

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE FARMÁCIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FARMACÊUTICAS

**IDENTIFICAÇÃO DE BIOMARCADORES DE SARS-CoV-2 E
CARACTERIZAÇÃO DO MICROBIOMA DE PACIENTES COM COVID-19**

OTÁVIO VON AMELN LOVISON

Porto Alegre, 2024

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE FARMÁCIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FARMACÊUTICAS

**IDENTIFICAÇÃO DE BIOMARCADORES DE SARS-COV-2 E
CARACTERIZAÇÃO DO MICROBIOMA DE PACIENTES COM COVID-19**

Tese apresentada por Otávio von Ameln
Lovison para a obtenção do TÍTULO DE
DOUTOR em Ciências Farmacêuticas

Orientadora: Profa. Dra. Andreza Francisco Martins

Porto Alegre, 2024

Tese a ser apresentada ao Programa de Pós-Graduação em Ciências Farmacêuticas, em nível de Doutorado da Faculdade de Farmácia da Universidade Federal do Rio Grande do Sul para a Banca Examinadora constituída por:

Profa. Dra. Fernanda de Paris
Hospital de Clínicas de Porto Alegre (HCPA)

Profa. Dra. Juliana Caierão
Universidade Federal do Rio Grande do Sul (UFRGS)

Prof. Dr. Luis Caetano Martha Antunes
University of Kansas (KU)

CIP - Catalogação na Publicação

Lovison, Otávio von Ameln
 IDENTIFICAÇÃO DE BIOMARCADORES DE SARS-COV-2 E
 CARACTERIZAÇÃO DO MICROBIOMA DE PACIENTES COM COVID-19
 / Otávio von Ameln Lovison. -- 2024.
 179 f.
 Orientadora: Andreza Francisco Martins.

Tese (Doutorado) -- Universidade Federal do Rio
 Grande do Sul, Faculdade de Farmácia, Programa de
 Pós-Graduação em Ciências Farmacêuticas, Porto Alegre,
 BR-RS, 2024.

1. COVID-19. 2. SARS-CoV-2. 3. microbioma. 4. NGS.
 5. Ômicas. I. Martins, Andreza Francisco, orient. II.
 Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Este trabalho foi desenvolvido no Laboratório de Pesquisa em Resistência Bacteriana e no Núcleo de Bioinformática do Hospital de Clínicas de Porto Alegre (HCPA), com financiamento da CAPES, do CNPq (Instituto Nacional de Pesquisa em Resistência Antimicrobiana - INPRA), da FAPERGS e do Laboratório Federal de Defesa Agropecuária (LFDA/RS), utilizando amostras fornecidas pelo Laboratório Exame e pelo biobanco do HCPA. O autor recebeu bolsa de estudos da CAPES.

DEDICATÓRIA

Já se passaram quatro anos desde o surgimento do SARS-CoV-2. A pandemia da COVID-19 foi uma das maiores emergências de saúde da história, a maior do nosso tempo. Meu projeto de doutorado inicial envolvia o estudo de virulência e resistência. Sempre fui um amante da resistência bacteriana. Tudo mudou com a pandemia, menos isso. Lutamos nesta guerra infinita contra virulentos e resistentes, e ainda temos que combater em outra frente - nós contra nós mesmos. Humanos contra humanos. E ao perceber que o melhor resultado para nós é o empate... ao perceber que não seremos nós a testemunharmos a última aurora... eu dedico esta tese a todas as vítimas do SARS-CoV-2 e dos erros da humanidade, mas, dedico, principalmente, aos sobreviventes deste período tão sombrio que vivemos. Resistentes somos nós.

AGRADECIMENTOS

À minha orientadora, profa. Andreza, pela parceria, pelos ensinamentos, pela compreensão e pelas oportunidades. Nos momentos em que a mente obscureceu, tu foste a luz que iluminou o meu caminho, e juntos nós construímos o profissional que eu te falei que eu queria ser anos atrás.

Ao prof. Afonso, coordenador do LABRESIS, orientador da minha orientadora e o responsável por nos aproximar. O senhor é o alfa, a origem, o princípio de tudo. Não é à toa que os seus alunos te veem como um “paizão”.

Aos colegas do LabMIMOS e do LABRESIS, pelos bons momentos, pelas trocas, pela diversão. Um agradecimento especial a duas colegas: a Fabi, pela amizade, por todo suporte e parceria constante nos meus experimentos. Raros os momentos na bancada em que você não esteve comigo. Não é por nada que teu nome está em vários dos artigos que aparecem nesta tese. Tu és insubstituível; a Rafa, minha “duplinha”, como o prof. Afonso costumava dizer. Nossa amizade me ensinou algo precioso, que eu vou levar para sempre: equilíbrio, ponderação.

Ao Dr. Fabiano Barreto, coordenador do LFDA/RS, sempre disponível e disposto a ajudar.

Às minhas famílias, von Ameln Lovison e Haase Pacheco. Vocês são o alicerce, a base, aquilo pelo qual vale a pena viver.

À minha esposa Maíra, por todo suporte, incentivo, apoio e amor. Tu és a minha motivação e minha maior motivadora! Maravilhosa! Nesta batalha que é a vida, não poderia existir melhor guerreira para lutar ao meu lado. Os melhores momentos da minha vida são contigo, e tu me torna uma pessoa melhor a cada dia. Se não fosse tu, eu não sei o que seria. Te amo!

“Messieurs, c'est les microbes qui auront le dernier mot.” Louis Pasteur

RESUMO

O impacto da pandemia causada pelo vírus SARS-CoV-2 acometeu todas as nações e provocou uma crise sanitária global. A falta de infraestrutura - incluindo a carência de métodos de diagnóstico rápidos e custo-efetivos - ficou muito evidente, e foi uma importante limitação no controle da disseminação da COVID-19. Apesar da redução significativa no número de mortes e hospitalizações, devido à vacinação, o SARS-CoV-2 continua pandêmico, fato evidenciado pelos frequentes aumentos no número de casos. Além disso, o desenvolvimento científico impulsionado pela pandemia é aplicável em diversas outras situações, fazendo com que ainda seja importante o desenvolvimento de pesquisas acerca deste tema. As ciências “ômicas” apresentam grande potencial para aplicação no diagnóstico e monitoramento da COVID-19 e de outras doenças, infecciosas ou não. Por isso, este trabalho teve como objetivo identificar marcadores da COVID-19 e do SARS-CoV-2 com base em abordagens de proteômica e metagenômica visando o diagnóstico rápido e diferencial, bem como a relação do perfil funcional da comunidade microbiana com a doença. Foram incluídas neste estudo amostras de nasofaringe e orofaringe provenientes do Biobanco do Hospital de Clínicas de Porto Alegre (HCPA) e do Laboratório Exame. As amostras foram selecionadas de acordo com o resultado positivo/negativo previamente obtido no RT-qPCR para SARS-CoV-2. Para o desenvolvimento do método de diagnóstico, foi utilizada a tecnologia de MALDI-TOF MS para detecção de proteínas específicas de SARS-CoV-2 e marcadores de resposta imune do hospedeiro. Após um rápido pré-processamento para extração das proteínas virais da amostra, o lisado proteico foi submetido à aquisição em MALDI-TOF MS. Os espectros foram tratados e analisados utilizando ferramentas de bioinformática, e posteriormente utilizados no treinamento e validação de algoritmos de aprendizado de máquina, que obteve a acurácia de 78% na classificação das amostras entre positivas e negativas para SARS-CoV-2. Para as análises de microbioma, foram realizadas extrações de DNA das amostras e, após, amplificação da região V3V4 do gene 16S rRNA, os amplicons foram sequenciados na plataforma Illumina MiSeq. Um pipeline adaptado de pré-processamento e análise foi executado majoritariamente em R, na interface de desenvolvimento RStudio (utilizando principalmente os pacotes DADA2, vegan, *phyloseq*, ANCOMB2 e ggplot2) e o software PICRUST2, a fim de caracterizar as comunidades microbianas

em termos de diversidade, composição, abundância diferencial, predição metabólica e associação com parâmetros laboratoriais. Nesta análise foi possível evidenciar uma menor alfa-diversidade e abundância diferencial, principalmente de táxons eubióticos, na COVID-19 severa. Além disso, uma série de rotas metabólicas relacionadas à modulação de resposta imune também apresentaram abundância diferencial reduzida neste grupo. O método em MALDI-TOF MS, implementado na prática diagnóstica, tem perspectiva de redução de custos na ordem de 80% em comparação com o RT-qPCR. A análise de microbioma tem o potencial de ampliar os conhecimentos sobre a patogênese da doença através da caracterização de biomarcadores microbianos, que podem ser utilizados como preditores diagnósticos ou para o desenvolvimento de novos fármacos probióticos.

Palavras-chave: COVID-19; SARS-CoV-2; microbioma; NGS; ômicas

ABSTRACT

IDENTIFICATION OF SARS-COV-2 BIOMARKERS AND CHARACTERIZATION OF THE MICROBIOME OF PATIENTS WITH COVID-19

The impact of the pandemic caused by the SARS-CoV-2 virus affected all nations and caused a global health crisis. The lack of infrastructure - including the lack of rapid and cost-effective diagnostic methods - was evident and an important limitation in controlling the spread of COVID-19. Despite the significant reduction in the number of deaths and hospitalizations due to vaccination, SARS-CoV-2 is still pandemic, a fact evidenced by the frequent increases in the number of cases. Furthermore, the scientific development driven by the pandemic is applicable in several other situations, making the research development on this topic still important. "Omics" sciences have great potential for application in the diagnosis and monitoring of COVID-19 and other diseases, infectious or not. Therefore, this work aims to identify markers of COVID-19 and SARS-CoV-2 based on proteomics and metagenomics approaches aiming for rapid and differential diagnosis and the relationship between the functional profile of the microbial community and the disease. Nasal and oropharyngeal samples from the Biobank of the Hospital de Clínicas de Porto Alegre (HCPA) and Laboratório Exame were included in this study. The samples were selected according to the positive/negative result previously obtained in the RT-qPCR for SARS-CoV-2. MALDI-TOF MS technology was used to develop the diagnostic method to detect SARS-CoV-2 specific proteins and host immune response markers. After a quick pre-processing to extract the viral proteins from the sample, the protein lysate was subjected to acquisition in MALDI-TOF MS. The spectra were treated and analyzed using bioinformatics tools and subsequently used in the training and validation of machine learning algorithms, which achieved accuracy of 78% in classifying samples between positive and negative for SARS-CoV-2. For microbiome analyses, DNA extractions were performed from the samples, and after amplification of the V3V4 region of the 16S rRNA gene, the amplicons were sequenced on the Illumina MiSeq platform. An adapted pre-processing and analysis pipeline was run mainly in R, in the RStudio development interface (mainly using the DADA2, vegan, phyloseq, ANCOMBC2, and ggplot2 packages) and the PICRUSt2 software to characterize the microbial

communities in terms of diversity, composition, differential abundance, metabolic prediction, and association with laboratory parameters. In this analysis, it was possible to highlight a lower alpha-diversity and differential abundance, mainly of eubiotic taxa, in severe COVID-19. Furthermore, a series of metabolic pathways related to immune response modulation also showed reduced differential abundance in this group. The MALDI-TOF MS method, if implemented in diagnostic practice, has the prospect of a cost reduction of around 80% in comparison with RT-qPCR. Microbiome analysis can expand knowledge about the pathogenesis of the disease through the characterization of microbial biomarkers, which can be used as diagnostic predictors or for the development of new probiotic drugs.

Keywords: COVID-19; SARS-CoV-2; microbiome; NGS; omics

SUMÁRIO

1	INTRODUÇÃO	15
2	OBJETIVOS	19
2.1	Objetivo geral	19
2.2	Objetivos específicos	19
3	REVISÃO DE LITERATURA	21
3.1	SARS-CoV-2 e a COVID-19	21
3.2	Proteômica aplicada à detecção de SARS-CoV-2	24
3.3	Microbioma	28
3.4	Análise de microbioma	29
3.4.1	Características dos dados de microbioma	36
3.4.2	Análises de diversidade	38
3.4.3	Abundância diferencial	41
3.4.4	Predição metabólica/funcional	43
3.5	O microbioma do trato respiratório e a COVID-19	45
4	MATERIAIS E MÉTODOS	53
4.1	Delineamento do estudo e tamanho amostral	53
5	RESULTADOS	55
6	DISCUSSÃO GERAL	99
7	CONCLUSÕES E PERSPECTIVAS	103
8	OUTROS ARTIGOS PUBLICADOS	105
9	REFERÊNCIAS	107
10	ANEXOS	121

1 INTRODUÇÃO

O *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) é um vírus causador da *coronavirus disease 2019* (COVID-19), uma doença respiratória com potencial letal, ampla e rápida disseminação, que já causou milhões de infecções e mortes ao redor do mundo e, apesar da desaceleração da pandemia, mesmo indivíduos imunizados continuam suscetíveis à infecção. Em função das modificações genéticas que o vírus sofre ao longo do tempo, o consenso atual é de que o SARS-CoV-2 continuará circulando por vários anos (BARREIRO; SAN ROMÁN, 2022; KISSLER *et al.*, 2020; ROSSMAN *et al.*, 2021). Conforme o painel oficial da Organização Mundial da Saúde (OMS), até o dia 12 de janeiro de 2024, somavam-se 773.819.856 casos confirmados, com 7.010.568 mortes, sendo admitida a possibilidade de que esses números estejam subestimados (WHO, 2020a). No Brasil, até a mesma data, a COVID-19 já causou 37.519.960 infecções e 702.116 óbitos (MINISTÉRIO DA SAÚDE, 2020).

Atualmente existem diversas técnicas para o diagnóstico de vírus respiratórios, sendo o padrão-ouro a reação em cadeia da polimerase e transcriptase reversa em tempo real (*real-time reverse transcription polymerase chain reaction* – RT-qPCR) (SIVANESAN *et al.*, 2022). Contudo, estas técnicas possuem importantes limitações e, dessa forma, múltiplos testes são necessários para o diagnóstico dessas infecções, principalmente nos casos de infecções mistas (BINNICKER *et al.*, 2013; MAJCHRZYKIEWICZ-KOEHORST *et al.*, 2015; YANG *et al.*, 2014).

Na pandemia, estes métodos foram rapidamente adaptados para o diagnóstico da COVID-19 (UDUGAMA *et al.*, 2020). Mas a triagem em massa da população requer métodos altamente sensíveis, específicos, minimamente invasivos, rápidos e de baixo custo. No caso de demanda global, a necessidade destes métodos se torna ainda mais pronunciada, visto que, em momentos de elevada transmissibilidade, os métodos atuais não permitem um controle adequado da pandemia em países com menor aporte financeiro (ILES *et al.*, 2020). Mais importante, se faz necessário o desenvolvimento de métodos que não possuam as limitações dos utilizados atualmente e que possam ser adaptados e ampliados para o diagnóstico de outros patógenos.

As tecnologias envolvendo espectrometria de massas (MS) são uma alternativa confiável em relação ao RT-qPCR e vem demonstrando potencial como recurso diagnóstico ou para investigação da patogênese de infecções virais (SIVANESAN *et al.*, 2022). Dentre estas, a tecnologia de MALDI-TOF MS se destaca como método rápido, de baixo custo, com menor necessidade de mão de obra especializada, e cada vez mais disponível nos laboratórios de microbiologia.

Em geral, houve uma falta de eficácia no enfrentamento da pandemia, especialmente no aspecto da realização de testes. Quando realizados de forma precisa, os testes foram cruciais para conter a propagação do vírus. Em momentos em que o mundo busca soluções rápidas e acessíveis para a detecção precoce de doenças como a COVID-19, a tecnologia de MALDI-TOF MS pode ser desenvolvida e empregada como o método de referência mais confiável (SIVANESAN *et al.*, 2022).

O microbioma desempenha funções essenciais tanto dentro quanto fora do organismo, incluindo a quebra de carboidratos e proteínas que não são digeridas, assim como a absorção de nutrientes, a síntese de vitaminas e o papel crucial na regulação e funcionamento do sistema imunológico do hospedeiro (ROOKS; GARRETT, 2016; ZHENG; LIWINSKI; ELINAV, 2020).

Certamente, há evidências indicando que essa interação desempenha um papel crucial na determinação dos estados de saúde e doença no corpo humano, devido à significativa influência dos microbiomas na regulação das funções imunológicas e metabólicas, podendo tanto melhorá-las quanto piorá-las (OGUNRINOLA *et al.*, 2020). Assim, compreender essas interações pode ser valioso para o prognóstico e tratamento de diversas doenças.

Estudos anteriores revelaram que as bactérias comensais presentes no trato respiratório desempenham um papel na prevenção da colonização e disseminação de patógenos nas superfícies mucosas. Esse processo, denominado "resistência à colonização", também parece ser eficaz contra vírus respiratórios, sugerindo que a microbiota respiratória humana pode contribuir, em certa medida, para a resistência à infecção pelo SARS-CoV-2 (BOGAERT; DE GROOT; HERMANS, 2004; WANG *et al.*, 2022). Dado que a microbiota desempenha um papel crucial na manutenção da homeostase e saúde do seu ambiente, é razoável esperar que haja uma relação entre a microbiota e a infecção pelo SARS-CoV-2 (WANG *et al.*, 2022).

Deste modo, torna-se evidente a necessidade de estudos que possam melhorar o acesso ao diagnóstico e ainda utilizar as informações de perfil proteico e microbiano para compreender melhor a patogênese da doença. Para isso, este projeto se propõe a aplicar as abordagens proteômicas, baseadas em MALDI-TOF MS, e metagenômicas, baseada no sequenciamento do gene codificador do 16S rRNA, para o desenvolvimento de métodos de diagnóstico e a construção do conhecimento acerca da relação microbioma x patógeno x hospedeiro para um melhor entendimento do SARS-CoV-2 e da COVID19.

2 OBJETIVOS

2.1 Objetivo geral

Identificar e caracterizar biomarcadores do vírus SARS-CoV-2 e da doença COVID-19 com base em análises de proteômica e metagenômica.

2.2 Objetivos específicos

1. Identificar biomarcadores específicos de SARS-CoV-2 em MALDI-TOF MS;
2. Caracterizar a diversidade, estrutura e composição do microbioma do trato respiratório superior de pacientes com COVID-19 moderada/severa;
3. Correlacionar os resultados obtidos da composição do microbioma com os parâmetros clínicos associados à COVID-19 moderada/severa;
4. Realizar a predição metabólica do microbioma do trato respiratório superior de pacientes com COVID-19 moderada/severa e identificar rotas metabólicas características desses graus da doença.

3 REVISÃO DE LITERATURA

3.1 SARS-CoV-2 e a COVID-19

O *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) é um novo beta coronavírus de potencial pandêmico, do gênero *Coronavirus*, família *Coronaviridae*, subfamília *Orthocoronavirineae* (KIM *et al.*, 2022). A *coronavirus disease 2019* (COVID-19), uma doença respiratória causada pelo SARS-CoV-2, já causou milhões de infecções e mortes ao redor do mundo e, apesar da desaceleração da pandemia, mesmo indivíduos imunizados continuam suscetíveis à infecção. Em função das modificações genéticas que o vírus sofre ao longo do tempo, o consenso atual é de que o SARS-CoV-2 continuará circulando por vários anos (BARREIRO; SAN ROMÁN, 2022; KISSLER *et al.*, 2020; ROSSMAN *et al.*, 2021). De acordo com o painel oficial da Organização Mundial da Saúde (OMS), até o dia 12 de janeiro de 2024, somavam-se 773.819.856 casos confirmados, com 7.010.568 mortes, sendo admitida a possibilidade de que esses números estejam subestimados (WHO, 2020a). No Brasil, até a mesma data, a COVID-19 já causou 37.519.960 infecções e 702.116 óbitos (MINISTÉRIO DA SAÚDE, 2020).

Para causar infecção, os vírus se ligam aos seus receptores celulares específicos e induzem a estimulação de receptores de reconhecimento padrão (RRPs) nas células epiteliais, resultando na secreção de moléculas, enzimas, peptídeos, espécies reativas de oxigênio (EROs) e quimiocinas com atividade antimicrobiana. Coletivamente, essas moléculas com efeito antimicrobiano atuam no recrutamento de células de defesa, contribuindo com a resposta imune, essencial para o combate da infecção (INVERNIZZI; LLOYD; MOLYNEAUX, 2020).

O SARS-CoV-2 codifica proteínas estruturais de membrana (M), nucleocapsídeo (N), spike (S) e envelope (E), além de diversas outras proteínas não-estruturais (WU *et al.*, 2020). A proteína S é considerada o principal fator de virulência do vírus, por ser fundamental para a entrada do vírus na célula do hospedeiro (LETKO; MARZI; MUNSTER, 2020).

No vírion, a proteína S é um homotrímero que compreende as subunidades S1 e S2. A subunidade S1 é a que se liga na enzima conversora de angiotensina 2 (ECA-2), enquanto a S2 promove a fusão da membrana celular do hospedeiro, capturando proteases de superfície celular, como a protease serina transmembrana 2 (TMPRSS2), o que ativa a proteína S, cliva os receptores ECA-2 e facilita a adesão

do vírus na membrana da célula hospedeira (SHANG *et al.*, 2020). Além disso, o SARS-CoV-2 ainda pode utilizar os processos de fagocitose e endocitose para invadir certos tipos celulares, como macrófagos (WANG *et al.*, 2020).

A ECA-2, usada como receptor para entrada do SARS-CoV-2 nas células, é altamente expressa nas células alveolares tipo II, e a sua ligação com o vírus perturba importantes funções dessas células, promovendo a desestabilização da barreira epitelial, impedindo a regeneração do tecido após lesão e prejudicando a produção de surfactante pulmonar. Estas células também desempenham um papel importante na resposta imune, através da secreção de citocinas em resposta ao dano alveolar e a presença de patógenos, o que promove o recrutamento e ativação de macrófagos e outras células de defesa no alvéolo (HALL *et al.*, 2022).

Essa resposta imune tem a capacidade de conter a infecção inicial e restabelecer a homeostase no trato respiratório, e, por isso, os indivíduos assintomáticos ou com sintomas leves são a maioria. Mas o SARS-CoV-2 possui a capacidade de escapar dessa resposta primária, inibindo as respostas dos Interferons I e III (IFN-I e IFN-III) e de citocinas e quimiocinas pró inflamatórias, o que permite que o vírus replique sem controle no trato respiratório, alcançando uma maior carga viral e eventualmente contribuindo para sua eficiente transmissão pré-sintomática (WEI *et al.*, 2020).

Paradoxalmente, essa supressão da resposta imune no tecido pulmonar está associada a níveis aumentados de citocinas circulantes em pacientes com pneumonia associada à COVID-19. O dano pulmonar severo causado pela intensa replicação viral pode levar a quebra da barreira epitelial, levando ao dano alveolar e aumento de permeabilidade microvascular (CHEN *et al.*, 2020). Então ocorre um extravasamento de citocinas pró inflamatórias, como interleucina-6 (IL-6), IL-8 e IL-1 β , do tecido pulmonar, em conjunto com um recrutamento de células de resposta imune para o pulmão. Uma resposta contra regulatória é ativada simultaneamente, resultando em um aumento local e sistêmico de citocinas supressoras, como IL-10 e fator de crescimento transformante- β (*transforming growth factor* - TGF- β) (HALL *et al.*, 2022; TISONCIK *et al.*, 2012).

Essa elevação sistêmica concorrente de citocinas pró e anti-inflamatórias é frequentemente chamada de 'tempestade de citocinas', onde a ação pró inflamatória pode promover uma má perfusão, prejudicando o funcionamento do pulmão e de outros órgãos, enquanto a ação anti-inflamatória pode resultar na severa supressão

da função dos leucócitos, ou “imunoparalisia”. Essa ação pró inflamatória no pulmão pode sobrepujar as funções de reparo tecidual, levando a dano irreversível e depleção dos macrófagos alveolares (WONG *et al.*, 2019). Durante este processo, fibroblastos ativos podem depositar excesso de colágeno, prejudicando a troca gasosa. A morte de células epiteliais expõe a membrana basal à patógenos microbianos secundários, promovendo seu acesso à circulação sanguínea. Esse conjunto de eventos pode culminar em falha respiratória refratária, fenômeno frequentemente encontrado nos quadros severos da COVID-19 (HALL *et al.*, 2022).

As manifestações clínicas são as mais diversas, sendo as mais comuns, febre, fadiga, tosse seca, dor de garganta, diarreia e anosmia. A maior parte dos indivíduos são assintomáticos ou desenvolvem apenas sintomas leves, mas alguns casos evoluem para SARS e/ou pneumonia, necessitando intubação e podendo evoluir para o óbito (FERREIRA-SANTOS; MARANHÃO; MONTEIRO-SOARES, 2020). Além do comprometimento pulmonar, a infecção por SARS-CoV-2 pode causar dano renal agudo, trombose, lesão endotelial e choque (PIVA *et al.*, 2020). É em função desses quadros graves que ocorre o aumento das taxas de mortalidade, prolongamento dos tempos de internação e sobrecarga dos sistemas de saúde (QUAH; LI; PHUA, 2020).

A OMS classificou o grau de severidade da COVID-19 em 4 níveis: 1 – Leve (sem evidência radiográfica de pneumonia); 2 – Moderado (com evidência de pneumonia, febre, e sintomas respiratórios); 3 – Severo (frequência respiratória \geq 30/min., saturação de oxigênio \leq 93% ao respirar ar ambiente, ou $\text{PaO}_2/\text{FiO}_2 \leq 300$ mmHg; 4 – Crítico (quando existe falência respiratória que necessite ventilação mecânica, choque, ou falência de órgãos que requeira cuidado intensivo) (WHO, 2020b). Além disso, um substancial número de indivíduos infectados vem apresentando sintomas residuais (COVID longa), que se prolongam por muito tempo além do quadro agudo, o que gerou um novo desafio para clínicos e pesquisadores (DIRICAN; BAL, 2022).

A probabilidade do desenvolvimento de COVID-19 severa tem mudado ao longo da pandemia, principalmente em função de 3 fatores: 1 – vacinação; 2 – a seleção de variantes menos letais (Ômicron e suas subvariantes); e 3 – a menor severidade das reinfecções. Dados recentes de um estudo realizado na Espanha demonstram que a probabilidade do desenvolvimento de COVID-19 severa e morte na população

vacinada são estimadas em 0,015% e 0,003%, respectivamente (BARREIRO; SAN ROMÁN, 2022).

3.2 Proteômica aplicada à detecção de SARS-CoV-2

Os tempos recentes têm testemunhado a virulência de uma ampla gama de vírus respiratórios, incluindo os subtipos atualmente circulantes de influenza A (H1N1 e H3N2), influenza B, vírus sincicial respiratório (RSV), parainfluenza, adenovírus e rinovírus. Além destes, outros vírus importantes emergiram, como SARS, *Middle East Respiratory Syndrome* (MERS), influenza A aviária (H5N1), metapneumovírus humano (hMPV), coronavírus NL63 e HKU1 e bocavírus humano (BINNICKER *et al.*, 2013; YANG *et al.*, 2014). A rápida identificação de vírus respiratórios é extremamente importante na contenção de surtos e epidemias, bem como para definição de abordagem terapêutica e necessidade de profilaxia (SIVANESAN *et al.*, 2022).

Atualmente, cultura celular, sorologia e reação em cadeia da polimerase e transcriptase reversa em tempo real (*real-time reverse transcription polymerase chain reaction* – RT-qPCR) são as técnicas utilizadas para o diagnóstico de vírus respiratórios (SIVANESAN *et al.*, 2022). Contudo, estas técnicas possuem importantes limitações: 1 - sorologia e RT-qPCR são métodos direcionados e não detectam patógenos não incluídos na técnica ou emergentes desconhecidos (BINNICKER *et al.*, 2013; YANG *et al.*, 2014); 2 – métodos baseados em cultura são laboriosos e demorados; 3 - métodos sorológicos não são aplicáveis em fase aguda e possuem baixa sensibilidade (SIVANESAN *et al.*, 2022). Dessa forma, múltiplos testes são necessários para o diagnóstico dessas infecções, principalmente nos casos de infecções mistas (BINNICKER *et al.*, 2013; MAJCHRZYKIEWICZ-KOEHORST *et al.*, 2015; YANG *et al.*, 2014).

No início da pandemia, houve a necessidade do desenvolvimento de diversas abordagens para testagem da população, visto que boa parte dos sistemas de saúde do mundo não estavam preparados para dar uma resposta adequada à crescente disseminação do SARS-CoV-2 (MARDIAN *et al.*, 2021). Testagem em larga escala, rastreamento de contatantes e medidas abrangentes de controle se mostraram efetivas em conter o avanço da COVID-19 na Coreia do Sul e Taiwan (FITZGERALD; WONG, 2020; MACINTYRE, 2020; PETO, 2020). Contudo, a limitação de recursos e a frágil sistemática de validação dos novos testes, desenvolvidos às pressas para

suprir a demanda mundial, impôs desafios para contenção e mitigação da pandemia (CARTER *et al.*, 2020).

De fato, no curso da pandemia, por diversos momentos os laboratórios – principalmente em países de baixa e média renda – enfrentaram a escassez de insumos para diagnóstico (ALCOBA-FLOREZ *et al.*, 2020). Isso fez com que múltiplos métodos de diagnóstico fossem implementados, para aumentar a capacidade de processamento de amostras. Diferentes plataformas requerem diferentes técnicas e *expertise*, o que geralmente requer a contratação adicional de colaboradores. Isso resultou em uma maior aglomeração das equipes nas áreas técnicas dos laboratórios, aumento nas taxas de erro laboratorial e aumento no tempo para liberação de resultado (OBERMEIER *et al.*, 2020).

Na era pós-vacina contra a COVID-19, a necessidade de ampla testagem populacional diminuiu. Indivíduos sem sintomas ou com elevado grau de exposição não são mais priorizados para testagem, apesar da eventual necessidade de rastreamento por outros motivos, como monitoramento epidemiológico, busca ativa ou cumprimento de normas e legislações locais (MARDIAN *et al.*, 2021).

A demanda por testes diagnósticos e a disponibilidade de diferentes tecnologias permitiu que os pesquisadores adaptassem rapidamente estes métodos ao diagnóstico da COVID-19. Mas somente a disponibilidade destes testes não garante o controle de uma pandemia (UDUGAMA *et al.*, 2020). A triagem em massa da população no curso de uma pandemia requer métodos altamente sensíveis, específicos, minimamente invasivos, rápidos e de baixo custo. Quando a demanda é global, fica ainda mais evidente a necessidade destes métodos, visto que, em momentos em que houve elevada transmissibilidade, os métodos atuais não permitem um controle adequado da pandemia em países com menor aporte financeiro (ILES *et al.*, 2020).

Ensaio imunocromatográfico de fluxo lateral são amplamente disponíveis e de fácil aplicação, mas a sua capacidade de detectar indivíduos assintomáticos ou estágios iniciais da infecção por SARS-CoV-2 já foi demonstrada como sendo limitada, e, conseqüentemente, ineficaz como medida de contenção de surtos (DEEKS *et al.*, 2021; LAMB *et al.*, 2021; LEBER *et al.*, 2021).

Para SARS-CoV-2, da mesma forma que para outros vírus respiratórios, as técnicas consideradas padrão-ouro, como o RT-qPCR e seus similares (*e.g.* LAMP)

são altamente sensíveis e específicos, mas não podem ser utilizados para detecção de novos vírus ou podem apresentar resultados falso-negativos contra novas variantes, por exemplo, se uma mutação alterar a região de reconhecimento de *primer*, como ocorreu com os *primers* para a proteína Spike de SARS-CoV-2 nas variantes Delta e Ômicron (MARDIAN *et al.*, 2021; WEISSLEDER *et al.*, 2020). Para minimizar a ocorrência desses falso-negativos, ensaios *multitarget* foram desenvolvidos (HARRINGTON *et al.*, 2020; REIJNS *et al.*, 2020).

Contudo, CDC (*Centers for Disease Control and Prevention*) e OMS recomendaram a separação das reações de RT-qPCR ao invés de multiplexá-las, visto que as elevadas concentrações de RNase P humana comparada com o RNA viral de SARS-CoV-2 pode reduzir a sensibilidade da reação para os genes alvo de SARS-CoV-2 quando multiplexadas em uma única reação. Infelizmente, esta prática não é comum nos países de baixa e média renda (PREMRAJ *et al.*, 2020).

Apesar disso, mesmo com mutações em regiões de ligação de *primer* ainda é possível que a reação ocorra adequadamente, e ainda não está claro com que frequência será necessário redesenhar os *primers* para cobrir as mutações ocorridas nas novas variantes (MARDIAN *et al.*, 2021).

As taxas de falso-negativos reportadas para RT-qPCR foram em torno de 30% (10 - 40%) (WEISSLEDER *et al.*, 2020). Dentre as principais causas, estão as importantes diferenças nas taxas de positividade dependendo do material biológico utilizado na análise (MARTINEZ, 2020), do momento de coleta da amostra (KUCIRKA *et al.*, 2020), técnica de coleta de amostra, inadequada preservação da amostra e limitações inerentes à técnica de RT-qPCR (MARDIAN *et al.*, 2021). Além disso, o custo dessas técnicas é elevado, necessitam de mão de obra especializada, e dependem de insumos importados para sua execução (ILES *et al.*, 2022), o que ficou evidenciado em diversos momentos de escassez de insumos no curso da pandemia.

As tecnologias envolvendo espectrometria de massas (MS) são uma alternativa confiável em relação ao RT-qPCR e vem sendo usadas como recurso diagnóstico ou para investigação da patogênese de infecções virais (SIVANESAN *et al.*, 2022). Abordagens proteômicas já foram aplicadas com sucesso em diversas doenças como dengue, chikungunya, zika, Ebola, gripe suína, H1N1 e outros influenza, MERS-CoV, SARS, e SARS-CoV-2 (MAHMUD; GARRETT, 2020).

Dentre estas, a tecnologia de MALDI-TOF MS se destaca como método rápido, de baixo custo, com menor necessidade de mão de obra especializada, e cada vez mais disponível nos laboratórios de microbiologia, apesar do custo elevado do equipamento. Esta tecnologia já se provou revolucionária ao ser aplicada na identificação de microrganismos, e vem sendo empregada com sucesso em muitos hospitais há aproximadamente duas décadas (CROXATTO; PROD'HOM; GREUB, 2012; DO *et al.*, 2022; HOLLAND *et al.*, 1996; ILES *et al.*, 2020; ORSBURN *et al.*, 2020).

As aplicações essenciais de MALDI-TOF MS em virologia constituem, além da detecção e identificação dos vírus, a identificação de mutações e de cepas, o que pode contribuir para um diagnóstico rápido e assertivo (PIAO *et al.*, 2012; SJÖHOLM; DILLNER; CARLSON, 2008; YI *et al.*, 2011). Contudo, seu uso nessa área está muito atrasado em comparação com a sua ampla gama de aplicações.

ROCCA *et al.* (2020) estão entre os primeiros a reportar suas tentativas de utilizar a tecnologia de MALDI-TOF MS diretamente em amostras de swab de naso e orofaringe para discriminar pacientes COVID-19 positivos de negativos. A abordagem utilizou estatística multivariada e aprendizado de máquina (*machine learning – ML*) nessas tentativas. Apesar da interessante proposta metodológica, o desempenho do método não foi adequado e o grupo trabalhava na tentativa de otimizar o método, aumentando o número de amostras, implementando protocolos de extração e melhorias nos algoritmos de ML. Apesar disso, este estudo deu luz à ideia de utilizar MALDI-TOF MS e ML como um ensaio revolucionário para o rastreamento de casos suspeitos da COVID-19.

Outros dois estudos também utilizaram a abordagem de MALDI-TOF MS e ML para o diagnóstico rápido da COVID-19, ambos obtendo acurácia, sensibilidade e especificidade superiores a 90% (DEULOFEU *et al.*, 2021; TRAN *et al.*, 2021).

Atualmente, a detecção de SARS-CoV-2 tem o RT-qPCR como padrão ouro. Na fase aguda da infecção, o RNA do SARS-CoV-2 é detectado no trato respiratório superior (swabs) ou inferior (lavado broncoalveolar), sendo que resultados positivos meramente indicam a presença do RNA do vírus, enquanto outros parâmetros clínicos e sorológicos, como d-dímeros, proteína C reativa, IL-6, dentre outros, são necessários para determinar o grau de severidade da doença e nível de intervenção terapêutica (SIVANESAN *et al.*, 2022).

Em um estudo recente, amostras de soro de 146 indivíduos com diagnóstico de COVID-19 em diferentes graus de severidade foram comparadas com amostras de 152 indivíduos negativos para COVID-19, mas que estavam apresentando sintomas semelhantes. Os espectros de massa foram gerados por MALDI-TOF MS e 25 picos de massa foram estatisticamente significativos para diferenciar pacientes COVID-19 de não-COVID-19. Utilizando diferentes algoritmos de ML, um modelo foi estabelecido, demonstrando 99% de acurácia na identificação de pacientes COVID-19 positivos, com uma sensibilidade de 98% e uma especificidade de 100% em uma coorte de validação de 100 amostras. Este estudo demonstrou o potencial do uso de proteômica *shotgun*, MALDI-TOF MS e ML em amostras de soro para o diagnóstico diferencial da COVID-19 (YAN *et al.*, 2021).

Visando o desenvolvimento de um método de diagnóstico que fosse tão sensível quanto o RT-qPCR, mas não tão específico de modo a detectar novas variantes emergentes, ou ainda, com potencial para detecção simultânea de outros vírus; com potencial para detecção de marcadores de resposta humoral; que tivesse um menor custo e fosse menos dependente de reagentes e mão de obra especializada; que fosse rápido e com a capacidade de detectar indivíduos assintomáticos, Iles *et al.* propuseram um racional metodológico na tentativa de atender todas essas necessidades (ILES *et al.*, 2020, 2022).

De maneira geral, o combate à pandemia falhou e ainda vem falhando no quesito testagem. Quando assertiva, a testagem foi a chave para cercear a disseminação do vírus. Em momentos em que o mundo estiver procurando por soluções rápidas para detecção precoce, sensível, rápida e de baixo custo, para diagnóstico de doenças como a COVID-19, talvez a tecnologia de MALDI-TOF MS possa ser desenvolvida e usada como o verdadeiro 'padrão-ouro' (SIVANESAN *et al.*, 2022).

3.3 Microbioma

O microbioma é uma comunidade microbiana, incluindo bactérias, vírus, fungos e protozoários, que ocupam um habitat bem definido com características físico-químicas distintas (BERG *et al.*, 2020). O termo 'microbioma' não se refere só aos microrganismos, mas inclui todas as suas atividades e interações, sejam entre si, com o microambiente ou com o macro ecossistema, incluindo seus hospedeiros eucarióticos, fundamentais para o seu funcionamento e manutenção de sua saúde

(BERG *et al.*, 2020). Nesta definição estão incluídas as estruturas microbianas, metabólitos, elementos genéticos móveis e o DNA relíquia contido no ambiente (BERG *et al.*, 2020). As interações hospedeiro-microbioma moldam a saúde, fenótipo, metabolismo, sendo que a fisiologia do hospedeiro e do microbioma tem influência no desfecho da interação. Assim, o estado de doença é caracterizado pela disbiose (patobioma), enquanto eubiose faz referência a um estado equilibrado da interação hospedeiro-microbioma e é caracterizada por um microbioma dito 'saudável' (BOSCH; MCFALL-NGAI, 2011).

Estudos anteriores sugeriam que existem 10 vezes mais células e 100 vezes mais genes bacterianos do que humanos no nosso corpo (THE HUMAN MICROBIOME PROJECT CONSORTIUM, 2012), mas um estudo mais recente estimou que o número de células humanas e bacterianas são quase equivalentes (SENDER; FUCHS; MILO, 2016). Essas bactérias habitam todas as superfícies do corpo humano, incluindo trato gastrointestinal e respiratório (THE HUMAN MICROBIOME PROJECT CONSORTIUM, 2012). O microbioma executa importantes funções dentro e fora do corpo, como a decomposição de carboidratos e proteínas não digeríveis, digestão e absorção de nutrientes, biossíntese de vitaminas e indução, instrução e funcionamento da imunidade do hospedeiro, incluindo a resistência à colonização (ROOKS; GARRETT, 2016; ZHENG; LIWINSKI; ELINAV, 2020).

De fato, já existem evidências de que esta interação desempenha um papel decisivo ao determinar os estados de saúde e doença no corpo humano, dada a importante contribuição do microbioma na melhora ou piora de funções imunológicas e metabólicas (OGUNRINOLA *et al.*, 2020). Portanto, entender essas interações pode ser útil para o diagnóstico e tratamento de doenças.

3.4 Análise de microbioma

O sequenciamento de nova geração (NGS) vem sendo utilizado de maneira crescente, especialmente para caracterização taxonômica de comunidades microbianas (CALLAHAN; MCMURDIE; HOLMES, 2017). Para tal, a metodologia mais utilizada é o sequenciamento de genes marcadores, sendo as regiões hipervariáveis do gene 16S rRNA o principal alvo molecular desses estudos (CALLAHAN *et al.*, 2016b).

Tradicionalmente, esse tipo de análise usava por definição o agrupamento das leituras em unidades taxonômicas operacionais (*operational taxonomic units* – OTUs): *clusters* de *reads* agrupadas utilizando um ponto de corte fixo de similaridade de, geralmente, 97% (SCHLOSS *et al.*, 2009; WESTCOTT; SCHLOSS, 2015). Este método não leva em conta os critérios de qualidade e as informações estatísticas contidas nas *reads*, fazendo com que os algoritmos frequentemente errassem na filtragem de erros, gerando um grande número de OTUs espúrias (CALLAHAN *et al.*, 2016a; EDGAR, 2018).

Recentemente, um novo método foi desenvolvido, utilizando um ponto de corte de 100% de similaridade para o agrupamento das sequências. Essas sequências foram chamadas por diversos nomes: ZOTUs (*zero radius*) (EDGAR, 2016), *exacts sequence variants* (ESVs) (CALLAHAN; MCMURDIE; HOLMES, 2017), mas a denominação atualmente mais utilizada é a de *amplicon sequence variants* (ASVs) (CALLAHAN *et al.*, 2016a; EDGAR, 2018). Esse método é capaz de inferir sequências biológicas incorporando os erros de amplificação e sequenciamento, além da frequência das sequências, em um modelo probabilístico com robustez estatística para inferir uma sequência como verdadeira a partir de duas ocorrências (ASVs não toleram *singletons*), distinguindo sequências variantes que diferem em apenas um único nucleotídeo (CALLAHAN *et al.*, 2016a; EDGAR, 2016).

Dentre as principais vantagens desta abordagem, estão: uma maior sensibilidade para anotação taxonômica, proporcionando um bom percentual de identificação em nível de espécie; substancial redução de falso-positivos (a análise por OTUs tende a gerar muitas sequências falsas); a possibilidade de comparar e incorporar dados gerados separadamente (com OTUs, caso seja necessário adicionar uma nova amostra à análise, a análise toda precisa ser refeita, e OTUs geradas em análises separadas não podem ser comparadas, ou seja, a análise de OTUs não é reprodutível) (CALLAHAN *et al.*, 2016a; CALLAHAN; MCMURDIE; HOLMES, 2017; EDGAR, 2018).

Por isso, nos últimos anos, vem sendo pautado que as análises utilizando ASVs devem substituir as de OTUs em estudos de microbioma (CALLAHAN; MCMURDIE; HOLMES, 2017; EDGAR, 2018), visto que os falso-positivos gerados nas análises de OTUs tendem a causar falsas significâncias estatísticas (CHIARELLO *et al.*, 2022). A tabela 1, extraída de Callahan *et al* (2016a) compara a acurácia dos principais

métodos de inferência de sequências em 3 diferentes *datasets* produzidos com cepas de referência (*mock community*).

Tabela 1: Acurácia do DADA2, UPARSE, MED, Mothur e QIIME em 3 diferentes *datasets* produzidos com cepas de referência

			Output de reads (%)	Output de sequências					Cepas de referência*	
				Total	Reference	Exacts	One off	Other		
<i>Balanced</i>	Fwd	DADA2	99,2	93	59	33	1	0	57	
		UPARSE	99,1	81	48	29	2	2	53	
		MED	95,5	86	59	5	22	0	57	
		Mothur	96,3	249	44	25	15	165	49	
		QIIME	99,2	378	51	34	3	290	54	
<i>Merged</i>		DADA2	96,2	87	57	29	1	0	55	
		UPARSE	94,2	76	45	27	2	2	50	
		MED	91,1	64	56	6	2	0	54	
		Mothur	94,1	108	42	27	11	28	47	
		QIIME	94,1	170	45	28	4	93	50	
HMP	Fwd	DADA2	95,1	151	23	112	8	8	21	
		UPARSE	96,7	161	20	123	10	8	21	
		MED	80,9	83	23	2	58	0	21	
		Mothur	95,4	849	20	177	47	605	21	
		QIIME	97,4	1375	20	177	60	1118	21	
<i>Merged</i>		DADA2	92,3	67	23	40	2	2	21	
		UPARSE	67,7	94	20	59	2	13	21	
		MED	64,8	32	23	3	6	0	21	
		Mothur	62,1	121	20	82	9	10	21	
		QIIME	67,6	290	20	71	8	191	21	
<i>Extreme</i>	Fwd	DADA2	99,5	68	26	35	3	4	23	
		UPARSE	99,5	74	21	40	0	13	21	
		MED	96,4	95	16	0	79	0	13	
		Mothur	-	-	-	-	-	-	-	-
		QIIME	99,5	3237	20	44	73	3100	20	
<i>Merged</i>		DADA2	97,6	25	24	1	0	0	21	
		UPARSE	69,9	23	18	4	0	1	18	
		MED	67,6	32	17	0	15	0	14	
		Mothur	94,3	44	23	14	0	7	23	
		QIIME	69,9	36	19	8	1	8	19	

Sequências *output* foram classificadas entre 'Reference' ou 'Exacts' (verdadeiros positivos) e 'One Off' ou 'Other' (falsos positivos) pela comparação com as sequências conhecidas dessas comunidades microbianas (cepas de referência) e a comparação com nucleotídeos (BLAST) para identificar contaminantes. *Balanced*: *mock community* que contém 57 bactérias e arqueias em frequências nominalmente iguais. HMP (*Human Microbiome Project*): *mock community* que contém 21 bactérias em frequências nominalmente iguais. *Extreme*: *mock community* contém 27 cepas bacterianas em frequências abrangendo cinco ordens de magnitude e diferindo na região sequenciada em apenas 1 nucleotídeo (nt). Mothur falhou ao completar a inferência das 'Extreme Forward Reads'. *Cepas de Referência: A coluna cepas

de referência demonstra quantas cepas de referência, dentre as que foram utilizadas para confecção da *mock community*, foram corretamente identificadas pelo algoritmo. Adaptado de Callahan et al, 2016a.

Diferentes objetivos podem ser propostos em estudos de microbioma, e não existe uma única abordagem que possa atender a todos os objetivos e às diferentes variabilidades inerentes a cada tipo de dado. Cabe ao pesquisador e ao bioinformata ter em conta as peculiaridades de cada estudo, realizar as adaptações e adequações necessárias, e traçar a melhor estratégia para o tratamento e análise dos dados, de forma a minimizar as incertezas e vieses, maximizando a informação extraída e diminuindo a probabilidade de gerar falsas significâncias (CALLAHAN et al., 2016b; CHIARELLO et al., 2022; HOLMES, 2019). Apesar disso, algumas etapas fundamentais da análise devem ser obedecidas de forma que se obtenham resultados acurados e reproduzíveis (CALLAHAN et al., 2016b; HOLMES, 2019).

A figura 1 demonstra um *pipeline* básico de pré-processamento de sequências para análise de microbioma usando o DADA2. Esse *pipeline* pode e deve ser adaptado conforme necessário.

A análise se inicia pela avaliação da qualidade das *reads*, que será utilizada como parâmetro para trinar e truncar as *reads*. Além disso, uma etapa de fundamental importância é a tolerância de, no máximo, 2 erros esperados por *read*, filtrando *reads* que excedam este parâmetro (EDGAR; FLYVBJERG, 2015). A trimagem e filtragem é executada de maneira pareada, de forma que ambas as *reads*, *forward* (fwd) e *reverse* (rev), devem passar pelo filtro para permanecerem na análise (CALLAHAN et al., 2016b).

Após, as sequências devem ser deduplicadas para remoção de redundâncias (CALLAHAN et al., 2016b) e submetidas a avaliação por um modelo parametrizado de erros de substituição, para distinguir erros de sequenciamento de variação biológica real. Os parâmetros do modelo são obtidos a partir do próprio dado analisado, utilizando um algoritmo de ML não supervisionado, que faz a estimativa das taxas de erro que, posteriormente, é utilizada para realizar a inferência das sequências (CALLAHAN et al., 2016a).

Em seguida, é realizada a inferência das ASVs, que pode ser realizada a partir de um *pool* de todas as amostras ou a partir de cada amostra individualmente. Utilizando o *pool*, ocorre uma maior demanda computacional, mas otimiza a detecção de ASVs raras – aquelas encontradas em baixa proporção nas amostras individuais,

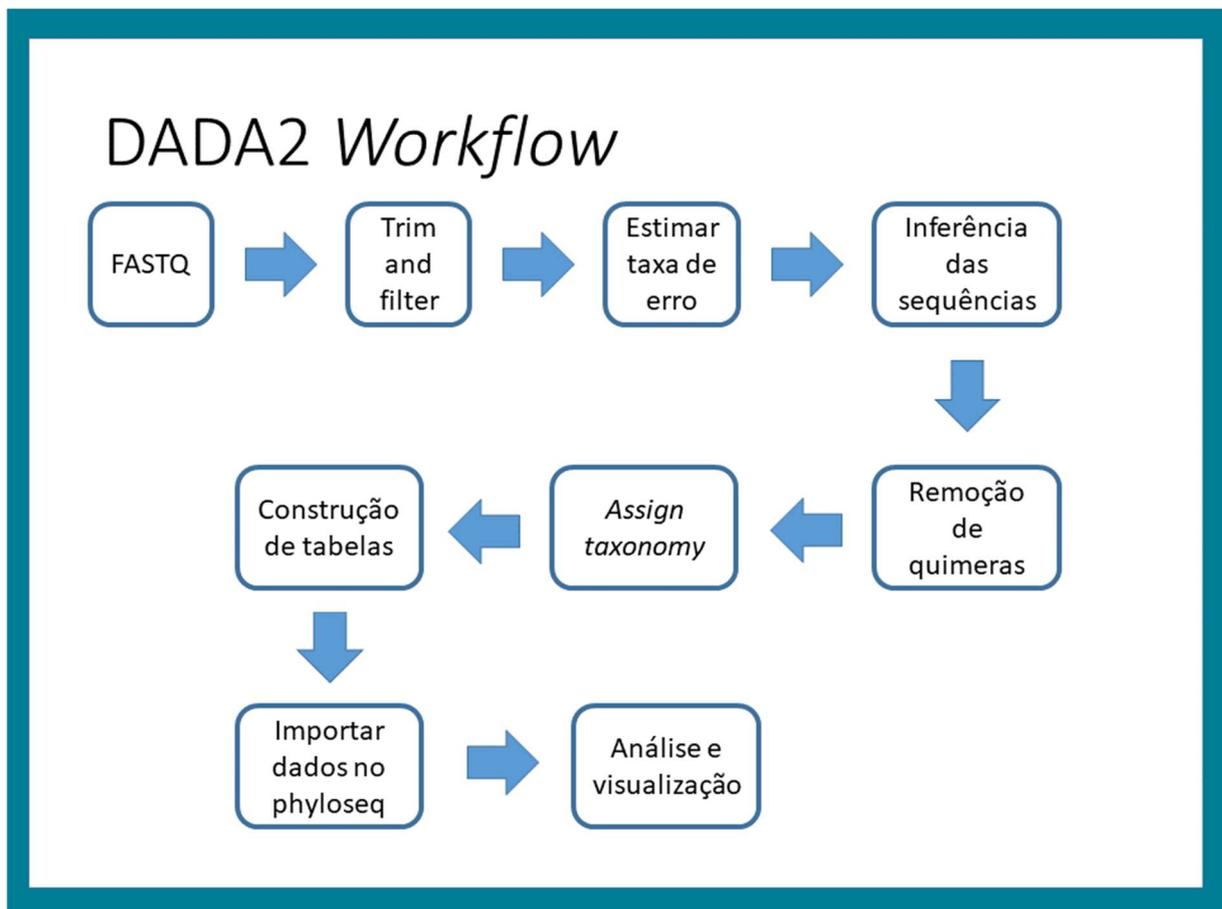
mas amplamente distribuídas em um número maior de amostras. Em casos em que o *dataset* possui dezenas de milhões de *reads*, a inferência em *pool* pode gerar um dado intratável (CALLAHAN; MCMURDIE; HOLMES, 2017).

O algoritmo de inferência de sequências do DADA2 remove quase todos os erros de substituição e *indels* do dado (CALLAHAN *et al.*, 2016a). A partir daí, pode-se juntar as sequências fwd e rev inferidas, removendo sequências pareadas que não se complementam perfeitamente, como uma última etapa de controle contra erros residuais (CALLAHAN *et al.*, 2016b). Após a complementação das sequências fwd e rev, é gerada a tabela de ASVs, uma versão análoga, mas com maior poder de resolução, que a conhecida tabela de OTUs (CALLAHAN *et al.*, 2016b).

Quimeras são artefatos de sequenciamento que ocorrem quando o segmento esquerdo e o segmento direito de duas sequências parentais mais abundantes se unem para formar uma sequência falsa (CALLAHAN *et al.*, 2016b). O algoritmo de inferência de sequências não possui um componente para remoção de quimeras, portanto, é esperado que a tabela de sequências possua um número considerável de quimeras. Felizmente, a elevada acurácia do algoritmo de inferência do DADA2 faz com que a identificação de quimeras seja executada muito facilmente, comparando todas as sequências inferidas da tabela entre si (CALLAHAN *et al.*, 2016b).

Para anotação taxonômica, o DADA2 utiliza um método classificador de *Naïve Bayesian*. Esse método compara as ASVs com um banco de dados de treinamento de sequências previamente classificadas (WANG *et al.*, 2007).

Figura 1: *DADA2 Workflow*: pré-processamento de sequências para análise de microbioma



FASTq: formato de arquivo comumente utilizado para análises metagenômicas; *Trim and Filter*: etapa de trimagem e filtragem do dado; *Assign taxonomy*: anotação taxonômica das sequências

Concluídas estas etapas iniciais, é recomendado que os dados gerados sejam agrupados em um único objeto de análise, contendo as informações das amostras, sequências e suas respectivas abundâncias, anotação taxonômica e metadados (HOLMES, 2019). Para tal, o pacote recomendado é o *phyloseq* (MCMURDIE; HOLMES, 2013). Este objeto geralmente é utilizado para todos os processos subsequentes. Na presença de variações residuais, efeito *batch* (e.g. múltiplas corridas de sequenciamento) e fatores de confusão não controláveis, é recomendado realizar uma pré-normalização do dado (LEEK; STOREY, 2007; OLBRICH; KÜNSTNER; BUSCH, 2023).

Em situações em que o NGS é executado a partir de amostras com baixa biomassa, talvez a execução de uma etapa adicional de descontaminação do dado seja necessária. Isso é obtido através do uso de controles negativos e amostras

obtidas a partir de swab do ambiente e/ou equipamentos utilizados (HOLMES, 2019). O pacote *decontam* (DAVIS *et al.*, 2018) realiza uma descontaminação baseada em um modelo probabilístico e foi recomendado recentemente em uma revisão (KARSTENS *et al.*, 2018).

Pode ser tentador remover características desconhecidas ou dados não anotados da análise. Contudo, os métodos estatísticos costumam ganhar poder quando o dado é mantido íntegro, e a ‘manipulação prematura do dado é a raiz de todos os males em estatística’ (HOLMES, 2019). Deve-se tomar cuidado para não deixar tudo que for desconhecido cair no esquecimento. É recomendado que sequências não anotadas sejam nomeadas, seja utilizando a própria ASV como nome, ou por alguma outra forma de nomenclatura que permita que esse dado possa ser utilizado em análises posteriores (CALLAHAN; MCMURDIE; HOLMES, 2017).

Uma profundidade desigual nas bibliotecas cria heterocedasticidade (variâncias desiguais para as prevalências em cada amostra), contudo, existem maneiras de renormalizar o dado, remover o efeito da profundidade, mantendo a informação útil sobre tamanho de amostra, vieses e erros padrão, para que sejam associados às estimativas de parâmetro (HOLMES, 2019).

Recentemente foi publicado um pacote em R chamado de *Microbiome Batch Effect Correction Suite (MBECS)* (OLBRICH; KÜNSTNER; BUSCH, 2023). Este pacote propõe um *workflow* com uma etapa de avaliação estatística do impacto de vieses não controláveis e efeito *batch* em dados de microbioma, seguido de 2 métodos de normalização (*centered log ratio – clr* e *Total-Sum Scaling – TSS*, que serão melhor detalhados abaixo), 6 métodos de tratamento de vieses controláveis (efeito *batch*) e 2 métodos de tratamento de vieses não controláveis, com subsequente renormalização e avaliação estatística do método que proporcionou os melhores resultados (OLBRICH; KÜNSTNER; BUSCH, 2023). A análise e tratamento de efeito *batch* não é uma tarefa trivial e deve ser realizada apenas quando necessário e com extremo cuidado.

Após a renormalização, geralmente se realiza uma etapa de filtragem supervisionada no dado. A condução desse processo vai depender do tipo de dado, dos objetivos do estudo, dentre outros fatores. É possível dividir a análise em duas abordagens diferentes: uma, mantendo o dado íntegro, inclusive sequências raras, e tratando-o como uma matriz de presença/ausência (dado binário), utilizando a

distância de Jaccard para estas análises; e outra, realizando etapas de processamento no dado (HOLMES, 2019). Dentre as diversas possibilidades, as mais utilizadas são a filtragem por prevalência (são removidas as ASVs de baixa prevalência, sendo que o ponto de corte é arbitrário – geralmente se utiliza algo em torno de 5% ou ocorrência entre 3 a 10 amostras), a remoção de ASVs não anotadas em nível de filo (elevada probabilidade de que sejam artefatos no dado) e a aglomeração filogenética e taxonômica (em função da elevada redundância funcional de táxons intimamente relacionados, a aglomeração costuma prover ganho de poder estatístico com mínimas perdas do ponto de vista funcional) (CALLAHAN *et al.*, 2016b; HOLMES, 2019).

3.4.1 Características dos dados de microbioma

Muitos métodos de análise amplamente utilizados assumem que o dado gerado por NGS é equivalente a um dado ecológico, e acabam por utilizar abordagens baseadas em contagens absolutas, como a dissimilaridade de Bray-Curtis, modelos Gaussianos zero-inflados ou binomiais negativos (MCMURDIE; HOLMES, 2014; WEISS *et al.*, 2017). Em estudos ecológicos, as contagens das espécies costumam ser normalizadas a uma determinada área, sendo que as espécies coexistem livremente, e, de maneira geral, a abundância de uma espécie não interfere na abundância de outra (GLOOR *et al.*, 2017).

Esta verdadeira independência não ocorre no dado gerado por NGS, onde o número de *reads* observadas é limitada à capacidade do instrumento. Por exemplo, para termos uma sequência de *E. coli* incluída no espaço reservado para uma determinada amostra, alguma sequência precisará deixar de ser sequenciada. Logo, o dado gerado por NGS é uma amostragem aleatória de abundâncias relativas de tamanho fixo e não deve ser tratado como abundâncias absolutas (GLOOR *et al.*, 2017).

O dado que naturalmente é descrito como proporções ou probabilidades e sua soma é constante ou irrelevante, é chamado de composicional. Este dado contém a informação da relação entre as partes que o compõem (AITCHISON, 1982). Amostras com diferentes contagens podem demonstrar proporções iguais no dado de NGS, e não é possível prever os valores absolutos contidos em uma amostra a partir das abundâncias relativas obtidas por NGS (GLOOR *et al.*, 2017).

Diversos problemas advêm da utilização de uma abordagem não composicional na análise de microbiomas. O número total de leituras obtidas para uma mesma amostra não é reprodutível, e este número é um importante fator de confusão para as métricas de distância ou dissimilaridade utilizadas para ordenações multivariadas (GLOOR *et al.*, 2017; MCMURDIE; HOLMES, 2014). O uso de rarefação/subamostragem e diversas outras estratégias de normalização não corrigem este problema, além de provocar a perda de informação e precisão (GLOOR *et al.*, 2017; MCMURDIE; HOLMES, 2014).

Neste contexto, é sabido que problemas graves ocorrem com o uso de correlação em um dado composicional (PEARSON, 1896), fato redescoberto em estudos de microbioma (KURTZ *et al.*, 2015; MORTON *et al.*, 2018). Admitir o problema com a correlação é fundamental, visto que correlação e covariação são conceitos-chave para ordenação, clusterização, análise de redes e abundância diferencial. O dado composicional possui um viés de correlação e uma estrutura de correlação diferente de seu dado absoluto originário, e a situação fica ainda pior ao se utilizar práticas comuns de análise de microbioma como a subamostragem ou agregação (GLOOR *et al.*, 2017).

A análise de dados composicional (CODA) inicia com a transformação do dado em razões. As razões capturam a relação entre as variáveis que compõem o dado e não variam, independentemente de estar representado como contagens ou proporções. Ao tirarmos o logaritmo dessas razões (log-ratios), o dado se torna simétrico e linearmente relacionado (PAWLOWSKY-GLAHN; EGOZCUE; TOLOSANA-DELGADO, 2015). Assim, obtemos informações sobre as abundâncias *log-ratio* de um táxon relativa a outros táxons do conjunto de dados, e esta informação se relaciona diretamente com a amostra (GLOOR *et al.*, 2017). A transformação *log-ratio* mais utilizada em estudos de microbioma é a *centered log-ratio* (clr) (AITCHISON, 1982; GLOOR *et al.*, 2017).

A partir da transformação clr, é possível calcular a distância de Aitchison, que nada mais é que a distância Euclidiana entre amostras após a transformação (AITCHISON, 1982). Esta distância é superior às tradicionalmente utilizadas Jensen-Shannon e Bray-Curtis, sendo mais estável, reprodutível, informativa e verdadeiramente linear (AITCHISON *et al.*, 2000; BARS-CORTINA, 2022; GLOOR *et al.*, 2017). A análise de PCA a partir da distância de Aitchison é a substituição

adequada para a exploração da beta-diversidade por PCoA (AITCHISON; GREENACRE, 2002).

Recentemente um pacote em R chamado *coda4microbiome* (CALLE; PUJOLASSOS; SUSIN, 2023) foi publicado trazendo diversas análises utilizando a abordagem CODA. O diferencial deste pacote é o foco na predição, ou seja, identificar uma assinatura microbiana que contenha o menor número de *features* com o máximo de poder preditivo. Apesar das importantes diferenças metodológicas entre este pacote e as demais abordagens para abundância diferencial/assinatura microbiana, o destaque fica para a metodologia proposta para a medida de associação entre variáveis dependentes e *log ratios*. A importância de um determinado táxon é inferida a partir da acurácia de predição agregada de todas as razões envolvendo este táxon, obtida por um modelo linear generalizado. Para variáveis contínuas, é calculado o coeficiente de correlação de Spearman entre a variável e as predições do modelo. Esta abordagem visa minimizar a ocorrência de correlações espúrias, muito comuns em estudos de microbioma que não utilizam abordagem CODA (BARS-CORTINA, 2022; CALLE, 2019; CALLE; PUJOLASSOS; SUSIN, 2023; GLOOR *et al.*, 2017).

3.4.2 Análises de diversidade

É frequentemente necessário realizar transformações nos dados de microbioma, para minimizar o impacto de diferenças de tamanho de biblioteca, variância, escala, dentre outros vieses. As transformações mais utilizadas são *log ratio* e abundância relativa (conversão do dado em percentual), que transforma números absolutos em proporções (GREENACRE; MARTÍNEZ-ÁLVARO; BLASCO, 2021). Através dessas transformações é possível avaliar diferenças em escala e comparar a distribuição das abundâncias absoluta e relativa, permitindo avaliar a melhor estratégia de transformação para as análises subsequentes e a detecção de distribuições anormais no dado, como a distribuição bimodal (CALLAHAN *et al.*, 2016b).

Beta-diversidade

A clusterização de amostras de microbioma baseada em sua composição revela a heterogeneidade da comunidade microbiana dentre os pacientes. A formação de subgrupos na clusterização é o que chama a atenção dos pesquisadores para

possíveis divergências, que podem ser exploradas para pautar diagnóstico, tratamento, dentre outros (COSTEA *et al.*, 2018).

Para classificar adequadamente as amostras em subgrupos de acordo com o seu perfil composicional, é necessário medir a dissimilaridade entre estas, o que é conhecido como beta-diversidade na comunidade microbiana (ANDERSON *et al.*, 2011).

Análises estatísticas multivariadas são particularmente úteis para exploração da beta-diversidade em microbioma. Projeções multivariadas são facilmente geradas ao *plotar* as ordenações de dissimilaridade das amostras utilizando métodos de redução dimensional de coordenadas principais (HOLMES, 2019). Existem mais de 50 métricas de distância disponíveis no pacote *phyloseq* (MCMURDIE; HOLMES, 2013), por isso, a escolha adequada das métricas utilizadas é fundamental para o sucesso e adequação da análise (HOLMES, 2019).

Os métodos de redução dimensional mais utilizados para essas análises são a análise de componentes principais (PCA) (HOTELLING, 1933), análise de coordenadas principais dupla (DPCoA) (PAVOINE; DUFOUR; CHESSEL, 2004), escalonamento multidimensional (MDS), análise de coordenadas principais (PCoA) (GOWER, 1966) e escalonamento multidimensional não-métrico (NMDS) (MINCHIN, 1987).

Dentre as métricas de distância mais utilizadas estão Unifrac/Unifrac ponderada, que utiliza informação filogenética para comparar as amostras (LOZUPONE *et al.*, 2011) e possui boa sensibilidade para *outliers* (CALLAHAN *et al.*, 2016b), Bray-Curtis, métrica de dissimilaridade que leva em conta presença/ausência e abundância (BRAY; CURTIS, 1957) e Jaccard, que mede a dissimilaridade entre atributos binários (presença/ausência) (JACCARD, 1912). Tanto Bray-Curtis quanto Jaccard são métricas derivadas da dissimilaridade de Manhattan (CHEN *et al.*, 2021).

A divergência de Kullback-Leibler (KL) reflete a diferença entre duas medidas probabilísticas (KULLBACK; LEIBLER, 1951). Das métricas derivadas de KL, a mais utilizada é a divergência de Jensen-Shannon (JSD) e a sua raiz quadrada, rJSD, é a mais utilizada dentre as métricas baseadas em KL para clusterização de subgrupos de microbioma (ENDRES; SCHINDELIN, 2003).

Diferentes métricas podem levar a diferentes padrões de clusterização, e a métrica adequada vai depender do tipo de dado e dos objetivos pretendidos. Autores

que orientam a utilização das métricas de diversidade, apontam que em geral os estudos de microbioma utilizam as métricas de maneira empírica e arbitrária, e boa parte destes obtém resultados inadequados (KOREN *et al.*, 2013; SHI *et al.*, 2022). Recentemente foi demonstrado que divergências baseadas em Manhattan são mais robustas para clusterização de subgrupos quando as diferenças composicionais são provocadas por elementos de maior abundância, enquanto divergências baseadas em KL favorecem a clusterização quando estas diferenças são causadas por táxons de menor abundância (CHEN *et al.*, 2021).

Alfa-diversidade

Outra característica frequentemente avaliada em estudos de microbioma é o número de táxons presentes em determinada amostra e a sua respectiva abundância. Essa análise de diversidade bacteriana em um determinado nicho é conhecida como alfa-diversidade (KIM *et al.*, 2017). De maneira geral, a diversidade é constituída de duas métricas: riqueza (*richness*) – número de táxons realmente observados na amostra; e uniformidade (*evenness*) – avalia a uniformidade com que cada táxon presente na amostra se apresenta em termos de abundância relativa (KIM *et al.*, 2017).

Quando a riqueza e uniformidade aumentam, também aumenta a alfa-diversidade. Dentre os índices mais utilizados para este cálculo estão ACE (*Abundance-based Coverage Estimator*), Shannon e Simpson, sendo o primeiro mais utilizado como métrica de estimativa de riqueza, e os dois últimos os mais recomendados como índice de diversidade microbiana (CHAO *et al.*, 2006; HAEGEMAN *et al.*, 2013).

A alfa-diversidade por si só não é informativa. Se faz necessário avaliar o quanto esta métrica, individualmente, reflete o ambiente que a comunidade microbiana habita, e o quanto se relaciona com as características que estão sendo estudadas (CALLAHAN *et al.*, 2016b). Não existe consenso sobre qual o melhor índice de diversidade para estudos de microbioma. Sendo estes índices derivados da ecologia, e não especificamente desenvolvidos para estudos de microbioma, todos são fontes de viés em maior ou menor grau. Mesmo assim, comparações ainda podem ser feitas. Múltiplas abordagens podem ser utilizadas, na tentativa de aumentar a robustez das análises (KOREN *et al.*, 2013; KOWALCHUK, 2004; SHI *et al.*, 2022).

3.4.3 Abundância diferencial

Definimos um táxon como 'diferencialmente abundante' entre 2 ecossistemas quando a média das suas abundâncias absolutas é diferente entre estes ecossistemas. Na prática, utilizamos a análise de abundância diferencial para identificar táxons nos quais a média das abundâncias absolutas é significativamente diferente com relação às variáveis de interesse, como, por exemplo, entre grupos amostrais (LIN; PEDDADA, 2020b).

Uma etapa fundamental ao avaliar abundância diferencial é a normalização do dado. Isso ocorre em função de vieses como tamanho da amostra e número de leituras da biblioteca. Falhar na normalização do dado pode aumentar a taxa de descobertas falsas (*False Discovery Rate* – FDR) e diminuir o poder estatístico da análise (LIN; PEDDADA, 2020b).

Uma das estratégias mais utilizadas para normalização de dados é a rarefação. Entretanto, alguns autores consideram isso inadmissível (BARS-CORTINA, 2022; LIN; PEDDADA, 2020b; MCMURDIE; HOLMES, 2014), pois essa estratégia induz à omissão de dados válidos disponíveis, a introdução de incerteza na subamostragem, a escolha arbitrária do número de leituras mínimo de cada biblioteca e desafios na estimativa da sobredispersão. Além da rarefação, outra estratégia bastante utilizada para normalização destes dados é a de *scaling*. A ideia é dividir a abundância observada por um fator de normalização para eliminar os vieses resultantes de amostragem desigual (LIN; PEDDADA, 2020b).

Dentre os métodos de *scaling* propostos estão o *Cumulative-Sum Scaling* (CSS) implementado no pacote metagenome-Seq (PAULSON *et al.*, 2013), quartil superior (*Upper Quartile* - UQ) (BULLARD *et al.*, 2010), *Trimmed Mean of M-values* (TMM) no pacote edgeR (ROBINSON; MCCARTHY; SMYTH, 2010) e a mediana (MED) no pacote DESeq2 (LOVE; HUBER; ANDERS, 2014). Estes métodos são derivados da análise de expressão diferencial e estão sujeitos a viés. Por exemplo, o método do DESeq2 (LOVE; HUBER; ANDERS, 2014) parte do pressuposto de que táxons de média abundância não serão diferentemente abundantes. Isso pode ser verdade para expressão gênica, mas nem sempre será verdade em estudos de microbioma. Além disso, estes métodos não levam em conta a composicionalidade do dado gerado por NGS (LIN; PEDDADA, 2020b).

Recentemente, um novo método chamado *Analysis of Compositions of Microbiome with Bias Correction* (ANCOM-BC), que como o próprio nome diz, foi desenvolvido especificamente para abundância diferencial em microbioma, foi proposto para resolver os vieses resultantes da amostragem desigual (LIN; PEDDADA, 2020a). ANCOM-BC parte do pressuposto de que a abundância observada e a abundância não observada de um táxon são proporcionais. Essa proporção é definida como uma fração da amostra, e pode variar entre as amostras. Então ANCOM-BC leva em conta essa fração ao introduzir um fator de compensação específico para cada amostra em um modelo de regressão linear que é estimado a partir da abundância observada. Esse fator é que atua na correção do viés (LIN; PEDDADA, 2020a).

Para análise estatística da abundância diferencial, diversos métodos já foram propostos. Alguns comumente encontrados utilizam testes não-paramétricos clássicos, como Mann–Whitney/Wilcoxon rank-sum test e Kruskal–Wallis. Estes métodos não levam em conta a natureza composicional do dado de microbioma e não devem ser utilizados (LIN; PEDDADA, 2020b).

Dentre os métodos baseados em RNA-seq, os mais populares são os métodos propostos pelos pacotes DESeq2 (LOVE; HUBER; ANDERS, 2014) e edgeR (ROBINSON; MCCARTHY; SMYTH, 2010). Esses métodos modelam a abundância observada utilizando um modelo binomial negativo após a normalização do dado utilizando *scaling* (LOVE; HUBER; ANDERS, 2014; ROBINSON; MCCARTHY; SMYTH, 2010).

Já a análise de ANCOM-BC (LIN; PEDDADA, 2020a) modela as abundâncias observadas utilizando uma compensação baseada em um modelo log-linear. Essa abordagem testa a abundância diferencial de cada táxon individualmente enquanto estima e corrige o viés causado pela amostragem desigual. ANCOM-BC, além de controlar para FDR de maneira satisfatória, fornece intervalos de confiança e valor de “*p*” para a abundância diferencial de cada táxon entre dois grupos amostrais (LIN; PEDDADA, 2020b). Além disso, por ser baseado em regressão linear, permite múltiplos delineamentos, correção para fatores de confusão e avaliação da abundância diferencial em estudos longitudinais e de dose-resposta (LIN; PEDDADA, 2020a). Muito semelhante em termos metodológicos, o algoritmo LinDA (*linear models for differential abundance analysis of microbiome compositional data*) juntamente com

a nova versão do algoritmo ANCOM-BC2 são consideradas as melhores abordagens em termos de controle de FDR e poder estatístico (ZHOU *et al.*, 2022).

3.4.4 Predição metabólica/funcional

Uma das limitações dos estudos de metagenômica direcionada é que esta abordagem não fornece informações a respeito do perfil funcional daquela comunidade microbiana (DOUGLAS *et al.*, 2020). Para contornar tal limitação, diversos métodos já foram propostos (ASSHAUER *et al.*, 2015; IWAI *et al.*, 2016; JUN *et al.*, 2015; LANGILLE *et al.*, 2013; WEMHEUER *et al.*, 2018).

Dentre as ferramentas mais robustas e atualizadas, está o software PICRUSt2 (DOUGLAS *et al.*, 2020). Este software, que é um upgrade do já conhecido *Phylogenetic Investigation of Communities by Reconstruction of Unobserved States* PICRUSt (LANGILLE *et al.*, 2013), possui um banco de dados de genes e genomas de referência maior e mais atualizado, o que permite uma predição fenotípica mais acurada de uma comunidade microbiana (DOUGLAS *et al.*, 2020). O algoritmo do PICRUSt2 possui etapas que otimizam a predição dos genomas; posicionando as sequências em uma filogenia de referência, ao invés de contar somente com a predição por taxonomia; utilizando um banco de dados muito maior de genes e genomas de referência; utilizando critérios de abundância na predição das rotas metabólicas; realizando a predição funcional de fenótipos complexos; e permitindo a integração de bancos de dados customizados (DOUGLAS *et al.*, 2020).

O PICRUSt2 integra ferramentas já existentes de código aberto para realizar a predição de genomas a partir de sequências do gene 16S rRNA. As ASVs são posicionadas em uma árvore filogenética de referência, que serve como base para as predições funcionais, e que contém mais de 20 mil sequências do 16S rRNA completo obtidas do banco *Integrated Microbial Genomes* (IMG) (DOUGLAS *et al.*, 2020; MARKOWITZ *et al.*, 2012).

O posicionamento filogenético no PICRUSt2 é baseado no *output* de 3 ferramentas: HMMER (<http://www.hmmer.org>) (FINN; CLEMENTS; EDDY, 2011) para alinhar as ASVs, EPA-NG (BARBERA *et al.*, 2018) para determinar a posição ótima das ASVs em uma filogenia de referência e GAPP (CZECH; STAMATAKIS, 2019) para exportar a nova árvore com o posicionamento das ASVs incorporado. Isso resulta em uma árvore filogenética contendo os genomas de referência e os microrganismos

obtidos a partir das amostras, o que é utilizado para prever o número de cópias de cada família de genes individualmente para cada ASV (DOUGLAS *et al.*, 2020).

No PICRUSt2, a inferência do conteúdo genômico das sequências das amostras ocorre em estado oculto (conhecido como *hidden-state* em programação) (DOUGLAS *et al.*, 2020). O pacote Castor para R (LOUCA; DOEBELI, 2018) é utilizado para essas funções de predição. Nessa abordagem, as ASVs são corrigidas pelo seu número de cópias do gene 16S rRNA e então multiplicadas pelas suas predições funcionais para produzir um metagenoma. O PICRUSt2 também fornece a contribuição de cada ASV para cada função predita, permitindo uma análise estatística baseada em taxonomia.

Por fim, a abundância das rotas metabólicas é inferida a partir de mapeamentos estruturais (*structured pathway mappings*) (DOUGLAS *et al.*, 2020). Para isso, o PICRUSt2 aceita a utilização de diversos bancos de dados de família de genes, sendo o *Kyoto Encyclopedia of Genes and Genomes* (KEGG) (KANEHISA *et al.*, 2012) *orthologs* (KOs) um dos mais utilizados.

Atualmente, o MetaCyc (CASPI *et al.*, 2016) é o principal banco de dados para predição da abundância de rotas metabólicas no PICRUSt2. Esse banco de dados é uma alternativa ao KEGG, e atualmente é o principal foco do amplamente utilizado perfilador funcional de metagenômica, o HUMAnN2 (FRANZOSA *et al.*, 2018).

A predição de fenótipos microbianos também é uma função do PICRUSt2. Baseada nos genomas do IMG (MARKOWITZ *et al.*, 2012), o PICRUSt2 pode prever 41 fenótipos, que representam atividades metabólicas de alto nível, como utilização de glicose ou utilização de nitrogênio, que estão anotadas nos genomas de referência (DOUGLAS *et al.*, 2020).

Existem duas principais críticas à predição metabólica baseada em amplicon (e.g. 16S rRNA). A primeira, é que as predições são enviesadas pela presença/ausência de genomas de referência. Essa limitação vem decrescendo conforme o número de genomas de alta qualidade vem aumentando nos bancos de dados (DOUGLAS *et al.*, 2020). A segunda, é que as predições baseadas em amplicon não possuem resolução suficiente para distinguir a funcionalidade específica em nível de cepa. Essa é uma limitação importante que se relaciona diretamente com o poder de resolução taxonômica do amplicon utilizado (DOUGLAS *et al.*, 2020), visto que mesmo dentro de uma mesma espécie, cepas diferentes podem ter perfis

funcionais/metabólicos diferentes, e isso não pode ser evidenciado através da predição baseada em amplicon.

3.5 O microbioma do trato respiratório e a COVID-19

A troca de gases (oxigênio/gás carbônico) é a principal função do trato respiratório. Para isso, as vias aéreas do ser humano possuem uma área aproximadamente 40 vezes superior à área da pele (WEIBEL, 1979). Toda essa área de superfície tecidual proporciona numerosos habitats para os microrganismos.

No trato respiratório superior, que compreende a cavidade nasal, nasofaringe e orofaringe, são encontradas elevadas densidades bacterianas (10^3 – 10^6 U⁻¹), enquanto no pulmão, que pertence ao trato respiratório inferior, as densidades bacterianas são mais modestas ($\sim 10^2$ U⁻¹). Quando em estado de eubiose, esses microrganismos atuam como ‘guardiões da saúde respiratória’ (MAN; DE STEENHUIJSEN PITERS; BOGAERT, 2017; WYPYCH; WICKRAMASINGHE; MARSLAND, 2019). Os gêneros considerados típicos de um trato respiratório superior saudável são *Staphylococcus*, *Propionibacterium*, *Leptotrichia*, *Rothia*, *Dolosigranulum*, *Haemophilus*, *Moraxella*, *Veillonella* e *Corynebacterium*, enquanto *Veillonella*, *Fusobacterium* e *Haemophilus* são típicos habitantes de um pulmão saudável. Já *Prevotella* e *Streptococcus* ocorrem tanto no trato superior quanto inferior (MAN; DE STEENHUIJSEN PITERS; BOGAERT, 2017; WYPYCH; WICKRAMASINGHE; MARSLAND, 2019).

Evidências anteriores já demonstraram que as bactérias comensais do trato respiratório ajudam a prevenir patógenos de estabelecer infecção e se espalhar pelas superfícies mucosas. Esse fenômeno, conhecido como “resistência à colonização”, também é evidenciado contra vírus respiratórios, o que leva a crer que a microbiota respiratória humana pode contribuir em certo grau com a resistência à infecção por SARS-CoV-2 (BOGAERT; DE GROOT; HERMANS, 2004; WANG *et al.*, 2022).

Sendo a microbiota responsável pela manutenção da homeostase e saúde do seu habitat, é esperado que exista uma conexão entre a microbiota e a infecção por SARS-CoV-2 (WANG *et al.*, 2022). Estudos recentes evidenciaram que tanto o microbioma intestinal quanto o respiratório sofrem importante disbiose na COVID-19, e que a composição do microbioma se correlaciona com o grau de gravidade da doença e desfecho (FERREIRA; VIANA; REIS, 2020; LLORÉNS-RICO *et al.*, 2021).

Já foi demonstrado que o microbioma intestinal pode não ter se restabelecido, mesmo 3 meses após a infecção por SARS-CoV-2 (TIAN *et al.*, 2021). Além disso, a infecção por SARS-CoV-2 predispõe os pacientes a infecções por patógenos secundários, sendo estes responsáveis por boa parte da morbimortalidade associada à COVID-19 (COX *et al.*, 2020; YEOH *et al.*, 2021).

No pulmão, ao provocar disbiose, a infecção diminui a proporção de microrganismos probióticos e aumenta a proporção de patógenos oportunistas, danifica as camadas celulares nas mucosas (AMORIM DOS SANTOS *et al.*, 2020) e destrói as junções (*Tight Junction*) entre as células epiteliais (TIAN *et al.*, 2021). Essas barreiras físicas atuam na prevenção da invasão por patógenos (OTANI; FURUSE, 2020). O transportador dependente de sódio B⁰AT1 ou SLC6A19 também podem estar relacionados com a ruptura das barreiras físicas e quebra da homeostase na infecção por SARS-CoV-2 (PENNINGER; GRANT; SUNG, 2021). B⁰AT1 também é uma chaperona da ECA-2 (KOWALCZUK *et al.*, 2008). Dessa forma, tanto ECA-2 quanto B⁰AT1 podem ser internalizadas no curso da infecção, fazendo com que a quantidade de B⁰AT1 na superfície das membranas celulares diminua. Microscopia crioeletrônica foi utilizada para avaliar as ultraestruturas do complexo ECA-2-B⁰AT1. A análise evidenciou que este complexo existe como um heterodímero, e que a proteína Spike (S1) pode interagir com este complexo (YAN *et al.*, 2020). A sub-regulação de B⁰AT1 induzida por SARS-CoV-2 na superfície das células epiteliais pode contribuir para a disbiose, o que, em contrapartida, possibilita a invasão de patógenos, facilita a “tempestade de citocinas” e promove a exacerbação da COVID-19 (VIANA; NUNES; REIS, 2020).

As células T associadas à mucosa (*Mucosal associated T cells - MAIT*) constituem um conjunto evolucionário conservado de células T com funções inatas semelhantes às das *natural killer T cells* (iNKT) (GODFREY *et al.*, 2019). Estas células estão localizadas principalmente no baço, linfonodos e fígado, mas também habitam tecidos de barreira como pele, intestino e pulmões (MEIEROVICS; YANKELEVICH; COWLEY, 2013). Estas células reagem aos patógenos via complexo maior de histocompatibilidade (MHC) mediado pela proteína 1 (MR1) que atua no reconhecimento de derivados da riboflavina produzidos por microrganismos da microbiota intestinal, tais como *Bifidobacterium animalis*, *Bacteroides thetaiotaomicron*, *Lactobacillus casei* e *Enterobacter cloacae* (CORBETT *et al.*, 2014).

Essa sinalização proveniente da microbiota afeta a biologia das células MAIT em todos os níveis, incluindo o desenvolvimento intra tímica, expansão periférica e funcionamento nos órgãos (LEGOUX; SALOU; LANTZ, 2020).

Nos tecidos, as células MAIT integram múltiplos sinais e funções efetoras associadas à defesa contra patógenos, estando altamente envolvidas na resposta imune do hospedeiro contra a COVID-19. Estas células atuam tanto de maneira local quanto sistêmica, nas fases iniciais da infecção. Elas são recrutadas por sinais pró-inflamatórios e migram do sangue para o trato respiratório, promovendo rápida resposta imune inata contra o SARS-CoV-2 (PARROT *et al.*, 2020).

Recentemente, foi demonstrado que deaminotirosina (DAT - metabólito bacteriano derivado dos flavonoides) protege o hospedeiro do Influenza através da iniciação de um *loop* de amplificação de sinalização de interferon tipo I (IFN). Os camundongos em que DAT foi administrado exibiram menor mortalidade, menor expressão de genes virais e uma menor proporção de células apoptóticas no trato respiratório (STEED *et al.*, 2017).

Outro estudo demonstrou que a sinalização de IFN nas células estromais do pulmão protege o hospedeiro contra a infecção por Influenza (BRADLEY *et al.*, 2019). SARS-CoV-2 e Influenza são parecidos em muitos quesitos, portanto, estes metabólitos derivados da microbiota podem ter um importante papel na inibição da infecção por SARS-CoV-2 (FLERLAGE *et al.*, 2021).

Outros componentes derivados da microbiota, como os lipopolissacarídeos (LPS), também ajudam na proteção dos pulmões contra infecções virais (STEFAN *et al.*, 2020). Evidências recentes apontam que estes componentes atuam na programação das células dendríticas, deixando estas células em um estado de prontidão, favorecendo uma resposta mais rápida contra patógenos, incluindo vírus respiratórios (SCHAUPP *et al.*, 2020), como o SARS-CoV-2.

A “tempestade de citocinas” pró-inflamatórias da COVID-19 está associada à doença severa e maior mortalidade. Diversos metabólitos e componentes derivados da microbiota, como vitaminas, carboidratos, aminoácidos e glicolipídeos têm o potencial para modular diferentes tipos de células imunológicas, através de diferentes mecanismos. Esses efeitos imunomodulatórios são alvos em potencial para o desenvolvimento de medidas terapêuticas ou profiláticas contra SARS-CoV-2 (WANG *et al.*, 2022).

Numerosos estudos já demonstraram que metabólitos microbianos possuem ação anti-inflamatória, tais como os ácidos graxos de cadeia curta (*short chain fat acids* - SCFA) que são produzidos por diversos grupos bacterianos. Entre esses, podemos citar acetato (50-70% - produzidos por diversos táxons), propionato (10-20% - sintetizados por Bacteroidetes e alguns Firmicutes) e butirato (10-40% - gerados por alguns Clostridia) (DALILE *et al.*, 2019). Esses SCFAs influenciam a resposta imune, tanto em nível local quanto sistêmico (FUKUDA *et al.*, 2011; SMITH *et al.*, 2013). Um estudo recente demonstrou que os SCFAs produzidos pela microbiota otimizam o metabolismo de células B e a expressão de genes associados à diferenciação de células do plasma (incluindo *Aicda*, *Xbp1*, *Irf4*, *Prdm1* e *Sdc1*), aumentando a expressão de imunoglobulinas do tipo IgG3, IgG1, IgG2b, IgG2a e IgA, dando suporte a homeostase e resposta específica de anticorpo a patógenos (KIM *et al.*, 2016). É possível que os SCFAs promovam a produção de anticorpos anti-SARS-CoV-2, inibindo o desenvolvimento da COVID-19. Pacientes recuperados da COVID-19 apresentam o metabolismo dos SCFAs pela microbiota prejudicado, o que pode indicar uma relação direta entre a gravidade da COVID-19 e o prejuízo persistente dessas importantes rotas metabólicas (VENZON; CADWELL, 2022).

O eixo intestino-pulmão (*gut-lung axis*) é uma relação já bem estabelecida em estudos de microbioma (BUDDEN *et al.*, 2017). Por isso, apesar de majoritariamente produzidos pela microbiota intestinal, existem evidências robustas de que os SCFAs podem afetar a homeostase imunológica pulmonar ao atuar de maneira sistêmica, influenciando no desenvolvimento de diversas doenças, dentre elas, a COVID-19, indicando um possível potencial terapêutico para estes componentes (ASHIQUE *et al.*, 2022). SCFAs também inibem a inflamação através da modulação de diversas células do sistema imune. O butirato promove a polarização de macrófagos M2-like, e, por consequência, atividade anti-inflamatória pela supra regulação da arginase 1 (ARG1) e sub regulação de TNF, IL-6, IL-12b e outras citocinas (SCOTT *et al.*, 2018).

Neste contexto, as células T regulatórias (Treg) possuem importante efeito imunossupressor. Uma variedade de citocinas anti-inflamatórias secretadas pelas células Treg podem inibir respostas autoinflamatórias, atuando na prevenção do dano causado pela resposta inflamatória exacerbada (ESENSTEN *et al.*, 2018). O mau funcionamento ou ausência das células Treg podem resultar em doença inflamatória (CARE *et al.*, 2018). O butirato pode promover a diferenciação de células T imaturas

em Treg, inibindo a histona desacetilase ou aumentando a transcrição do promotor Foxp3 em células T imaturas (ARPAIA *et al.*, 2013). O propionato ativa o receptor GPR43 nas células Treg, promovendo sua proliferação (MASLOWSKI *et al.*, 2009).

Outros componentes e metabólitos da microbiota também modulam as células Treg. O ácido retinóico (AR) (vitamina A) derivado de *Bifidobacterium infantis* regula a *Aldh1a2* que codifica a desidrogenase 2 retinal nas células dendríticas (CDs) (ZENG *et al.*, 2016). CDs que expressam *Aldh1a2* produzem altos níveis de AR e essa substância colabora com o fator de crescimento transformante- β (*transforming growth factor- β* - TGF- β) para promover a diferenciação de células T imaturas em células Treg FOXP3+ (YOKOTA *et al.*, 2009). O polissacarídeo A (PSA), componente capsular de *Bacteroides fragilis*, pode ser capturado no lúmen intestinal por proteínas relacionadas à autofagia (CHU *et al.*, 2016) e reconhecido pelo receptor Toll-like 2 (TLR2) nas células Treg FOXP3+, o que induz a proliferação destas células, produção de IL-10 e promove um estado anti-inflamatório (ROUND; MAZMANIAN, 2010). O polissacarídeo de superfície beta glucana/galactana produzido por *Bifidobacterium bifidum* promove a regeneração das células Treg FOXP3+ (WANG *et al.*, 2022).

Além disso, o receptor do AR relacionado ao receptor órfão gamma t (ROR γ t, um receptor hormonal nuclear) pode induzir a diferenciação das células T *helper* 17 (T_H17) (ZHOU *et al.*, 2008). Estudos recentes demonstraram que certos comensais intestinais como *Helicobacter* spp. e *Clostridium ramosum* podem induzir a expressão de ROR γ t nas células Treg Foxp3+. Células Treg ROR γ t+ Foxp3+ sub regulam T_H1-, T_H2- e respostas imunes produzidas por células T_H17 (XU *et al.*, 2018).

Devido a essas características, terapias baseadas em microbioma vem sendo utilizadas no tratamento de diversas doenças como diabetes, obesidade, câncer, colite ulcerativa, doença de Crohn e certas infecções virais (SORBARA; PAMER, 2022). Alguns desses estudos atuaram na manipulação do microbioma para tratar a COVID-19 e suas complicações (CONTE; TORALDO, 2020; INFUSINO *et al.*, 2020; MAK; CHAN; NG, 2020).

Recentemente, um ensaio clínico (ClinicalTrials.gov Identifier No. NCT04824222) propôs validar a eficácia do transplante de microbiota fecal (FMT – *fecal microbiota transplantation*) como imunomodulador para reduzir o risco de agravamento da COVID-19 (TANG *et al.*, 2021). O estágio do estudo é desconhecido e os resultados ainda não foram publicados, mas é possível que o FMT seja uma

abordagem terapêutica viável na supressão da tempestade de citocinas e inflamação induzidas pela COVID-19 (WANG *et al.*, 2022).

A suplementação com prebióticos e probióticos, mais seguros e fáceis de administrar que o FMT, também vem sendo considerada uma abordagem terapêutica promissora contra a COVID-19 (GIANNONI *et al.*, 2020; MAK; CHAN; NG, 2020). A Comissão Nacional de Saúde da China recomenda a administração de probióticos para pacientes com COVID-19 severa com o propósito de restabelecer e manter a homeostase da microbiota e prevenir infecções secundárias. De fato, diversos ensaios clínicos estão investigando a eficácia dos probióticos e/ou prebióticos na redução dos sintomas e do tempo de curso da COVID-19. Apesar de ainda não publicados, já foi empiricamente demonstrado que certas cepas probióticas possuem atividade antiviral contra outros coronavírus, demonstrando o potencial desse tipo de terapia na prevenção ou como adjuvante no tratamento da COVID-19 (WANG *et al.*, 2022).

Os avanços na biologia sintética e manipulação genética vem facilitando e permitindo o *design* de microrganismos especificamente para tratamento da COVID-19. Já é possível produzir microrganismos com as funções que desejamos, como a capacidade de produzir certos metabólitos, ou de atingir determinado alvo terapêutico (WANG *et al.*, 2022). Uma cepa de *Lactococcus lactis* foi desenhada para expressar e secretar a citocina anti-inflamatória IL-10 para o tratamento de colite. A biossegurança dessa cepa foi garantida ao fazer com que essa cepa necessite de timidina exógena para sobrevivência e produção da IL-10 (BRAAT *et al.*, 2006). Como a severidade e mortalidade da COVID-19 possui relação direta com a tempestade de citocinas, é possível utilizar abordagens semelhantes para criar cepas que produzam metabólitos anti-inflamatórios na intenção de desenvolver uma nova abordagem terapêutica para esta doença. Enquanto muito ainda se faz necessário para garantir a segurança e eficácia dessa abordagem, a suplementação direta de metabólitos derivados da microbiota, como SCFAs, constitui uma abordagem terapêutica mais próxima e factível (WANG *et al.*, 2022).

Alguns estudos também demonstram que a microbiota possui um papel importante na resposta de anticorpo à vacinação (HAGAN *et al.*, 2019; HARRIS *et al.*, 2018; SANTANA-MEDEROS *et al.*, 2022). Ou seja, além da possibilidade de uso das terapias baseadas em microbioma para prevenção e tratamento da COVID-19, deve

ser considerada a possibilidade desse tipo de terapia na modulação da resposta imune no intuito de otimizar a efetividade das vacinas (WANG *et al.*, 2022).

Dado o robusto corpo de evidências, é bem possível que o microbioma, tanto respiratório quanto intestinal, possuam um papel fundamental na infecção por SARS-CoV-2, seja na proteção contra a infecção inicial, desenvolvimento, agravamento da doença ou na modulação da resposta às medidas terapêuticas, podendo servir como biomarcadores prognósticos, de resposta imune e tratamento para a COVID-19 (ZHU *et al.*, 2022).

4 MATERIAIS E MÉTODOS

4.1 Delineamento do estudo e tamanho amostral

O projeto contemplou duas abordagens distintas: um estudo observacional analítico de microbioma e um estudo de desenvolvimento de método de diagnóstico (abordagem proteômica), conforme apresentado na figura 2. As amostras usadas neste estudo foram obtidas a partir de *swab* combinado de nasofaringe e orofaringe cujo resultado do RT-qPCR para SARS-CoV-2 estava disponível (CDC, 2020). Estas amostras foram fornecidas pelo Biobanco do Hospital de Clínicas de Porto Alegre (Biobanco/HCPA) e pelo Laboratório Exame.

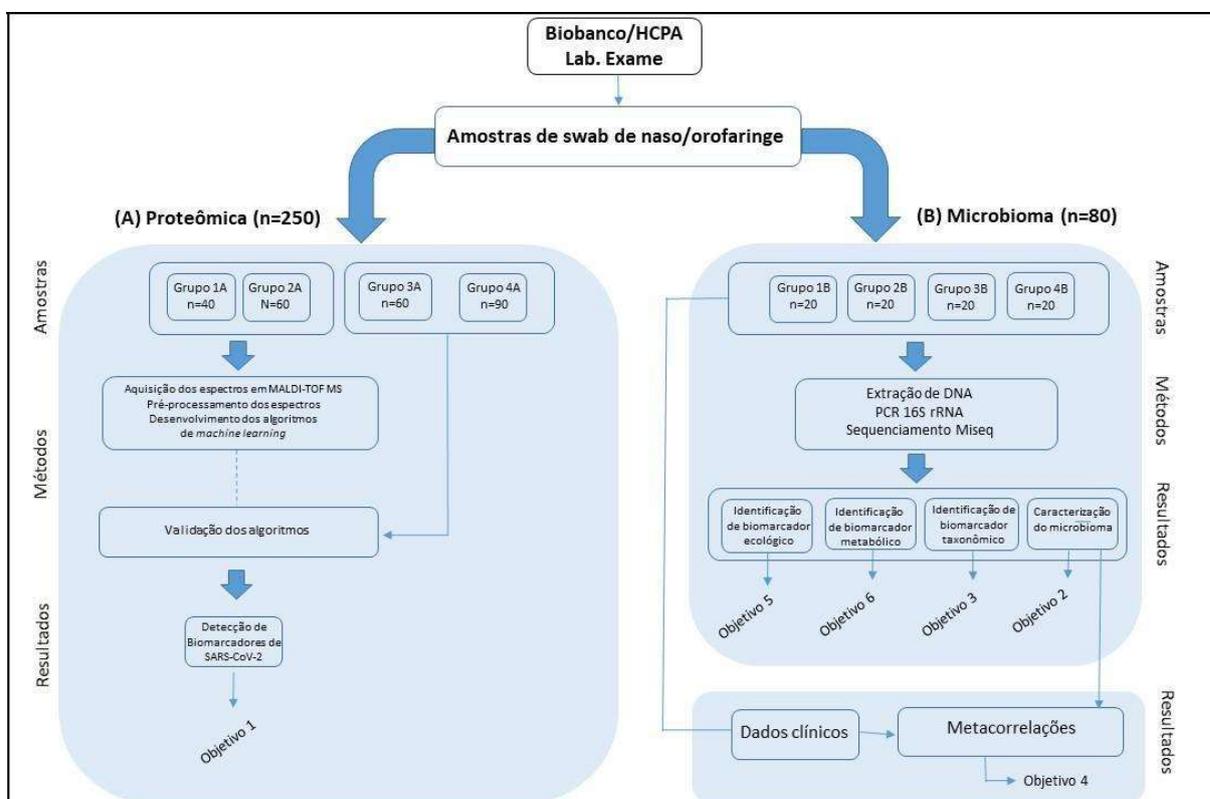


Figura 2: Fluxograma representativo do delineamento do estudo e tamanho amostral. A figura apresenta os grupos amostrais e os fluxos de trabalho que foram adotados nas análises Proteômica (A) e Metagenômica (B; Microbioma).

A metodologia detalhada de cada estudo está apresentada nos encartes das publicações a seguir.

5 RESULTADOS

Manuscrito 1 (Objetivo Especifico 1)

Título: *Validation of a MALDI-TOF MS method for SARS-CoV-2 detection on the Bruker Biotyper and nasopharyngeal swabs. A Brazil - UK collaborative study.*

Intervalo de páginas suprimido: 55 - 65

Referência

Lovison, O.A.; Grigaite, R.; Volpato, F.C.Z.; Iles, J.K.; Lacey, J.; Barreto, F.; Pandiri, S.R.; Balzan, L.d.L.R.; Cantarelli, V.V.; Barth, A.L.; Iles, R.K.; Martins, A.F. Validation of a MALDI-TOF MS Method for SARS-CoV-2 Detection on the Bruker Biotyper and Nasopharyngeal Swabs: A Brazil - UK Collaborative Study. *Diagnostics* 2023, 13, 1470. <https://doi.org/10.3390/diagnostics13081470>

Manuscrito 2 (Objetivos Específicos 2 a 4).

Artigo submetido para revista *Nature Microbiology* (IF 30.964)

Título: *Upper Respiratory Tract Microbiome Association to Covid-19 Severity*

Intervalo de páginas suprimido: 67 - 98

O microbioma humano influencia significativamente as funções imunes e metabólicas do hospedeiro. Estudos anteriores sugerem que as bactérias comensais no trato respiratório superior podem proteger contra a colonização e infecção por patógenos, incluindo vírus respiratórios. Dada a modulação imune mediada pelo microbioma, espera-se uma ligação entre o microbioma do trato respiratório superior e a susceptibilidade e gravidade da COVID-19. Este estudo transversal avaliou a composição do microbioma do trato respiratório superior, a predição metabólica e a associação com biomarcadores laboratoriais em pneumonia não-COVID-19 (NO-CoV), COVID-19 moderada (M-CoV), COVID-19 grave (S-CoV) e indivíduos altamente expostos assintomáticos e negativos para COVID-19 (NC). Os pacientes com S-CoV apresentam redução na diversidade do microbioma, principalmente devido à diminuição na abundância de táxons eubióticos. Alguns desses táxons também estavam associados a biomarcadores inflamatórios. Múltiplas vias metabólicas ligadas à resposta imune, atividade antiviral e susceptibilidade do hospedeiro mostraram diminuição na abundância em S-CoV. Essas vias podem sugerir potenciais alternativas para o arsenal terapêutico contra a COVID-19.

6 DISCUSSÃO GERAL

De 11 de março de 2020 a 5 de maio de 2023. Três anos e 3 meses. Foi o tempo que durou a maior emergência de saúde pública mundial do nosso tempo. Ao contrário do que muitos pensam, a pandemia da COVID-19 não acabou, apenas o seu status de emergência global (HARRIS, 2023). O SARS-CoV-2 continuará circulando de maneira endêmica, e a vigilância epidemiológica permanente se faz necessária, para que possamos monitorar o número de casos, surgimento de novas variantes, e antecipar eventuais tomadas de decisão (BARREIRO; SAN ROMÁN, 2022). Para além da epidemiologia, a busca por novos métodos de diagnóstico e novos alvos terapêuticos não deve ser interrompida, tanto pela possibilidade da ocorrência de novas emergências globais, quanto pelo desenvolvimento em si, visto que o desenvolvimento gerado pela força da pandemia será aproveitado para inúmeras outras situações.

No curso da pandemia, houve diversos momentos de escassez de insumos, impactando principalmente países de baixa e média renda, que, em função da sua menor infraestrutura tecnológica, são altamente dependentes de importação (VOLPATO *et al.*, 2021). É aqui que entra a necessidade do desenvolvimento de alternativas diagnósticas aplicáveis ao SUS.

A tecnologia de MALDI-TOF MS se enquadra perfeitamente a este cenário. É uma tecnologia de fácil operação, baixo custo, rápida e seus insumos são de fácil aquisição (SIVANESAN *et al.*, 2022). Neste trabalho, demonstramos ser possível a detecção de um vírus, o SARS-CoV-2, utilizando esta tecnologia. Além disso, foi possível detectar a presença de marcadores de resposta imune, o que abre a possibilidade para o desenvolvimento de métodos que utilizem essa capacidade para realizar inferências diversas. Ao aliarmos o poder do MALDI-TOF MS com aprendizagem de máquina, foi possível evidenciar os ganhos de sensibilidade e especificidade proporcionados por essa estratégia. Sem dúvidas o uso da inteligência artificial (IA) será cada vez mais presente na nossa rotina, e muito provavelmente em breve teremos amplos modelos de IA realizando diagnósticos automatizados a partir dos mais variados tipos de dados.

Outro ponto fundamental do nosso trabalho foi a colaboração internacional. Ao aliarmos a propriedade intelectual e a expertise de diversos pesquisadores de alto nível, foi possível, mesmo à distância, obter sucesso em uma demanda que, caso

essa colaboração não existisse, com certeza seria muito mais difícil de atingir. O resultado foi um método de detecção de SARS-CoV-2 em MALDI-TOF MS com potencial para ser amplamente utilizado em rotinas laboratoriais, e ainda com a possibilidade de ser otimizado, não só em performance, mas para detecção de alvos adicionais. Além disso, a mesma estratégia pode vir a ser utilizada no desenvolvimento de métodos de diagnóstico para outros vírus endêmicos no Brasil, seja na clínica humana, veterinária ou no setor produtivo agropecuário.

A relação entre o microbioma e a saúde do hospedeiro está bem estabelecida, apesar de pouco explorada. As evidências apontam para a possibilidade do uso da modulação do microbioma para a prevenção e/ou tratamento de diversas doenças (OGUNRINOLA *et al.*, 2020; ROOKS; GARRETT, 2016; ZHENG; LIWINSKI; ELINAV, 2020). Além disso, os estudos de microbioma tem o potencial para apontar rotas metabólicas ou seus produtos como terapias a serem exploradas, gerando um corpo de evidências de proporção quase inigualável visto o investimento necessário para execução deste tipo de estudo.

Evidências anteriores apontam que tanto o microbioma intestinal quanto o respiratório sofrem importante disbiose na COVID-19, e que a composição do microbioma se correlaciona com a gravidade da doença e desfecho (FERREIRA; VIANA; REIS, 2020; LLORÉNS-RICO *et al.*, 2021).

Neste estudo, foi avaliada a composição do microbioma do trato respiratório superior e a predição metabólica para pacientes com COVID-19 e não-COVID-19 usando uma abordagem CODA para levar em conta a natureza composicional intrínseca dos dados de sequenciamento (GLOOR *et al.*, 2017). A abordagem CODA foi relatada como o padrão ouro para analisar dados de microbiomas (e quaisquer outros dados gerados a partir de plataformas de sequenciamento de próxima geração (NGS)) (GREENACRE; MARTÍNEZ-ÁLVARO; BLASCO, 2021; SWARTE *et al.*, 2023), uma vez que apresenta coerência estatística (os métodos estatísticos tradicionais podem não levar em conta a natureza composicional, levando a resultados tendenciosos e correlações espúrias (AITCHISON, 1982; AITCHISON *et al.*, 2000; PEARSON, 1896), evita o problema da inferência composicional (mudanças na abundância de um táxon podem afetar artificialmente a abundância aparente de outros), melhor interpretabilidade (em vez de lidar com valores de abundância absoluta, os pesquisadores podem interpretar as mudanças nas razões logarítmicas

entre táxons microbianos, proporcionando uma compreensão mais clara das mudanças de composição), lida com a dispersão (grande número de táxons de baixa prevalência) e dados zero-inflados, características inerentes aos conjuntos de dados de microbiomas (BARS-CORTINA, 2022; CALLE, 2019; GLOOR *et al.*, 2017; GREENACRE; MARTÍNEZ-ÁLVARO; BLASCO, 2021).

Dentre os nossos achados, conseguimos evidenciar táxons, rotas metabólicas e componentes derivados da microbiota com sugestivo potencial protetor do trato respiratório superior, atividade antiviral, atividade imunomoduladora, dentre outros, que apesar de relacionados com a suscetibilidade do hospedeiro ou com a severidade da doença, são ainda pouco conhecidos e com mecanismos de ação a serem esclarecidos. Ainda foi possível relacionar, de maneira exploratória, a associação de alguns destes táxons com biomarcadores laboratoriais previamente correlacionados com a severidade da COVID-19.

O desenho do nosso estudo nos permitiu observar que as principais alterações na composição e predição metabólica do microbioma se relacionam mais com a severidade do quadro do que com o patógeno em si. Faz sentido, se pensarmos na natureza inata do microbioma. Se isso de fato se replicar ao longo de outros estudos com diferentes patógenos, os achados ganharão robustez significativa, identificando potenciais marcadores e novas abordagens terapêuticas. Infelizmente ainda são poucos os estudos que realizam o adequado pareamento entre caso e controle, principalmente nas ciências ômicas.

Ao corroborar nossos achados com o extenso corpo de evidências da literatura, apesar da falta de padronização tecnológica e metodológica, existem achados que são reprodutíveis, o que torna a análise de microbioma uma importante ferramenta na busca por alternativas terapêuticas. Ainda não está claro quais alterações são características de um estado prévio de doença/disbiose, e quais alterações ocorrem em resposta à doença, podendo de fato serem benéficas ao hospedeiro, apesar de relacionadas a maior severidade. Um melhor entendimento acerca da relação microbioma - hospedeiro - doença se faz necessário para um melhor direcionamento de novas abordagens terapêuticas e estabelecimento de biomarcadores.

7 CONCLUSÕES E PERSPECTIVAS

A tecnologia de MALDI-TOF MS é uma revolução no campo da medicina laboratorial e pesquisa. Apesar de ainda ser encontrada somente em grandes centros urbanos no Brasil, aos poucos vem se popularizando, e tem seu potencial ainda pouco explorado. O fato de ter sido possível identificar biomarcadores virais (e.g. fragmentos da proteína Spike) e de resposta imune (e.g. cadeias de imunoglobulinas) em um único espectro, abre um leque imenso de possibilidades para o uso desta tecnologia. Por ser uma tecnologia rápida, de baixo custo, de fácil operação, e que utiliza reagentes genéricos, se apresenta como uma alternativa aos métodos moleculares em momentos de elevada demanda e escassez de recursos. Além disso, em função da característica *shotgun* de MALDI-TOF MS, o método pode ser multiplexado com certa facilidade, e pode se apresentar particularmente valioso para alvos em que a elevada taxa de mutação possa vir a provocar perda de sensibilidade nos métodos moleculares.

A relação entre microbioma, resposta imune e desfecho já está bem estabelecida para uma série de patologias. No que tange a COVID-19, foi possível evidenciar uma relação da severidade da doença com a diversidade alfa, sendo que táxons considerados eubióticos apresentaram abundância diferencial reduzida na COVID-19 severa. Alguns desses táxons, como *Corynebacteriales* e *Peptostreptococcaceae*, aparentam estar mais relacionados com a suscetibilidade do hospedeiro, atuando como uma primeira linha de defesa contra infecções do trato respiratório. Este achado encontra respaldo em importantes rotas metabólicas potencialmente expressas por estes táxons, relacionadas com a produção de mucinas e ácidos graxos, que atuam na resistência à colonização no trato respiratório superior.

Outros táxons, como *Lachnospiraceae*, *Neisseria* sp. e *Propionibacteriaceae* aparentam estar mais relacionados com a modulação da resposta imune do hospedeiro. Seja atuando na produção de ácidos graxos de cadeia curta, de anticorpos anti-SARS-CoV-2, ou na biossíntese de vitaminas com conhecida atividade antiviral e imunomoduladora, as evidências apontam para relação destes táxons com o grau de severidade da doença. Estes achados abrem caminho para que um potencial uso destes táxons como probióticos seja explorado.

Além disso, uma ampla gama de componentes e rotas metabólicas, como aminoácidos, metabolismo de purinas, ou o próprio lipopolissacarídeo, conhecido

antígeno bacteriano, se apresentaram como alvos terapêuticos em potencial para o tratamento ou prevenção da COVID-19, sendo que muitos deles já vêm sendo testados, inclusive em ensaios clínicos randomizados.

O corpo de evidências gerado a partir destes estudos é amplo e diverso, e a tarefa mais difícil é traduzir estes achados para a prática clínica. A geração de evidências reprodutíveis e aplicáveis em larga escala é fundamental para a manutenção do mundo acadêmico e da fé da sociedade na ciência, mas isto vem sendo dificultado por uma academia produtivista, onde quantidade tem mais valor que qualidade. Esperamos que, em breve, alguns dos nossos achados possam, de fato, ser alternativas viáveis no diagnóstico, prevenção ou tratamento de doenças infecciosas.

8 OUTROS ARTIGOS PUBLICADOS

Em 18 de março de 2020 a pandemia da COVID-19 chegou até nós. Todas as atividades foram suspensas, sem previsão de retorno. Em meados de maio de 2020, 6 meses após o meu ingresso como doutorando, surgiu a proposta de mudança de projeto. Ao longo do ano de 2020 e boa parte de 2021 (totalizando mais de 1 ano de paralisação de projeto), em função do cancelamento de atividades por causa da pandemia; e mesmo após, em função das medidas de contingência, que determinavam um número máximo de pessoas por laboratório; e da orientação recebida de que a prioridade deveria ser dos alunos que possuíam os prazos mais próximos para suas defesas (e não dos projetos relacionados à COVID-19); as atividades relacionadas ao meu projeto ficaram bastante limitadas.

Felizmente, em outubro de 2020 a professora Dra. Andreza Francisco Martins passa a integrar o Núcleo de Bioinformática do HCPA, como líder do time de metagenômica. Junto com ela, dois de seus alunos, Otávio von Ameln Lovison, autor desta tese, e Rafaela Ramalho Guerra, passaram a atender demandas de metagenômica internas e externas ao HCPA.

Este fato permitiu que eu me envolvesse em diversos projetos, me aprofundando em bioinformática, aprendendo as ferramentas e desenvolvendo *pipelines* que viriam a ser utilizados nas análises deste projeto. Abaixo estão citados os artigos publicados não relacionados com a tese, em que participei ativamente desde as etapas laboratoriais, de bancada, quanto nas etapas de análise. Os encartes das publicações estão disponíveis em anexos.

Título: *Genomic Surveillance of SARS-CoV-2 Lineages Indicates Early Circulation of P.1 (Gamma) Variant of Concern in Southern Brazil.* (Anexo 1)

Estudo já publicado no periódico *Microbiology Spectrum* (IF 9.043). O destaque para esta demanda fica no intenso trabalho de bioinformática envolvido. A análise genômica e a modelagem do relógio molecular são processos extremamente complexos que exigem bastante tempo, raciocínio e dedicação do pesquisador. Através dessas análises foi possível inferir que a evolução da variante Gamma de SARS-CoV-2 foi um processo gradual e contínuo, em acompanhamento à disseminação comunitária do vírus.

Título: Identification of pesticides in water samples by solid-phase extraction and liquid chromatography–electrospray ionization mass spectrometry. (Anexo 2)

Artigo publicado como primeiro autor na revista *Water Environment Research* (IF 3.3) que tem como destaque a modelagem estatística utilizada, que realizou a inferência de um coeficiente de risco e delineou um *worst case scenario* para a presença de pesticidas em água para consumo humano.

Título: NBioinfo: Establishing a Bioinformatics Core in a University-based General Hospital in South Brazil. (Anexo 3)

Artigo escrito em conjunto com outros times do NBioinfo a convite da *Journal of Information and Data Management* para uma *special issue* abordando serviços de bioinformática implementados no Brasil. Aceito para publicação.

Título: Can the data of microbiome be used to predict the presence of Burkholderia spp. in pulmonary microbiota of Cystic Fibrosis patients? (Anexo 4)

Artigo publicado no *Brazilian Journal of Infectious Diseases* (IF 3.4) que levanta a possibilidade do uso do NGS como ferramenta de diagnóstico precoce para infecções do trato respiratório, geralmente polimicrobiana, de pacientes com fibrose cística.

Título: Gut microbiota transfer evidence from mother to newborn. (Anexo 5)

Artigo publicado na *European Journal of Pediatrics* (IF 3.6) que tem como destaque uma modelagem estatística baseada em permutação, em que foram computadas as probabilidades de transferência dos componentes da microbiota da mãe para o bebê, a probabilidade de transferência relacionada com a abundância do microrganismo, e a probabilidade de transferência ponderada, correlacionando com a via de parto.

9 REFERÊNCIAS

AITCHISON, J. The Statistical Analysis of Compositional Data. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 44, n. 2, p. 139–160, 1982.

AITCHISON, J. et al. Logratio Analysis and Compositional Distance. **Mathematical Geology**, 2000.

AITCHISON, J.; GREENACRE, M. Biplots of Compositional Data. **Journal of the Royal Statistical Society Series C: Applied Statistics**, v. 51, n. 4, p. 375–392, 1 out. 2002.

ALCOBA-FLOREZ, J. et al. Fast SARS-CoV-2 detection by RT-qPCR in preheated nasopharyngeal swab samples. **International Journal of Infectious Diseases**, v. 97, p. 66–68, ago. 2020.

AMORIM DOS SANTOS, J. et al. Oral mucosal lesions in a COVID-19 patient: New signs or secondary manifestations? **International Journal of Infectious Diseases**, v. 97, p. 326–328, ago. 2020.

ANDERSON, M. J. et al. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist: Roadmap for beta diversity. **Ecology Letters**, v. 14, n. 1, p. 19–28, jan. 2011.

ARPAIA, N. et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. **Nature**, v. 504, n. 7480, p. 451–455, dez. 2013.

ASHIQUE, S. et al. Short Chain Fatty Acids: Fundamental mediators of the gut-lung axis and their involvement in pulmonary diseases. **Chemico-Biological Interactions**, v. 368, p. 110231, dez. 2022.

ASSHAUER, K. P. et al. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data: Fig. 1. **Bioinformatics**, v. 31, n. 17, p. 2882–2884, 1 set. 2015.

BARBERA, P. et al. **Data from: EPA-ng: massively parallel evolutionary placement of genetic sequences**. Dryad, , 2018. Disponível em: <<http://datadryad.org/stash/dataset/doi:10.5061/dryad.kb505nc>>. Acesso em: 23 dez. 2022

BARREIRO, P.; SAN ROMÁN, J. Vulnerability to SARS-CoV-2 infection and disease: ripping the curl after the storm. **Revista Española de Quimioterapia**, v. 35, n. Suppl3, p. 2–5, 24 out. 2022.

BARS-CORTINA, D. **Alpha and beta-diversities performance comparison between different normalization methods and centered log-ratio transformation in a microbiome public dataset**. [s.l.] Microbiology, 7 nov. 2022. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/2022.11.07.512066>>. Acesso em: 14 jan. 2024.

BERG, G. et al. Microbiome definition re-visited: old concepts and new challenges. **Microbiome**, v. 8, n. 1, p. 103, dez. 2020.

- BINNICKER, M. J. et al. Identification of an Influenza A H1N1/2009 Virus with Mutations in the Matrix Gene Causing a Negative Result by a Commercial Molecular Assay. **Journal of Clinical Microbiology**, v. 51, n. 6, p. 2006–2007, jun. 2013.
- BOGAERT, D.; DE GROOT, R.; HERMANS, P. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. **The Lancet Infectious Diseases**, v. 4, n. 3, p. 144–154, mar. 2004.
- BOSCH, T. C. G.; MCFALL-NGAI, M. J. Metaorganisms as the new frontier. **Zoology**, v. 114, n. 4, p. 185–190, set. 2011.
- BRAAT, H. et al. A Phase I Trial With Transgenic Bacteria Expressing Interleukin-10 in Crohn's Disease. **Clinical Gastroenterology and Hepatology**, v. 4, n. 6, p. 754–759, jun. 2006.
- BRADLEY, K. C. et al. Microbiota-Driven Tonic Interferon Signals in Lung Stromal Cells Protect from Influenza Virus Infection. **Cell Reports**, v. 28, n. 1, p. 245–256.e4, jul. 2019.
- BRAY, J. R.; CURTIS, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. **Ecological Monographs**, v. 27, n. 4, p. 325–349, 1957.
- BUDDEN, K. F. et al. Emerging pathogenic links between microbiota and the gut–lung axis. **Nature Reviews Microbiology**, v. 15, n. 1, p. 55–63, jan. 2017.
- BULLARD, J. H. et al. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. **BMC Bioinformatics**, v. 11, n. 1, p. 94, dez. 2010.
- CALLAHAN, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. **Nature Methods**, v. 13, n. 7, p. 581–583, jul. 2016a.
- CALLAHAN, B. J. et al. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. **F1000Research**, v. 5, p. 1492, 2 nov. 2016b.
- CALLAHAN, B. J.; MCMURDIE, P. J.; HOLMES, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. **The ISME Journal**, v. 11, n. 12, p. 2639–2643, dez. 2017.
- CALLE, M. L. Statistical Analysis of Metagenomics Data. **Genomics & Informatics**, v. 17, n. 1, p. e6, 31 mar. 2019.
- CALLE, M. L.; PUJOLASSOS, M.; SUSIN, A. coda4microbiome: compositional data analysis for microbiome cross-sectional and longitudinal studies. **BMC Bioinformatics**, v. 24, n. 1, p. 82, 6 mar. 2023.
- CARE, A. S. et al. Reduction in Regulatory T Cells in Early Pregnancy Causes Uterine Artery Dysfunction in Mice. **Hypertension**, v. 72, n. 1, p. 177–187, jul. 2018.
- CARTER, L. J. et al. Assay Techniques and Test Development for COVID-19 Diagnosis. **ACS Central Science**, v. 6, n. 5, p. 591–605, 27 maio 2020.

- CASPI, R. et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. **Nucleic Acids Research**, v. 44, n. D1, p. D471–D480, 4 jan. 2016.
- CHAO, A. et al. Abundance-Based Similarity Indices and Their Estimation When There Are Unseen Species in Samples. **Biometrics**, v. 62, n. 2, p. 361–371, jun. 2006.
- CHEN, B. et al. Comparison of beta diversity measures in clustering the high-dimensional microbial data. **PLOS ONE**, v. 16, n. 2, p. e0246893, 18 fev. 2021.
- CHEN, N. et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. **The Lancet**, v. 395, n. 10223, p. 507–513, fev. 2020.
- CHIARELLO, M. et al. Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. **PLOS ONE**, v. 17, n. 2, p. e0264443, 24 fev. 2022.
- CHU, H. et al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. **Science**, v. 352, n. 6289, p. 1116–1120, 27 maio 2016.
- CONTE, L.; TORALDO, D. M. Targeting the gut–lung microbiota axis by means of a high-fibre diet and probiotics may have anti-inflammatory effects in COVID-19 infection. **Therapeutic Advances in Respiratory Disease**, v. 14, p. 175346662093717, jan. 2020.
- CORBETT, A. J. et al. T-cell activation by transitory neo-antigens derived from distinct microbial pathways. **Nature**, v. 509, n. 7500, p. 361–365, maio 2014.
- COSTEA, P. I. et al. Enterotypes in the landscape of gut microbial community composition. **Nature Microbiology**, v. 3, n. 1, p. 8–16, jan. 2018.
- COX, M. J. et al. Co-infections: potentially lethal and unexplored in COVID-19. **The Lancet Microbe**, v. 1, n. 1, p. e11, maio 2020.
- CROXATTO, A.; PROD’HOM, G.; GREUB, G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. **FEMS Microbiology Reviews**, v. 36, n. 2, p. 380–407, mar. 2012.
- CZECH, L.; STAMATAKIS, A. Scalable methods for analyzing and visualizing phylogenetic placement of metagenomic samples. **PLOS ONE**, v. 14, n. 5, p. e0217050, 28 maio 2019.
- DALILE, B. et al. The role of short-chain fatty acids in microbiota–gut–brain communication. **Nature Reviews Gastroenterology & Hepatology**, v. 16, n. 8, p. 461–478, ago. 2019.
- DAVIS, N. M. et al. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. **Microbiome**, v. 6, n. 1, p. 226, dez. 2018.

- DEEKS, J. J. et al. Letter to the Editor regarding Peto T; UK COVID-19 Lateral Flow Oversight Team: COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay. **eClinicalMedicine**, v. 38, p. 101037, ago. 2021.
- DEULOFEU, M. et al. Detection of SARS-CoV-2 Infection in Human Nasopharyngeal Samples by Combining MALDI-TOF MS and Artificial Intelligence. **Frontiers in Medicine**, v. 8, p. 661358, 1 abr. 2021.
- DIRICAN, E.; BAL, T. COVID-19 disease severity to predict persistent symptoms: a systematic review and meta-analysis. **Primary Health Care Research & Development**, v. 23, p. e69, 2022.
- DO, T. et al. Use of MALDI-TOF mass spectrometry for virus identification: a review. **Analyst Review Articles 2022**, v. 14, n. 147, p. 3131–3154, 2022.
- DOUGLAS, G. M. et al. PICRUSt2 for prediction of metagenome functions. **Nature Biotechnology**, v. 38, n. 6, p. 685–688, jun. 2020.
- EDGAR, R. C. **UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing**. [s.l.] Bioinformatics, 15 out. 2016. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/081257>>. Acesso em: 5 jan. 2023.
- EDGAR, R. C. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. **Bioinformatics**, v. 34, n. 14, p. 2371–2375, 15 jul. 2018.
- EDGAR, R. C.; FLYVBJERG, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. **Bioinformatics**, v. 31, n. 21, p. 3476–3482, 1 nov. 2015.
- ENDRES, D. M.; SCHINDELIN, J. E. A new metric for probability distributions. **IEEE Transactions on Information Theory**, v. 49, n. 7, p. 1858–1860, jul. 2003.
- ESENSTEN, J. H. et al. Regulatory T-cell therapy for autoimmune and autoinflammatory diseases: The next frontier. **Journal of Allergy and Clinical Immunology**, v. 142, n. 6, p. 1710–1718, dez. 2018.
- FERREIRA, C.; VIANA, S. D.; REIS, F. Is Gut Microbiota Dysbiosis a Predictor of Increased Susceptibility to Poor Outcome of COVID-19 Patients? An Update. **Microorganisms**, v. 9, n. 1, p. 53, 28 dez. 2020.
- FERREIRA-SANTOS, D.; MARANHÃO, P.; MONTEIRO-SOARES, M. Identifying common baseline clinical features of COVID-19: a scoping review. **BMJ Open**, v. 10, n. 9, p. e041079, set. 2020.
- FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: interactive sequence similarity searching. **Nucleic Acids Research**, v. 39, n. suppl, p. W29–W37, 1 jul. 2011.
- FITZGERALD, D. A.; WONG, G. W. K. COVID-19: A tale of two pandemics across the Asia Pacific region. **Paediatric Respiratory Reviews**, v. 35, p. 75–80, set. 2020.

- FLERLAGE, T. et al. Influenza virus and SARS-CoV-2: pathogenesis and host responses in the respiratory tract. **Nature Reviews Microbiology**, v. 19, n. 7, p. 425–441, jul. 2021.
- FRANZOSA, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. **Nature Methods**, v. 15, n. 11, p. 962–968, nov. 2018.
- FUKUDA, S. et al. Bifidobacteria can protect from enteropathogenic infection through production of acetate. **Nature**, v. 469, n. 7331, p. 543–547, jan. 2011.
- GIANNONI, E. et al. Probiotics and COVID-19. **The Lancet Gastroenterology & Hepatology**, v. 5, n. 8, p. 720–721, ago. 2020.
- GLOOR, G. B. et al. Microbiome Datasets Are Compositional: And This Is Not Optional. **Frontiers in Microbiology**, v. 8, p. 2224, 15 nov. 2017.
- GODFREY, D. I. et al. The biology and functional importance of MAIT cells. **Nature Immunology**, v. 20, n. 9, p. 1110–1128, set. 2019.
- GOWER, J. C. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. **Biometrika**, v. 53, n. 3/4, p. 325–338, 1966.
- GREENACRE, M.; MARTÍNEZ-ÁLVARO, M.; BLASCO, A. Compositional Data Analysis of Microbiome and Any-Omics Datasets: A Validation of the Additive Logratio Transformation. **Frontiers in Microbiology**, v. 12, p. 727398, 11 out. 2021.
- HAEGEMAN, B. et al. Robust estimation of microbial diversity in theory and in practice. **The ISME Journal**, v. 7, n. 6, p. 1092–1101, jun. 2013.
- HAGAN, T. et al. Antibiotics-Driven Gut Microbiome Perturbation Alters Immunity to Vaccines in Humans. **Cell**, v. 178, n. 6, p. 1313- 1328.e13, set. 2019.
- HALL, M. W. et al. Immune Immunomodulation in Coronavirus Disease 2019 (COVID-19): Strategic Considerations for Personalized Therapeutic Intervention. **Clinical Infectious Diseases**, v. 74, n. 1, p. 144–148, 7 jan. 2022.
- HARRINGTON, A. et al. Comparison of Abbott ID Now and Abbott m2000 Methods for the Detection of SARS-CoV-2 from Nasopharyngeal and Nasal Swabs from Symptomatic Patients. **Journal of Clinical Microbiology**, v. 58, n. 8, p. e00798-20, 23 jul. 2020.
- HARRIS, E. WHO Declares End of COVID-19 Global Health Emergency. **JAMA**, v. 329, n. 21, p. 1817, 2023.
- HARRIS, V. C. et al. Effect of Antibiotic-Mediated Microbiome Modulation on Rotavirus Vaccine Immunogenicity: A Human, Randomized-Control Proof-of-Concept Trial. **Cell Host & Microbe**, v. 24, n. 2, p. 197- 207.e4, ago. 2018.
- HOLLAND, R. D. et al. Rapid Identification of Intact Whole Bacteria Based on Spectral Patterns using Matrix-assisted Laser Desorption/Ionization with Time-of-flight Mass Spectrometry. **Rapid Communications in Mass Spectrometry**, v. 10, n. 10, p. 1227–1232, 31 jul. 1996.

- HOLMES, S. Successful strategies for human microbiome data generation, storage and analyses. **Journal of Biosciences**, v. 44, n. 5, p. 111, out. 2019.
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. **The Journal of Educational Psychology**, v. 24(6), p. 417–441, 1933.
- ILES, R. K. et al. Development of a clinical MALDI-ToF Mass Spectrometry assay for SARS-CoV-2: Rational design and multi-disciplinary team work. **Diagnostics**, p. 16, 2020.
- ILES, R. K. et al. A How to Guide: Clinical Population Test Development and Authorization of MALDI-ToF Mass Spectrometry-Based Screening Tests for Viral Infections. **Viruses**, v. 14, n. 9, p. 1958, 3 set. 2022.
- INFUSINO, F. et al. Diet Supplementation, Probiotics, and Nutraceuticals in SARS-CoV-2 Infection: A Scoping Review. **Nutrients**, v. 12, n. 6, p. 1718, 8 jun. 2020.
- INVERNIZZI, R.; LLOYD, C. M.; MOLYNEAUX, P. L. Respiratory microbiome and epithelial interactions shape immunity in the lungs. **Immunology**, v. 160, n. 2, p. 171–182, jun. 2020.
- IWAI, S. et al. Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. **PLOS ONE**, v. 11, n. 11, p. e0166104, 7 nov. 2016.
- JACCARD, P. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. **New Phytologist**, v. 11, n. 2, p. 37–50, fev. 1912.
- JUN, S.-R. et al. PanFP: pangenome-based functional profiles for microbial communities. **BMC Research Notes**, v. 8, n. 1, p. 479, dez. 2015.
- KANEHISA, M. et al. KEGG for integration and interpretation of large-scale molecular data sets. **Nucleic Acids Research**, v. 40, n. D1, p. D109–D114, 1 jan. 2012.
- KARSTENS, L. et al. Community profiling of the urinary microbiota: considerations for low-biomass samples. **Nature Reviews Urology**, v. 15, n. 12, p. 735–749, dez. 2018.
- KIM, B.-R. et al. Deciphering Diversity Indices for a Better Understanding of Microbial Communities. **Journal of Microbiology and Biotechnology**, v. 27, n. 12, p. 2089–2093, 28 dez. 2017.
- KIM, J. Y. et al. Diagnostic usefulness of subgenomic RNA detection of viable SARS-CoV-2 in patients with COVID-19. **Clinical Microbiology and Infection**, v. 28, n. 1, p. 101–106, jan. 2022.
- KIM, M. et al. Gut Microbial Metabolites Fuel Host Antibody Responses. **Cell Host & Microbe**, v. 20, n. 2, p. 202–214, ago. 2016.
- KISSELER, S. M. et al. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. **Science**, v. 368, p. 860–868, 2020.

- KOREN, O. et al. A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. **PLoS Computational Biology**, v. 9, n. 1, p. e1002863, 10 jan. 2013.
- KOWALCHUK, G. A. (ED.). **Molecular microbial ecology manual**. 2. ed ed. Dordrecht: Kluwer Acad. Publ, 2004.
- KOWALCZUK, S. et al. A protein complex in the brush-border membrane explains a Hartnup disorder allele. **The FASEB Journal**, v. 22, n. 8, p. 2880–2887, ago. 2008.
- KUCIRKA, L. M. et al. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. **Annals of Internal Medicine**, v. 173, n. 4, p. 262–267, 18 ago. 2020.
- KULLBACK, S.; LEIBLER, R. A. On Information and Sufficiency. **The Annals of Mathematical Statistics**, v. 22, n. 1, p. 79–86, mar. 1951.
- KURTZ, Z. D. et al. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. **PLOS Computational Biology**, v. 11, n. 5, p. e1004226, 7 maio 2015.
- LAMB, G. et al. Real-world evaluation of COVID-19 lateral flow device (LFD) mass-testing in healthcare workers at a London hospital; a prospective cohort analysis. **Journal of Infection**, v. 83, n. 4, p. 452–457, out. 2021.
- LANGILLE, M. G. I. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. **Nature Biotechnology**, v. 31, n. 9, p. 814–821, set. 2013.
- LEBER, W. et al. Comparing the diagnostic accuracy of point-of-care lateral flow antigen testing for SARS-CoV-2 with RT-PCR in primary care (REAP-2). **eClinicalMedicine**, v. 38, p. 101011, ago. 2021.
- LEEK, J. T.; STOREY, J. D. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. **PLOS Genetics**, v. 3, n. 9, p. 1724–1735, 2007.
- LEGOUX, F.; SALOU, M.; LANTZ, O. MAIT Cell Development and Functions: the Microbial Connection. **Immunity**, v. 53, n. 4, p. 710–723, out. 2020.
- LETKO, M.; MARZI, A.; MUNSTER, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. **Nature Microbiology**, v. 5, n. 4, p. 562–569, abr. 2020.
- LIN, H.; PEDDADA, S. D. Analysis of compositions of microbiomes with bias correction. **Nature Communications**, v. 11, n. 1, p. 3514, 14 jul. 2020a.
- LIN, H.; PEDDADA, S. D. Analysis of microbial compositions: a review of normalization and differential abundance analysis. **npj Biofilms and Microbiomes**, v. 6, n. 1, p. 60, 2 dez. 2020b.
- LLORÉNS-RICO, V. et al. Clinical practices underlie COVID-19 patient respiratory microbiome composition and its interactions with the host. **Nature Communications**, v. 12, n. 1, p. 6243, dez. 2021.

- LOUCA, S.; DOEBELI, M. Efficient comparative phylogenetics on large trees. **Bioinformatics**, v. 34, n. 6, p. 1053–1055, 15 mar. 2018.
- LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 550, dez. 2014.
- LOZUPONE, C. et al. UniFrac: an effective distance metric for microbial community comparison. **The ISME Journal**, v. 5, n. 2, p. 169–172, fev. 2011.
- MACINTYRE, C. R. Case isolation, contact tracing, and physical distancing are pillars of COVID-19 pandemic control, not optional choices. **The Lancet Infectious Diseases**, v. 20, n. 10, p. 1105–1106, out. 2020.
- MAHMUD, I.; GARRETT, T. J. Mass Spectrometry Techniques in Emerging Pathogens Studies: COVID-19 Perspectives. **Journal of the American Society for Mass Spectrometry**, v. 31, n. 10, p. 2013–2024, 7 out. 2020.
- MAJCHRZYKIEWICZ-KOEHORST, J. A. et al. Rapid and generic identification of influenza A and other respiratory viruses with mass spectrometry. **Journal of Virological Methods**, v. 213, p. 75–83, mar. 2015.
- MAK, J. W. Y.; CHAN, F. K. L.; NG, S. C. Probiotics and COVID-19: one size does not fit all. **The Lancet Gastroenterology & Hepatology**, v. 5, n. 7, p. 644–645, jul. 2020.
- MAN, W. H.; DE STEENHUIJSEN PITERS, W. A. A.; BOGAERT, D. The microbiota of the respiratory tract: gatekeeper to respiratory health. **Nature Reviews Microbiology**, v. 15, n. 5, p. 259–270, maio 2017.
- MARDIAN, Y. et al. Review of Current COVID-19 Diagnostics and Opportunities for Further Development. **Frontiers in Medicine**, v. 8, p. 615099, 7 maio 2021.
- MARKOWITZ, V. M. et al. IMG: the integrated microbial genomes database and comparative analysis system. **Nucleic Acids Research**, v. 40, n. D1, p. D115–D122, 1 jan. 2012.
- MARTINEZ, R. M. Clinical Samples for SARS-CoV-2 Detection: Review of the Early Literature. **Clinical Microbiology Newsletter**, v. 42, n. 15, p. 121–127, ago. 2020.
- MASLOWSKI, K. M. et al. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. **Nature**, v. 461, n. 7268, p. 1282–1286, out. 2009.
- MCMURDIE, P. J.; HOLMES, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. **PLoS ONE**, v. 8, n. 4, p. e61217, 22 abr. 2013.
- MCMURDIE, P. J.; HOLMES, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. **PLoS Computational Biology**, v. 10, n. 4, p. e1003531, 3 abr. 2014.

- MEIEROVICS, A.; YANKELEVICH, W.-J. C.; COWLEY, S. C. MAIT cells are critical for optimal mucosal immune responses during in vivo pulmonary bacterial infection. **Proceedings of the National Academy of Sciences**, v. 110, n. 33, 13 ago. 2013.
- MINCHIN, P. R. An evaluation of the relative robustness of techniques for ecological ordination. **Vegetatio**, v. 69, p. 89–107, 1987.
- MINISTÉRIO DA SAÚDE, M. DA S. **Painel CORONAVIRUS**. Disponível em: <<https://covid.saude.gov.br/>>. Acesso em: 17 jul. 2020.
- MORTON, J. T. et al. Correction for Morton et al., “Uncovering the Horseshoe Effect in Microbial Analyses”. **mSystems**, v. 3, n. 1, p. e00006-18, 27 fev. 2018.
- OBERMEIER, M. et al. Improved molecular laboratory productivity by consolidation of testing on the new random-access analyzer Alinity m. **Journal of Laboratory Medicine**, v. 44, n. 6, p. 319–328, 16 dez. 2020.
- OGUNRINOLA, G. A. et al. The Human Microbiome and Its Impacts on Health. **International Journal of Microbiology**, v. 2020, p. 1–7, 12 jun. 2020.
- OLBRICH, M.; KÜNSTNER, A.; BUSCH, H. MBECS: Microbiome Batch Effects Correction Suite. **BMC Bioinformatics**, v. 24, n. 1, p. 182, 3 maio 2023.
- ORSBURN, B. C. et al. ***In silico* approach toward the identification of unique peptides from viral protein infection: Application to COVID-19**. [s.l.] Biochemistry, 10 mar. 2020. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/2020.03.08.980383>>. Acesso em: 25 ago. 2020.
- OTANI, T.; FURUSE, M. Tight Junction Structure and Function Revisited. **Trends in Cell Biology**, v. 30, n. 12, p. 1014, dez. 2020.
- PARROT, T. et al. MAIT cell activation and dynamics associated with COVID-19 disease severity. **Science Immunology**, v. 5, n. 51, p. eabe1670, 18 set. 2020.
- PAULSON, J. N. et al. Differential abundance analysis for microbial marker-gene surveys. **Nature Methods**, v. 10, n. 12, p. 1200–1202, dez. 2013.
- PAVOINE, S.; DUFOUR, A.-B.; CHESSEL, D. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. **Journal of Theoretical Biology**, v. 228, n. 4, p. 523–537, jun. 2004.
- PAWLOWSKY-GLAHN, V.; EGOZCUE, J. J.; TOLOSANA-DELGADO, R. **Modeling and Analysis of Compositional Data**. London: JohnWiley & Sons, 2015.
- PEARSON, K. VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia. **Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character**, v. 187, p. 253–318, 31 dez. 1896.
- PENNINGER, J. M.; GRANT, M. B.; SUNG, J. J. Y. The Role of Angiotensin Converting Enzyme 2 in Modulating Gut Microbiota, Intestinal Inflammation, and Coronavirus Infection. **Gastroenterology**, v. 160, n. 1, p. 39–46, jan. 2021.

PETO, J. Covid-19 mass testing facilities could end the epidemic rapidly. **BMJ**, p. m1163, 22 mar. 2020.

PIAO, J. et al. Simultaneous Detection and Identification of Enteric Viruses by PCR-Mass Assay. **PLoS ONE**, v. 7, n. 8, p. e42251, 1 ago. 2012.

PIVA, S. et al. Clinical presentation and initial management critically ill patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in Brescia, Italy. **Journal of Critical Care**, v. 58, p. 29–33, ago. 2020.

PREMRAJ, A. et al. **Nucleic Acid and Immunological Diagnostics for SARS-CoV-2: Processes, Platforms and Pitfalls**. [s.l.] LIFE SCIENCES, 23 set. 2020.

Disponível em: <<https://www.preprints.org/manuscript/202009.0526/v1>>. Acesso em: 5 jan. 2023.

QUAH, P.; LI, A.; PHUA, J. Mortality rates of patients with COVID-19 in the intensive care unit: a systematic review of the emerging literature. **Critical Care**, v. 24, n. 1, p. 285, dez. 2020.

REIJNS, M. A. M. et al. A sensitive and affordable multiplex RT-qPCR assay for SARS-CoV-2 detection. **PLOS Biology**, v. 18, n. 12, p. e3001030, 15 dez. 2020.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139–140, 1 jan. 2010.

ROCCA, M. F. et al. **A Combined approach of MALDI-TOF Mass Spectrometry and multivariate analysis as a potential tool for the detection of SARS-CoV-2 virus in nasopharyngeal swabs**. [s.l.] Microbiology, 7 maio 2020. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/2020.05.07.082925>>. Acesso em: 25 ago. 2020.

ROOKS, M. G.; GARRETT, W. S. Gut microbiota, metabolites and host immunity. **Nature Reviews Immunology**, v. 16, n. 6, p. 341–352, jun. 2016.

ROSSMAN, H. et al. COVID-19 dynamics after a national immunization program in Israel. **Nature Medicine**, v. 27, n. 6, p. 1055–1061, jun. 2021.

ROUND, J. L.; MAZMANIAN, S. K. Inducible Foxp3⁺ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. **Proceedings of the National Academy of Sciences**, v. 107, n. 27, p. 12204–12209, 6 jul. 2010.

SANTANA-MEDEROS, D. et al. A COVID-19 vaccine candidate composed of the SARS-CoV-2 RBD dimer and *Neisseria meningitidis* outer membrane vesicles. **RSC Chemical Biology**, v. 3, n. 2, p. 242–249, 2022.

SCHAUPP, L. et al. Microbiota-Induced Type I Interferons Instruct a Poised Basal State of Dendritic Cells. **Cell**, v. 181, n. 5, p. 1080–1096.e19, maio 2020.

SCHLOSS, P. D. et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial

Communities. **Applied and Environmental Microbiology**, v. 75, n. 23, p. 7537–7541, 1 dez. 2009.

SCOTT, N. A. et al. Antibiotics induce sustained dysregulation of intestinal T cell immunity by perturbing macrophage homeostasis. **Science Translational Medicine**, v. 10, n. 464, p. eaao4755, 24 out. 2018.

SENDER, R.; FUCHS, S.; MILO, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. **PLOS Biology**, v. 14, n. 8, p. e1002533, 19 ago. 2016.

SHANG, J. et al. Structural basis of receptor recognition by SARS-CoV-2. **Nature**, v. 581, n. 7807, p. 221–224, 14 maio 2020.

SHI, Y. et al. Performance determinants of unsupervised clustering methods for microbiome data. **Microbiome**, v. 10, n. 1, p. 25, dez. 2022.

SIVANESAN, I. et al. Consolidating the potency of matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) in viral diagnosis: Extrapolating its applicability for COVID diagnosis? **TrAC Trends in Analytical Chemistry**, v. 150, p. 116569, maio 2022.

SJÖHOLM, M. I. L.; DILLNER, J.; CARLSON, J. Multiplex Detection of Human Herpesviruses from Archival Specimens by Using Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. **Journal of Clinical Microbiology**, v. 46, n. 2, p. 540–545, fev. 2008.

SMITH, P. M. et al. The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic T_{reg} Cell Homeostasis. **Science**, v. 341, n. 6145, p. 569–573, 2 ago. 2013.

SORBARA, M. T.; PAMER, E. G. Microbiome-based therapeutics. **Nature Reviews Microbiology**, v. 20, n. 6, p. 365–380, jun. 2022.

STEED, A. L. et al. The microbial metabolite desaminotyrosine protects from influenza through type I interferon. **Science**, v. 357, n. 6350, p. 498–502, 4 ago. 2017.

STEFAN, K. L. et al. Commensal Microbiota Modulation of Natural Resistance to Virus Infection. **Cell**, v. 183, n. 5, p. 1312–1324.e10, nov. 2020.

SWARTE, J. C. et al. Health-related quality of life is linked to the gut microbiome in kidney transplant recipients. **Nature Communications**, v. 14, n. 1, p. 7968, 2 dez. 2023.

TANG, H. et al. Randomised, double-blind, placebo-controlled trial of Probiotics To Eliminate COVID-19 Transmission in Exposed Household Contacts (PROTECT-EHC): a clinical trial protocol. **Open access**, p. 8, 2021.

THE HUMAN MICROBIOME PROJECT CONSORTIUM. Structure, function and diversity of the healthy human microbiome. **Nature**, v. 486, n. 7402, p. 207–214, jun. 2012.

TIAN, Y. et al. Gut Microbiota May Not Be Fully Restored in Recovered COVID-19 Patients After 3-Month Recovery. **Frontiers in Nutrition**, v. 8, p. 638825, 13 maio 2021.

TISONCIK, J. R. et al. Into the Eye of the Cytokine Storm. **Microbiology and Molecular Biology Reviews**, v. 76, n. 1, p. 16–32, mar. 2012.

TRAN, N. K. et al. Novel application of automated machine learning with MALDI-TOF-MS for rapid high-throughput screening of COVID-19: a proof of concept. **Scientific Reports**, v. 11, n. 1, p. 8219, dez. 2021.

UDUGAMA, B. et al. Diagnosing COVID-19: The Disease and Tools for Detection. **ACS Nano**, v. 14, n. 4, p. 3822–3835, 28 abr. 2020.

VENZON, M.; CADWELL, K. COVID-19 and the Forgotten Organ: Prolonged Changes to the Metabolic Output of the Gut Microbiome. **Gastroenterology**, v. 162, n. 2, p. 394–396, fev. 2022.

VIANA, S. D.; NUNES, S.; REIS, F. ACE2 imbalance as a key player for the poor outcomes in COVID-19 patients with age-related comorbidities – Role of gut microbiota dysbiosis. **Ageing Research Reviews**, v. 62, p. 101123, set. 2020.

VOLPATO, F. et al. Pooling of samples to optimize SARS-CoV-2 diagnosis by RT-qPCR: comparative analysis of two protocols. **European Journal of Clinical Microbiology & Infectious Diseases**, v. 40, n. 4, p. 889–892, abr. 2021.

WANG, B. et al. Alterations in microbiota of patients with COVID-19: potential mechanisms and therapeutic interventions. **Signal Transduction and Targeted Therapy**, v. 7, n. 1, p. 143, dez. 2022.

WANG, Q. et al. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. **Applied and Environmental Microbiology**, v. 73, n. 16, p. 5261–5267, 15 ago. 2007.

WANG, Q. et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. **Cell**, v. 181, n. 4, p. 894–904.e9, maio 2020.

WEI, W. E. et al. Presymptomatic Transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. **MMWR. Morbidity and Mortality Weekly Report**, v. 69, n. 14, p. 411–415, 10 abr. 2020.

WEIBEL, E. Morphometry of the human lung: the state of the art after two decades. **Bull Eur Physiopathol Respir.**, 5. v. 15, p. 999–1013, out. 1979.

WEISS, S. et al. Normalization and microbial differential abundance strategies depend upon data characteristics. **Microbiome**, v. 5, n. 1, p. 27, dez. 2017.

WEISSLEDER, R. et al. COVID-19 diagnostics in context. **Science Translational Medicine**, v. 12, n. 546, p. eabc1931, 3 jun. 2020.

WEMHEUER, F. et al. **Tax4Fun2: a R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S**

rRNA gene marker gene sequences. [s.l.] *Bioinformatics*, 9 dez. 2018. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/490037>>. Acesso em: 23 dez. 2022.

WESTCOTT, S. L.; SCHLOSS, P. D. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ*, v. 3, p. e1487, 8 dez. 2015.

WHO. **WHO COVID-19 Dashboard.** Geneva: Geneva: World Health Organization, 2020a. Disponível em: <<https://covid19.who.int/>>. Acesso em: 19 nov. 2022.

WHO, W. H. O. (ED.). **Clinical Management of COVID-19.** , 27 maio 2020b. Disponível em: <<https://www.who.int/publications/i/item/clinical-management-of-covid-19>>. Acesso em: 26 ago. 2020

WONG, J. J. M. et al. Insights into the immuno-pathogenesis of acute respiratory distress syndrome. *Annals of Translational Medicine*, v. 7, n. 19, p. 504–504, out. 2019.

WU, A. et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host & Microbe*, v. 27, n. 3, p. 325–328, mar. 2020.

WYPYCH, T. P.; WICKRAMASINGHE, L. C.; MARSLAND, B. J. The influence of the microbiome on respiratory health. *Nature Immunology*, v. 20, n. 10, p. 1279–1290, out. 2019.

XU, M. et al. c-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature*, v. 554, n. 7692, p. 373–377, fev. 2018.

YAN, L. et al. Rapid Detection of COVID-19 Using MALDI-TOF-Based Serum Peptidome Profiling. *Analytical Chemistry*, v. 93, n. 11, p. 4782–4787, 23 mar. 2021.

YAN, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, v. 367, n. 6485, p. 1444–1448, 27 mar. 2020.

YANG, J.-R. et al. Newly Emerging Mutations in the Matrix Genes of the Human Influenza A(H1N1)pdm09 and A(H3N2) Viruses Reduce the Detection Sensitivity of Real-Time Reverse Transcription-PCR. *Journal of Clinical Microbiology*, v. 52, n. 1, p. 76–82, jan. 2014.

YEOH, Y. K. et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*, v. 70, n. 4, p. 698–706, abr. 2021.

YI, X. et al. A New PCR-Based Mass Spectrometry System for High-Risk HPV, Part I. *American Journal of Clinical Pathology*, v. 136, n. 6, p. 913–919, 1 dez. 2011.

YOKOTA, A. et al. GM-CSF and IL-4 synergistically trigger dendritic cells to acquire retinoic acid-producing capacity. *International Immunology*, v. 21, n. 4, p. 361–377, abr. 2009.

ZENG, R. et al. Generation and transcriptional programming of intestinal dendritic cells: essential role of retinoic acid. **Mucosal Immunology**, v. 9, n. 1, p. 183–193, jan. 2016.

ZHENG, D.; LIWINSKI, T.; ELINAV, E. Interaction between microbiota and immunity in health and disease. **Cell Research**, v. 30, n. 6, p. 492–506, jun. 2020.

ZHOU, H. et al. LinDA: linear models for differential abundance analysis of microbiome compositional data. **Genome Biology**, v. 23, n. 1, p. 95, dez. 2022.

ZHOU, L. et al. TGF- β -induced Foxp3 inhibits TH17 cell differentiation by antagonizing ROR γ t function. **Nature**, v. 453, n. 7192, p. 236–240, maio 2008.

ZHU, T. et al. The impact of infection with COVID-19 on the respiratory microbiome: A narrative review. **Virulence**, v. 13, n. 1, p. 1076–1087, 31 dez. 2022.

HOSPITAL DE CLÍNICAS DE
PORTO ALEGRE DA
UNIVERSIDADE FEDERAL DO
RIO GRANDE DO SUL - HCPA
UFRGS



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Abordagem Proteômica e Metagenômica para Caracterização e Identificação de Biomarcadores da COVID-19

Pesquisador: Andreza Francisco Martins

Área Temática:

Versão: 1

CAAE: 38972220.3.0000.5327

Instituição Proponente: Hospital de Clínicas de Porto Alegre

Patrocinador Principal: Hospital de Clínicas de Porto Alegre

DADOS DO PARECER

Número do Parecer: 4.355.906

Apresentação do Projeto:

O impacto da nova pandemia causada pelo vírus SARS-CoV-2, conhecida por Coronavirus Disease 2019 (COVID-19), vem acometendo todas as nações e em todos os níveis da sociedade. O acesso a testes de diagnóstico rápidos e custo-efetivos é certamente crucial para controlar a propagação do vírus e compreender os mecanismos moleculares envolvidos na evolução da COVID-19 e identificar biomarcadores associados a cada estágio da doença é determinante para combater a ameaça que esta pandemia representa para nossa sociedade. Esse conhecimento sobre a patogênese de COVID-19 ainda está incompleto, mas essa barreira pode ser superada com tecnologias emergentes que venham a caracterizar a doença ao nível molecular, como as ciências "ômicas", que apresentam grande potencial para aplicação no diagnóstico e monitoramento da COVID-19. Por isso, este estudo tem como objetivo identificar marcadores da COVID-19 e do SARS-CoV-2 com base em abordagens de proteômica e metagenômica visando o diagnóstico rápido e diferencial. Serão incluídas neste estudo amostras de naso e orofaringe provenientes do Biobanco do Hospital de Clínicas de Porto Alegre (HCPA) e do Laboratório Exame. As amostras serão selecionadas de acordo com o resultado positivo/negativo previamente obtido no RT-PCR para SARS-CoV-2. Também utilizaremos um grupo de amostras provenientes de pacientes com suspeita de reinfecção. Serão utilizadas duas abordagens proteômicas em MALDI-TOF MS para o

Endereço: Rua Ramiro Barcelos 2.350 sala 2229

Bairro: Santa Cecília

CEP: 90.035-903

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3359-7640

Fax: (51)3359-7640

E-mail: cep@hcpa.edu.br

HOSPITAL DE CLÍNICAS DE
PORTO ALEGRE DA
UNIVERSIDADE FEDERAL DO
RIO GRANDE DO SUL - HCPA
UFRGS



Continuação do Parecer: 4.355.906

desenvolvimento dos métodos de diagnóstico diretamente da amostra clínica, sendo uma baseada na detecção de marcadores virais específicos, tais como proteína S e seus derivados e a outra baseada no perfil proteico característico da COVID-19. Os espectros serão adquiridos, pré-processados e analisados utilizando o pacote R MaldiQuant, para posterior desenvolvimento dos algoritmos de detecção. Para as análises de microbioma, será realizada extração do DNA das amostras e, após, amplificação da região V3V4 do gene 16S rRNA com os primers 336F-806R. O sequenciamento do microbioma será realizado na plataforma Illumina MiSeq. Um pipeline customizado em Mothur será usado para análise das sequências. As análises subsequentes serão realizadas em R (usando pacotes vegan, phyloseq e ggplot2) a fim de se caracterizar as comunidades microbianas em termos de diversidade e composição, predição metabólica, análise de coocorrência, e correlação com parâmetros clínicos e moleculares. Os resultados obtidos permitirão a ampla utilização do método em MALDI-TOF na prática diagnóstica, com perspectiva de redução de custos na ordem de 80%, além de ampliar os conhecimentos sobre a patogênese da doença através da caracterização de biomarcadores microbianos.

Objetivo da Pesquisa:

Objetivo Geral

Identificar marcadores microbiológicos e proteicos da COVID-19 e do SARS-CoV-2 com base em abordagens de proteômica e metagenômica visando o diagnóstico rápido e diferencial bem como uma compreensão mais ampla da patogênese da doença.

Objetivos Específicos

1. Desenvolver um método diagnóstico em MALDI-TOF MS baseado em biomarcadores específicos de SARS-CoV-2;
2. Desenvolver um algoritmo em MALDI-TOF MS para o diagnóstico diferencial da COVID-19 através da identificação do perfil proteico característico da doença;
3. Caracterizar a diversidade, estrutura e composição do microbioma de pacientes com COVID-19 moderada, severa e suspeitos de reinfecção;
4. Identificar gêneros bacterianos (abundância diferencial) característicos de pacientes com COVID-19 moderada, severa e suspeitos de reinfecção;
5. Caracterizar as rotas metabólicas bacterianas nas amostras de pacientes com COVID-19 (moderada, severa e suspeitos de reinfecção) e identificar rotas metabólicas características desses

Endereço: Rua Ramiro Barcelos 2.350 sala 2229

Bairro: Santa Cecília

CEP: 90.035-903

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3359-7640

Fax: (51)3359-7640

E-mail: cep@hcpa.edu.br

HOSPITAL DE CLÍNICAS DE
PORTO ALEGRE DA
UNIVERSIDADE FEDERAL DO
RIO GRANDE DO SUL - HCPA
UFRGS



Continuação do Parecer: 4.355.906

estágios da doença;

6. Caracterizar as redes de interação microbiana (redes de coocorrência) de amostras de pacientes com COVID-19 (moderada, severa e suspeitos de reinfecção) e identificar as espécies-chave (biomarcadores ecológicos) características desses estágios da doença;

7. Correlacionar os resultados obtidos do microbioma e proteoma com os parâmetros clínicos e moleculares associados à COVID-19 moderada, severa e na suspeita de reinfecção.

Avaliação dos Riscos e Benefícios:

Riscos:

O trabalho será desenvolvido no Laboratório de pesquisa em Resistência Bacteriana – LABRESIS - (Centro de Pesquisa Experimental - HCPA), sendo que as amostras passarão por etapa prévia de inativação no Laboratório de Diagnóstico Molecular para COVID-19. Serão utilizados procedimentos padronizados de biossegurança em todas as etapas experimentais, incluindo o uso de luvas, avental, máscara, face shield e capela de fluxo laminar. Materiais contaminados e/ou potencialmente tóxicos serão descartados conforme as regras de biossegurança recomendadas para cada caso, garantindo a segurança dos pesquisadores envolvidos e do meio-ambiente. Por se tratar de estudo com amostras do BioBanco, não há risco de contágio referente à COVID-19. Um dos principais riscos, o qual é inerente a projetos inovadores, é a avaliação de técnicas e marcadores moleculares frente a novas condições clínicas ou doenças (como é o caso da COVID-19), as quais, por serem inovadoras em termos conceituais e operacionais, estão sujeitas a não apresentarem resultados (ou avaliações) que garantam sua recomendação para uso no diagnóstico da condição em estudo. Em nosso projeto, este risco é minimizado pois as tecnologias ômicas aqui propostas vêm apresentando resultados promissores para a identificação de alvos moleculares envolvidos no Desenvolvimento da COVID-19. Além disso, por se tratar de um estudo baseado em tecnologias sofisticadas, necessitando equipamentos de última geração, ter a disposição infraestrutura moderna se torna um desafio, o que será superado através de colaborações com laboratórios da própria instituição e colaborações com outras instituições de pesquisa já estabelecida pelo proponente, como o Laboratório de Pesquisa em Resistência Bacteriana - LABRESIS/HCPA (que possui o sequenciador Miseq e MALDI-TOF MS Microflex LT) e o Laboratório Federal de Defesa Agropecuária (que possui o sistema de cromatografia líquida acoplada à espectrometria de massas em Tandem - LC-MS/MS e MALDI-TOF Autoflex Speed).

Endereço: Rua Ramiro Barcelos 2.350 sala 2229

Bairro: Santa Cecília

CEP: 90.035-903

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3359-7640

Fax: (51)3359-7640

E-mail: cep@hcpa.edu.br

HOSPITAL DE CLÍNICAS DE
PORTO ALEGRE DA
UNIVERSIDADE FEDERAL DO
RIO GRANDE DO SUL - HCPA
UFRGS



Continuação do Parecer: 4.355.906

Outro grande desafio para o desenvolvimento deste projeto se refere à aplicação de bioinformática, uma vez que esta é uma área recente e envolve a integração de conhecimento analítico em diversas áreas do conhecimento. Esse conjunto de desafios que envolve a aplicação de bioinformática será superado pela ampla expertise da equipe interdisciplinar e interinstitucional (com sólida formação em microbiologia, biologia molecular, bioinformática, biologia de sistemas e estatística avançada, bem como atuação em diversos projetos envolvendo microbioma e proteômica), o que incentiva o intercâmbio de estudantes e pesquisadores entre as Instituições que compõem esta proposta e a formação de recursos humanos com expertise na área através da orientação integrada e capacitação de alunos de mestrado e doutorado de modo a promover a transferência de conhecimentos entre os diferentes grupos de pesquisa envolvidos.

Benefícios:

Os resultados obtidos neste projeto permitirão o avanço do conhecimento sobre os mecanismos moleculares envolvidos na patogênese da COVID-19, com especial atenção aos estágios moderado e severo (de maior impacto ao SUS) e na suspeita de reinfecção (um novo quadro clínico recentemente notificado em relação à doença), com a consequente identificação de biomarcadores associados à doença, potenciais alvos terapêuticos e, principalmente, um método de diagnóstico rápido, custo-efetivo e de ampla aplicabilidade no SUS. Apesar das análises de proteômica e microbioma serem sofisticadas, necessitando equipamentos de última geração, os resultados refinados obtidos por estas tecnologias são de direta translacionalidade ao diagnóstico e monitoramento da COVID-19 e fácil aplicabilidade na rotina laboratorial, fornecendo novos marcadores moleculares envolvidos no desenvolvimento da doença, baratos e de fácil acesso ao sistema público de saúde. Como resultados diretos, pretendemos contribuir com: 1. Identificação de biomarcadores específicos de SARS-CoV-2 em MALDI-TOF MS diretamente da amostra clínica de pacientes com COVID-19. 2. Desenvolvimento de algoritmos de machine learning com a capacidade de diagnosticar a COVID-19 no espectro de proteínas obtido diretamente da amostra clínica, através da tecnologia de MALDI-TOF MS. 3. Identificação de biomarcadores microbianos (taxonômicos, metabólicos e ecológicos) característicos de pacientes com COVID-19 moderada, severa e na suspeita de reinfecção. Como resultados indiretos do projeto, pretendemos contribuir para: 1. Ampliação da capacidade de screening dos serviços de diagnóstico, ao implementarem os métodos desenvolvidos em MALDI-TOF MS. 2. Estratégias mais eficazes para o combate da doença,

Endereço: Rua Ramiro Barcelos 2.350 sala 2229

Bairro: Santa Cecília

CEP: 90.035-903

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3359-7640

Fax: (51)3359-7640

E-mail: cep@hcpa.edu.br

HOSPITAL DE CLÍNICAS DE
PORTO ALEGRE DA
UNIVERSIDADE FEDERAL DO
RIO GRANDE DO SUL - HCPA
UFRGS



Continuação do Parecer: 4.355.906

com o manejo e tratamento dos pacientes conduzidos de forma mais precisa e confiável e com custos reduzidos.3. Desenvolvimento de vacinas: Com a caracterização dos mecanismos moleculares e marcadores envolvidos na resposta do hospedeiro às vacinas contra COVID-19.4. Desenvolvimento de medicamentos: Com a identificação de novos alvos moleculares para drogas contra a COVID-19, contribuindo, inclusive, com os esforços de reaproveitamento de medicamentos já desenvolvidos contra esses novos alvos (caracterizados em outras doenças).5. Formação de recursos humanos com expertise na área através da orientação integrada e capacitação de alunos de mestrado e doutorado de modo a promover a transferência de conhecimentos entre os diferentes pesquisadores envolvidos. Os resultados do projeto serão divulgados através de:1. Apresentação de palestras e seminários sobre o tema e projeto em instituições da área da saúde (e.g. universidades, hospitais, etc).2. Apresentação de 5 resumos em Congressos.3. Publicação de 5 artigos científicos em revista Qualis A1/A2.4. Elaboração de material acessível às instituições e profissionais de saúde.5. Elaboração de Relatório Final.

Comentários e Considerações sobre a Pesquisa:

Metas

1. Desenvolver um método diagnóstico de baixo custo, que possa ser amplamente utilizado de modo a permitir um controle mais efetivo da propagação do SARS-CoV-2.
2. Identificar biomarcadores associados à COVID-19 moderada, severa e na suspeita de reinfecção que auxiliem na escolha de intervenções precoces e mais eficazes para os pacientes acometidos pela doença e reduzam o tempo ou previnam a internação.
3. Disponibilizar os resultados de forma compreensiva e acessível (Webinar, Lives, Vídeos) à diferentes instituições e profissionais de saúde, principalmente as/aos que integram a rede pública de saúde, SUS.
4. Capacitar recursos humanos para as análises de proteômica e microbioma.
5. Publicar pelo menos 4 artigos científicos.

4. Metodologia e Estratégia de Ação

4.1 Delineamento do estudo

Este projeto desenvolverá duas abordagens: um estudo observacional analítico de caso-controle (abordagem metagenômica/microbioma) e um estudo de desenvolvimento de métodos de diagnóstico (abordagem proteômica), conforme apresentado no fluxograma da Figura 1. As

Endereço: Rua Ramiro Barcelos 2.350 sala 2229

Bairro: Santa Cecília

CEP: 90.035-903

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3359-7640

Fax: (51)3359-7640

E-mail: cep@hcpa.edu.br

HOSPITAL DE CLÍNICAS DE
PORTO ALEGRE DA
UNIVERSIDADE FEDERAL DO
RIO GRANDE DO SUL - HCPA
UFRGS



Continuação do Parecer: 4.355.906

amostras a serem usadas nesse estudo serão obtidas a partir de swab combinado de nasofaringe e orofaringe cujo resultado do RT-PCR para SARS-CoV-2 está disponível. Estas amostras serão provenientes do Biobanco do Hospital de Clínicas de Porto Alegre (Biobanco/HCPA) bem como do Laboratório Exame (LE). Até o momento, o HCPA já analisou mais de 10.000 amostras com uma taxa de positividade de 18,9%. O Laboratório Exame já analisou 100.000 amostras com uma taxa de positividade de 20%.

Considerações sobre os Termos de apresentação obrigatória:

Apresentados. Solicita dispensa de TCLE.

Conclusões ou Pendências e Lista de Inadequações:

Não apresenta pendências.

Considerações Finais a critério do CEP:

Lembramos que a presente aprovação (projeto versão 07/10/2020 e demais documentos que atendem às solicitações do CEP) refere-se apenas aos aspectos éticos e metodológicos do projeto.

Os pesquisadores devem atentar ao cumprimento dos seguintes itens:

- a) Este projeto está aprovado para inclusão e revisão de registros de 520 participantes no Centro HCPA, de acordo com as informações do projeto. Qualquer alteração deste número deverá ser comunicada ao CEP e ao Serviço de Gestão em Pesquisa para autorizações e atualizações cabíveis.
- b) O projeto está cadastrado no sistema AGHUse () Pesquisa para fins de avaliação logística e financeira e somente poderá ser iniciado após aprovação final do Grupo de Pesquisa e Pós-Graduação.
- c) Qualquer alteração nestes documentos deverá ser encaminhada para avaliação do CEP.
- d) Deverão ser adicionados relatórios semestrais e um relatório final do projeto no cadastro do mesmo, no Sistema AGHUse Pesquisa.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_P ROJETO_1602184.pdf	07/10/2020 13:37:02		Aceito

Endereço: Rua Ramiro Barcelos 2.350 sala 2229

Bairro: Santa Cecília

CEP: 90.035-903

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3359-7640

Fax: (51)3359-7640

E-mail: cep@hcpa.edu.br

HOSPITAL DE CLÍNICAS DE
PORTO ALEGRE DA
UNIVERSIDADE FEDERAL DO
RIO GRANDE DO SUL - HCPA
UFRGS



Continuação do Parecer: 4.355.906

Projeto Detalhado / Brochura Investigador	ProjetoOtaviovonAmelnLovisonPPSUS2 020_071020Submetido.docx	07/10/2020 11:50:31	OTAVIO VON AMELN LOVISON	Aceito
Outros	Figura_PPSUS_2020.pptx	05/10/2020 20:48:08	Andreza Francisco Martins	Aceito
Folha de Rosto	Folha_Rosto_PPSUS2020.pdf	05/10/2020 20:45:50	Andreza Francisco Martins	Aceito
Declaração de concordância	LabExame_termo.pdf	05/10/2020 20:11:37	Andreza Francisco Martins	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

PORTO ALEGRE, 22 de Outubro de 2020

Assinado por:
Têmis Maria Félix
(Coordenador(a))

Endereço: Rua Ramiro Barcelos 2.350 sala 2229

Bairro: Santa Cecília

CEP: 90.035-903

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3359-7640

Fax: (51)3359-7640

E-mail: cep@hcpa.edu.br