



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu



ANÁLISE DO PERFIL TRANSCRICIONAL DO SECRETOMA DO LAVADO BRONCO ALVEOLAR EM PACIENTES COM RISCO DE COVID-19 SEVERA

ANA LUIZA LABBATE BONALDO



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Campus de Botucatu



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

ANÁLISE DO PERFIL TRANSCRICIONAL DO SECRETOMA DO LAVADO BRONCO ALVEOLAR EM PACIENTES COM RISCO DE COVID-19 SEVERA

Mestranda: Ana Luiza Labbate Bonaldo

Orientador: Robson Francisco Carvalho

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas (Genética), como uma das exigências para a obtenção do título de Mestre.

**BOTUCATU - SP
2022**

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Bonaldo, Ana Luiza Labbate.

Análise do perfil transcricional do secretoma do lavado
bronco alveolar em pacientes com risco de COVID-19 severa
/ Ana Luiza Labbate Bonaldo. - Botucatu, 2022

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

Orientador: Robson Francisco Carvalho

Capes: 20205007

1. COVID-19 (Doença). 2. Secretoma. 3. Transcriptoma.
4. Análise da expressão gênica de célula única.

Palavras-chave: COVID-19 severa; Infecção por SARS-CoV-2;
LRP11; Secretoma; scrNA-Seq.

Dedicatória

Aos que perderam
alguém e aos que se foram
para a COVID-19, dedico.

Agradecimentos

Aos meus pais, *Carlos e Rita*, meu grande alicerce, que sempre estiveram ao meu lado me dando apoio e colo quando necessário. Obrigada por se orgulharem de mim e me deixarem bater asas. Sempre saberei para onde voltar. Eu amo vocês.

Ao meu noivo, *João Luís*, essencial nos bons e nos maus momentos. Sou eternamente grata pelos puxões de orelha e por ser minha metade pé no chão. Obrigada por me ouvir e se interessar pelo meu trabalho, por me distrair e me acolher. Te amo.

Aos meus colegas de laboratório *Amanda, Jakeline, Sarah, Victória, Jeferson, Caio e Fernando*. Vocês me tiraram risos e sorrisos em dias nublados. Sou extremamente grata por tudo que aprendo com vocês.

Aos meus amigos *Verônica, Felipe, Gabriel, Luiza, Sabrina e Laura*, que mesmo sem saber, deixavam os dias mais fáceis e leves. Tenho muita sorte em ter vocês ao meu lado.

Às minhas queridas companheiras de vôlei. Vocês foram essenciais pra esse momento. Obrigada por torcerem por mim. Estarei sempre torcendo por vocês.

Aos membros da banca, professoras *Flávia Delella e Célia Regina*. Obrigada por aceitarem estar presentes em um momento tão importante para mim, tanto profissional e quanto pessoal.

Ao Programa de Pós-graduação em Ciências Biológicas (Genética) e a CAPES pela infraestrutura e bolsa concedidos para a realização deste trabalho.

Ao meu orientador, professor *Robson Carvalho*. Por cada conselho e por acreditar em mim, mesmo quando as coisas não estavam tão bem. É uma honra poder aprender diariamente com você. Que esse trabalho seja o primeiro de vários. Obrigada por não desistir de mim.

Por fim, agradeço à *Deus* por sua infinita bondade sobre a minha vida (Habacuque 3:18).

Epígrafe

Para ser grande, sê inteiro: nada
Teu exagera ou exclui.
Sê todo em cada coisa. Põe quanto és
No mínimo que fazes.
(Ricardo Reis)

Lista de Ilustrações

Figura 1. Análise da expressão gênica do secretoma do BALF de pacientes nas comparações estudadas (idade, gênero e carga viral).	31
Figura 2. Caracterização do grupo de 43 genes exclusivos em comum no grupo de melhor prognóstico (mulheres, jovens e baixa carga viral).	34
Figura 3. <i>Violin plot</i> demonstrando os níveis de expressão dos ligantes e receptores presentes no cluster A usando os dados do <i>conjunto de dados</i> GSE145926 com adicional da amostra de BALF saudável GSM3660650 do <i>conjunto de dados</i> GSE128033.....	34
Figura 4. <i>Box plot</i> do <i>time course</i> dos níveis de expressão dos ligantes e receptores presentes no cluster A nas amostras de BALF de nove macacos rhesus adultos infectados com SARS-CoV-2 (GSE156701).	35
Figura 5. Gráfico UMAP representando a distribuição de expressão dos ligantes e receptores do cluster A nos tipos celulares presentes no BALF	37
Figura 6. <i>Dotplot</i> representando a expressão de cada ligante e receptor presentes no cluster A em cada tipo celular encontrado no BALF	37

Lista de Tabelas

Tabela 1. Comparações para cada análise de expressão diferencial e as novas nomenclaturas, com seus respectivos números de pacientes (n).....	25
Tabela 2. Número total e de cada grupo analisado de DEGs com aumento e diminuição de expressão.....	29

Lista de Abreviações

COVID-19	Doença causada pelo novo coronavírus 2
SARS-CoV-2	Coronavírus 2 da síndrome respiratória aguda grave
GEO	<i>Gene Expression Omnibus</i>
RNA-Seq	Sequenciamento de RNA
DEGs	<i>Differentially expressed genes</i>
OMS	Organização Mundial da Saúde
RT-qPCR	PCR quantitativo em tempo real
CT	Tomografia computadorizada
DNA	Ácido desoxirribonucleico
mRNA	RNA mensageiro
SARS-CoV	<i>Severe Acute Respiratory Syndrome</i>
MERS-CoV	<i>Middle East respiratory syndrome</i>
RNA	Ácido ribonucleico
ORF	<i>Open Read Frames</i>
scRNA-Seq	Sequenciamento de RNA de célula única
IUPHAR	<i>The International Union of Basic and Clinical Pharmacology</i>
STRING	<i>Search Tool for Retrieval of Interacting Genes</i>
PPI	<i>Protein-protein interaction</i>
PCA	<i>Principal Components Analysis</i>
BALF	<i>Bronchoalveolar lavage fluid</i>
UMAP	<i>Uniform manifold approximation and projection</i>
t-SNE	<i>t-distributed stochastic neighbor embedding</i>
Ct	<i>Cycle threshold</i>
GO	<i>Gene ontology</i>
MP	Melhor prognóstico

PP	Pior prognóstico
mDC	<i>Myeloid dendritic cell</i>
NK	<i>Natural killer</i>
LDL	<i>Low-density lipoprotein</i>
LDLR	<i>Low-density lipoprotein receptor</i>

Resumo

A doença causada pelo novo coronavírus (COVID-19) se tornou uma pandemia e já atingiu mais de 595 milhões de pessoas em todo o mundo, ultrapassando os 6 milhões e 400 mil mortos após dois anos e meio. Pacientes com comorbidades, como diabetes e hipertensão, ou do sexo masculino e idosos são descritos como grupo de pior prognóstico. Já mulheres, jovens e pacientes com carga viral baixa são considerados grupo de melhor prognóstico. A produção excessiva de citocinas pró-inflamatórias, conhecida como tempestade de citocinas (*cytokine storm*), também está relacionada com a severidade e mortalidade da COVID-19. Portanto, a análise do conjunto de moléculas secretadas pelas células (secretoma) de tecidos alvos do SARS-CoV-2 é de extrema importância para o prognóstico e para a identificação de mediadores dessa doença. Diante disso, o presente trabalho teve como objetivo analisar o perfil transcricional do secretoma de amostras de lavado bronco alveolar em pacientes com risco distintos de desenvolvimento de COVID-19 severa. A hipótese desse trabalho é que a análise do perfil do secretoma utilizando dados de transcriptoma, permitir predizer genes envolvidos com o prognóstico da infecção pelo SARS-CoV-2 e com a progressão da doença. Foi reanalisado um conjunto de dados de transcriptoma (RNA-Seq) com 430 pacientes positivos para COVID-19 e 54 pacientes negativos, disponível no repositório *Gene Expression Omnibus* (GEO; GSE152075), separados por gênero, idade (<60 anos, jovens; ≥60 anos, idosos) e carga viral (baixa e alta). Os genes diferencialmente expressos (DEGs) entre as condições foram utilizados para identificação de genes que codificam proteínas secretadas, utilizando os dados disponíveis na plataforma *The Human Protein Atlas*. Identificamos um conjunto de DEGs exclusivo do grupo de melhor prognóstico (mulheres, jovens e baixa carga viral). Desse conjunto de genes, caracterizamos quais codificam ligantes e receptores e, posteriormente, analisamos, através do método de análise de sequenciamento de células únicas (*single-cell RNA-Seq*), a expressão dos genes selecionados em células únicas do lavado bronco-alveolar de pacientes com COVID-19. Como método de validação dos nossos achados, utilizamos o conjunto de dados GSE145926 (GEO) para análise temporal (dias 1, 2, 4, 7, 10 e 12) do lavado bronco alveolar de macacos rhesus adultos infectados com SARS-CoV-2. Nossos dados demonstraram a relevância do receptor LRP11, como um potencial biomarcador de melhor prognóstico para COVID-19. Esse receptor faz parte da família de receptores de LDL, o qual altera sua expressão em infecções virais. Além disso, observamos que esse receptor tem sua expressão predominantemente em células epiteliais, que expressam o receptor ACE2 e, portanto, são mais susceptíveis às infecções pelo SARS-CoV-2. Esses resultados são relevantes pois os genes identificados podem determinar o prognóstico dos pacientes com COVID-19 e constituem potenciais alvos de drogas para redução da mortalidade pela doença.

Palavras-chave: Infecção por SARS-CoV-2, COVID-19 severa, Secretoma, LRP11, scRNA-Seq

Abstract

The disease caused by the new coronavirus (COVID-19) has become a pandemic and has already reached more than 595 million people worldwide, exceeding 6 million and 400 thousand dead after two and a half years. Patients with comorbidities such as diabetes and hypertension, or male and elderly patients are described as groups with the worst prognosis. On the other hand, women, young people and patients with low viral loads are considered groups with the best prognosis. Excessive production of pro-inflammatory cytokines, known as cytokine storm, is also linked to the severity and mortality of COVID-19. Therefore, the analysis of the set of molecules secreted by the cells (secretome) of SARS-CoV-2 target tissues is extremely important for the prognosis of the disease. Our study analyzed the transcriptional profile of the secretome of bronchoalveolar lavage samples in patients at different risks of developing severe COVID-19. We hypothesized that a transcriptomic-based secretome analysis allows the prediction of genes involved in the prognosis of SARS-CoV-2 infection and the progression of the disease. We applied a transcriptome dataset (RNA-Seq) with 430 COVID-19 positive patients and 54 negative patients, available in the Gene Expression Omnibus repository (GEO; GSE152075), which was separated by gender, age (<60 years, young; ≥60 years, elderly), and viral load (low and high). Differently expressed genes (DEGs) between conditions were used to identify genes encoding secreted proteins, using data available on The Human Protein Atlas platform. We identified a set of DEGs exclusive to the group with the best prognosis (women, young people, and low viral load). From this set of genes, we characterized those encoding ligands and receptors. Next, we analyzed, through the single-cell sequencing analysis method (single-cell RNA-Seq), the expression of the selected genes in single cells from the bronchoalveolar lavage of patients with COVID-19. To validate our findings, we used the GSE145926 (GEO) dataset for a time course analysis (days 1, 2, 4, 7, 10, and 12) of bronchoalveolar lavage samples of adult rhesus monkeys infected with SARS-CoV-2. Our data demonstrated the relevance of the LRP11 receptor as a potential biomarker of better prognosis for COVID-19. This receptor is part of the LDL receptor family, which presents gene expression changes-induced by viral infections. We also observed that this receptor is predominantly expressed in epithelial cells, which express the ACE2 receptor and, consequently, are more susceptible to SARS-CoV-2 infections. These results are relevant because the genes identified in our analysis can determine the prognosis of COVID-19 patients and constitute potential drug targets to reduce disease mortality.

Keywords: SARS-CoV-2 infection, severe COVID-19, Secretome, LRP11, scRNA-Seq

SUMÁRIO

1 - Introdução	13
1.1 Doença do Coronavírus 2019 (COVID-19)	13
1.2 A infecção pelo SARS-CoV-2	14
1.3 Resposta transcricional das células pulmonares à infecção pelo SARS-CoV-2	16
1.4 O secretoma na infecção viral	17
2 - Objetivos	20
2.1 Objetivo Geral	20
2.2 Objetivos Específicos	20
3 – Artigo científico	21
4 - Conclusão	40
5 – Referências bibliográficas	42
6 – Apêndices A	51
6.1 Figuras suplementares	51

1 - Introdução

1.1 Doença do Coronavírus 2019 (COVID-19)

O primeiro caso conhecido da síndrome respiratória aguda grave pelo novo coronavírus 2 (SARS-CoV-2) foi relatado na cidade de Wuhan, na China, em dezembro de 2019 (1). Desde então, a doença causada pelo novo coronavírus (COVID-19) tornou-se uma pandemia (2). No fim de agosto de 2022, dois anos e meio após o início da pandemia, o número de infectados ultrapassa os 596 milhões, com mais de seis milhões de mortos pela doença em todo o mundo (3). Nesse mesmo período, o Brasil já alcançou a marca de 34 milhões e 264 mil casos confirmados da doença, e mais de 682 mil mortes (3).

A COVID-19 é uma doença infecciosa, e a transmissão do vírus SARS-CoV-2 ocorre por transmissão direta, por contato com pessoas infectadas por meio de pequenas partículas líquidas, lançadas pela boca ou nariz; ou indiretamente, por contato em superfícies ou objetos utilizados por pessoas infectada (4), o que justifica a alta dispersão do vírus em um curto período de tempo. Para prevenção da infecção, a vacinação, o distanciamento social e o uso de máscaras são amplamente recomendados (5).

Segundo a Organização Mundial da Saúde (OMS), a COVID-19 é caracterizada por uma variedade de sintomas (5). Os sintomas clínicos mais frequentes são febre, tosse, produção de escarro, fadiga e dificuldade para respirar (6). Sintomas gastrointestinais, tais como náusea, diarreia, diminuição do apetite e dor abdominal ocorrem com menor frequência (7). No entanto, existem alguns sintomas raros ou poucos conhecidos da COVID-19, como diminuição ou perda do paladar (disgeusia e ageusia, respectivamente) (8,9), e manifestações neurológicas, como comprometimento do olfato (anosmia) (10). Os sintomas no sistema nervoso central estão associados com a principal forma de lesão neurológica em pacientes com COVID-19, e os casos mais graves podem apresentar doenças cerebrovasculares agudas, como acidente vascular cerebral isquêmico e hemorragia cerebral, e consciência prejudicada (10).

Os graus da doença podem variar e o paciente pode ser assintomático, ou apresentar sintomas leves, médios ou graves, podendo evoluir para óbito nos casos mais graves (11,12). Alguns grupos de riscos foram estipulados com maiores possibilidades de piores sintomas e severidade na COVID-19 (13). Pacientes do sexo masculino (14–17) e idosos (> 60 anos) (13,16,17) foram associados com piores desfechos da doença. A carga viral também tem sido

relevante para determinar a mortalidade na infecção pelo SARS-CoV-2. Pacientes com morte decorrente de COVID-19 possuem carga viral mais alta quando comparados a pacientes que não morreram em decorrência da doença (18). A presença de algumas comorbidades também pode agravar os casos de COVID-19, como doenças cardiovasculares e respiratórias (incluindo a asma), obesidade, diabetes e doença renal crônica (13,16,17). Os pacientes com câncer também são considerados bastante vulneráveis e apresentam maior risco de eventos clínicos graves do que os pacientes não oncológicos (19,20).

Como estratégia de diagnóstico para a COVID-19 existem métodos moleculares, como a Reação em Cadeia da Polimerase Quantitativa após Transcrição Reversa (RT-qPCR), que detecta a presença do RNA viral através de amostras coletadas através do *swab* da nasofaringe e da orofaringe (21). Além de métodos sorológicos, que foram desenvolvidos para detectar imunoglobulinas (IgM e IgG) específicas contra o vírus (22), métodos de imagem, como a radiografia do tórax e a tomografia computadorizada (CT), que identificam anormalidades pulmonares (23) também são utilizados. Por fim, existe o chamado teste rápido, que visa também detectar as imunoglobulinas IgM e IgG através do sangue total, plasma ou soro, tendo um tempo de reação de dez minutos (24).

No entanto, como medida mais eficaz a prevenção a COVID-19, a colaboração mundial fez com que, rapidamente, fossem desenvolvidas vacinas para a COVID-19. Vacinas de DNA, de mRNA, de vetor viral não replicante e de vírus inativado têm sido as mais testadas e aprovadas pela Organização Mundial da Saúde (25). Segundo a Universidade Johns Hopkins, até 8 de setembro de 2022 já foram administradas mundialmente mais de 12 bilhões de doses de vacinas (3).

1.2 A infecção pelo SARS-CoV-2

O SARS-CoV-2 é um membro da família *Coronaviridae*, fazendo parte do gênero Betacoronavírus, juntamente com o SARS-CoV e o MERS-CoV, dois vírus altamente patogênicos, devido à relação com síndromes respiratórias (26,27). Em 2002-2003, o SARS-CoV surgiu na China, se espalhando rapidamente pelo mundo, infectando 8000 pacientes e ocasionando 800 mortes (28). O MERS-CoV foi identificado no Oriente Médio em 2012, e se espalhou por vários países, infectando 1110 pacientes e ocasionando 422 mortes (23,29). Esses vírus são pleiomórficos e envelopados, com diâmetro de 60 a 140 nm e possuem RNA de fita

simples positiva como material genético, o qual varia de 26 a 32 kb e contém, pelo menos, seis “*Open Read Frames*” (ORF) (22).

Para que o SARS-CoV-2 infecte a célula hospedeira e se replique, são necessários diversos fatores celulares. A enzima conversora de angiotensina 2 (do inglês, *Angiotensin-converting enzyme 2*, ACE2), é o receptor canônico do domínio de ligação ao receptor da glicoproteína homotrimérica *Spike* (S), composta por duas unidades sub-funcionais, que são responsáveis pela ligação do vírus ao receptor da célula hospedeira (subunidade S1) e pela fusão do vírus com a membrana celular (24,30,31). Após a entrada do vírus na célula, a proteína S é inicializada pela *transmembrane serine protease 2* (TMPRSS2), responsável pela clivagem das subunidades S1 e S2, que permite a fusão do vírus com as membranas celulares do hospedeiro (32).

A proteína ACE2 é altamente expressa nos pneumócitos tipo I e tipo II (33,34). Há um aumento de duas a três vezes da expressão de ACE2 nas células epiteliais das vias aéreas em pacientes com a COVID-19 (35). Ao estudar o potencial tropismo do SARS-CoV-2 pela expressão gênica de genes associados à entrada do vírus na célula, através do método de sequenciamento de RNA de células únicas (scRNA-Seq, do inglês, *single-cell RNA sequencing*), Sungnak *et al.* (34) encontraram expressão enriquecida da ACE2 em células epiteliais nasais, mais especificamente em células secretoras. Em estudo de scRNA-Seq realizado por Muus *et al.* (36), demonstrou-se a co-expressão de ACE2 e TMPRSS2 em subconjuntos específicos de células do epitélio respiratório e nos alvéolos (36).

Os lipídeos também são fundamentais para o processo de infecção, podendo atuar como membranas de organelas e como componentes da membrana celular, participando da regulação do tráfego molecular transmembrana, assim como a internalização de materiais infecciosos, como vírus (35,37). Após a replicação, as partículas virais recém-sintetizadas saem da célula, atravessando a membrana rica em lipídeos (35). Diversos estudos já demonstraram o papel fundamental do colesterol na infecção do SARS-CoV-2 (38–40), onde o colesterol intracelular pode aumentar a infectividade do vírus (41).

Nosso grupo de pesquisa realizou diversas pesquisas, utilizando de bioinformática, que visavam compreender a infecção do SARS-CoV-2. Em estudo para predição de rotas não-canônicas de infecção pelo SARS-CoV-2 em células placentárias, observaram que essas células expressam altos níveis de dipeptidil peptidase-4 (DPP4) e de catepsina L (CTLS), os quais

foram preditos como mediadores não-canônicos de entrada do vírus na célula (42). Visando compreender o mecanismo pelo qual os monócitos respondem à infecção pelo SARS-CoV-2 em contexto de suprimento elevado de glicose, observaram que níveis elevados de glicose e glicólise promovem a replicação do vírus e a produção de citocinas pelos monócitos resulta na disfunção de células T e morte de células epiteliais (43). Por fim, nosso grupo de estudo observou que o gene *TRIB3*, que anteriormente já foi relatado para diminuir a infecção e a replicação viral, tem a expressão diminuída em pulmões envelhecidos, o que pode explicar o porquê de pacientes idosos estarem relacionados com casos mais graves da COVID-19 (44).

1.3 Resposta transcricional das células pulmonares à infecção pelo SARS-CoV-2

O pulmão possui uma grande variedade de mecanismos protetores, incluindo um conjunto altamente especializado de células imunes inatas e adaptativas para combater as infecções, e várias populações de células-tronco e progenitoras que fornecem ao pulmão a capacidade regenerativa após lesões (45). Na infecção do pulmão pelo SARS-CoV-2, foi demonstrado que os principais alvos do vírus são as células epiteliais das vias aéreas, células epiteliais alveolares, endoteliais vasculares e macrófagos dos pulmões, todas expressando a *ACE2* (33,46,47). Após a destruição das células pulmonares por conta da infecção, há uma resposta imune local com a presença de macrófagos e monócitos, que liberam citocinas, e resposta imune com a participação de células T e B (48). Essa resposta pode ser disfuncional em alguns casos, causando uma doença pulmonar grave e até mesmo sistêmica (48).

A resposta contra a infecção viral é geralmente iniciada no momento da replicação, no citoplasma das células (49,50). Após o reconhecimento de estruturas de RNAs aberrantes produzidas durante a replicação viral por vários *pattern recognition receptors* (PPRs), ocorre a ativação de fatores de transcrição, principalmente *interferon regulator factors* (IRFs) e *nuclear factor kB* (NF-kB) (49,51,52). Em seguida, duas respostas antivirais são liberadas: a primeira é o engajamento das defesas antivirais, mediada pela indução de transcrição de interferons tipo I e III (IFN-I e IFN-III, respectivamente) e, posteriormente, aumento da expressão de genes estimulados por IFN (49,53). A segunda resposta envolve o recrutamento e coordenação de subconjuntos específicos de leucócitos, que são organizados principalmente pela secreção de quimiocinas (49,54,55).

Em estudo que comparou a resposta transcricional do SARS-CoV-2 com outras viroses respiratórias, os autores identificaram uma resposta reduzida de INF-I e INF-III, porém com uma assinatura consistente de quimiocinas (49). Com a inflamação local, há um aumento da secreção de citocinas e quimiocinas pró-inflamatórias, como a IL-6, IFN- γ , MCP-1 e CXCL10 (15,56). Em alguns pacientes, a resposta imune pode ser anormal, desencadeando uma tempestade de citocinas (em inglês, *cytokine storm*), a qual é mediadora para uma inflamação pulmonar generalizada (48).

Nosso grupo de pesquisa demonstrou que a análise da expressão de genes do secretoma é uma estratégia que pode ser utilizada para identificar por que tipos tumorais específicos possuem uma probabilidade maior de desenvolver perda de massa muscular na caquexia associada ao câncer (57). Isso demonstra que a utilização de análises do secretoma também constituem uma ótima ferramenta para identificar potenciais biomarcadores, mediadores de severidade e alvos terapêuticos para tratamento da COVID-19.

1.4 O secretoma na infecção viral

O termo secretoma foi primeiramente usado por Tjalsma *et al.* (58) para caracterizar todos os fatores secretados por uma célula e também os constituintes da via secretória, tendo sua definição posteriormente revisada para incluir apenas proteínas secretadas no meio extracelular (59,60). Portanto, o secretoma representa o conjunto de macromoléculas secretadas por células, as quais possibilitam a comunicação celular (61,62).

As proteínas secretadas desempenham papel considerável em diversos processos biológicos importantes, sendo essenciais para a comunicação célula-célula, proliferação, metabolismo (63), e em mecanismos de defesa (64). Através dessas proteínas é possível identificar mediadores para diversas doenças, e alguns desses componentes do secretoma podem ser utilizados como alvos terapêuticos (65). O secretoma humano também pode revelar biomarcadores para serem utilizados no diagnóstico ou prognóstico e, assim, contribuir de forma significativa no tratamento de diversas doenças (66,67).

Um meio de prever o secretoma através de dados de transcriptoma, que vem sendo amplamente utilizado, é o sequenciamento de RNA (RNA-Seq). A utilização do RNA-Seq e de algoritmos emergentes para predição de sequências de peptídeos-sinais tornaram-se ferramentas favoráveis para avaliação do perfil do secretoma (68). Com base na estratégia de predição do secretoma através de dados transcriptômicos, Robinson *et al.* (69) realizaram, com

componentes do secretoma, uma análise *Pan-Cancer* de expressão gênica. Isso resultou em listas ordenadas de possíveis candidatos a biomarcadores diagnósticos detectáveis em fluidos biológicos (69). Esse estudo também revelou funções e padrões biológicos associados à alterações na expressão de proteínas secretadas em diferentes tipos tumorais (69).

Outra forma de caracterizar o transcriptoma é através do sequenciamento de RNA de células única (scRNA-Seq). Essa técnica já vem sendo usada para investigar a resposta imune e seus mecanismos associados à severidade da COVID-19 (70,71) e também já revelou associações específicas entre o tipo celular e idade, sexo e tabagismo com níveis de expressão de ACE2, TPMRSS2 e CTSL (36). Essas descobertas mostram que o scRNA-Seq tem grande potencial de inferir o secretoma na infecção pelo SARS-CoV-2.

A determinação do secretoma pode ser muito relevante pois pode elencar um perfil de moléculas secretadas pelas células infectadas pelo SARS-CoV-2, uma vez que a severidade da COVID-19 não se dá apenas pela infecção viral, mas também pela resposta do hospedeiro. Como exemplo, o perfil lipídico de pacientes pode ser considerado um biomarcador de inflamação na COVID-19 (35). A grande parte dos estudos que observaram o perfil lipídico em pacientes com COVID-19 observaram que a diminuição nos níveis de lipídeos aumenta a gravidade e mortalidade da doença (72).

O avanço em dados transcriptômicos também pode fornecer informação relevantes sobre a comunicação célula-célula. A comunicação celular tem papel fundamental em funções biológicas (73), coordenando diferentes atividades celulares, como homeostase (73), respostas imunológicas (74) e até mesmo na carcinogênese (73). O estudo da comunicação célula-célula é um tipo de aplicação de análise que pode auxiliar na elucidação de mecanismos do secretoma, através da expressão gênica dos fatores secretados em células e tecidos-alvo (75). Na COVID-19, as células epiteliais infectadas secretam quimiocinas que recrutam e ativam diversas populações de células imunes no local da infecção (76), e a expressão dessas quimiocinas e seus respectivos receptores varia de acordo com a severidade da doença (70).

Visto isso e diante da gravidade e emergência em encontrar tratamentos para a COVID-19, a descoberta de fatores secretados e relacionados a infecção do SARS-CoV-2 seria de grande relevância para a identificação de possíveis mediadores no secretoma de células do sistema respiratório que estão associadas à COVID-19, visto que mais de 6 milhões de pessoas vieram a óbito por ação direta ou indireta da doença (3). Para o nosso conhecimento, ainda não foi realizado um estudo para análise do perfil de expressão gênica do secretoma em células do

fluido do lavado bronco alveolar de pacientes com COVID-19, considerando-se as variáveis carga viral (SARS-CoV-2), idade e gênero.

2 - Objetivos

2.1 Objetivo Geral

Analisar o perfil transcricional do conjunto de proteínas secretadas pelas células do fluido do lavado bronco alveolar de pacientes com risco de desenvolver COVID-19 severa.

2.2 Objetivos Específicos

- Analisar o perfil de expressão gênica em 430 amostras do fluido do lavado bronco alveolar de pacientes com COVID-19;
- Estabelecer o perfil de expressão gênica do secretoma em amostras de pacientes com COVID-19, considerando-se as variáveis de carga viral, gênero e idade;
- Realizar a análise de enriquecimento funcional dos genes diferencialmente expressos do secretoma de pacientes com COVID-19;
- Identificar potenciais mediadores da COVID-19 a partir da análise transcricional do secretoma;
- Correlacionar o perfil de expressão de genes do secretoma de pacientes com COVID-19 com as variáveis de carga viral, gênero e idade;
- Identificar alterações de expressão através da análise temporal;
- Realizar a análise de scRNA-Seq para identificar tipos celulares presentes nas amostras analisadas e a expressão dos genes em cada um.

3 – Artigo científico

Low Density Lipoprotein Receptor Related Protein 11 (LRP11) as a protective role in the severity of COVID-19

Ana Luiza Labbate Bonaldo¹, Jakeline Santos Oliveira¹, Sarah Santiloni Cury¹, Jeferson dos Santos Souza¹, Amanda Piveta Schnepfer¹, Victória Larissa Schmidt Camargo¹, Caio Fernando Ferreira Mussato¹ & Robson Francisco Carvalho¹.

¹ Departamento de Biologia Estrutural e Funcional, Instituto de Biociências – Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP)

RESUMO

Background

A COVID-19 se espalhou rapidamente pelo mundo, fazendo milhões de vítimas fatais. Pacientes do sexo masculino, idosos, com alta carga viral e com comorbidades como diabetes e hipertensão, caracterizam um grupo de pior prognóstico para a COVID-19. A produção excessiva de mediadores inflamatórios e citocinas, conhecida como tempestade de citocinas (do inglês, *cytokine storm*), também está associada com o aumento da mortalidade. Visto isso, é de extrema importância determinar o conjunto de moléculas secretadas (secretoma) pelas células de tecidos alvos do SARS-CoV-2 de pacientes com COVID-19, a fim de identificar mediadores e biomarcadores para a doença. No presente trabalho, nós analisamos o perfil transcricional do secretoma das células do fluido do lavado bronco alveolar de pacientes com risco de desenvolver COVID-19 severa, buscando determinar os principais fatores que modulam a severidade e mortalidade nestes pacientes.

Métodos

Foram analisadas amostras de fluido do lavado bronco alveolar (BALF) de 430 pacientes positivos para COVID-19 e 54 negativos. O conjunto de dados (GSE152075; GEO) (77) de *RNA sequencing* (RNA-Seq) foi processado usando o *NetworkAnalyst* (70) para identificar os genes diferencialmente expressos (DEGs; de $\text{Log2 Fold Change} \geq |1.5|$ e $p\text{-value} \leq 0.05$) nos grupos de acordo com gênero, idade e carga viral, quando comparados com os respectivos controles (negativos). Posteriormente, os DEGs foram filtrados pela lista de proteínas secretadas do *The Human Protein Atlas* (71). Utilizando a ferramenta online *Intervene* (72), foram visualizadas intersecções de genes com alteração de expressão nos grupos analisados, de acordo com o risco para desenvolvimento de COVID-19 severa. Foi caracterizado o perfil de expressão gênica do interactoma (ligantes e respectivos receptores) em comum no grupo de melhor prognóstico (mulheres, jovens e baixa carga viral). Por fim, através de sequenciamento de RNA célula única (scRNA-Seq), usando a linguagem R com o pacote Seurat (v.3), foi analisada a expressão gênica dos genes identificados ao nível de célula única do lavado bronco alveolar. Foi utilizado o conjunto de dados GSE145926 (GEO), com 12 amostras de BALF, sendo 6 pacientes severos, 3 moderados e 6 saudáveis, e também foi adicionada uma amostra de BALF saudável (GSM3660650) do conjunto de dados GSE128033 (GEO). Para validação dos achados, foi feito uma análise temporal do BALF de 9 macacos rhesus adultos infectados com SARS-CoV-2 nos dias 1, 2, 4, 7, 10 e 12 (GSE156701).

Resultados

Nossos resultados demonstraram um perfil de expressão dos genes do secretoma distinto para cada comparação analisada, de acordo com o risco de severidade para o desenvolvimento da COVID-19. O enriquecimento funcional das vias dos genes do secretoma indicaram alteração em vias relacionadas à resposta imune e inflamação, onde o grupo de alta carga viral apresentou genes com expressão aumentada para essas vias. Por outro lado, o grupo mulheres mostrou uma diminuição na expressão dos genes nessas mesmas vias. O grupo de melhor prognóstico (mulheres, jovens e baixa carga viral) apresentam alteração na expressão de 43 genes do secretoma compartilhados entre si. O gene *LRP11*, que codifica o receptor de mesmo nome, apresenta diminuição de expressão no grupo de melhor prognóstico. Esse receptor faz parte da família dos receptores de LDL, molécula importante para a entrada do SARS-CoV-2 nas células. Através das análises de scRNA-Seq, foi observado que esse gene é expresso principalmente em células epiteliais.

Conclusão

Nossos resultados demonstram que a diminuição da expressão gênica do receptor *LRP11* tem um potencial efeito protetivo contra a infecção pelo SARS-CoV-2 e, consequentemente, diminuição da severidade da doença. Também observamos que a expressão do gene *LRP11* ocorre principalmente majoritariamente em células epiteliais, as quais tem uma alta expressão de ACE2, um fator celular importante na infecção pelo vírus SARS-CoV-2.

Palavras-chave: Infecção pelo SARS-CoV-2, COVID-19 severa, Secretoma, *LRP11*, scRNA-Seq

INTRODUÇÃO

A pandemia de COVID-19, que teve seu primeiro caso confirmado em Wuhan, na China (26), se alastrou rapidamente pelo planeta, fazendo mais de 6 milhões de vítimas fatais (3). O espectro clínico dos pacientes pode variar de assintomáticos até casos graves, com evolução para o óbito (26,78). Além de algumas comorbidades já estudadas, como hipertensão, doenças cardiovasculares e diabetes (16,79), pacientes idosos (acima de 60 anos) (80–82), do sexo masculino (77,80) e com carga viral alta (18,77,83) tem risco aumentado de severidade e mortalidade (84).

O SARS-CoV-2 necessita de fatores celulares do hospedeiro para infecção, como a enzima ACE2 (enzima conversora de angiotensina 2) (85). Essa enzima encontra-se abundantemente expressa no epitélio do pulmão (33). Diferentes estudos já demonstraram um aumento na expressão dessa enzima em células epiteliais, com destaque para as do epitélio respiratório, em pacientes positivos para COVID-19 (34,36,70). Após a morte das células pulmonares pela infecção, há uma resposta imune local com a presença de macrófagos e monócitos, os quais são responsáveis pela secreção de citocinas (48). Alguns pacientes respondem à essa infecção com uma produção excessiva de citocinas pró-inflamatórias, evento conhecido como *cytokine storm*, acarretando em uma inflamação pulmonar generalizada (86).

Considerando-se esses aspectos da infecção, a determinação do secretoma, que é o conjunto das macromoléculas secretadas pelas células (58), tem grande potencial de revelar biomarcadores, mediadores e potenciais alvos terapêuticos para a COVID-19. Um das formas de análise do secretoma é o sequenciamento de RNA (RNA-Seq) (68), que é um método de alto rendimento para analisar o perfil transcricional de um organismo ou de células individuais (70,87,88). No presente estudo, nós analisamos dados de RNA-Seq de 430 pacientes positivos e 54 pacientes negativos para a COVID-19. Nossa hipótese é que o perfil de expressão gênica desses pacientes poderia revelar potenciais mediadores e biomarcadores da doença. Os dados foram validados utilizando a análise de scRNA-Seq, usando o *conjunto de dados* GSE145926 (89), com 12 amostras, sendo 6 de pacientes severos, 3 moderados e 6 saudáveis, com o adicional da amostra de BALF saudável (GSM3660650) do *conjunto de dados* GSE128033 (90). Por fim, realizamos uma análise temporal do BALF de 9 macacos rhesus adultos infectados com SARS-CoV-2 nos dias 1, 2, 4, 7, 10 e 12 (GSE156701) (91).

MÉTODOS

Seleção de dados e análise da expressão diferencial dos dados de RNA-Seq

Foi selecionado um estudo de RNA-Seq de *swabs* nasofaríngeos de 430 pacientes positivos para COVID-19 e 54 pacientes negativos (GSE152075) (78). Esses dados foram coletados no banco de dados GEO (do inglês, *Gene Expression Omnibus*) (87). Foi utilizada a plataforma *NetworkAnalyst* 3.0 (70,88–92) para identificação dos genes diferencialmente expressos (do inglês *Differentially expressed genes* - DEGs; seguindo os parâmetros de *Log2 Fold Change* $\geq |1.5|$ e *p-value* ≤ 0.05). Para análise, foi utilizado *Log2-counts per million* como

forma de normalização dos dados, e *Limma*, como método estatístico, conforme a *pipeline* da plataforma *NetworkAnalyst* 3.0. Os indivíduos analisados foram separados por idade, segundo a classificação da Organização Mundial da Saúde (jovens, < 60 anos; e idosos, > 60 anos), gênero (mulheres e homens) e carga viral, (alta, N1 Ct < 24; e baixa, N1 Ct > 24), de acordo com os valores *cycle threshold* (Ct) das análises de RT-qPCR (93). Cada grupo foi analisado com suas respectivas amostras negativas, e as comparações realizadas em cada análise e a nova nomenclatura utilizada estão apresentadas na **Tabela 1**.

Tabela 1. Comparações para cada análise de expressão diferencial e as novas nomenclaturas, com seus respectivos números de pacientes (n).

Comparações entre os grupos experimentais	Denominação
COVID-19 (n = 430) x Pacientes negativos (n = 54)	COVID-19
Mulheres positivas (n = 201) x Mulheres negativas (n = 30)	Mulheres
Homens positivos (n = 176) x Homens negativos (n = 24)	Homens
Jovens positivos (n = 241) x Homens negativos (n = 39)	Jovens
Idosos positivos (n = 172) x Idosos negativos (n = 15)	Idosos
Baixa carga viral (n = 99) x Pacientes negativos (n = 54)	Baixa carga viral
Alta carga viral (n = 108) x Pacientes negativos (n = 54)	Alta carga viral

A análise do secretoma humano foi realizada filtrando as listas de DEGs obtidas em cada comparação (mulheres, homens, jovens, idosos, baixa carga viral e alta carga viral), pela lista de genes do secretoma disponível no *The Human Protein Atlas* (71). As proteínas do secretoma humano foram preditas por uma varredura de proteoma usando pelo menos dois de três métodos a seguir para a predição do peptídeo sinal: SignalP4.0, Phobius e SPOCTOPUS.

Para a análise temporal do perfil de expressão gênica de alvos selecionados após a infecção pelo SARS-CoV-2, foram utilizados os dados do estudo de Aid *et al.* (94) (GSE156701, disponível no GEO). Esse estudo possui dados de transcriptoma do BALF coletados nos dias 1, 2, 4, 7, 10 e 14 de nove macacos rhesus adultos após a infecção com SARS-CoV-2 (94). Os dados de RNA-Seq foram gerados pela plataforma *Illumina NextSeq*

6000, como descrito no trabalho original (94). Para identificação dos DEGs, foi empregada a mesma metodologia utilizada para análise do GSE152075.

Análise do enriquecimento funcional

Para a análise do enriquecimento funcional, foi utilizada a ferramenta on-line *EnrichR* (95–98). Essa ferramenta foi usada na lista de DEGs filtrados pelo secretoma. Os bancos de dados *WikiPathways* e *GO Biological Process* foram usados para determinação de vias e ontologias enriquecidas. Os valores de $p\text{-adj} < 0.05$ foram classificados do maior para o menor, e considerados os 5 melhores pontuados (top 5) para a lista de DEGs filtrados pelo secretoma em cada condição.

Análise de interactoma utilizando dados de RNA-Seq

Foi usado o conjunto de potenciais interações ligante-receptor da IUPHAR (do inglês, *International Union of Pharmacology*) (99) e o banco de dados de conectoma, descrito previamente por Ramilowski et al. (100), que incluem 695 ligantes, 652 receptores e 2695 interações ligante-receptor. Posteriormente, foram utilizadas estratégias de análise descritas no estudo de Armingol et al. (101) para analisar essas interações e aplicá-las em diferentes contextos. Foram consideradas as interações positivas dos 43 genes exclusivos do grupo de melhor prognóstico. Para o cálculo da interação ligante-receptor, consideramos a função $\text{Ligante} \times 0,33$ e $\text{Receptor} \times 0,33$ pelo valor de $\text{Log}_{10}P_{adj} < 0,05$, onde 1 para pontuação de comunicação foi considerado como positivo.

Rede de interação proteína-proteína (PPI)

As proteínas identificadas no interactoma foram submetidas à base de dados STRING (do inglês, *Search Tool for Retrieval of Interacting Genes*, versão 11.5) (102) para a construção de uma rede de interação proteína-proteína de genes identificados como potencialmente associados ao prognóstico da doença. Para a construção da rede, consideramos experimentos, bancos de dados, co-expressão, vizinhança, e co-ocorrência como fontes de interações ativas. A pontuação mínima de interação exigida foi 0,700 (alta confiança), e os nós desconectados da rede foram ocultados para simplificar a exibição da rede PPI. O valor de p de enriquecimento de PPI indica a significância estatística fornecida pelo *STRING*. Para visualizar a PPI,

utilizamos o software *Cytoscape* v3.8.2 (*Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks*).

Análise e representação dos dados

Os gráficos de barra foram construídos com o *GRAPHPAD PRISM* (*GRAPHPAD Software* v8). A análise de Componentes Principais (do inglês, *Principal Component Analysis* - PCA) foi construída pela ferramenta on-line *ClustVis* (biit.cs.ut.ee/clustvis) (103). Os genes com aumento e diminuição de expressão foram visualizados com *Volcano plots* gerados com a ferramenta *VolcanoR* (huygens.science.uva.nl/VolcanoR) (104). O gráfico *UpSet*, para visualização das intersecções de genes com alteração de expressão entre as comparações, foi obtido com a ferramenta *Intervene* (intervene.shinyapps.io/intervene) Os *heatmaps* foram gerados com o software *Morpheus* (software.broadinstitute.org/morpheus) (109). O *heatmap* com *k-means* foi utilizado como método de agrupamento de *clusters*.

Obtenção dos dados de scRNA-Seq de pacientes com COVID-19

O conjunto de dados de sequenciamento de células únicas (GSE145926) selecionado contém 12 amostras de BALF: 6 pacientes severos e 3 pacientes moderados com COVID-19, e 3 controles saudáveis. A gravidade da doença dos pacientes foi definida como moderada, grave e severa, de acordo com o "Protocolo de Diagnóstico e Tratamento de COVID-19" pela Comissão Nacional de Saúde da China, emitido em 3 de março de 2020, como descrito em Liao *et al.* (105). Também foi adicionada uma amostra de BALF saudável (GSM3660650) do conjunto de dados GSE128033, totalizando 13 amostras com dados de sequenciamento de RNA de células únicas.

Análise dos dados de scRNA-Seq

A demultiplexação das amostras dos genes do cluster A (*LTF*, *MST1*, *SLIT1*, *GPR162*, *LRP1*, *LRP11*, *MST1R*, *GPC1* e *SDC1*), o processamento do código de barras e a contagem do identificador molecular exclusivo 5' de célula única (UMI, do inglês, *unique molecular identifier*) foram realizadas utilizando o *Cell Ranger Software Suite* (v.3.1.0) (106). O alinhamento das leituras de SARS-CoV-2 e do transcriptoma em células individuais foi realizado utilizando a ferramenta *STAR* (107), a partir do genoma customizado integrando os genomas humano (GRCh38) e total do vírus (*GenBank*: MN908947.3). A análise de células

únicas foi feita por meio de linguagem R, com o pacote de desenvolvimento para análise de células únicas *Seurat* (v.3) (108). Para identificar os agrupamentos de células, foi efetuada a análise estatística UMAP (do inglês, *uniform manifold approximation and projection*). Para a avaliação do perfil de expressão gênica, foram utilizados os gráficos *dotplot* e *violinplot*. As figuras foram geradas utilizando as funções *DotPlot* e *VlnPlot*, respectivamente, do pacote *Seurat* (v.3).

RESULTADOS

Os genes do secretoma revelam perfis de expressão distintos em pacientes positivos com diferentes susceptibilidades para COVID-19 severa

Nós reanalisamos os dados de RNA-Seq de amostras de *swabs* nasofaríngeos de 430 pacientes com COVID-19 e 54 pacientes saudáveis, descritos em Lieberman *et al.* (77), para caracterização dos genes diferencialmente expressos (DEGs) (Tabela Suplementar S1). Nós dividimos as amostras dos pacientes de acordo com o sexo (mulheres e homens), idade (jovens, < 60 anos; e idosos, > 60 anos) e carga viral (alta, N1 Ct < 24; e baixa, N1 Ct > 24) (Figura Suplementar 1A). Nossas listas de DEGs de cada comparação foram filtradas pela lista de 2641 proteínas secretadas disponível no *The Human Protein Atlas* (Tabela Suplementar S2). Nós identificamos um total de 927 genes do secretoma diferencialmente expressos em pelo menos uma das comparações analisadas (COVID-19, mulheres, homens, jovens, idosos, baixa carga viral e alta carga viral) ($\log_2 \text{Fold Change} \geq 1.5$ e $p\text{-adj} \leq 0.05$) (Tabelas Suplementares S3-S9). O *heatmap* e as análises de agrupamento *k-means* revelaram perfis distintos de expressão gênica para cada grupo avaliado, em relação ao seu respectivo controle (amostra negativa) (Figura Suplementar 1B).

A análise de componentes principais (PCA) revelou que homens, idosos, baixa carga viral e alta carga viral distinguem-se do conjunto composto por COVID-19, mulheres e jovens (Figura 1A). A comparação alta carga viral apresentou uma maior amplitude de *FoldChange* ($\text{FoldChange} > 1.5$, corrigido por *Z-Score*) do perfil secretor (Figura 1A). Nós obtivemos 257, 177, 87, 281, 33, 67 e 375 genes do secretoma com expressão aumentada e 301, 270, 120, 357, 110, 503 e 230 genes do secretoma com expressão diminuída nos grupos de COVID-19, mulheres, homens, jovens, idosos, baixa carga viral e alta carga viral, respectivamente (**Tabela 2**). Em relação ao número total de genes do secretoma com alteração de expressão, observamos

que homens e idosos possuem uma menor quantidade de genes com alteração na expressão em relação às demais comparações (Figura 1B). Através de *Volcano Plots*, encontramos os top 10 DEGs com maior variação de expressão em cada uma das comparações (Figura Suplementar 2A-G).

Tabela 2. Número total e de cada grupo analisado de DEGs com aumento e diminuição de expressão.

Grupos analisados	Aumento de expressão (n total = 1277)	Diminuição de expressão (n total = 1891)
COVID-19	257	301
Mulheres	177	270
Homens	87	120
Jovens	281	357
Idosos	33	110
Baixa carga viral	67	503
Alta carga viral	375	230

Genes do secretoma com potencial papel em vias de inflamação e resposta imune para a infecção por SARS-CoV-2 em grupos de risco para a COVID-19 severa

Os DEGs das comparações de idade, gênero e carga viral foram selecionados para análise de enriquecimento funcional, usando as bibliotecas *WikiPathways* (Figura 1C) e *GO Biological Process* (Figura 1D). Observamos o enriquecimento de vias e termos ontológicos relacionados à resposta imunológica e à inflamação (Figura 1C-D). Mulheres e pacientes com baixa carga viral apresentam uma menor expressão de genes associados à resposta imunológica e à inflamação (Figura 1C-D). Pacientes com alta carga viral aumentam a expressão de vias e termos ontológicos relacionados a resposta de citocinas, como *cytokine-mediated signaling pathway*, *chemokine signaling pathway* e *cellular response to cytokine stimulus* (Figura 1C-D), e em vias de resposta imunológica e inflamatória, como *SARS-CoV-2 innate immunity evasion and cell-specific immune response*, *inflammatory response*, *complement activation and regulation of T cell proliferation* (Figura 1C-D).

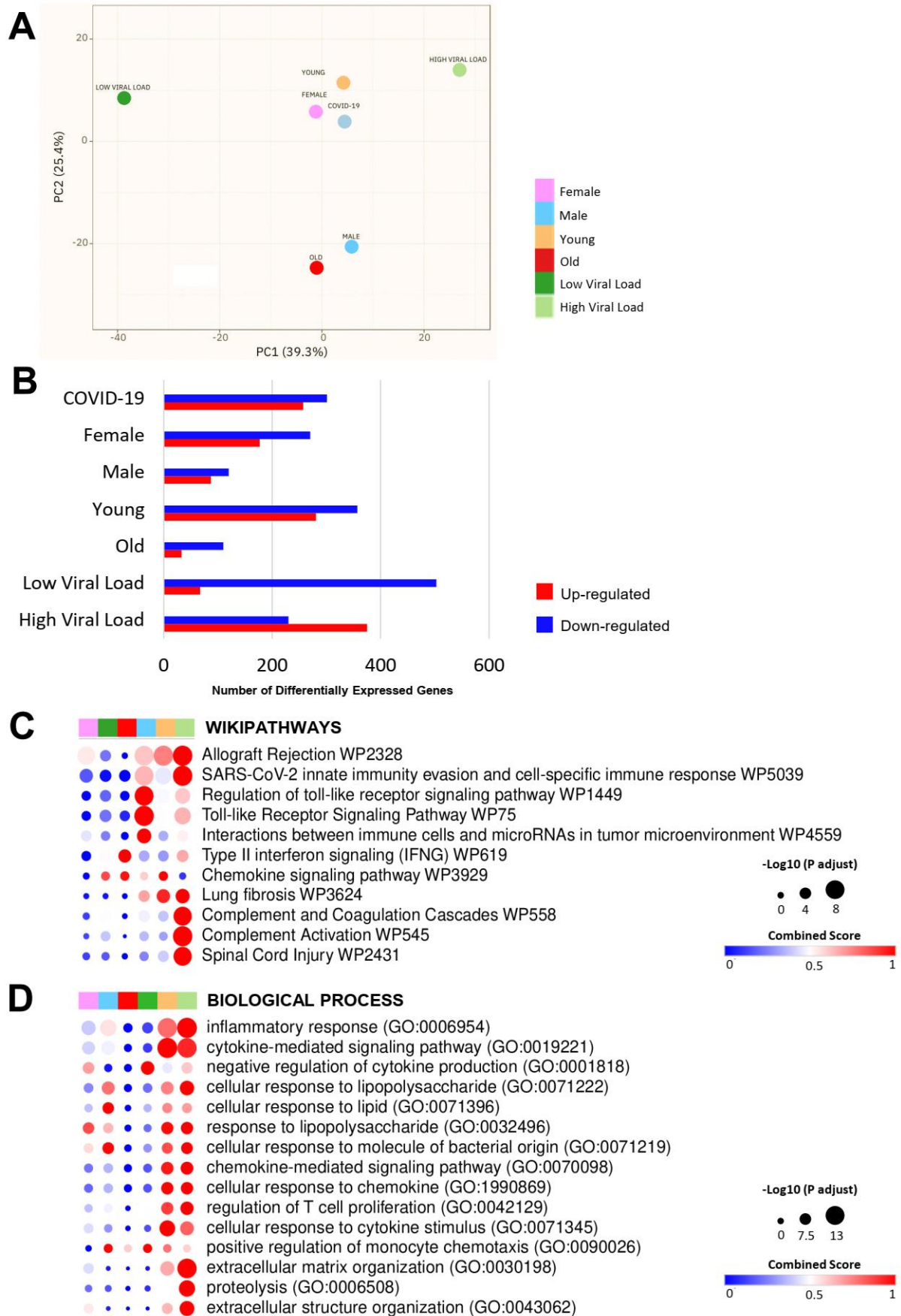


Figura 1. Análise da expressão gênica do secretoma do BALF de pacientes nas comparações estudadas (idade, gênero e carga viral). (A) Análise dos componentes principais (PCA) dos dados de expressão dos genes do secretoma das 6 comparações. (B) Caracterização do número total de DEGs com aumento e diminuição de expressão em cada comparação. (C) *Bubble heatmap* das 5 vias melhores pontuadas (top 5) no *combined score* da biblioteca *WikiPathways* e (D) da biblioteca *GO Biological Process*. O tamanho da bolha e a intensidade da cor são representados, respectivamente, pelo valor de p ajustado transformado em $-\log_{10}$ e pelo valor de *combined score*.

Análise do interactoma dos 43 genes revelam diferenças nos perfis de expressão entre os

Calculamos a sobreposição de genes do secretoma entre as comparações analisadas através da ferramenta *UpSet Plot*. O grupo de melhor prognóstico para a COVID-19, que é composto por mulheres, jovens e baixa carga viral, possui 43 genes exclusivos (Figura 2A). Esses genes estão caracterizados na Tabela Suplementar S10. Encontramos que os 43 genes possuem perfis de expressão gênica diferentes entre si em relação ao grupo de melhor prognóstico e ao grupo de pior prognóstico. Dentre esse conjunto de 43 genes, o grupo de melhor prognóstico demonstra um perfil com mais genes com diminuição de expressão, enquanto o grupo de pior prognóstico apresenta um perfil com menos genes apresentando diminuição da expressão (Figura 2B). Em pacientes com baixa carga viral, alguns genes possuem maior significância estatística, como o *MYDGF*, *HSD11B1L*, *SLIT1*, *MANF*, *B3GNT9*, *SPON2* e *C4B* (Figura 2B). Partindo do pressuposto de co-expressão para identificação de genes relevantes para a doença, fizemos uma rede ligante-receptor para potenciais interações encontradas nesse grupo de 43 genes (Figura 2C), para identificar, além das moléculas secretadas, seus respectivos receptores. Observamos que os perfis de expressão dos pares de ligantes-receptores encontrados distinguem-se entre os grupos de melhor e pior prognóstico. (Figura 2D). O agrupamento por *Euclidian distance* nos retornou 3 *clusters*. Os pares de ligantes-receptores do *cluster A* (*LTF_GPR162*, *LTF_LRP1*, *LFT_LRP11*, *MST1_MST1R*, *SLIT1_GPC1* e *SLIT1_SDC1*) encontram-se com menor expressão no grupo de melhor prognóstico em comparação ao grupo de pior prognóstico (Figura 2D). Os pares dos grupos de melhor e pior prognóstico do *cluster B* (*EFEMP2_PLSCR4*, *LTF_TFRC*, *CALM3_GRM5*, *TIMP1_CD63*, *ZG16B_TLR5*, *NTS_NGFR*, *FN1_IL17RC*, *SLIT1_ROBO1*, *NTS_SORT1* e *SPON2_ITGB*) possuem expressão heterogênea entre si (Figura 2D). Por fim, observamos que os pares do *cluster C* (*C4B_CR1*, *SPON2_ITGAM*, *ZG16B_TLR6*, *ZG16B_TLR2* e *ZG16B_TLR4*) têm sua expressão aumentada no grupo de melhor prognóstico

(Figura 2D). Ao realizar a análise funcional dos pares de ligantes-receptores, vimos que a lactoferrina (*LTF*), gene presente no *cluster* A, possui atividade antiviral (Tabela Suplementar S10), podendo assim ter relação com a severidade da COVID-19. Visto isso, selecionamos o *cluster* A para análises posteriores.

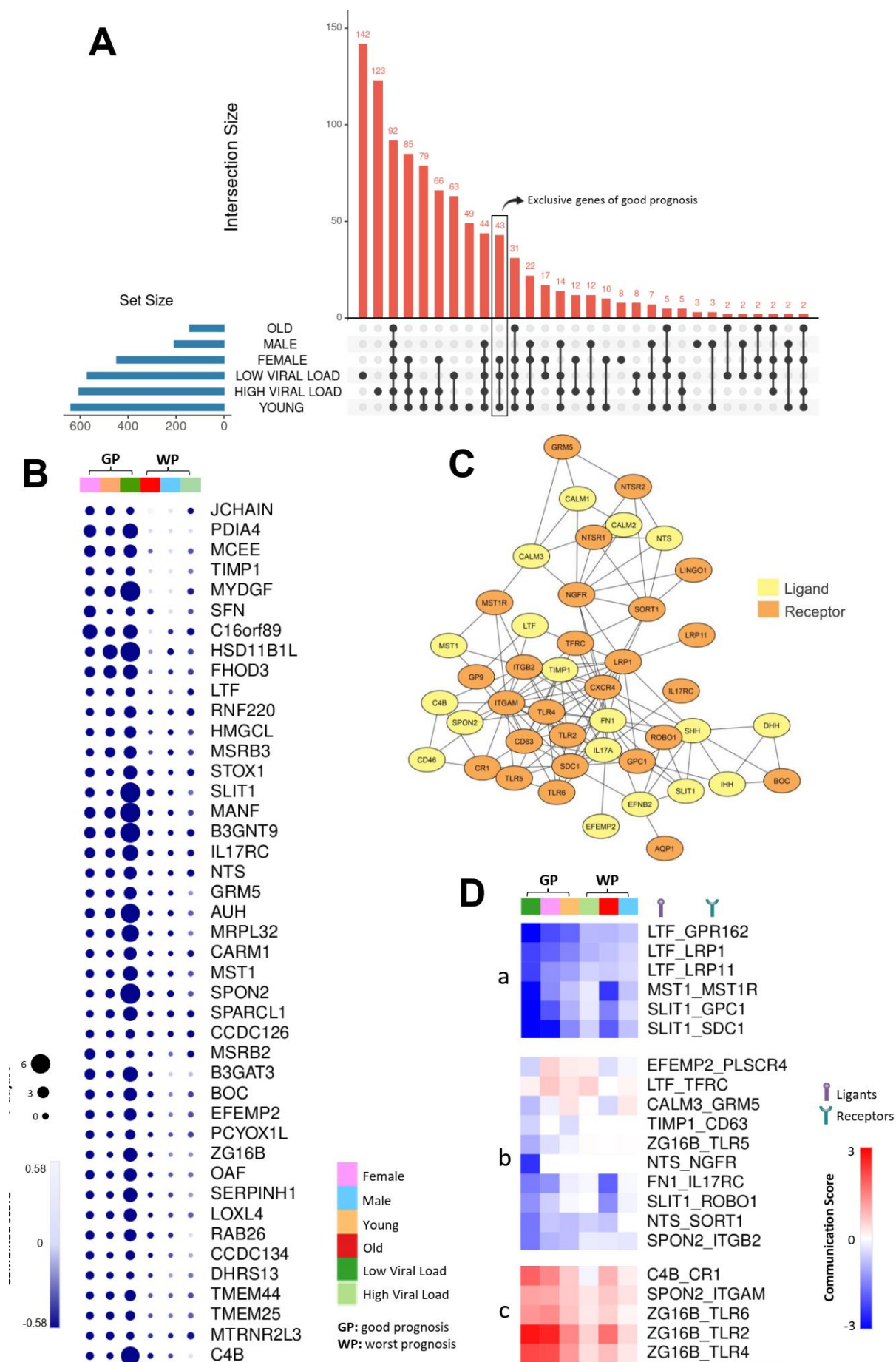


Figura 2. Caracterização do grupo de 43 genes exclusivos em comum no grupo de melhor prognóstico (mulheres, jovens e baixa carga viral). (A) *UpSet plot* dos DEGs do secretoma do BALF compartilhados entre as comparações. (B) *Bubble heatmap* dos 43 genes exclusivos das comparações de bom prognóstico. A intensidade da cor das bolhas representa o valor de *p* ajustado transformado em $-\log_{10}$, enquanto que o tamanho da bolha representa o valor de *combined Score*. A clusterização foi feita através de *Euclidian distance*. (C) Interactoma dos 43 genes. (D) *Heatmap* dos pares de ligantes e receptores do interactoma dos 43 genes. A cor azul representa os genes com expressão diminuída, enquanto que a cor vermelha, os genes com aumento de expressão. Quanto mais escura a cor, maior o valor de *combined score* do par de ligante-receptor. A clusterização foi feita através de *Euclidian distance*.

Relacionamos os níveis de expressão dos genes de ligantes (*LTF*, *MST1* e *SLIT1*) e receptores (*GPR162*, *LRP1*, *LRP11*, *MST1R*, *GPC1* e *SDC1*) presentes no cluster A com a carga viral do secretoma através da análise de scRNA-Seq, utilizando o conjunto de dados GSE145926, com adição de uma amostra saudável (GSM36600650), disponíveis no GEO (Figura 3). O gene *LRP11* tem um nível de expressão aumentado em pacientes com a carga viral alta se comparado com a expressão em pacientes com baixa carga viral, onde o nível é diminuído (Figura 3).

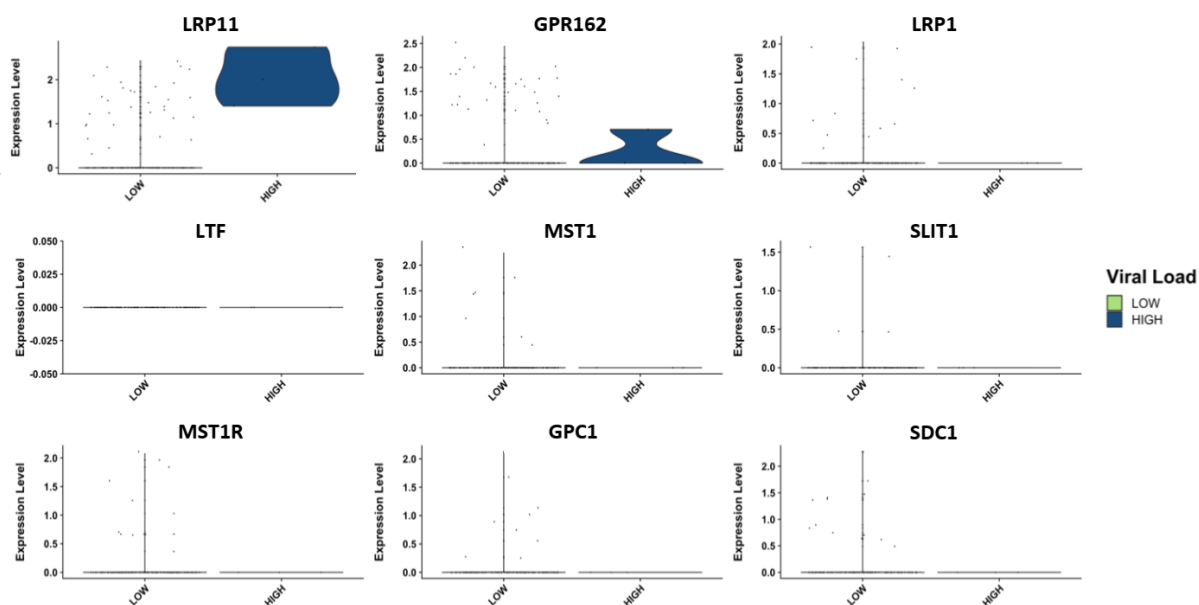


Figura 3. Violin plot demonstrando os níveis de expressão dos ligantes e receptores presentes no cluster A usando os dados do conjunto de dados GSE145926 com adicional da amostra de BALF saudável GSM3660650 do conjunto de dados GSE128033. Os genes *LTF* e *SLIT1*, que estão presentes em mais de um par de ligante-receptor, foram representados apenas uma vez.

Para confirmar esses nossos achados, correlacionamos os níveis de expressão gênica desses 3 ligantes e 6 receptores com a carga viral de SARS-CoV-2, a partir da reanálise dos dados de Aid *et al.* (GSE156701) (91), que avaliou a ruptura endotelial e a trombose vascular. Para confirmar esses nossos achados, correlacionamos os níveis de expressão gênica desses 3 ligantes e 6 receptores com a carga viral de SARS-CoV-2, a partir da reanálise dos dados de Aid *et al.* (GSE156701) (91), que avaliou a ruptura endotelial e a trombose vascular em macacos rhesus infectados com SARS-CoV-2 (91). Observamos que o gene *LRP11* aumenta seus níveis de expressão com a progressão dos dias pós-infecção (dias 1, 2, 4, 7, 10 e 14) e aumento da carga viral (Figura 4). Dentre os genes de ligantes e receptores do cluster A, o gene *LRP11* foi a única que demonstrou esse tipo de perfil variável durante os dias avaliados, tendo seus níveis de expressão aumentando de acordo com os dias avaliados. (Figura 4).

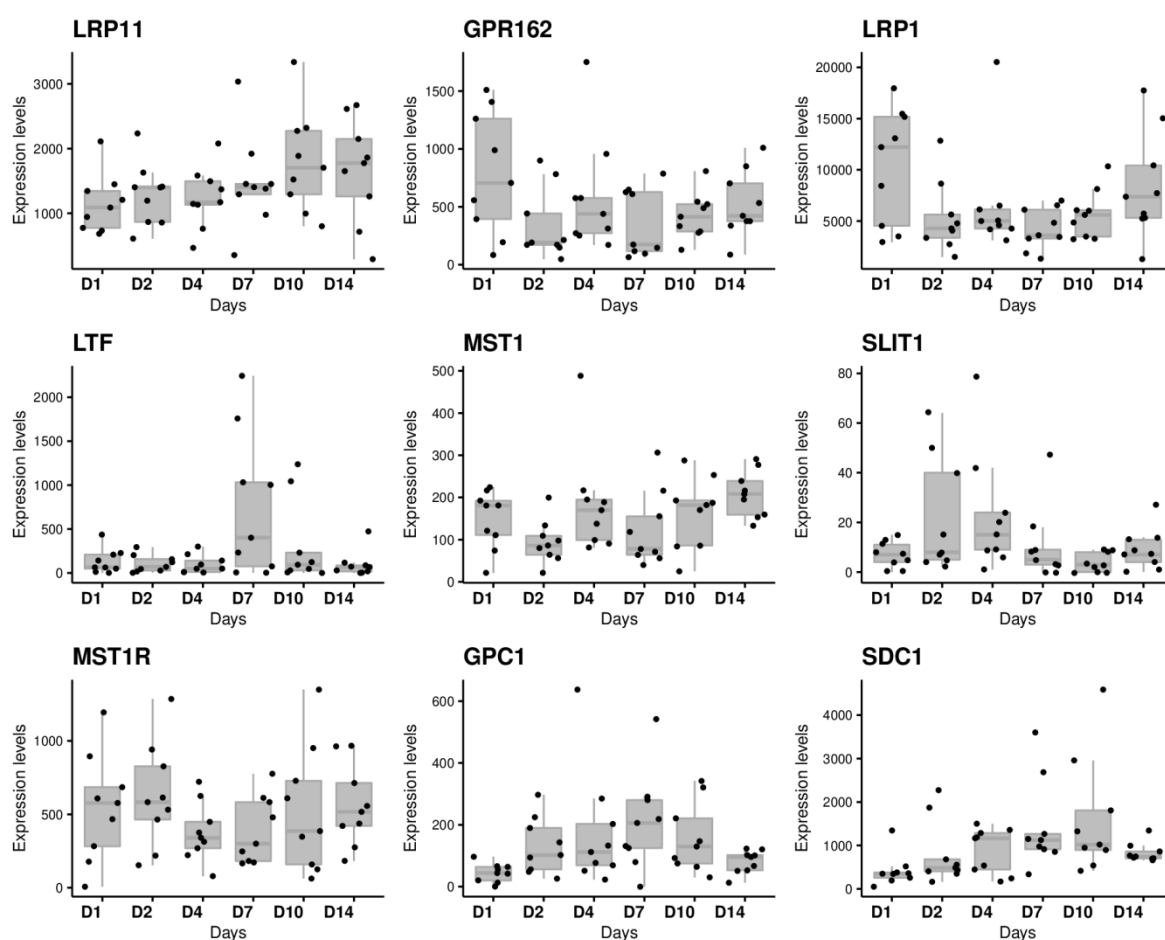


Figura 4. Box plot do time course dos níveis de expressão dos ligantes e receptores presentes no *cluster A* nas amostras de BALF de nove macacos rhesus adultos infectados com SARS-CoV-2 (GSE156701). As amostras de BALF de cada macaco infectado foram coletadas nos dias 1, 2, 4, 7, 10 e 14 após a infecção pelo vírus. Os genes *LTF* e *SLIT1*, que estão presentes em mais de um par de ligante-receptor, foram representados apenas uma vez. D1: dia 1; D2: dia 2; D4: dia 4; D7: dia 7; D10: dia 10; e D14: dia 14.

Expressão variável de LRP11 em células epiteliais de BALF

Para revelar a expressão de cada ligante e receptor do cluster A nos tipos celulares presentes no BALF, foi feita a análise de scRNA-Seq utilizando 12 amostras de BALF, sendo 6 pacientes com COVID-19 severa, 3 pacientes moderados e 6 saudáveis, com adicional de uma amostra de BALF saudável. O gene *GPR162* também se encontra expressa em células epiteliais, enquanto o *LRP1* tem sua maior expressão em macrófagos (Figura 5). Esse mesmo resultado pode ser visualizado na Figura 6, onde através de um *dotplot*, observamos que a *LRP11* possui um percentual de expressão e uma expressão média predominantemente em células epiteliais. Comparando com a expressão dos genes que codificam os 13 genes pertencentes à família dos receptores de LDL, a *LRP11* se destaca das demais em questão de expressão média e porcentagem de expressão (Figura Suplementar 4).

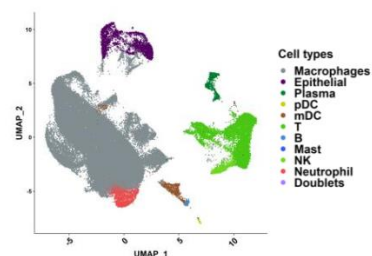
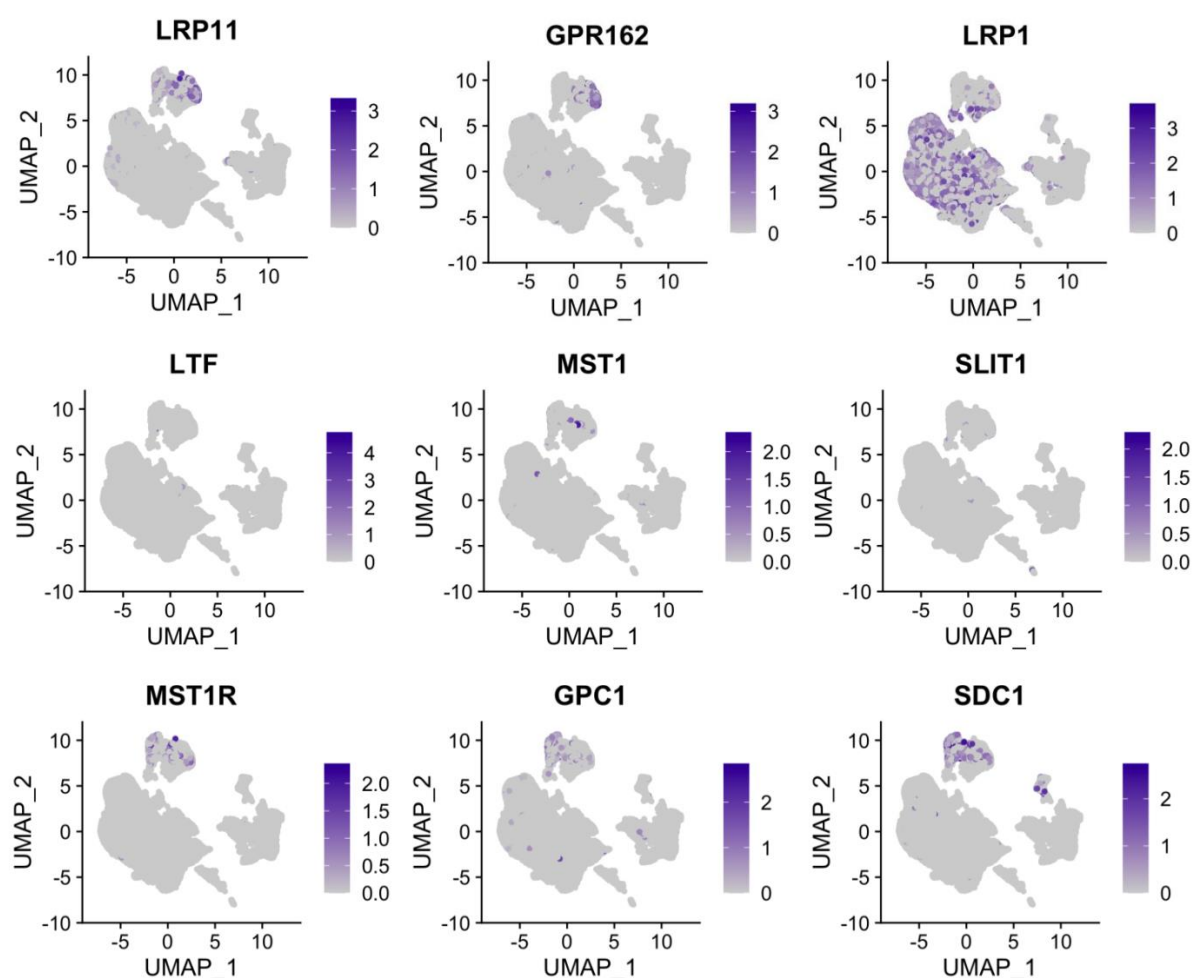


Figura 5. Gráfico UMAP representando a distribuição de expressão dos ligantes e receptores do *cluster A* nos tipos celulares presentes no BALF, utilizando os dados do *conjunto de dados* GSE145926 com adicional da amostra de BALF saudável GSM3660650 do *conjunto de dados* GSE128033. Quanto mais escuro o roxo, maior a expressão do gene no tipo celular correspondente àquela zona. Na legenda da figura encontram-se os respectivos tipos celulares, determinados pelas cores, categorizados em 11 *clusters*. Todas as amostras foram analisadas utilizando o pacote Seurat (v. 3) da linguagem R.

Através das análises baseadas em *scRNA-Seq*, pudemos prever em quais tipos celulares do fluido do lavado bronco alveolar nosso grupo de genes está expresso. A *LRP11* é expressa principalmente em células epiteliais (Figura 5). 70% de sua expressão ocorre em células epiteliais, 22% em macrófagos, 6% em células T, e apenas 1% da expressão em células *natural killer* (NK) e células dendríticas mieloides (mDC) (Figura Suplementar 3).

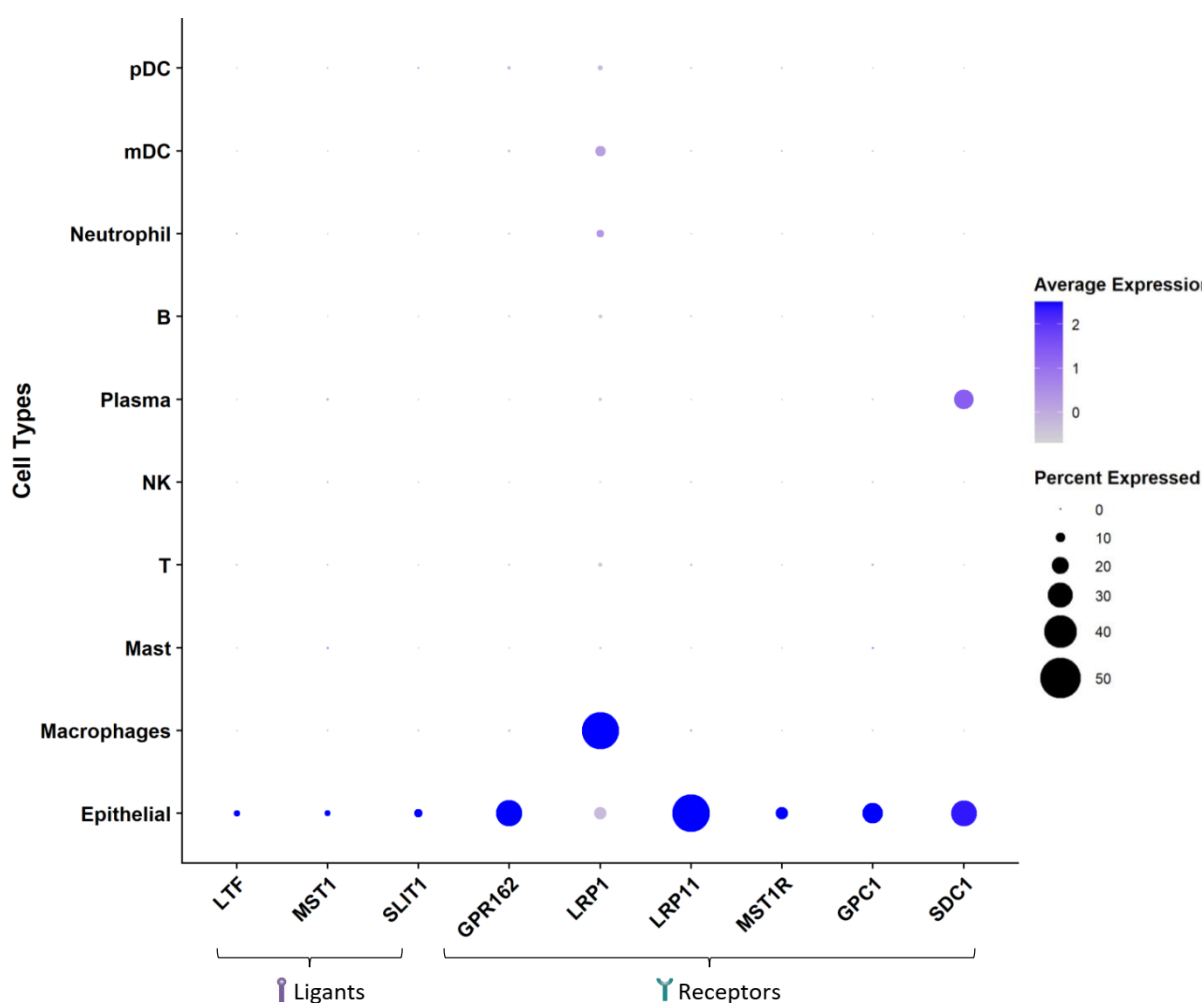


Figura 6. Dotplot representando a expressão de cada ligante e receptor presentes no *cluster A* em cada tipo celular encontrado no BALF. A intensidade da cor e o tamanho da bola são, respectivamente, a expressão média de cada gene e o percentual expresso em cada tipo celular. Para essa representação foram utilizados os dados do *conjunto de dados* GSE145926 com adicional da amostra saudável de BALF (GSM3660650) do *conjunto de dados* GSE128033, utilizando o pacote Seurat (v. 3) da linguagem R.

DISCUSSÃO

Nesse estudo, nós utilizamos análises de bioinformática como método de investigação para identificar alterações em genes presentes no secretoma humano que pudessem prever o prognóstico de pacientes positivos para COVID-19. Para as análises, usamos os GSE152075, GSE145926, uma amostra de BALF saudável (GSM36600650) do *conjunto de dados* GSE120833 e o GSE156701, disponíveis no banco de dados GEO (92). Essa estratégia revelou perfis de expressão do secretoma diferentes entre as variáveis de sexo, idade e carga viral.

Pacientes do sexo masculino e com idade acima de 60 anos estão associados a um pior prognóstico para a COVID-19 (49,80). O estudo de Freire *et al.* (57) demonstrou que mulheres e jovens possuem uma modulação inversa de citocinas e quimiocinas e da transcrição de neutrófilos, sugerindo uma melhor resposta imunorregulatória nesses dois grupos de pacientes (57). Nossas análises de enriquecimento funcional revelaram genes com expressão diminuída em vias relacionadas à resposta imunológica e inflamação em pacientes do sexo feminino. Outra variável relacionada ao pior prognóstico é a carga viral de SARS-CoV-2, onde pacientes com uma carga viral alta tem aumento da severidade e mortalidade da COVID-19 (18,83). Em nosso estudo, os genes relacionados às vias de sinalização de citocinas e quimiocinas, resposta inflamatória e regulação da proliferação de células T encontraram-se com aumento de expressão em pacientes com alta carga viral. Foram detectados em estudos anteriores que altos níveis de expressão de interleucina-1B (IL-1B), CXCL10, interferon- γ (IFN- γ) e proteína quimioatrativa de monócitos 1 (MCP-1), também conhecida como CCL2, em pacientes com COVID-19 (93). Essa produção excessiva de moléculas pró-inflamatórias é chamada *cytokine storm* (94), responsável por acarretar uma hiper inflamação pulmonar generalizada (86).

Inicialmente, hipotetizamos identificar genes relacionados a um pior prognóstico. Porém, nossos dados identificaram um conjunto de genes que, tendo sua expressão diminuída, podem oferecer um fator protetivo contra a infecção pelo SARS-CoV-2. Esses genes foram identificados ao analisarmos ligantes e seus respectivos receptores, com diferença estatística apenas no grupo de melhor prognóstico, ou seja, pacientes mulheres, jovens e com baixa carga viral. Isso revela a importância de examinar não somente as moléculas secretadas pelas células, mas também para seus receptores. Os genes em questão são *LTF*, *MST1*, *SLIT1* (ligantes), e *GPR162*, *LRP1*, *LRP11*, *MST1R*, *GPC1* e *SDC1* (receptores).

Os genes *LRP1* e *LRP11* são membros da família de receptores de lipoproteína de baixa densidade (LDL), juntamente com mais 12 outros genes (95). Em consequência da infecção por patógenos, pode haver alterações na homeostase de lipoproteína, bem como na inflamação

decorrente dessa infecção (96). A maioria das infecções, sendo a mais investigada a infecção pelo HIV, estão associadas com níveis reduzidos de lipoproteínas plasmáticas (96). Na COVID-19, há o aumento da angiotensina 2 devido a disfunção da ACE2, o que causa o aumento dos níveis de LDLR (97). Em estudo de Pillay *et al.* (98), em que os autores visaram encontrar proteínas essenciais para a infecção por vírus adeno-associados (VAA), o LDLR produziu infecciosidade reduzida em 2 vezes, mostrando que o LDLR é importante para a infecção viral (98). O *LRP1* também já foi descrito como um fator derivado do hospedeiro em resposta à infecção pelo SARS-CoV-2 (99). Esses autores observaram que a ausência de *LRP1*, em células Calu-3 de pulmão humano, reduziu de forma estável o nível de RNA viral (99). Entretanto, ainda não existem estudos conhecidos que relacionam a expressão gênica de *LRP11* com a infecção pelo SARS-CoV-2.

Nossos resultados demonstraram que, no grupo de melhor prognóstico, o gene *LRP11* encontra-se com a expressão gênica mais reduzida se comparado ao grupo de pior prognóstico. Também observamos, através da análise de *single-cell* RNA-Seq, que a expressão de *LRP11* é predominante em células epiteliais. A maioria das células que expressam *ACE2*, proteína importante para a infecção pelo SARS-CoV-2, são pneumócitos tipo 2 (100). Sungnak *et al.* (34) observaram uma expressão enriquecida de *ACE2* em células epiteliais nasais, que são a principal via de infecção do SARS-CoV-2 e também de transmissão do vírus (101).

Em conclusão, nosso estudo demonstrou que a investigação de expressão do secretoma e a caracterização da comunicação célula-célula é uma estratégia válida para identificar moléculas relevantes para o prognóstico da COVID-19. Nossos dados revelaram um conjunto de 43 genes com seu perfil de expressão diminuído no grupo de pacientes de melhor prognóstico. Entre esses genes, o *LRP11* se destacou por, além da expressão diminuída no grupo de melhor prognóstico, é expresso principalmente em células epiteliais. Portanto, concluímos que o gene *LRP11*, em expressão diminuída, possui um fator protetivo contra a infecção pelo SARS-CoV-2, podendo ser utilizado, em futuros estudos, como alvo de drogas para essa doença.

4 - Conclusão

A COVID-19 é uma doença que fez muitas vítimas, principalmente nos anos em que foi considerada uma pandemia. Pacientes do sexo feminino, jovens e com carga viral baixa são indicadores de melhor prognóstico como consequência da diferença da resposta transcricional de genes pró-inflamatórios e da idade pulmonar. Visto isso, a compreensão dos mecanismos moleculares envolvidos na piora ou na não-piora da COVID-19 é de extrema importância para a predição de moléculas com potencial de novos biomarcadores e alvos terapêuticos.

Pacientes do sexo masculino e com idade acima de 60 anos estão associados a um pior prognóstico para a COVID-19 (49,79). O estudo de Freire *et al.* (57) demonstrou que mulheres e jovens possuem uma modulação inversa de citocinas e quimiocinas e da transcrição de neutrófilos, sugerindo uma melhor resposta imunorregulatória nesses dois grupos de pacientes (57). Nossas análises de enriquecimento funcional revelaram genes com expressão diminuída em vias relacionadas à resposta imunológica e inflamação em pacientes do sexo feminino. Outra variável relacionada ao pior prognóstico é a carga viral de SARS-CoV-2, onde pacientes com uma carga viral alta tem aumento da severidade e mortalidade da COVID-19 (18,83). Em nosso estudo, os genes relacionados às vias de sinalização de citocinas e quimiocinas, resposta inflamatória e regulação da proliferação de células T encontraram-se com aumento de expressão em pacientes com alta carga viral. Foram detectados em estudos anteriores que altos níveis de expressão de interleucina-1B (IL-1B), CXCL10, interferon- γ (IFN- γ) e proteína quimioatrativa de monócitos 1 (MCP-1), também conhecida como CCL2, em pacientes com COVID-19 (93). Essa produção excessiva de moléculas pró-inflamatórias é chamada *cytokine storm* (94), responsável por acarretar uma hiper inflamação pulmonar generalizada (86).

Inicialmente, hipotetizamos identificar genes relacionados a um pior prognóstico. Porém, nossos dados identificaram um conjunto de genes que, tendo sua expressão diminuída, podem oferecer um fator protetivo contra a infecção pelo SARS-CoV-2. Esses genes foram identificados ao analisarmos ligantes e seus respectivos receptores, com diferença estatística apenas no grupo de melhor prognóstico; ou seja, pacientes mulheres, jovens e com baixa carga viral. Isso revela a importância de examinar não somente as moléculas secretadas pelas células, mas também para seus receptores. Os genes em questão são *LTF*, *MST1*, *SLT1* (ligantes) *GPR162*, *LRP1*, *LRP11*, *MST1R*, *GPC1* e *SDC1* (receptores).

Os genes *LRP1* e *LRP11* são membros da família de receptores de lipoproteína de baixa densidade (LDL), juntamente com mais 12 outros genes (95). Em consequência da infecção

por patógenos, pode haver alterações na homeostase de lipoproteína, bem como na inflamação decorrente dessa infecção (96). A maioria das infecções, sendo a mais investigada a infecção pelo HIV, estão associadas com níveis reduzidos de lipoproteínas plasmáticas (96). Na COVID-19, há o aumento da angiotensina 2 devido a disfunção da ACE2, o que causa o aumento dos níveis de LDLR (97). Em estudo de Pillay *et al.* (98), em que os autores visaram encontrar proteínas essenciais para a infecção por vírus adeno-associados (VAA), o LDLR produziu infecciosidade reduzida em 2 vezes, mostrando que o LDLR é importante para a infecção viral (98). O *LRP1* também já foi descrito como um fator derivado do hospedeiro em resposta à infecção pelo SARS-CoV-2 (99). Esses autores observaram que a ausência de *LRP1*, em células Calu-3 de pulmão humano, reduziu de forma estável o nível de RNA viral (99). Entretanto, ainda não existem estudos conhecidos que relacionam a expressão gênica de *LRP11* com a infecção pelo SARS-CoV-2.

Nossos resultados demonstraram que, no grupo de melhor prognóstico, o gene *LRP11* encontra-se com a expressão gênica mais reduzida se comparado ao grupo de pior prognóstico. Também observamos, através da análise de *single-cell* RNA-Seq, que a expressão de *LRP11* é predominante em células epiteliais. A maioria das células que expressam *ACE2*, proteína importante para a infecção pelo SARS-CoV-2, são pneumócitos tipo 2 (100). Sungnak *et al.* (34) observaram uma expressão enriquecida de *ACE2* em células epiteliais nasais, que são a principal via de infecção do SARS-CoV-2 e também de transmissão do vírus (101).

Desta forma, concluímos que a análise do perfil do secretoma pelo transcriptoma em dados de RNA-Seq é um método eficaz para identificar moléculas secretadas durante a infecção pelo SARS-CoV-2. Pacientes do sexo feminino, jovens e com baixa carga viral apresentam um conjunto de genes do interactoma com expressão diminuída que predizem um melhor prognóstico para esse grupo. Dentre esses genes, o receptor *LRP11*, com diminuição da expressão, tem efeito protetivo contra a infecção pelo SARS-CoV-2. Esse gene encontra-se expresso principalmente em células epiteliais alveolares, as quais constituem o principal tipo celular alvo desse vírus.

5 – Referências bibliográficas

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. março de 2020;579(7798):265–9.
2. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Military Med Res*. dezembro de 2020;7(1):11.
3. COVID-19 Map - Johns Hopkins Coronavirus Resource Center [Internet]. Disponível em: <https://coronavirus.jhu.edu/map.html>
4. Coronavirus disease (COVID-19): How is it transmitted? - World Health Organization (WHO) [Internet]. Disponível em: <https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted>
5. World Health Organization (COVID-19) [Internet]. Disponível em: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
6. Guan W jie, Ni Z yi, Hu Y, Liang W hua, Ou C quan, He J xing, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*. 30 de abril de 2020;382(18):1708–20.
7. Zhang J jin, Dong X, Cao Y yuan, Yuan Y dong, Yang Y bin, Yan Y qin, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy*. 27 de fevereiro de 2020;all.14238.
8. Gautier J, Ravussin Y. A New Symptom of COVID-19: Loss of Taste and Smell. *Obesity*. maio de 2020;28(5):848–848.
9. Xydakis MS, Dehgani-Mobaraki P, Holbrook EH, Geisthoff UW, Bauer C, Hautefort C, et al. Smell and taste dysfunction in patients with COVID-19. *The Lancet Infectious Diseases*. abril de 2020;S1473309920302930.
10. Mao L, Jin H, Wang M, Hu Y, Chen S, He Q, et al. Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China. *JAMA Neurol*. 1º de junho de 2020;77(6):683.
11. Burki TK. Coronavirus in China. *The Lancet Respiratory Medicine*. março de 2020;8(3):238.
12. Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, Evaluation and Treatment Coronavirus (COVID-19). Em: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 [citado 4 de agosto de 2020]. Disponível em: <http://www.ncbi.nlm.nih.gov/books/NBK554776/>
13. Palaiodimos L, Kokkinidis DG, Li W, Karamanis D, Ognibene J, Arora S, et al. Severe obesity, increasing age and male sex are independently associated with worse in-hospital outcomes, and higher in-hospital mortality, in a cohort of patients with COVID-19 in the Bronx, New York. *Metabolism*. julho de 2020;108:154262.

14. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med*. 26 de março de 2020;382(13):1199–207.
15. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. fevereiro de 2020;395(10223):497–506.
16. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ*. 22 de maio de 2020;m1985.
17. The OpenSAFELY Collaborative, Williamson E, Walker AJ, Bhaskaran KJ, Bacon S, Bates C, et al. Factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. [Internet]. *Epidemiology*; 2020 maio [citado 17 de junho de 2020]. Disponível em: <http://medrxiv.org/lookup/doi/10.1101/2020.05.06.20092999>
18. Pujadas E, Chaudhry F, McBride R, Richter F, Zhao S, Wajnberg A, et al. SARS-CoV-2 viral load predicts COVID-19 mortality. *The Lancet Respiratory Medicine*. agosto de 2020;S2213260020303544.
19. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. *The Lancet Oncology*. março de 2020;21(3):335–7.
20. Zhang L, Zhu F, Xie L, Wang C, Wang J, Chen R, et al. Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. *Annals of Oncology*. julho de 2020;31(7):894–901.
21. Alsharif W, Qurashi A. Effectiveness of COVID-19 diagnosis and management tools: A review. *Radiography*. maio de 2021;27(2):682–7.
22. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol*. junho de 2009;7(6):439–50.
23. Bermingham A, Chand MA, Brown CS, Aarons E, Tong C, Langrish C, et al. Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill*. 4 de outubro de 2012;17(40):20290.
24. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. abril de 2020;181(2):271–280.e8.
25. Sharma O, Sultan AA, Ding H, Triggler CR. A Review of the Progress and Challenges of Developing a Vaccine for COVID-19. *Front Immunol*. 14 de outubro de 2020;11:585354.
26. Ge H, Wang X, Yuan X, Xiao G, Wang C, Deng T, et al. The epidemiology and clinical information about COVID-19. *Eur J Clin Microbiol Infect Dis*. junho de 2020;39(6):1011–9.

27. Harapan H, Itoh N, Yufika A, Winardi W, Keam S, Te H, et al. Coronavirus disease 2019 (COVID-19): A literature review. *Journal of Infection and Public Health*. maio de 2020;13(5):667–73.
28. Weinstein RA. Planning for Epidemics — The Lessons of SARS. *N Engl J Med*. 3 de junho de 2004;350(23):2332–4.
29. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N Engl J Med*. 8 de novembro de 2012;367(19):1814–20.
30. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. Gallagher T, organizador. *J Virol*. 29 de janeiro de 2020;94(7):e00127-20, /jvi/94/7/JVI.00127-20.atom.
31. Li M, Chen L, Zhang J, Xiong C, Li X. The SARS-CoV-2 receptor ACE2 expression of maternal-fetal interface and fetal organs by single-cell transcriptome study. Chan RWY, organizador. *PLoS ONE*. 16 de abril de 2020;15(4):e0230295.
32. Yang X, Zhao H, Li Y, Zhang P. Angiotensin Converting Enzyme 2 in Coronavirus Disease (COVID-19): Evidence from Bioinformatics Analysis [Internet]. *MEDICINE & PHARMACOLOGY*; 2020 abr [citado 29 de junho de 2020]. Disponível em: <https://www.preprints.org/manuscript/202004.0466/v1>
33. Hamming I, Timens W, Bulthuis M, Lely A, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol*. junho de 2004;203(2):631–7.
34. HCA Lung Biological Network, Sungnak W, Huang N, Bécavin C, Berg M, Queen R, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med*. maio de 2020;26(5):681–7.
35. the STACOV-XULA research group, Masana L, Correig E, Ibarretxe D, Anoro E, Arroyo JA, et al. Low HDL and high triglycerides predict COVID-19 severity. *Sci Rep*. dezembro de 2021;11(1):7217.
36. The NHLBI LungMap Consortium, The Human Cell Atlas Lung Biological Network, Muus C, Luecken MD, Eraslan G, Sikkema L, et al. Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat Med*. março de 2021;27(3):546–59.
37. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*. 25 de agosto de 2020;324(8):782.
38. Daniels LB, Sitapati AM, Zhang J, Zou J, Bui QM, Ren J, et al. Relation of Statin Use Prior to Admission to Severity and Recovery Among COVID-19 Inpatients. *The American Journal of Cardiology*. dezembro de 2020;136:149–55.
39. Wang S, Li W, Hui H, Tiwari SK, Zhang Q, Croker BA, et al. Cholesterol 25-Hydroxylase inhibits SARS -CoV-2 and other coronaviruses by depleting membrane cholesterol.

- EMBO J [Internet]. 2 de novembro de 2020 [citado 31 de agosto de 2022];39(21). Disponível em: <https://onlinelibrary.wiley.com/doi/10.15252/emboj.2020106057>
40. Sanders DW, Jumper CC, Ackerman PJ, Bracha D, Donlic A, Kim H, et al. SARS-CoV-2 requires cholesterol for viral entry and pathological syncytia formation. *eLife*. 23 de abril de 2021;10:e65962.
 41. Wang H, Yuan Z, Pavel MA, Jablonski SM, Jablonski J, Hobson R, et al. The role of high cholesterol in age-related COVID19 lethality [Internet]. *Cell Biology*; 2020 maio [citado 31 de agosto de 2022]. Disponível em: <http://biorxiv.org/lookup/doi/10.1101/2020.05.09.086249>
 42. Constantino FB, Cury SS, Nogueira CR, Carvalho RF, Justulin LA. Prediction of Non-canonical Routes for SARS-CoV-2 Infection in Human Placenta Cells. *Front Mol Biosci*. 8 de novembro de 2021;8:614728.
 43. Codo AC, Davanzo GG, Monteiro LB, Souza G, Muraro S, Carregari V, et al. Elevated Glucose Levels Favor Sars-Cov-2 Infection and Monocyte Response Through a Hif-1 α /Glycolysis Dependent Axis. *SSRN Journal* [Internet]. 2020 [citado 28 de junho de 2020]; Disponível em: <https://www.ssrn.com/abstract=3606770>
 44. Moraes D de, Paiva BVB, Cury SS, Ludwig RG, Junior JPA, Mori MA da S, et al. Prediction of SARS-CoV Interaction with Host Proteins during Lung Aging Reveals a Potential Role for TRIB3 in COVID-19. *Aging and disease*. 2021;12(1):42.
 45. Angelidis I, Simon LM, Fernandez IE, Strunz M, Mayr CH, Greiffo FR, et al. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nat Commun*. dezembro de 2019;10(1):963.
 46. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, et al. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci*. dezembro de 2020;12(1):8.
 47. Jia HP, Look DC, Shi L, Hickey M, Pewe L, Netland J, et al. ACE2 Receptor Expression and Severe Acute Respiratory Syndrome Coronavirus Infection Depend on Differentiation of Human Airway Epithelia. *JVI*. 15 de dezembro de 2005;79(23):14614–21.
 48. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*. junho de 2020;20(6):363–74.
 49. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*. maio de 2020;181(5):1036-1045.e9.
 50. tenOever BR. The Evolution of Antiviral Defense Systems. *Cell Host & Microbe*. fevereiro de 2016;19(2):142–9.
 51. Janeway CA, Medzhitov R. Innate Immune Recognition. *Annu Rev Immunol*. abril de 2002;20(1):197–216.

52. Hur S. Double-Stranded RNA Sensors and Modulators in Innate Immunity. *Annu Rev Immunol.* 26 de abril de 2019;37(1):349–75.
53. Lazear HM, Schoggins JW, Diamond MS. Shared and Distinct Functions of Type I and Type III Interferons. *Immunity.* abril de 2019;50(4):907–23.
54. Sokol CL, Luster AD. The Chemokine System in Innate Immunity. *Cold Spring Harb Perspect Biol.* maio de 2015;7(5):a016303.
55. Proudfoot AEI. Chemokine receptors: multifaceted therapeutic targets. *Nat Rev Immunol.* fevereiro de 2002;2(2):106–15.
56. Zhang B, Zhou X, Qiu Y, Song Y, Feng F, Feng J, et al. Clinical characteristics of 82 cases of death from COVID-19. Jin X, organizador. *PLoS ONE.* 9 de julho de 2020;15(7):e0235458.
57. Freire PP, Marques AHC, Baiocchi GC, Schimke LF, Fonseca DLM, Salgado RC, et al. The relationship between cytokine and neutrophil gene network distinguishes SARS-CoV-2–infected patients by sex and age. *JCI Insight.* 24 de maio de 2021;6(10):e147535.
58. Tjalsma H, Bolhuis A, Jongbloed JDH, Bron S, van Dijl JM. Signal Peptide-Dependent Protein Transport in *Bacillus subtilis*: a Genome-Based Survey of the Secretome. *Microbiol Mol Biol Rev.* 1º de setembro de 2000;64(3):515–47.
59. Talbert EE, Lewis HL, Farren MR, Ramsey ML, Chakedis JM, Rajasekera P, et al. Circulating monocyte chemoattractant protein-1 (MCP-1) is associated with cachexia in treatment-naïve pancreatic cancer patients: A biomarker analysis in pancreatic adenocarcinoma-induced cachexia. *Journal of Cachexia, Sarcopenia and Muscle.* abril de 2018;9(2):358–68.
60. Penafuerte CA, Gagnon B, Sirois J, Murphy J, MacDonald N, Tremblay ML. Identification of neutrophil-derived proteases and angiotensin II as biomarkers of cancer cachexia. *Br J Cancer.* março de 2016;114(6):680–7.
61. Makridakis M, Vlahou A. Secretome proteomics for discovery of cancer biomarkers. *Journal of Proteomics.* novembro de 2010;73(12):2291–305.
62. Patel S, Wetie AGN, Darie CC, Clarkson BD. Cancer Secretomes and Their Place in Supplementing Other Hallmarks of Cancer. Em: Woods AG, Darie CC, organizadores. *Advancements of Mass Spectrometry in Biomedical Research* [Internet]. Cham: Springer International Publishing; 2014 [citado 28 de junho de 2020]. p. 409–42. (Advances in Experimental Medicine and Biology; vol. 806). Disponível em: http://link.springer.com/10.1007/978-3-319-06068-2_20
63. Zhang H, Wu P, Chen F, Hao Y, Lao Y, Ren L, et al. SILAC-based quantitative proteomic analysis of secretome between activated and reverted hepatic stellate cells. *Proteomics.* setembro de 2014;14(17–18):1977–86.
64. Bauer JW, Baechler EC, Petri M, Batliwalla FM, Crawford D, Ortmann WA, et al. Elevated Serum Levels of Interferon-Regulated Chemokines Are Biomarkers for Active Human Systemic Lupus Erythematosus. Sontheimer R, organizador. *PLoS Med.* 19 de dezembro de 2006;3(12):e491.

65. Schaaij-Visser TBM, de Wit M, Lam SW, Jiménez CR. The cancer secretome, current status and opportunities in the lung, breast and colorectal cancer context. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. novembro de 2013;1834(11):2242–58.
66. Brown KJ, Seol H, Pillai DK, Sankoorikal BJ, Formolo CA, Mac J, et al. The human secretome atlas initiative: Implications in health and disease conditions. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. novembro de 2013;1834(11):2454–61.
67. Chen G, Chen J, Liu H, Chen S, Zhang Y, Li P, et al. Comprehensive Identification and Characterization of Human Secretome Based on Integrative Proteomic and Transcriptomic Data. *Front Cell Dev Biol*. 21 de novembro de 2019;7:299.
68. Mukherjee P, Mani S. Methodologies to decipher the cell secretome. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. novembro de 2013;1834(11):2226–32.
69. Robinson JL, Feizi A, Uhlén M, Nielsen J. A Systematic Investigation of the Malignant Functions and Diagnostic Potential of the Cancer Secretome. *Cell Reports*. março de 2019;26(10):2622-2635.e5.
70. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, et al. COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nat Biotechnol*. agosto de 2020;38(8):970–9.
71. Shaath H, Vishnubalaji R, Elkord E, Alajez NM. Single-Cell Transcriptome Analysis Highlights a Role for Neutrophils and Inflammatory Macrophages in the Pathogenesis of Severe COVID-19. *Cells*. 29 de outubro de 2020;9(11):2374.
72. Zinellu A, Paliogiannis P, Fois AG, Solidoro P, Carru C, Mangoni AA. Cholesterol and Triglyceride Concentrations, COVID-19 Severity, and Mortality: A Systematic Review and Meta-Analysis With Meta-Regression. *Front Public Health*. 18 de agosto de 2021;9:705916.
73. Song D, Yang D, Powell CA, Wang X. Cell–cell communication: old mystery and new opportunity. *Cell Biol Toxicol*. abril de 2019;35(2):89–93.
74. Lin YT, Salihovic S, Fall T, Hammar U, Ingelsson E, Ärnlov J, et al. Global Plasma Metabolomics to Identify Potential Biomarkers of Blood Pressure Progression. *ATVB* [Internet]. agosto de 2020 [citado 9 de setembro de 2022];40(8). Disponível em: <https://www.ahajournals.org/doi/10.1161/ATVBAHA.120.314356>
75. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet*. fevereiro de 2021;22(2):71–88.
76. Newton AH, Cardani A, Braciale TJ. The host immune response in respiratory virus infection: balancing virus clearance and immunopathology. *Semin Immunopathol*. julho de 2016;38(4):471–82.

77. Lieberman NAP, Peddu V, Xie H, Shrestha L, Huang ML, Mears MC, et al. In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. Cadwell K, organizador. PLoS Biol. 8 de setembro de 2020;18(9):e3000849.
78. Song JW, Zhang C, Fan X, Meng FP, Xu Z, Xia P, et al. Immunological and inflammatory profiles in mild and severe cases of COVID-19. Nat Commun. dezembro de 2020;11(1):3410.
79. Guan W jie, Liang W hua, Zhao Y, Liang H rui, Chen Z sheng, Li Y min, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. Eur Respir J. maio de 2020;55(5):2000547.
80. Pastor-Barriuso R, Pérez-Gómez B, Hernán MA, Pérez-Olmeda M, Yotti R, Oteo-Iglesias J, et al. Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: nationwide seroepidemiological study. BMJ. 27 de novembro de 2020;m4509.
81. Ward H, Atchison C, Whitaker M, Ainslie KEC, Elliott J, Okell L, et al. SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. Nat Commun. dezembro de 2021;12(1):905.
82. Perez-Saez J, Lauer SA, Kaiser L, Regard S, Delaporte E, Guessous I, et al. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. The Lancet Infectious Diseases. abril de 2021;21(4):e69–70.
83. The Massachusetts Consortium for Pathogen Readiness, Fajnzylber J, Regan J, Coxen K, Corry H, Wong C, et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. Nat Commun. dezembro de 2020;11(1):5493.
84. Mallapaty S. The coronavirus is most deadly if you are older and male — new data reveal the risks. Nature. 3 de setembro de 2020;585(7823):16–7.
85. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. novembro de 2003;426(6965):450–4.
86. Messner CB, Demichev V, Wendisch D, Michalick L, White M, Freiwald A, et al. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. Cell Systems. junho de 2020;S2405471220301976.
87. Pernitzsch SR, Sharma CM. Transcriptome Complexity and Riboregulation in the Human Pathogen *Helicobacter pylori*. Front Cell Inf Microbio [Internet]. 2012 [citado 15 de agosto de 2022];2. Disponível em: <http://journal.frontiersin.org/article/10.3389/fcimb.2012.00014/abstract>
88. Peng Z, Cheng Y, Tan BCM, Kang L, Tian Z, Zhu Y, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol. março de 2012;30(3):253–60.
89. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat Med. junho de 2020;26(6):842–4.

90. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, Valenzi E, et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J*. agosto de 2019;54(2):1802441.
91. Aid M, Busman-Sahay K, Vidal SJ, Maliga Z, Bondoc S, Starke C, et al. Vascular Disease and Thrombosis in SARS-CoV-2-Infected Rhesus Macaques. *Cell*. novembro de 2020;183(5):1354-1366.e13.
92. Home - GEO - NCBI [Internet]. Disponível em: ncbi.nlm.nih.gov/geo
93. Ye Q, Wang B, Mao J. The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. *Journal of Infection*. junho de 2020;80(6):607–13.
94. Ferrara JL, Abhyankar S, Gilliland DG. Cytokine storm of graft-versus-host disease: a critical effector role for interleukin-1. *Transplant Proc*. fevereiro de 1993;25(1 Pt 2):1216–7.
95. Champion O, Al Khalifa T, Langlois B, Thevenard-Devy J, Salesse S, Savary K, et al. Contribution of the Low-Density Lipoprotein Receptor Family to Breast Cancer Progression. *Front Oncol*. 30 de julho de 2020;10:882.
96. Sviridov D, Bukrinsky M. Interaction of pathogens with host cholesterol metabolism: Current Opinion in Lipidology. outubro de 2014;25(5):333–8.
97. Cure E, Cumhur Cure M. Strong relationship between cholesterol, low-density lipoprotein receptor, Na⁺/H⁺ exchanger, and SARS-COV-2: this association may be the cause of death in the patient with COVID-19. *Lipids Health Dis*. dezembro de 2021;20(1):179, s12944-021-01607–5.
98. Pillay S, Meyer NL, Puschnik AS, Davulcu O, Diep J, Ishikawa Y, et al. An essential receptor for adeno-associated virus infection. *Nature*. 4 de fevereiro de 2016;530(7588):108–12.
99. Devignot S, Sha TW, Burkard T, Schmerer P, Hagelkruys A, Mirazimi A, et al. Low Density Lipoprotein Receptor-Related Protein 1 (LRP1) is a host factor for RNA viruses including SARS-CoV-2 [Internet]. *Microbiology*; 2022 fev [citado 13 de julho de 2022]. Disponível em: <http://biorxiv.org/lookup/doi/10.1101/2022.02.17.480904>
100. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-Cell RNA Expression Profiling of ACE2, the Receptor of SARS-CoV-2. *Am J Respir Crit Care Med*. 1º de setembro de 2020;202(5):756–9.
101. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. março de 2020;579(7798):270–3.

6 – Apêndices A

6.1 Figuras suplementares

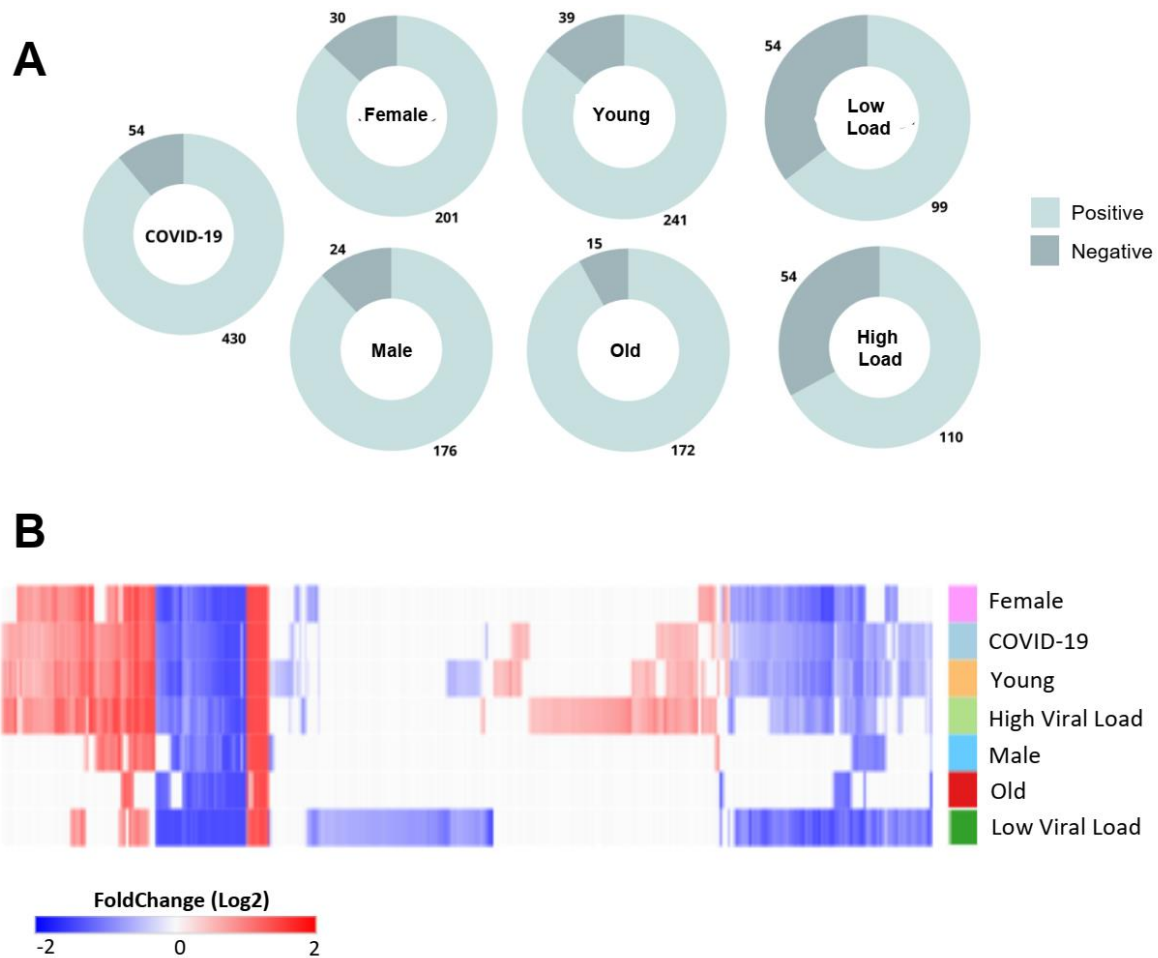


Figura Suplementar 1. Análises transcriptômicas de amostras de *swabs* nasofaríngeos de pacientes positivos para COVID-19 do conjunto de dados GSE152075. (A) Número total de amostras positivas e negativas para COVID-19 nos respectivos grupos: COVID-19, sexo (mulheres e homens), idade (jovens, < 60 anos; idosos ≥ 60 anos) e carga viral (baixa e alta). (B) *Heatmap* dos valores de expressão de logFC em cada comparação analisada. As linhas (genes do secretoma) e as colunas (comparações) foram agrupadas através de *k-means*.

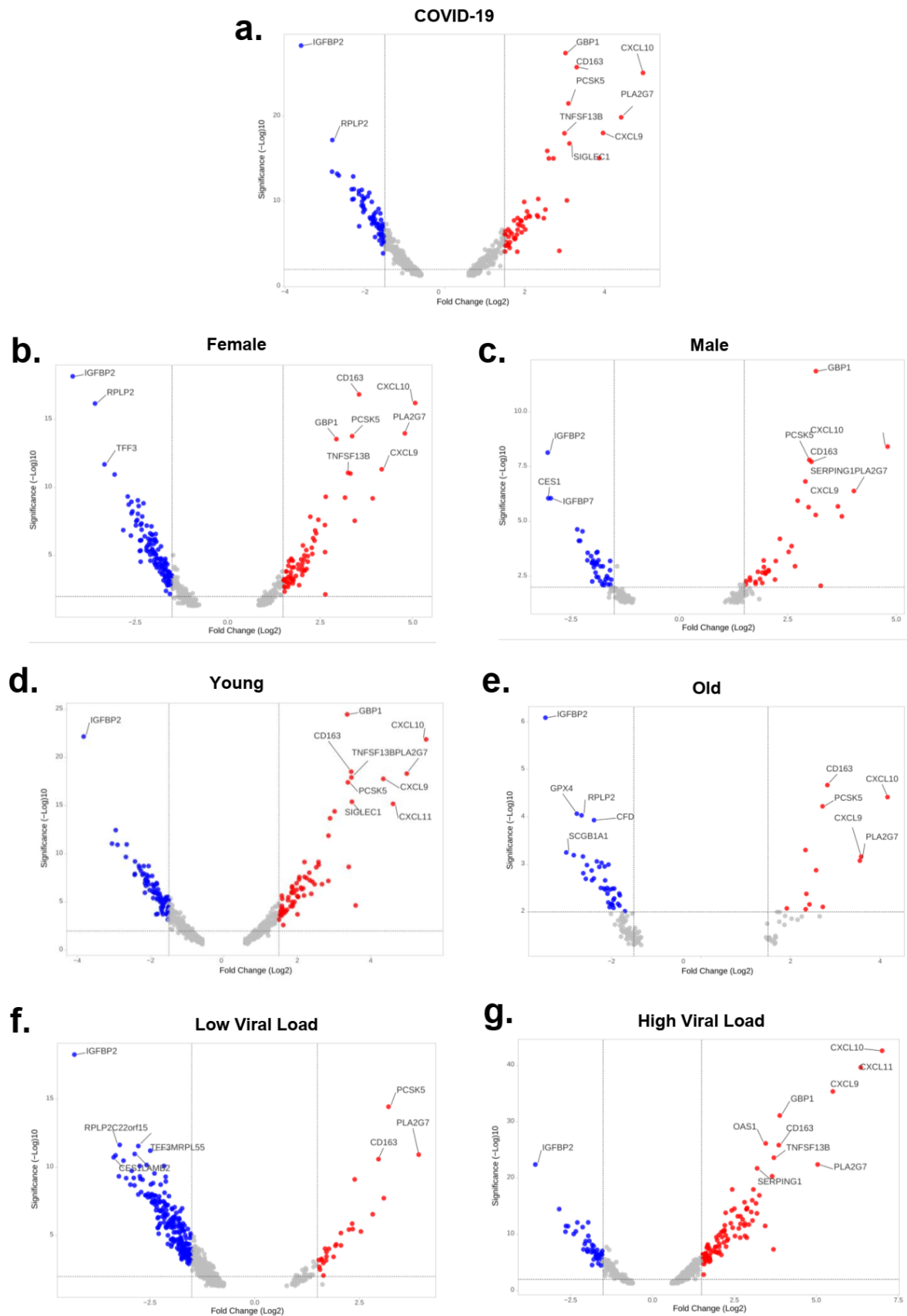


Figura Suplementar 2. *Volcano plots* dos DEGs do secretoma de BALF de pacientes com COVID-19 com alteração na expressão entre as comparações (COVID-19, sexo, idade e carga viral). Cada ponto representa um gene. Os pontos vermelhos representam os genes com aumento de expressão, enquanto que os pontos azuis, os genes com diminuição da expressão. (A) COVID-19. (B) Mulheres. (C) Homens. (D) Jovens. (E) Idosos. (F) Baixa carga viral. (G) Alta carga viral.

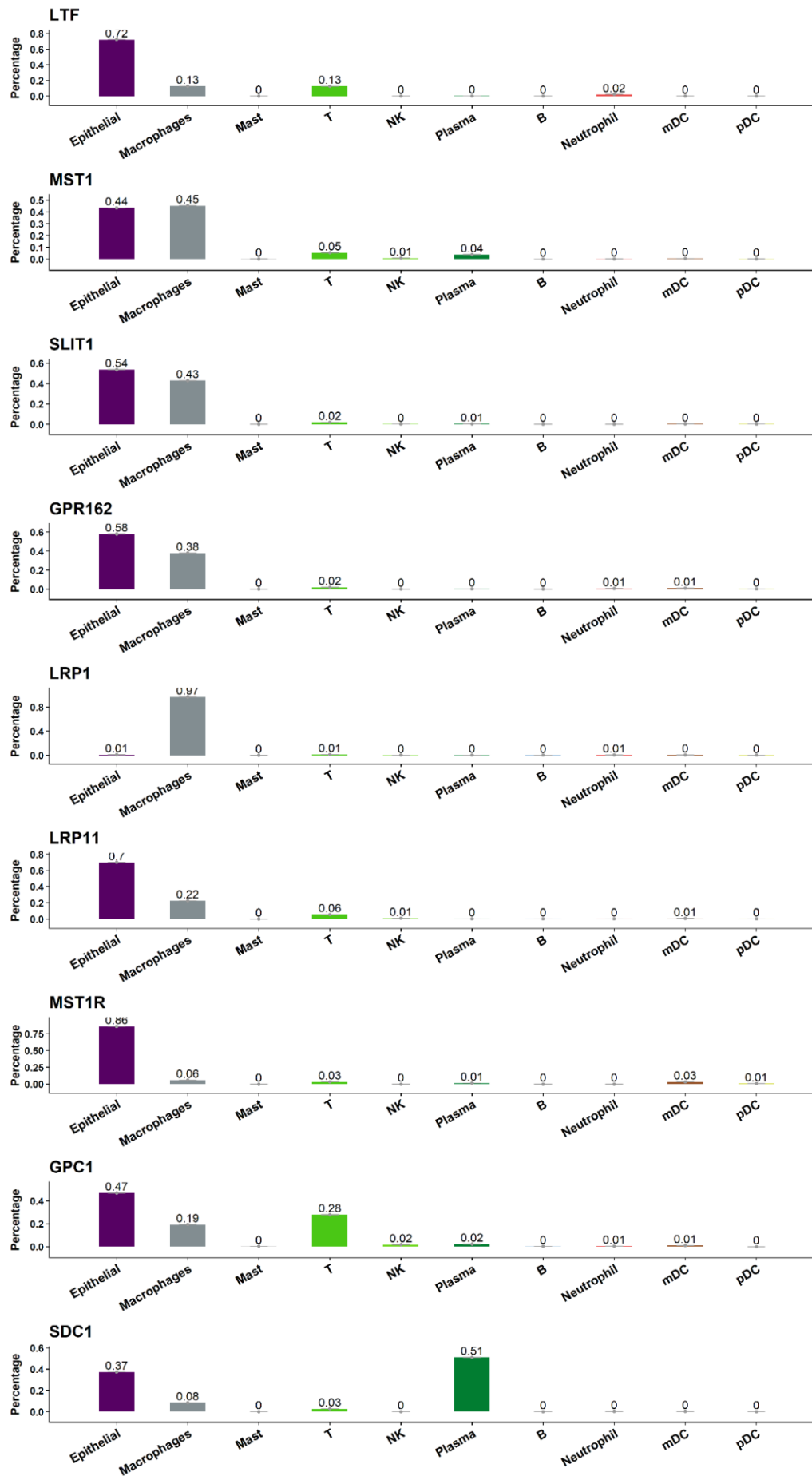


Figura Suplementar 3. *Percentage plot* demonstrando a porcentagem de expressão de cada ligante e receptor presente no cluster A em cada tipo celular encontrado nas amostras de BALF, utilizando os dados do *conjunto de dados* GSE145926 com adicional da amostra saudável de BALF (GSM3660650) do *conjunto de dados* GSE128033, utilizando o pacote Seurat (v. 3) da linguagem R.

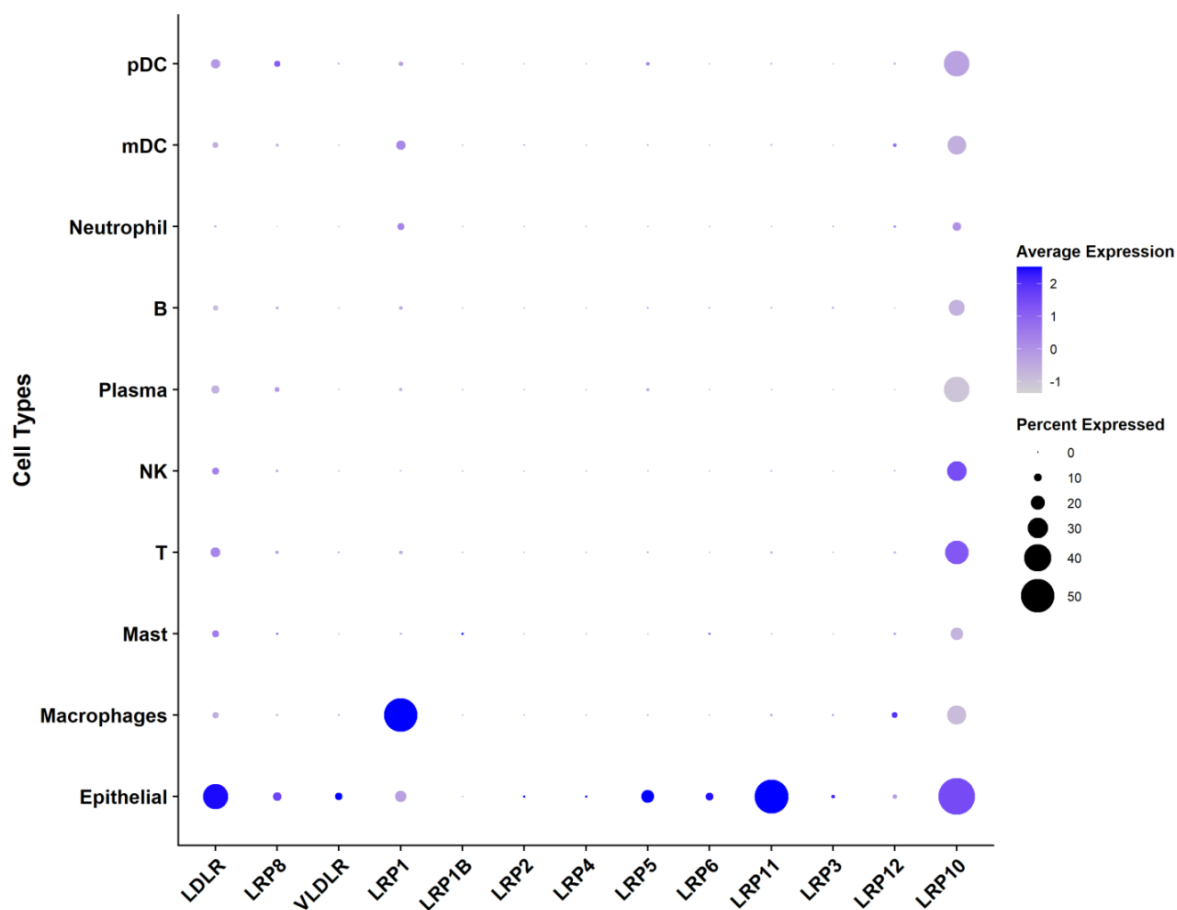


Figura Suplementar 4. *Dotplot* representando os níveis de expressão dos genes participantes da família de receptores de LDL em cada tipo celular encontrado nas amostras de BALF, utilizando os dados do *conjunto de dados* GSE145926 com adicional da amostra saudável de BALF (GSM3660650) do *conjunto de dados* GSE128033, utilizando o pacote Seurat (v. 3) da linguagem R.