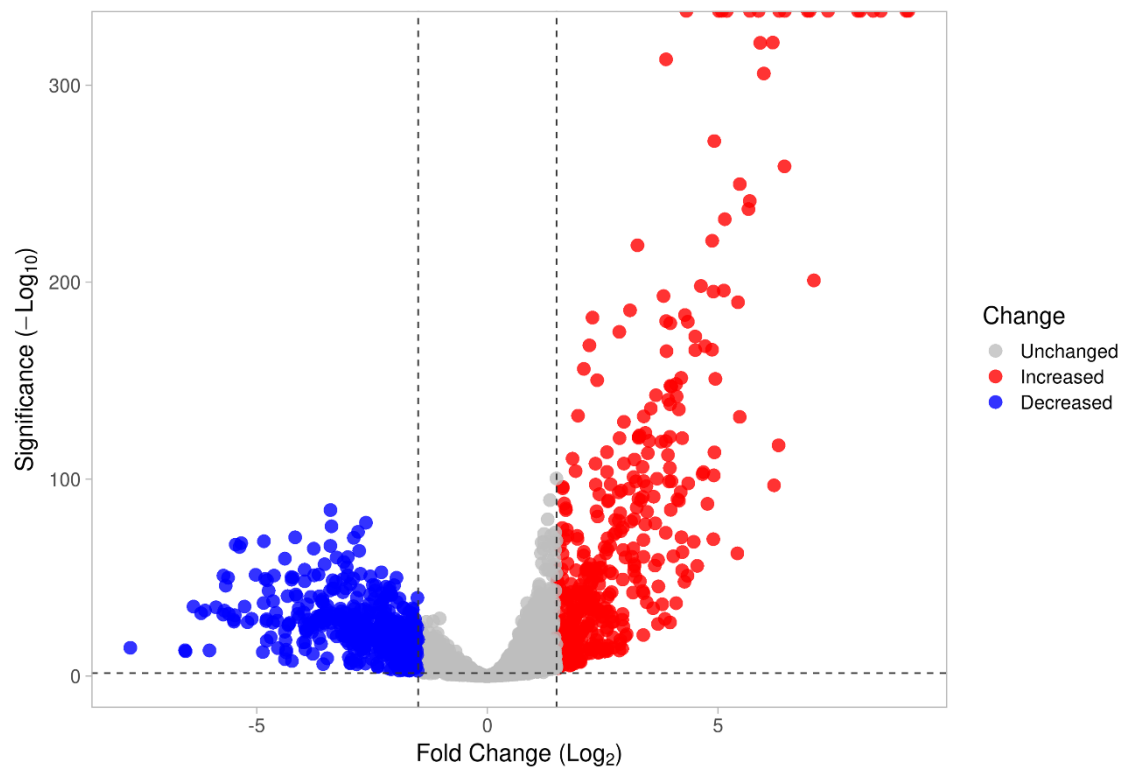


Figura 1: *Volcano Plots* dos dados de expressão de COAD (A) e de READ (B). Os pontos vermelhos representam genes na qual houve aumento de expressão e os azuis genes na qual houve redução. Dados representados em Log2FC e -log10 de *p-value*.

Posteriormente, foram comparados os genes up apresentados em cada um dos grupos. Neste aspecto, notou-se que há diferença considerável entre os DEGs up regulados em cada tecido. Isso fica evidente ao vermos que apenas 76 genes up são compartilhados entre cólon e reto (**Figura 2**), demonstrando comportamento específico entre a expressão das células tumorais de cada tecido. Isso ficou demonstrado ao analisarmos PPI de cada um dos tecidos, onde após clusterização com o cytoHubba, foram obtidas redes distintas (**Figura Suplementar 1 A-B**).

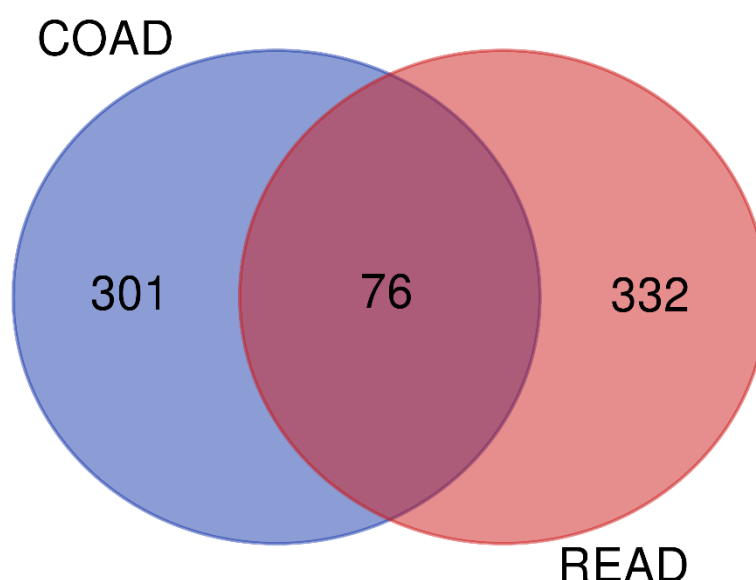


Figura 2: Diagrama de Venn dos genes up regulados detectados nas amostras tumorais de cólon e reto do TCGA.

2.2.2 As interações mais significativas variam entre os diferentes tecidos analisados

Os genes up regulados totais em ambos os estudos, ao serem comparados com a lista do conectoma de Ramilowski 2016, apresentou 30 pares na qual ligante e receptor formavam um com aumento de expressão significativo ($FC > 1.5$ e $p\text{-value} < 0.05$). Esta lista foi submetida ao STRING, onde obteve-se o interatoma (**Figura 3B**) destes tumores. Ao se analisar o *communication score* dos pares, foi possível notar padrões de interação diferentes entre cólon e reto, onde o primeiro apresentou como pares mais significativos: MMP9_EPHB2 e TNFSF11_TNFRSF11B. Já para reto os pares mais significativos foram: FGF2_FGFR3, SERPINA1_LRP1. Ainda tivemos pares, cujo o *communication score* ficou semelhante,

revelando expressão similar em ambos, como: MMP12_PLAUR e PLAUR_PLAUR (**Figura 3A**).

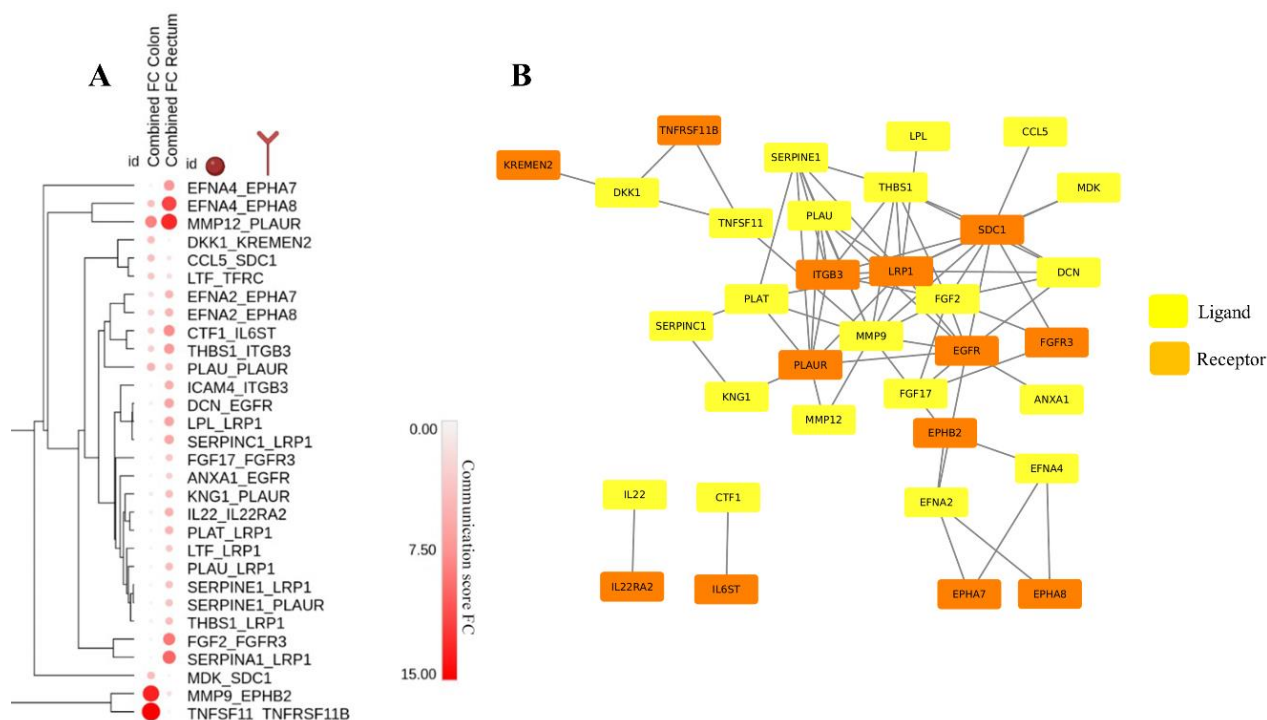


Figura 3: Caracterização dos 30 pares detectados no secretoma das amostras de tumor colorretal. (A) Bubble Heatmap dos 20 pares up regulados nos tumores de Cólon e Reto. A intensidade do vermelho da bolha representa os valores combinados de Log2 Fold Change e o tamanho destas o FDR. (B) Interactoma dos genes up regulados detectados nos 30 pares identificados como ligantes e receptores.

2.2.3 Os genes destacados dos pares se relacionam com vias fundamentais de tumorigênese

O enriquecimento dos genes presentes no interactoma formado demonstrou que grande parte destes se relacionam com processos muito bem descritos na progressão tumoral (**Tabela Suplementar 2**). Podemos destacar alguns como *Positive Regulation of cell population proliferation* (LTF, THBS1, FGF2, EGFR, CTF1, FGFR3, FGF17, TFRC, MMP9, ANXA1, IL6ST, MDK, ITGB3, MDK, MMP12, CCL5), *Positive regulation of Cell Migration* (SERPINE1, LRP1, THBS1, FGF2, EGFR, MMP9, PLAUR, ANXA1, IL6ST, MDK, ITGB3, MMP12 e CCL5). Além disso, houve enriquecimento da regulação de processos inflamatórios, como visto em *Inflammatory response* (LRP1, THBS1, KNG1, EGFR, SERPINC1, ANXA1, SDC1, MDK, SERPINA1, IL22, CCL5). Tais dados de enriquecimento estão demonstrados na **Figura 4**. Além disso, a clusterização através do método *degree*, revelou que os genes que apresentam o maior número de relações na lista são: LRP1, SDC1, ITGB3, THBS1, MMP9, FGF2, PLAUR, PLAUR, SERPINE1 e EGFR (**Figura Suplementar 2**).

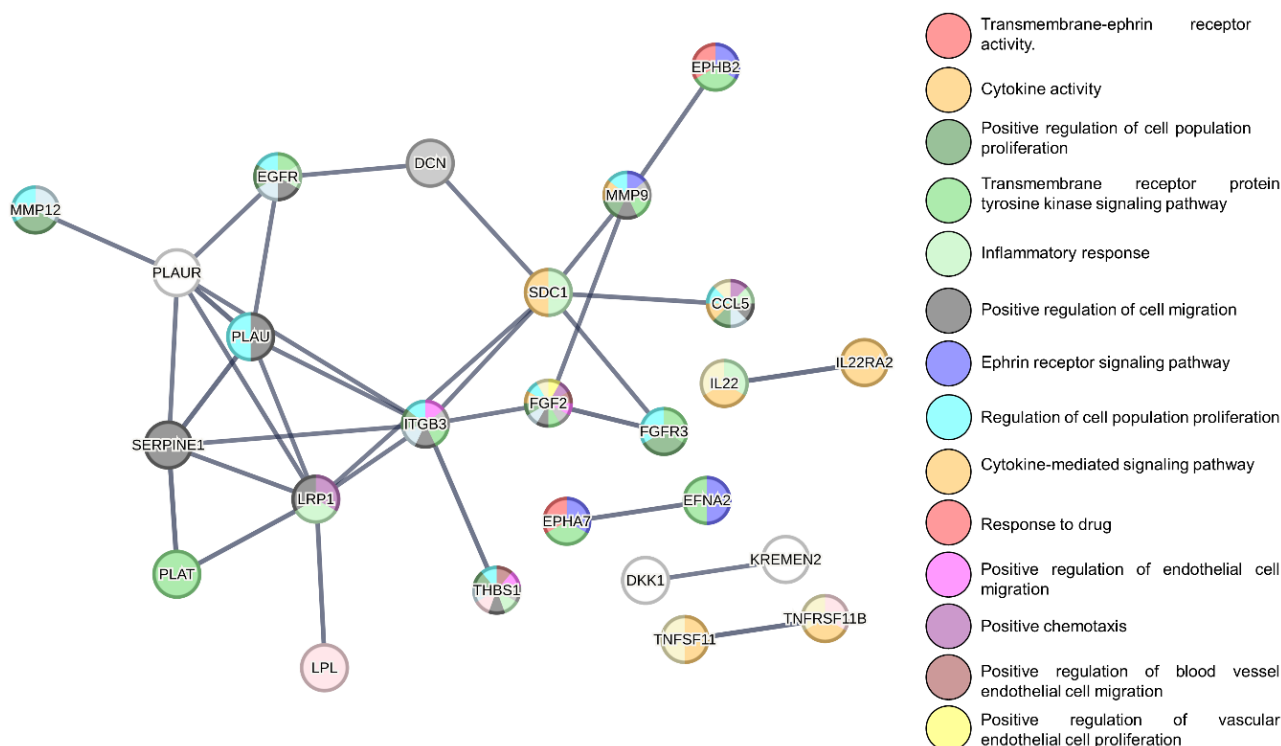


Figura 4: Enriquecimento funcional do interactoma gerado, contendo os ligantes e receptores identificados. As diferentes cores dentro dos nodos indicam o termo com o qual o gene está relacionado.

2.2.4 Os genes candidatos a biomarcadores apresentam expressão diferencial em diferentes populações celulares do intestino grosso.

Para verificar com maior especificidade a expressão dos genes candidatos a biomarcadores, utilizamos dados de scRNA-Seq, oriundos do estudo de Lee et.al 2020. Ao realizarmos o processamento dos dados, ficou evidente que PLAU apresenta expressão aumentada em células intestinais positivas para o marcador DCLK1⁺. Tal marcador é descrito como associado à progressão tumoral do CRC, ao participar de processos como transição epitélio mesênquima, bem como ao processo inflamatório atrelado à carcinogênese. Relacionado a este, temos justamente que EGFR apresentou expressão aumentada em Monócitos, que são descritos como associados à células DCLK1⁺, ocasionando a perda de função de células CD8⁺, e assim suprimindo a resposta imune (CHHETRI *et al.*, 2022). MMP9 apresentou expressão exacerbada em células de Paneth, enquanto FGF2 não apresentou detecção nos diferentes tipos celulares, corroborando o visto na Figura 5-D (**Figura 6**). Os dados contendo a porcentagem de expressão e nível estão presentes nas Figuras Suplementares 2 e 3.

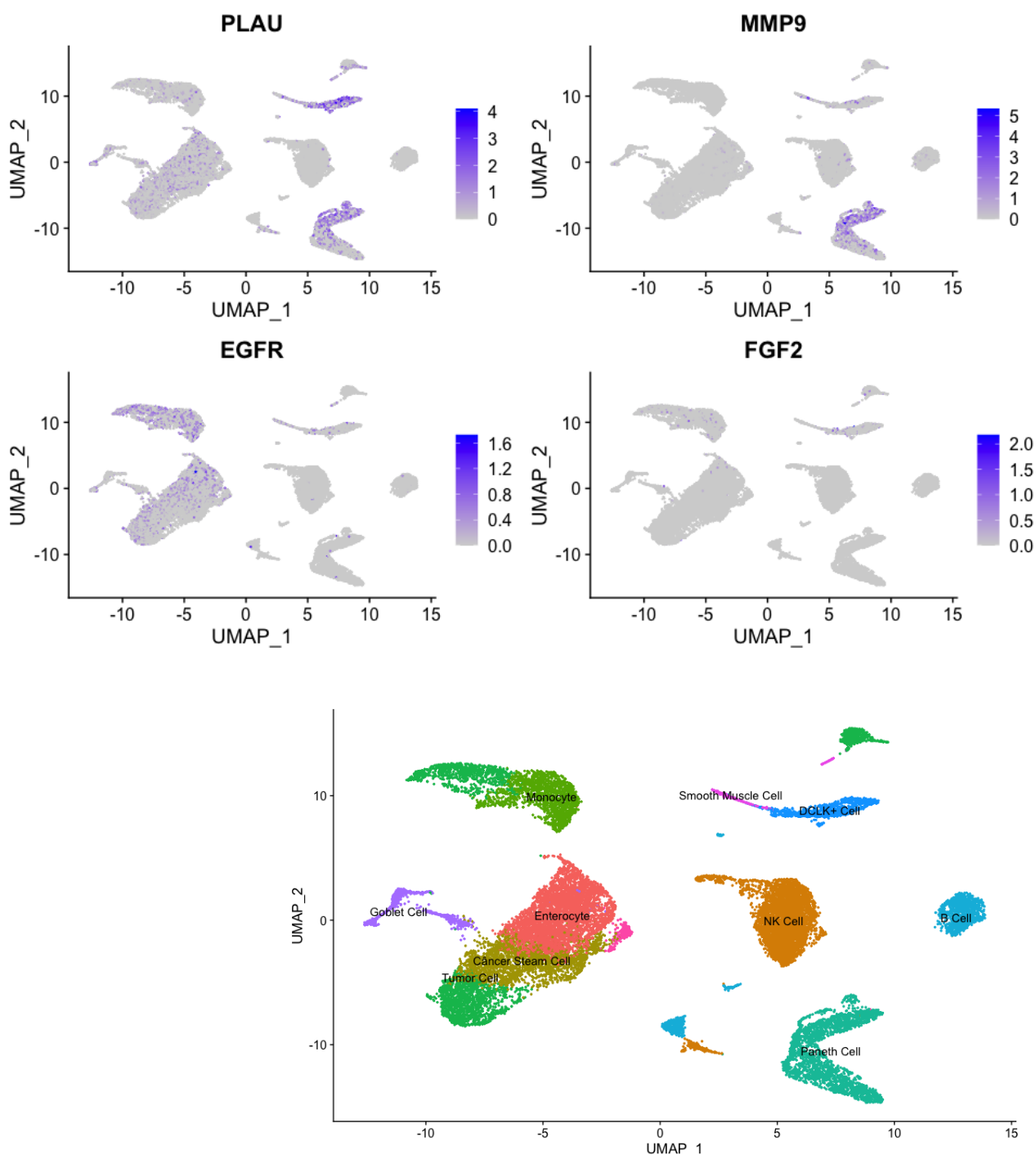
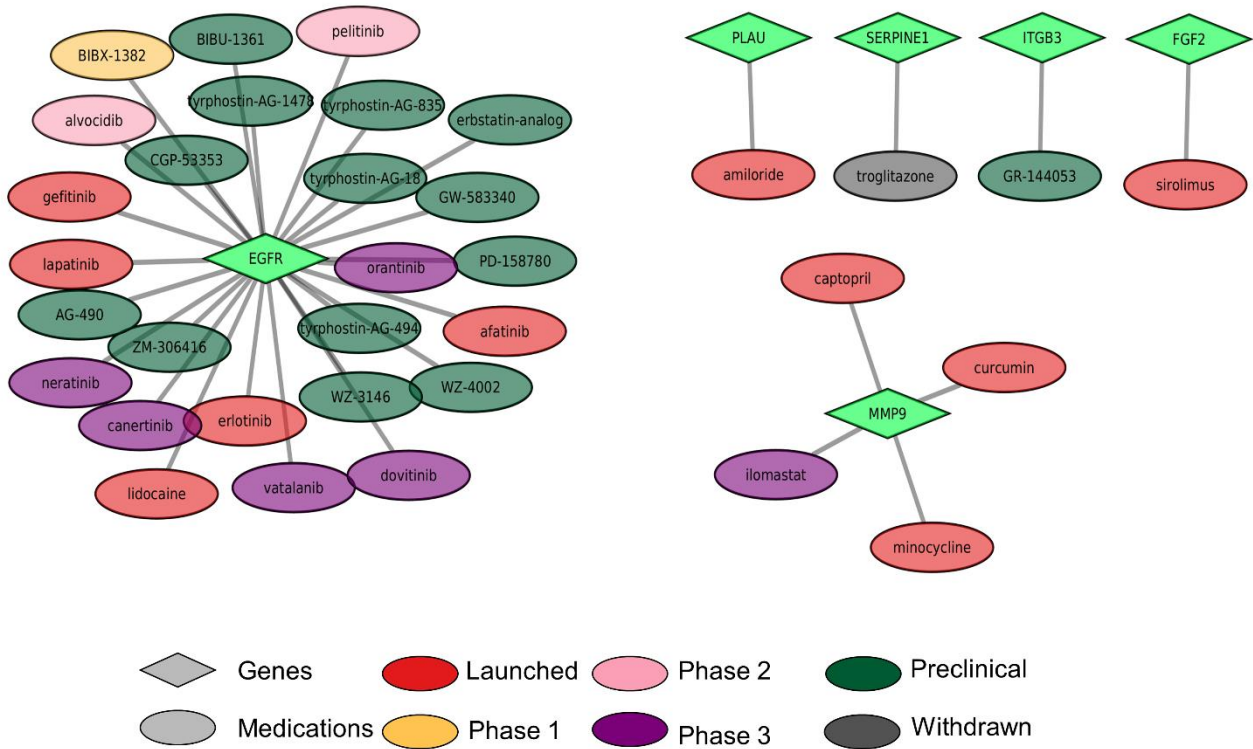


Figura 6: UMAP plot, representando a distribuição dos genes identificados como potenciais biomarcadores nos diferentes tipos celulares identificados nos dados de scRNA-Seq. A intensidade da marcação azul nas diferentes populações celulares indica a expressão do gene destacado naquela população. Abaixo os 10 tipos celulares, destacados em cores distintas. As análises foram realizadas utilizando o pacote Seurat (v.4.0.6), baseado na linguagem de programação R.

2.2.5 Os genes com expressão diferencial qualitativa e quantitativa possuem interações descritas com fármacos

Para a detecção da interação entre as assinaturas genéticas de *COAD* e *READ*, e a interação destas com drogas, foi inicialmente utilizada a plataforma OCTAD. As listas obtidas para ambos os tipos tumorais foram iguais (**Tabela Suplementar 3**). A partir disso, foram buscados os 10 genes detectados no ranqueamento por *degree*, na qual 6 (EGFR, MMP9, PLAU, SERPINE1, ITGB3 e FGF2) apresentavam interações com fármacos (**Figura 6A**). Contudo, a lista dos 10 genes foi submetida à plataforma DGIdb (**Tabela Suplementar 4**) e foram selecionados apenas os genes que tinham interações em comum com as drogas descritas. Logo, obtivemos uma rede de interação entre drogas e genes (**Figura 6B**), na qual foram descritos 4 genes (EGFR, FGF2, MMP9 e PLAU), que apresentavam ligações comuns com 6 fármacos (Curcumina, Sirolimus, Cisplatina, Geldamicina, Celecoxib e Bevacizumab).

A



B

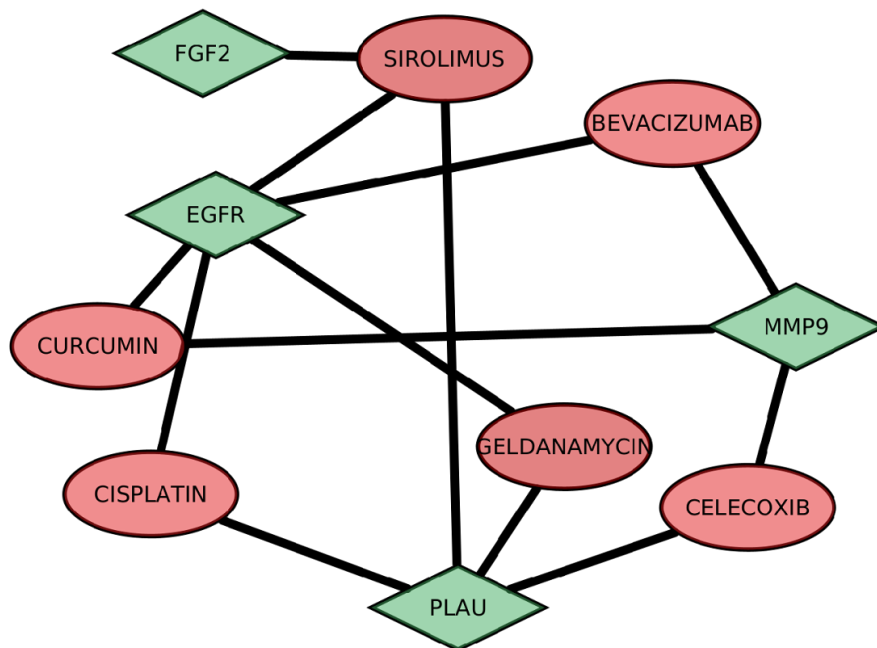


Figura 7: Interação detectada entre genes e drogas através da assinatura genética de *COAD* e *READ* (A) Interações detectadas através da plataforma OCTAD. Os losangos indicam os genes e as elipses as drogas que apresentaram interações. As cores das elipses refletem a etapa de teste na qual a droga se encontra. (B) Genes que compartilham interações com drogas entre si. Os losangos indicam os genes e as elipses as drogas.

2.3. Discussão

O nosso estudo utilizou a integração das listas do secretoma e do conectoma a fim de buscar biomarcadores prognósticos e terapêuticos no CRC. Para isso, utilizamos dados originados das plataformas GTEx e TCGA, verificando se existem alterações nos genes do secretoma humano que são ligantes e/ou receptores. Foram identificados 3 genes como potenciais biomarcadores, PLAU, EGFR e MMP9. Posteriormente, foram utilizados dados de scRNA-Seq para verificar em quais células específicas do tecido intestinal, a expressão dos genes identificados, para enfim verificar a interação destes com fármacos, onde identificamos 6 fármacos que interagem com os genes destacados, Sirolimus, Curcumina, Cisplatina, Bevacizumab, Celecoxib e Geldanamicina.

Dos genes encontrados, PLAU apresentou expressão exacerbada em células DCLK1+ presentes no cólon. Tal gene sintetiza a *Double Cortin Like Kinase*, proteína da família quinase amplamente descrita como um marcador tumoral em diversos tipos de cânceres (CHHETRI et al., 2022). No CRC, este marcador está atrelado com processos de imunomodulação, estando diretamente associados com a atração de macrófagos associados ao tumor, que auxiliam na degradação de matriz, além de criar um ambiente de imunossupressão (CAO et al., 2020; KITAMURA; QIAN; POLLARD, 2015).

Enfatizando o gene PLAU, este já é amplamente descrito como um possível fator de risco pela *American Society of Clinical Oncology* (ASCO), ocasionando pior prognóstico, sendo descrito pela mesma como alvo terapêutico (DING et al., 2022). Isso se deve ao fato de PLAU estar associado a fatores críticos da progressão tumoral, como invasão, migração e consequente metástase tumoral, amplamente descritos para o CRC (MAURO et al., 2017). Tal padrão foi igualmente relatado em linhagens deste mesmo tumor, onde a supressão de PLAU, através do MIR-193a-3p, ocasionou menor migração, invasão e angiogênese em células HT-29, HCT116 e LOVO (LIN et al., 2019). Outro fator atrelado a PLAU é a modulação do microambiente tumoral, fazendo com que fibroblastos adquiram um fenótipo maligno e trabalhem a favor da progressão do tumor (FANG et al., 2021). Contudo, uma ação modulatória fundamental do PLAU no microambiente tumoral, é na atividade de macrófagos associados ao câncer, inclusive no CRC, o que leva a uma imunossupressão da resposta imune específica, para um processo inflamatório pró tumoral (CUI et al., 2021).

Outro DEG encontrado em nossos resultados é a MMP9, descrita como uma proteína responsável pela degradação da matriz extracelular, mais precisamente uma gelatinase, atuando

principalmente na remodelação da membrana basal em processos patológicos, degradando colágenos do tipo IV, V, VII, X e XIV. Contudo, na condição patológica tumoral, a MMP9 atua contribuindo na progressão de diversos tipos tumorais, tais como mama, pulmão e colorretal (GOBIN et al., 2019; MONDAL et al., 2020). A MMP9, é descrita como proteína participante de processos chave da progressão, como metástase, invasão e angiogênese, visto que para que tais processos ocorram, há a necessidade de degradação da membrana basal, principalmente do colágeno IV (MONDAL et al., 2020). Contudo um achado importante, é que a liberação de MMP9 no CRC, depende da presença de neutrófilos associados ao tumor, demonstrando novamente o papel da imunidade inespecífica no processo de progressão deste tumor, como visto anteriormente com PLAUI (MIZUNO et al., 2019). Ainda, Moirangthem; Bondhopadhyay; Mukherjee, 2016, demonstraram que a inibição conjunta de PLAUI e MMP9 reduziu a progressão do tumor de mama ao gerar maior adesão célula-célula, bem como ao modular genes envolvidos na transição epitélio mesênquima.

O gene EGFR é amplamente descrito como envolvido na progressão do CRC, pois este modula diferenciação e proliferação celular. Inclusive, a quimioterapia padrão utilizada para este tumor, depende da combinação do 5-FU com diversos fármacos, dentre eles inibidores do EGFR (MODEST; PANT; SARTORE-BIANCHI, 2019). Tais inibidores são descritos por aumentar a permeabilidade tumoral a células T-CD8, e auxiliar no combate ao tumor. Contudo, já são descritos processos de quimiorresistência aos tratamentos baseados na inibição do EGFR (ZHOU; JI; LI, 2021). Assim, a busca por novos fármacos que atuem em um gene que possui tão crucial é fundamental. Nossos resultados demonstraram que além do Bevacizumab, que já é descrito no tratamento do CRC (XIE, Y. H.; CHEN; FANG, 2020), outras moléculas como Cisplatina, Curcumina, Sirolimus e Geldanamycin podem ser utilizadas como tratamento combinado, ou adjuvante para a quebra desse processo de resistência.

A respeito das novas drogas descritas em nosso trabalho, a curcumina tem recebido bastante atenção no tratamento de diversos tumores. Esse destaque se deve ao fato deste ser um composto natural, que demonstrou ter atividades antitumorais diversas, inclusive no CRC, onde já é demonstrado que a curcumina apresentou efeito antiproliferativo, anti-angiogênico e consequente antitumoral (CALIBASI-KOCAL et al., 2019). Além disso, no CRC a curcumina apresentou atuar modulando os principais alvos do tratamento deste tumor, VEGF e EGFR, o que acentua o efeito anti-angiogênico, bem como atuar na quebra da resistência aos tratamentos anti- EGFR (PRICCI et al., 2020; RUIZ DE PORRAS; LAYOS; MARTÍNEZ-BALIBREA, 2021). Além disso, este composto atua na regulação da microbiota intestinal, o que pode ser

benéfico no combate ao CRC, visto que a microbiota possui papel fundamental, tanto na progressão quanto no combate ao tumor (JABCZYK *et al.*, 2021; WENG; GOEL, 2022; XU *et al.*, 2018). Por fim há de se destacar, que já é demonstrada a ação da curcumina inibindo a invasão do CRC, através da modulação de PLA2 e MMP, logo há evidências relacionando o composto e genes propostos como biomarcadores neste estudo. (TONG *et al.*, 2016).

O Celecoxib, outro composto destacado em nossos resultados, trata-se de um inibidor específico da Ciclo Oxigenase 2 (COX-2), inicialmente prescrito como um fármaco anti-inflamatório. Contudo já é descrita atividade antitumoral associada ao fármaco, visto que no microambiente tumoral, o aumento de atividade da COX-2, gera aumento local de prostaglandinas, o que contribui para desbalanço imune, crescimento tumoral e angiogênese (SAXENA; SHARMA; PUROHIT, 2020). Especificamente falando sobre a ação deste no CRC, já é demonstrado que a inibição de COX-2 ocasiona melhor prognóstico e representa um alvo chave na progressão deste tipo tumoral, visto que esta se encontra muito ativa em macrófagos associados ao tumor, levando à maior síntese de prostaglandina-2 e aumento da inflamação associada à progressão tumoral (TUYNMAN; PEPPELENBOSCH; RICHEL, 2004).

No caso da Cisplatina, trata-se de um antineoplásico já utilizado em diversos tumores sólidos, como ovário, testículo e cabeça e pescoço. Seu mecanismo de ação se baseia na formação de adutos de DNA, que interrompem a divisão celular (MAKOVEC, 2019). Contudo, a resistência ao tratamento com este fármaco é descrita para diversos tumores, dentre eles o CRC (GHOSH, 2019). Ainda assim, já ficou demonstrado que tal resistência pode ser diminuída com aumento da expressão de IRF6, através de transfecção (TAN *et al.*, 2022). Assim, verifica-se a necessidade de maiores estudos para o desenvolvimento de terapias combinadas que quebrem essa resistência.

Sirolimus, que também é conhecido como Rapamicina, é um fármaco inicialmente prescrito como um imunossupressor, mas sempre houve interesse em verificar a sua atividade antitumoral. Isso ocorre pelo fato da ação deste composto inibir mTOR, uma serina treonina quinase, cuja expressão está aumentada em diversos tumores, atuando principalmente na proliferação exacerbada celular, pois leva a maior expressão de oncogenes como PI3K e Akt, e inibição de supressores, como PTEN, LKB 1 e 2 (LAW, 2005; LI, JING; KIM; BLENIS, 2014). Pouco se sabe sobre a atuação deste no CRC, mas um estudo realizado por Fujishita *et al.*, 2008

demonstrou que a inibição de mTOR, através da Rampamicina levou ao menor desenvolvimento de pólipos em camundongos e consequente redução da proliferação tumoral.

Acerca da Geldanamicina, esta se trata de uma molécula inibidora da HSP90, uma proteína que atua em diversos processos celulares, mas principalmente na regulação da célula para sobreviver em situações de estresse, tais como alterações de pH, lesões, entre outras. Este fármaco foi inicialmente descrito como um antibiótico, contudo estudos indicam o aumento de expressão da HSP90 em diversos cânceres, fazendo com que esse se torne um possível alvo terapêutico (FRANKE et al., 2013; FUKUYO; HUNT; HORIKOSHI, 2010). No caso de CRC, estudo utilizando esta molécula demonstraram que ela é capaz de gerar apoptose nas células tumorais, através do aumento da expressão de genes pró-apoptóticos, como caspase-9, Apf-1 e citocromo-c, e decréscimo de genes anti-apoptóticos, como Bcl-2 (KOSOVA et al., 2017). Além disso são descritas interações desse fármaco com outros alvos anteriormente descritos no trabalho, como por exemplo a inibição de Cox-2 através da Geldanamicina, gerando menor produção de prostaglandina-2, em células HT-29 (MOHAMMADI et al., 2016), bem como a diminuição MMP9 e VEGF através da combinação desta molécula com o Irinotecano (ZEYNALI-MOGHADDAM et al., 2019).

Conclusão

Desta forma, o nosso trabalho demonstrou uma nova abordagem para a busca de alvos terapêuticos e prognósticos para o CRC. Através da integração de plataformas e filtros, demonstramos que MMP9, PLAU, e EGFR podem ser utilizados para detecção precoce deste tumor, bem como serem suscetíveis a drogas para tratamento. A respeito das drogas, a integração de plataforma permitiu que encontrássemos 6 drogas que possuem interações descritas com tais genes, das quais 5 não são utilizadas como tratamento padrão para este tumor. Desta forma, nosso estudo demonstra que a integração de dados pode ser um fator crucial para a descoberta de novas moléculas bioativas no tratamento do CRC, abordando uma variedade de substâncias, desde compostos naturais, representados neste estudo pela Curcumina e Geldanamicina, até mesmo por fármacos que possuem funções descritas para outras doenças, mas podem ser testados para o tratamento tumoral, na chamada medicina reversa.

Referências Bibliográficas

AKGÜL, Ö. *et al.* Role of surgery in colorectal cancer liver metastases. *World Journal of Gastroenterology*, v. 20, n. 20, p. 6113–6122, 2014.

ALESSANDRINO, F. *et al.* 5-Fluorouracil induced liver toxicity in patients with colorectal cancer: role of computed tomography texture analysis as a potential biomarker. *Abdominal Radiology*, v. 44, n. 9, p. 3099–3106, 2019. Disponível em: <<https://doi.org/10.1007/s00261-019-02110-3>>.

ARAN, D.; SIROTA, M.; BUTTE, A. J. Systematic pan-cancer analysis of tumour purity. *Nature Communications*, v. 6, p. 1–11, 2015. Disponível em: <<http://dx.doi.org/10.1038/ncomms9971>>.

ARMINGOL, E. *et al.* Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics*, v. 22, n. 2, p. 71–88, 2021. Disponível em: <<http://dx.doi.org/10.1038/s41576-020-00292-x>>.

BEKLEN, H. *et al.* Differential Interactome Based Drug Repositioning Unraveled Abacavir, Exemestane, Nortriptyline Hydrochloride, and Tolcapone as Potential Therapeutics for Colorectal Cancers. *Frontiers in Bioinformatics*, v. 1, n. September, p. 1–13, 2021.

BENSON, A. B. *et al.* Journal of the National Comprehensive Cancer Network, Colon Cancer, Version 1.2021, NCCN Clinical Practice Guidelines in Oncology. *NCCN Clinical Practice Guidelines in Oncology*, v. 19, n. 3, p. 329–359, 2021.

BILLER, L. H.; SCHRAG, D. Diagnosis and treatment of metastatic colorectal cancer: A review. *JAMA - Journal of the American Medical Association*, v. 325, n. 7, p. 669–685, 2021.

CAO, Z. *et al.* Tuft and cancer stem cell marker dcl1: A new target to enhance anti-tumor immunity in the tumor microenvironment. *Cancers*, v. 12, n. 12, p. 1–17, 2020.

CHHETRI, D. *et al.* Pleiotropic effects of DCLK1 in cancer and cancer stem cells. *Frontiers in Molecular Biosciences*, v. 9, n. September, p. 1–25, 2022.

CHIN, C. H. *et al.* cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, v. 8, n. 4, p. S11, 2014. Disponível em: <<http://www.biomedcentral.com/1752-0509/8/S4/S11>>.

CLARKE, D. J. B. *et al.* Article Appyters : Turning Jupyter Notebooks into data- driven web

apps Appyters : Turning Jupyter Notebooks into data-driven web apps. *Patterns*, v. 2, n. 3, p. 100213, 2021. Disponível em: <<https://doi.org/10.1016/j.patter.2021.100213>>.

COTTO, K. C. *et al.* DGIdb 3.0: A redesign and expansion of the drug-gene interaction database. *Nucleic Acids Research*, v. 46, n. D1, p. D1068–D1073, 2018.

CUI, D. *et al.* Identification of colorectal cancer-associated macrophage biomarkers by integrated bioinformatic analysis. *International Journal of Clinical Pathology*, v. 14, n. 1, p. 1–8, 2021.

DE OLIVEIRA, G. *et al.* An integrated meta-analysis of secretome and proteome identify potential biomarkers of pancreatic ductal adenocarcinoma. *Cancers*, v. 12, n. 3, 2020.

DEKKER, E. *et al.* Colorectal cancer. *The Lancet*, v. 394, n. 10207, p. 1467–1480, 2019.

DEPETRIS, I. *et al.* Fluoropyrimidine-induced cardiotoxicity. *Critical Reviews in Oncology/Hematology*, v. 124, n. January, p. 1–10, 2018. Disponível em: <<https://doi.org/10.1016/j.critrevonc.2018.02.002>>.

DING, Y. *et al.* Plasminogen activator , urokinase enhances the migration , invasion , and proliferation of colorectal cancer cells by activating the Src / ERK pathway. *Journal of Gastrointestinal Oncology*, v. 13, n. 6, p. 3100–3111, 2022.

DONCHEVA, N. T. *et al.* Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research*, v. 18, n. 2, p. 623–632, 2019.

DUMOULIN, F. L.; HILDENBRAND, R. Endoscopic resection techniques for colorectal neoplasia: Current developments. *World Journal of Gastroenterology*, v. 25, n. 3, p. 300–307, 2019.

EISTERER, W.; PRAGER, G. Chemotherapy, Still an Option in the Twenty-First Century in Metastatic Colorectal Cancer? *CardioVascular and Interventional Radiology*, v. 42, n. 9, p. 1213–1220, 2019. Disponível em: <<https://doi.org/10.1007/s00270-019-02278-7>>.

FANG, L. *et al.* PLAU directs conversion of fibroblasts to inflammatory cancer-associated fibroblasts , promoting esophageal squamous cell carcinoma progression via uPAR / Akt / NF- κ B / IL8 pathway. *Cell Death Discovery*, 2021. Disponível em: <<http://dx.doi.org/10.1038/s41420-021-00410-6>>.

FASANO, C. *et al.* Short Linear Motifs in Colorectal Cancer Interactome and Tumorigenesis.

Cells, v. 11, n. 23, p. 1–23, 2022.

FRIEDENREICH, C. M.; RYDER-BURBIDGE, C.; MCNEIL, J. Physical activity, obesity and sedentary behavior in cancer etiology: epidemiologic evidence and biologic mechanisms. *Molecular Oncology*, v. 15, n. 3, p. 790–800, 2021.

GOBIN, E. *et al.* A pan-cancer perspective of matrix metalloproteases (MMP) gene expression profile and their diagnostic / prognostic potential. *BMC Cancer*, v. 19, n. 581, p. 1–10, 2019.

GOEDHART, J. VolcanoR is a web app for creating , exploring , labeling and sharing volcano plots. *Scientific Reports*, v. 10, p. 1–5, 2020. Disponível em: <<https://doi.org/10.1038/s41598-020-76603-3>>.

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

HAO, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell*, v. 184, n. 13, p. 3573–3587.e29, 2021. Disponível em: <<https://doi.org/10.1016/j.cell.2021.04.048>>.

HU, C. *et al.* CellMarker 2.0 : an updated database of manually curated cell markers in human / mouse and web tools based on scRNA-seq data. *Nucleic Acids Research*, v. 51, n. October 2022, p. 870–876, 2023.

INCA. *Estimativa 2020 Incidência de Câncer no Brasil*. Rio de Janeiro: INCA, 2019.

JAMES, K.; WIPAT, A.; COCKELL, S. J. Expanding interactome analyses beyond model eukaryotes. *Briefings in Functional Genomics*, v. 21, n. 4, p. 243–269, 27 jul. 2022. Disponível em: <<https://academic.oup.com/bfg/article/21/4/243/6583213>>.

JEAN-QUARTIER, C. *et al.* In silico cancer research towards 3R. *BMC Cancer*, v. 18, n. 1, p. 1–12, 2018.

JOHDI, N. A.; SUKOR, N. F. Colorectal Cancer Immunotherapy: Options and Strategies. *Frontiers in Immunology*, v. 11, n. September, p. 1–18, 2020.

KALKAT, M. *et al.* MYC Protein Interactome Profiling Reveals Functionally Distinct Regions that Cooperate to Drive Tumorigenesis. *Molecular Cell*, v. 72, n. 5, p. 836– 848.e7, 2018. Disponível em: <<https://doi.org/10.1016/j.molcel.2018.09.031>>.

KITAMURA, T.; QIAN, B.-Z.; POLLARD, J. W. Immune cell promotion of metastasis. *Nat Rev Immunol*, v. 15, n. 2, p. 73–86, 2015. Disponível em:

<<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4470277/pdf/nihms693262.pdf>>.

LA VECCHIA, S.; SEBASTIÁN, C. Metabolic pathways regulating colorectal cancer initiation and progression. *Seminars in Cell and Developmental Biology*, v. 98, n. May 2019, p. 63–70, 2020. Disponível em: <<https://doi.org/10.1016/j.semcdb.2019.05.018>>.

LEE, H. O. *et al.* Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nature Genetics*, v. 52, n. 6, p. 594–603, 2020. Disponível em: <<http://dx.doi.org/10.1038/s41588-020-0636-z>>.

LEE, J. J. *et al.* Elucidation of tumor-stromal heterogeneity and the ligand-receptor interactome by single-cell transcriptomics in real-world pancreatic cancer biopsies. *Clinical Cancer Research*, v. 27, n. 21, p. 5912–5921, 2021.

LEMBERGER, T. Systems biology in human health and disease. *Molecular Systems Biology*, v. 3, n. 136, p. 1–2, 2007.

LI, J. *et al.* Genetic and biological hallmarks of colorectal cancer. *Genes and Development*, v. 38, n. 11–12, p. 787–820, 2021.

LI, K. *et al.* Bioinformatics Approaches for Anti-cancer Drug Discovery. *Current Drug Targets*, v. 21, n. 1, p. 3–17, 2019.

LIN, M. *et al.* MicroRNA-193a-3p suppresses the colorectal cancer cell proliferation and progression through downregulating the PLA2 expression. *Cancer Management and Research*, v. 11, p. 5353–5363, 2019.

MARIN, J. J. G. *et al.* Chemoprevention, chemotherapy, and chemoresistance in colorectal cancer. *Drug Metabolism Reviews*, v. 44, n. 2, p. 148–172, 2012.

MARQUES, N. P. *et al.* Cancer diagnosis in Brazil in the COVID-19 era. *Seminars in Oncology*, v. 48, n. 2, p. 156–159, abr. 2021. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0093775420301238>>.

MASDOR, N. A. *et al.* The Link between Food Environment and Colorectal Cancer: A Systematic Review. *Nutrients*, v. 14, n. 19, p. 3954, 23 set. 2022. Disponível em: <<https://www.mdpi.com/2072-6643/14/19/3954>>.

MAURO, C. DI *et al.* Urokinase-type plasminogen activator receptor (uPAR) expression enhances invasion and metastasis in RAS mutated tumors. *Scientific Reports*, v. 7, n. March, p.

1–12, 2017.

MIZUNO, R. *et al.* The Role of Tumor-Associated Neutrophils in Colorectal Cancer. *International Journal of Molecular Sciences*, v. 20, n. 529, p. 1–14, 2019.

MODEST, D. P.; PANT, S.; SARTORE-BIANCHI, A. Treatment sequencing in metastatic colorectal cancer. *European Journal of Cancer*, v. 109, p. 70–83, 2019. Disponível em: <<https://doi.org/10.1016/j.ejca.2018.12.019>>.

MOIRANGTHEM, A.; BONDHOPADHYAY, B.; MUKHERJEE, M. Simultaneous knockdown of uPA and MMP9 can reduce breast cancer progression by increasing cell-cell adhesion and modulating EMT genes. *Nature Publishing Group*, v. 6, n. January, p. 1–11, 2016. Disponível em: <<http://dx.doi.org/10.1038/srep21903>>.

MONDAL, S. *et al.* Matrix metalloproteinase-9 (MMP-9) and its inhibitors in cancer : A minireview. *European Journal of Medicinal Chemistry*, v. 194, p. 112260, 2020. Disponível em: <<https://doi.org/10.1016/j.ejmech.2020.112260>>.

NIEDERREITER, L.; ADOLPH, T. E.; TILG, H. Food, microbiome and colorectal cancer. *Digestive and Liver Disease*, v. 50, n. 7, p. 647–652, 2018. Disponível em: <<https://doi.org/10.1016/j.dld.2018.03.030>>.

NUSSINOV, R.; TSAI, C. J.; JANG, H. Anticancer drug resistance: An update and perspective. *Drug Resistance Updates*, v. 59, 2021.

POLASTRO, L.; EL HACHEM, G.; HENDLISZ, A. Pseudoadjuvant chemotherapy in resectable metastatic colorectal cancer. *Current Opinion in Oncology*, v. 30, n. 4, p. 269–275, 2018.

RAMILOWSKI, J. A. *et al.* A draft network of ligand-receptor-mediated multicellular signalling in human. *Nature Communications*, v. 6, 2015.

RAWLA, P.; SUNKARA, T.; BARSOUK, A. Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Przegląd Gastroenterologiczny*, v. 14, n. 2, p. 89–103, 2019.

SHANNON, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, v. 13, n. 11, p. 2498–2504, nov. 2003. Disponível em: <<http://ci.nii.ac.jp/naid/110001910481/>>.

- SIEGEL, R. L. *et al.* Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, v. 72, n. 1, p. 7–33, 2022.
- SIMON, K. Colorectal cancer development and advances in screening. *Clinical Interventions in Aging*, v. 11, p. 967–976, 2016.
- SZKLARCZYK, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, v. 49, n. D1, p. D605–D612, 2021.
- THANIKACHALAM, K.; KHAN, G. Colorectal cancer and nutrition. *Nutrients*, v. 11, n. 1, 2019.
- TSIMBERIDOU, A. M. *et al.* Review of precision cancer medicine: Evolution of the treatment paradigm. *Cancer Treatment Reviews*, v. 86, p. 1–24, 2020.
- UHLÉN, M. *et al.* The human secretome. *Science Signaling*, v. 12, n. 609, p. 1–9, 2019.
- UHLÉN, M. *et al.* Tissue-based map of the human proteome. *Science*, v. 347, n. 6220, 2015.
- VAN CUTSEM, E. *et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Annals of Oncology*, v. 27, n. 8, p. 1386–1422, 2016.
- WEI, L. *et al.* The emerging role of noncoding RNAs in colorectal cancer chemoresistance. *Cellular Oncology*, v. 42, n. 6, p. 757–768, 2019.
- XIE, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols*, v. 1, n. 3, p. 1–51, 2021.
- YIN, Y. *et al.* The immune-microenvironment confers chemoresistance of colorectal cancer through macrophage-derived IL6. *Clinical Cancer Research*, v. 23, n. 23, p. 7375–7387, 2017.
- YU, T. *et al.* *Fusobacterium nucleatum* Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell*, v. 170, n. 3, p. 548–563, 2017.
- ZAMANIAN-AZODI, M.; REZAEI-TAVIRANI, M. Investigation of health benefits of cocoa in human colorectal cancer cell line, HT-29 through interactome analysis. *Gastroenterology and Hepatology from Bed to Bench*, v. 12, n. 1, p. 67–73, 2019.
- ZENG, B. *et al.* OCTAD: an open workspace for virtually screening therapeutics targeting precise cancer patient groups using gene expression features. *Nature Protocols*, v. 16, n. 2, p.

728–753, 23 fev. 2021. Disponível em: <<http://www.nature.com/articles/s41596-020-00430-z>>.

ZHENG, H. *et al.* Comprehensive Review of Web Servers and Bioinformatics Tools for Cancer Prognosis Analysis. *Frontiers in Oncology*, v. 10, n. February, 2020.

Apêndices

Figuras Suplementares

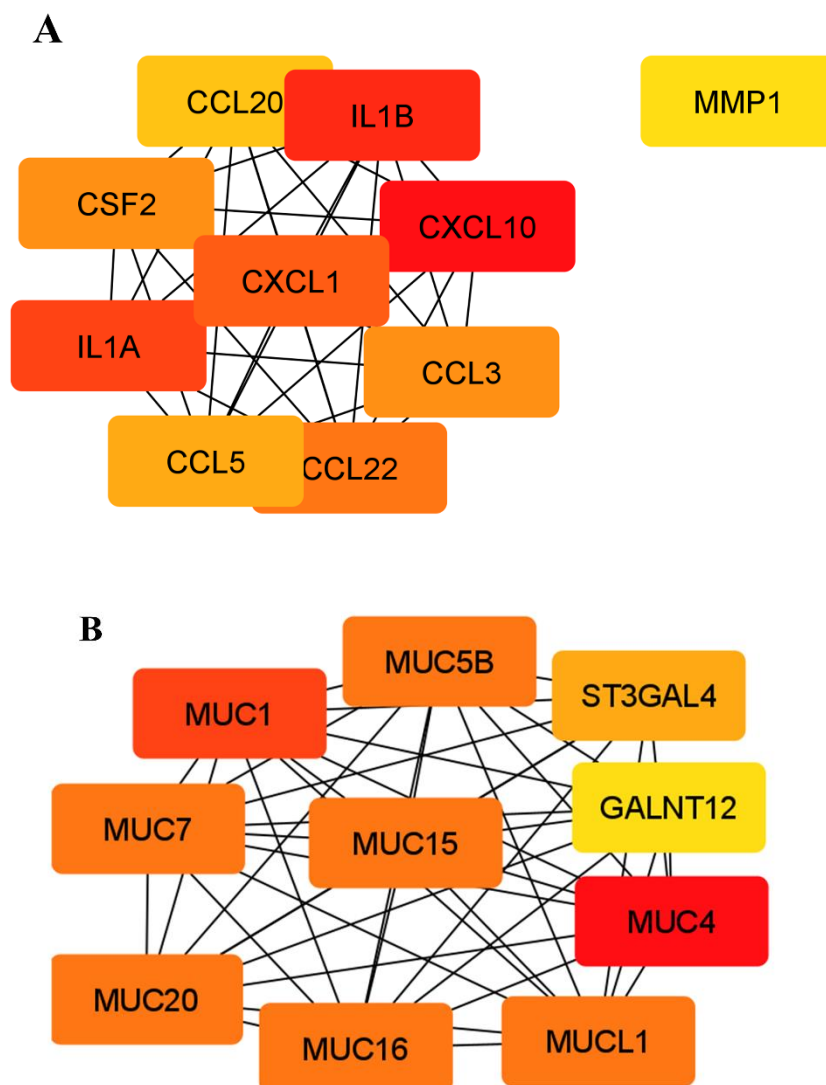


Figura Suplementar 1: Redes de interação proteína/proteína obtidas a partir dos genes up regulados ranqueados por *degree* em cólon (A) e reto (B).

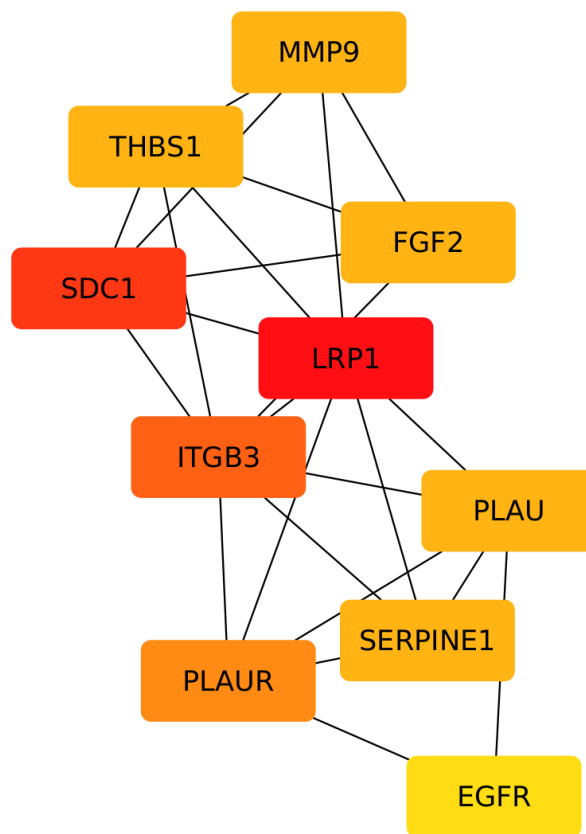


Figura Suplementar 2: Top 10 genes que apresentam o maior número de interações, clusterizados pelo método *degree*. As cores são relativas as posições destes genes no rank.

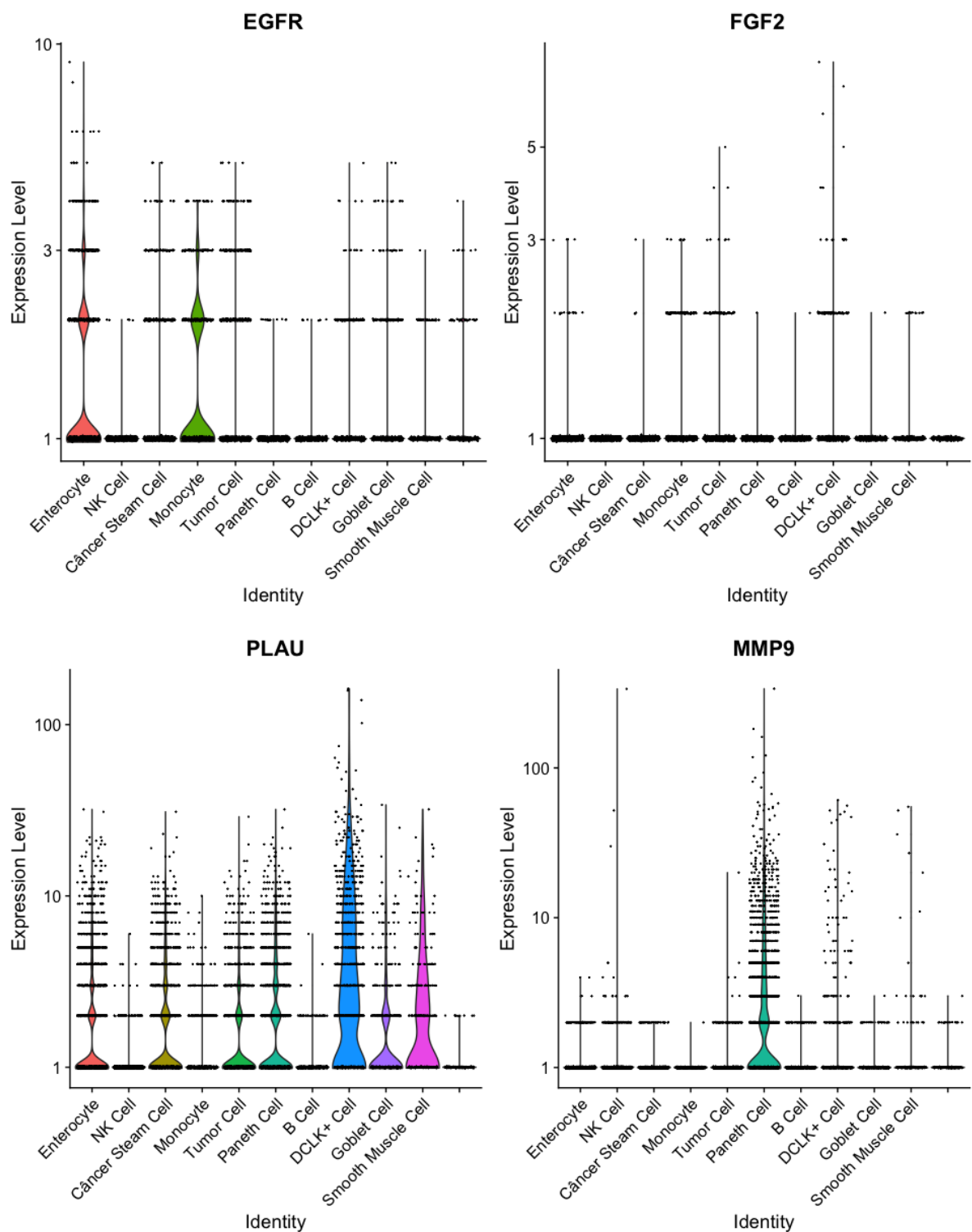


Figura Suplementar 3: Violin plots demonstrando os níveis de expressão nos diferentes tipos celulares dos genes elencados como potenciais biomarcadores prognósticos e diagnósticos do CRC, através da análise de scRNA-Seq.

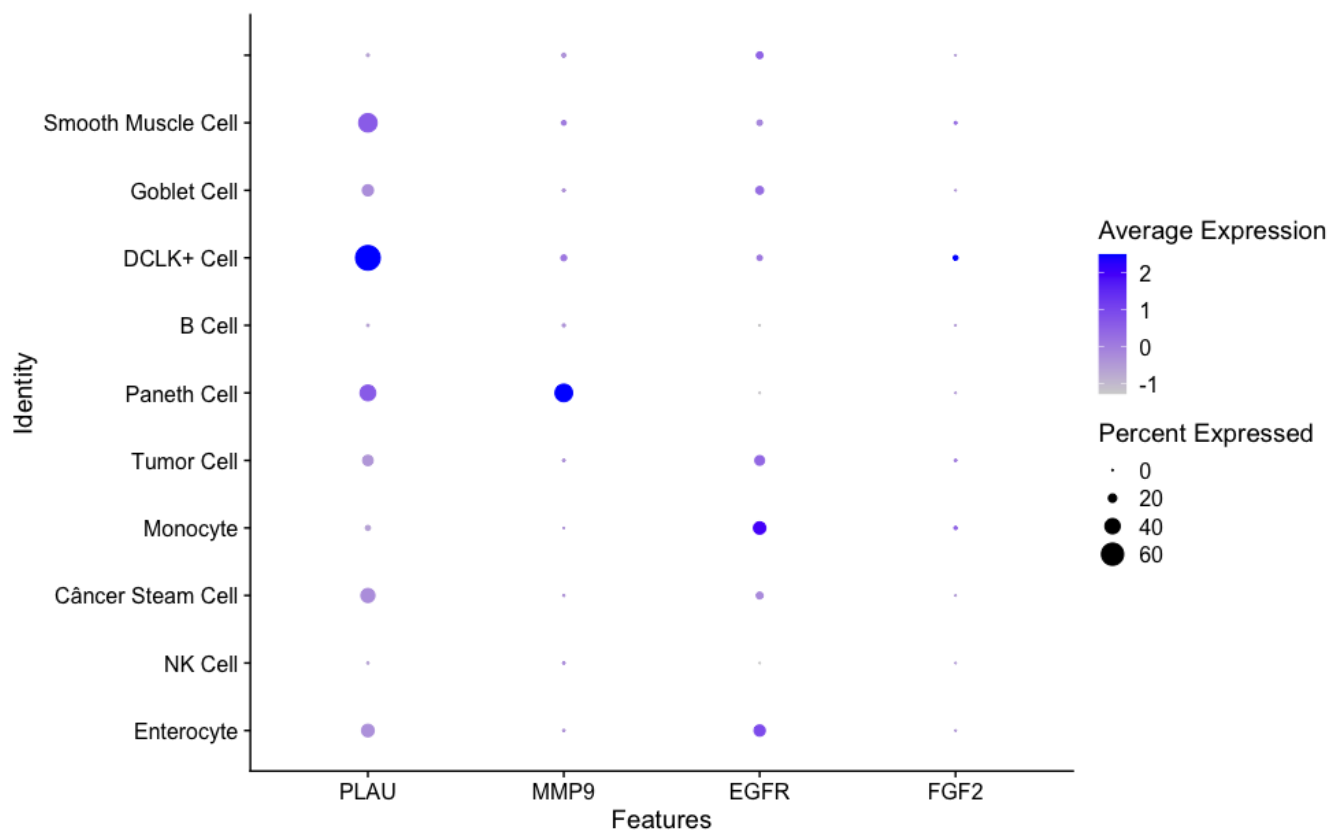


Figura Suplementar 4: *Dotplot* indicando a expressão dos genes candidatos a biomarcadores diagnósticos e prognósticos do CRC, nos diferentes tipos celulares identificados na análise de dados de scRNA-Seq. A intensidade da cor e o tamanho da bola são, respectivamente, a expressão média de cada gene e o percentual expresso em cada tipo celular.