UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Yasmin Moreto Guaitolini

Fatores Genéticos de Risco e Proteção para Sequelas Neurológicas da COVID-19

Yasmin Moreto Guaitolini

Vitória, 2025

Centro de Ciências da Saúde,PPG em Biotecnologia
Av. Maruípe, 1468 – Campus Universitário
CEP 29047-185 Vitória ES
pgbiotecnologia@gmail.com
www.biotecnologia.ufes.br

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Yasmin Moreto Guaitolini

Fatores Genéticos de Risco e Proteção para Sequelas Neurológicas da COVID-19

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia do Centro de Ciências da Saúde da Universidade Federal do Espírito Santo, como requisito final para obtenção do título de Mestre em Biotecnologia.

Orientador: Prof. Dr. Iúri Drumond Louro
Coorientadora: Profª. Drª. Débora Dummer Meira

Vitória, 2025

Centro de Ciências da Saúde,PPG em Biotecnologia
Av. Maruípe, 1468 – Campus Universitário
CEP 29047-185 Vitória ES
pgbiotecnologia@gmail.com
www.biotecnologia.ufes.br

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

Moreto Guaitolini, Yasmin, 1998-
M844f Fatores genéticos de risco e proteção para sequelas neurológicas da COVID-19 / Yasmin Moreto Guaitolini. - 2025.
71 p. : il.

Orientador: Iúri Drumond Louro.
Coorientadora: Débora Dummer Meira.
Dissertação (Mestrado em Biotecnologia) - Universidade Federal do Espírito Santo, Centro de Ciências da Saúde.

1. COVID-19 (Doença). 2. COVID-19, Pandemia de, 2020 2023. 3. Nervos periféricos. 4. Neurogenética. 5. Polimorfismo (Genética). 6. Vírus - Genética. I. Drumond Louro, Iúri. II. Dummer Meira, Débora. III. Universidade Federal do Espírito Santo. Centro de Ciências da Saúde. IV. Título.

CDU: 61

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Centro de Ciências da Saúde
Programa de Pós-Graduação em Biotecnologia

Ata da 245ª sessão de Defesa de Dissertação do Programa de Pós-Graduação em Biotecnologia, do Centro de Ciências da Saúde da Universidade Federal do Espírito Santo, da discente de mestrado **Yasmin Moreto Guaitolini**, realizada às quatorze horas do dia quatro de agosto de dois mil e vinte e cinco. A sessão pública foi realizada em formato de videoconferência, no link: https://meet.google.com/axf-igqk-fys. O presidente da Banca, Prof. Dr. Iúri Drumond Louro (orientador), apresentou os demais membros da comissão examinadora constituída pelos Doutores: Débora Dummer Meira, coorientadora; Flávia de Paula, examinadora interna; Elizeu Fagundes de Carvalho, examinador externo; Adriana Madeira Álvares da Silva, membro suplente interno; Bartolomeu Acioli dos Santos, membro suplente externo. Em seguida, passou a palavra à aluna que apresentou a sua proposta de dissertação intitulada **"Fatores Genéticos de Risco e Proteção para Sequelas Neurológicas da COVID-19"**. Terminada a apresentação, o presidente retomou a palavra e a cedeu aos membros da Comissão Examinadora, um a um, para procederem à arguição. Em seguida, convidou a Comissão Examinadora a se reunir em separado para deliberação. Ao final, a Comissão Examinadora retornou e o presidente informou aos presentes que a Dissertação havia sido **APROVADA**. O Presidente, então, deu por encerrada a sessão, e lavrou a presente ata, que é assinada pelos membros da Comissão Examinadora. Vitória, 04 de agosto de 2025.

Iúri Drumond Louro
Universidade Federal do Espírito Santo – Orientador

gov.br Documento assinado digitalmente
IURI DRUMOND LOURO
Data: 04/08/2025 13:56:53-0300
Verifique em https://validar.iti.gov.br

Débora Dummer Meira
Universidade Federal do Espírito Santo – Coorientadora

gov.br Documento assinado digitalmente
DEBORA DUMMER MEIRA
Data: 04/08/2025 14:12:46-0300
Verifique em https://validar.iti.gov.br

Flavia de Paula
Universidade Federal do Espírito Santo – Examinadora interna

gov.br Documento assinado digitalmente
FLAVIA DE PAULA
Data: 04/08/2025 16:08:36-0300
Verifique em https://validar.iti.gov.br

Elizeu Fagundes de Carvalho
Universidade Estadual do Rio de Janeiro - Examinador extern

gov.br Documento assinado digitalmente
ELIZEU FAGUNDES DE CARVALHO
Data: 04/08/2025 15:26:16-0300
Verifique em https://validar.iti.gov.br

Adriana Madeira Álvares da Silva
Universidade Federal do Espírito Santo - Membro suplente interno

Bartolomeu Acioli dos Santos,
Instituto Aggeu Magalhães|Fiocruz Pernambuco - Membro suplente externo

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Yasmin Moreto Guaitolini

Fatores Genéticos de Risco e Proteção para Sequelas Neurológicas da COVID-19

Data da realização do exame: 04/08/2025

Banca examinadora:

Prof. Dr. Iúri Drumond Louro (Orientador)
Doutor em Bioquímica e Genética Molecular pela *University of Alabama at Birmingham*

Profª. Drª. Débora Dummer Meira (Coorientadora)
Doutora em Biociências pela Universidade do Estado do Rio de Janeiro (UERJ)

Profª. Drª. Flavia de Paula (Membra interna)
Doutora em Biologia Genética pela Universidade de São Paulo (USP)

Prof. Dr. Elizeu Fagundes de Carvalho (Membro externo)
Doutor em Ciências pela Universidade Federal do Rio de Janeiro (UFRJ)

Dedicatória

À Deus, por sempre me guiar de acordo com os seus propósitos;
aos meus pais, meu irmão, meus avós, tios e à minha tia, por todo amor e suporte;
aos meus orientadores, pela criteriosa orientação;
e a todos que, com gestos de carinho, auxílio técnico e palavras de encorajamento, contribuíram para que eu prosseguisse.

Com gratidão e ternura, essa conquista também é de vocês.

Agradecimentos

Agradeço primeiramente a Deus, que, em meio a incertezas, me guiou para esta trajetória.

Ao meu orientador, Prof. Dr. Iúri Drumond Louro, e à minha coorientadora, Profª. Drª. Débora Dummer Meira, instrumentos de Deus nesse percurso, expresso minha profunda gratidão pela dedicação, paciência e orientação generosa, fundamentais ao longo desta jornada.

Aos meus pais, irmão, avós e demais familiares, agradeço pelo apoio constante, pelas orações e por sempre acreditarem em mim e me instruírem com sabedoria.

Aos amigos que a universidade me concedeu, verdadeiros alicerces de apoio e inspiração, minha sincera gratidão. Sem vocês, este trabalho não teria sido possível. Em especial, agradeço ao Matheus Casotti, por dividir responsabilidades acadêmicas e tornar a caminhada mais leve e prazerosa, e ao Felipe Mion, pelo auxílio essencial nas análises de bioinformática e pela ajuda incansável, e por vezes fora dos horários convencionais, na compreensão da dimensão computacional deste estudo.

À FAPES, CAPES e CNPq, agradeço pelo fomento à ciência e pelos recursos que viabilizaram este projeto.

Aos parceiros institucionais — CIAS, HEMOES e Hospital Estadual Dr. Jayme dos Santos Neves —, agradeço pela colaboração com esta pesquisa.

À FIOCRUZ Pernambuco, agradeço especialmente ao Dr. Bartolomeu Acioli dos Santos, ao Prof. Dr. Flávio Rosendo da Silva Oliveira e ao Dr. Túlio de Lima Campos, pela generosidade em compartilhar conhecimentos fundamentais ao processamento do grande volume de dados genômicos gerados neste projeto.

Epígrafe

"*Porquanto a sabedoria entrará no teu coração, e o conhecimento será suave à tua alma.* "

— Provérbios 2:10.

Resumo

A COVID-19, causada pelo SARS-CoV-2, gerou uma crise global com impactos sanitários, sociais e econômicos. Embora a maioria dos infectados se recupere, alguns indivíduos desenvolvem sequelas persistentes conhecidas como condições pós-COVID, incluindo manifestações neurológicas que afetam o sistema nervoso periférico (SNPer). Evidências recentes sugerem que a genética do hospedeiro pode influenciar a suscetibilidade e a gravidade dessas sequelas, tornando fundamental a investigação de biomarcadores genéticos associados à sua manifestação. Este é um estudo de delineamento caso-controle que busca identificar variações genéticas associadas ao desenvolvimento de sequelas no SNPer pós-COVID-19 utilizando dados de sequenciamento completo do exoma (WES). A coorte inclui 312 indivíduos sem esquema vacinal completo antes da infecção, sendo 161 com sequelas (grupo caso) e 151 sem sequelas (grupo controle). Foram analisadas características clínicas, sociodemográficas e genéticas. Para a predição do risco genético, foi implementado um modelo de aprendizado de máquina (*machine learning* - ML), no qual diferentes classificadores foram testados. O modelo de regressão logística (LR) apresentou o melhor desempenho (AUC-ROC = 0,90, acurácia = 82% e F1-*score* = 0,83), destacando 20 SNPs mais influentes na predição do risco de sequelas neurológicas. As análises evidenciaram, predominantemente, a presença de vias relacionadas à regulação imunológica, destacando-se a expressiva participação do gene *HLA-A* (*Antigen Peptide Transporter*) nesse contexto. Também foi identificado o gene *PAQR5* (*Progestin and AdipoQ Receptor Family Member 5*), associado à sinalização de hormônios esteroides. Além disso, foram observados outros genes com função indefinida ou pouco caracterizada, a exemplo do *NPIPB15* (*Nuclear Pore Complex Interacting Protein Family Member B15*), possivelmente relacionado ao transporte nuclear. Esses achados sugerem que a resposta imunológica, a inflamação e alterações no metabolismo lipídico e hormonal podem exercer influência relevante na predisposição a sequelas neurológicas pós-COVID-19. Os resultados obtidos até o momento já fornecem evidências importantes sobre a base

genética dessas sequelas, contribuindo para a identificação de biomarcadores de suscetibilidade e potenciais alvos terapêuticos, o que possibilita avanços, principalmente no manejo clínico, das condições pós-COVID-19.

Palavras-chave: SARS-CoV-2; COVID longa; sequelas neurológicas; aprendizado de máquina; biomarcadores genéticos.

Abstract

COVID-19, caused by SARS-CoV-2, has triggered a global crisis with significant health, social, and economic impacts. Although most infected individuals recover, some develop persistent sequelae known as post-COVID conditions, including neurological manifestations affecting the peripheral nervous system (PNS). Recent evidence suggests that host genetics may influence the susceptibility and severity of these sequelae, highlighting the importance of investigating genetic biomarkers associated with their occurrence. This is a case-control study aiming to identify genetic variants linked to the development of PNS sequelae following COVID-19, using whole-exome sequencing (WES) data. The cohort comprises 312 individuals without a complete vaccination scheme prior to infection, including 161 with sequelae (case group) and 151 without sequelae (control group). Clinical, sociodemographic, and genetic characteristics were analyzed. For genetic risk prediction, a machine learning (ML) model was implemented, testing different classifiers. The logistic regression (LR) model showed the best performance (AUC-ROC = 0.90, accuracy = 82%, and F1-score = 0.83), highlighting 20 SNPs most influential in predicting the risk of neurological sequelae. Analyses predominantly revealed pathways related to immune regulation, with the *HLA-A* (Antigen Peptide Transporter) gene playing a prominent role in this context. The *PAQR5* gene (Progestin and AdipoQ Receptor Family Member 5), associated with steroid hormone signaling, was also identified. Additionally, other genes with undefined or poorly characterized functions, such as *NPIPB15* (Nuclear Pore Complex Interacting Protein Family Member B15), possibly involved in nuclear transport, were observed. These findings suggest that immune response, inflammation, and alterations in lipid and hormonal metabolism may play a relevant role in the predisposition to neurological sequelae. The results obtained thus far provide important evidence on the genetic basis of these sequelae, contributing to the identification of susceptibility biomarkers and potential therapeutic targets, which may support advances, particularly in the clinical management of post-COVID-19 conditions.

Keywords: SARS-CoV-2; long COVID; neurological sequelae; machine learning; genetic biomarkers.

Sumário

Lista de tabelas

Lista de figuras

1. Introdução

Causada pelo vírus SARS-CoV-2, a pandemia de COVID-19 iniciou-se em dezembro de 2019, em Wuhan, China. Essa doença apresentou uma rápida disseminação global, o que resultou em uma grande crise nos âmbitos sanitário, social e econômico. Em setembro de 2024, a Organização Mundial da Saúde (OMS) registrou mais de 776 milhões de casos confirmados e mais de 7 milhões de mortes atribuídas à COVID-19 (World Health Organization, 2024).

Embora a apresentação clínica da COVID-19 varie de casos assintomáticos a quadros graves, os sintomas mais comuns na fase aguda incluem, mas não se restringem a, febre, tosse, pneumonia e dificuldade respiratória aguda (Pastor *et al.*, 2023). Nesse contexto, a OMS define como "condições pós-COVID" os sintomas ou sequelas que persistem ou surgem, geralmente, três meses após a infecção inicial pelo SARS-CoV-2, com duração mínima de dois meses e sem explicação por outro diagnóstico (World Health Organization, 2023). Embora este tenha sido o termo padrão adotado neste estudo, sendo a caracterização das sequelas pós-COVID realizada conforme os critérios estabelecidos pela OMS, outras denominações, como "COVID longa", "COVID-19 pós-aguda" e "síndrome pós-COVID", também são encontradas na literatura científica (Ely; Brown; Fineberg, 2024) e frequentemente utilizadas na linguagem coloquial. Além disso, é relevante destacar que no Brasil o termo 'condições pós-COVID-19' foi adotado pelo Ministério da Saúde, conforme a Nota Técnica nº 57/2023-DGIP/SE/MS (Brasil, 2023).

Conforme ressaltado pelo Ministério da Saúde, as condições pós-COVID-19 podem apresentar oscilações ao longo do tempo, variando entre melhora, piora ou recorrência dos sintomas. Essas condições podem evoluir para complicações graves ou fatais e persistir por longos períodos após a infecção (Brasil, 2023). Sabe-se que dentre os fatores de risco para desfechos mais severos durante a fase aguda da infecção (COVID-19 grave) estão inclusos idade, sexo e comorbidades

pré-existentes, como diabetes, hipertenção e obesidade (Zsichla; Müller, 2023). Estudos mais recentes têm demonstrado que a genética do vírus e a predisposição genética do hospedeiro também estão associados a desfechos críticos pós-COVID-19 (Li *et al.*, 2020; Niemi *et al.*, 2021; Asteris *et al.*, 2022; Pastor *et al.*, 2023). Já a imunidade adaptativa do vírus é um fator muito importante na recorrência das infecções (Zsichla; Müller, 2023).

O quadro pós-COVID-19 caracteriza-se por um amplo espectro de manifestações neurológicas que acometem tanto o sistema nervoso central (SNC) quanto o sistema nervoso periférico (SNPer). Dentre as sequelas neurológicas associadas à infecção pelo SARS-CoV-2, destacam-se fadiga, cefaleia, comprometimento cognitivo e alterações no humor, as quais são predominantemente subjetivas e inespecíficas, dificultando a determinação de uma relação causal direta com a infecção. Em contraste, outras sequelas frequentemente relatadas, como déficits sensório-motores, anosmia, ageusia, mialgias e disautonomia, apresentam características mais objetivas e menos susceptíveis a influências de condições preexistentes, tornando-se, assim, mais passíveis de uma correlação causal com a COVID-19 (Carvalho *et al.*, 2024).

Os mecanismos subjacentes às manifestações neurológicas persistentes da COVID-19 ainda não estão completamente elucidados. No entanto, sabe-se que a infecção pelo SARS-CoV-2 pode desencadear uma cascata neuroinflamatória, especialmente quando o vírus acessa o sistema nervoso central (SNC). Existem duas principais hipóteses para a entrada viral no cérebro: a invasão direta, na qual o vírus atravessa a barreira hematoencefálica (BHE) e infecta diretamente o tecido neural, e a invasão indireta, em que o patógeno primeiramente infecta células do sistema imunológico, que posteriormente facilitam sua chegada ao SNC. Em ambas as situações, a ruptura da BHE é um evento necessário para permitir a presença viral no cérebro (Armocida *et al.*, 2020; Jamil Al-Obaidi; Desa, 2023; Carvalho *et al.*, 2024).

Entretanto, a resposta à infecção pelo SARS-CoV-2 não é homogênea entre os indivíduos. Enquanto alguns desenvolvem complicações neurológicas persistentes após a fase aguda da doença – possivelmente em decorrência da intensa resposta inflamatória desencadeada pelo vírus –, outros sequer apresentam sintomas durante a fase aguda, tampouco sequelas a longo prazo. Essa discrepância clínica sugere a existência de fatores individuais que modulam a progressão da doença e suas consequências neurológicas. Destaca-se, ainda, a influência da variabilidade genética do próprio SARS-CoV-2, reconhecida na literatura como um importante modulador da gravidade da infecção (Biswas; Mudi, 2020; Carabelli *et al.*, 2023).

Em seu artigo de revisão, Carabelli *et al*. (2023) apresentam uma análise abrangente dos mecanismos biológicos que impulsionam a evolução das variantes do SARS-CoV-2. O vírus gerou variantes altamente mutadas — denominadas *variants of concern* (VOCs), como Alpha, Beta, Gamma, Delta e Omicron — que surgiram de forma independente e demonstraram maior transmissibilidade em relação às variantes anteriores, principalmente devido a alterações nas propriedades funcionais intrínsecas do vírus. Como exemplo, mutações no sítio de clivagem da furina na proteína spike aumentam sua infectividade, enquanto alterações em proteínas não spike contribuem para a aptidão viral, favorecendo tanto a replicação quanto a evasão das respostas imunes. Isso resulta em modificações na antigenicidade do vírus, que passa a evadir com maior eficácia as respostas imunológicas previamente estabelecidas. Complementando o cenário evolutivo, Biswas e Mudi (2020) identificaram em um estudo observacional que as mutações *D614G* e *P323L* no SARS-CoV-2 podem estar associadas a formas mais graves da COVID-19. Isso possivelmente ocorre em decorrência do aumento da infectividade viral e de sua taxa de replicação.

Dentre as explicações existentes para a resposta diferenciada à infecção, destaca-se de forma relevante a susceptibilidade genética individual do hospedeiro, amplamente corroborada por estudos recentes (Lammi *et al.* 2025; Zhang *et al.*

2024; Ercegovac *et al.* 2022; Wang *et al.*, 2021; Cruz *et al.*, 2021; Hussain *et al.*, 2020). Esta pode exercer influência tanto sobre a vulnerabilidade do sistema nervoso central à infecção pelo SARS-CoV-2 quanto sobre a intensidade da resposta inflamatória subsequente.

Segundo Wang *et al.* (2021), o polimorfismo ε4 da apolipoproteína E (ApoE4) pode facilitar a entrada do SARS-CoV-2 nas células neuronais e intensificar os danos neurológicos associados à infecção. Essa isoforma altera significativamente a estrutura e a função da proteína ApoE, sendo associada a uma maior carga viral intracelular, disfunção da barreira hematoencefálica e aumento da apoptose neuronal. Além disso, observou-se uma resposta inflamatória exacerbada, sugerindo que indivíduos portadores do alelo ε4 podem apresentar maior susceptibilidade a manifestações neurológicas graves da COVID-19 e no contexto da COVID longa também (Wang *et al.*, 2021).

Variantes genéticas como ECA2 rs73635825 e rs143936283 demonstraram interação direta com a proteína spike do SARS-CoV-2, favorecendo a infecção neuronal (Cruz *et al.*, 2021; Hussain *et al.*, 2020), uma vez que os receptores ECA2 também são expressos em células neurais e na BHE. De acordo com Carvalho *et al.* (2024), Sideratou; Papaneophytou (2023) e Shafqat *et al.* (2023), essa interação pode contribuir para a ativação da cascata inflamatória, favorecendo o desenvolvimento de sequelas neurológicas persistentes.

Ercegovac *et al.* (2022) identificaram que a combinação dos genótipos *GSTM1*-nulo e *GPX1*Leu/Leu está significativamente associada à ocorrência de névoa cerebral pós-COVID-19, atuando também como fatores independentes de risco. Esse risco é ainda mais elevado na presença do alelo *Nrf2 A*, sugerindo um efeito cumulativo dessas variantes genéticas na predisposição à essa sequela. Já Zhang *et al.* (2024) apresentaram evidências robustas de que padrões mutacionais no DNA — ou

assinaturas genômicas — podem exercer um papel determinante na predisposição ao desenvolvimento de sequelas pós-COVID-19.

Nesse sentido, também destaca-se o estudo conduzido por Lammi *et al.* (2025), que representa um dos maiores estudos de associação genômica ampla (GWAS) sobre COVID longa realizados até o momento. Essa pesquisa identificou variantes no locus do gene *FOXP4* significativamente associadas à susceptibilidade à condição, independentemente da gravidade da infecção aguda por SARS-CoV-2. O alelo de risco rs9367106-C apresentou associação robusta com maior expressão de *FOXP4* em células alveolares tipo 2 e células imunes pulmonares, sugerindo um papel funcional na persistência dos sintomas respiratórios e sistêmicos característicos da COVID longa.

Além disso, análises de colocalização apontaram que essas variantes genéticas também se associam a outras condições pulmonares, como câncer de pulmão e hospitalização por COVID-19, indicando que mecanismos relacionados à fisiologia pulmonar e à imunorregulação local podem mediar essa predisposição genética (Lammi *et al.* 2025). Esses achados reforçam a hipótese de que fatores genéticos específicos do hospedeiro desempenham um papel relevante na etiologia da síndrome pós-COVID-19.

Nesse contexto, a aplicação de técnicas de inteligência artificial (IA) tem se mostrado fundamental em estudos relacionados à COVID-19, especialmente na investigação dos aspectos genéticos associados ao seu desenvolvimento, uma vez que se trata de uma condição multifatorial e complexa, influenciada por múltiplos loci genéticos (Pastor *et al.*, 2023). Diante desse cenário, abordagens baseadas em *machine learning* (ML) e *deep learning* (DL) têm se destacado na análise e interpretação dessa patologia (Pastor *et al.*, 2023, Fallerini *et al.*, 2022; Wang *et al.*, 2021).

Por exemplo, Fallerini *et al.* (2022) aplicaram ML na análise de dados de sequenciamento completo de exoma (*Whole Exome Sequencing* – WES) de cinco coortes distintas, identificando associações entre a gravidade da COVID-19 e genes relacionados ao sistema imunológico, como *TLR7*, *TLR3*, *TICAM1*, *TLR8*, *IRAK* e *RNASEL*. Além disso, técnicas de DL têm sido amplamente aplicadas na análise de imagens médicas, como radiografias de tórax e tomografias computadorizadas (Fang *et al.*, 2021), auxiliando a identificação de pacientes com maior risco de evolução para o quadro de COVID-19 grave. Ainda, modelos baseados em IA foram desenvolvidos para detectar anormalidades pulmonares associadas à infecção, possibilitando a avaliação da gravidade e prognóstico da doença. Outros estudos nessa mesma abordagem utilizaram dados clínicos, laboratoriais e radiológicos dos pacientes para prever o risco de evolução para formas mais graves da COVID-19 (Wang *et al.*, 2021).

A integração de grandes volumes de dados clínicos e genéticos por meio de algoritmos avançados de ML permite uma compreensão mais aprofundada dos padrões de disseminação viral, aprimora a velocidade e precisão diagnóstica, além de contribuir para o desenvolvimento de novas abordagens terapêuticas. Essas ferramentas também são fundamentais para identificar indivíduos mais suscetíveis com base em características genéticas e fisiológicas personalizadas, possibilitando estratégias de proteção mais eficazes (Alimadadi *et al.*, 2020).

As diferentes manifestações e desfechos da COVID-19 refletem a complexidade da doença e sua interação com diversos sistemas orgânicos, ressaltando a necessidade de investigações aprofundadas sobre os mecanismos subjacentes às suas complicações. A literatura científica tem se concentrado, em grande parte, na identificação de marcadores moleculares associados à ação fisiológica do vírus, com o objetivo de compreender as causas das sequelas persistentes observadas em determinados grupos de pacientes. Além disso, muitos estudos exploram a associação genética com desfechos de gravidade da COVID-19 durante a fase

aguda da infecção (Li *et al.*, 2020; Fallerini *et al.*, 2022; Pastor *et al.*, 2023; Zsichla; Müller, 2023). Essas investigações são fundamentais para avançar na identificação precoce de pacientes com maior risco de hospitalização, possibilitando intervenções mais eficazes e personalizadas.

Nessa perspectiva, estudos voltados à identificação de marcadores genéticos associados às sequelas persistentes pós-COVID-19 também são essenciais para aprimorar as estratégias terapêuticas, possibilitando um manejo mais eficaz e individualizado. Além de contribuir para a melhoria da qualidade de vida dos pacientes, essas pesquisas podem otimizar a gestão hospitalar, permitindo a estratificação de risco com base nas predisposições genéticas de cada indivíduo, um fator fundamental para a personalização dos cuidados em saúde.

Em síntese, este estudo busca aprofundar a compreensão dos fatores genéticos associados ao desenvolvimento de sequelas no SNPer decorrentes da COVID-19 em determinados indivíduos em detrimento de outros que não manifestam tais complicações (Carvalho *et al.*, 2024). Trata-se de um estudo representativo da população do Espírito Santo, que possibilita a correlação com resultados de investigações semelhantes conduzidas em outras coortes. Ainda, é uma investigação que busca identificar biomarcadores genéticos para a predição do risco de desenvolver sequelas no sistema nervoso periférico no pós-COVID-19. A pesquisa se insere e se justifica no cenário científico global, contribuindo para a elucidação de questões ainda não resolvidas sobre a COVID-19 e suas implicações.

2. Objetivos

2.1. Objetivo geral

Identificar a possível existência de fatores genéticos predisponentes ao desenvolvimento de sequelas no SNPer, a fim de contribuir para a compreensão dos

mecanismos biológicos subjacentes e para o aprimoramento das estratégias de diagnóstico e manejo clínico.

2.2. Objetivos específicos

- Realizar a triagem, seleção e caracterização sociodemográfica dos voluntários para a composição da coorte do estudo.
- Selecionar, processar e gerar dados de sequenciamento completo do exoma (WES) das amostras dos participantes, assegurando a qualidade e a integridade dos dados genômicos.
- Investigar possíveis associações entre fatores sociodemográficos e o desenvolvimento de sequelas no sistema nervoso periférico (SNPer) pós-COVID-19, por meio do teste qui-quadrado.
- Investigar possíveis associações entre variantes genéticas e o desenvolvimento de sequelas no SNPer pós-COVID-19, através da aplicação do classificador de *machine learning* (ML) e da análise de regressão logística (LR).
- Identificar fatores de risco e proteção genéticos e não genéticos associados à progressão das sequelas no SNPer pós-COVID-19.

3. Metodologia

3.1. Aprovação do estudo e centros parceiros

Este estudo integra o projeto intitulado "A Genética da COVID Longa", o qual constitui um desdobramento da pesquisa em andamento "Patogênese Molecular da COVID-19: O Envolvimento de Fatores Genéticos, Hemostáticos e Imunológicos na Evolução de Quadros Graves, Desenvolvimento de Coagulopatias e COVID Longa". O protocolo do estudo foi aprovado pelo Comitê de Ética em Pesquisa com Seres

Humanos do Centro de Ciências da Saúde da Universidade Federal do Espírito Santo (UFES), sob o número CAAE 37094020.6.0000.5060, e registrado como projeto institucional da UFES sob o número 10789/2020. A pesquisa conta com financiamento proveniente do Edital Universal FAPES nº 03/2021 e da Chamada CNPq/MCTI/CT-Saúde nº 53/2022 – Linha Temática II: Mecanismos e Fatores de Risco da COVID Longa.

A execução do mesmo foi viabilizada por meio de parcerias institucionais em âmbito estadual e nacional, envolvendo o Centro de Hematologia e Hemoterapia do Estado do Espírito Santo (Hemoes), em Vitória (ES); o Hospital Dr. Jayme Santos Neves (HJSN), localizado na Serra (ES); o Hospital Unimed Vitória (HU), em Cariacica (ES); o Instituto Aggeu Magalhães, da Fundação Oswaldo Cruz (Fiocruz), em Pernambuco (PE); e o Instituto Nacional de Cardiologia (INC), no Rio de Janeiro (RJ).

O apoio estratégico dos hospitais metropolitanos do Estado, Hospital Dr. Jayme Santos Neves (HJSN) e Hospital Unimed Vitória (Cias-HU), consistiu na disponibilização de prontuários médicos de potenciais voluntários para a composição da coorte do estudo. De forma complementar, o Centro de Hematologia e Hemoterapia do Espírito Santo (Hemoes) desempenhou um papel fundamental na formação do grupo amostral, ao viabilizar o acesso aos contatos de doadores do centro, além de fornecer infraestrutura para a aplicação de questionários e a coleta de material biológico destinado às análises genéticas.

No cenário nacional, instituições de referência desempenharam papel fundamental na estruturação e condução das análises de dados. O Instituto Aggeu Magalhães (Fiocruz) contribuiu ativamente com a categorização das variáveis, incluindo a conversão de variáveis categóricas em binárias, além de oferecer suporte técnico em todas as etapas das análises bioinformáticas. Por sua vez, o Instituto Nacional de Cardiologia (INC) foi responsável pela realização do sequenciamento completo

do exoma humano, bem como pelo apoio técnico na interpretação dos dados brutos gerados.

3.2. Desenho do estudo e delineamento amostral

Trata-se de um estudo de associação genética com delineamento caso-controle. O grupo caso foi composto por indivíduos com diagnóstico prévio de COVID-19 que desenvolveram, após a fase aguda da doença, ao menos uma das seguintes sequelas no SNPer: perda ou diminuição do olfato, perda ou diminuição do paladar, zumbido, neuropatia de fibras pequenas, câimbra, tosse crônica (superior a seis meses), disautonomia e aumento da sensibilidade a estímulos luminosos ou sonoros. O grupo controle, por sua vez, incluiu indivíduos também infectados pelo SARS-CoV-2, mas que não apresentaram sequelas relacionadas ao SNPer após a fase aguda da infecção.

Os critérios de inclusão para ambos os grupos foram: idade entre 18 e 65 anos, diagnóstico confirmado de infecção pelo SARS-CoV-2, ausência de sintomas associados à fase aguda da doença por um período mínimo de 30 dias e presença de condições pós-COVID no SNPer.

3.3. Triagem, seleção e coleta de amostras biológicas

Os voluntários foram recrutados entre novembro de 2020 e julho de 2024. No período de novembro de 2020 a dezembro de 2021, foram analisados 4.807 prontuários médicos provenientes do Hospital Estadual Dr. Jayme Santos Neves (HJSN) e do Hospital Unimed Vitória (CIAS-HU), o que resultou na seleção de 134 participantes para compor a coorte inicial.

Posteriormente, foram incorporados ao estudo 92 voluntários recrutados por meio de redes sociais e televisão, 71 doadores do Centro de Hematologia e Hemoterapia do Espírito Santo (Hemoes), além de 81 indivíduos provenientes de instituições

diversas, incluindo o Hospital Universitário Cassiano Antônio de Moraes (HUCAM), o Centro de Referência em Especialidades de Saúde (CREFES), o Hospital Unimed Vitória (CIAS), a Escola Superior de Ciências da Santa Casa de Misericórdia de Vitória (EMESCAM) e a população geral.

Adicionalmente, quatro participantes da Fundação Oswaldo Cruz (Fiocruz), no Rio de Janeiro, também foram incluídos, totalizando 378 participantes na coorte do estudo.

As estratégias de recrutamento envolveram contato por telefone, mensagens via WhatsApp e e-mails institucionais, visando otimizar o alcance e a adesão dos participantes. Todos receberam informações detalhadas sobre os objetivos, procedimentos e implicações da pesquisa, e manifestaram concordância por meio da assinatura presencial do Termo de Consentimento Livre e Esclarecido (TCLE), requisito fundamental para a participação no estudo.

Após o consentimento, foi aplicado aos participantes um questionário estruturado abordando aspectos relacionados ao estado de saúde pós-COVID-19. Os dados coletados foram armazenados e gerenciados na plataforma Research Electronic Data Capture (REDCap), acessível em https://redcap.saude.es.gov.br. O questionário foi elaborado com base em instrumentos validados pela Organização Mundial da Saúde (OMS) (2021; 2023), contemplando questões sobre manifestações sintomatológicas pós-COVID-19 e qualidade de vida dos participantes. Além disso, diretrizes do National Institutes of Health (NIH) (Institutos Nacionais de Saúde), 2023, foram adaptadas para a avaliação das sequelas no SNPer.

Por fim, foram coletados 10 mL de sangue periférico de cada participante, distribuídos em dois tubos contendo EDTA (5 mL cada), os quais foram

armazenados em temperatura controlada (2°C – 8°C) para posterior realização de análises moleculares.

3.4. Aquisição de dados genéticos

3.4.1. Extração e análise qualitativa e quantitativa do DNA

A extração do DNA genômico foi realizada utilizando o QIAamp DNA Mini Kit (Qiagen®), seguindo o protocolo recomendado pelo fabricante, com exceção da etapa de eluição, na qual o material genético foi eluído em água ultrapura.

As amostras de DNA foram quantificadas por espectrofotometria, utilizando o Nanodrop One (Thermo Scientific®), e por fluorimetria, com o Qubit 4 (Thermo Fisher®). A integridade do material genético foi avaliada por meio de eletroforese em gel de agarose a 1%, corado com 2 µL de GelRed (Biotium®) e visualizado sob luz ultravioleta (UV).

Amostras que apresentaram baixa concentração de DNA (<1,5 ng/µL na razão 280/230) e/ou evidências de fragmentação foram submetidas a um novo processo de extração para garantir a qualidade do material genético.

3.4.2. Sequenciamento de nova geração (NGS) do exoma humano

O estudo utilizou o sequenciamento do exoma humano sem restrição prévia a genes específicos, permitindo uma análise abrangente de todas as regiões codificantes do genoma. Essa abordagem possibilitou a identificação de variantes genéticas potencialmente associadas ao desenvolvimento de sequelas no SNPer em indivíduos que tiveram COVID-19, após a fase aguda da infecção.

As amostras que atenderam aos critérios de concentração e integridade adequados (n = 378) foram diluídas a 6,67 ng/µL e enviadas ao Instituto Nacional de Cardiologia (INC) para sequenciamento do DNA. No INC, foram realizadas as

etapas de amplificação e qualificação da biblioteca genômica, seguidas pelo sequenciamento do exoma utilizando a plataforma NovaSeq 6000 (Illumina®).

O sequenciamento de exoma humano é um método baseado na tecnologia de sequenciamento de nova geração (NGS), que envolve as etapas:

I. Construção da biblioteca genômica, com fragmentação do DNA em segmentos unifilamentares curtos, reparo de extremidades, ligação de adaptadores e amplificação por *polymerase chain reaction* (PCR) (reação em cadeia da polimerase);
II. Enriquecimento de alvos, realizado pelo sequenciamento de regiões específicas de interesse, por meio da tecnologia *synthesis by synthesis* (SBS) (síntese por síntese). Essa etapa consiste na hibridização dos segmentos unifilamentares com sondas marcadas com fluoróforos, complementares às regiões dos éxons, permitindo a identificação dos duplexes por um sistema computadorizado;
III. Armazenamento dos dados contendo as sequências de nucleotídeos (A, T, C e G) no formato *fastq format* (FASTQ), amplamente utilizado em bioinformática para o processamento e análise de dados do exoma (Illumina, 2024).

Após a conclusão das análises, a equipe do INC disponibilizou os arquivos FASTQ na plataforma BaseSpace Sequence Hub (Illumina, 2024), permitindo o acesso dos pesquisadores responsáveis pelo estudo.

3.4.3. Tratamento dos dados pós-sequenciamento

Os arquivos FASTQ gerados a partir do sequenciamento foram alinhados ao genoma de referência humano (*Genome Reference Consortium Human Build 38* – GRCh38) utilizando a metodologia *small variant caller: germline*. A identificação de

variantes foi realizada por meio do software DRAGEN Enrichment, seguida pela anotação das variantes empregando o software Nirvana.

Os arquivos resultantes nos formatos FASTQ, *binary alignment/map* (BAM) e *variant call format* (VCF) foram armazenados na plataforma BaseSpace Sequence Hub. Todos os softwares utilizados no processamento dos dados são fornecidos pela empresa Illumina®.

O processo de chamada de variantes identificou variantes de nucleotídeo único (*single nucleotide variants* – SNVs), bem como inserções e deleções (*indels*).

3.4.4. Pré-processamento dos dados: construção da matriz

O pré-processamento dos dados teve início com a construção de uma matriz, utilizando a linguagem de programação Python (versão 3.9.13). Essa matriz foi composta por três conjuntos de informações: (i) a identificação do paciente (ID de estudo), (ii) a presença (1) ou ausência (0) de sequelas no SNPer e (iii) as variantes genéticas. Para reduzir potenciais vieses relacionados à imunização, foram incluídos apenas indivíduos que não haviam recebido o esquema vacinal completo contra a COVID-19 (duas doses da vacina) antes da infecção.

Para garantir a confiabilidade e precisão das análises subsequentes, foram aplicados controles de qualidade no processamento dos dados genéticos:

I. Seleção de variantes com um escore de qualidade (*QUAL*) superior a 20 na escala *Phred*, garantindo que a probabilidade de erro na chamada do alelo alterado fosse inferior a 1% ($p < 0,01$);
II. Inclusão apenas de variantes que apresentaram "*pass*" no filtro de qualidade (*filter*);
III. Profundidade de leitura (*total depth*) mínima de 20, assegurando alta confiabilidade na detecção das variantes.

Essa abordagem possibilitou a avaliação da relação entre fatores genéticos e as sequelas no SNPer na manifestação dos desfechos clínicos.

3.5. Análise estatística

No modelo estatístico, as variantes genéticas foram consideradas variáveis explicativas ou preditoras (independentes), enquanto o conjunto de sequelas no SNPer foi definido como a variável desfecho (dependente). A associação entre essas variáveis foi avaliada da forma descrita a seguir, com o objetivo de identificar possíveis correlações entre os fatores genéticos e os desfechos clínicos investigados.

3.5.1. Classificador genômico para predição de sequelas no SNPer

A base de dados utilizada para a modelagem com ML inicialmente continha informações de 378 pacientes. No entanto, 66 indivíduos foram excluídos por apresentarem esquema vacinal completo (duas doses da vacina) antes da infecção, resultando em um total de 312 pacientes analisados. Desses, 161 pertenciam à classe "Positivo para sequelas no sistema nervoso periférico (SNPer-Positivo)" e 151 à classe "Negativo para sequelas no sistema nervoso periférico (SNPer-Negativo)". No total, estavam disponíveis 4052 polimorfismos de nucleotídeo único (SNPs).

Para garantir a qualidade dos dados, foram mantidas apenas as variantes com até 10% de valores ausentes, resultando em um conjunto final de 3805 variantes. A imputação de valores ausentes foi realizada utilizando o método do valor mais frequente.

Além disso, os dados genéticos extraídos do arquivo VCF foram codificados conforme o seguinte esquema:

- ➢ Homozigoto para o alelo de referência (0/0) → 0

- Heterozigoto (1/0 ou 0/1) → 1
- Homozigoto para o alelo variante (1/1) → 2

Essa abordagem permitiu a estruturação dos dados para a aplicação dos modelos de ML, garantindo a padronização das informações genotípicas para as análises subsequentes.

Neste estudo, foi desenvolvido um modelo de *machine learning* (ML) baseado em regressão logística para prever o risco associado a genótipos específicos. O treinamento do modelo foi realizado na linguagem Python (versão 3.12.5), utilizando as seguintes bibliotecas especializadas:

- *pandas*, para análise e manipulação dos dados;
- *sklearn.feature_selection*, para seleção de variáveis (variantes genéticas) utilizando a técnica *recursive feature elimination* (RFE) (Chen; Jeong, 2007);
- *sklearn.metrics*, para avaliação do desempenho dos modelos de ML;
- *sklearn.linear_model*, para implementação do classificador de LR.

Inicialmente, os dados genotípicos pré-processados foram carregados e submetidos a um processo de limpeza, no qual identificadores de pacientes foram removidos para minimizar riscos de viés e sobreajuste. O conjunto de dados foi então dividido em conjuntos de treinamento e teste, permitindo a avaliação da performance do modelo de maneira robusta e estatisticamente válida.

Posteriormente, foi implementada a técnica RFE, utilizando um classificador do tipo LR linear, com o objetivo de identificar as variantes genéticas (também denominadas características ou *features*) mais influentes no desempenho do modelo. O RFE variou de 1 a 50, permitindo avaliar o impacto da quantidade de *features* na capacidade preditiva do modelo.

Cada configuração foi testada em um conjunto de dados transformado, tanto para treinamento quanto para teste. O modelo foi treinado no conjunto de treino transformado e avaliado no conjunto de teste transformado. Para mensurar o seu desempenho, foram calculadas as seguintes métricas de avaliação (Santos, 2024; Sokolova; Lapalme, 2009):

- Sensibilidade (*sensitivity)*: Também chamada de revocação (*recall*), mede a capacidade do modelo de identificar corretamente os casos positivos, sendo crítica em diagnósticos médicos.

- Especificidade (*specificity*): Avalia a proporção de verdadeiros negativos corretamente classificados, essencial para reduzir falsos positivos, principalmente em triagens clínicas.

- Área sob a Curva ROC (AUC-ROC – *area under the receiver operating characteristic curve*): Mede a capacidade do modelo em distinguir entre classes positivas e negativas.

- Pontuação F1 (F1-*score*): É a média harmônica entre precisão e sensibilidade, equilibrando essas métricas e sendo recomendada para conjuntos de dados desbalanceados.

- Precisão (*precision*): Indica a proporção de previsões positivas corretas em relação ao total de previsões positivas, sendo útil quando o custo de falsos positivos é alto, como em casos na área da saúde.

- Acurácia (*accuracy*): Mede a proporção de previsões corretas em relação ao total de previsões feitas pelo modelo. É uma métrica global que reflete o desempenho geral do classificador.

- Matriz de Confusão (*confusion matrix*): Tabela que compara previsões do modelo com valores reais, detalhando verdadeiros positivos (VP), falsos

positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN). O ideal é que se tenha um alto número de VP e VN, com poucos FP e FN.

É fundamental destacar que nenhuma métrica deve ser analisada isoladamente, uma vez que cada uma fornece uma perspectiva limitada sobre o desempenho de um modelo de classificação e apresenta restrições específicas quando considerada individualmente. De modo geral, valores superiores a 70%–75% são frequentemente classificados como satisfatórios (Santos, 2024), porém sua adequação depende do contexto clínico ou do cenário específico em estudo.

Além do classificador LR, outros modelos foram treinados para comparação dos resultados, incluindo k-nearest neighbors (*KNeighborsClassifier*), árvore de decisão (*DecisionTreeClassifier*) e máquina de vetores de suporte (*Support Vector Machine*) todos da biblioteca scikit-learn (*sklearn*). O tratamento de valores ausentes foi realizado com o imputador simples (*SimpleImputer – sklearn*), e a serialização dos modelos foi conduzida utilizando a biblioteca *pickle*. Os resultados obtidos foram registrados e armazenados em um arquivo no formato valores separados por vírgula (*comma-separated values* – CSV). Por fim, a comparação entre os diferentes modelos foi realizada por meio de validação cruzada (*cross-validation*), garantindo maior robustez estatística na avaliação do desempenho preditivo.

3.5.2. Análise de regressão logística (LR) *stepwise*

Uma vez que o modelo LR demonstrou o melhor desempenho entre os classificadores de ML, foi realizada uma regressão logística com seleção *stepwise*, de modo a avaliar a contribuição individual e ajustada das variantes genéticas que permaneceram no modelo final, permitindo a interpretação dos efeitos em termos de risco ou proteção associados ao desfecho clínico. Nesta etapa, foram inseridos os 20 SNPs retornados pelo classificador, além dos fatores sociodemográficos espectro clínico, sexo, índice de massa corporal (IMC) e idade, que foram os que apresentaram significância em análise prévia.

4. Resultados

4.1. Caracterização da coorte

A coorte do presente estudo é composta por 312 indivíduos sem esquema vacinal completo antes da infecção, sendo 161 com sequelas (grupo caso) e 151 sem sequelas no SNPer (grupo controle). As características dos participantes foram examinadas com base nas seguintes variáveis: comorbidades preexistentes, sexo, etnia, prática de exercício pré-COVID-19, tabagismo, idade, índice de massa corporal (IMC) e espectro clínico da doença. A tabela a seguir apresenta as inferências realizadas a partir das informações obtidas do questionário aplicado aos participantes:

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Tabela 1 - Distribuição dos pacientes por fatores sociodemográficos

		Caso (n=161)		Controle (n=151)		Total (n=312)		Dados Faltantes		Valor-p (<0,05)
Comorbidades preexistentes	Sim	94	58%	81	54%	175	56%	0	0%	0.47
	Não	67	42%	70	46%	137	44%			
Sexo	Feminino	101	63%	76	50%	177	57%	0	0%	0.04*
	Masculino	60	37%	75	50%	135	43%			
Etnia	Branco	91	57%	93	62%	184	59%	1	0%	0.47
	Não-branco	69	43%	58	38%	127	41%			
Prática de exercício pré-COVID-19	Sim	88	55%	85	56%	173	55%	2	1%	0.86
	Não	72	45%	65	43%	137	44%			
Tabagismo	Sim	20	12%	13	9%	33	11%	43	14%	0.39
	Não	120	75%	116	77%	236	76%			
Idade	≤ 39 anos	47	29%	59	39%	106	34%	2	1%	0.11
	40 - 49 anos	42	26%	40	26%	82	26%			
	≥ 50 anos	71	44%	51	34%	122	39%			
IMC	Obeso	73	45%	53	35%	126	40%	0	0%	0.08
	Não obeso	88	55%	98	65%	186	60%			
Espectro clínico	Grave	73	45%	46	30%	119	38%	0	0%	0.01*
	Não grave	88	55%	105	70%	193	62%			

Legenda: Comorbidades preexistentes incluem pelo menos uma das seguintes condições: doenças pulmonares, cardiovasculares, renais, hepáticas, diabetes, imunodeficiência, neoplasias, acidente vascular cerebral (AVC), obesidade e tabagismo. A categoria *Não-branco* na variável etnia compreende indivíduos autodeclarados negros, pardos ou asiáticos. No índice de massa corporal (IMC), indivíduos com valores ≥30 foram classificados como *Obeso*, enquanto aqueles com

valores inferiores a esse limiar foram designados como *Não obeso*. A classificação *Grave* no espectro clínico inclui pelo menos uma das seguintes condições: necessidade de internação em unidade de terapia intensiva (UTI), ventilação de alto fluxo e/ou permanência hospitalar ≥10 dias. Valores com * apresentaram significância estatística. **Fonte:** Elaboração própria.

No grupo caso, composto por indivíduos com sequelas, a maioria era do sexo feminino (63%), enquanto no grupo controle a distribuição entre homens e mulheres foi mais equilibrada, com menor frequência de mulheres (50%). A proporção significativamente maior de mulheres no grupo caso em comparação ao controle (p = 0,04) sugere uma possível maior predisposição do sexo feminino a sequelas no SNPer.

No grupo caso, 45% dos pacientes apresentaram a forma grave da doença na fase aguda, enquanto 55% tiveram a forma não grave. No grupo controle, a proporção de casos graves foi menor (30%), com 70% apresentando a forma não grave. Essa diferença foi estatisticamente significativa (p = 0,01), sugerindo uma associação entre a gravidade da condição clínica e o *status* de caso ou controle. Esses dados indicam que indivíduos que desenvolveram a forma grave da COVID-19 podem ter maior predisposição a sequelas no SNPer.

Embora os fatores idade e IMC não tenham atingido significância estatística, seus valores de p estiveram mais próximos do limiar de significância em comparação aos demais fatores não significativos. Portanto, eles não foram descartados na análise de regressão logística subsequente. A regressão logística possibilita compreender, de maneira mais aprofundada, a influência de características sobre a presença ou a ausência de sequelas. Esse método estatístico gera estimativas como a razão de chances (OR), que quantifica a força da associação entre esses fatores e a ocorrência das sequelas. Além disso, os limites inferior (L95) e superior (U95) do intervalo de confiança indicam a margem de incerteza da estimativa, permitindo avaliar a precisão e a confiabilidade do valor obtido para a OR.

Por fim, as variáveis comorbidades preexistentes (p = 0,47), etnia (p = 0,47), prática de exercício pré-COVID-19 (p = 0,86) e tabagismo (p = 0,39) não foram estatisticamente significativas, indicando que essas características não influenciaram a distinção entre os grupos caso e controle.

4.2. Classificador genômico

Neste estudo, o classificador complexo baseado em dados do exoma apresentou resultados promissores. Diferentes modelos de aprendizado de máquina (ML) foram treinados, e a seleção do modelo para as análises subsequentes considerou o valor da métrica área sob a curva ROC (AUC-ROC). Conforme ilustrado na Figura 1, o modelo regressão logística (LR) obteve o melhor desempenho, com um AUC-ROC de 0,89 no conjunto de teste. Esse valor indica que o classificador possui uma boa capacidade de discriminação entre classes positivas e negativas, sendo significativamente superior ao desempenho esperado pelo acaso (AUC-ROC = 0,5).

Observa-se que o desempenho do modelo LR melhora com o aumento do número de SNPs (*features*). No entanto, esse ganho se estabiliza progressivamente após o aumento inicial, sugerindo que as variantes genéticas adicionais contribuem cada vez menos para a melhoria do modelo.

Esse comportamento é esperado em aprendizado de máquina, onde a identificação do ponto de saturação é essencial para evitar a inclusão de SNPs redundantes, que não agregam ao desempenho do modelo, e minimizar o risco de *overfitting* (Oliveira, 2015; Takahashi *et al.*, 2020). O gráfico indica que a inclusão de até cerca de 20 SNPs proporciona o melhor desempenho, motivo pelo qual esse valor foi estabelecido como critério de corte para a análise das variantes genéticas.

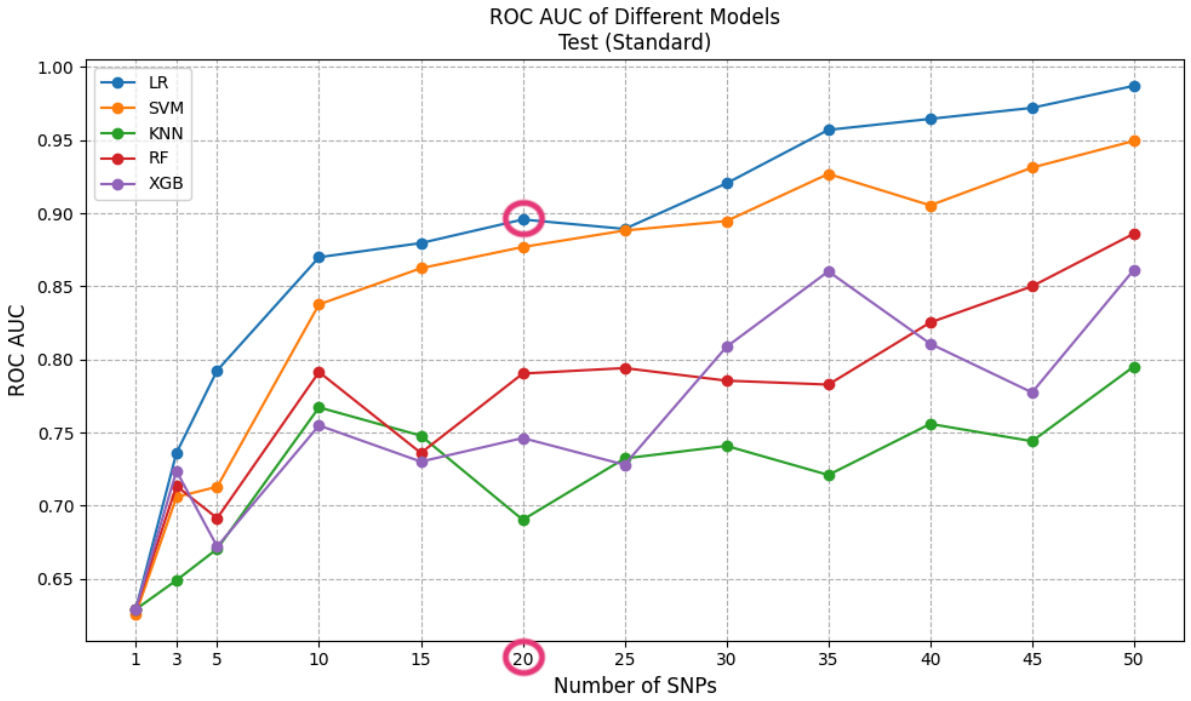

Figura 1: Performance do modelo em função do valor da métrica AUC-ROC e do número de SNPs (*features*). O modelo de regressão logística (LR) apresentou o melhor desempenho no número de 20 *features* estabelecido. O valor de AUC-ROC tende à estabilização conforme o número de SNPs aumenta. Fonte: Elaboração própria.

O modelo LR também apresentou desempenho estatisticamente robusto para todas as demais métricas analisadas, conforme a matriz de confusão e a Tabela 2 dispostas a seguir:

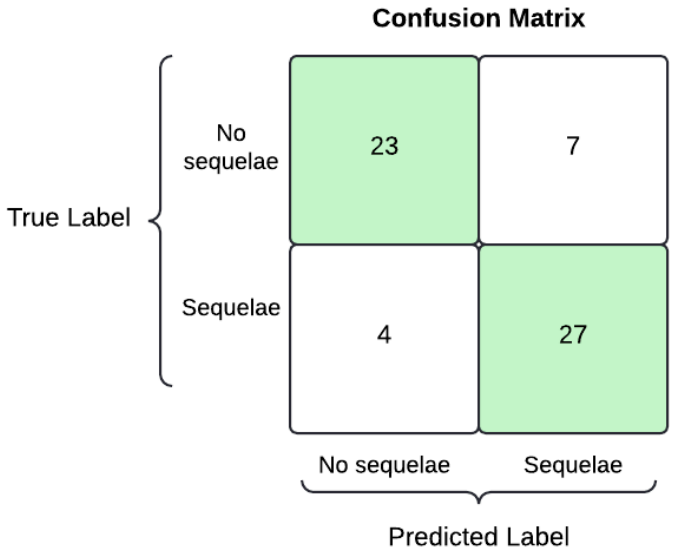

Figura 2: Matriz de confusão. Desempenho do classificador baseado no exoma humano e aprendizado de máquina para prever a presença ou ausência de sequelas da COVID-19 no SNPer, utilizando 61 amostras de teste. Os resultados referem-se ao modelo LR, que apresentou as melhores métricas, considerando as 20 *features* selecionadas. "*No sequelae*" se refere à classe negativa e "*Sequelae*" se refere à classe positiva. A primeira coluna exibe as previsões negativas (verdadeiros e falsos negativos), enquanto a segunda coluna representa as previsões positivas (falsos e verdadeiros positivos). Os quadrantes em verde indicam as previsões verdadeiras do modelo. Fonte: Elaboração própria.

Tabela 2 - Métricas alcançadas para o modelo LR

Métrica	Valor do conjunto de teste
Sensibilidade	0,87
Especificidade	0,77
AUC-ROC	0,90
Pontuação F1	0,83
Precisão	0,79
Acurácia	0,82

Fonte: Elaboração própria.

O classificador apresentou um bom desempenho em termos de sensibilidade, evidenciando sua eficácia na identificação correta das amostras positivas, isto é, com a presença de sequelas no SNPer. Quanto à especificidade, o modelo demonstrou uma taxa de acerto de 77%, refletindo sua capacidade de distinguir os pacientes sem sequelas no SNPer, reduzindo os falsos positivos. A Pontuação F1, que equilibra precisão e sensibilidade, foi de 83%, indicando um desempenho consistente. Já a acurácia do modelo foi de 82%, demonstrando sua capacidade global de classificação correta das amostras.

No entanto, classificadores baseados em LR não permitem a interpretação direta da relação entre as entradas e as previsões, nem do impacto individual de cada característica na classificação. Para elucidar o funcionamento interno do classificador, empregou-se a técnica SHAP (Lundberg; Lee, 2017), um método avançado de explicabilidade em aprendizado de máquina. Essa abordagem quantifica a contribuição de cada característica para a previsão do modelo,

proporcionando uma análise mais transparente do seu processo de tomada de decisão.

A figura abaixo apresenta uma versão adaptada de um gráfico de violino, gerado pela API SHAP (*Shapley Additive Explanations*), compatível com o *Scikit-learn*, uma das bibliotecas mais amplamente utilizadas em aprendizado de máquina:

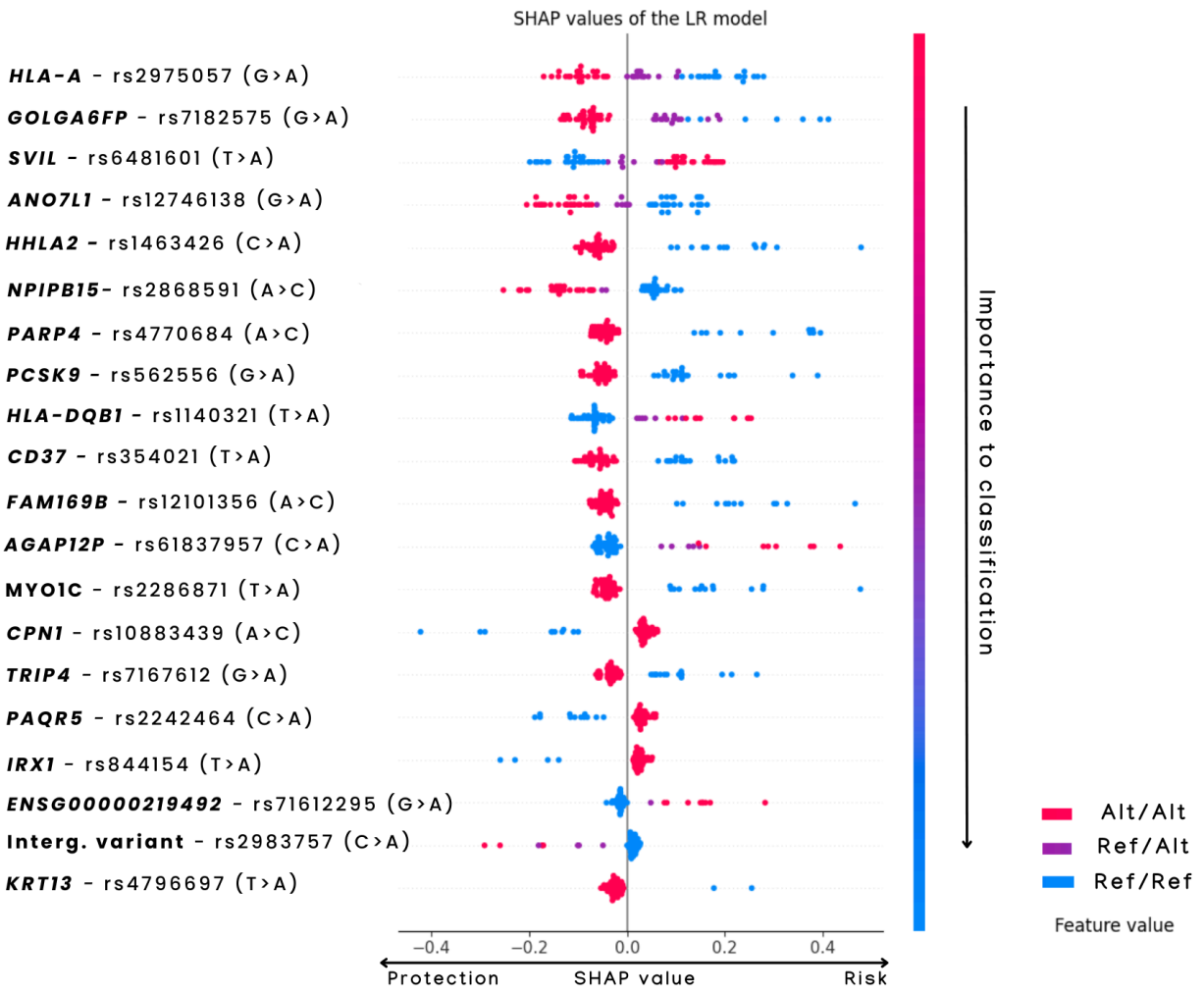

Figura 3: Impacto geral de cada genótipo dos polimorfismos genéticos identificados pelo classificador de aprendizado de máquina (ML) no prognóstico das sequelas da COVID-19, analisado por SHAP (*Shapley Additive Explanations*). As alterações alélicas e as associações entre rsIDs e genes foram determinadas a partir das posições cromossômicas brutas, utilizando a anotação funcional fornecida pela ferramenta snpXplorer. Valores SHAP positivos (>0) indicam maior risco, enquanto valores SHAP negativos (<0) indicam efeito protetor em relação ao desenvolvimento de sequelas associadas ao SNPer. Os polimorfismos genéticos estão ordenados de forma decrescente conforme sua importância relativa para o modelo preditivo. Em azul, homozigotos para o alelo de referência (Ref/Ref); em roxo heterozigotos (Ref/Alt); em rosa, homozigotos para o alelo alternativo (Alt/Alt). A figura representa a análise SHAP realizada sobre o conjunto de teste – alguns alelos podem não apresentar os três genótipos. Fonte: Elaboração própria.

A partir deste gráfico, é possível fazer algumas observações relevantes. Os 20 SNPs selecionados foram ordenados de acordo com sua importância para a classificação do modelo, com os mais influentes posicionados no topo.

Cada ponto no gráfico representa o genótipo de um indivíduo, posicionado ao longo da linha horizontal correspondente. Na parte inferior da figura, é possível visualizar o impacto de cada SNP na predição do modelo, indicando sua influência na presença ou ausência de sequelas no SNPer associadas a essas variantes genéticas.

4.3. Regressão logística (LR) *stepwise*

A regressão logística submetida ao procedimento *stepwise* constitui uma abordagem que realiza a pré-seleção das variáveis com base no Critério de Informação de Akaike (*Akaike Information Criterion – AIC*), em vez de incluir todos os preditores simultaneamente no modelo. O AIC é um método que compara diferentes modelos estatísticos considerando, ao mesmo tempo, o grau de ajuste aos dados e a complexidade do modelo, penalizando a inclusão excessiva de variáveis irrelevantes (Cavanaugh; Neath, 2019). Durante o processo *stepwise*, as variantes são adicionadas ou retiradas automaticamente conforme contribuam para reduzir o valor do AIC, o que indica um modelo com melhor capacidade explicativa. Essa estratégia permite priorizar os SNPs com maior relevância estatística e evitar interpretações baseadas em efeitos artificiais decorrentes do grande número de variáveis genéticas analisadas simultaneamente. Por esse motivo, apenas as variantes que

permaneceram após essa etapa são consideradas no modelo final, o que explica a redução do número de SNPs observados, conforme apresentado na tabela abaixo:

Tabela 3 - Resultados da análise por regressão logística com seleção *stepwise*

Característica/Variante	Gene	Alteração alélica	Tipo de mutação	Característica/Genótipo	OR	IC95%	p-valor	Interpretação
Espectro clínico	-	-	-	Grave	3.3492	1.8438 – 6.2622	<0,001	Risco
Sexo	-	-	-	Masculino	0.3733	0.2061 – 0.6610	<0,001	Proteção
rs12746138	*ANO7L1*	G>A	regulatory	Heterozigoto	0.3864	0.1554 – 0.9330	0.036	Proteção
				Homozigoto Alterado	0.3671	0.2035 – 0.6494	<0,001	
rs61837957	*AGAP12P*	C>A	non_coding_exon	Heterozigoto	4.4402	1.5686 – 14.002	0.007	Risco
rs61837957		C>A	non_coding_exon	Homozigoto Alterado	0.7530	0.2959 – 1.8968	0.500	Proteção
rs2983757	Variante intergênica	C>A	regulatory	Heterozigoto	1.1226	0.3998 – 3.1671	0.800	Risco
rs2983757		C>A	regulatory	Homozigoto Alterado	3.2080	1.1018 – 10.426	0.039	Risco
rs2868591	*NPIPB15*	A>C	missense	Heterozigoto	0.2418	0.0662 – 0.7920	0.023	Proteção
				Homozigoto Alterado	0.4993	0.2725 – 0.8993	0.022	
rs354021	*CD37*	T>A	regulatory	Heterozigoto	1.3571	0.3531 – 5.0092	0.600	Risco
				Homozigoto Alterado	0.6242	0.1733 – 2.1210	0.500	Proteção
rs2242464	*PAQR5*	C>A	regulatory	Heterozigoto	20.3350	1.7344 – 520.69	0.026	Risco
				Homozigoto Alterado	18.9550	1.6798 – 475.58	0.027	
rs7182575	*GOLGA6FP*	G>A	downstream	Heterozigoto	0.2876	0.1087 – 0.7352	0.010	Proteção
				Homozigoto Alterado	0.3621	0.1563 – 0.8108	0.015	
rs2286871	*MYO1C*	T>A	intron	Heterozigoto	3.3733	0.4612 – 24.342	0.200	Risco
rs2286871		T>A	intron	Homozigoto Alterado	1.0807	0.1596 – 7.1235	>0.9	Risco

rs2975057	*HLA-A*	G>A	regulatory	Heterozigoto	0.9592	0.4481 – 2.0503	>0.9	Proteção
rs2975057		G>A	regulatory	Homozigoto Alterado	0.4807	0.2499 – 0.9094	0.026	Proteção

Legenda: Os nomes e valores riscados referem-se aos resultados que não atingiram o limiar de significância estatística. A interpretação do espectro clínico considera a comparação do grupo “Grave” em relação ao grupo “Leve”, enquanto a variável sexo foi analisada comparando indivíduos do grupo “Masculino” em relação ao grupo “Feminino”. Os rsIDs, os genes associados, as alterações alélicas e os tipos de mutação foram obtidos por meio da anotação funcional realizada através da ferramenta snpXplorer. Para todos os SNPs, a interpretação baseia-se na comparação dos genótipos heterozigoto e homozigoto alterado em relação ao alelo selvagem correspondente. Valores com * apresentaram significância estatística. **Fonte:** Elaboração própria.

Observou-se que indivíduos do sexo masculino apresentaram menor chance no desenvolvimento de sequelas no SNPer. Em relação ao espectro clínico, casos graves mostraram maior probabilidade nesse desfecho quando comparados aos leves.

Dos 20 SNPs retornados pelo classificador, apenas 7 permaneceram significativos após a LR. São eles: rs12746138 (*ANO7L1*), que apresentou efeito protetor tanto no genótipo heterozigoto quanto no homozigoto alterado; rs61837957 (*AGAP12P*), cujo genótipo heterozigoto associou-se a aumento do risco; rs2983757 (região sem genes), com efeito de risco observado no homozigoto alterado; rs2868591 (*NPIPB15*), que demonstrou efeito protetor nos genótipos heterozigoto e homozigoto alterado; rs2242464 (*PAQR5*), com associação a risco em ambos os genótipos heterozigoto e homozigoto alterado; rs7182575 (*GOLGA6FP*), mostrando efeito protetor nos genótipos heterozigoto e homozigoto alterado; e rs2975057 (*HLA-A*), que apresentou efeito protetor no genótipo homozigoto alterado.

4.4. Associação dos SNPs significativos com traços GWAS

O gráfico da figura a seguir foi gerado pela ferramenta snpXplorer (Tesi *et al.*, 2021). Ele exibe a associação entre os polimorfismos genéticos identificados e os traços fenotípicos catalogados no *Genome-Wide Association Studies* (GWAS). O GWAS é uma abordagem que investiga, de maneira ampla, o genoma de grandes populações para identificar variantes genéticas associadas a características específicas, como doenças, parâmetros clínicos e outros fenótipos complexos (Visscher *et al.*, 2017). Esses estudos reúnem informações provenientes de diversos consórcios e bancos de dados públicos, permitindo contextualizar os SNPs detectados em relação a achados já descritos na literatura científica. A visualização apresentada contribui para compreender se as variantes associadas às sequelas neurológicas também estão envolvidas em outros processos biológicos ou condições clínicas previamente estudadas:

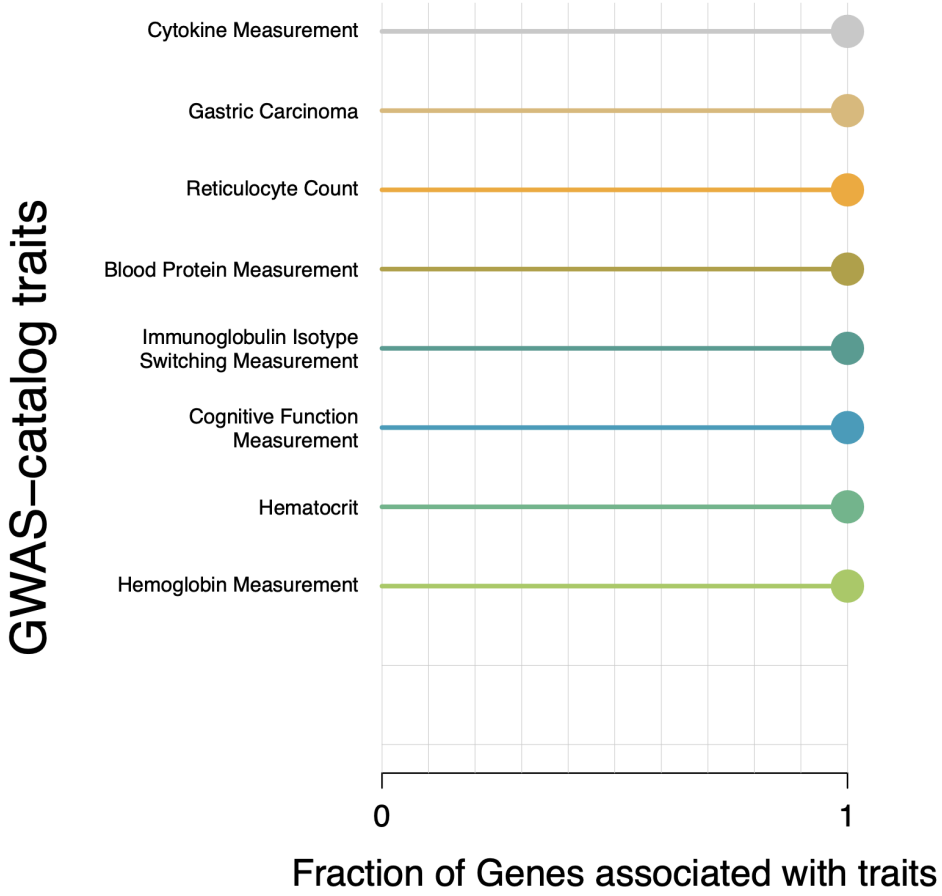

Figura 4: Traços fenotípicos catalogados no GWAS associados aos genes mapeados a partir dos SNPs significativos. O eixo vertical apresenta as diferentes características fenotípicas, e o eixo horizontal indica a proporção de genes, identificados por meio dos SNPs, que já foram relacionados a cada traço. **Fonte:** Elaboração própria.

A análise de sobreposição com o catálogo GWAS revelou que uma fração significativa dos genes mapeados pelos SNPs identificados apresenta associações com fenótipos imunológicos (ex.: níveis circulantes de citocinas, troca de isotipo de imunoglobulina), hematológicos (ex.: contagem de reticulócitos, hemoglobina) e cognitivos. Tais achados sustentam a plausibilidade biológica dos *loci* detectados estarem associados a perturbações inflamatórias e hematológicas, contribuindo para as manifestações clínicas observadas.

4.5. Análise de enriquecimento de vias

Para aprofundar a interpretação dos achados genéticos identificados, foi realizada a análise de enriquecimento de vias biológicas também por meio da ferramenta snpXplorer (Tesi *et al.*, 2021). Essa abordagem possibilita associar as variantes genéticas identificadas a processos biológicos relevantes e identificar vias significativamente enriquecidas no banco de dados Reactome (REAC). Dessa forma, torna-se possível contextualizar os mecanismos moleculares potencialmente relacionados aos SNPs significativos. A tabela abaixo apresenta as vias biológicas enriquecidas, organizadas por grupo temático, a fim de facilitar a visualização e a compreensão dos resultados:

Tabela 4 - Enriquecimento de vias Reactome (REAC)

Grupo temático	Vias Reactome (p<0.05)
Apresentação antigênica e processamento de antígenos	Apresentação de antígenos: dobramento, montagem e carregamento de peptídeos MHC classe I*; Via endossômica/vacuolar*; Processamento - Apresentação cruzada de antígenos; Via do retículo endoplasmático–fagossomo.
Sinalização Interferônica e resposta antiviral	Sinalização de interferon alfa/beta; Sinalização de interferon gama; Interações do SARS-CoV-2 com o hospedeiro; O SARS-CoV-2 ativa/modula respostas imunes inata e adaptativa.
Interações do HIV e evasão imune	Regulação negativa mediada por Nef da expressão de MHC classe I na superfície celular*; A proteína Nef promove a modulação negativa de receptores da superfície celular ao recrutá-los para adaptadores de clatrina*; Interações dos fatores do HIV com o hospedeiro; O papel da proteína Nef na replicação do HIV-1 e na patogênese da doença.
Ubiquitinação e regulação pós-traducional	Ubiquitinação de proteínas; Ligases E3 de ubiquitina ubiquitinam proteínas-alvo.

Legenda: Vias biológicas enriquecidas que apresentaram significância estatística (p < 0,05) agrupadas conforme a função biológica. As vias com * correspondem às que apresentaram os menores valores p. **Fonte:** Elaboração própria.

5. Discussão

5.1. Acerca dos fatores sociodemográficos

O presente estudo sugere uma maior predisposição do sexo feminino ao desenvolvimento de sequelas no SNPer. Esses achados estão em concordância com pesquisas anteriores, como o estudo de Saxena e Mautner (2025), que indica um risco aumentado de COVID longa em mulheres em comparação aos homens. Embora ainda não haja uma explicação definitiva para essa maior suscetibilidade, a desregulação hormonal tem sido sugerida como um fator relevante.

Além disso, de acordo com Jensen *et al.* (2022), os mecanismos do sistema imune inato feminino resultam em uma resposta imunológica mais eficiente, promovendo uma eliminação mais rápida de patógenos e maior eficácia vacinal. No entanto, essa hiperativação do sistema imune também pode aumentar o risco de respostas desreguladas, tornando as mulheres mais propensas ao desenvolvimento de doenças autoimunes e da inflamação crônica persistente, que desempenha um papel chave no desenvolvimento ou persistência de sequelas pós-COVID-19.

Corroborando essas evidências, o estudo de Shah *et al.* (2025) evidenciou que o sexo feminino está significativamente associado ao maior risco de desenvolvimento de COVID longa, mesmo após ajuste por fatores demográficos, clínicos e sociais. O risco foi particularmente elevado em mulheres entre 40 e 54 anos (RR = 1,48; razão de risco que indica um aumento de 48% no risco em comparação aos homens), especialmente naquelas não menopausadas, sugerindo um papel modulador dos hormônios sexuais femininos. A pesquisa também identificou que sintomas como fadiga, mal-estar pós-esforço, confusão mental e palpitações foram mais prevalentes entre mulheres com COVID longa. Esses dados fortalecem a hipótese de que fatores hormonais, imunológicos e inflamatórios específicos do sexo feminino contribuem para uma trajetória clínica distinta e mais propensa à cronicidade pós-infecção pelo SARS-CoV-2.

Por esse motivo, pesquisadores têm postulado que diferenças sexuais na resposta imunológica podem desempenhar um papel central na evolução da COVID-19. Homens, de modo geral, apresentam uma resposta imune menos robusta, o que os torna mais suscetíveis à forma grave da infecção. Em contraste, mulheres tendem a ter uma resposta imunológica mais vigorosa e eficiente, favorecendo a eliminação do vírus durante a fase aguda. No entanto, esse padrão de ativação imune elevado e sustentado, característico do sexo feminino, pode aumentar a predisposição ao desenvolvimento de sequelas pós-infecciosas, com traços inflamatórios e autoimunes persistentes, como observado em casos de COVID longa (Jensen *et al.*, 2022).

Além disso, os dados analisados neste estudo indicam que a gravidade da infecção inicial está associada a um risco aumentado de sequelas no SNPer. Indivíduos que evoluem para formas graves da COVID-19 tendem a apresentar um estado inflamatório sistêmico (MIS, do inglês *multisystem inflammatory syndrome*) exacerbado, caracterizado por tempestade de citocinas, disfunção endotelial, alterações na coagulação e lesões teciduais multissistêmicas (Booker *et al.*, 2025). Tais alterações fisiopatológicas intensas podem desencadear danos duradouros no sistema nervoso periférico. Assim, tanto a inflamação persistente associada à resposta imunológica intensa em mulheres quanto os efeitos da inflamação aguda grave contribuem, por mecanismos distintos, para o surgimento de sequelas pós-COVID-19.

5.2. Acerca das vias enriquecidas e dos genes relacionados

A análise de regressão logística identificou SNPs estatisticamente significativos, que foram mapeados aos seus respectivos genes candidatos. A integração desses achados com dados de enriquecimento funcional e da literatura possibilitou a construção de algumas hipóteses biológicas plausíveis.

A análise de enriquecimento funcional revelou o predomínio de vias associadas à apresentação antigênica e processamento de antígenos mediados por MHC classe I

e de interações do HIV e evasão imune, consistentes com descrições prévias de processos envolvidos na COVID-19 (Rana; Ignatz-Hoover; Driscoll, 2023; Zhang *et al.*, 2021). É importante ressaltar que a maior parte dessas vias foi relacionada principalmente ao gene *HLA-A* (rs2975057), reconhecido por seu papel central na regulação da resposta imunológica adaptativa (Bouayad, 2021).

HLA-A - Human Leukocyte Antigen A (rs2975057)

O gene *HLA-A* codifica uma proteína do complexo principal de histocompatibilidade (MHC) classe I, fundamental para a apresentação de antígenos e ativação de linfócitos T CD8+, desempenhando um papel essencial na resposta imune contra patógenos intracelulares, incluindo o SARS-CoV-2 (*Gangaev et al.*, 2021). A sua expressão é modulada por citocinas como interferon-gama e interferon-alfa/beta, vias que também foram enriquecidas na análise funcional. Estudos prévios demonstraram que variações na região de *HLA-A* podem influenciar na suscetibilidade e na evolução de doenças infecciosas e autoimunes, corroborando a relevância do presente achado (Langton *et al.*, 2021).

O polimorfismo *HLA-A* rs2975057 demonstrou, no presente estudo, um possível efeito protetor contra o desenvolvimento de sequelas no SNPer. A maioria dos alelos clássicos do gene *HLA-A* é definida por variações nos éxons 2 e 3, que codificam a região de ligação ao peptídeo (Robinson *et al.*, 2017). No entanto, o rs2975057 está localizado em uma região intrônica, e, até o momento, não foi descrito como determinante de alelos clássicos, tampouco como marcador funcional de risco ou proteção em doenças neurológicas. Isso indica que o efeito protetor observado pode configurar uma evidência original ainda não relatada.

A complexidade estrutural do *locus HLA* dificulta o estabelecimento de vínculos precisos entre o SNP identificado e haplótipos funcionais. Contudo, esse mapeamento alélico não foi o foco do presente estudo. Estudos funcionais adicionais, incluindo análises de expressão gênica e caracterização de haplótipos, são essenciais para elucidar o papel biológico de rs2975057 e confirmar se esse

efeito protetor observado realmente representa um mecanismo previamente não descrito.

A substituição nucleotídica observada (G>A) corresponde a uma variante de natureza regulatória. Embora esse polimorfismo específico ainda não tenha sido descrito como marcador funcional em doenças neurológicas, a associação do gene *HLA-A* com efeitos imunoprotetores já é amplamente documentada na literatura, especialmente em contextos infecciosos. Estudos populacionais e metanálises indicam que alelos como *HLA-A*02*, *HLA-A*31* e *HLA-A*11* estão associados a uma resposta imunológica mais eficaz contra o SARS-CoV-2 e a menor risco de evolução para formas graves da COVID-19 (Dobrijević *et al.*, 2023; Littera *et al.*, 2020; Migliorini *et al.*, 2021).

Nos alelos *HLA-A*02* e *HLA-A*11*, observa-se um mecanismo imunológico particularmente relevante. Estes apresentam elevada capacidade de apresentação de epítopos do SARS-CoV-2 aos linfócitos T $CD8^+$, promovendo uma resposta citotóxica mais eficiente. Essa maior cobertura na apresentação antigênica está associada à redução do risco de formas graves da COVID-19, bem como de complicações pós-infecciosas (Migliorini *et al.*, 2021). Segundo Migliorini *et al.* (2021), a capacidade de apresentação de antígenos mediada por moléculas HLA constitui um dos fatores genéticos mais importantes para explicar a variabilidade interindividual na gravidade da infecção.

Estudos indicam que determinados alelos do sistema *HLA*, como o *HLA-A*02*, estão associados a respostas imunológicas mais eficazes e à menor gravidade da COVID-19 em algumas populações, conforme observado em análises realizadas na Itália e nos Emirados Árabes Unidos (Migliorini *et al.*, 2021; Tay *et al.*, 2023). De forma complementar, outras investigações sugerem que o alelo *HLA-A*23*, predominante em populações do Leste Asiático, está relacionado a um papel protetor contra a infecção pelo SARS-CoV-2, enquanto o alelo *HLA-A*02*, mais frequente em europeus, associa-se a um risco reduzido de mortalidade entre

pacientes infectados. Esses achados reforçam as evidências de que variantes específicas do *HLA* podem exercer efeitos protetores diferenciados entre populações globalmente distintas (Soko *et al.*, 2022). Tal constatação também corrobora os resultados de Fakhkhari; Caidi; Sadki (2023), que destacam que a associação entre alelos de *HLA* e os desfechos clínicos da COVID-19 pode ser modulada por fatores genéticos próprios de cada população, considerando o elevado grau de polimorfismo desse sistema entre diferentes grupos étnicos.

Tais achados evidenciam que determinados polimorfismos ou alelos do gene *HLA-A* podem influenciar a apresentação antigênica e a ativação de linfócitos T $CD8^+$, contribuindo para uma eliminação viral mais eficiente e, por conseguinte, para a mitigação de complicações de longo prazo. A convergência dessas evidências reforça a importância funcional de variantes em *HLA-A* como potenciais moduladores genéticos da susceptibilidade a desfechos pós-infecciosos.

Nesse contexto, levanta-se a hipótese de que o polimorfismo rs2975057 possa estar associado a uma regulação positiva da expressão de *HLA-A*, promovendo maior ativação das respostas imune inata e adaptativa. Especificamente, um aumento na expressão de moléculas *HLA-A* na superfície celular poderia favorecer a apresentação mais eficiente de antígenos aos linfócitos T $CD8^+$, intensificando a resposta imune citotóxica e limitando os danos prolongados, como aqueles observados nas sequelas neurológicas pós-COVID-19.

Por fim, é relevante destacar que o predomínio do gene *HLA-A* nas vias identificadas pode ser atribuído ao seu papel central na resposta imune. Esse gene codifica diretamente moléculas do complexo principal de histocompatibilidade de classe I (MHC I), elementos essenciais no processamento e apresentação de antígenos aos linfócitos T citotóxicos. Sua expressão é regulada por citocinas, como os interferons, e pode ser alvo de estratégias de evasão viral, a exemplo da modulação promovida pela proteína Nef do HIV (Mwimanzi *et al.*, 2012). Diante de sua elevada relevância funcional, é plausível que as análises de enriquecimento funcional apresentem forte

influência de *HLA-A*, refletindo sua participação proeminente em diversas vias imunorregulatórias.

Pseudogenes *ANO7L1 - Anoctamin 7 Like 1* (rs12746138), *AGAP12P - ArfGAP With GTPase Domain, Ankyrin Repeat And PH Domain 12* (rs61837957) e *GOLGA6FP - Golgin A6 Family Member F* (rs7182575)

Embora os genes *ANO7L1*, *AGAP12P* e *GOLGA6FP* não tenham sido associados diretamente às vias enriquecidas na análise funcional, estes correspondem a pseudogenes que não codificam proteínas funcionais. O gene *ANO7L1* encontra-se sobre-expresso na mucosa do esôfago (GeneCards – The Human Gene Database, 2024) e seu SNP identificado relaciona-se à proteção. O gene *AGAP12P*, com SNP associado a risco, é expresso em diversos tecidos, com destaque para a sobre-expressão nos testículos e presença em células-tronco germinativas masculinas e no ápice do coração (GeneCards – The Human Gene Database, 2024). Já o gene *GOLGA6FP*, com SNP associado à proteção, localiza-se no cromossomo 15, pertence à família *golgin A6* e é anotado como pseudogene expresso em tecidos do trato reprodutivo masculino, embora sem função proteica caracterizada (GeneCards – The Human Gene Database, 2024). Apesar de suas funções permanecerem pouco caracterizadas, é plausível que variações nessas regiões genômicas exerçam efeitos indiretos sobre mecanismos regulatórios de expressão gênica ou modulação epigenética (An *et al.*, 2017; Milligan & Lipovich, 2015), aspectos ainda não explorados de forma sistemática na literatura.

Variante intergênica (rs2983757)

Até o momento, também não foram identificadas informações na literatura científica que relacionem o SNP intergênico associado a risco identificado (rs2983757) a qualquer fenótipo descrito. A ausência de associação com vias imunológicas clássicas não invalida sua potencial relevância, mas reforça a necessidade de investigações adicionais que possam esclarecer os papéis dessas variantes na COVID-19 e na COVID longa.

***NPIPB15 - Nuclear Pore Complex Interacting Protein Family Member B15* (rs286859)**

O gene *NPIPB15* não foi diretamente associado às vias enriquecidas na análise funcional. Entretanto, considerando sua função relacionada ao complexo do poro nuclear (GeneCards – The Human Gene Database, 2024), é plausível hipotetizar que a variante identificada exerce efeito protetor contra sequelas no SNPer pós-COVID-19 possivelmente por modular mecanismos regulatórios do transporte nucleocitoplasmático. Especula-se que essa modulação possa limitar a translocação de fatores pró-inflamatórios ou preservar a homeostase celular em neurônios periféricos, uma vez que o poro nuclear desempenha papel central na entrada e saída de mediadores inflamatórios e elementos do ciclo viral (Guo *et al.*, 2023, Zhang *et al.*, 2025). Assim, polimorfismos nessa família gênica poderiam influenciar a magnitude e a duração da resposta inflamatória local, mitigando processos neurodegenerativos ou alterações na função axonal. Embora essa hipótese careça de confirmação experimental, ela sugere que variantes em genes envolvidos no transporte nuclear podem representar potenciais marcadores genéticos de suscetibilidade ou não a complicações neurológicas pós-infecciosas.

***PAQR5 - Progestin and AdipoQ Receptor Family Member 5* (rs2242464)**

O gene *PAQR5* também não foi diretamente ligado às vias identificadas, mas codifica um receptor de membrana relacionado à sinalização de progestogênios, moléculas que apresentam expressão detectável em diversos tecidos, incluindo alguns segmentos do sistema nervoso. Esses ligantes exercem funções anti-inflamatórias, neuroprotetoras e participam da manutenção da integridade tecidual (Melcangi; Panzica, 2009). Considerando que progestinas e seus receptores nucleares desempenham funções reconhecidamente neuroprotetoras e moduladoras da inflamação em diversos modelos experimentais (Bassani *et al.*, 2023; Giatti; Melcangi; Pesaresi, 2016), é plausível supor que *PAQR5*, enquanto membro da família de receptores de progestina de membrana, possa influenciar respostas

inflamatórias sistêmicas ou a sinalização hormonal em tecidos neurais.

Assim, variantes neste gene poderiam, indiretamente, alterar a suscetibilidade ou a recuperação frente a danos induzidos por processos infecciosos, como observado em sequelas neurológicas pós-COVID-19. O polimorfismo associado ao risco encontrado neste estudo situa-se em sua região regulatória, podendo afetar sua expressão gênica reduzindo a proteção fisiológica frente à inflamação induzida pela COVID-19. Consequentemente, tal situação pode favorecer processos degenerativos ou retardar a regeneração axonal, contribuindo para a persistência das manifestações neurológicas reportadas pelos pacientes. Essa hipótese, contudo, carece de validação experimental, uma vez que não há demonstração funcional direta da participação de *PAQR5* na homeostase do sistema nervoso.

6. Conclusão

Os resultados obtidos no presente estudo fornecem evidências preliminares de que fatores sociodemográficos, clínicos e genéticos podem atuar de maneira integrada na determinação do risco ou proteção frente a sequelas no sistema nervoso periférico pós-COVID-19.

A maior suscetibilidade observada nas mulheres e em indivíduos que apresentaram formas graves da doença corrobora hipóteses já estabelecidas de que a inflamação sistêmica exacerbada e a ativação imune persistente sejam elementos centrais na patogênese das manifestações neurológicas prolongadas.

As análises conduzidas por meio do classificador complexo, da regressão logística e do mapeamento funcional dos SNPs identificados evidenciaram o papel potencial e sem precedentes descritos na literatura da variante rs2975057 do gene *HLA-A*. Esta variante, possivelmente envolvida nos mecanismos clássicos de apresentação antigênica e na resposta imunológica adaptativa, demonstrou contribuição protetora frente ao desenvolvimento de neuropatias periféricas.

A associação de variantes nos genes *NPIPB15* e *PAQR5* também se destaca como um achado relevante. Observou-se que a variante nos gene *NPIPB15* conferiu efeito protetor, enquanto a variante no gene *PAQR5* esteve associada a maior risco de neuropatias periféricas. Esses resultados sugerem que mecanismos relacionados ao transporte nucleocitoplasmático e à sinalização de progestogênios podem exercer influência significativa sobre a suscetibilidade individual a danos neuronais e processos de desmielinização.

As variantes localizadas em pseudogenes e regiões intergênicas, como *ANO7L1*, *AGAP12P, GOLGA6FP* e rs2983757, ainda que não vinculadas diretamente a vias imunológicas reconhecidas, apresentaram associações relevantes: *ANO7L1 e GOLGA6FP* conferiram efeito protetor, enquanto *AGAP12P* e rs2983757 estiveram associadas a maior risco de neuropatias periféricas. Esses achados sugerem que tais variantes podem exercer efeitos regulatórios indiretos sobre processos biológicos envolvidos na suscetibilidade a danos neuronais, destacando a necessidade de investigação aprofundada.

Tais hipóteses, embora ainda careçam de confirmação experimental, apontam novos caminhos de pesquisa que podem contribuir para o esclarecimento da fisiopatologia da COVID longa e para o desenvolvimento de estratégias de prevenção e monitoramento de sequelas neurológicas prolongadas em populações vulneráveis.

7. Contribuição biotecnológica e perspectivas futuras

Os resultados deste estudo oferecem uma contribuição significativa para a área da biotecnologia, pois identificaram um conjunto de variantes genéticas que podem compor painéis de rastreamento genômico voltados à predição do risco de sequelas neurológicas pós-COVID-19.

A caracterização de SNPs associados a risco e proteção, combinada ao uso de modelos de aprendizado de máquina, fortalece o desenvolvimento de biomarcadores de suscetibilidade com potencial aplicação em estratégias de medicina

personalizada, programas de vigilância genômica e identificação de novos alvos terapêuticos. Essa abordagem integrada evidencia o potencial translacional dos achados, aproximando a pesquisa básica das práticas clínicas inovadoras no contexto da saúde pública e da biotecnologia aplicada.

Entre os possíveis desdobramentos, destacam-se a criação de kits diagnósticos baseados em painéis de genotipagem, o desenvolvimento de plataformas digitais de apoio à decisão clínica para estratificação de risco, o armazenamento estruturado dos dados genéticos em biobancos integrados, além da implementação de programas de monitoramento genômico populacional que possam subsidiar políticas públicas direcionadas ao acompanhamento de indivíduos vulneráveis a sequelas neurológicas prolongadas.

8. Limitações

A ferramenta snpXplorer também foi empregada para o mapeamento de variantes, a fim de atribuir a cada SNP, previamente identificado por posição cromossômica pelo classificador, um rsID e um gene correspondente. Para isso, além do mapeamento posicional clássico, a ferramenta integra inferências derivadas de *expression Quantitative Trait Loci* (eQTL), *splicing Quantitative Trait Loci* (sQTL), escores preditivos como CADD, impactos em regiões codificantes e elementos regulatórios, bem como efeitos de co-localização genômica, que permitem relacionar variantes a genes próximos ou funcionalmente conectados. Essa abordagem exploratória, que combina múltiplas bases de dados secundárias, possibilita a associação das variantes a genes de forma mais abrangente do que métodos exclusivamente posicionais. No entanto, tais anotações possuem caráter preditivo e, portanto, demandam validação adicional, que inclua: (i) a confirmação em bancos de dados posicionais clássicos; e (ii) análises funcionais complementares, como testes de expressão gênica, edição genômica e ensaios de atividade regulatória, a fim de garantir maior confiabilidade na atribuição variante-gene realizada.

Além disso, é importante destacar que a diferença entre os métodos de análise de dados do classificador e da regressão logística ajuda a explicar por que surgiram resultados aparentemente opostos para a variante intergênica rs2983757. Essa variante apareceu como protetora na análise do classificador, mas como fator de risco na LR.

Isso possivelmente aconteceu porque o classificador utiliza técnicas de inteligência artificial chamadas algoritmos de aprendizado de máquina, que têm como objetivo principal prever com mais precisão quem pertence a cada grupo, caso e controle. Para isso, o classificador identifica padrões complexos nos dados, incluindo combinações entre diferentes variantes genéticas, que ajudam a separar melhor os participantes com e sem sequelas. Assim, ele pode indicar que uma determinada variante contribui para risco ou proteção porque, no conjunto de todas as informações avaliadas, sua presença aumenta a capacidade de distinção entre os grupos.

Por esse motivo, a análise com o classificador é realizada primeiro, mas não representa a etapa conclusiva. Após essa triagem inicial, seguem-se análises estatísticas mais tradicionais e detalhadas, neste caso, a regressão logística, que permite confirmar ou aprofundar o entendimento desses achados.

Diferentemente do classificador, a regressão logística tem como principal objetivo estimar o efeito individual de cada variante genética, considerando sua contribuição isolada ou ajustada pelas demais variáveis incluídas no modelo. Esse método é mais interpretativo e é tradicionalmente empregado para verificar se uma variante específica, por si só, está associada ao aumento ou à redução do risco de determinada condição. Nessa etapa, caso ainda permaneçam, mesmo após a triagem inicial com o classificador, variáveis altamente correlacionadas entre si (situação conhecida como multicolinearidade), o modelo de regressão pode produzir resultados distintos daqueles observados no classificador.

Essa divergência é particularmente frequente em análises genômicas, uma vez que o genoma apresenta situações como o desequilíbrio de ligação, em que variantes próximas tendem a ser herdadas em conjunto, e as interações epistáticas, nas quais o efeito de uma variante depende da presença de outras. Essas particularidades genômicas contribuem para padrões complexos nos dados e tornam mais difícil isolar e interpretar os efeitos individuais de cada variante na análise estatística.

Portanto, o resultado divergente entre o classificador e a regressão logística observado para a variante rs2983757 pode refletir essencialmente uma diferença conceitual entre os métodos. Pode-se considerar que, de forma isolada, essa variante esteja associada ao aumento do risco, mas, em combinação com outras variantes genéticas e fatores não genéticos, sua presença pode contribuir para um efeito protetor.

Essa é justamente a maior dificuldade de se trabalhar com dados genômicos, em razão da alta multicolinearidade e da complexidade das interações entre variantes. Estudos futuros com amostras mais amplas e metodologias que combinem abordagens preditivas e interpretativas, como *elastic net regression*, *random forests*, *gradient boosting machines* e análise de redes de interação gênica, poderão contribuir para esclarecer de forma mais robusta a relevância de cada associação identificada no desenvolvimento e/ou persistência de sequelas neurológicas periféricas.

O grupo de pesquisa do Núcleo de Genética Humana e Molecular (NGHM) da Universidade Federal do Espírito Santo (UFES) encontra-se ainda em processo de capacitação técnica nessa área, que constitui uma fronteira do conhecimento científico e carece de profissionais especializados. O modelo metodológico empregado neste estudo seguiu protocolos descritos em investigações previamente publicadas, como o trabalho de Pastor *et al.*, (2023), conduzido por pesquisadores com ampla experiência na área. A exploração de estratégias analíticas mais inovadoras e complexas, como as mencionadas anteriormente, depende do

fortalecimento contínuo da capacitação técnica da equipe, bem como do estabelecimento de colaborações multidisciplinares e interinstitucionais que possam ampliar o domínio e a aplicação dessas abordagens no contexto da genômica aplicada à saúde.

9. Referências bibliográficas

ALADAWI, Mohammad *et al.* Guillain Barre Syndrome as a Complication of COVID-19: A Systematic Review. **Canadian Journal of Neurological Sciences**, [*s. l.*], v. 49, n. 1, p. 38–48, 2022.

ALIMADADI, Ahmad *et al.* Artificial intelligence and machine learning to fight COVID-19. **Physiological Genomics**, [*s. l.*], v. 52, n. 4, p. 200–202, 2020.

AN, Yang; FURBER, Kendra L.; JI, Shaoping. Pseudogenes regulate parental gene expression via ceRNA network. **Journal of Cellular and Molecular Medicine**, [*s. l.*], v. 21, n. 1, p. 185–192, 2017.

ARMOCIDA, Daniele *et al.* How SARS-Cov-2 can involve the central nervous system. A systematic analysis of literature of the department of human neurosciences of Sapienza University, Italy. **Journal of Clinical Neuroscience**, [*s. l.*], v. 79, p. 231–236, 2020.

ASTERIS, Panagiotis G. *et al.* Genetic prediction of ICU hospitalization and mortality in COVID-19 patients using artificial neural networks. **Journal of Cellular and Molecular Medicine**, [*s. l.*], v. 26, n. 5, p. 1445–1455, 2022.

BASSANI, Taysa Bervian *et al.* Progestogen-Mediated Neuroprotection in Central Nervous System Disorders. **Neuroendocrinology**, [*s. l.*], v. 113, n. 1, p. 14–35, 2023.

BEREZHNOY, Georgy *et al.* Maintained imbalance of triglycerides, apolipoproteins, energy metabolites and cytokines in long-term COVID-19 syndrome patients. **Frontiers in Immunology**, [*s. l.*], v. 14, 2023. Disponível em: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2023.1144224/full. Acesso em: 6 mar. 2025.

BINIAZ-HARRIS, Nicholas; KUVALDINA, Mara; FALLON, Brian A. Neuropsychiatric Lyme Disease and Vagus Nerve Stimulation. **Antibiotics**, [*s. l.*], v. 12, n. 9, p. 1347, 2023.

BISWAS, Subrata K.; MUDI, Sonchita R. Spike protein D614G and RdRp P323L: the SARS-CoV-2 mutations associated with severity of COVID-19. **Genomics & Informatics**, [*s. l.*], v. 18, n. 4, p. e44, 2020.

BOOKER, Anthony *et al.* Editorial: Multisystem inflammatory syndrome observed post-COVID-19: the role of natural products, medicinal plants and nutrients and the use of prediction tools supporting traditional forms of diagnosis. **Frontiers in Pharmacology**, [*s. l.*], v. 16, 2025. Disponível em: https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2025.1539793/full. Acesso em: 5 mar. 2025.

BOUAYAD, Abdellatif. Features of HLA class I expression and its clinical relevance in SARS-CoV-2: What do we know so far?. **Reviews in Medical Virology**, [*s. l.*], v. 31, n. 6, p. e2236, 2021.

BRASIL. Ministério da Saúde. *Nota Técnica nº 57/2023* – Atualizações acerca das condições pós-COVID no âmbito do Ministério da Saúde. Brasília, DF, 2023. Disponível em:<https://bvsms.saude.gov.br/bvs/publicacoes/nota_tecnica_n57_atualizacoes_condicoes_poscovid.pdf>. Acesso em: 16 fev. 2025.

CAMPEN, C. (Linda) M. C. Van; ROWE, Peter C.; VISSER, Frans C. Orthostatic Symptoms and Reductions in Cerebral Blood Flow in Long-Haul COVID-19 Patients: Similarities with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. **Medicina**, [*s. l.*], v. 58, n. 1, p. 28, 2021.

CARABELLI, Alessandro M. *et al.* SARS-CoV-2 variant biology: immune escape, transmission and fitness. **Nature Reviews Microbiology**, [*s. l.*], 2023. Disponível em: https://www.nature.com/articles/s41579-022-00841-7. Acesso em: 6 maio 2025.

CARVALHO, Elizeu Fagundes De *et al.* **Além da pandemia: desvendando a covid longa e suas múltiplas facetas**. 1. ed. [*S. l.*]: Atena Editora, 2024. Disponível em: https://atenaeditora.com.br/catalogo/ebook/alem-da-pandemia-desvendando-a-covid-longa-e-suas-multiplas-facetas. Acesso em: 15 jan. 2025.

CAVANAUGH, Joseph E.; NEATH, Andrew A. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. **WIREs Computational Statistics**, [*s. l.*], v. 11, n. 3, p. e1460, 2019.

CHAO, Julie; CHAO, Lee. Kallikrein–kinin in stroke, cardiovascular and renal disease. **Experimental Physiology**, [*s. l.*], v. 90, n. 3, p. 291–298, 2005.

CHEN, Xue-wen; JEONG, Jong Cheol. Enhanced recursive feature elimination. *In*: SIXTH INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS (ICMLA 2007), 2007, Cincinnati, OH, USA. **Sixth International Conference on Machine Learning and Applications (ICMLA 2007)**. Cincinnati, OH, USA: IEEE, 2007. p. 429–435. Disponível em: http://ieeexplore.ieee.org/document/4457268/. Acesso em: 18 fev. 2025.

CRUZ, Juliana de O. et al. Functional prediction and frequency of coding variants in human ECA2 at binding sites with SARS‐CoV‐2 spike protein on different populations. **Journal of Medical Virology**, v. 93, n. 1, p. 71, 2021.

CRUZ-TAPIAS, Paola; CASTIBLANCO, John; ANAYA, Juan-Manuel. Major histocompatibility complex: Antigen processing and presentation. *In*: AUTOIMMUNITY: FROM BENCH TO BEDSIDE [INTERNET]. [*S. l.*]: El Rosario University Press, 2013. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK459467/. Acesso em: 7 mar. 2025.

DOBRIJEVIĆ, Z. et al. The association of human leucocyte antigen (HLA) alleles with COVID-19 severity: a systematic review and meta-analysis. **Reviews in Medical Virology**, v. 33, n. 3, p. e2378, 2023. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1002/rmv.2378. Acesso em: 9 jul. 2025.

ELY, E. Wesley; BROWN, Lisa M.; FINEBERG, Harvey V. Long Covid Defined. **New England Journal of Medicine**, [*s. l.*], v. 391, n. 18, p. 1746–1753, 2024.

ENDO, Yusuke; KANNO, Toshio; NAKAJIMA, Takahiro. Fatty acid metabolism in T-cell function and differentiation. **International Immunology**, [*s. l.*], v. 34, n. 11, p. 579–587, 2022.

ERCEGOVAC, Marko et al. Antioxidant Genetic Profile modifies probability of developing neurological sequelae in Long-COVID. **Antioxidants**, v. 11, n. 5, p. 954, 2022.

FAKHKHARI, Meryem; CAIDI, Hayat; SADKI, Khalid. HLA alleles associated with COVID-19 susceptibility and severity in different populations: a systematic review. **Egyptian Journal of Medical Human Genetics**, [*s. l.*], v. 24, n. 1, p. 10, 2023.

FALLERINI, Chiara *et al.* Common, low-frequency, rare, and ultra-rare coding variants contribute to COVID-19 severity. **Human Genetics**, [*s. l.*], v. 141, n. 1, p. 147–173, 2022.

FANG, Cong *et al.* Deep learning for predicting COVID-19 malignant progression. **Medical Image Analysis**, [*s. l.*], v. 72, p. 102096, 2021.

GANGAEV, Anastasia *et al.* Identification and characterization of a SARS-CoV-2 specific CD8+ T cell response with immunodominant features. **Nature Communications**, [*s. l.*], v. 12, n. 1, p. 2593, 2021.

GENECARDS – THE HUMAN GENE DATABASE. AGAP12P Gene – ArfGAP with GTPase Domain, Ankyrin Repeat and PH Domain 12, Pseudogene. Rehovot: Weizmann Institute of Science, 2024. Disponível em: https://www.genecards.org/cgi-bin/carddisp.pl?gene=AGAP12P. Acesso em: 3 jul. 2024.

GENECARDS – THE HUMAN GENE DATABASE. ANO7L1 Gene – Anoctamin 7 Like 1. Rehovot: Weizmann Institute of Science, 2024. Disponível em: https://www.genecards.org/cgi-bin/carddisp.pl?gene=ANO7L1&keywords=ANO7L1. Acesso em: 3 jul. 2024.

GENECARDS – THE HUMAN GENE DATABASE. GOLGA6FP (Golgin A6 Family Member F, Pseudogene). Rehovot: Weizmann Institute of Science, atualizado em 17 jul. 2025. Disponível em: https://www.genecards.org/cgi-bin/carddisp.pl?gene=GOLGA6FP&keywords=GOLGA6FP. Acesso em: 27 ago. 2025.

GENECARDS – THE HUMAN GENE DATABASE. NPIPB15 Gene – Nuclear Pore Complex Interacting Protein Family Member B15. Rehovot: Weizmann Institute of Science, 2024. Disponível em: https://www.genecards.org/cgi-bin/carddisp.pl?gene=NPIPB15. Acesso em: 3 jul. 2024.

GIATTI, Silvia; MELCANGI, Roberto Cosimo; PESARESI, Marzia. The other side of progestins: effects in the brain. **Journal of Molecular Endocrinology**, [*s. l.*], v. 57, n. 2, p. R109–R126, 2016.

GIGLI, Gian Luigi *et al.* HLA and immunological features of SARS-CoV-2-induced Guillain-Barré syndrome. **Neurological Sciences**, [*s. l.*], v. 41, n. 12, p. 3391–3394, 2020.

GUO, Jiayin *et al.* Virus Infection and mRNA Nuclear Export. **International Journal of Molecular Sciences**, [*s. l.*], v. 24, n. 16, p. 12593, 2023.

GUTIÉRREZ-BAUTISTA, Juan Francisco *et al.* Study of HLA-A, -B, -C, -DRB1 and -DQB1 polymorphisms in COVID-19 patients. **Journal of Microbiology, Immunology and Infection**, [*s. l.*], v. 55, n. 3, p. 421–427, 2022.

HUSSAIN, Mushtaq et al. Structural variations in human ECA2 may influence its binding with SARS‐CoV‐2 spike protein. **Journal of medical virology**, v. 92, n. 9, p. 1580-1586, 2020.

ISMAIL, Ismail Ibrahim; SALAMA, Sara. Association of CNS demyelination and COVID-19 infection: an updated systematic review. **Journal of Neurology**, [*s. l.*], v. 269, n. 2, p. 541–576, 2022.

JAMIL AL-OBAIDI, Mazen M.; DESA, Mohd Nasir Mohd. A review of the mechanisms of blood-brain barrier disruption during COVID-19 infection. **Journal of Neuroscience Research**, [*s. l.*], v. 101, n. 11, p. 1687–1698, 2023.

JENSEN, Adelaide *et al.* Sex and gender differences in the neurological and neuropsychiatric symptoms of long COVID: a narrative review. [*s. l.*], 2022.

LAMMI, Vilma *et al.* Genome-wide association study of long COVID. **Nature Genetics**, [*s. l.*], v. 57, n. 6, p. 1402–1417, 2025.

LANGTON, David J. *et al.* The influence of HLA genotype on the severity of COVID-19 infection. **HLA**, [*s. l.*], v. 98, n. 1, p. 14–22, 2021.

LESNAK, Joseph B.; MAZHAR, Khadijah; PRICE, Theodore J. Neuroimmune Mechanisms Underlying Post-acute Sequelae of SARS-CoV-2 (PASC) Pain, Predictions from a Ligand-Receptor Interactome. **Current Rheumatology Reports**, [*s. l.*], v. 25, n. 9, p. 169–181, 2023.

LI, Xiaochen *et al.* Risk factors for severity and mortality in adult COVID-19 inpatients

in Wuhan. **The Journal of Allergy and Clinical Immunology**, [*s. l.*], v. 146, n. 1, p. 110–118, 2020.

LITTERA, R. et al. Human leukocyte antigen complex and other immunogenetic and clinical factors influence susceptibility or protection to SARS-CoV-2 infection and severity of the disease course. **Frontiers in Immunology**, v. 11, p. 605688, 2020. Disponível em: https://www.frontiersin.org/articles/10.3389/fimmu.2020.605688/full. Acesso em: 9 jul. 2025.

LUNDBERG, Scott M; LEE, Su-In. A Unified Approach to Interpreting Model Predictions. *In*: , 2017. **Advances in Neural Information Processing Systems**. [*S. l.*]: Curran Associates, Inc., 2017. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html. Acesso em: 3 mar. 2025.

MCFARLAND, Amelia J. *et al.* Neurobiology of SARS-CoV-2 interactions with the peripheral nervous system: implications for COVID-19 and pain. **PAIN Reports**, [*s. l.*], v. 6, n. 1, p. e885, 2021.

MELCANGI, Roberto Cosimo; PANZICA, Giancarlo. Neuroactive steroids: an update of their roles in central and peripheral nervous system. **Psychoneuroendocrinology**, [*s. l.*], v. 34 Suppl 1, p. S1-8, 2009.

MIGLIORINI, F. et al. Association between HLA genotypes and COVID-19 susceptibility, severity and progression: a comprehensive review of the literature. **European Journal of Medical Research**, v. 26, n. 1, p. 84, 2021. Disponível em: https://link.springer.com/article/10.1186/s40001-021-00563-1. Acesso em: 9 jul. 2025.

MILLIGAN, Michael J.; LIPOVICH, Leonard. Pseudogene-derived lncRNAs: emerging regulators of gene expression. **Frontiers in Genetics**, [*s. l.*], v. 5, 2015. Disponível em: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2014.00476/full. Acesso em: 9 jul. 2025.

MWIMANZI, Philip *et al.* Human Leukocyte Antigen (HLA) Class I Down-Regulation by Human Immunodeficiency Virus Type 1 Negative Factor (HIV-1 Nef): What Might We Learn From Natural Sequence Variants?. **Viruses**, [*s. l.*], v. 4, n. 9, p. 1711–1730, 2012.

NIEMI, Mari E. K. *et al.* Mapping the human genetic architecture of COVID-19. **Nature**, [*s. l.*], v. 600, n. 7889, p. 472–477, 2021.

OLIVEIRA, Fabrízzio Condé de. Um método para seleção de atributos em dados genômicos. [*s. l.*], 2015.

PASTOR, André Filipe *et al.* Human Genome Polymorphisms and Computational Intelligence Approach Revealed a Complex Genomic Signature for COVID-19

Severity in Brazilian Patients. **Viruses**, [*s. l.*], v. 15, n. 3, p. 645, 2023.

RANA, Priyanka S.; IGNATZ-HOOVER, James J.; DRISCOLL, James J. Targeting Proteasomes and the MHC Class I Antigen Presentation Machinery to Treat Cancer, Infections and Age-Related Diseases. **Cancers**, [*s. l.*], v. 15, n. 23, p. 5632, 2023.

ROBINSON, James *et al.* Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. **PLOS Genetics**, [*s. l.*], v. 13, n. 6, p. e1006862, 2017.

SANTOS, Vinicius De Souza. Comparison and selection of machine learning algorithms for diabetes prediction: An exploratory quantitative study based on medical data analysis. *In*: MULTIDISCIPLINARY PERSPECTIVES: INTEGRATING KNOWLEDGE. 1. ed. [*S. l.*]: Seven Editora, 2024. Disponível em: https://sevenpublicacoes.com.br/index.php/editora/article/view/4102. Acesso em: 26 fev. 2025.

SAXENA, Apoorva; MAUTNER, Josef. A Disease Hidden in Plain Sight: Pathways and Mechanisms of Neurological Complications of Post-acute Sequelae of COVID-19 (NC-PASC). **Molecular Neurobiology**, [*s. l.*], v. 62, n. 2, p. 2530–2547, 2025.

SCHULTZ, Verena *et al.* Zika Virus Infection Leads to Demyelination and Axonal Injury in Mature CNS Cultures. **Viruses**, [*s. l.*], v. 13, n. 1, p. 91, 2021.

SHAFQAT, Areez et al. Neutrophil extracellular traps and long COVID. **Frontiers in Immunology**, v. 14, 2023.

SIDERATOU, Christina-Michailia; PAPANEOPHYTOU, Christos. Persisting Shadows: Unraveling the Impact of Long COVID-19 on Respiratory, Cardiovascular, and Nervous Systems. **Infectious Disease Reports**, v. 15, n. 6, p. 806-830, 2023.

SOKO, Nyarai D. *et al.* The COVID-19 Pandemic and Explaining Outcomes in Africa: Could Genomic Variation Add to the Debate?. **OMICS: A Journal of Integrative Biology**, [*s. l.*], v. 26, n. 11, p. 594–607, 2022.

SOKOLOVA, Marina; LAPALME, Guy. A systematic analysis of performance measures for classification tasks. **Information Processing & Management**, [*s. l.*], v. 45, n. 4, p. 427–437, 2009.

TAY, Guan K. *et al.* HLA class I associations with the severity of COVID-19 disease in the United Arab Emirates. **PLOS ONE**, [*s. l.*], v. 18, n. 9, p. e0285712, 2023.

TAKAHASHI, Yuta *et al.* Machine learning for effectively avoiding overfitting is a crucial strategy for the genetic prediction of polygenic psychiatric phenotypes. **Translational Psychiatry**, [*s. l.*], v. 10, n. 1, p. 1–11, 2020.

TESI, Niccolo *et al.* snpXplorer: a web application to explore human SNP-associations and annotate SNP-sets. [*s. l.*], 2021.

VISSCHER, Peter M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. **The American Journal of Human Genetics**, [*s. l.*], v. 101, n. 1, p. 5–22, 2017.

WANG, Cheng et al. ApoE-isoform-dependent SARS-CoV-2 neurotropism and cellular response. **Cell stem cell**, v. 28, n. 2, p. 331-342. e5, 2021.

WANG, Lian *et al.* Artificial Intelligence for COVID-19: A Systematic Review. **Frontiers in Medicine**, [*s. l.*], v. 8, p. 704256, 2021.

WORLD HEALTH ORGANIZATION. Coronavirus disease (COVID-19): post COVID-19 condition. 2023. Disponível em: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-post-covid-19-condition>. Acesso em: 18 fev. 2025.

WORLD HEALTH ORGANIZATION. WHO COVID-19 Dashboard. 2024. Disponível em: <https://data.who.int/dashboards/covid19/cases>. Acesso em: 16 fev. 2025.

ZHANG, Nan *et al.* Genomic Patterns are Associated with Different Sequelae of Patients with Long-Term COVID-19. **Advanced Science**, [*s. l.*], p. 2407342, 2024.

ZHANG, Xin *et al.* Strategies for the Viral Exploitation of Nuclear Pore Transport Pathways. **Viruses**, [*s. l.*], v. 17, n. 2, p. 151, 2025.

ZHANG, Yiwen *et al.* The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. **Proceedings of the National Academy of Sciences**, [*s. l.*], v. 118, n. 23, p. e2024202118, 2021.

ZSICHLA, Levente; MÜLLER, Viktor. Risk Factors of Severe COVID-19: A Review of Host, Viral and Environmental Factors. **Viruses**, [*s. l.*], v. 15, n. 1, p. 175, 2023.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

APÊNDICE A - PRODUÇÕES DURANTE O MESTRADO

Categoria	Descrição
Capítulos de livro	MEIRA, Débora Dummer *et al.* Chapter 6 - How does understanding epigenetics help circumvent HER-2 antibody resistance? In: OVERCOMING CANCERS RESISTANT TO HER-2 ANTIBODIES. [S.l.]: Academic Press, 2024. Disponível em: https://doi.org/10.1016/B978-0-12-816408-2.00001-4. Acesso em: 21 jul. 2025. SIQUEIRA, S. *et al.* COVID-19 e o sistema imunológico. In: ALÉM DA PANDEMIA: desvendando a covid longa. [S.l.]: Atena Editora, 2024. Disponível em: https://doi.org/10.22533/at.ed.508240503. Acesso em: 21 jul. 2025. GIACINTI, G. M. *et al.* Os efeitos da CL em outros sistemas corporais. In: ALÉM DA PANDEMIA: desvendando a covid longa. [S.l.]: Atena Editora, 2024. Disponível em: https://doi.org/10.22533/at.ed.508240503. Acesso em: 21 jul. 2025. LOIOLA, Graziela Moreira *et al.* Importância fitoquímica dos alimentos na sinalização celular e modulação epigenética. In: AS CIÊNCIAS BIOLÓGICAS. [S.l.]: Atena Editora, 2024. Disponível em: https://doi.org/10.22533/at.ed.975232509. Acesso em: 21 jul. 2025.
Artigos: primeira autoria	GUAITOLINI, Yasmin Moreto *et al.* Biotechnology and genetic engineering. RECIMA21, [S.l.], v. 5, n. 2, 2024. Disponível em: https://doi.org/10.47820/recima21.v5i2.4797. Acesso em: 21 jul. 2025. GUAITOLINI, Yasmin Moreto *et al.* Biblical nutrition and epigenetics. Journal of Religion and Health, [S.l.], [s.n.], [s.d.]. Ressubmetido.
Artigos: coautoria	CASOTTI, Matheus Correia *et al.* Syncytia: from a historical resumption to epigenetic advances. DNA and Cell Biology Reports, [S.l.], 2025. Disponível em: https://doi.org/10.1089/dcbr.2024.0037. Acesso em: 21 jul. 2025. COUTINHO, Emanuelle *et al.* Biofilme oral e doenças sistêmicas. Brazilian Journal of Health Review, [S.l.], v. 8, n. 4, 2025. Disponível em: https://doi.org/10.34119/bjhrv8n4-051. Acesso em: 21 jul. 2025. DOS SANTOS ALVARENGA, Flávio *et al.* Pós-COVID e transtornos depressivos. Brazilian Journal of Health Review, [S.l.], v. 8, n. 1, 2025. Disponível em: https://doi.org/10.34119/bjhrv8n1-131. Acesso em: 21 jul. 2025. SILVA FILHO, Luiz Claudio Gobbi Da *et al.* Neuroregenerative protein networks. Brazilian Archives of Biology and Technology, [S.l.], 2024. Disponível em: https://doi.org/10.1590/1678-4324-2024240133. Acesso em: 21 jul. 2025. CASOTTI, Matheus Correia *et al.* Integrating frontiers: a holistic, quantum and evolutionary approach to conquering cancer through systems biology and multidisciplinary synergy. Frontiers in Oncology, [S.l.], 2024. Disponível em: https://doi.org/10.3389/fonc.2024.1419599. Acesso em: 21 jul. 2025. MEIRA, Débora Dummer *et al.* Prognostic factors in lung cancer. Genes, [S.l.], v. 14, n. 10, 2023. Disponível em: https://doi.org/10.3390/genes14101906. Acesso em: 21 jul. 2025.
Artigos: submetidos	PEGOS, Arthur Ribeiro *et al.* Long COVID and fibromyalgia. Clinical Rheumatology, [S.l.], [s.n.], [s.d.]. Ressubmetido.

	CASOTTI, Matheus C. *et al.* Defining life. BioSystems, [S.l.], [s.n.], [s.d.]. Preprint disponível em: http://dx.doi.org/10.2139/ssrn.5105814.
Organização de eventos	Organização do I Simpósio de Genética Aplicada do Espírito Santo. 2024. Organização do I Simpósio de Ciências Forenses da LAIGGES e GenES. 2023.
Cursos/Treinamentos	Mulher Digital Júnior Achievment: segurança cibernética, desenvolvimento web e computação em nuvem (AWS Cloud Practitioner e Cisco CCST). Em andamento. Inova Ufes: Jornada Start Up. 03/06 a 22/07/2025. CH: 18h. Treinamento: Mapeamento da Ciência e Produção Científica. 2024. Curso de verão: Aplicação biotecnológica de fungos. Instituto Butantan, 01 a 02/10/2024. CH: 6h. Curso: Uso da Propriedade Intelectual em Negócios de Base Tecnológica. 05/2024. CH: 20h.
Participação em trabalhos e trabalhos apresentados	Resumo: Biossegurança e OGMs na Alimentação. 9º Encontro Bienal de Biossegurança. 2023. Resumo: Fronteiras na biossegurança do controle de pragas: uma revolução biotecnológica liderada pelos "fungos zumbis". 10º Encontro Bienal de Biossegurança. 2025. Prevenção inteligente: biossegurança e modelagem preditiva para infecções fúngicas. 10º Encontro Bienal de Biossegurança. 2025. Revolução simbiogênica: fungos & mixomicetos na fronteira da biossegurança e biotecnologia. 10º Encontro Bienal de Biossegurança. 2025. Resumo: A Influência da Dieta Mediterrânea na Modulação do Câncer. 1º Simpósio de Genética Aplicada do Espírito Santo. 2024. Resumo: O câncer como um estrategista regenerativo: descrições genéticas e holísticas. 1º Simpósio de Genética Aplicada do Espírito Santo. 2024 Resumo: Bioprospecção e Bioeconomia em Ecossistemas Terrestres e Marinhos no Combate de Doenças Complexas e Infecciosas: Da Genética à Sintética. 1º Simpósio de Genética Aplicada do Espírito Santo. 2024. Resumo: Genetic study of patients with persistent neurocognitive sequelae after COVID-19. Brazilian Symposium on Bioinformatics. 2024. Resumo: From Equipment Miniaturization to Universal Microbial Testing: A Real-Time, Personalized and Portable Planetary Protection. Science and Planetary Protection in Advance of Human Missions Seminar. 2024. Resumo: Grafos e Fluxos: Redes Fúngicas sob um olhar da Modelagem e Estatística. SEAGRO. 2025. Resumo: Virtual cells as modeling tools for epigenetics, ploidy, and cell fate under quantum principles. XV Encontro de Física Aplicada. 2025.