

Universidade Estadual Paulista Júlio de Mendonça Filho

Caio Fernando Ferreira Mussatto

**Análises Transcriptômicas em Células Únicas de Fatores Indutores de  
Caquexia em linhagens Celulares de Câncer**

Botucatu

2023

Caio Fernando Ferreiro Mussatto

**Análises Transcriptômicas em Células Únicas de Fatores Indutores de  
Caquexia em linhagens Celulares de Câncer**

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do título de Bacharel, junto ao Curso de Graduação em Ciências Biológicas, do Instituto de Biociências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Botucatu.

Orientador: Prof. Dr. Robson Francisco Carvalho

Botucatu

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 A. CRUZ E SANTOS-CRB 8/10188

Mussatto, Caio Fernando Ferreira.

Análises transcriptômicas em células únicas de fatores indutores de caquexia em linhagens celulares de câncer / Caio Fernando Ferreira Mussatto. - Botucatu, 2023

Trabalho de conclusão de curso (bacharelado - Ciências Biológicas) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Robson Francisco Carvalho

Capes: 20205007

1. Caquexia. 2. Linhagem Celular. 3. Análise da Expressão Gênica de Célula Única.

Palavras-chave: CIFS; Caquexia; Linhagens celulares; scRNA-seq.

## Resumo

A caquexia, uma síndrome multifatorial caracterizada pela perda contínua de massa muscular esquelética, representa um desafio significativo na gestão do câncer, uma vez que afeta aproximadamente 80% dos pacientes com câncer avançado e tem impacto direto na sobrevivência. Este estudo aborda a necessidade premente de selecionar linhagens celulares apropriadas para investigações *in vitro* e pesquisa pré-clínica relacionadas à caquexia.

Utilizando ferramentas de mineração de texto, Geneshot e Open Targets Platform, foram identificados 123 fatores indutores de caquexia (CIFs). Em seguida, sequenciamentos de células únicas (Single Cell RNA-seq) de 198 linhagens celulares foram obtidos, onde, usando o Celligner, selecionamos 84 linhagens celulares que possuíam uma maior correlação com tumores primários. Em seguida, utilizando a linguagem R e o pacote de ferramentas, Seurat, realizamos a caracterização dos 101 CIFs presente nas linhagens. Em seguida, realizamos uma análise de correlação de Pearson, entre 10 tipos tumorais e 54 linhagens (linhagens dos mesmos tipos tumorais previamente selecionados), usando a expressão média escalada de 89 CIFs em comum, o que permitiu a identificação de 10 linhagens celulares que melhor recapitulam a caquexia, sendo uma para cada tipo tumoral.

Os dados de RNA-seq revelaram 10 linhagens celulares altamente correlacionadas com a expressão dos CIFs em vários tipos de câncer, incluindo ovário, pulmão, nasofaringe, rim, cólon e mama. Essas linhagens surgiram como candidatas promissoras para estudos futuros devido as correlações estatisticamente significativas. A seleção criteriosa de linhagens celulares é de extrema importância para estudos *in vitro* sobre a caquexia associada ao câncer, uma vez que uma escolha inadequada pode levar a resultados insatisfatórios. Estudos anteriores destacaram a necessidade de aprimorar os métodos de seleção para garantir a representatividade das linhagens em relação aos tumores de interesse. O enfoque na expressão gênica dos CIFs mostrou ser uma estratégia valiosa na identificação das linhagens celulares mais relevantes. Este estudo contribui para o entendimento e tratamento da caquexia associada ao câncer, fornecendo uma base sólida para futuras pesquisas. A identificação de linhagens celulares representativas e a caracterização de genes associados à caquexia abrem caminho para abordagens terapêuticas mais direcionadas e avanços significativos no controle da síndrome.

## Sumário

1. Introdução.....	7
2. Metodologia.....	8
3. Resultados.....	9
4. Discussão.....	12
5. Conclusão .....	14
6. Referências Bibliográficas.....	14

## **Agradecimentos**

Agradeço ao meu orientador Prof<sup>o</sup> Dr<sup>o</sup> Robson Francisco Carvalho pelo enorme apoio, incentivo e dedicação ao meu trabalho desde o início em que entrei no laboratório.

Também agradeço aos meus colegas do CacaoLab, graças as vocês os dias se tornam mais leves e o trabalho mais prazeroso. Aos meus queridos pais, Antonio e Roseney, que sempre estiveram ao meu lado me apoiando, mesmo com uma distância de 200 quilômetros. Ao meu irmão, Diego, que foi a minha companhia durante essa jornada, a sua presença tranquilizou os momentos difíceis e melhorou os momentos bons. À UNESP, ao Departamento de Biologia funcional e estrutural, mais especificamente o setor da fisiologia de Botucatu.

Por fim, mas não menos importante, agradeço ao programa de permanência do Instituto de Biociências de Botucatu (IBB) e ao Programa Institucional de Bolsa de Iniciação Científica (PIBIC), sem o auxílio de vocês, a graduação não teria sido possível

## 1. Introdução

A caquexia é uma síndrome multifatorial caracterizada por uma perda contínua de massa muscular esquelética, com ou sem perda de tecido adiposo<sup>1</sup>, que impacta a tolerância aos tratamentos radio- e quimioterápicos e reduz a sobrevida dos pacientes com câncer<sup>2</sup>. A síndrome afeta aproximadamente 80% dos pacientes com câncer em estádios avançados, e representa 20% de todas as mortes pela doença<sup>3</sup>. A caquexia é comumente associada aos cânceres de pâncreas, esôfago, estômago, pulmão e fígado; porém, pacientes com câncer de tireoide, melanoma, mama ou próstata têm menor risco de desenvolver a síndrome<sup>4,5</sup>.

O conjunto de proteínas que são secretadas pelas células do microambiente tumoral – também conhecido como secretoma tumoral<sup>6</sup> – é responsável por um estado pró-inflamatório sistêmico que contribui para o desenvolvimento da caquexia<sup>7</sup>. Nosso grupo de pesquisa caracterizou o perfil de expressão de genes do secretoma em 12 tipos de cânceres (4651 amostras) em comparação com seus tecidos normais correspondentes (2737 amostras)<sup>8</sup>. Esse estudo identificou que, dentre os componentes do secretoma tumoral, os fatores indutores de caquexia (CIFs, do inglês, *cachexia inducing factors*) apresentam um perfil de expressão gênica específico para cada tipo de câncer. Nesse perfil de expressão gênica, o número de genes de CIFs regulados positivamente está correlacionado com a prevalência de caquexia e perda de peso associado com cada tipo tumoral.

Embora o perfil de expressão gênica de CIFs seja capaz de explicar por que tipos específicos de cânceres são mais propensos a desenvolver caquexia<sup>8</sup>, o microambiente tumoral é caracterizado por uma alta de diversidade celular e por interações e comunicações celulares induzidas pelas células cancerosas<sup>9</sup>. Nesse sentido, linhagens de células cancerosas têm sido utilizadas como importantes modelos “*in vitro*” para estudo da biologia tumoral<sup>10,11</sup>. Para selecionar linhagens celulares mais correlacionadas com amostras de tumor primário, Yu *et al.*<sup>12</sup>, através de um estudo pan-câncer, ou seja, estudo que analisa diversos tumores, realizaram uma comparação dos perfis transcriptômicos de linhagens celulares e amostras de tumores primários dentre 22 tipos tumorais disponíveis nos bancos de dados TCGA (*The Cancer Genome Atlas*)<sup>13</sup> e CCLE (*Cancer Cell Line Encyclopedia*)<sup>10,11</sup>. Essa comparação definiu um novo painel de linhagens celulares (TCGA-110-CL) para estudos abrangentes utilizando diferentes tipos celulares (estudos “*Pan-Cancer*”)<sup>12</sup>. Warren *et al.*<sup>14</sup>, comparando perfis transcriptômicos, com dados obtidos no TCGA, CCLE, TARGET (Therapeutically Applicable Research to Generate Effective Treatments) e Treehouse<sup>10,11,14,15</sup>, desenvolveu a ferramenta online Celligner, utilizando o sequenciamento de RNA, que indica quais são as linhagens celulares que apresentam a maior correlação com tumores primários, a partir dos dados transcritômicos<sup>14</sup>.

Recentemente, Kinker *et al.*<sup>16</sup>, utilizaram o sequenciamento de RNA de células-únicas (scRNA-Seq) - uma ferramenta valiosa para estudar a heterogeneidade celular do microambiente tumoral - para definir as linhagens celulares do CCLE que melhor recapitulam os programas de heterogeneidade intra-tumoral<sup>16</sup>. Portanto, uma caracterização abrangente do perfil de expressão gênica do secretoma e de CIFs em linhagens celulares do CCLE utilizando dados scRNA-Seq tem o potencial de revelar padrões transcricionais do secretoma das células cancerosas e, conseqüentemente, contribuir para o entendimento da variação na prevalência e gravidade da caquexia induzida pelas células cancerosas dentre os diferentes cânceres humanos. O presente projeto analisou o perfil de expressão gênica do secretoma e de CIFs utilizando dados scRNA-Seq em 198 linhagens de células cancerosas do CCLE. Essas análises exploratórias de dados de scRNA-Seq permitiram: 1) determinar padrões e funções biológicas associadas às mudanças na expressão gênica do secretoma entre linhagens celulares de cânceres que possuem prevalência distinta de caquexia e perda de peso; 2) avaliar o perfil de expressão gênica de CIFs para linhagens celulares do CCLE que mais se relacionam com tumores primário; 3) Determinar, por meio de expressão gênica dos CIFs, qual são as linhagens celulares que mais se correlacionam com tumores primário e, conseqüentemente, as melhores linhagens para estudar caquexia associada ao câncer *in vitro*.

## 2. Metodologia

Para a identificação e construção de uma lista com os fatores indutores de caquexia (CIFS), utilizamos as ferramentas de mineração de texto Geneshot (<https://maayanlab.cloud/geneshot/>)<sup>17</sup> e Open Targets Plataforma (<https://platform.opentargets.org/>)<sup>18</sup>, com os termos de pesquisa “Cancer Cachexia” e “Cachexia”, respectivamente. Em seguida, as listas foram filtradas para genes codificadores de proteínas secretadas com base na lista do secretoma humano, disponível no *The Human Protein Atlas* (<https://www.proteinatlas.org/humanproteome/secretome>)<sup>13</sup>, resultando em uma lista final com 123 CIFS.

Os dados de sequenciamento de scRNA-seq de 198 células do CCLE (estudo SCP542)<sup>16</sup> foram obtidos no site *Curated Cancer Cell Atlas* (3CA)<sup>19</sup>. Para cada tipo tumoral (n = 18), foram selecionadas as cinco linhagens celulares mais correlacionadas com amostras de tumores primários, determinada por um baixo valor de distância do tumor na análise do Cellinger<sup>14</sup>, resultando em 84 linhagens celulares e 20,691 células.

Para a análise de correlação, foram obtidos dados de scRNA-seq de 10 tipos tumorais: Pulmão, Pele, Pâncreas, Ovário, Rim, Cabeça e Pescoço, Próstata, Mama, Cérebro e Colorretal, do banco de dados 3CA<sup>19</sup>, onde os diferentes estudos sobre o mesmo tipo de câncer foram integrados com a ferramenta Seurat, utilizando sua configuração padrão<sup>20</sup>. Para a seleção dos estudos, um filtro foi aplicado e apenas aqueles que apresentavam tumores humanos primários, sem tratamento anterior à coleta, tecnologia de scRNA-Seq da *10x Genomics*, e dados disponíveis publicamente foram selecionados, resultando em 906,884 células

únicas. Por fim, um último filtro foi realizado, o primeiro para conter somente as células neoplásicas primárias foram selecionadas para a análise, resultando em um total de 259,431 células, e o segundo, para seleção de linhagens dos mesmos tipos tumorais previamente selecionados, resultando num total de 54 linhagens e 22,073 células únicas.

As análises dos dados scRNA-Seq foram reanalisadas na plataforma RStudio (<https://www.rstudio.com/>), com a última versão disponível do pacote de ferramentas Seurat (v4) (<https://github.com/satijalab/seurat>)<sup>20</sup>. Os genes diferencialmente expressos, extraídos dos resultados gerados pela função *DotPlot* em cada linhagem celular e das células neoplásicas primárias, foram filtrados para genes de CIFs, como descrito em Freire *et al.*, 2020<sup>8</sup>. O método de correlação utilizado foi o de correlação Pearson. “Heatmaps” foram gerados e analisados com a ferramenta Morpheus (<https://software.broadinstitute.org/morpheus>)<sup>21</sup>

### 3. Resultados

Os dados de scRNA-seq das linhagens foram analisados com a plataforma Rstudio, utilizando o pacote Seurat<sup>20</sup>, para normalização, escalonamento e análise de componentes principais (PCA). A partir disso, foram identificados 15 componentes principais (PCs) significativos, que apresentaram forte enriquecimento de recursos com baixo “p-value”. Com base nos tipos tumorais, realizamos a clusterização das células e aplicamos uma redução dimensional não linear (non-linear dimensional reduction - UMAP), resultando em 58 agrupamentos de células (Figura 1).

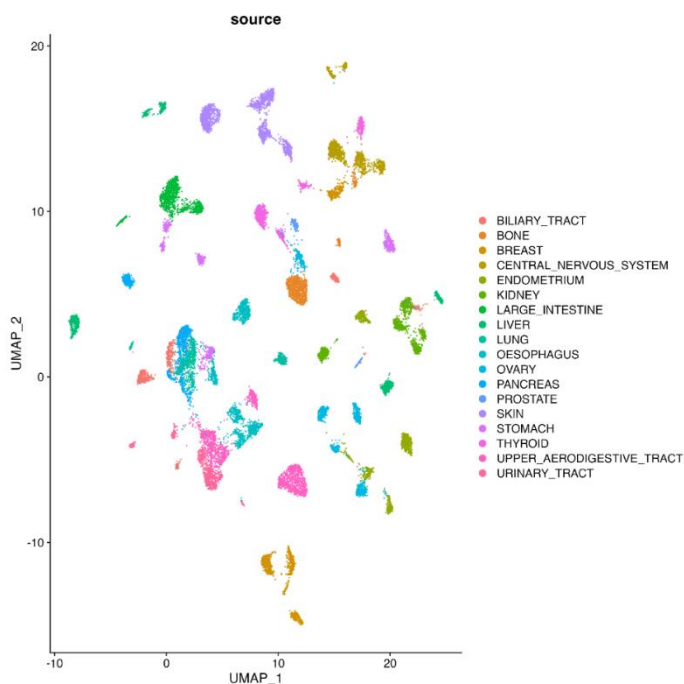




Figura 2: Heatmap da expressão média escalada de células única de Fatores Indutores de Caquexia em 84 Linhagens celulares que mais se assemelham a tumores primários. O Heatmap exibe a expressão em escala média de 101 Fatores Indutores de Caquexia em 84 linhagens celulares, representando um total de 20691 células. A escala de cores representa os níveis de expressão relativos, variando baixa (azul) a alta (vermelho) expressão. Os dados de células únicas foram obtidos no site *Câncer Curated Cell Atlas* (3CA).

Para determinar as linhagens que mais se correlacionam com os tumores primários em relação à expressão de CIFs, foi obtida a expressão média escalada de células malignas de diversos tipos de câncer (pâncreas, pulmão, próstata, ovário, mama, pele, cérebro e colón, nasofaringe e rim). Dos 101 CIFs expressos nas linhagens celulares, 89 também mostram expressão nas células neoplásicas primárias. Em seguida, uma análise de correlação entre o perfil de expressão das células malignas e o perfil de expressão de 54 linhagens celulares, contendo o mesmo tipo tumoral que das células malignas, foi realizada (Figura 3), demonstrando que as linhagens celulares que mais se correlacionam são as da próstata, mama e pulmão.

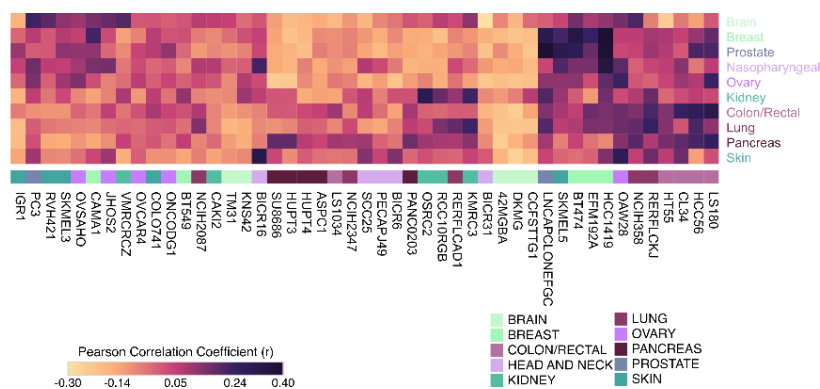


Figura 3: Heatmap de correlação de Pearson da Expressão Média Escalada entre Linhagens Celulares e Células Neoplásicas Primárias Humanas, colorido por tipo de câncer. O *Heatmap* exibe os coeficientes de correlação entre a expressão média escalada em 54 linhagens celulares e 10 *Datasets* de células neoplásicas primárias. A escala de cores representa o valor do coeficiente de correlação, variando de baixo (amarelo) a alto (roxo). O *Heatmap* mostra que o câncer de próstata, mama e pulmão apresentam os coeficientes de correlação mais altos, representados pelas cores cinza, verde claro e marrom, respectivamente. Cada tipo de câncer é representado por uma cor diferente, como indicado na legenda da figura.

Por fim, para cada tipo tumoral, foram selecionadas as linhagens que apresentaram o menor valor de  $p$  e o maior valor de  $r$ , resultando em 10 linhagens (Figura 4) que melhor recapitulam o perfil transcricional de CIFs, indicando as linhagens celulares para estudos *in vitro* da síndrome, visando a obtenção de resultados mais promissores e efetivos.

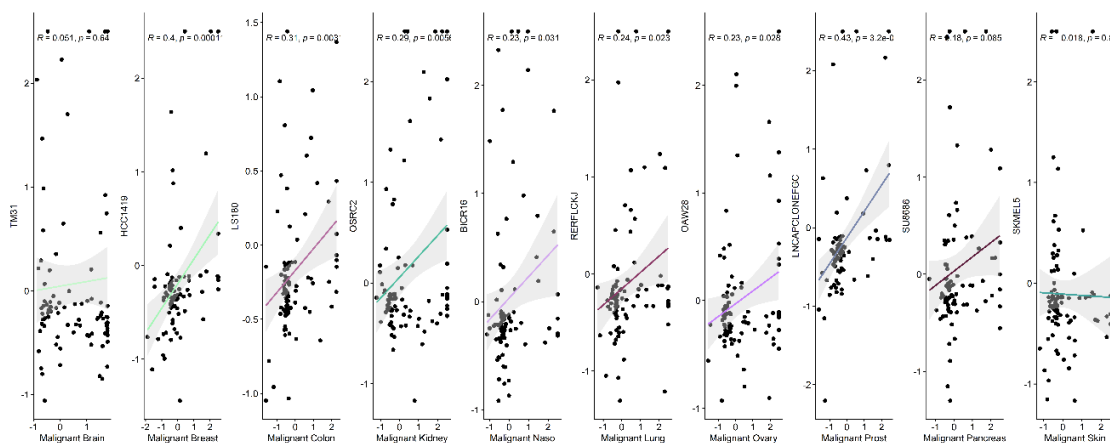


Figura 4: Gráfico de correlação Pearson da Expressão Média Escalada entre Linhagens Celulares e Células Neoplásicas Primárias Humanas, colorido por tipo de câncer. O Gráfico exibe as linhagens (eixo Y) que apresentaram o maior valor de r e o menor valor de p em cada tipo tumoral (eixo X), com a ordem do tipo tumoral sendo (da esquerda para direita): Cérebro, Mama, Cólon, Rim, Nasofaringe, Pulmão, Ovário, Próstata, Pâncreas e Pele.

#### 4. Discussão

Com intuito de encontrar as linhagens celulares que mais recapitulam o perfil transcricional da caquexia associada ao câncer, o presente estudo identificou 10 linhagens celulares: TM31 (câncer de cérebro), HCC1419 (câncer de mama), LS180 (câncer de cólon), OSRC2 (câncer de rim), BICR16 (câncer de nasofaringe), RERFLCKJ (câncer pulmão), OAW28 (câncer de ovário), LNCAPCLONEFGC (câncer de próstata), SU8686 (câncer pâncreas) e SKMEL5 (câncer pele).

Estes resultados possuem uma alta importância e relevância, uma vez que criteriosa de linhagens celulares desempenha um papel crucial em estudos *in vitro* e na pesquisa pré-clínica, oferecendo oportunidades para o desenvolvimento de estratégias terapêuticas<sup>12,14,22-24</sup>. É importante ressaltar que a escolha inadequada de linhagens celulares pode comprometer a relevância clínica e resultar em resultados insatisfatórios<sup>22</sup>.

Estudos anteriores revelaram que as linhagens celulares frequentemente usadas em estudos de câncer de fígado<sup>24</sup> e de ovário<sup>25</sup> nem sempre refletiam de maneira adequada os tumores primários correspondentes. Isso destaca a necessidade de aprimorar os métodos de seleção das linhagens celulares para garantir sua representatividade e relevância em relação aos tumores de interesse. Nesse contexto, analisamos dados transcriptômicos públicos obtidos por meio do sequenciamento de célula única (scRNA-seq) para identificar as linhagens celulares mais apropriadas para estudos *in vitro* sobre a caquexia associada ao câncer. Utilizamos um conjunto de dados composto por 281.504 células, incluindo 22.073 células de linhagens celulares e 259.504 células neoplásicas, das quais obtivemos valores de expressão média escalados para os fatores indutores de caquexia (CIFs).

Essa abordagem baseada na expressão gênica foi escolhida devido à sua ampla utilização em estudos anteriores para identificar linhagens celulares relevantes<sup>12,14,25</sup>. Além disso, a correlação entre o número de genes de CIFs regulados positivamente e a prevalência de caquexia reforça a validade dessa estratégia<sup>8</sup>.

É interessante notar que, ao comparar nossos resultados com o banco de dados do Cellinger<sup>14</sup>, não encontramos nenhuma linhagem que se destacasse como a que mais recapitula o câncer. Essa diferença nos resultados pode ser explicada por dois fatores importantes. Primeiro, nosso foco específico na caquexia associada ao câncer, que pode ter influenciado a seleção das linhagens celulares em comparação com um conjunto de dados mais amplo. Em segundo, a tecnologia de sequenciamento utilizada na construção do banco de dados do Cellinger, sequenciamento de RNA-seq convencional (bulk RNA), possui limitações para a captura da heterogeneidade tumoral, uma vez que captura a expressão média dessa heterogeneidade, o que compromete o perfil de expressão, levando a alterações na expressão gênica<sup>12,26,27</sup>.

A análise da literatura sobre essas 10 linhagens identificadas revelou que já foram amplamente utilizadas, onde diversos linhagem TM31 de câncer de cérebro apresentam uma expressão elevada do gene *VIM*<sup>28</sup>, o qual foi previamente relacionado com o crescimento tumoral acelerado, invasividade e pior prognóstico em cânceres de próstata, gastrointestinal e do sistema nervoso central<sup>29</sup>. Já a linhagem HCC1419 de câncer de mama possui o gene *PREX1*, relacionado com proliferação celular por meio de um loop de feedback positivo, amplificado<sup>30,31</sup>.

Zheng<sup>32</sup> et al. evidenciou que a linhagem de câncer de cólon LS180 (cólon) apresenta uma redução na expressão do gene *DIRAS1*, causada pela hipermetilação da região promotora, resultando em carcinogênese e progressão do câncer<sup>28</sup>. Em outro estudo, Chen et al<sup>33</sup>. demonstrou que a linhagem de câncer de rim OSRC2 (rim) apresenta expressão aumentada da enzima COX-2, enzima relacionada à tumorigênese e à progressão tumoral, algo que difere da sua expressão insignificante em rins saudáveis<sup>33,34</sup>.

Em estudos que utilizaram a linhagem BICR16 (nasofaringe), demonstraram que há expressão amplificada do gene *FGFR1*, gene relacionado com à tumorigênese<sup>35</sup>. Já em estudo de RERFLCKJ (pulmão), os pesquisadores encontram uma alta expressão do gene *TMEPAI*, conhecido por converter o sinal supressor do TGF- $\beta$  em um sinal promotor<sup>36</sup>.

Para as linhagens OAW28 (ovário), estudo anteriores já consideravam ela um bom modelo para o estudo do câncer de ovário<sup>37</sup>. Mitchell et al. encontraram que a linhagem LNCAPCLONEFGC (próstata) é conhecida por não expressar o gene de reparo *hMSH2*<sup>38</sup>.

Estudos indicam que a linhagem SU8686 (pâncreas) demonstram a maior expressão de mRNA da enzima sialoproteína óssea (BSP), enzima que desempenha um papel importante no crescimento

celular, migração e invasão<sup>39</sup> e que a linhagem SKMEL5 (pele) exhibe níveis elevados de alguns CIFs, como IL6, IL8 e VEGF<sup>40</sup>.

Por fim, considerando esses resultados, as linhagens de (LS180) e pulmão (RERFLCKJ) emergem como candidatos potenciais para estudos futuros, devido a seus valores de correlação (0.31 e 0.24, respectivamente), por estarem estatisticamente significante ( $p < 0.05$ ), e, principalmente, devido à alta prevalência de caquexia nesses tipos de câncer.

## 5. Conclusão

Neste estudo, identificamos as melhores linhagens celulares para estudar a caquexia in vitro. Utilizamos o perfil de expressão escalada de 89 fatores indutores de caquexia (CIFs), que, primeiramente foram identificados através das ferramentas de mineração de texto, Geneshot e Open Targets, e posteriormente, filtrados utilizando o TCGA. Com base no perfil de expressão CIFs presentes tanto em linhagens celulares quanto em dados humanos, realizamos uma análise de correlação de Pearson para identificar qual linhagem apresenta a maior correlação com o tipo tumoral. Esse processo resultou em 10 linhagens celulares que melhor recapitulam o perfil transcricional da caquexia associada ao câncer.

Os resultados deste estudo sugerem que as linhagens celulares de câncer de mama e pulmão são os melhores modelos para estudar a caquexia in vitro. Essas linhagens apresentam uma alta correlação com os fatores indutores de caquexia presentes em pacientes com câncer. Os resultados também sugerem que as linhagens celulares de câncer de ovário (OAW28), nasofaringe (BICR16), rim (OSRC2) e colón (LS180) são potenciais bons modelos para estudos futuros<sup>4,5</sup>.

## 6. Referências Bibliográficas

1. Fearon, K. *et al.* Definition and classification of cancer cachexia: an international consensus. *Lancet Oncol* **12**, 489–495 (2011).
2. Fearon, K., Arends, J. & Baracos, V. Understanding the mechanisms and treatment options in cancer cachexia. *Nature Reviews Clinical Oncology* vol. 10 90–99 Preprint at <https://doi.org/10.1038/nrclinonc.2012.209> (2013).
3. Fearon, K. C. H., Glass, D. J. & Guttridge, D. C. Cancer cachexia: Mediators, signaling, and metabolic pathways. *Cell Metabolism* vol. 16 153–166 Preprint at <https://doi.org/10.1016/j.cmet.2012.06.011> (2012).
4. Pressoir, M. *et al.* Prevalence, risk factors and clinical implications of malnutrition in french comprehensive cancer centres. *Br J Cancer* **102**, 966–971 (2010).
5. Hébuterne, X. *et al.* Prevalence of malnutrition and current use of nutrition support in patients with cancer. *Journal of Parenteral and Enteral Nutrition* **38**, 196–204 (2014).
6. Robinson, J. L., Feizi, A., Uhlén, M. & Nielsen, J. A Systematic Investigation of the Malignant Functions and Diagnostic Potential of the Cancer Secretome. *Cell Rep* **26**, 2622–2635.e5 (2019).

7. Argilés, J. M., Stemmler, B., López-Soriano, F. J. & Busquets, S. Inter-tissue communication in cancer cachexia. *Nature Reviews Endocrinology* vol. 15 9–20 Preprint at <https://doi.org/10.1038/s41574-018-0123-0> (2018).
8. Freire, P. P. *et al.* The expression landscape of cachexia-inducing factors in human cancers. *J Cachexia Sarcopenia Muscle* **11**, 947–961 (2020).
9. Whiteside, T. L. The tumor microenvironment and its role in promoting tumor growth. *Oncogene* vol. 27 5904–5912 Preprint at <https://doi.org/10.1038/onc.2008.271> (2008).
10. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
11. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
12. Yu, K. *et al.* Comprehensive transcriptomic analysis of cell lines as models of primary tumors across 22 tumor types. *Nat Commun* **10**, (2019).
13. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (1979)* **347**, (2015).
14. Warren, A. *et al.* Global computational alignment of tumor and cell line transcriptional profiles. *Nat Commun* **12**, (2021).
15. Carvalho, R. F. *et al.* Drug repositioning based on the reversal of gene expression signatures identifies top2a as a therapeutic target for rectal cancer. *Cancers (Basel)* **13**, (2021).
16. Kinker, G. S. *et al.* Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet* **52**, 1208–1218 (2020).
17. Lachmann, A. *et al.* Geneshot: search engine for ranking genes from arbitrary text queries. *Nucleic Acids Res* **47**, W571–W577 (2019).
18. Koscielny, G. *et al.* Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res* **45**, D985–D994 (2017).
19. Gavish, A. *et al.* The transcriptional hallmarks of intra-tumor heterogeneity across a thousand tumors. doi:10.1101/2021.12.19.473368.
20. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
21. Starruß, J., De Back, W., Bruschi, L. & Deutsch, A. Morpheus: A user-friendly modeling environment for multiscale and multicellular systems biology. *Bioinformatics* **30**, 1331–1332 (2014).
22. Gillet, J. P. *et al.* Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. *Proc Natl Acad Sci U S A* **108**, 18708–18713 (2011).
23. Gillet, J. P., Varma, S. & Gottesman, M. M. The clinical relevance of cancer cell lines. *Journal of the National Cancer Institute* vol. 105 452–458 Preprint at <https://doi.org/10.1093/jnci/djt007> (2013).
24. Chen, B., Sirota, M., Fan-Minogue, H., Hadley, D. & Butte, A. J. Relating hepatocellular carcinoma tumor samples and cell lines using gene expression data in translational research. *BMC Med Genomics* **8**, (2015).
25. Jiang, G. *et al.* Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. *BMC Genomics* **17**, (2016).
26. Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611–1624.e24 (2017).
27. Guruprasad, P., Lee, Y. G., Kim, K. H. & Ruella, M. The current landscape of single-cell transcriptomics for cancer immunotherapy. *Journal of Experimental Medicine* **218**, (2021).

28. Kurimoto, M. *et al.* Establishment and characterization of a novel malignant astrocytoma cell line derived from a tumor removed in a patient with neurofibromatosis Type 1. *J Neurosurg* **94**, 301–308 (2001).
29. Satelli, A. & Li, S. Vimentin in cancer and its potential as a molecular target for cancer therapy. *Cellular and Molecular Life Sciences* **68**, 3033–3046 (2011).
30. Liu, H.-J. *et al.* PtdIns(3,4,5)P<sub>3</sub>-dependent Rac Exchanger 1 (PREX1) Rac-Guanine Nucleotide Exchange Factor (GEF) Activity Promotes Breast Cancer Cell Proliferation and Tumor Growth via Activation of Extracellular Signal-regulated Kinase 1/2 (ERK1/2) Signaling. *Journal of Biological Chemistry* **291**, 17258–17270 (2016).
31. Dillon, L. M. *et al.* P-REX1 creates a positive feedback loop to activate growth factor receptor, PI3K/AKT and MEK/ERK signaling in breast cancer. *Oncogene* **34**, 3968–3976 (2015).
32. Zheng, R. *et al.* Methylation of DIRAS1 promotes colorectal cancer progression and may serve as a marker for poor prognosis. *Clin Epigenetics* **9**, 50 (2017).
33. CHEN, Q., SHINOHARA, N., ABE, T., HARABAYASHI, T. & NONOMURA, K. IMPACT OF CYCLOOXYGENASE-2 GENE EXPRESSION ON TUMOR INVASIVENESS IN A HUMAN RENAL CELL CARCINOMA CELL LINE. *Journal of Urology* **172**, 2153–2157 (2004).
34. Hashemi Goradel, N., Najafi, M., Salehi, E., Farhood, B. & Mortezaee, K. Cyclooxygenase-2 in cancer: A review. *J Cell Physiol* **234**, 5683–5699 (2019).
35. Koole, K. *et al.* FGFR1 Is a Potential Prognostic Biomarker and Therapeutic Target in Head and Neck Squamous Cell Carcinoma. *Clinical Cancer Research* **22**, 3884–3893 (2016).
36. Vo Nguyen, T. T. *et al.* TMEPAI/PMEPA1 enhances tumorigenic activities in lung cancer cells. *Cancer Sci* **105**, 334–341 (2014).
37. Černe, K. *et al.* Characteristics of Extracellular Vesicles from a High-Grade Serous Ovarian Cancer Cell Line Derived from a Platinum-Resistant Patient as a Potential Tool for Aiding the Prediction of Responses to Chemotherapy. *Pharmaceuticals* **16**, 907 (2023).
38. Mitchell, R. J. Mismatch Repair Genes hMLH1 and hMSH2 and Colorectal Cancer: A HuGE Review. *Am J Epidemiol* **156**, 885–902 (2002).
39. Kaye, H. *et al.* Effects of bone sialoprotein on pancreatic cancer cell growth, invasion and metastasis. *Cancer Lett* **245**, 171–183 (2007).
40. Singh, L. *et al.* ILT3 (LILRB4) Promotes the Immunosuppressive Function of Tumor-Educated Human Monocytic Myeloid-Derived Suppressor Cells. *Molecular Cancer Research* **19**, 702–716 (2021).