



Universidade de Brasília

Instituto de Ciências Biológicas

Pós-Graduação em Biologia Molecular

**Prospecção de marcadores transcricionais na
COVID-19**

Pedro Henrique Aragão Barros

Brasília
2023

Pedro Henrique Aragão Barros

**Prospecção de marcadores transcricionais na
COVID-19**

Orientador: Prof. Dr. Marcelo de Macedo Brígido

Co-orientadora: Prof. Dr. Waldeyr Mendes Cordeiro da Silva

Dissertação apresentada ao Programa de Pós-graduação em Biologia Molecular do Instituto de Ciências Biológicas da Universidade de Brasília para a obtenção do Título de Mestre em Biologia Molecular.

Brasília
2023

Pedro Henrique Aragão Barros

Prospecção de marcadores transcricionais na COVID-19

Trabalho desenvolvido no laboratório de Imunologia Molecular da Universidade de Brasília, sob a orientação do Prof. Dr. Marcelo de Macedo Brígido e do Prof. Dr. Waldeyr Mendes Cordeiro da Silva.

BANCA EXAMINADORA

Dr. Marcelo de Macedo Brígido (Presidente da Banca)

Dr. Georgios Joannis Pappas Júnior (Membro titular)

Dra. Gloria Regina Franco (Membro titular)

Dra. Andréa Queiroz Maranhão (Membro suplente)

LISTA DE ABREVIATURAS

| | |
|----------------|--|
| 7SL | Partícula reconhedora de sinal |
| A3SS | Sítio doador 3' alternativo de splicing |
| A5SS | Sítio doador 5' alternativo de splicing |
| ACE2 | Angiotensin Converting Enzyme 2 |
| ADAR | Adenosine deaminase acting on RNA |
| AngII | Angiotensina 2 |
| APOBEC | Apolipoprotein B mRNA Editing Catalytic Polypeptide-like |
| ATAC-seq | Transposase-Accessible Chromatin |
| BAM | Binary alignment maps |
| BPS | Região de ramificação |
| CALU-3 | Adenocarcinoma epitelial de pulmão humano |
| CARD | Caspase activation and recruitment domain |
| CDC | Centers for Disease Control and Prevention |
| ChIP-seq | Chromatin immunoprecipitation |
| CoV | Coronavírus |
| COVID-19 | Coronavirus disease 2019 |
| DNA | Ácido desoxirribonucleico |
| dsRNA | RNA dupla fita |
| E | Envelope |
| ER | Edição de RNA |
| ES | Exclusão exônica |
| FDR | False discovery rate |
| GO | Gene Ontology |
| GTF | Gene transfer format |
| HBV | Vírus da hepatite B |
| HIF1- α | Hypoxia-inducible factor 1-alpha |
| HPV | Vírus do papiloma humano |
| HR1 | Heptad repeat 1 |
| HR2 | Heptad repeat 2 |

| | |
|-----------------|---|
| HTLV-1 | Human T cell leukaemia virus type 1 |
| HuR | Human antigen R) |
| IAV | Vírus da Influenza A |
| ICTV | International Committee on Taxonomy of Viruses |
| IFIT2 | Interferon-induced protein with tetratricopeptide repeats 2 |
| IFITM1 | Interferon-induced transmembrane protein 1 |
| IFITM2 | Interferon-induced transmembrane protein 2 |
| IFN | Interferon |
| IFN-1 | IFN tipo 1 |
| IFN-3 | IFN tipo 3 |
| IFN- κ | Interferon kappa |
| IFN- α | Interferon alpha |
| IFN- β | Interferon beta |
| IFN- γ | Interferon gamma |
| IFN- ϵ | Interferon epsilon |
| IFN- ω | Interferon omega |
| IFNAR | Interferon- α/β receptor |
| IL | Interleucina |
| IR | Retenção intrônica |
| IRF | Interferon Regulatory Factor |
| ISG | Interferon-stimulated gene |
| JAK | Janus kinase |
| M | Membrana |
| MDA5 | melanoma differentiation-associated protein 5 |
| MERS | Middle east respiratory syndrome |
| miRNA | Micro RNA |
| mRNA | RNA mensageiro |
| MX2 | MX Dynamin Like GTPase 2 |
| MXE | Éxons mutuamente excludentes |
| N | Nucleocapsídeo |
| NCBI | National Center for Biotechnology Information |

| | |
|----------------|--|
| NF- κ B | Nuclear factor kappa B |
| NPS | Polimorfismo de nucleotídeo único |
| NSP | Nonstructural protein |
| ORF | Open Reading Frame |
| PAMPS | Pathogen-associated molecular pattern |
| PBMC | Células mononucleares de sangue periférico |
| PPR | Pattern recognition receptor |
| R0 | Basic reproductive number |
| RBD | Receptor binding domain |
| RBPs | RNA-binding proteins |
| RNA | Ácido Ribonucleico |
| S | Spike |
| SAM | Sequence alignment maps |
| SARS | Severe acute respiratory syndrome |
| SARS-CoV-2 | Severe acute respiratory syndrome coronavírus |
| SDRA | Síndrome do desconforto respiratório agudo |
| SF1 | Fator de splicing 1 |
| SF3B1 | Splicing factor 3b subunit 1 |
| SI | Sistema imune |
| SINEs | Short interspersed elements |
| snRNA | Ribonucleoproteínas nucleares pequenas |
| SRP | Signal recognition particle |
| ssRNA | RNA de fita simples |
| STAR | Spliced Transcripts Alignment to a Reference |
| STAT1 | Signal transducer and activator of transcription 1 |
| STAT2 | Signal transducer and activator of transcription 2 |
| STAT3 | Signal transducer and activator of transcription 3 |
| TLR | Toll-like receptor |
| TMPRSS2 | Protease serina transmembranar 2 |
| TNF- α | Tumor Necrosis Factor alpha |
| TYK2 | Tyrosine kinase 2 |

| | |
|---------------|------------------------------|
| U2AF2 | Fator auxiliar U2 snRNP 2 |
| UMIs | Unique molecular identifier |
| UNG | Uracil N-glycosylase |
| WHO | Organização Mundial da Saúde |
| α -CoV | alpha Coronavírus |
| β -CoV | beta Coronavírus |
| γ -CoV | gamma Coronavírus |
| δ -CoV | delta Coronavírus |

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1. Taxonomia dos Nidovirales..... | 18 |
| Figura 2. Estrutura do virion, do genoma do SARS-CoV-2 e sua proteína spike (S)..... | 20 |
| Figura 3. Ciclo de infecção do SARS-CoV-2..... | 22 |
| Figura 4. Mecanismo de escape immune pelo SARS-CoV-2..... | 24 |
| Figura 5. Resposta atrasada do interferon..... | 25 |
| Figura 6. Processamento do splicing..... | 27 |
| Figura 7. Interferência no splicing pelo SARS-CoV-2..... | 29 |
| Figura 8. Edição de base nucleotídica realizada pela proteína ADAR por processo de desaminação da adenosina para inosina..... | 30 |
| Figura 9. Possíveis regiões editadas pela proteína ADAR e suas consequências para o transcrito editado..... | 31 |
| Figura 10. Transcritos contendo o elemento ALU..... | 33 |
| Figura 11. RNA dupla fita (dsRNA) afeta via MDA5-MAVS-IFN..... | 33 |
| Figura 12. Passos do sequenciamento utilizando a plataforma Illumina..... | 35 |
| Figura 13. fluxo de trabalho e as etapas do pipeline transcriptAnalysis..... | 51 |
| Figura 14. Qualidade média das reads por par de base..... | 57 |
| Figura 15. Gráfico de vulcão com a dispersão do fold change pelo FDR..... | 59 |

| | |
|---|----|
| Figura 16. Gráfico de Venn mostrando a intersecção entre os genes diferencialmente expressos entre os grupos de Covid-19 e dengue..... | 59 |
| Figura 17. 30 termos mais enriquecidos em genes aumentados..... | 61 |
| Figura 18. 30 termos mais enriquecidos entre as interseções de genes diferencialmente expressos de COVID-19 e dengue..... | 62 |
| Figura 19. Contagem de eventos de splicing alternativo (SA)..... | 66 |
| Figura 20. Isoforma do gene CD74..... | 67 |
| Figura 21. Isoforma do gene MX2..... | 68 |
| Figura 22. Gráfico de vulcão com a dispersão do fold change pelo FDR..... | 71 |
| Figura 23. Análise de enriquecimento com enrichR dos monócitos nos estados leve e grave..... | 72 |
| Figura 24. Índice de edição de RNA (ER) no tratamento e controle..... | 74 |
| Figura 25. Significância da sobreposição de regiões editadas nos transcritos anotados..... | 76 |
| Figura 26. Contagem de regiões editadas em elementos ALU..... | 77 |
| Figura 27. Comparação do número de edições de RNA (ER)..... | 78 |
| Figura 28. Comparação do número de edições de RNA (ER)..... | 79 |
| Figura 29. Análise de rede dos termos enriquecidos, em genes editados, feita com Metascape..... | 80 |
| Figura 30. Distribuição dos grupos de edição..... | 80 |
| Figura 31. Comparação dos termos enriquecidos dos genes editados..... | 81 |
| Figura 32. 18 genes com maior número de edições na COVID-19 grave e MAVS..... | 82 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1. Identificação das amostras dispostas no SRA..... | 53 |
| Tabela 2. Desenho experimental dos dados obtidos contendo diferentes características..... | 55 |
| Tabela 3. Genes com expressão aumentada na dengue, mas diminuídos na COVID-19, e enriquecidos no Reactome pathway com relação à via de sinalização de IFN..... | 64 |
| Tabela 4. Termos enriquecidos no MDSig2020 pelos genes com SA..... | 66 |
| Tabela 5. Distribuição de edições para cada grupo..... | 72 |

SUMÁRIO

| | |
|---|-----------|
| 1. INTRODUÇÃO..... | 16 |
| 1.1 COVID-19..... | 16 |
| 1.1.1. Família coronaviridae e variantes do SARS-CoV-2..... | 17 |
| 1.1.2. Estrutura genômica do SARS-CoV-2..... | 18 |
| 1.1.3. Infecção e ciclo viral..... | 21 |
| 1.1.4. Resposta imune e escape viral..... | 22 |
| 1.1.5. Splicing alternativo e perturbação de isoformas pela COVID-19..... | 26 |
| 1.1.6. Edição de RNAs e papel na COVID-19..... | 29 |
| 1.1.6.1. <u>Elementos ALU como alvo da proteína ADAR</u> | 31 |
| 1.1.7. NGS E SUAS APLICAÇÕES..... | 34 |
| 2. OBJETIVOS..... | 37 |
| 2.1. Objetivos gerais..... | 37 |
| 2.2. Objetivos específicos..... | 37 |
| 3. MÉTODOS..... | 38 |
| 3.1. Recursos utilizados..... | 38 |
| 3.2. Obtenção das amostras de transcrito e pré-processamento..... | 38 |
| 3.3. Alinhamento das Reads ao genoma de referência..... | 40 |
| 3.4. Matriz de contagem..... | 41 |
| 3.5. Análise de expressão diferencial..... | 42 |
| 3.6. Processamento de lncRNA..... | 44 |
| 3.7. Quantificação e detecção de splicing alternativo..... | 44 |
| 3.7.1. rMATS..... | 44 |
| 3.8. Detecção e análise diferencial de edição de RNA..... | 46 |
| 3.9. Análise funcional..... | 47 |
| 3.9.1. ClusterProfiler..... | 47 |
| 3.9.2. enrichR..... | 48 |
| 3.9.3. Metascape..... | 48 |
| 3.10. Análise de via..... | 49 |
| 3.11. Pipeline para análise de transcrito..... | 49 |
| 3.12. Visualização dos resultados..... | 49 |
| 4. RESULTADOS E DISCUSSÃO..... | 51 |
| 4.1. Workflow e pipeline transcript analysis..... | 51 |
| 4.2. Caracterização das amostras utilizadas..... | 52 |
| 4.3. PBMC: COMPARANDO VIROSES..... | 58 |
| 4.3.1. Expressão e análise diferencial..... | 58 |

| | | |
|----------|---|-----------|
| 4.3.2 | Análise funcional..... | 60 |
| 4.3.3 | Splicing alternativo..... | 64 |
| 4.4 | Monócitos..... | 69 |
| 4.4.1 | Expressão e análise funcional..... | 69 |
| 4.4.2 | Edição de RNA..... | 72 |
| 4.4.2.1 | Edições gerais nas amostras..... | 73 |
| 4.4.2.2 | Regiões editadas no transcrito..... | 74 |
| 4.4.2.3 | Edições de RNA diferenciais em monócitos..... | 76 |
| 4.4.2.4 | Análise funcional de genes de transcritos editados..... | 78 |
| 4.4.2.5 | Análise dos genes que possuem maior razão de edição de RNA..... | 80 |
| 5 | CONCLUSÕES E PERSPECTIVAS..... | 83 |
| 6 | REFERÊNCIAS..... | 85 |

AGRADECIMENTOS

Agradeço primeiramente à minha família, em especial minha esposa que me deu força, apoio e coragem para chegar até o final.

Aos meus orientadores, professor Marcelo e professor Waldeyr que sempre tiveram paciência para me auxiliar nos muitos momentos em que precisei, sempre com muita calma e boa vontade.

À Universidade de Brasília e seus excelentes professores, que com muito empenho e maestria me deram as bases para que esse trabalho fosse criado.

Ao programa de pós-graduação em Biologia Molecular, que com grande excelência forma mestres e doutores, há mais de 50 anos.

Agradeço, por fim, à agência de fomento CAPES que com muita luta torna possível o desenvolvimento de pesquisas no país.

RESUMO

O SARS-CoV-2, vírus causador da COVID-19 foi primeiramente descrito em Wuhan, China em 2019, de onde rapidamente se espalhou pelo mundo, causando a pandemia anunciada em março de 2020. Diversos estudos encontraram mecanismos virais de escape imunológico que perturbam o desenvolvimento da resposta gerada pelo sistema imune inato e atrasam a resposta do sistema imune adaptativo. Esses mecanismos incluem a capacidade do vírus de reprimir resposta antiviral causada por interferons (IFNs), e a capacidade de interferir em processos de splicing. Ambos mecanismos são realizados com o auxílio das proteínas virais expressas na célula hospedeira e geram resposta complexa, necessitando análises em larga escala, para elucidar as consequências das interação patógeno-hospedeiro. Neste trabalho, a modulação da expressão gênica, padrões de splicing alternativo (SA) e edição de RNA foram avaliados a partir do transcrito de células imunes disponibilizado em bancos de dados públicos. Análises da expressão de transcritos em células mononucleares do sangue periférico (PBMC) de pacientes com COVID-19, em estágio moderado, mostraram pouca resposta do sistema imune e termos enriquecidos relacionados com a replicação viral. Operações entre conjunto de genes diferencialmente expressos, na COVID-19 e na dengue, mostraram uma resposta inflamatória induzida por NF- κ B, indicando uma possível resposta inflamatória, com diminuída resposta antiviral pela via dos IFNs, na COVID-19, que é prejudicial para o hospedeiro. Assinatura relacionada à resposta ao IFN estava enriquecida em genes com SA diferencial. Os genes CD74 e MX2, relacionados com as vias de IFN do tipo 1 (IFN-1), e com ação antiviral, mostraram splicing alternativo diferencial, com perda de função proteica. Análise de expressão de monócitos circulantes de pacientes infectados, em estado severo, mostrou diminuição da resposta de IFN-1 em relação aos casos leves, correlação com estado de hipóxia, aumento da resposta inflamatória via NF- κ B e padrões de edição de RNA que indicam diminuição da resposta de IFNs e podem ajudar a compreender a desregulação e infecção desse tipo celular.

Palavras-chave: SARS-CoV-2, expressão gênica diferencial, SA, edição de RNA.

ABSTRACT

SARS-CoV-2, the virus causing COVID-19 was first described in Wuhan, China in 2019, from where it quickly spread worldwide, causing the pandemic announced in March 2020. Several studies have found viral mechanisms of immune escape that disrupt the development of the response generated by the innate immune system and delay the response of the adaptive immune system. These mechanisms include the ability of the virus to repress antiviral response caused by interferons (IFNs), and the ability to interfere with splicing processes. Both mechanisms are accomplished with the aid of viral proteins expressed in the host cell and generate a complex response, requiring large-scale analysis to elucidate the consequences of pathogen-host interactions. In this work, modulation of gene expression, alternative splicing patterns and RNA editing were evaluated from the immune cell transcriptome available in public databases. Transcript expression analyses in peripheral blood mononuclear cells (PBMC) from patients with moderate-stage COVID-19 showed little immune system response and enriched terms related to viral replication, operations between set of differentially expressed genes in COVID-19 and dengue showed similar NF-kB-induced inflammatory response, indicating a possible inflammatory response, with diminished antiviral, host-damaging response, in COVID-19. Signature related to IFN response was enriched in genes with differential alternative splicing. CD74 and MX2 genes related to IFN-1 pathways showed differential alternative splicing with loss of function. Expression analysis of circulating monocytes from infected patients in severe state showed decreased IFN-1 response compared to mild cases, correlation with hypoxic state, increased inflammatory response via NF-kB and RNA editing patterns that indicate decreased IFNs response and may help to understand downregulation and infection of this cell type.

Keywords: SARS-CoV-2, differential gene expression, alternative splicing, RNA editing.

1. INTRODUÇÃO

1.1 COVID-19

O agente etiológico da doença do coronavírus 2019 (COVID-19), foi nomeado SARS-CoV-2 (do inglês, *severe acute respiratory syndrome coronavirus*) como uma nova espécie de Coronavírus, em fevereiro de 2020 pela WHO (*organização mundial da saúde*), devido às suas características filogenéticas semelhantes a outras espécies pertencentes a essa família de vírus (Lauxmann et al., 2020). Houve rápida disseminação após as primeiras infecções que ocorreram em um mercado de comida em Wuhan, na China, em dezembro de 2019, de onde o vírus se espalhou para diversos países em poucos meses, resultando em mais de 4 milhões de infecções e mais de 279 mil mortos em 187 países até maio de 2020, revelando rápido contágio e alta transmissibilidade (Lauxmann et al., 2020).

A família *coronaviridae* infecta uma ampla variedade de mamíferos e pássaros e já foi responsável por duas epidemias no passado: SARS (do inglês, *severe acute respiratory syndrome*) na província de Guangdong na China e MERS (do inglês, *Middle East respiratory syndrome*) em países do oriente médio. Desde então, o surgimento de novas epidemias passou a ser de preocupação global, estando evidenciado a capacidade dos vírus da família coronavírus de atravessar a barreira de espécies, sendo transferidos de espécies que são reservatórios virais, para a população humana. Morcegos como reservatórios da família coronavírus tornam maior as chances de uma nova pandemia, visto o potencial de infecção humana (Lauxmann et al., 2020). MERS teve origem em morcegos, sendo os camelos dromedários o hospedeiro do reservatório que desencadeou a infecção em humanos, tendo 99.2-99.8% de similaridade com o MERS-CoV, enquanto os civetas de palma mascaradas ou musang e cães-guaxinins surgiram como possíveis intermediários da infecção pelo Sars-CoV (Alsafi et al., 2022). O vírus do SARS-CoV-2 provavelmente também teve o morcego como origem, tendo 96.3% de similaridade com a linhagem de morcego CoV-RaTG13. Martas e Pangolins foram os prováveis reservatórios intermediários, antes da transmissão para humanos (Alsafi et al., 2022).

A resposta global à pandemia de COVID-19 incluiu medidas de saúde pública, como lockdown, quarentenas e testagem em massa, bem como o desenvolvimento de tratamentos e

vacinas contra a COVID-19. Embora tenham surgido soluções para o problema, a pandemia continua a afetar a saúde, a economia e a sociedade em todo o mundo em 2023. A transmissão do vírus ocorre principalmente através de gotículas respiratórias, sendo importante o uso de máscaras, lavagem frequente das mãos e distanciamento social para prevenir a propagação da doença (World Health Organization [WHO], 2022).

1.1.1. Família *coronaviridae* e variantes do SARS-CoV-2

A família *coronaviridae* faz parte da ordem *Nidovirales*, vírus que possuem genoma de RNA fita simples, de polaridade positiva (ssRNA+), com aproximadamente 29.9 kb (V'kovski et al., 2021). Sendo a subfamília *Orthocoronavirinae*, pertencente à família *coronaviridae*, dividida em coronavírus alpha (α -CoV), beta (β -CoV), gamma (γ -CoV), e delta (δ -CoV) de acordo com o ICTV (do inglês, *International Committee on Taxonomy of Viruses*), em 2019 (Figura 1). Os gêneros α -CoV e β -CoV são capazes de infectar mamíferos, e análises evolutivas mostram seu surgimento a partir de morcegos e roedores, enquanto que γ -CoV e δ -CoV surgiram de pássaros (Saberian et al., 2022). Os β -CoV são capazes de infectar humanos e são divididos nas linhagens A, B, C e D. A linhagem B abarca as linhagens do SARS-CoV e SARS-CoV-2, enquanto que o MERS-CoV faz parte da linhagem C. Por mais que a origem do vírus causador da COVID-19 ainda não seja completamente esclarecida, seu genoma completo apresenta 96% de similaridade com o BatCoV-RaTG13, vírus que infecta morcegos, indicando sua possível origem (Saberian et al., 2022).

As variantes genéticas do SARS-CoV-2 são classificadas de acordo com seu fenótipo, pela WHO e CDC (do inglês, *Centers for Disease Control and Prevention*). Até 31 de janeiro de 2023 mais de 14.8 milhões de genomas do vírus foram depositados no banco de dados EpiCoV do GISAID, contribuindo para a vigilância das variantes circulantes. As variantes Alfa (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2) e Omicron (B.1.1.529), mostraram aumento de transmissibilidade, diferentes interações no sistema imune, incluindo a dificuldade no reconhecimento pela resposta humoral e aumento da patogenicidade (Saberian et al., 2022). A variante Omicron foi reportada pela primeira vez na África do sul em 2021 e apresentou maior número de reinfecções em relação à variante Delta (fold ~5.4), possuindo R_0 (do inglês “*basic reproductive number*”), medida que indica o número médio de novas

infecções por uma pessoa em uma população *naive* e estima o potencial infeccioso do agente causador da pandemia, de 2.5 para a cepa original, 7 para a Delta e 10 para a Omicron, desde fevereiro, tornou-se a variante dominante (Liu et al., 2022; Zmasek et al., 2022). Ainda para a Omicron, existem mais de 50 mutações em relação à cepa original de Wuhan, sendo cerca de 30 delas localizadas na ORF que transcreve a proteína Spike (Liu et al., 2022).

Portanto, a família Coronavírus possui potencial para causar novas pandemias, com alto impacto na saúde e economia, sendo necessário o desenvolvimento de técnicas para executar ações públicas e fazer a correta vigilância desse grupo de vírus. Entender os mecanismos de ação no sistema imune se torna igualmente necessário, visto que diversos mecanismos relatados atuam de forma similar entre o SARS-CoV-2, MERS e SARS e serão abordados adiante.

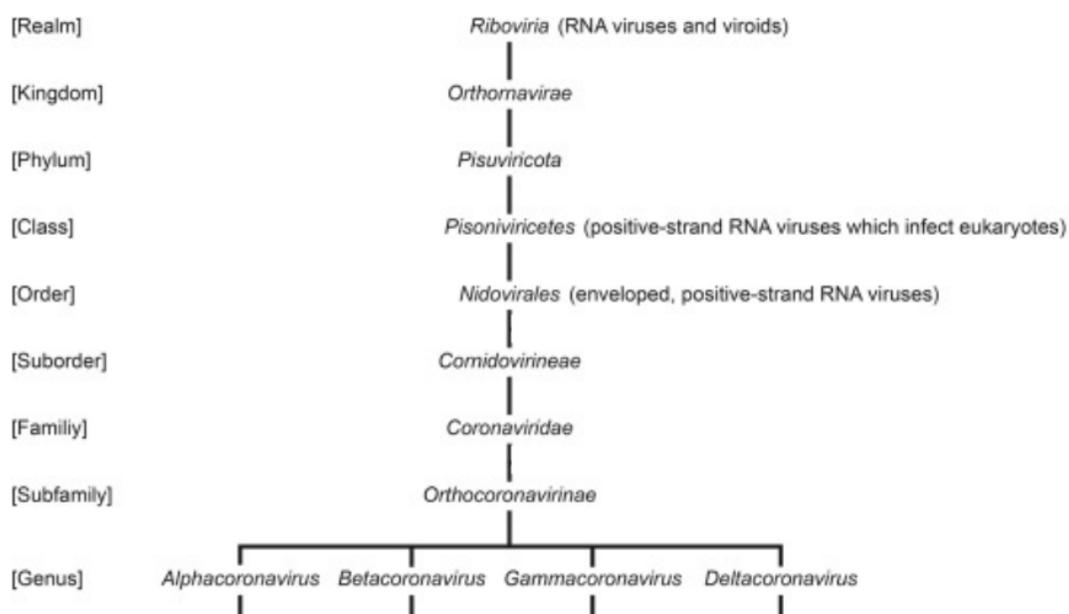


Figura 1. Taxonomia dos Nidovirales.

Fonte: Zmasek et al. Genomic evolution of the Coronaviridae family. *Virology*, 570, 123–133, 2022; Adaptado

1.1.2. Estrutura genômica do SARS-CoV-2

O genoma do SARS-CoV-2 consiste em cinco principais fases abertas de leitura (ORFs), sendo essas as ORF1a, ORF1ab e as quatro proteínas estruturais que constituem o

vírião, sendo elas a proteína do nucleocapsídeo (N) (419 aa), que forma o nucleocapsídeo para acomodar o genoma viral, coberta pelas proteínas do envelope (E) (75 aa), de membrana (M) (222 aa) e a proteína S (1273 aa). As quantidades de aminoácidos indicadas são referentes à cepa original de Wuhan (NCBI accession number NC 045512.2) (Troyano-Hernández et al., 2020). A proteína S é a responsável pelo reconhecimento do receptor ACE2 (angiotensin converting enzyme 2) para entrada do vírião na célula, sendo formada por homotrímeros que se projetam da superfície viral, contendo duas subunidades funcionais S1 e S2 (Figura 2). A proteína S permite a ancoragem com a célula a ser infectada, a proteína N empacota o genoma viral e as proteínas E, M e S formam o envelope viral (Troyano-Hernández et al., 2020).

A proteína S é uma glicoproteína trimérica que possui duas subunidades, a subunidade S1 é formada pelo domínio N-terminal e o RBD (do inglês, *receptor binding domain*), o RBD é o domínio responsável pela ligação ao receptor do hospedeiro, diferentes coronavírus reconhecem diferentes receptores do hospedeiro, ambos S1 de Sars-Cov e SARS-CoV-2 reconhecem a ACE2. A subunidade S2 é constituída do peptídeo de fusão, região HR1 (do inglês, *heptad repeat 1*), a hélice central, o domínio conector HR2 (do inglês, *heptad repeat 2*), o domínio transmembrana e a região N-terminal citoplasmática, sendo o último o responsável pela fusão do capsídeo viral com a membrana do hospedeiro a partir da clivagem em uma região conservada entre os sítios S1 e S2, por proteases do hospedeiro, como TMPRSS2. A clivagem ativa a proteína S, com uma mudança conformacional irreversível, levando ao estado aberto da proteína S. A partir da clivagem da ligação covalente entre S1 e S2, mas com formação posterior de ligações não covalentes entre as duas subunidades, origina-se nova conformação da proteína em estado aberto, necessário para a fusão com a membrana e entrada do vírus (Peacock et al., 2021).

Além das proteínas estruturais, o vírus possui 16 proteínas não estruturais (NSP1-16) formadas pela clivagem das poliproteínas pp1a e pp1b, que são originárias do ORF1a/b. As NSPs são necessárias para o ciclo viral, formam o complexo replicase transcriptase e estão relacionadas com interações patógeno-hospedeiro. Diversas dessas proteínas interagem com RNAs e proteínas do hospedeiro e podem suprimir resposta imune. Por exemplo, a NSP1 inibe a tradução de proteínas pelo hospedeiro, sem inibir a síntese de proteínas virais. As NSP8 e NSP9 interferem na secreção de proteínas ao se ligarem à partícula reconhecidora de

signal (PRS). A NSP16 interfere com mecanismos pós-transcricionais e será abordada adiante (Banerjee et al., 2020).

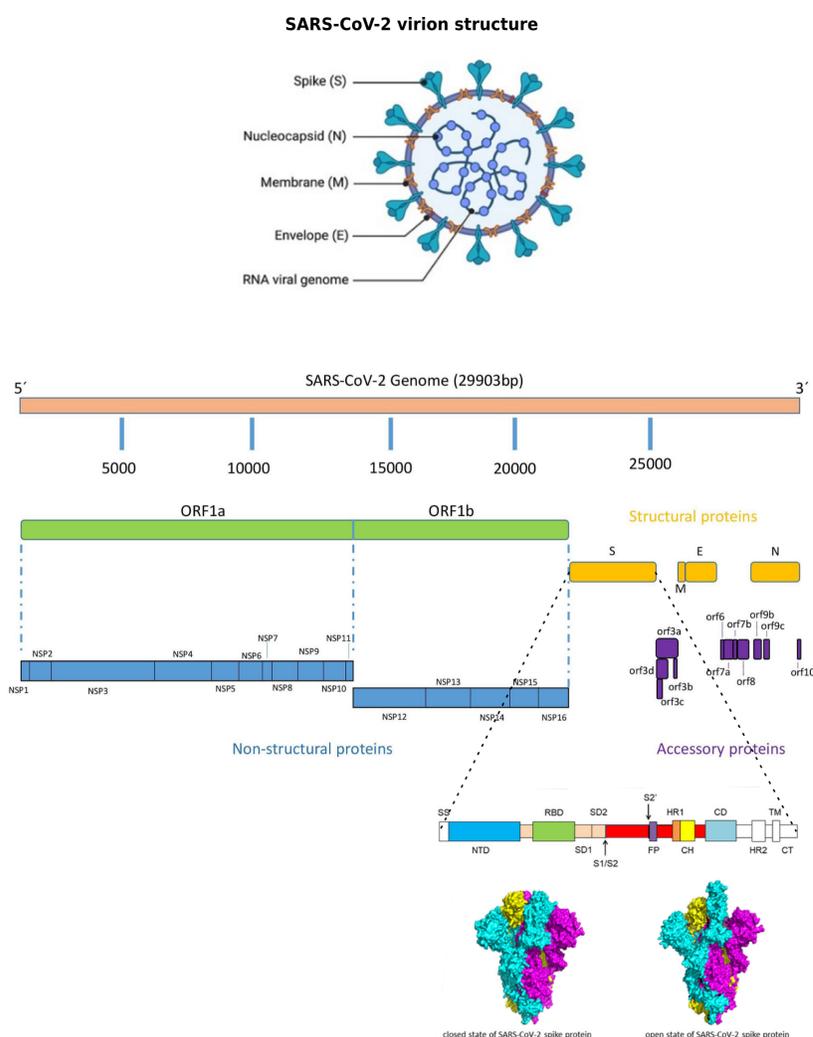


Figura 2. Estrutura do virion, do genoma do SARS-CoV-2 e sua proteína spike (S). As 16 proteínas não estruturais, (NSP1-NSP16), importantes para replicação viral e escape imunológico, são codificadas pelas duas ORFs ORF1a e ORF1b. As proteínas estruturais codificadas são as proteínas S, envelope (E), membrana (M), e nucleocapsídeo (N). A estrutura da proteína S contém diferentes domínios: N-terminal domain (NTD); receptor-binding domain (RBD); subdomain 1 (SD1); subdomain 2 (SD2); sítio S1/S2 de clivagem por protease (S1/S2); sítio S2' de clivagem por protease (S2'); fusion peptide (FP); heptad repeat 1 (HR1); central helix (CH); connector domain (CD); heptad repeat 2 (HR2); transmembrane domain (TM); cytoplasmic tail (CT). Os sítios de clivagem por protease para ativação da proteína, estão indicados por setas.

Fonte: Alanagreh et al (2021), Redondo et al (2021) e Wang et al (2020); Adaptado

1.1.3. Infecção e ciclo viral

O SARS-CoV-2, assim como o Sars-CoV, utiliza o receptor ACE2 (do inglês “*Angiotensin-converting enzyme 2*”), o domínio terminal NH2 é o componente principal da proteína, com seu sítio catalítico disposto extracelular, ademais, faz parte do sistema renina-angiotensina-aldosterona, que é responsável pelo controle da pressão arterial. Ainda que seja encontrado em quase todos os tecidos, sua alta expressão ocorre no íleo e rim, seguido de tecido adiposo, coração, tronco encefálico, pulmão, vasculatura, estômago, fígado, testículos e mucosa nasal e oral. Angiotensina 2 (AngII) é produto de clivagem da angiotensina 1 pela proteína ACE, após sua clivagem passa ao seu estado ativo com ação pró inflamatória e vasoconstritora. A ACE2 pode posteriormente clivar AngII para formação do peptídeo Ang1-7, que possui ação anti-flamatória antagônica da AngII (Bhalla et al., 2021). A infecção da célula pelo SARS-CoV-2 leva à diminuição dos níveis de ACE2 e subsequente processo que estimula depleção de Ang1-7 com ação pró-inflamatória decorrente (Harb et al., 2022; Gong et al., 2023). Os processos decorrentes da diminuição da ACE2 com aumento da pressão arterial, inflamação e fibrose são fatores agravantes em casos de comorbidades.

TMPRSS2 (do inglês “*transmembrane protease serine 2*”) vem sendo evidenciado como a principal protease com ação ativadora da proteína S, via clivagem inicial da região S1/S2, para entrada do vírus na célula. TMPRSS2 é substancialmente expressa em células epiteliais humanas, incluindo vias aéreas superiores, brônquios e pulmões (Gong et al., 2023). Sem a clivagem da proteína S o vírus passa a depender de entrada dependente de baixo pH, através de endossomos tardios e catepsinas. Durante o processo de acoplamento viral a proteína IFITM2 é um possível inibidor ao impedir esse processo, principalmente via endossomos tardios (Winstone et al., 2021). Após fusão com membrana, o genoma viral de fita simples positiva é liberada no citosol da célula infectada, onde utiliza maquinaria de tradução do hospedeiro para traduzir as poliproteínas pp1a e pp1ab, diferencialmente produzidas por sítio de frameshift localizado entre as ORF1a e ORF1b. As lipoproteínas são processadas para formação de peptídeos menores e formam complexo de replicação viral para conversão da fita viral positiva em negativa. A fita negativa codifica todos os transcritos que geram as proteínas estruturais e NSPs a partir de transcrição descontínua pela RNA polimerase, contendo o sítio iniciador da transcrição 5’ líder em cada fragmento, seguida de

tradução, e pode, além disso, a fita senso negativa ser novamente convertida para a fita positiva, que será incorporada em um novo virion. As proteínas estruturais são traduzidas no retículo endoplasmático rugoso e a progênie viral é formada pela via de secreção através do complexo de Golgi (figura 3) (V'kovski et al., 2021; Martina et al., 2020).

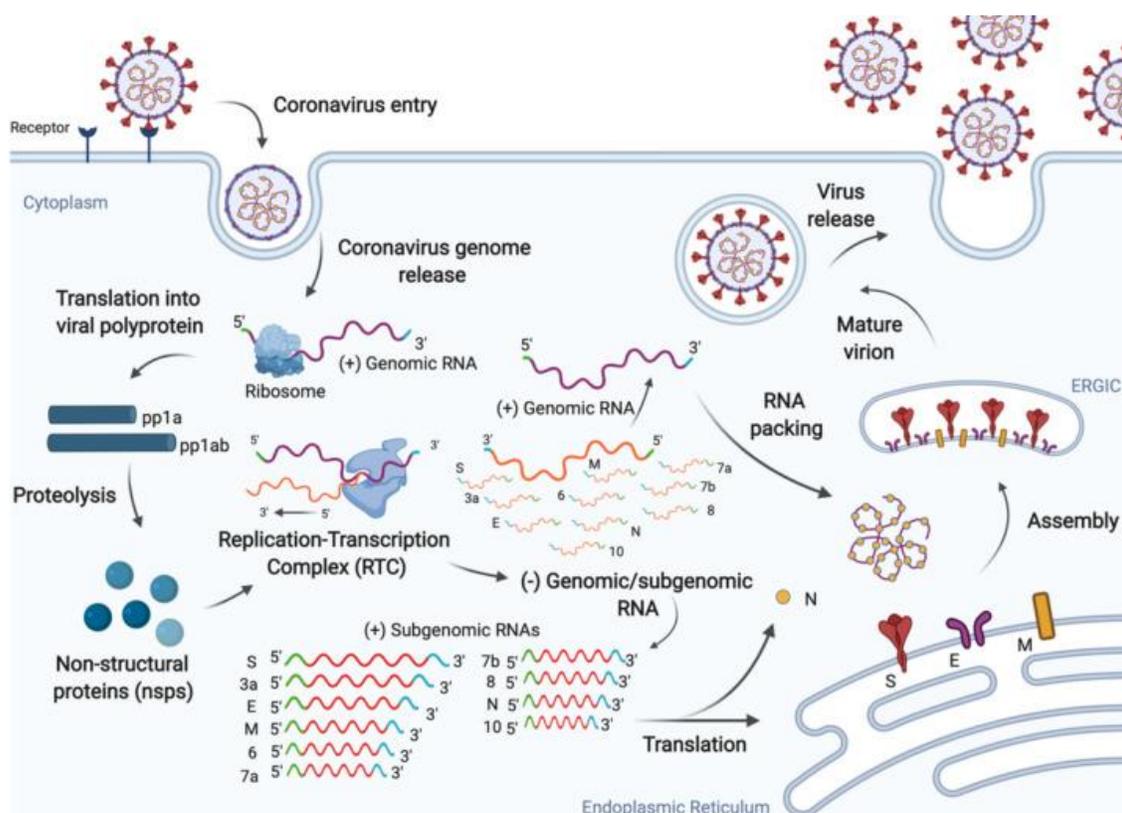


Figura 3. Ciclo de infecção do SARS-CoV-2. O vírus entra na célula através do reconhecimento da proteína ACE2 e libera seu genoma para formação das proteínas que irão participar da formação de novos virions, completando seu ciclo de infecção.

Fonte: Martina et al. Amilorides inhibit SARS-CoV-2 replication in vitro by targeting RNA structures. *bioRxiv*: the preprint server for biology. 2020

1.1.4. Resposta imune e escape viral

O paciente infectado pelo SARS-CoV-2 pode não apresentar sintomas, em quadro assintomático, ou em casos leves, apresentar sintomas que incluem febre, tosse, dor de garganta, mal-estar, dor de cabeça, dor muscular, náusea, vômito, diarreia, anosmia ou disgeusia, mas sem falta de ar, casos moderados apresentam evidência radiológica de doença

do trato respiratório, como pneumonia, mas com saturação de oxigênio acima de 94%, enquanto que casos graves apresentam infiltração pulmonar e baixa saturação de oxigênio, casos mais críticos da doença se apresentam com insuficiência respiratória aguda, choque séptico e possível disfunção de múltiplos órgãos, podendo desenvolver a síndrome do desconforto respiratório agudo (SDRA), configurada por edema pulmonar não cardiogênico e síndrome da inflamação pulmonar difusa, que tende a ocorrer aproximadamente uma semana após o início dos sintomas, sendo uma das principais causas de letalidade por COVID-19, iniciada pela inflamação sistêmica e tempestade de citocinas, estado em que há elevação desenfreada dos níveis de citocinas pró-inflamatórias. (Sofi et al., 2020; Cascella et al., 2022; World Health Organization [WHO], 2022; Montazersaheb et al., 2022).

Sinais de reconhecimento viral desencadeiam respostas ao vírus. As principais formas de reconhecimento são realizadas pelo reconhecimento de PAMPS (do inglês, *pathogen-associated molecular pattern*) pelos PRR (do inglês, *pattern recognition receptor*), receptores do tipo Toll-like (TLR) que podem reconhecer fragmentos da superfície viral, como o TLR4 que reconhecem a glicoproteína S (Zhao et al., 2021), ou os TLR TLR3, TLR7, TLR8, que reconhecem RNA viral, mas enquanto TLR3 reconhece dsRNA, TLR7 e TLR8 reconhecem ssRNA, sendo TLR7 o principal detector do ssRNA do SARS-CoV-2. Sensores citoplasmáticos do RNA como RIG-I e MDA5 fazem reconhecimento do RNA viral. As vias de sinalização estimulam a produção de citocinas pró-inflamatórias típicas na COVID-19 como IL-6, TNF- α e IL-1 via NF- κ B, enquanto que a liberação de IFN-1 e IFN-3 são estimulados pelos fatores reguladores de interferon (IRF-*Interferon Regulatory Factor*), como IRF3 e IRF7, que estimulam a expressão de IFNs no núcleo (Tolomeo; Cavalli; Cascio., 2022).

A partir do reconhecimento da infecção, a principal via de ação antiviral é por meio da liberação do IFN que podem ser do tipo I (IFN- α , IFN- β , IFN- ϵ , IFN- κ , IFN- ω) ou do tipo III (IFN- γ). Os IFNs são liberados de forma autócrina e parácrina de forma a sinalizar células próximas, desencadeando resposta antiviral através dos receptor de IFN do tipo I e III: IFNAR e IFNLR respectivamente (Park et al., 2020). A resposta ao IFN é mediada via Jak-STAT, pela STAT1 e STAT2 que transduzem o sinal e ativam a transcrição de genes estimulados por interferon (*ISG-interferon-stimulated gene*), que agem inibindo o ciclo viral (figura 4) (Tolomeo; Cavalli; Cascio., 2022).

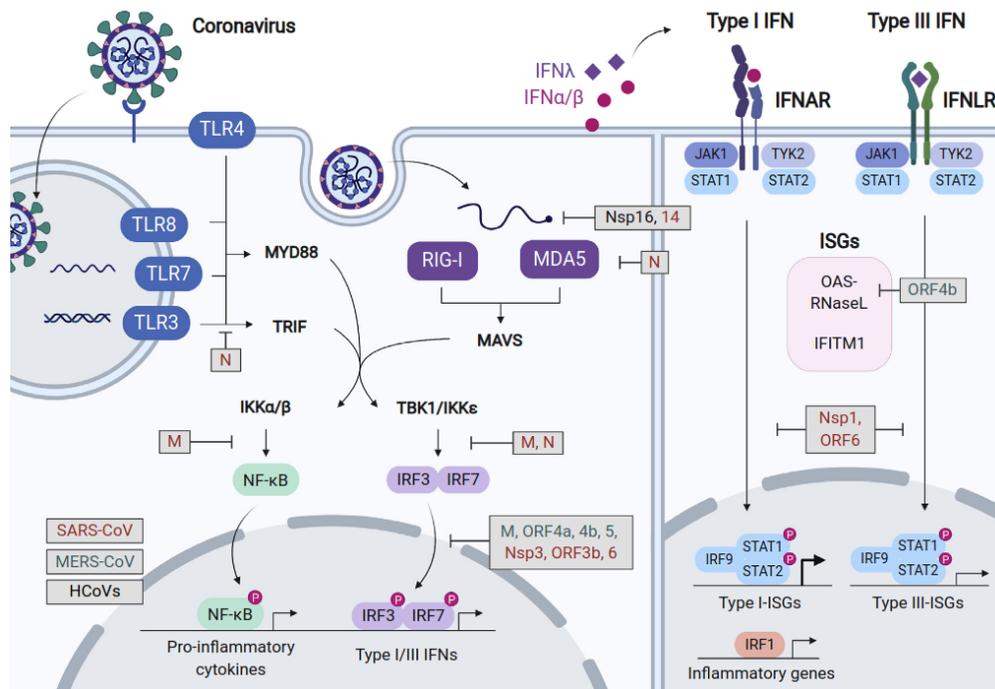


Figura 4. Mecanismo de escape immune pelo SARS-CoV-2. Mecanismos de defesa antivirais por via de reconhecimento do vírus pelos Toll-like receptors (TLR) e Retinoic acid-inducible gene I (*RIG-I*)-like receptors (RLRs) liberam interferons (IFN) e citocinas pró-inflamatórias. As proteínas virais agem em várias etapas da via e interferem na resposta imune adequada.

Fonte: Park; Iwasaki. Type I and Type III Interferons - Induction, Signaling, Evasion, and Application to Combat COVID-19. *Cell host & microbe*, 27(6), 870–878. 2020.

Os casos mais graves de COVID-19 estão relacionados com a alta produção de citocinas pró-inflamatórias, como IL-6, IL-2, IL-7, IL-12, IFN- γ e TNF- α . A principal hipótese é a de que a tempestade de citocinas advém por mecanismos infecciosos e de escape do SI pelo SARS-CoV-2 (Montazersaheb et al., 2022). Primeiramente, a replicação do vírus leva ao rompimento da célula por piroptose, uma forma de morte celular programada, altamente inflamatória, que pode comprometer o correto funcionamento de macrófagos e linfócitos, levando-os à liberação desenfreada das citocinas pró-inflamatórias em processo de retroalimentação positiva. Ainda, existem diversos mecanismos virais pelos quais interagem com a célula hospedeira e perturbam o sistema imune, modulam vias do IFN e interferem em mecanismo de SA do hospedeiro, em que esse último será abordado no item 1.1.4.

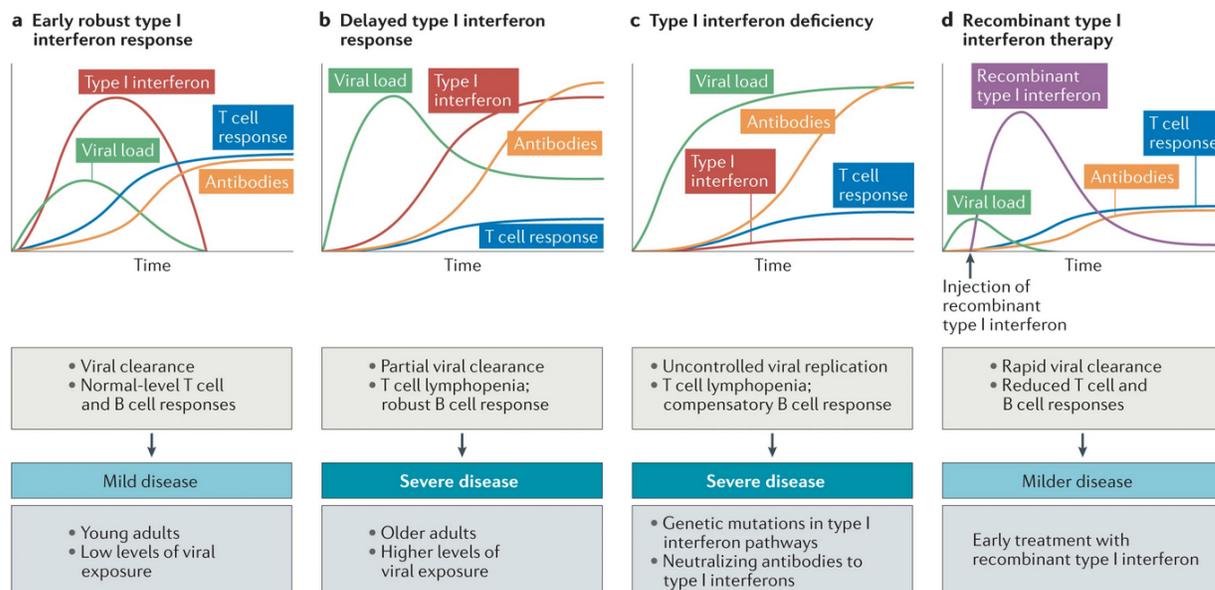


Figura 5. Resposta atrasada do interferon. Resposta robusta do IFN do tipo I nos estágios iniciais da COVID-19 levam à resposta imune adequada com controle da carga viral. Resposta atrasada ou restringida acarretam em descontrole da carga viral, resposta imune desregulada, com possibilidade de quadros clínicos mais graves.

Fonte: Carvalho; Krammer; Iwasaki. The first 12 months of COVID-19: a timeline of immunological insights. *Nat Rev Immunol* 21, 245–256. 2021.

O IFN do tipo I mostrou papel fundamental para prevenir casos graves na COVID-19. Resposta ao IFN do tipo I nos estágios iniciais estão mais relacionados a casos leves, com resposta normal do sistema imune (SI) e recuperação do paciente. De outro modo, atraso na resposta ou mutações em genes da via do IFN do tipo I, que dificultem sua resposta, acarretam em resposta imune aberrante, desregulação dos monócitos e comprometimento da resposta imune adaptativa, podendo levar ao agravamento da doença (Figura 5) (Carvalho; Krammer; Iwasaki., 2021). Notoriamente, ambas NSPs e proteínas estruturais virais da família Coronavírus lesam vias do IFN, desde a sinalização até a exportação das proteínas, via interações de biopolímeros virais-hospedeiro: RNA-RNA (Zhang et al., 2022), proteína-RNA (Banerjee et al., 2020) e proteína-proteína (PPI-*protein-protein interactions*) (Zhou et al., 2023). Algumas das interações PPI, que interferem nas vias de IFN estão mostradas na figura 4.

1.1.5. Splicing alternativo e perturbação de isoformas pela COVID-19

O splicing ou processamento de introns é um mecanismo no qual se realiza a retirada de introns e junção dos éxons do pré-mRNA para formação do RNA maduro que será utilizado para codificação de determinada proteína. Durante a evolução, houve surgimento do processo de SA que leva a geração de diferentes transcritos a partir de um mesmo gene, ou controlar o nível de expressão gênica a partir de geração de transcritos produtivos ou não produtivos, dessa forma, aumentando a complexidade dos organismos pelo aumento da diversidade proteômica e controle da eficiência de resposta aos estímulos. A diversidade proteômica pelo SA pode ser observada no genoma humano, mais de 250 mil transcritos estão anotados pelo projeto GENCODE, enquanto que apenas cerca de 20 mil genes codificadores de proteínas, com cerca de 95% destes genes apresentando processamento de mais de um tipo de transcrito que codifica proteína (Park et al., 2018; GENCODE, 2023). O splicing pode ocorrer por autocatálise feita pelo próprio RNA ou com auxílio do spliceossomo, complexo formado por proteínas e RNAs (Bessa et al., 2020) e consiste de duas reações de transesterificação consecutivas definidas por elementos de ação em cis presentes no pré-mRNA e que definem as junções exon-intron, tendo evoluído de processos autocatalíticos da sequência feita por introns do tipo II, para sequências que passaram a utilizar elementos atuantes em trans, que catalizam a reação de splicing. Esse elemento, denominado spliceossomo, é formado pelas ribonucleoproteínas nucleares pequenas (snRNP): U1, U2, U3, U4, U5, U6. Em um primeiro momento, U1 se liga na região 5' e, em seguida, o fator de splicing 1 (SF1) se liga na região de ramificação (BPS-*branch point*), o fator auxiliar U2 snRNP 2 (U2AF2) se liga na poli-pirimidina à jusante do BPS e U2AF1 reconhece o dinucleotídeo AG na região de splice 3'. A ocupação dessas três regiões induz a associação do snRNP U2 no BPS, que é posteriormente estabilizado pelo componente SF3B1. Conseqüentemente, o reconhecimento do intron dirige o recrutamento e formação do trímero U4/U6/U5 com a subsequente formação do complexo, que catalisa a reação de splicing (Figura 6A) (Borao; Ayté; Hümmer., 2021; Bessa et al., 2020).

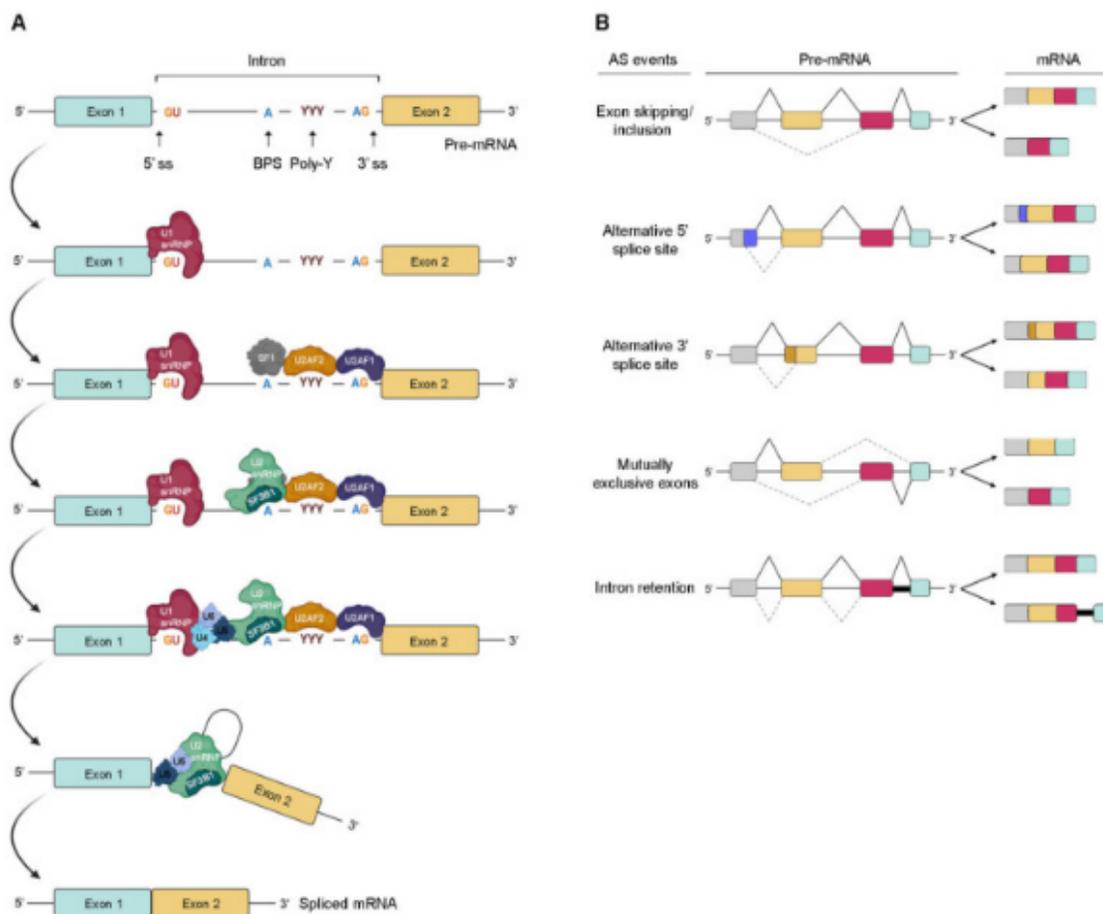


Figura 6. Processamento do splicing. (A) Formação do spliceossomo e remoção dos introns. (B) Estruturas mais comuns de transcritos contendo splicing alternativo (SA): Exon skipping (ES), Alternative 5' splice site (A5SS), Alternative 3' splice site (A3SS), Mutually exclusive exons (MXE) e intron retention (IR).

Fonte: Bessa et al. Alternative Splicing: Expanding the Landscape of Cancer Biomarkers and Therapeutics. *International journal of molecular sciences*, 21(23), 9032. 2020. Adaptado.

Cinco formas principais de SA, que possuem modificações na composição dos éxons e introns em relação à sua estrutura canônica, foram descritas na literatura. A exclusão de exon é a forma predominante em mamíferos e representa cerca de 40% dos transcritos alternativos em humanos, origina-se por exclusão de um ou mais exons do transcrito final. Sítios 3' ou 5' alternativos de splicing resultam em região exônica inicial ou terminal variada. Em éxons mutuamente exclusivos a presença de um dos exons exclui o outro e vice versa. Retenção de introns ocorre quando um intron é mantido no transcrito final (Figura 6B) (Borao et al., 2021; Wang et al., 2015).

No curso de infecções, diversos vírus comprometem os mecanismos de splicing através do sequestro e manipulação das vias do spliceossomo, para facilitar sua replicação ou evadir o sistema imune. Ambos os vírus da Influenza A (IAV-*Influenza A virus*) e o da COVID-19 causam doença respiratória, possuem um mecanismo similar de propagação, e o IAV também tem mecanismo de evasão da resposta de IFN do tipo 1 com comprometimento da sua produção em células epiteliais e células dendríticas. A proteína NS1 de IAV mostrou interagir com U2 e U6 *in vitro* e interferir no splicing (Wang; Krug., 1998). A dengue também possui similares sintomas com a COVID-19 em casos mais graves como a tempestade de citocinas, atraso da resposta do IFN do tipo 1 e atraso da resposta do SI adaptativo (Malavige; Jeewandara; Ogg., 2022), sua proteína NSP5 se liga aos componentes da U5 e modula SA e modula fatores antivirais (De Maio et al., 2016). Dentre as NSPs do SARS-CoV-2, a NSP16 mostrou se ligar a ambos U1 e U2 e, do mesmo modo, interferir no SA, mas em células Calu3 infectadas. Essa interferência no splicing pela NSP16 mostrou diminuir a resposta de IFN (Banerjee et al., 2020) (figura 7). Outros estudos ômicos já analisaram o perfil de SA durante a infecção pelo SARS-CoV-2. Amostras *post mortem* provindas de pulmão de pacientes, que foram infectados, indicaram modificação aberrante do SA, no transcrito de diferentes tipos celulares, que perturbou a tradução de proteínas importantes para a defesa do hospedeiro. Esses achados sugerem que o SA anômalo esteja relacionado a sintomas clínicos, como a inflamação sistêmica na COVID-19 (Wang et al., 2022)

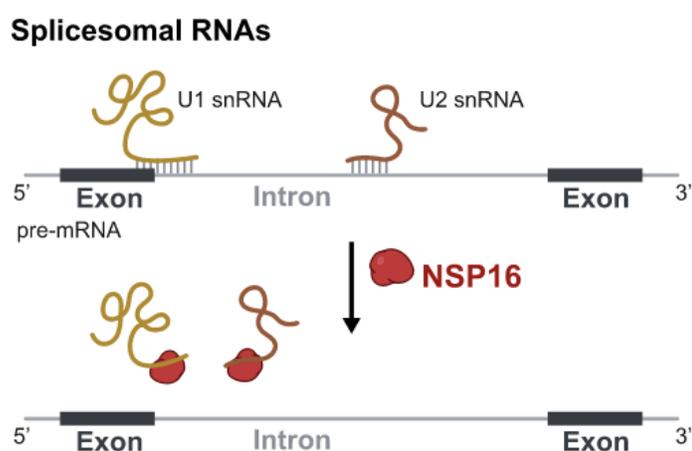


Figura 7. Interferência no splicing pelo SARS-CoV-2. A NSP16 pode afetar reconhecimento de mRNA pelas snRNA U1/U2 e prejudicar os processos normais de splicing.

Fonte: Banerjee et al. SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell*, 183(5), 1325–1339.e21. 2020. Adaptado.

1.1.6. Edição de RNAs e papel na COVID-19

A edição do RNA é uma modificação pós transcricional que consiste na modificação de uma base nucleotídica no RNA e pode aumentar a variabilidade de transcritos e de proteínas quando em regiões codificadoras de proteínas, assim como pode intermediar os níveis de expressão do transcrito por diferentes mecanismos, ou também inviabilizar RNA exógeno ao modificar seu RNA, podendo, então, funcionar como mecanismo de defesa da célula. Em mamíferos, essa edição de bases pode ocorrer por duas famílias distintas de proteínas: A APOBEC (do inglês, *Apolipoprotein B mRNA Editing Catalytic Polypeptide-like*) e pela ADAR (do inglês, *adenosine deaminase acting on RNA*). Algumas dessas proteínas são capazes de modificar DNA, como APOBECA3A (A3A) e A3G, podendo criar SNPs (do inglês, *single nucleotide polymorphism*) (Pecori et al., 2022). Edição de RNA (ER) diferencial está presente em doenças inflamatórias (Li et al., 2022), desordens neurológicas do sistema nervoso central (Cuddleston et al., 2022), câncer (Kurkowiak et al., 2021; Wang et al., 2021) e tem desdobramentos nessas e em outras condições (Song et al., 2022), sendo, portanto, de importância para a saúde humana.

APOBECs convertem citosina em timina (C-para-T), no DNA, e citosina para uracila (C-para-U), no RNA, por processo de desaminação, sendo capazes de restringir o ciclo viral, o que foi verificado primeiramente em HIV, onde leva a um processo de hipermutação catastrófica (Erro-catástrofe). Parte do cDNA viral convertido em T pode então ser reconhecido pela proteína UNG (do inglês, *uracil N-glycosylase*), que inicia via de reparo por excisão de bases e remove base nitrogenada, esse processo pode levar o genoma viral ao processo de degradação (Pecori et al., 2022; Siriwardena et al., 2016). Além de outros, DNAs e RNAs virais editados pela APOBEC, como o vírus da hepatite B (HBV) e o vírus do papiloma humano (HPV), a proteína também mostrou ser capaz de editar RNA de SARS-CoV-2 e outros como HTLV-1 (human T cell leukaemia virus type 1) e rubéola (Pecori et al., 2022; Riva et al., 2021; Venkatesan et al., 2018).

A edição mais prevalente encontrada em RNA endógeno humano é do tipo adenosina para inosina (A-para-I), convertida pelas proteínas da família ADAR, por processo de desaminação de base (figura 8). Uma característica dessa reação é a necessidade de reconhecimento de dsRNA (RNA dupla fita) para ligação e edição (Tomaselli et al., 2013). Apenas duas proteínas dessa família foram mostradas ativas em mamíferos, a proteína ADAR1 e ADAR2. ADAR3 não possui atividade de edição, porém estudos sugerem que essa proteína compete por regiões alvo de edições e coíbe edições em sítios dsRNA, por exemplo, foi mostrada impedir edição em MAVS, aumentando a expressão do transcrito em células de glioblastoma (Raghava et al., 2022). ADAR1 possui duas isoformas, ADAR1p110 e ADAR1p150, possuindo a p150 domínio $Z\alpha$ adicional contendo sequência de exportação do núcleo, que permite que essa isoforma edite sequências no citoplasma e está associada com a resposta ao IFN (Nakahama; Kawahara., 2021).

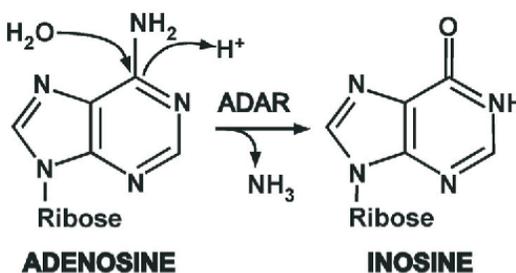


Figura 8. Edição de base nucleotídica realizada pela proteína ADAR por processo de desaminação da adenosina para inosina.

Fonte: Jain; Jantsch; Licht, K. The Editor's I on Disease Development. *Trends in genetics : TIG*, 35(12), 903–913. 2019. Adaptado.

As edições A-para-I são reportadas em diferentes regiões do transcrito, com diferentes implicações em ambos genes codificadores e não codificadores de proteínas (Figura 9). A edição frequentemente leva à abertura da fita dsRNA para ssRNA, tendo implicações na estrutura secundária e na estabilidade do transcrito. Edições em regiões de íntrons podem modular as isoformas do transcrito e sua diversidade proteica após processo de tradução, assim como edições em éxons podem gerar diferentes aminoácidos na proteína (Tang et al., 2020). As edições feitas pela ADAR são enriquecidas em regiões 3' UTR e intrônicas de transcritos, e têm potencial para alterar a ligação de miRNAs e RBPs (do inglês, *RNA-binding*

proteins) (Jain; Jantsch; Licht., 2019). A ligação de RBPs pode depender tanto da estrutura em simples ou dupla fita quanto do motivo de sequência alterado pela ADAR (Hu et al., 2022).

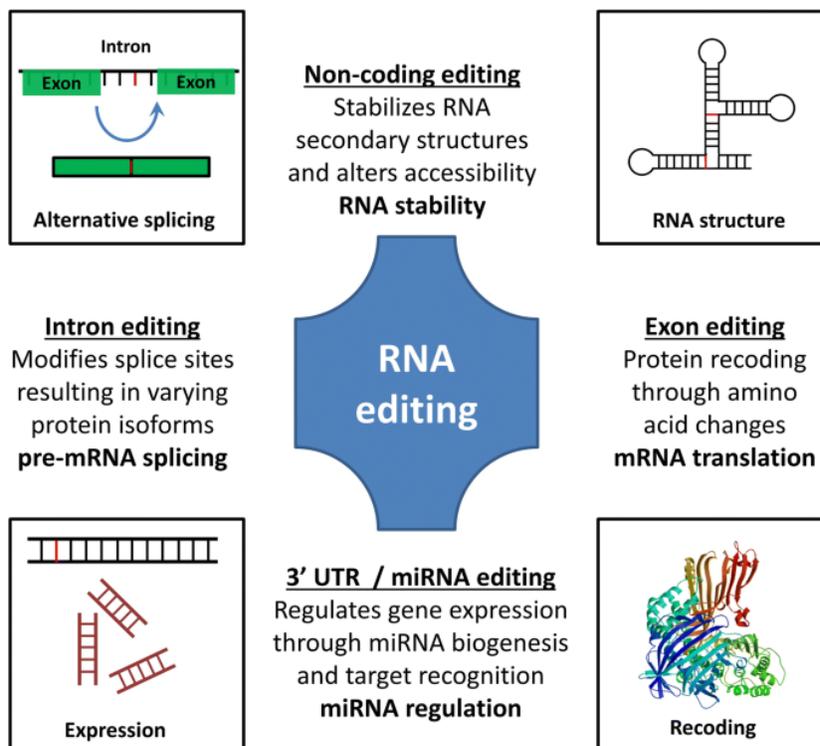


Figura 9. Possíveis regiões editadas pela proteína ADAR e suas consequências para o transcrito editado.

Fonte: Christofi; Zaravinos. RNA editing in the forefront of epitranscriptomics and human health. *Journal of translational medicine*, 17(1), 319. 2019.

1.1.6.1. Elementos ALU como alvo da proteína ADAR

Elementos ALU fazem parte de família de retroelementos determinados SINEs (do inglês, *short interspersed elements*), regiões repetitivas intercaladas curtas, presentes apenas em primatas. Os ALUs cobrem cerca de 11% do genoma humano e $\frac{1}{3}$ dos genes codificadores de proteínas com mais de 1 milhão de cópias no genoma (Schaffer; Levanon., 2021 ,Deininger., 2011). Diversas subfamílias desses elementos foram geradas ao longo da evolução, em que muitos desses perderam ou diminuíram a atividade de retrotransposons. Os elementos ALU mais antigos são representados pela subfamília J, enquanto que a mais recente e dominante em humanos é representado pela subfamília Y (Deininger., 2011).

Em relação à origem, não se sabe ao certo como se deu a formação dos primeiros elementos ALU, mas supõe-se que sejam originários do RNA 7SL (partícula reconhecidora de sinal), formadora da ribonucleoproteína SRP (do inglês, *signal recognition particle*). Em sua estrutura possui cerca de 300 pares de bases, dois dímeros, ligados por sequência rica em adenina (A) curta e cauda 3' poli-A, que tem papel em sua replicação. Sua transcrição pode ocorrer de maneira independente a partir de seu sítio promotor da RNA polimerase III, usando sequência poli-T para término da transcrição. De outra forma, pode estar incorporado em gene transcrito, estando normalmente associado às regiões 3' UTR ou intrônicas que formam dsRNA (Figura 10) (Daniel et al., 2014). As regiões repetitivas ALU tendem a formar sequências invertidas, que levam ao pareamento de bases por ligação Watson-Crick, criando sítio de ligação e edição pela proteína ADAR (Figura 11A) (Song et al., 2022).

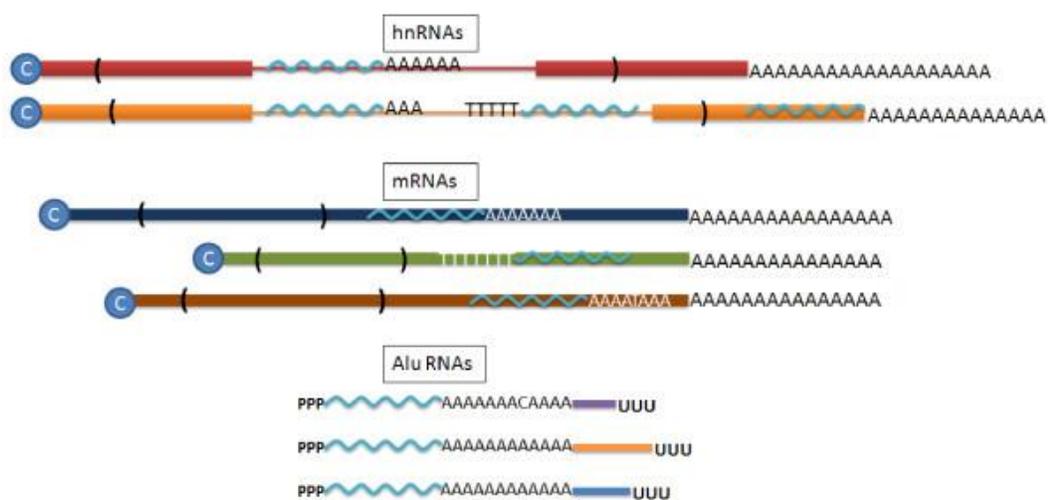


Figura 10. Transcritos contendo o elemento ALU. Elementos ALU em transcritos podem estar na forma de RNAs ALU quando transcritos a partir do seu promotor interno, ou presentes em heterogeneous nuclear RNA (hnRNAs) ou mRNAs, quando transcritos a partir do gene em que estão contidos.

Fonte: Deininger. Alu elements: know the SINEs. *Genome biology*, 12(12), 236. 2011.

Mutações com perda de função da proteína ADAR1 estão relacionadas com a síndrome de Aicardi-Goutières, em que uma das características é a elevação de ISGs. A deleção do gene ADAR1 em camundongos causou letalidade embrionária precoce como resultado do aumento dos níveis de IFN e ISG, mas a letalidade foi revertida pela inativação

do sistema de reconhecimento de dsRNA pela deleção de MAVS ou MDA5 (Jain; Jantsch; Licht., 2019). Em câncer de mama, pulmão, fígado e esôfago e na leucemia mielóide crônica foi encontrado aumento de ADAR1 relacionado à diminuição da resposta de IFN do tipo I (Xu; Öhman., 2018), a deleção do gene em células tumorais levou à ativação da via de infiltração por células imunes mediada por MAVS (Liu et al., 2019).

Sensores de dsRNA pela via MAVS-MDA5 (do inglês, *melanoma differentiation antigen 5*) formam uma das principais vias de ativação de IFN do tipo I e do tipo III, o reconhecimento de dsRNA pode advir de ambos dsRNA endógeno e exógeno. Em uma célula saudável a ADAR reduz a disponibilidade de dsRNA, porque a conversão A-para-I reduz a formação de ligações de hidrogênio entre bases pareadas, resultando no aumento de energia livre para formação do dsRNA e aumentando a fração de ssRNA (Figura 11). Essa redução do dsRNA diminui a sensibilidade de disparo de ativação da via de IFN pelos receptores RIG1-like (RLR), como RIG1 e MAVS. Quando ativada, MDA5 forma estrutura de filamento ao redor do dsRNA, causando a oligomerização da MAVS, via sua região CARD (do inglês, *caspase activation and recruitment domain*), na mitocôndria, e eleva os níveis de IFN-1 via IRF, e citocinas pró-inflamatórias (Figura 11B) (Vitali; Scadden., 2010; Ren et al., 2020).

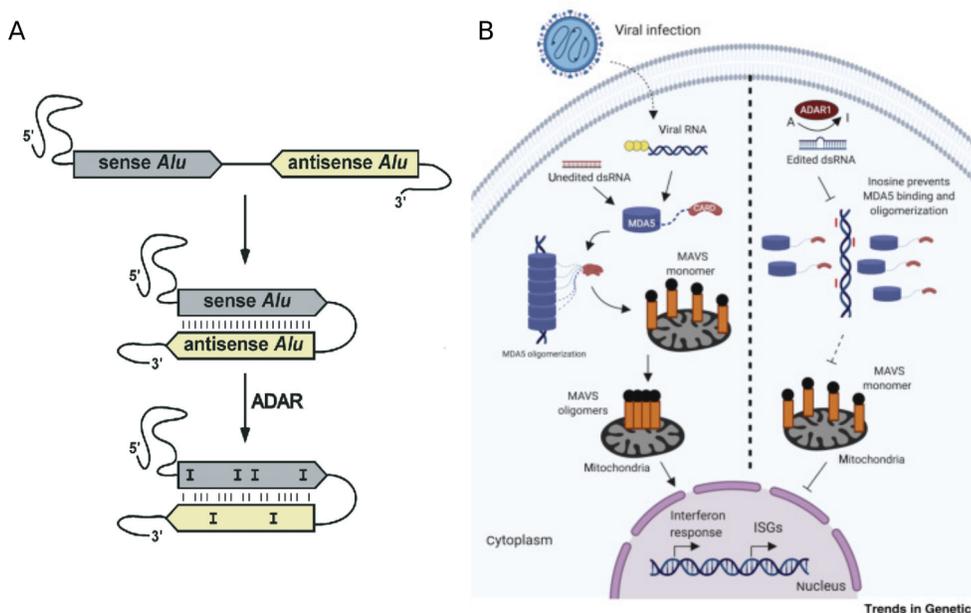


Figura 11. RNA dupla fita (dsRNA) afeta via MDA5-MAVS-IFN. (A) Formação de dsRNA pelos elementos ALU. A edição por ADAR reduz as ligações de hidrogênio entre bases pareadas e consequentemente a estabilidade de dsRNA (B) Reconhecimento de dsRNA endógeno estimula resposta do interferon. Adar regula dsRNA e modula resposta imune.

A ADAR1 possui papéis conflitantes no curso de diferentes viroses. Durante a infecção pelo HBV o aumento da ADAR1 via tratamento de IFN- α aumentou a edição de MAVS na região 3' UTR contendo ALU, essa edição causa a diminuição da expressão de MAVS pela perda de estabilidade do transcrito induzida pela proteína que se liga à RNA HuR (Human antigen R) (Li et al., 2021). Por outro lado, durante a infecção de COVID-19 nos estágios leves e severos houve diminuição global da edição pela ADAR1, em sítios ALU, em leucócitos. Essa diminuição é ampliada de forma inversamente proporcional com a gravidade da doença e foi relacionada com o aumento atrasado e prejudicial da resposta de IFN-I, somando-se aos efeitos inflamatórios típicos em estágios avançados da doença (Aune et al., 2022; Crooke et al., 2021).

1.1.7. NGS E SUAS APLICAÇÕES

Sequenciamento de próxima geração, ou do inglês *next sequencing generation* (NGS), se baseia no sequenciamento em larga escala de DNA e RNA, com possibilidade de milhões de leituras de sequências simultâneas e com retorno de grandes volumes de dados. Historicamente, o sequenciamento pode ser dividido em três gerações. A primeira geração se baseia no sequenciamento de Sanger, desenvolvido por Fred Sanger, em 1977, tendo sido utilizada no projeto genoma humano, que possibilitou identificar os cerca de 20 mil genes humanos e aproximados 3 bilhões de pares de bases (Moraes; Góes., 2016). Aplicações com Sanger ainda são rotineiramente utilizadas para sequenciamento em pequena escala. Trinta anos depois, deu-se início às tecnologias NGS com a segunda geração de sequenciadores, que têm como característica o sequenciamento de larga escala utilizando *reads* curtos e o sequenciamento por amplificação, ou seja, é necessária a amplificação para que haja leitura da sequência, e é necessário que as sequências de DNA ou RNA sejam fragmentadas. A plataforma illumina é o sequenciador mais utilizado atualmente. Para o RNA-seq, o RNA é primeiramente convertido em cDNA e em ambas as extremidades dos fragmentos são adicionados adaptadores, que se anexam aos seus complementares ligados à flowcell. Posteriormente, se emprega amplificação em ponte para geração de sequências clones agrupadas. Em cada ciclo são incorporados nucleotídeos, a ligação do nucleotídeo trava a

amplificação pela presença de terminador fluorescente na posição 3', cada tipo de nucleotídeo (A, T, C ou G) libera um sinal fluorescente diferente após estímulo, permitindo sua detecção através de um detector. A cada novo ciclo o terminador é removido e um novo nucleotídeo pode ser incorporado (Figura 12) (Zhong et al., 2021).

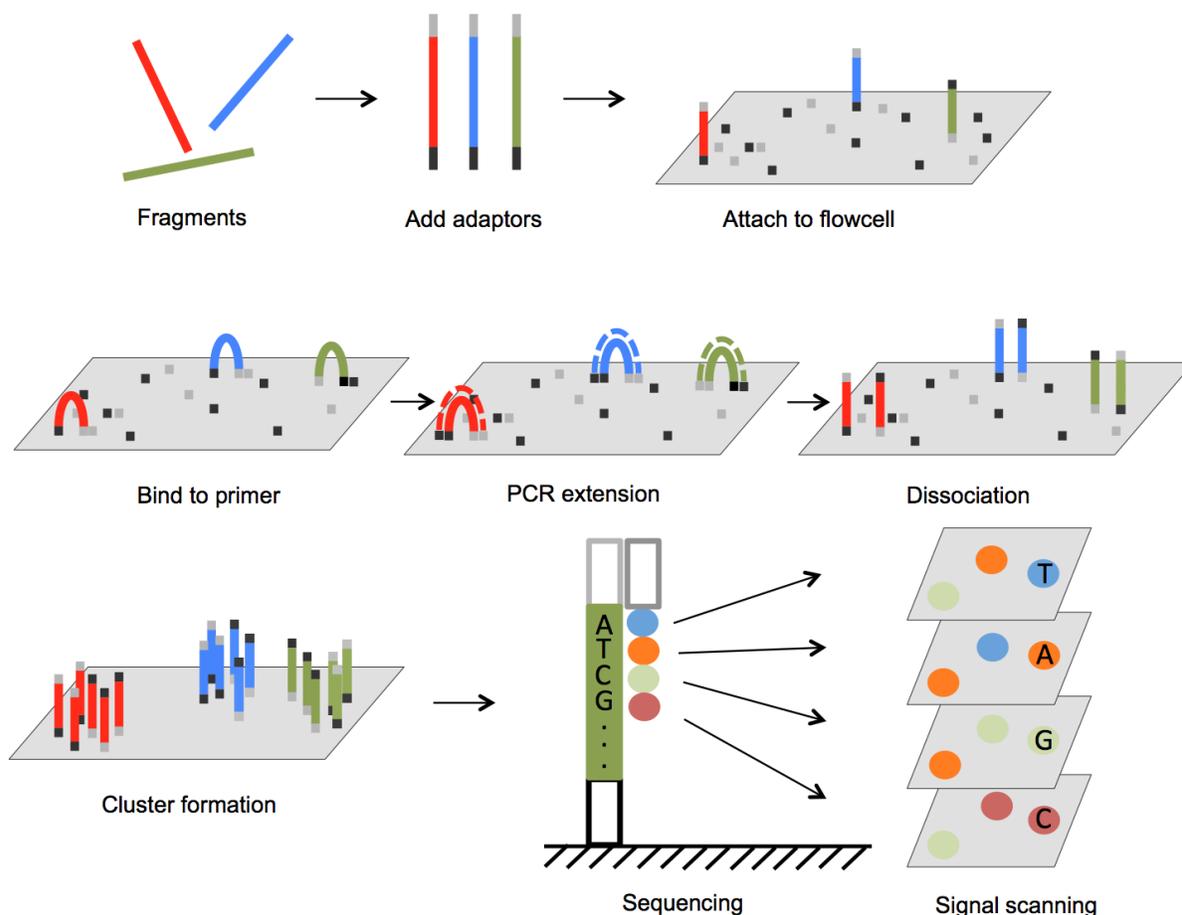


Figura 12. Passos do sequenciamento utilizando a plataforma Illumina.

Fonte: Lu; Shen; Warren; Walter. Next Generation Sequencing in Aquatic Models. Next Generation Sequencing - Advances, Applications and Challenges. 2016.

Os curtos fragmentos de *reads* lidos pela plataforma illumina possuem entre 75-300 pb, dificultando a cobertura total do genoma e a montagem das *reads*, por conta das regiões repetitivas do genoma. No RNA-seq, além da desvantagem para montagem das *reads*, há dificuldade para associação de isoformas aos transcritos, necessitando de métodos para quantificação e análise de isoformas. As tecnologias de sequenciamento de terceira geração

superam alguns dos problemas gerados pela de segunda ao permitir a leitura de longos reads com PacBio (Pacific Biosciences) e o Nanopore (Oxford), porém ainda possuem alto custo por sequenciamento por par de base e maior taxa de erros quando comparados com a geração anterior (Zhong et al., 2021; Pisapia et al., 2021; Hu et al., 2021).

Com a chegada do sequenciamento em larga escala, diversas aplicações baseadas no pré-processamento das amostras foram criadas, para estudos diversos. Aplicações com RNA-seq que não visam análise de RNA ribossômico (rRNA) normalmente fazem seleção de RNA por filtro das sequências de interesse por depleção de rRNA (ribo-minus) ou seleção de RNA de interesse, no caso de mRNA, utiliza-se seleção da cauda poli-A, presente em mRNA e alguns RNAs não codificadores (ncRNA). O sequenciamento pode utilizar ambas extremidades dos fragmentos para sequenciamento (*paired-end*) ou apenas uma das extremidades (*single-end*), além disso, os adaptadores contém as sequências de ligação à flowcell, os primers iniciadores do sequenciamento e podem conter UMIs (do inglês, *Unique molecular identifier*), sequências aleatórias adicionada aos fragmentos para identificação da origem de cada sequência amplificada. Após o processamento, a conversão dos sinais fluorescentes detectados para nucleotídeos, em processo computacional, é denominado base-call, são gerados arquivos do tipo fastq, contendo as sequências e o indicador de qualidade de cada nucleotídeo, que indica a probabilidade de erro para cada nucleotídeo, esses arquivos são filtrados e alinhados contra um genoma de referência para posterior análise e quantificação dos transcritos (Hu et al., 2021).

2. OBJETIVOS

2.1. Objetivos gerais

Analisar perfil de transcrito de leucócitos, de pacientes infectados pela COVID-19, a partir de dados públicos de sequenciamento de próxima geração (NGS), para encontrar eventos transcricionais relacionados com a patologia.

2.2. Objetivos específicos

- Encontrar, em repositório de dados online, como o Sequence Read Archive (SRA) e o Gene Expression Omnibus (GEO), arquivos de transcritomas sequenciados por Illumina e relacionados com células imunes, de pacientes infectados, ou células infectadas pelo SARS-CoV-2.
- Comparar processos biológicos com outras viroses ou entre os diferentes tipos celulares.
- Predizer formas alternativas de RNA e edições de RNA para relacionar com a infecção pelo SARS-CoV-2.
- Criar pipeline de análise diferencial das diferentes modificações pós transcricionais analisadas.

3. MÉTODOS

3.1. Recursos utilizados

Os recursos computacionais utilizados pertencem à estrutura do laboratório de Imunologia Molecular da Universidade de Brasília (UnB). O servidor utilizado possui 24 processadores Intel(R) Xeon(R) X5690 @ 3.47GHz (6 núcleos com 2 threads cada), e 94 gigabytes de memória RAM. Foram utilizados dois diferentes sistemas operacionais: *Ubuntu* 16.04 LTS e *Debian* GNU/Linux 11.

3.2. Obtenção das amostras de transcrito e pré-processamento

As amostras de transcrito foram obtidas através do banco de dados SRA (Sequence Read Archive). O banco de dados SRA é mantido pelo NCBI (National Center for Biotechnology Information) e armazena dados de sequenciamento de alto rendimento compartilhados em colaboração com o EBI (European Bioinformatics Institute) e o DDBJ (DNA Database of Japan). O NCBI SRA aceita diferentes tipos de arquivos, como os arquivos do tipo fastq e bam, também, são depositados dados que passam por diferentes metodologias relacionadas às tecnologias NSG e seleção de RNAs com características específicas (Katz et al., 2022). Os arquivos fastq são arquivos oriundos do base calling, e contém as sequências e os dados de qualidade. Por outro lado, o formato de arquivo BAM (do inglês, *binary alignment maps*) é a versão binária do arquivo SAM (do inglês, *sequence alignment maps*), que contém as informações relativas ao alinhamento de cada sequência contra o genoma de referência. Como filtro das bibliotecas a serem utilizadas, foram selecionadas apenas aquelas que continham sequenciamento com Illumina, single- ou paired-end, arquivos do tipo fastq e metodologias voltadas ao sequenciamento de mRNA com seleção da cauda poli-A. Primeiramente, foram selecionados estudos de células mononucleares de sangue periférico (PBMC) provindos de pacientes com a COVID-19, assim como diferentes dados de outras viroses para comparação das análises. Em uma segunda parte

do estudo foram selecionadas diferentes tipos celulares infectadas *in vitro* ou proveniente de pacientes infectados pela COVID-19.

Para o download das amostras no formato fastq foi utilizado o pacote de ferramentas SRA-toolkit (v2.9) (NCBI, [201-]), que entre suas ferramentas possibilita o acesso aos arquivos a partir do acesso ID de cada amostra e a conversão para seu formato padrão requisitado. O comando para realização dos downloads foi feito conforme se segue no terminal UNIX:

```
fastq-dump --split-files <acession id>
```

Em que a opção `--split-files` necessita ser utilizada para permitir a divisão dos arquivos paired-end, que serão utilizados nas próximas etapas de preparação. O controle de qualidade foi feito utilizando o programa fastQC (v0.12.1), que permite a análise da qualidade das *reads* a partir dos arquivos fastq ou bam. Os arquivos de qualidade são compilados e visualizados com o software MultiQC (v1.9) (Andrews., 2017, Ewels; Magnusson; Lundin; Källér., 2016). O programa é capaz de verificar os tipos de arquivos citados de qualquer diretório apontado a partir de uma varredura, podendo percorrer subdiretórios a partir do diretório apontado. Foram utilizados arquivos pré-processados no formato fastq para análise de qualidade com o comando:

```
multiqc <input files>
```

Em que “*input_files*” indica o diretório contendo os arquivos de saída do programa FastQC, que indicam a qualidade resultante das sequências de cada amostra.

Os arquivos com boa qualidade foram pré-processados usando o programa fastp, ferramenta rápida e eficiente desenvolvida em C++ que suporta multi-threading. A ferramenta permite realizar diversas operações de pré-processamento de dados de sequenciamento, Além de realizar a remoção dos adaptadores e filtragem de qualidade, o programa fastp também pode ser utilizado para realizar operações como correção de erros, remoção de *reads* duplicados e filtragem baseada no comprimento das sequências. Fornece relatórios estatísticos

detalhados e gráficos para análise e visualização dos dados. O comando utilizado foi realizado com a seguir:

```
fastp -m -L -i <I1.fastq> -I <I2.fastq> --merged_out <O merged.fastq>
```

Onde as opções -i e -I indicam os arquivos fastq de entrada, --merged_out indica o arquivo de saída correspondente. Esse comando permite processar arquivos fastq paired-end e realizar operações de trimming dos adaptadores e filtragem por qualidade. Foram utilizadas as definições padrões do programa.

3.3. Alinhamento das Reads ao genoma de referência

O programa de alinhamento STAR (Spliced Transcripts Alignment to a Reference) é um software desenvolvido para alinhamento de dados de RNA-seq de alta rendimento. Ele foi projetado para lidar com problemas gerados pelo sequenciamento de segunda geração, como a estrutura de transcritos não contígua, comprimentos de leitura curtos e capacidade crescente de geração de dados das NGS. Algumas das principais vantagens do STAR são: a sua velocidade de mapeamento, precisão e sensibilidade e possibilidade de alinhamento de ambos short e long-reads, *de novo* ou a partir de genoma de referência (Dobin et al., 2013).

O programa utiliza dois principais métodos para realização dos alinhamentos: *seed search* e *clustering, stitching and scoring* (Dobin et al., 2013). Na primeira etapa cada *read* é alinhado na melhor região possível a partir de método de busca denominado “*Maximal Mappable Prefix*”, que procura pela região do genoma de referência em que o *read* melhor é alinhado, *reads* que estão presentes em junções de splicing normalmente possuem uma parte bem alinhada (*seed*) e outra região mal pareada que pode alinhar melhor em regiões mais afastada do genoma, representado, por exemplo, uma junção exon-exon, portanto, uma segunda etapa do *seed search* procura por regiões que alinhem com a porção da *read* não pareada em um primeiro momento, dessa forma, pode encontrar regiões de splicing sem necessidade de prévia anotação. O método de *clustering, stitching and scoring* representa uma segunda etapa do alinhamento em que o programa agrupa as *reads* ancoradas, formando

blocos de contigs a partir da sobreposição e proximidade entre as *reads*. A união das *reads* é determinada por score de alinhamento local, não permitindo união de sequências com inserção ou deleção entre elas, mas permitindo nucleotídeos trocados (Dobin et al., 2013).

Foi utilizado o programa STAR (v2.7) para fazer o alinhamento de cada amostra pré-processada, o genoma de referência utilizado foi o GRCh38 (hg20), última versão do genoma humano lançada no início deste projeto e obtido em formato GTF (do inglês, *Gene transfer format*), que possui estrutura de tabela, através do projeto GENCODE (Frankish et al., 2018). A criação do *index* do genoma e alinhamento das sequências foram executadas como a seguir:

```
STAR --runMode genomeGenerate --genomeDir <output_dir> --genomeFastaFiles
<GRCh38.fa> --sjdbGTFfile <GRCh38.gtf>
```

Em que “--runMode” indica a geração do index através da opção “genomeGenerate”, “--genomeDir” indica o local do index, “--genomeFastaFiles” indica a sequência de referência no formato fasta e “--sjdbGTFfile” indica a anotação do genoma no formato GTF. O alinhamento foi dado pelo comando no terminal:

```
STAR --runThreadN n --genomeDir <genomeStarDir> --readFilesIn <I1.fastq> <I2.fastq>
--outFileNamePrefix <O> --outSAMtype BAM
```

Em que “--runThreadN” indica o número (n) de processadores utilizados, “--genomeDir” indica o diretório de saída do arquivo BAM. “--readFilesIn” indica os arquivos de entrada no formato fastq e “--outSAMtype” se refere ao tipo de arquivo de saída.

3.4. Matriz de contagem

Dado o arquivo de alinhamento gerado pelo software STAR, e a anotação do genoma fornecida pelo banco de dados GENCODE, foi utilizado o programa htseq-count (v2.0.2) para contar o número de reads que alinharam a cada região do genoma, e assim obter a matriz de

contagem, relacionada a cada gene, com fim de posterior expressão diferencial a ser realizada por outro programa estatístico como DESeq2 ou edgeR. O htseq-count conta os reads mapeados por gene nas regiões de exon, intron ou outra definida pelo usuário, utilizando uma lógica de ambiguidade que exclui reads que alinham a múltiplas regiões ou em mais de um gene diferente, objetivando diminuir viés na expressão diferencial. Este passo de contagem é fundamental para a realização de análise transcriptômica, permitindo analisar as diferenças na expressão gênica entre diferentes amostras biológicas e condições experimentais (Anders; Pyl; Huber., 2015). Para fim de quantificar a expressão em nível de genes, as matrizes de contagem contendo um gene por linha da tabela e as amostras em cada coluna foram obtidas com o comando:

```
htseq-count --additional-attr=gene_type -f bam <I.bam...> <GRCh38.gtf>
```

Em que a opção “-f” indica o tipo de arquivo a ser utilizado, podendo ser dado como entrada ambos arquivos no formato “SAM” ou “BAM”. Os arquivos de alinhamento (BAM) foram dados em sequência para as amostras de cada grupo em análise, indicados por <I.bam...>, da mesma forma, o arquivo de anotação utilizado no alinhamento também foi passado com argumento e está representado por <GRCh38.gtf>. A opção “--additional-attr=gene_type” foi passado ao comando para a inserção do atributo “gene_type” da anotação na matriz final. A coluna “gene_type” permite a identificação do transcrito como “exon” ou “lncRNA”, que foram posteriormente separados para análise de expressão diferencial usando o programa DESeq2.

3.5. Análise de expressão diferencial

A análise da expressão diferencial foi realizada utilizando o pacote R DESeq2 (v1.38.2), desenvolvido pelo projeto Bioconductor e disponível em <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>. O programa é especializado em normalização, visualização e análise de expressão gênica diferencial, e utiliza técnicas empíricas de Bayes para estimar o log do fold change, a dispersão e

estimativas relacionadas, como o p-valor e o FDR (do inglês, *false discovery rate*). O fold change é uma medida utilizada para avaliar a razão da expressão gênica entre duas condições, sendo definido como a razão entre as médias das expressões. O pacote utiliza as matrizes de contagem bruta das reads geradas pelo htseq-count para normalizar os dados. Com fim de corrigir possíveis distorções na distribuição de contagens entre os genes, a normalização é feita com relação à cobertura do sequenciamento de cada amostra, com relação ao tamanho dos genes, e também, em relação à composição de cada biblioteca. O programa estima a dispersão utilizando as contagens relacionadas à expressão de genes de um mesmo conjunto de réplicas, através de estimativa por máxima verossimilhança e ajusta a expressão média de cada amostra de acordo com a sua dispersão usando modelo linear generalizado (Love et al., 2014).

Dada a matriz de contagem “*rawCountsDF*” e a matriz contendo a identificação de cada amostra e seu respectivo grupo de estudo. O comando em R para criação da classe *DESeqDataSet*, contendo a matriz de contagem, o design experimental e as fórmulas utilizadas para calcular a dispersão e a expressão diferencial é dado por:

```
> deseq2Data <- DESeqDataSetFromMatrix(countData=rawCountsDF, colData=sampleDF,
design= ~ conditions)
```

Em que a variável “*countData*” indica a matriz inicial contendo a contagem bruta de reads, “*colData*” indica a matriz contendo informações relacionadas às amostras e *design* indica as colunas referidas ao “*colData*” a serem analisadas em relação aos grupos tratamento e controle, que serão comparados. Os genes com baixa expressão entre todas as amostras são excluídos e a análise final, contendo normalização, modelagem e teste de hipótese de Wald, que se baseia na distribuição assintótica normal dos estimadores de máxima verossimilhança dos parâmetros do modelo (Love et al., 2014) é dada pelo comando R “*DESeq(deseq2Data)*”, em que “*deseq2Data*” indica o objeto “*DESeqDataSet*”. O código foi simplificado para exemplo e o código completo pode ser encontrado em: <https://github.com/PHAB1/transcAnalysis>.

3.6. Processamento de lncRNA

A obtenção de lncRNA se deu através do arquivo de anotação gtf e com o programa htseq-count como mostrado no tópico 2.5 e a análise de expressão foi feita com o DESeq2 como descrito em 2.6.

A análise de enriquecimento de via foi feita com o pacote R LncRNAs2Pathways (v.1.1), desenvolvido para associar vias relacionadas com lncRNAs e está disponível em <https://cran.r-project.org/web/packages/LncPath/LncPath>. O pacote utiliza uma biblioteca em redes, que relaciona lncRNAs com a expressão de mRNAs. A rede de interação foi criada a partir da análise de 28 diferentes estudos e leva em conta interações proteína-proteína anotadas em diferentes banco de dados como REACTOME (Gary et al., 2022) e HPRD (Mishra et al., 2006), sendo capaz de relacionar os lncRNA diferencialmente expressos com vias "KEGG", "Reactome" ou "BioCarta". A execução em R para a análise de via dos lncRNAs foi realizado pela linha de comando:

```
> lncRNA_2pathway <- lncPath(lncRNA_genes_list, LncPathNet)
```

Em que lncRNA_genes_list indica a lista de lncRNA diferencialmente expressos e LncPathNet indica uma tabela contendo as informações relacionadas com a rede de interações entre lncRNA, expressão e PPI.

3.7. Quantificação e detecção de splicing alternativo

3.7.1. rMATS

Para detecção de SA diferencial foi utilizado o programa rMATS turbo (v4.1.2), disponível em <http://rnaseq-mats.sourceforge.net>. O programa se baseia em utilizar reads mapeados, em regiões que indiquem diferentes isoformas, para detectar e estimar a proporção de diferentes tipos de splicing entre diferentes condições. Aqui, os cinco tipos de splicing possíveis são identificadas, sendo eles, SE (Skipped exons), RI (Retained intron), MXE

(Mutually exclusive exon), A3SS (alternative 3' Splicing Site) e A5SS (alternative 5' Splicing Site). Como exemplo, em SE os reads mapeados nas regiões utilizadas como marcadores de SA são os reads de junção (S), posicionados entre o exon montante e o exon a jusante com salto de um exon, formando a isoforma, enquanto que os *reads* relacionados com a forma canônica do gene, com a inclusão do exon, são os *reads* de inclusão (I). Os níveis de inclusão (ψ) são definidos como a porcentagem da inclusão do transcrito $\psi = (I/LI) / (I/LI + S/LS)$ em que: ψ = Nível de inclusão (IncLevel), I = número de leituras mapeadas para a isoforma de inclusão, S = número de leituras mapeadas para a isoforma SE, LI = comprimento efectivo da isoforma de inclusão do exon, LS = comprimento efectivo da isoforma SE (Shen et al., 2014). O comando UNIX dado para execução do programa rMATS foi executado como a seguir:

```
rmats.py --b1 <rmats_files1.txt> --b2 <rmats_files2.txt> --gtf <GRCh38.gtf> -t paired  
--readLength <L> --nthread 4 --od <AltSpl_outDirectory>
```

Em que “--b1” e “--b2” indicam os arquivos *.txt*, contendo os caminhos de cada amostra em formato *.bam* alinhadas pelo programa STAR, em que “rmats_files1.txt” indica as amostras controle e “rmats_files2.txt” indica as amostras tratamento. “--gtf” indica o caminho do genoma de referência GRCh38 utilizado no alinhamento das amostras. “-t” indica o tipo de sequenciamento single ou paired-end. “--readLength” indica o tamanho das *reads* de interesse obtidas em cada experimento, determinada como “L” e “--od” indica o diretório de saída.

rMATS utiliza teste da razão de verossimilhança para calcular a diferença estatística entre os dois grupos analisados, assim, é testado, para cada evento, quando os grupos possuem média diferente entre eles, dado p-valor definido pelo usuário. Para filtro dos eventos de splicing, após identificação com rMATS, foi utilizado $FDR < 0.05$ e $|\psi_{it} - \psi_{ic}| > 0.1$, em que ψ_{ic} e ψ_{it} indicam as médias relativas ao psi em cada evento entre os grupos (t) tratamento e (c) controle (Shen et al., 2014).

3.8. Detecção e análise diferencial de edição de RNA

Para detecção de edições de RNAs foi utilizado o programa SPRINT (v0.1.8), disponível em <https://github.com/jumphone/SPRINT>. O programa SPRINT pré-processa as *reads* com a remoção das variantes genéticas de nucleotídeo único (SNPs) anotadas, visando remover falsos positivos provindos de variantes genéticas, e, posteriormente, identifica candidatos à ER (Feng et al., 2017). Apenas os candidatos contendo nucleotídeos com alta qualidade $q > 20$ e *reads* não mapeados em regiões com repetições indistinguíveis são mantidos, removendo aqueles com erro de mapeamento ou de sequenciamento. Além disso, foram mantidas apenas as edições presentes em todas as amostras para um mesmo grupo.

O software possui duas formas de aplicação, a primeira através dos arquivos fastq em que o alinhamento é realizado com o BWA (do inglês, *Burrows-Wheeler Aligner*) (v0.7.12), feito pelo pipeline do SPRINT, e posteriormente, as amostras são processadas para identificação das edições de RNA. Outra forma de utilização é a partir da identificação de ER direta a partir de sequências anteriormente alinhadas e fornecidas através de arquivo de alinhamento SAM ou BAM. No presente trabalho foi utilizado o pipeline com pré-alinhamento pelo BWA, por possuir maior precisão de detecção de candidatos à ER utilizando esse alinhador (Lo et al., 2020). O comando de execução no terminal UNIX foi:

```
sprint main -1 <fastq1.fastq> -2 <fastq2.fastq> <GRCh38.fa> <output_sprintDir> <bwa>  
<samtools>
```

Em que “-1” e “-2” indicam os arquivos fastq paired-end, <GRCh38.fa> indica o genoma humano em formato fasta, <output_sprintDir> indica o diretório de saída e <bwa> e <samtools> indicam os caminhos dos programas requisitados pelo programa *sprint*.

3.9. Análise funcional

Anotar as propriedades funcionais de genes é uma tarefa fundamental na análise transcritômica. A análise funcional de genes é um processo que consiste em classificar um

conjunto de genes diferencialmente expressos em termos biológicos, em que a anotação de um grupo de genes é feita por um banco de dados, usando *Gene Ontology* (GO), conceito que introduziu a associação de genes a termos (nomenclatura relacionada a um processo biológico). Bancos de dados mais recentes como o MsigDB e LINCS L1000 possibilitam novas associações de genes a termos. O primeiro faz sobreposição de conjuntos de genes relacionados com diferentes estados ou processos biológicos para gerar uma assinatura de expressão coerente. O segundo associa conjunto de genes a diferentes ensaios de perturbação de um determinado gene a partir de sua superexpressão ou deleção. O objetivo da análise funcional é analisar a distribuição desses genes em diferentes categorias biológicas e entre diferentes espécies. A GO divide os termos em três categorias: componente celular, processo biológico e função molecular. A significância desses termos é dada pelo enriquecimento de termos.

3.9.1. ClusterProfiler

O pacote R ClusterProfiler (v4.6.0) é um pacote desenvolvido pelo projeto Bioconductor e permite a automatização de classificação de termos biológicos utilizando GO e a análise estatística dos termos a partir de um conjunto de genes e está disponível em: <https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>. Para análise estatística é utilizado o enriquecimento, baseado em distribuição hipergeométrica para cálculo do p-valor e estimação do q-valor para controle do FDR (Yu et al., 2012). Diferentes métodos para classificação de GO são fornecidos pelo clusterProfiler, o método utilizado foi *enrichGO*. Para execução em R foi executado o seguinte comando através do ambiente R:

```
> enrichGO(geneSet, OrgDb, keyType = "ENTREZID", ont = "ALL", pvalueCutoff = 0.05,  
pAdjustMethod = "BH", qvalueCutoff = 0.05,)
```

Em que “*geneSet*” indica um vetor contendo os genes a serem analisados, “*OrgDB*” indica o pacote de anotação de genes humanos Bioconductor org.Hs.eg.db (v3.16.0), que faz a conexão entre identificador “ENTREZ” e código “*GenBank*” (Carlson., 2019). “*Ont*” indica o

tipo de GO utilizada, no caso, foram utilizadas todas as categorias, indicado por “*ALL*”, o valor limite máximo do p-valor indicado por “*pvalueCutoff*” foi de 0.05, o q-valor máximo, indicado por “*qvalueCutoff*”, foi de 0.05 e o método para cálculo do q-valor utilizado foi de Benjamini-Hochberg, usado para controlar a taxa de FDR.

3.9.2. enrichR

A ferramenta enrichR pode ser acessada através da sua plataforma: <https://maayanlab.cloud/Enrichr/> por onde possibilita a análise de enriquecimento dos termos, possuindo mais de 430.000 termos em 209 diferentes bibliotecas (Kuleshov et al., 2016). Uma das vantagens das bibliotecas disponibilizadas para análise com o enrichR é a possibilidade de comparação com outros experimentos de RNA-seq em determinadas condições, como perturbações de genes (por exemplo, indução ou deleção do gene), estímulos de drogas, doenças e patógenos. A análise de enriquecimento com enrichR é feita com diferentes métodos, entre eles: teste de proporção, Z-score ou escore combinado, que é calculado tomando o logaritmo do p-valor do teste de Fisher e multiplicando-o pela pontuação z do desvio da pontuação esperada. Visualização com R e outras análises performadas foram feitas utilizando o pacote *clusterProfiler* citado no item 2.9.1

3.9.3. Metascape

O Metascape é uma plataforma web voltada a facilitar a análise de enriquecimento a partir de uma interface simples e a possibilidade de integrar diferentes condições a serem comparadas a partir de listas contendo genes de interesse, cada uma relacionada a uma condição de interesse (Zhou et al., 2019). A página pode ser acessada em: <http://metascape.org>. A ferramenta primeiro identifica o gene submetido, permitindo a entrada dentre outros identificadores: Entrez Gene ID, gene symbol, RefSeq number e Ensembl ID. A análise com metascape faz análise de enriquecimento em diferentes bibliotecas de termos, incluindo GO de modo customizado ou por modelo pré-definido.

3.10. Análise de via

Reactome pathway é um banco de dados curado de vias e reações que ocorrem em humanos. A plataforma do reactome permite a consulta de genes envolvidos nos processos relacionados a termos através de análise de enriquecimento e pode ser acessado em: <https://reactome.org/> (Marc et al., 2021). A análise de via foi feita utilizando uma lista de genes diferencialmente expressos, submetidos na plataforma.

3.11. Pipeline para análise de transcrito

Para a automatização e reprodutibilidade das análise de transcrito feitas neste trabalho foi criado pipeline automatizado com o programa *snakemake*, que gerencia o workflow do pipeline, e permite integração com scripts que utilizem as linguagens de programação python e R (Mölder et al., 2021).

3.12. Visualização dos resultados

Os gráficos utilizados na sessão de resultados foram gerados utilizando os programas como indicado:

- ❖ Diagramas de Venn: VennDiagram, versão 1.7.3 (disponível em: cran.r-project.org/web/packages/VennDiagram)
- ❖ dotPlot: ggplot2, versão 3.4.0 (disponível em: <https://ggplot2.tidyverse.org/>)
- ❖ Sashimi plot: IGV, versão 2.13.2 (disponível em: broadinstitute.org/software/igv/)
- ❖ Gráfico de barras: ggplot2, versão 3.4.0 (disponível em: <https://ggplot2.tidyverse.org/>)
- ❖ Heatmaps gerados com clusterProfiler (v4.6.0): enrichplot, version 1.18.3 (disponível em: <https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>)

- ❖ Volcano plot: EnhancedVolcano, version 1.16.0 (disponível em: <https://bioconductor.org/packages/devel/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html>)
- ❖ Gráfico do tipo upset: UpSetR, versão 1.4.0 (disponível em: <https://github.com/hms-dbmi/UpSetR>)
- ❖ Manhattan plot: qqman, versão 0.18.0 (disponível em: <https://cran.r-project.org/web/packages/qqman/qqman.pdf>)

4. RESULTADOS E DISCUSSÃO

4.1. Pipeline *transcAnalysis*

O pipeline *transcAnalysis* executa a análise de expressão de mRNA, lncRNA, SA e ER a partir do arquivo BAM, como entrada, que provém do alinhamento das *reads* obtidas a partir de arquivos fastq do RNA-seq com Illumina e facilita o desenvolvimento ao permitir a execução dos passos com apenas uma linha de comando, o que é possível por conta da utilização do gerenciador de pipelines *snakemake* (*ver métodos*). O pipeline tem como saída as análises diferenciais relacionadas a cada evento e se encontra disponível em: <https://github.com/PHAB1/transcAnalysis>. O fluxo de trabalho e as etapas do pipeline estão apresentados na figura 13.

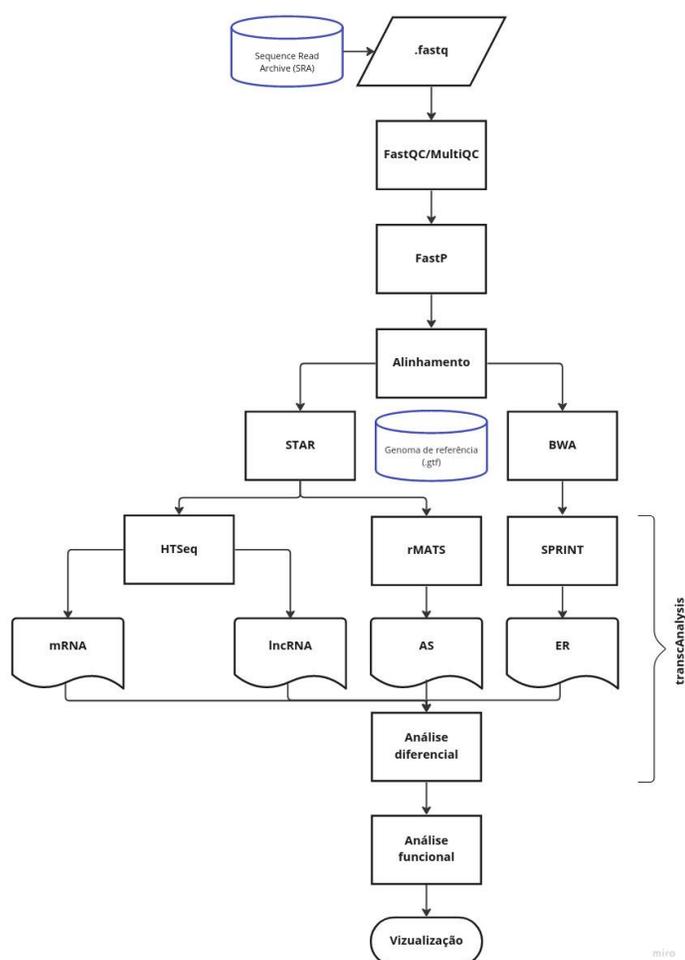


Figura 13. fluxo de trabalho e as etapas do pipeline transcriptAnalysis.

4.2. Caracterização das amostras utilizadas

A pesquisa no banco de dados SRA retornaram estudos com PBMC e outros estudos com diferentes tipos de células imunes. Os experimentos com PBMC podem ser encontrados no banco de dados SRA pelos IDs **SRP304889** e **SRP099676**, que se referem, respectivamente, a amostras de pacientes com Covid-19 e Dengue. Outros estudos com células imunes específicas relacionadas à infecção pelo SARS-CoV-2 incluem monócitos de pacientes infectados nos estágios leve e grave (**SRP323042**), monócitos clássicos de pacientes infectados (**SRP269501**) e células dendríticas de pacientes hospitalizados (**SRP273473**), a identificação de cada amostra e a sua origem pode ser encontrada na tabela 1.

O desenho experimental descrito e as características de cada estudo podem ser encontradas na tabela 2 e envolvem o uso de amostras de PBMC provenientes de pacientes com Covid-19 que retestaram positivo para o vírus, pacientes convalescentes (de estágio moderado) e pacientes saudáveis. Além disso, são utilizadas amostras de PBMC de pacientes com dengue em estágio leve da doença e pacientes controle para comparação de expressão e isoformas diferencialmente reguladas nas duas condições. Esse estudo foca no aproveitamento de dados disponibilizados em bancos de dados públicos, sem gastos adicionais com experimentos. No entanto, há limitações relacionadas ao uso de dados públicos disponíveis para download em relação consistente entre eles. Por mais que não inviabilize as análises, os dados de PBMC de Covid-19 foram obtidos de pacientes que retestaram positivo para a doença e o controle da dengue advém de pacientes com outras doenças febris com sintomas similares ao da dengue, mas que eram negativos para a dengue (NS1/IgM) e RNA viral, mas apresentavam sintomas semelhantes aos da dengue. Por mais que o SARS-CoV-2 e o vírus da dengue possuam diferentes vias de infecção, ambos possuem escape imune similar ao perturbar mecanismos de resposta viral via IFN do tipo 1, causam interferência da resposta imune inata, atraso da resposta de células T (Unterman et al., 2022) e podem levar à tempestade de citocinas em casos graves (Malavige et al., 2022), servindo, portanto, como controle para análise de eventos marcadores da doença que estejam diferencialmente representados especificamente em PBMC.

Os dados que contêm tipos celulares específicos foram analisados e comparados em relação à expressão gênica, bem como a vários tipos de SA (A3SS, A5SS, RI, MXE e SE) e

ER pela proteína ADAR (edições A-para-I), utilizando o workflow do pipeline *transcAnalysis* descrito na seção 3.1. As amostras obtidas diferem em relação às diferentes características detalhadas na tabela 2. O foco principal, na análise de célula específica, foi dado aos monócitos por mostrarem papel chave no decorrer da COVID-19 (Knoll; Schultze; Schulte-Schrepping., 2021). O estudo apresenta limitações ao utilizar dados públicos, em que o desenho original dos estudos foi realizado para diferentes finalidades. Neste estudo, aproveitamos esses experimentos para as diferentes análises, em que, a grande disponibilidade desses dados disponíveis oferece oportunidades para análise de informações não utilizadas em seus estudos correspondentes.

Tabela 1. Identificação das amostras dispostas no SRA

| Estudo | Bioprojeto SRA | Condição | Estado | Corrida SRA | Nº de Reads | | | | |
|----------------------|-----------------------|-------------------|------------------|------------------------|--------------------|-------------------|-------------------------|--------------------|-------------------|
| PBMC - Dengue | PRJNA374737 | Tratamento | Infectado | SRR5257026 | 49.411.022 | | | | |
| | | | | SRR5257027 | 39.390.445 | | | | |
| | | | | SRR5257028 | 30.941.580 | | | | |
| | | | | SRR5257029 | 43.988.815 | | | | |
| | | | | SRR5257030 | 38.859.747 | | | | |
| | | | | SRR5257031 | 71.930.269 | | | | |
| | | | | SRR5257018 | 56.541.565 | | | | |
| | | Controle | OD | SRR5257019 | 67.742.933 | | | | |
| | | | | SRR5257020 | 38.499.123 | | | | |
| | | | | SRR5257021 | 60.484.639 | | | | |
| | | | | SRR5257022 | 35.938.451 | | | | |
| | | | | SRR5257023 | 64.821.551 | | | | |
| | | | | PBMC - Covid-19 | PRJNA699856 | Tratamento | Reteste positivo | SRR13639995 | 20.258.782 |
| | | | | | | | SRR13639996 | 20.146.471 | |

| | | | | | |
|---------------------------------|--------------------|-------------------|----------------------|-------------|------------|
| | | | | SRR13639997 | 20.620.265 |
| | | | | SRR13639998 | 20.057.825 |
| | | | | SRR13639999 | 20.743.927 |
| | | | | SRR13634000 | 20.760.038 |
| PBMC Covid-19 | | Tratamento | Convalescente | SRR13639989 | 20.702.403 |
| | | | | SRR13639991 | 20.593.037 |
| | | | | SRR13639992 | 20.115.520 |
| | | | | SRR13639993 | 20.577.041 |
| | | | | SRR13639994 | 20.509.447 |
| | | | | SRR13639990 | 20.087.749 |
| | | Controle | Saudável | SRR13639979 | 20.178.060 |
| | | | | SRR13639980 | 20.569.607 |
| | | | | SRR13639981 | 20.576.482 |
| | | | | SRR13639982 | 20.675.095 |
| | | | | SRR13639983 | 20.642.726 |
| | | | | SRR13639984 | 20.433.942 |
| Monócitos - Covid-19 | PRJNA735729 | Tratamento | Grave | SRR14749963 | 33.493.754 |
| | | | | SRR14749964 | 39.930.793 |
| | | | | SRR14749970 | 38.052.731 |
| | | | | SRR14749971 | 33.401.539 |
| | | | | SRR14749972 | 32.854.752 |
| | | | | SRR14749975 | 30.552.812 |
| | | Tratamento | Leve | SRR14749965 | 39.528.534 |
| | | | | SRR14749966 | 28.321.591 |
| | | | | SRR14749967 | 35.024.230 |

| | | | | | |
|---------------------------------------|--------------------|-------------------|------------------|-------------|------------|
| | | | | SRR14749968 | 33.089.251 |
| | | | | SRR14749969 | 33.022.222 |
| | | | | SRR14749973 | 30.877.145 |
| | | Controle | Saudável | SRR14749982 | 37.447.689 |
| | | | | SRR14749983 | 28.066.312 |
| | | | | SRR14749984 | 41.511.565 |
| | | | | SRR14749985 | 29.459.397 |
| | | | | SRR14749986 | 35.588.269 |
| | | | | SRR14749987 | 35.701.682 |
| Monócitos clássicos - Covid-19 | PRJNA643420 | Tratamento | Infectado | SRR12124915 | 12.855.122 |
| | | | | SRR12124917 | 11.680.906 |
| | | | | SRR12124918 | 17.517.590 |
| | | | | SRR12124919 | 27.319.147 |
| | | Controle | Saudável | SRR12414261 | 16.721.906 |
| | | | | SRR12414262 | 14.274.892 |
| | | | | SRR12414263 | 23.149.634 |
| | | | | SRR12414264 | 17.116.604 |
| CD - Covid-19 | PRJNA648425 | Tratamento | grave | SRR12319414 | 26.584.091 |
| | | Controle | Saudável | SRR12319413 | 27.390.634 |

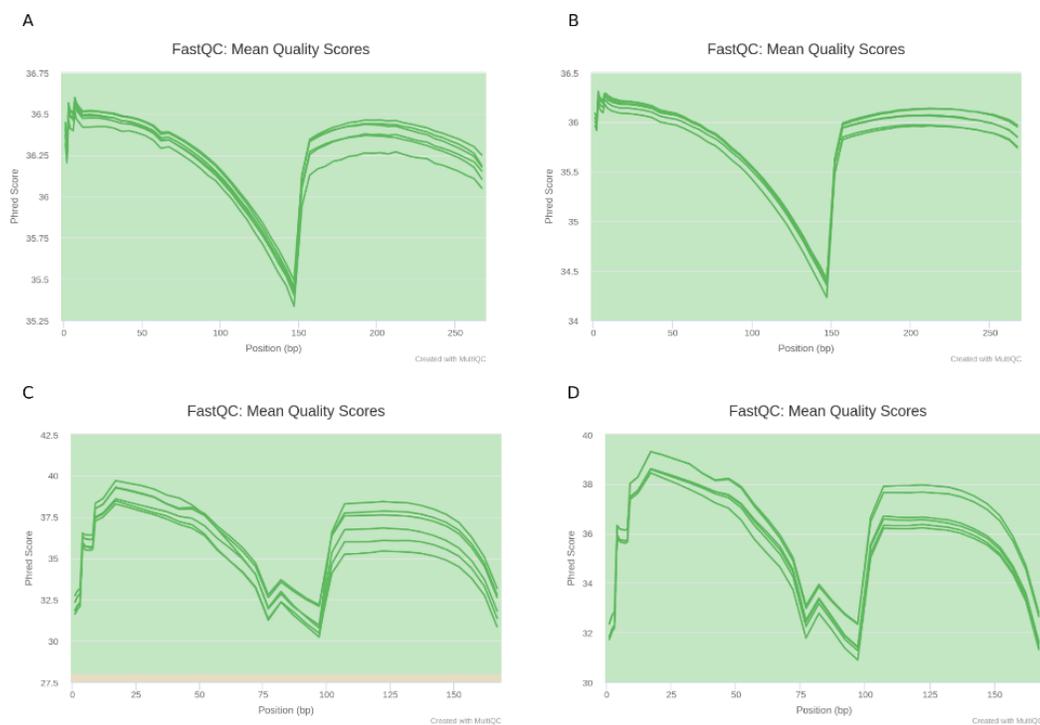
OD = other diseases; CD = célula dendrítica

Tabela 2. Desenho experimental dos dados obtidos contendo diferentes características.

| Características | PBMC - Dengue | PBMC - Covid-19 moderada | PBMC Covid-19 convalescente | Monócitos - Covid-19 grave | Monócitos - Covid-19 leve | Monócitos clássicos - Covid-19 | Monócitos - infectados | CD - Covid-19 grave |
|------------------------|----------------------|---------------------------------|------------------------------------|-----------------------------------|----------------------------------|---------------------------------------|-------------------------------|----------------------------|
|------------------------|----------------------|---------------------------------|------------------------------------|-----------------------------------|----------------------------------|---------------------------------------|-------------------------------|----------------------------|

| | | | | | | | | |
|---|------------------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Infecção <i>in vivo/in vitro</i> | <i>In vivo</i> | <i>In vivo</i> | <i>In vivo</i> | <i>In vivo</i> | <i>In vivo</i> | <i>In vivo</i> | <i>In vitro</i> | <i>In vitro</i> |
| Tipo celular | PBMC | PBMC | PBMC | CD14+/ CD16+ | CD14+/ CD16+ | CD14+ | | CD |
| Plataforma NGS | Illumina HiSeq 2500 | HiSeq X Ten | HiSeq X Ten | NovaSeq 6000 | NovaSeq 6000 | NovaSeq 6000 | | BGISEQ-500 |
| Seleção RNA | poli A | poli A | poli A | | | | | |
| Layout | paired-end | paired-end | paired-end | paired-end | paired-end | paired-end | paired-end | single-end |
| Tamanho das reads (pb) | 100 | 150 | 150 | 150 | 150 | 76 | 150 | 50 |

A qualidade de todas as amostras foi verificada com o programa fastQC e, em seguida, pré-processada com o fastP. Todas as amostras usadas no estudo apresentaram qualidade satisfatória mesmo antes do pré-processamento com fastP, o conjunto da média da qualidade para as amostras de cada estudo foram geradas com auxílio do programa multiQC e estão representados nas figuras 14A-k.



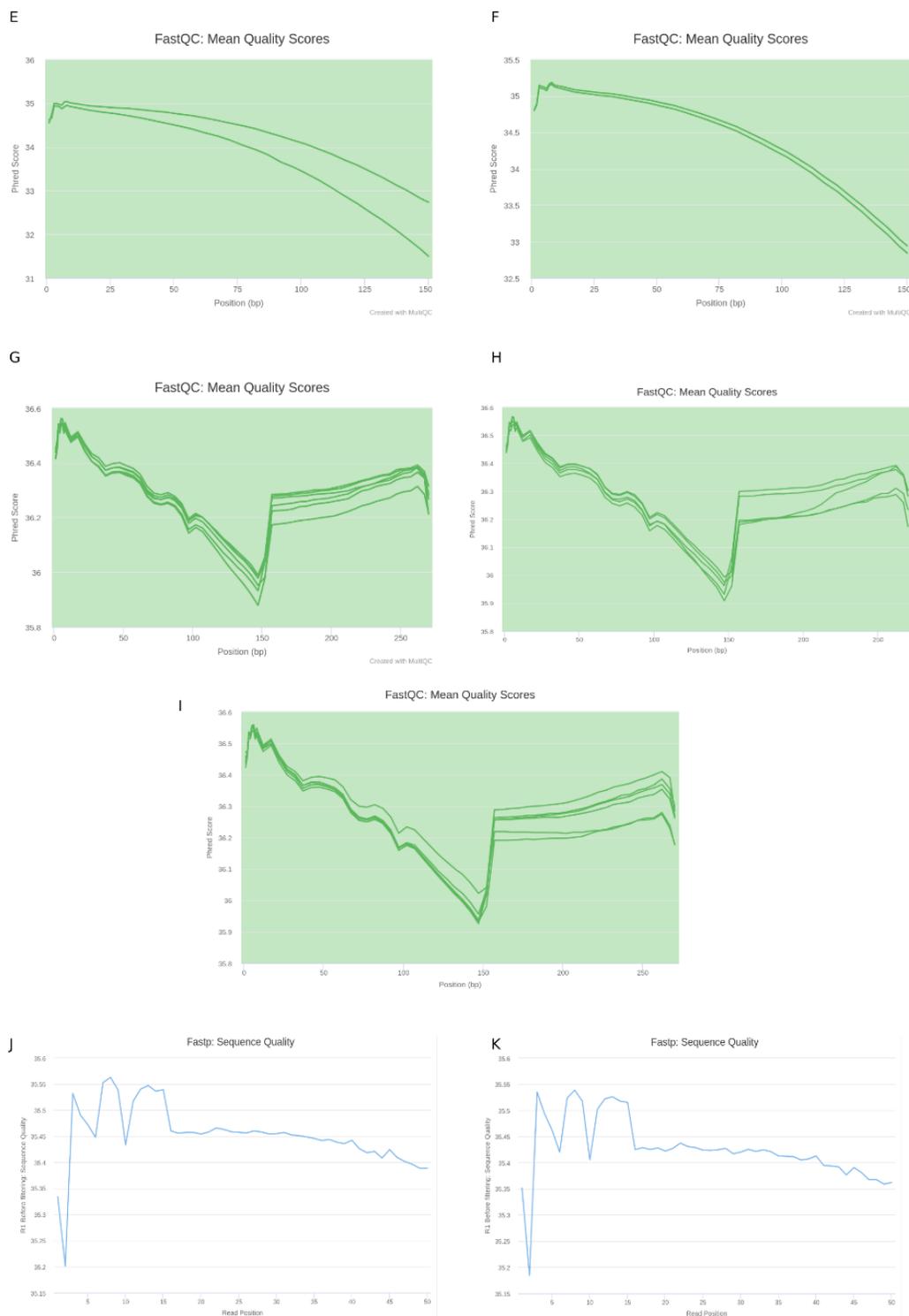


Figura 14. Qualidade média das *reads* por par de base. (A) Controle, PBMC da COVID-19 moderada. (B) Tratamento, PBMC da COVID-19 moderada. (C) Controle, PBMC da dengue (D) Tratamento, PBMC da dengue. (E) Controle, Monócitos infectados *in vitro*. (F) Tratamento, Monócitos infectados *in vitro*. (G) Controle, monócitos de pacientes com COVID-19 (H) Tratamento, monócitos de pacientes com COVID-19 severa. (I) Tratamento, monócitos de pacientes com COVID-19 severa.

4.3. PBMC: COMPARANDO VIROSES

4.3.1. Expressão e análise diferencial

A análise do transcrito de PBMC de pacientes com Covid-19 e Dengue revelou diferenças na expressão gênica entre esses pacientes. Foi observado que os pacientes com Covid-19 apresentaram um aumento na expressão de 5945 genes e uma diminuição na expressão de 5543 genes, enquanto os pacientes com Dengue mostraram um aumento na expressão de 1724 genes e uma diminuição na expressão de 1445 genes.

A Figura 15 A-B apresenta a expressão dos genes em gráficos do tipo vulcão, que é utilizado para visualizar a significância estatística (p-valor; eixo y) em relação à magnitude da expressão (Log Fold change; eixo x) entre duas condições. Na Figura 15A, a expressão gênica é comparada entre pacientes com Covid-19 e pacientes saudáveis, enquanto que na Figura 15B, compara-se a expressão gênica entre pacientes com dengue com a expressão de pacientes controle. O símbolo representativo dos genes com maior expressão e menor FDR estão mostrados nos gráficos (fig. 15A e 15B), os genes representados indicam resposta inflamatória e antiviral, como, por exemplo, mostrado por IFNG, IL6, CCL3 na COVID-19 e IFIT1, IFI6, ISG15, e CCL8 em dengue. Observamos que houve um maior número de genes diferencialmente expressos na Covid-19, com intersecção de 1605 genes entre as duas viroses (Fig. 16).

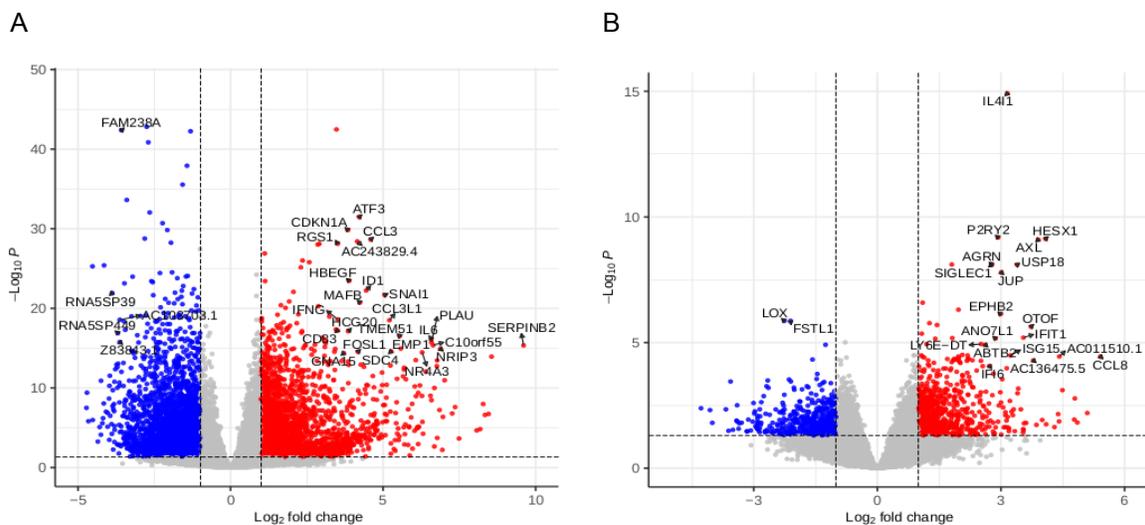


Figura 15. Gráfico de vulcão com a dispersão do *fold change* pelo *FDR*. Foram considerados os transcritos com $|\text{fold change}| > 1$ e $\text{FDR} < 0.05$. (A) Genes diferencialmente expressos em células mononucleares de sangue periférico (PBMC) de pacientes COVID-19 em estágio moderado. (B) Genes diferencialmente expressos em células mononucleares do sangue periférico (PBMC) de pacientes com dengue em estágio leve.

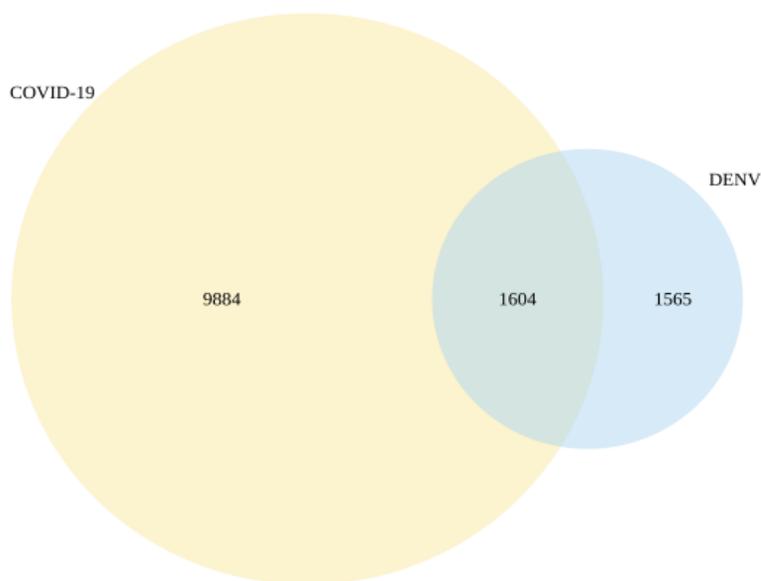


Figura 16. Gráfico de Venn mostrando a intersecção entre os genes diferencialmente expressos entre os grupos de Covid-19 e dengue.

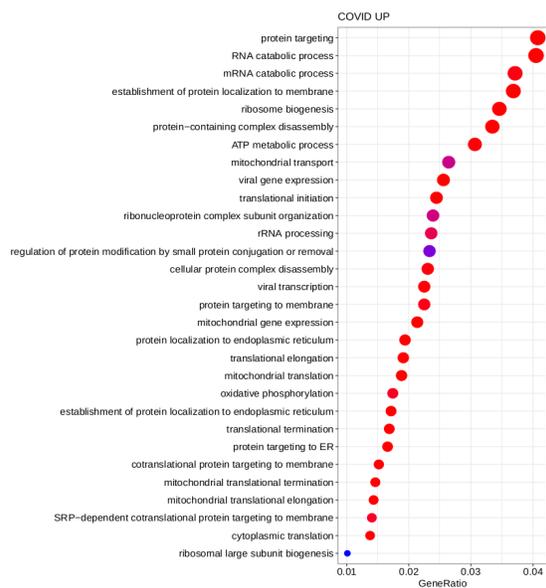
4.3.2 Análise funcional

A análise das ontologias gênicas também revelou diferenças na resposta do organismo às duas doenças virais. Para analisar a expressão dos genes com aumento ou diminuição na expressão, dividimos conjuntos de genes com aumento ou diminuição relacionados a cada doença e verificamos o enriquecimento de termos GO relacionados a processos biológicos, função molecular e componentes celulares através do pacote R *clusterProfiler* (Figuras 17 e 18).

Primeiramente, identificamos os transcritos que apresentaram aumento na expressão. Para garantir a significância estatística dos resultados, filtramos os termos com $FDR < 0,01$. Na Dengue, observamos termos que evidenciam uma resposta de defesa contra o vírus, como indicado pelos termos relacionados com a resposta ao vírus, tais como "*response to virus*", "*defense response to virus*", "*response to type 1 interferon*" e "*cellular response to virus*". Já na Covid-19, não observamos uma relação de aumento de transcritos com defesa viral, mas sim com o metabolismo do hospedeiro, modulação de processos mitocondriais e replicação viral (Figura 17). Dessa forma, é possível perceber uma diferença em relação aos processos imunes de combate ao vírus, em células imunes, que podem estar afetados na Covid-19, principalmente com relação aos IFN de tipo 1 e 2.

Em um segundo momento, caracterizamos os transcritos associados à intersecção dos conjuntos diferencialmente expressos entre as duas doenças. Transcritos que apresentaram aumento na expressão na Covid-19, mas diminuição na Dengue foram encontrados com padrões similares aos de genes regulados positivamente, porém com indicações de processamento de RNAs via ncRNA (RNA não codificador) e degradação mediada por mutação com mudança de sentido (NMD, sigla em inglês para nonsense-mediated decay). Termos de GO em intersecção entre transcritos aumentados nas duas viroses tiveram enriquecimento relacionados à epigenética, em especial, processos relacionados à modificação de histonas, organização da cromatina e termos relacionados com a sinalização de NF-kB (Figura 18). Resposta inflamatória sinérgica de NF-kB entre as viroses, mas resposta ao IFN-1 apenas em dengue, mostram resposta inflamatória que não tende a diminuir carga viral, mas que contribui para o agravamento da COVID-19.

A



B

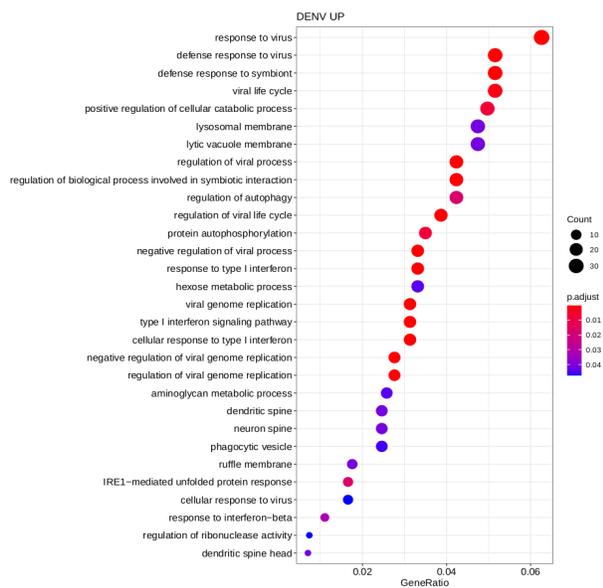
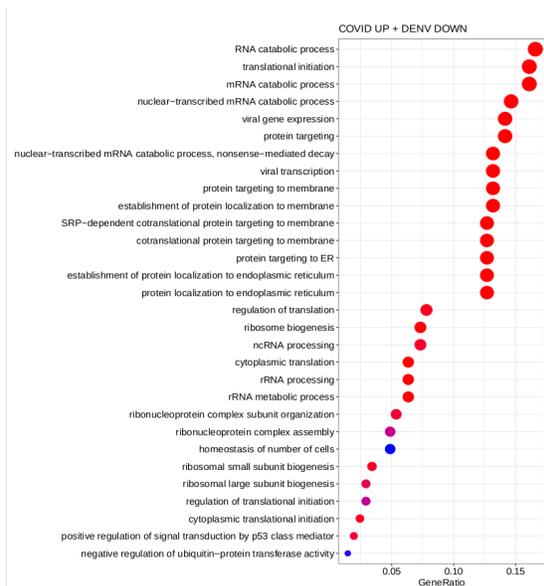
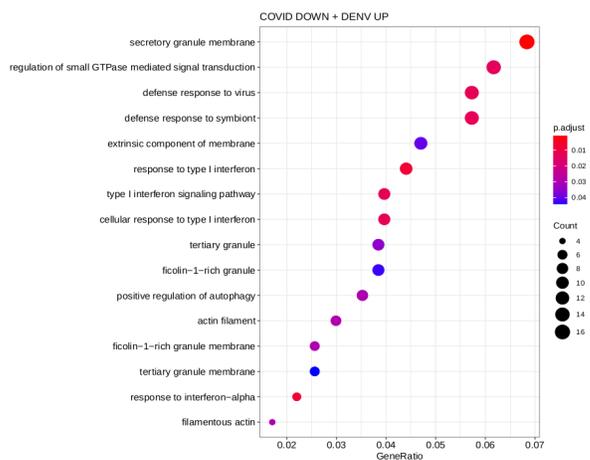


Figura 17. 30 termos mais enriquecidos em genes aumentados. (A) Termos enriquecidos na COVID-19. (B) Termos enriquecidos na dengue.

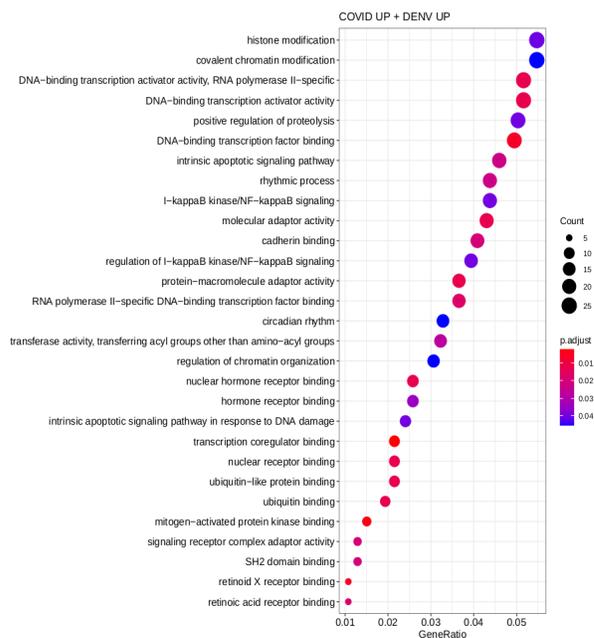
A



B



C



D

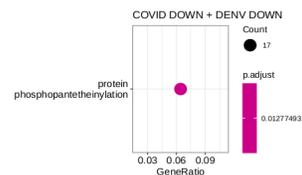


Figura 18. 30 termos mais enriquecidos entre as interseções de genes diferencialmente expressos de COVID-19 e dengue. (A) Enriquecimento de termos de genes aumentados na COVID-19 e diminuídos na dengue. (B) Enriquecimento de termos de genes aumentados na dengue e diminuídos na COVID-19. (C) Enriquecimento de termos de genes aumentados em dengue e COVID-19. (D) Enriquecimento de termos de genes diminuídos em dengue e COVID-19.

Ao realizar a análise de vias com o Reactome Pathways dos genes que apresentaram aumento de expressão na Dengue, mas diminuição na Covid-19, foram obtidas duas vias enriquecidas com $FDR < 0.05$. Via enriquecida relacionada com a sinalização do IFN ($FDR = 5.94E-3$) e pela sinalização de IFN-1 ($FDR = 0.0179$), que contém 15 genes representativos, dos quais, 7 estão representados na tabela 3. De acordo com a literatura, os genes IFITM1, IFITM2, IFIT1, IFIT2, MX2, TYK2 e ADAR desempenham um papel importante no controle da resposta imune mediada pelo IFN-1.

As proteínas IFITM (*Interferon-induced transmembrane proteins*) são uma família de proteínas que dificultam a entrada e a replicação viral ao desfazer a membrana de fusão do vírus com a célula do hospedeiro (Shi et al., 2021; Zhao et al., 2019). IFITM1 teve diminuição de sua expressão em leucócitos de pacientes em casos graves de influenza e foi associado com mortalidade, mas não foi achado diminuído em PBMC de casos graves de COVID-19, assim como IFITM2 (Regino-Zamarripa et al., 2022). Neste trabalho, a diminuição encontrada com relação a IFITM1 e IFITM2 pode estar relacionada com casos moderados ou o reteste positivo dos pacientes.

O gene IFIT2 (*interferon-induced protein with tetratricopeptide repeats 2*) também atua como inibidor do ciclo viral e pode inviabilizar o SARS-CoV-2 ao atuar sobre o seu mRNA, além disso, tem papel protetivo da neuropatogênese do vírus (Li et al., 2021). Em outro estudo foi mostrado ser alvo da proteína viral spike, o que diminuiu sua expressão em ensaio de imunoprecipitação em células HEK293T, sua diminuição já foi mostrada em plasma de pacientes em estágio grave (Freitas et al., 2022).

A expressão de TYK2 (tyrosine kinase 2), um importante indutor da sinalização de IFN do tipo 1, também já mostrou estar diminuída em amostras de sangue venoso de pacientes do sexo masculino com a Covid-19 (Akbari et al., 2022).

O gene ADAR (adenosine deaminase acting on RNA) é responsável pela edição de transcritos do vírus e do hospedeiro e tem mostrado um importante papel imunoregulador em diferentes doenças. Alguns estudos sugerem que a diminuição do gene ADAR pode aumentar os níveis de inflamação durante a infecção pelo SARS-CoV-2 (Crooke et al., 2021). A diminuição da edição nos sítios ALU por ADAR1 é um mecanismo imune que leva ao aumento do dsRNA em regiões 3' UTR de transcritos e à ativação da resposta inflamatória

pela activação da resposta do IFN através do reconhecimento do dsRNA pelo RIG-I, MDA5 e TLR3 (Philip et al., 2022).

Tabela 3. Genes com expressão aumentada na dengue, mas diminuídos na COVID-19, e enriquecidos no Reactome pathway com relação à via de sinalização do IFN.

| Gene | Ensembl | Log2 FC Covid | FDR |
|--------|-----------------|---------------|--------------|
| ADAR | ENSG00000160710 | -0.5430925 | 0.0002054483 |
| IFITM2 | ENSG00000185201 | -1.109285 | 0.0001651737 |
| IFIT1 | ENSG00000185745 | -1.3192614 | 0.02 |
| IFIT2 | ENSG00000119922 | -1.044371 | 0.01648842 |
| MX2 | ENSG00000183486 | -1.192354 | 7.723685e-05 |
| IFITM1 | ENSG00000185885 | -0.3980738 | 0.04091603 |
| TYK2 | ENSG00000105397 | -0.5923423 | 1.98271e-06 |

4.3.3. SPLICING ALTERNATIVO

Em contexto de transcrito de células imunes em pacientes infectados pelo SARS-CoV-2, os eventos de SA são aberrantes e a assinatura proteômica do splicing se mostrou reduzida, como foi mostrado em Wang et al (2022). O vírus também já mostrou ser capaz de desregular maquinaria de splicing através das suas NSPs (Banerjee et al., 2020). Com adição à análise de expressão de PBMC, se buscou relação de transcritos com splicing diferencial que possam estar relacionados com o escape imune e o aumento desenfreado da resposta inflamatória. Aqui, foi utilizado o programa rMATS para contar as diferentes categorias de splicing (retenção de intron, éxons mutuamente exclusivos, sítio alternativo 3', sítio alternativo 5' e exclusão de exon), considerou-se ambos eventos anotados e não anotados com $FDR < 0.05$. Foi observado um aumento expressivo de eventos de splicing na Covid-19, em comparação com o vírus da dengue, para todas as categorias de eventos de SA (Figura 19). A razão de aumento foi de mais de 125 vezes para exclusão exônica (SE) e 19 vezes para MXE (éxons mutuamente exclusivo). De acordo com a literatura, o SA em ambas as viroses

podem estar comprometidas pelas proteínas virais, necessitando verificar se esses eventos podem estar comprometidos em células imunes, na patologia da Covid-19.

Apesar do conhecimento sobre a interferência nos processos de splicing pelo SARS-CoV-2, ainda não foi elucidado como esse mecanismo pode afetar os leucócitos. Para verificar em contexto global como as isoformas presentes podem afetar o paciente, no curso da doença, foram verificados termos enriquecidos na biblioteca MsigDB 2020, que indicaram modulação imune relacionada aos transcritos contendo pelo menos uma das cinco isoformas com $FDR < 0.05$ e $\Delta\Psi > 0.1$. A tabela 4 indica os termos enriquecidos contendo os termos, o valor-p, a razão de chances (do inglês Odds ratio) e o escore combinado. Dentro os termos se destacam “Mitotic Spindle”, “*interferon alpha response*” e “*interferon gamma response*”. Apesar do aumento de SA observado em PBMC e de saber que o vírus compromete maquinaria de splicing, é evidente a necessidade de validação de que esses vias de fato estejam sendo prejudicadas pela ação do vírus, portanto, foram analisado com auxílio do website Ensembl e do programa IGV (v2.13.2), alguns genes com splicing comprometido na COVID-19, que estavam comprometidos em outros estudos (em células infectadas) e genes relacionados com a via dos IFNs (Figuras 20-21).

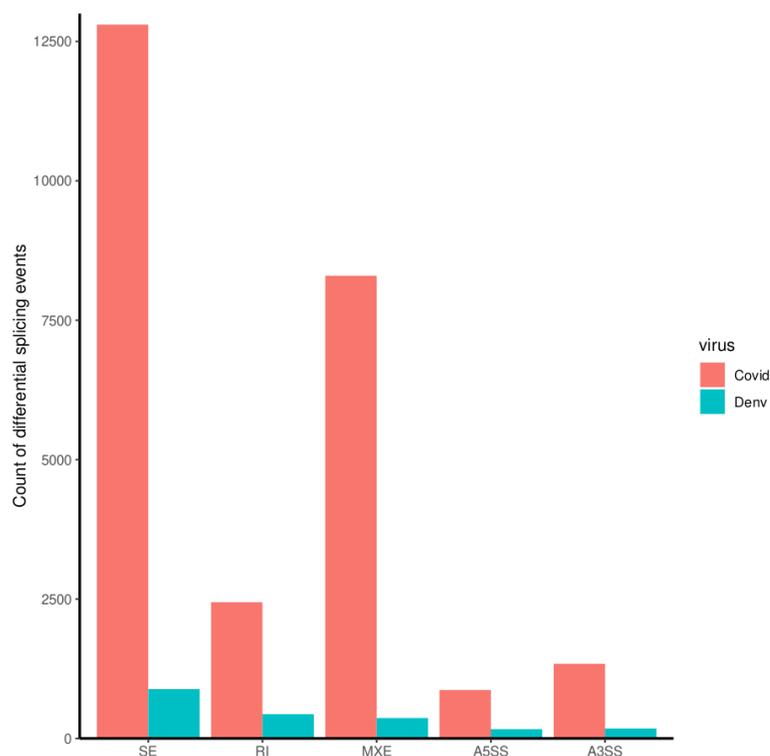


Figura 19. Contagem de eventos de splicing alternativo (SA). Comparação entre a contagem das principais categorias de SA entre a COVID-19 e dengue. Apenas os eventos de SA significativos (FDR<0.05) foram representados.

Tabela 4. Termos enriquecidos no MDSig2020 pelos genes com SA.

| Index | Name | P-value | Adjusted p-value | Odds Ratio | Combined score |
|-------|---------------------------|------------|------------------|------------|----------------|
| 1 | Mitotic Spindle | 1.406e-12 | 7.031e-11 | 2.94 | 80.31 |
| 2 | Interferon Gamma Response | 8.803e-10 | 2.201e-8 | 2.57 | 53.53 |
| 3 | Interferon Alpha Response | 4.049e-8 | 6.748e-7 | 3.25 | 55.37 |
| 4 | Myc Targets V1 | 0.00002064 | 0.00002581 | 2.10 | 27.51 |
| 5 | DNA Repair | 0.00002503 | 0.0002503 | 2.13 | 22.55 |
| 6 | Protein Secretion | 0.00005098 | 0.0004249 | 2.44 | 24.10 |
| 7 | G2-M Checkpoint | 0.0004602 | 0.002877 | 1.74 | 13.35 |
| 8 | heme Metabolism | 0.0004602 | 0.002877 | 1.74 | 13.35 |
| 9 | Apoptosis | 0.001297 | 0.007090 | 1.75 | 11.65 |
| 10 | Allograft Rejection | 0.001418 | 0.007090 | 1.65 | 10.84 |

O gene CD74 (ENSG00000019582) foi um dos genes com AS encontrados em células do tecido de pulmão de pacientes com a COVID-19 grave, no post mortem, em que o aumento do SE correlaciona com a diminuição da imunidade antiviral (Wang et al., 2022). O transcrito também possui SE nas células de PBMC de COVID-19 analisadas neste estudo, do mesmo éxon 7 (Figura 20). A proteína funcional codificada pelo gene CD74 se associa com MHC de classe II e regula a apresentação de antígenos durante a resposta imune, além disso, já apresentou função anti viral em infecções por ebola e COVID-19 ao bloquear via de entrada do virion (Bruchez et al., 2020). O SE do E7 forma proteína truncada ao perder domínio da tireoglobulina, a hipótese é a de que a perda desse exon facilite a entrada viral. Verificamos também que o SA relatado não ocorre na dengue, o que fortalece uma possível assinatura dessas isoformas na COVID-19, ocorrendo não só em células do pulmão, mas também em células de PBMC.

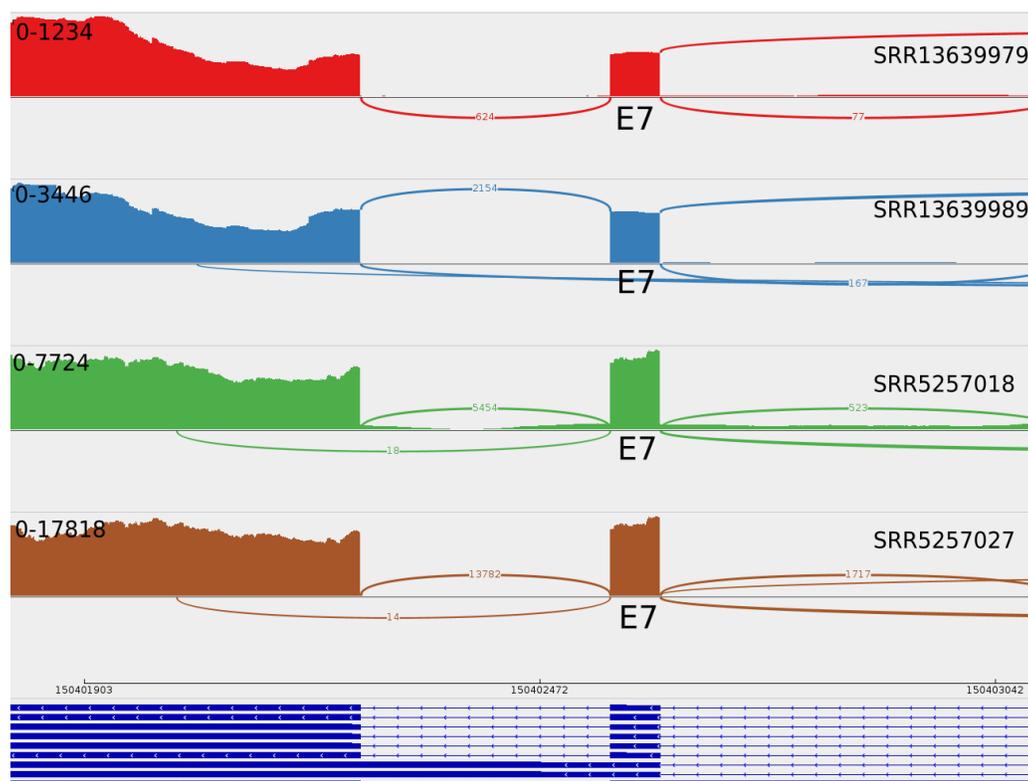


Figura 20. Isoforma do gene CD74. Ilustração do evento de *skipping exon* do éxon 7 do gene CD74. $\text{diffPSI} = -0.131$; $\text{FDR} < 0.01$ na Covid-19. Não houve diferença significativa em pacientes de dengue. Na parte inferior é mostrado as isoformas anotadas do gene, na região genômica correspondente. Amostras em vermelho: Paciente saudável. Azul: Covid-19. Verde: Controle da Dengue. Marrom: Dengue.

O gene MX2 (ENSG00000183486) é uma proteína com propriedades antivirais estimulada pelo interferon do tipo 1, o MX2 foi mostrado diminuído na Covid-19 e aumentado da dengue (Figura 21), o transcrito teve SE do E6, a perda desse exon está relacionada com o transcrito MX2-204, que leva a perda de função e decaimento pela via de NMD, essa perda pode explicar em parte a diminuição em sua expressão e foi o único dos citados na tabela 3 a possuir AS diferencial significativo.

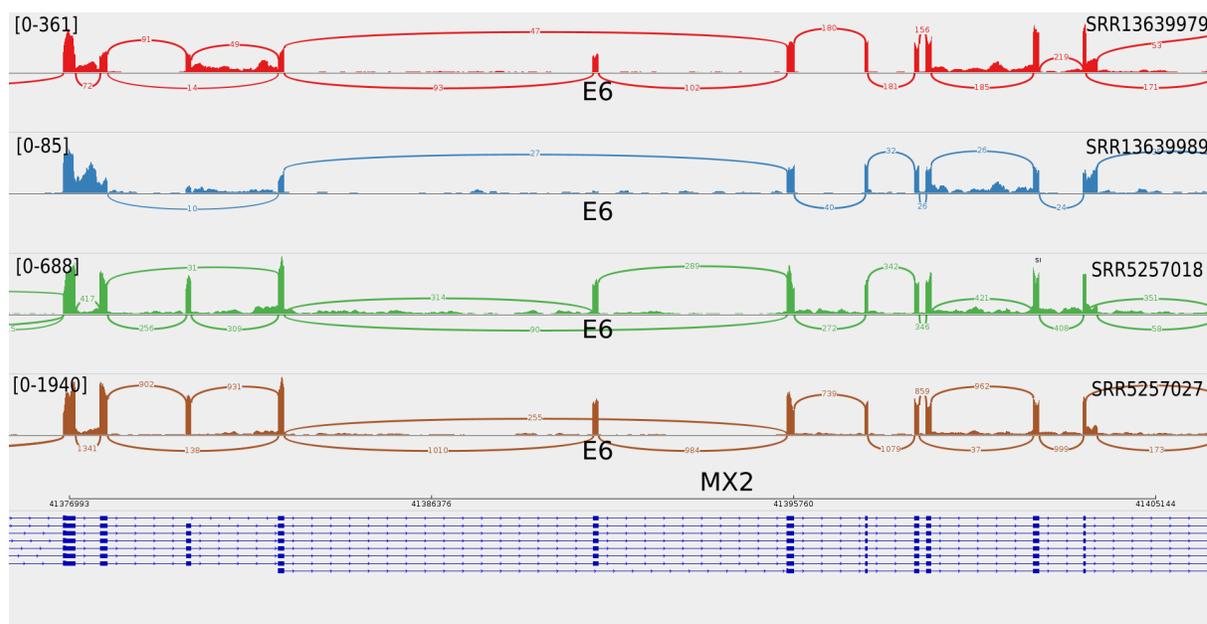


Figura 21. Isoforma do gene MX2. Ilustração do evento de *skipping exon* do éxon 6 do gene MX2. diffPSI = -0.294; FDR < 0.01 na Covid-19. Não houve diferença significativa em pacientes de dengue. Na parte inferior é mostrado as isoformas anotadas do gene, na região genômica correspondente. Amostras em vermelho: Paciente saudável. Azul: Covid-19. Verde: Controle da Dengue. Marrom: Dengue.

4.4. Monócitos

4.4.1. Expressão e análise funcional

Os genes diferencialmente expressos (DEGs) de monócitos circulantes de pacientes com covid-19 leve e grave foram analisados e 839 genes foram encontrados aumentados, enquanto 402 foram diminuídos em pacientes em estado grave da doença e 565 aumentados e 366 diminuídos em estado leve. Em concordância com a literatura, a expressão do gene HLA-DRA foi encontrado diminuído com o aumento do nível da gravidade e, também, HIF-1 α (*Hypoxia-inducible factor 1-alpha*) estava aumentado na covid-19 grave, o que sugere resposta à hipóxia, além disso, HIF-1 α é repressor do IFN do tipo 1 e interfere na resposta antiviral (Peng et al., 2021). A expressão de ADAR foi encontrada aumentada em monócitos de pacientes com covid leve e grave (Figura 22A e 22B), mas não em PBMC de pacientes, como mostrado em estado moderado (Figura 15A) da doença e, igualmente, em estado de convalescença, indicando aumento de expressão pelo menos em monócitos, mas não em outros tipos celulares. Além disso, o STAT3, um indutor da ADAR1, foi encontrado aumentado em ambas covid-19 leve e grave. Em monócitos de pacientes infectados, STAT3 já foi relatado ter papel na doença ao causar uma resposta inflamatória pró-viral ao elevar citocinas pró-inflamatórias, mas inibir resposta do IFN-1 e sequestrar STAT1, um mediador da resposta do IFN-1, que também é alvo da ORF6 e NSP1 do SARS-CoV-2 (Jafarzadeh; Nemati; Jafarzadeh., 2021). Os monócitos possuem importante papel na patologia da COVID-19, tendo sido mostrado estarem desregulados o que contribui para a resposta inflamatória sistêmica e enfraquecimento da resposta imune adaptativa, levando à persistência viral (Knoll; Schultze; Schulte-Schrepping., 2021). Diferentes marcadores transcricionais durante o curso da doença, em monócitos, foram propostos e verificados, entre esses, a diminuição de HLA-DRA, uma cadeia do HLA de classe II HLA-DR. Também foi verificado aumento da expressão de HIF-1 α , sinalizador de hipóxia que tem mostrado facilitar a replicação viral em monócitos, a hipóxia em si também vem sendo mostrada nessas células, o que pode prejudicar o funcionamento da célula e a resposta do IFN, que fica afetada (Peng et al., 2021, Codo et al., 2020).

Enquanto já se demonstrou que o aumento da hipóxia eleva os níveis de ER pela ADAR, pela consequente diminuição do pH que causa aumento de sua atividade catalítica (Malik et al., 2021), a ER por ADAR1 também eleva a sinalização de hipóxia ao estimular HIF-1 α através da inibição de seus repressores transcricionais (Ma et al., 2019). O acúmulo de HIF-1 α pode facilitar a replicação viral em monócitos e sua inibição mostrou bloquear o ciclo viral (Codo et al., 2020). IL-6, uma das marcas transcricionais relacionadas com a tempestade de citocinas, juntamente com ADAR1 são capazes de levar à hiperativação da via STAT3 (Teoh et al., 2020), que também é capaz de suprimir a resposta antiviral via IFN-1, mas induz resposta inflamatória que pode contribuir para a tempestade de citocinas e também contribui para o aumento da fibrose pulmonar e trombose (Jafarzadeh et al., 2021). Além disso, STAT3 e ADAR1 têm um mecanismo de retroalimentação positiva na qual STAT3 eleva a transcrição de ADAR1 e ADAR1 estabiliza STAT3, aumentando a sua concentração celular (Jafarzadeh et al., 2021). Assim, tanto a condição de hipóxia encontrada em monócitos como o aumento de STAT3, que estimula a expressão do ADAR1, podem explicar o aumento da edição encontrada e, além disso, podem ajudar a compreender a desregulação dos monócitos, na Covid-19, ao facilitar a entrada e a replicação viral e ao contribuir para o aumento da inflamação.

Para análise funcional foram utilizadas as bibliotecas LINCS-L1000 e HsigDB 2020, para verificar a assinatura molecular dos transcritos diferencialmente expressos, em monócitos de pacientes, entre os estados leve e grave. As análises de enriquecimento foram realizadas com o score combinado (*ver métodos*) e a significância foi estimada utilizando o teste unilateral de Mann-Whitney (U). Na Figura 23A cada ponto representa a correlação dos genes diferencialmente expressos encontrados neste estudo com outros fornecidos pelo LINCS-L1000 para cada tipo de assinatura. Em relação à resposta ao interferon, houve diminuição significativa da resposta ao IFN- α , por outro lado, não houve diferença significativa da resposta ao IFN- γ e TNF- α , apenas um estudo de expressão de IL-6 estava enriquecido, nos estados leve e severo, e mostrou aumento de resposta ao IL-6 aumentado nos monócitos de pacientes em estado grave.

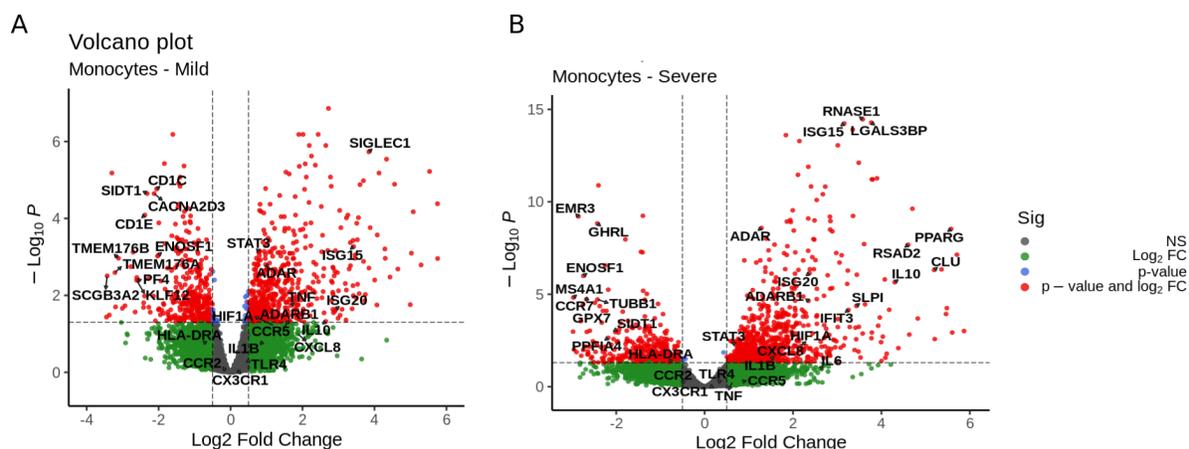


Figura 22. Gráfico de vulcão com a dispersão do *fold change* pelo *FDR*. (A) Genes diferencialmente expressos em monócitos de pacientes com COVID-19 em estágio leve. (B) Genes diferencialmente expressos em monócitos de pacientes com COVID-19 em estágio grave.

Analisando os perfis de assinatura com MSigDB foi encontrada assinatura de genes relacionados com hipóxia em covid-19 severa, mas não em Covid-19 leve, as assinaturas de resposta ao IFN corroboram com as comparações feitas com LINCS-L1000 em relação a ambos IFN- α e IFN- γ , porém, mesmo com diminuição da resposta inflamatória via IFN, houve simultâneo aumento da assinatura de resposta inflamatória e sinalização do TNF- α via NF-kB (Figuras 23A e B). Esses dados corroboram com a diminuição da resposta do IFN- α e aumento da hipóxia em monócitos de pacientes com COVID-19, já verificada em outros trabalhos (Peng et al., 2021). Esses resultados indicam um aumento da resposta inflamatória pelos monócitos de pacientes em estado grave, mas não pelas vias antivirais do IFN-1, portanto, esse padrão de enriquecimento pode explicar, em parte, uma possível via que facilite a infecção dos monócitos e a persistência viral com aumento da inflamação, em alguns pacientes, a partir da desregulação de uma resposta antiviral adequada e com aumento da resposta inflamatória via NF-kB.

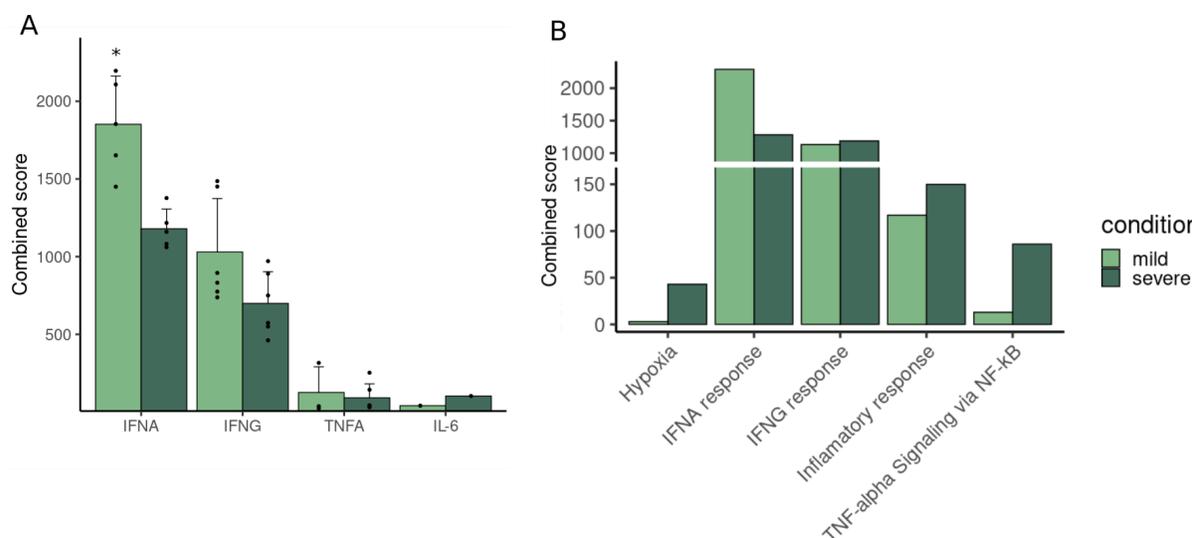


Figura 23. Análise de enriquecimento com enrichR dos monócitos nos estados leve e grave. (A) Cada ponto representa o escore combinado de cada um dos estudos relacionados com a assinatura da expressão do gene indicado no eixo x com a expressão de genes diferencialmente expressos nos monócitos. * $p < 0.05$. (B) Assinatura MsigDB indicada pelo escore combinado da expressão dos genes diferencialmente expressos.

4.4.2. Edição de RNA

As edições de RNA foram investigadas com o software SPRINT. A maioria das edições de RNA encontradas, na distribuição das edições (tabela 5), foram de A para I, em todos os grupos analisados, representadas como A-to-G ou T-to-C (Lo Giudice et al., 2019).

Posteriormente, filtramos as edições com pelo menos 5% de base trocada, em relação ao genoma de referência, e *reads* > 10 para cada posição como em Philip et al. (2022).

Tabela 5. Distribuição de edições para cada grupo. Contagem referente às edições que passaram no filtro de edição e presentes em todas as amostras do grupo. Contagem do controle entre parênteses.

| Edição | PBMC - COVID-19 | Monócitos - COVID-19 severa | Monócitos - COVID-19 leve | Monócitos infectados <i>in vitro</i> | Monócitos clássicos - COVID-19 | CD - COVID-19 grave |
|--------|-----------------|-----------------------------|---------------------------|--------------------------------------|--------------------------------|---------------------|
| AG | 206(485) | 989(596) | 680(596) | 1693(2362) | 139(120) | 780(1039) |
| TC | 317(701) | 1040(609) | 824(609) | 1528(1754) | 130(148) | 500(689) |
| CT | 8(5) | 10(10) | 5(10) | 24(37) | 0(0) | 60(48) |
| GA | 10(0) | 0(0) | 0(0) | 21(36) | 2(0) | 58(55) |

| | | | | | | |
|----|------|------|------|------|------|------|
| GT | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 3(0) |
| AT | 3(0) | 0(0) | 0(0) | 1(3) | 0(0) | 1(1) |
| TG | 0(0) | 0(0) | 0(0) | 8(6) | 0(0) | 9(0) |

4.4.2.1. Edições gerais nas amostras

Em conformidade com a literatura houve diminuição do Index de ER em Covid-19 (rooke et al., 2021), que foi confirmada, neste estudo, em PBMC convalescente (Figura 24A), células dendríticas infectadas in vitro (Figura 24D) e monócitos infectados in vitro (Figura 24C). No entanto, monócitos de pacientes com Covid-19 leve ou grave e de monócitos clássicos, de pacientes sem estado da doença definido, mostraram aumento do index de ER (Figuras 24B e E).

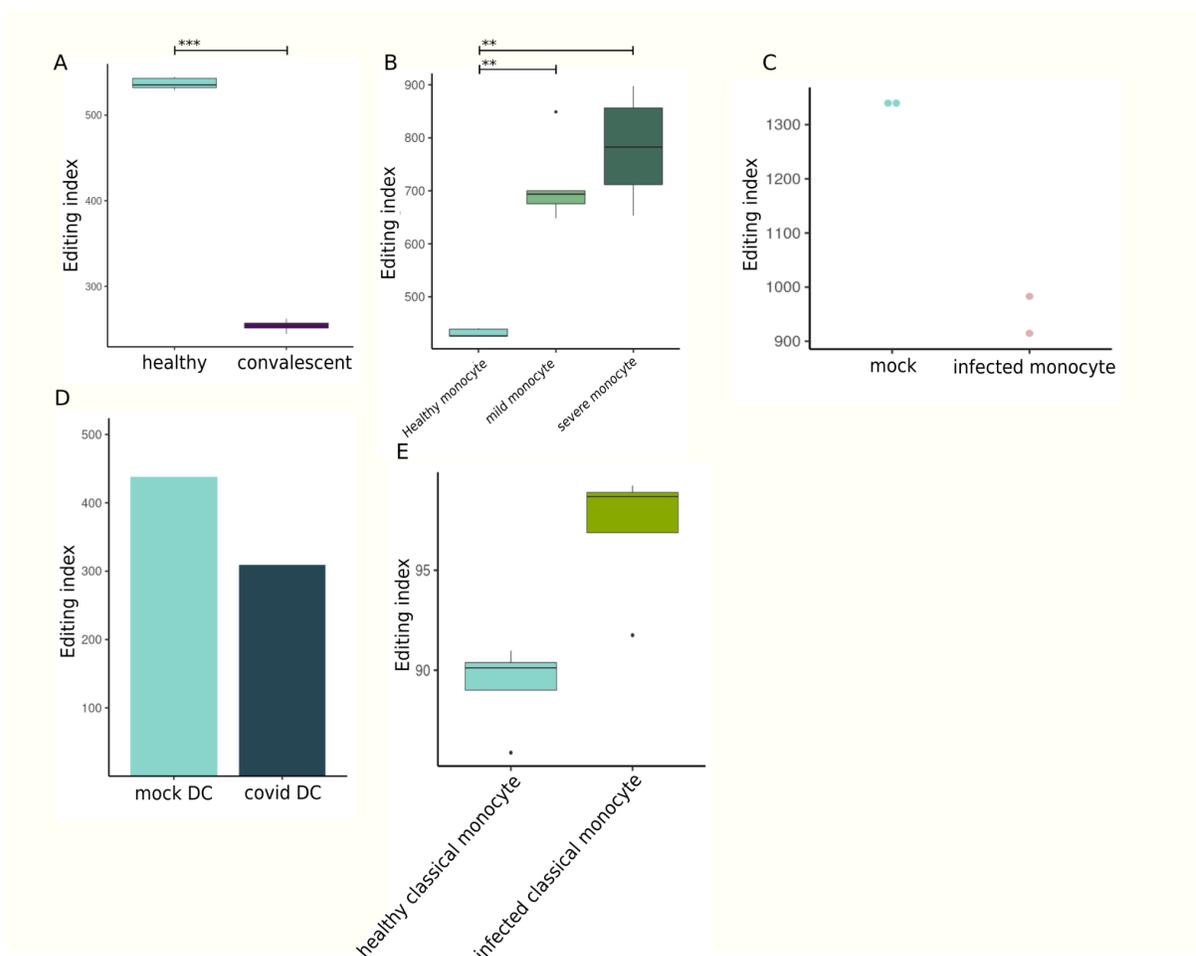


Figura 24. Índice de edição de RNA (ER) no tratamento e controle. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Teste t com correção com correção de Welch. (A) Índice de ER em PBMC de pacientes em estado convalescente da COVID-19. (B) Índice de ER em monócitos de pacientes nos estados leve e severo da COVID-19. (C) Índice de ER em monócitos infectados *in vitro* de pacientes nos estados leve e severo da COVID-19. (D) Índice de ER em células dendríticas (CD) infectadas *in vitro*. (E) Índice de ER em monócitos clássicos (CD14+) de pacientes com COVID-19.

4.4.2.2. Regiões editadas no transcrito

A sobreposição das regiões editadas com pelo menos 1% de edição, verificadas contra anotação de transcritos e utilizando o programa OLOGRAM (Ferré et al., 2019), mostrou que as edições estão significativamente sobrepostas em regiões UTR, enquanto que as regiões relacionadas com CDS (*coding sequence*) estavam sub-representadas (Figura 25), o que corrobora com as regiões UTR editadas descritas na literatura.

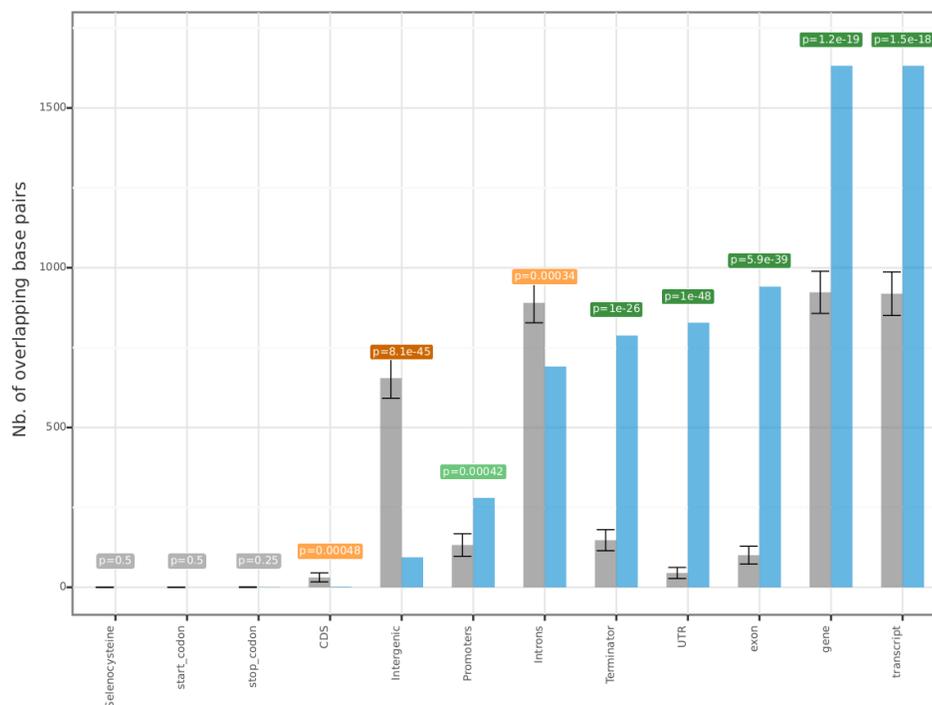


Figura 25. Significância da sobreposição de regiões editadas nas regiões de transcritos anotados. Cinza: regiões de sobreposição aleatória. Azul: sobreposição de de regiões editadas.

As edições de RNA com pelo menos 1% de edição, entre cada grupo, foram encontradas principalmente nos sítios ALU anotados. Regiões ALU anotadas também são descritas na literatura (Picardi et al., 2017). Essas regiões cobrem pelo menos 90% das edições em comparação com elementos não repetitivos de ALU (REP) ou elementos não repetitivos (NONREP) (Figura 26) em edições anotadas obtidas pelo rediportal (<http://srv00.recas.ba.infn.it/atlas/>) (Lo et al., 2020). A diminuição da edição em sítios ALU pela ADAR leva ao aumento de dsRNA endógeno, e ativa resposta inflamatória dos interferons pela via de reconhecimento de dsRNA pela RIG-I, MDA5 e TLR3 (Philip et al., 2022). Em certo nível, a ativação da via de reconhecimento do dsRNA endógeno por pequena diminuição da edição global é necessária durante o curso da infecção viral, ao ativar mecanismo de resposta antiviral, como já foi mostrado em células de PBMCs de diferentes viroses como influenza (Li et al., 2016). Conforme o aumento da severidade, há também diminuição da edição em PBMC de pacientes infectados pelo SARS-CoV-2, é hipotetizado em outros estudos que isso possa levar a uma resposta inflamatória prejudicial em pacientes em estágios avançados, visto que a resposta do IFN nesses casos passa a gerar resposta danosa no

hospedeiro pelo alta resposta inflamatória infligida. Neste trabalho foram confirmados a diminuição global da edição em PBMC de pacientes convalescentes, células dendríticas e monócitos infectados *in vitro*, mas por outro lado, monócitos circulantes mostraram inesperado aumento de edição, em ambos casos leve e severo, mostrando um padrão diferente do analisado por outros trabalhos. A assinatura transcricional do estado de hipóxia encontrada nos pacientes infectados com COVID-19, em monócitos, (Figura 23B) pode explicar o aumento de edição nesse tipo celular, visto que é um dos mecanismos que leva ao aumento da ER pela ADAR e pode estar relacionada com a desregulação dessas células na resposta imune à doença.

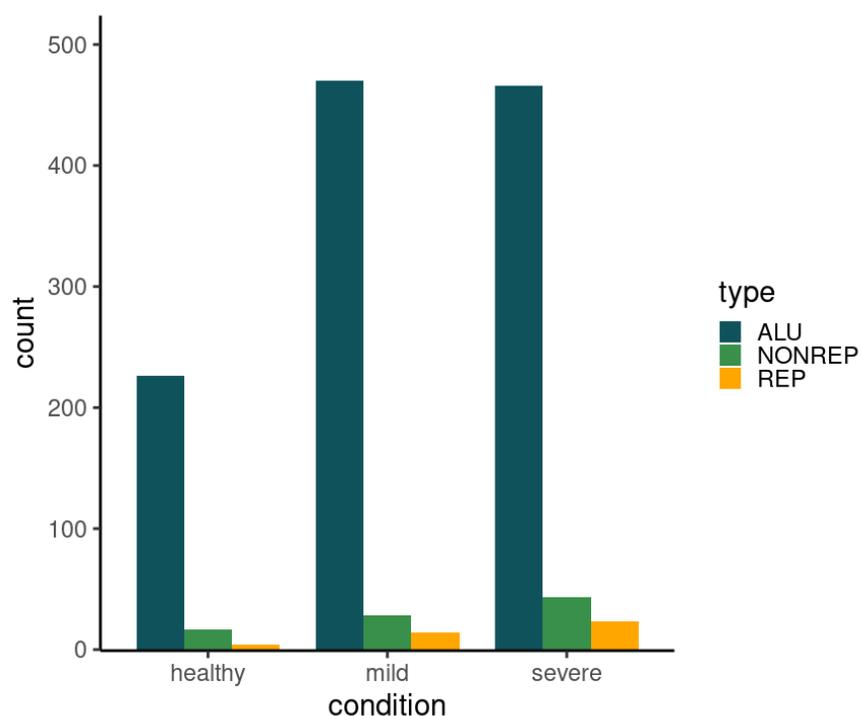


Figura 26. Contagem de regiões editadas em elementos ALU. As regiões editadas, e anotadas, de amostras de monócitos foram verificadas em relação a regiões contendo elemento ALU, repetições (REP) ou sítios sem repetições (NONREP).

4.4.2.3. Edições de RNA diferenciais em monócitos

Transcritos editados com pelo menos 5% de ER de monócitos infectados *in vitro* e monócitos de pacientes infectados com Covid-19 foram comparados. A interseção dos genes diferencialmente editados são vistos com maior frequência entre monócitos de doentes com

infecção leve e grave do que entre estes grupos contra células infectadas *in vitro* (Figura 27). Comparando os transcritos diferencialmente editados divididos, em aumento e diminuição da edição, foi observado uma maior quantidade de transcritos com aumento de edição em intersecção entre os grupos de pacientes com a Covid-19 leve e severa e na segunda maior intersecção entre diminuição de edição em células infectadas *in vitro* e aumento de edição em células de pacientes em estágio grave (Figura 28), o que sugere uma regulação diferencial entre os estados *in vivo* e *in vitro*. No entanto, houve também um aumento concordante de edições entre a Covid-19 grave e a *in vitro* na terceira maior intersecção e aumento de edição na condição severa em transcritos diminuídos na Covid-19 leve, na quarta posição (Figura 28). Assim, corroborando a nossa hipótese de que este aumento da edição está mais relacionado com as condições fisiológicas criadas pelo SARS-CoV-2 do que por infecção direta ou, pelo menos, por ambas as situações.

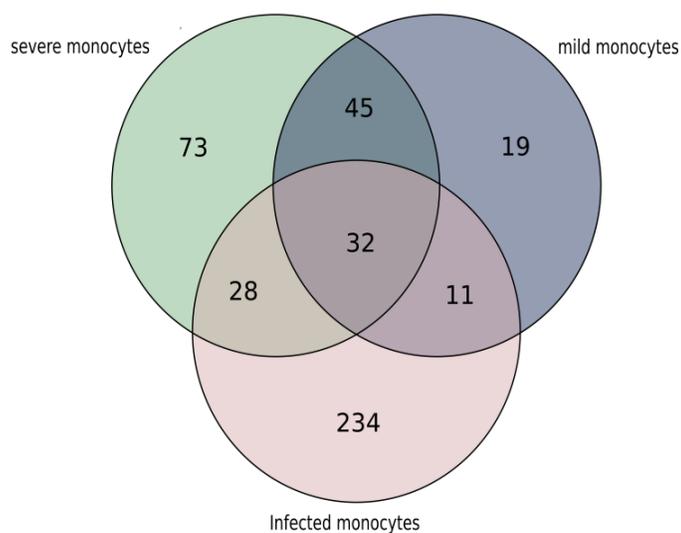


Figura 27. Comparação do número de edições de RNA (ER). Comparação dos conjuntos de ER para os grupos de monócitos de pacientes nos estágios leve, severo e infectado *in vitro*.

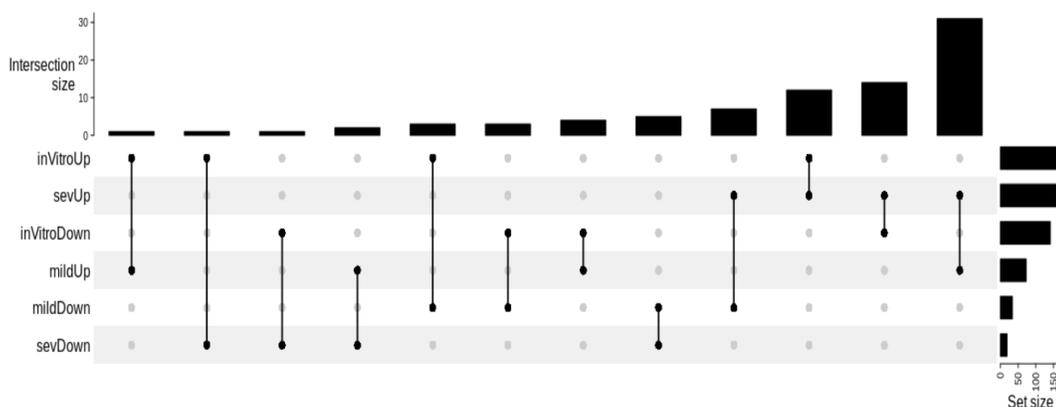


Figura 28. Comparação do número de edições de RNA (ER). Comparação dos conjuntos de ER para os grupos de monócitos de pacientes nos estágios leve, severo e infectado *in vitro*. Os conjuntos foram separados em edições aumentadas e diminuídas. São mostradas apenas as interseções, que estão classificadas por número de edições.

4.4.2.4. Análise funcional de genes de transcritos editados

Trabalhos anteriores mostraram a ER por ADAR envolvida no processo de resposta imune (Ahmad; Mu; Hur., 2021). Aqui, comparamos o enriquecimento de termos entre os grupos de monócitos com ER diferencial. Embora na análise geral tenham termos enriquecidos relacionados com a resposta imunológica, há uma diferença entre os termos enriquecidos entre os grupos em estudo. Análise de enriquecimento, com *Metascape*, entre monócitos *in vitro* e *in vivo* (condições leves e graves), dos genes editados, permitiu comparar a relação entre ontologias dos grupos em estudo, tanto em relação aos genes com aumento quanto com diminuição de edição. Termos que aparecem apenas na Covid-19 severa foram: Regulação da fosforilação, via reguladora da sinalização de resposta imune e via PDGFRB do PID, que interagem com outras ontologias de regulação da resposta ao estímulo biótico. Termos de sinalização de citocinas no SI, via de sinalização NF-kappa B e TNF aparecem em edição diferencial aumentada, mas também diminuída, nas três condições analisadas (Figura 29, 30 e 31). O que demonstra a participação da edição nos mecanismos de regulação imune. Do mesmo modo, percebemos integração entre genes relacionados com a infecção SARS-CoV-2 *in vitro* com as vias de sinalização do interferon, resposta ao vírus e sinalização, indicando um papel para a edição mediada por ADAR na infecção viral e resposta ao IFN.

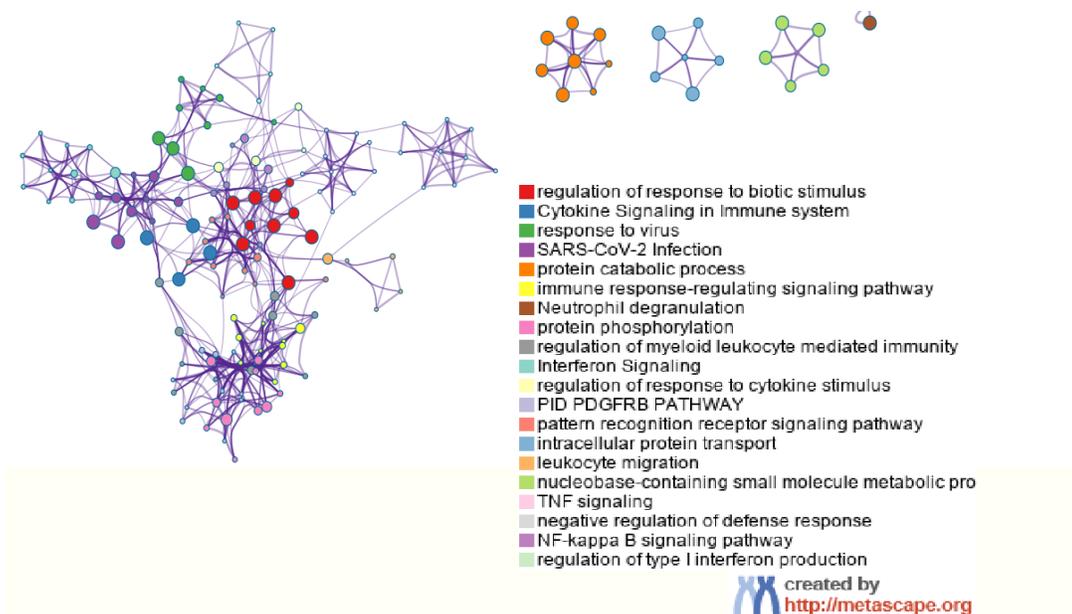


Figura 29. Análise de rede dos termos enriquecidos, em genes editados, feita com *Metascape*. Termos enriquecidos compartilhados entre os grupos de monócitos de pacientes infectados (estágios leve ou grave) ou infectados *in vitro*.

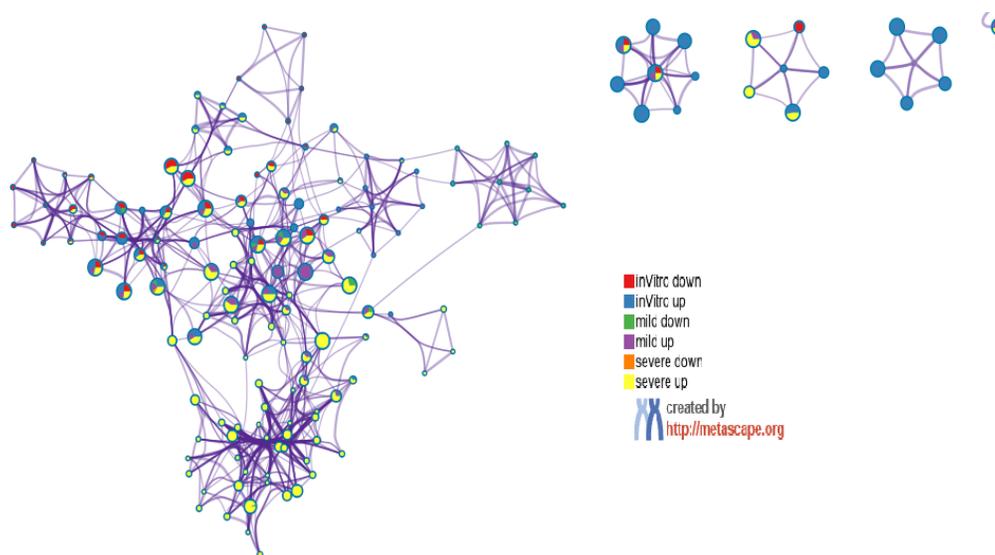


Figura 30. Distribuição dos grupos de edição. Cada cor representa um grupo de edição dos termos enriquecidos na figura 29.

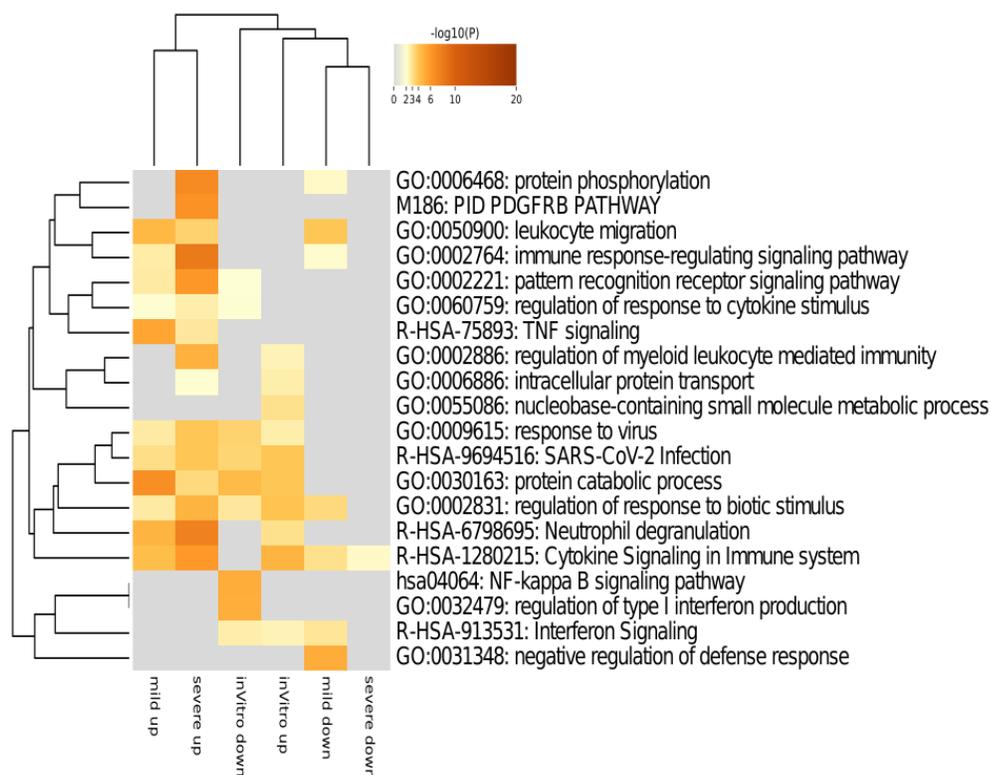


Figura 31. Comparação dos termos enriquecidos dos genes editados. Heatmap feito com *Metascape* para visualizar os termos mais enriquecidos em monócitos de pacientes em condição severa em comparação com as outras condições. Escala de cor indica a significância estatística.

4.4.2.5. Análise dos genes que possuem maior razão de edição de RNA

Para uma análise mais precisa das edições na COVID-19 grave, transcritos com alto número de edições, em diferentes sítios, na COVID-19 grave e no transcrito do gene MAVS, foram identificados nas amostras de monócitos e comparados (Figura 32). A maior parte desses genes selecionados possuem edições em regiões ALU e 3' UTR, tipicamente relacionadas com a formação de dsRNA. Aumento de edição nessas regiões indica diminuição da formação do RNA dupla fita, o que deve levar à supressão da ativação de IFN do tipo 1 pela via de reconhecimento de RNA dupla fita mediada por MDA5 e MAVS. Dos genes mostrados, na figura 32, MAVS é adicionalmente mostrado pelo fato de sua edição, na porção 3' UTR, já ter sido mostrado suprimir sua atividade (Li et al., 2020). Por mais que haja aumento de edição da MAVS nos monócitos infectados *in vitro*, a expressão em monócitos de pacientes infectados não teve expressão suficiente para se detectar níveis de edição (Figura 32).

Para entender o papel dos transcritos editados nas porções 3' UTR e intrônicas encontradas nos transcritos mais editados em COVID-19 grave, a edição inespecífica nessas regiões foram investigadas na literatura. Edição de DOCK2 em células-tronco de melanoma mostraram aumento de estabilidade e aumento de expressão com consequente ativação de RAC1 e fosforilação de Akt e NF- κ B, a edição nesta região também mostrou suprimir MDA5 via expressão de DOCK2 (Zhang et al., 2021). A edição da região 3' UTR em AhR (do inglês "*aryl hydrocarbon receptor*") em células derivadas de carcinoma hepatocelular Huh-7 criou sítio de ligação do miRNA miR-378, que leva à degradação do transcrito (Nakano et al., 2016). Na COVID-19 a ativação de AhR tem mostrado suprimir resposta antiviral durante a resposta imune inata pela via do IFN do tipo 1 (Giovannoni et al., 2020). Edição na posição 3' UTR de CTSS está relacionada com ligação da proteína HuR, proteína que pode estabilizar ou desestabilizar transcritos a depender da forma que se liga. No CTSS, a proteína HuR estabiliza o transcrito e mostrou aumentar sua expressão em células epiteliais, em Stellos et al. (2016), que verificou aumento da edição do transcrito por ambos casos: aumento da expressão de ADAR1 e pela indução de hipóxia na célula. O aumento de CTSS pode levar à clivagem do receptor de IL-6 (IL-6R) e induzir sua trans- sinalização, isso é, a IL-6R passa a poder ser integrado a outras células para interação com IL-6, podendo elevar a resposta desta interleucina em diferentes células e aumentar a resposta inflamatória, além disso, a catepsina S tem um papel importante nos processos inflamatórios vasculares por meio da degradação da matriz extracelular (Stellos et al., 2016; Flynn et al., 2020). O gene EIF2AK2 é um agente que previne a replicação viral, a sua edição cria alvo para os miRNAs miR-3622a-3p e miR-3622b-3p, que devem levá-lo à degradação (De Toma et al., 2021). Em doença de Parkinson APOL6 mostrou aumento de expressão com a edição de sua região 3' UTR através da perda da ligação dos miRNAs miR-19 and miR-4659, que possivelmente levariam à degradação do transcrito (Sijia et al., 2022, Preprint). O aumento de APOL6, na COVID-19, mostrou aumentar os níveis de espécies reativas de oxigênio (ROS), o que pode facilitar a infecção do vírus em monócitos (Codo et al., 2020; Zhaorigetu et al., 2011). ADAR1 também já foi mostrada editar posições do MDM2 tanto em linhagem celular WI38 quanto em células HEK293 quando comparadas com clones knockdown de ADAR1, a edição leva à competição de ligação com STAU1, que por sua vez induz retenção nuclear do transcrito, mostrando que ADAR1 pode interferir na distribuição espacial (Yang et al., 2016), MDM2 nuclear suprime

p53 via ubiquitinação, leva à sobrevivência celular e pode elevar os níveis de HIF-1 α , visto que p53 é seu inibidor, podendo a ADAR elevar os níveis de HIF-1 α diretamente e indiretamente pela edição de seus inibidores (Zauli et al., 2022). Por fim, todos os genes da figura 32 ocorrem na posição 3' UTR e podem levar à abertura de dsRNA, essa abertura das fitas sugere diminuição da ativação da resposta de IFN-1 via MAVS-MDA5, além disso, outros mecanismos induzidos pela edição podem diminuir a expressão de MDA5 e elevar os níveis de HIF-1 α . A maior parte das edições analisadas foram em regiões ALU (Figura 26), o que ressalta a importância dessas regiões para os mecanismos de edição.

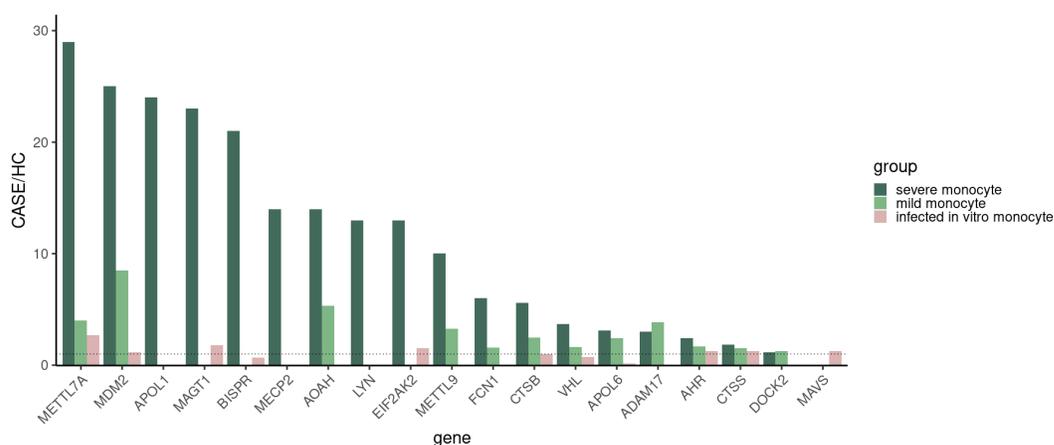


Figura 32. 18 genes com maior número de edições na COVID-19 grave e MAVS. A medida de ER é dada pela razão caso/controle. A linha pontilhada indica razão tratamento/controle igual a 1.

5. CONCLUSÕES E PERSPECTIVAS

Neste trabalho, avaliamos transcrito de leucócitos adquiridos de bancos de dados públicos. Para isso, utilizamos diferentes programas na análise dos principais eventos que podem ser extraídos de RNA-seq com seleção de RNAs pela cauda poli-A. Durante a execução do projeto, criou-se um pipeline com o gerenciador de pipeline *Snakemake*. O pipeline realiza análise estatística entre dois grupos distintos, incluindo análise diferencial de expressão gênica, lncRNA, splicing alternativo e edição de RNA. O pipeline foi pensado para integração desses eventos, permitindo uma análise complementar da análise de expressão convencional, e serão os próximos passos de seu desenvolvimento. Em um primeiro momento do projeto, foram realizadas análises PBMC de pacientes infectados pela COVID-19, para comparação com um controle, utilizou-se dados de pacientes infectados pela dengue devido a similaridade que essas viroses possuem em relação ao escape imunológico viral, que embargam processos relacionados com a resposta de IFN-1 e mecanismos de splicing. Em um segundo momento, foram utilizados dados de transcrito de monócitos, de pacientes infectados pela COVID-19, nos estados leve e grave, devido ao papel chave que essas células mostraram desempenhar no curso da doença ao desregular resposta imune.

A análise de PBMC mostrou enriquecimento de termos GO relacionados com a replicação viral em amostras de pacientes com a COVID-19 e pouca relação com resposta imune antiviral, isso fica evidente ao se comparar com os termos enriquecidos em dengue (tipo A), que mostrou resposta de IFN-1 e resposta de defesa contra o vírus em seus termos GO enriquecidos. Operações entre conjunto de genes diferencialmente expressos entre as viroses confirmaram essas diferenças ao se verificar os genes de cada conjunto conjunto menos sua interseção. Mostramos também, que dentre os genes com expressão aumentada em dengue, mas diminuída na COVID-19, houve enriquecimento de via relacionada com IFN-1. Esses genes: IFITM1, IFITM2, IFIT2, MX2, TYK2 e ADAR, possuem papel na atividade antiviral. Análise de genes contendo SA mostraram relação com assinatura de IFNs. Em especial, verificamos os genes diminuídos relacionados com IFN-1, que estavam com expressão aumentada em dengue, dentre eles, MX2 apresentou ES do E6, dando origem ao transcrito MX2-204, que é degradado pela via NMD. CD74 estava com diminuição do splicing do E7, outro estudo já mostrou a correlação desse splicing com diminuição da

atividade antiviral em células de pulmão. Aqui, reforçamos essas isoformas em PBMC, que levam à perda de função destas proteínas.

Análise de expressão de monócitos circulantes de pacientes infectados, em estado severo, mostrou padrões de assinatura relacionada com a diminuição da resposta de IFN-1 em relação aos casos leves, correlação com estado de hipóxia, aumento da resposta inflamatória via NF-kB e padrões de edição de RNA que indicam diminuição da resposta de IFNs e podem ajudar a compreender a desregulação e infecção desse tipo celular. O aumento da edição nessas células é contrária ao indicado na literatura pelas células imunes. Encontramos resultados similares aos de outros estudos em PBMCs e células dendríticas, mas monócitos não haviam sido verificados ainda. O aumento de hipóxia nessas células, na infecção por COVID-19, já foi descrito e se relaciona tanto com o aumento de ED, podendo explicar esse aumento diferenciado de outros leucócitos, e se relaciona, também, com aumento de HIF-1 α . HIF-1 α foi relacionado com a infecção dos monócitos em outros trabalhos, neste trabalho foi verificado seu aumento significativo em monócitos de pacientes em estado grave. ADAR, o principal editor de RNA já foi relacionada com o aumento de HIF-1 α e pode contribuir para a patogenicidade do SARS-CoV-2 por contribuir para o aumento de ambos HIF-1 α e STAT3, fator de transcrição pró-inflamatório, que embarga resposta do IFN-1. Grande parte dos genes editados pela ADAR se localizam em regiões ALU e podem estar relacionados com diminuição de dsRNA endógeno, essa perda diminui a resposta do IFN-1 via MAVS-MDA5. Genes mais editados possuem relação com resposta imune e podem ser controlados via edição de suas regiões 3' UTR, grande parte possui relação com resposta imune, sendo necessário estudos adicionais para verificar a relação dos genes editados com a sua expressão e a relação com a resposta desregulada de monócitos.

6. REFERÊNCIAS

Ahmad, S., Mu, X., & Hur, S. **The Role of RNA Editing in the Immune Response.** *Methods in molecular biology (Clifton, N.J.)*. 2021; 2181, 287–307. https://doi.org/10.1007/978-1-0716-0787-9_17

Akbari M, Akhavan-Bahabadi M, Shafigh N, Taheriazam A, Hussien BM, Sayad A, Fathi M, Taheri M, Ghafouri-Fard S, Fathi M. **Expression analysis of IFNAR1 and TYK2 transcripts in COVID-19 patients.** *Cytokine*. 2022 May;153:155849. doi: 10.1016/j.cyto.2022.155849. Epub 2022 Mar 4. PMID: 35339044; PMCID: PMC8894869.

Alanagreh L, Alzoughool F, Atoum M. **The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms.** *Pathogens*. 2020; 9(5):331. <https://doi.org/10.3390/pathogens9050331>

Alexander Lex, Nils Gehlenborg, Hendrik Strobel, Romain Vuillemot, Hanspeter Pfister. **UpSet: Visualization of Intersecting Sets** *IEEE Transactions on Visualization and Computer Graphics*. InfoVis. 2014. 20(12): 1983--1992, doi:10.1109/TVCG.2014.2346248.

Alsafi R. T. **Lessons from SARS-CoV, MERS-CoV, and SARS-CoV-2 Infections: What We Know So Far.** *The Canadian journal of infectious diseases & medical microbiology. Journal canadien des maladies infectieuses et de la microbiologie medicale*. 2022; 1156273. <https://doi.org/10.1155/2022/1156273>)

Anders, S., Pyl, P. T., & Huber, W. **HTSeq--a Python framework to work with high-throughput sequencing data.** *Bioinformatics (Oxford, England)*. 2015; 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>

Andrews, S. **FastQC: a quality control tool for high throughput sequence data.** 2017.

Aune, T. M., Tossberg, J. T., Heinrich, R. M., Porter, K. P., & Crooke, P. S., 3rd. **Alu RNA Structural Features Modulate Immune Cell Activation and A-to-I Editing of Alu RNAs Is Diminished in Human Inflammatory Bowel Disease.** *Frontiers in immunology*. 2022; 13, 818023. <https://doi.org/10.3389/fimmu.2022.818023>

Banerjee, A. K., Blanco, M. R., Bruce, E. A., Honson, D. D., Chen, L. M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S. A., Loney, C., Thai, J., Miller, Z. D., Lin, A. E., Schmidt, M. M., Stewart, D. G., Goldfarb, D., De Lorenzo, G., Rihn, S. J., Voorhees, R. M., Botten, J. W., ... Guttman, M. **SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses.** *Cell*. 2020; 183(5), 1325–1339.e21. <https://doi.org/10.1016/j.cell.2020.10.004>

Bessa, C., Matos, P., Jordan, P., & Gonçalves, V. **Alternative Splicing: Expanding the Landscape of Cancer Biomarkers and Therapeutics.** *International journal of molecular sciences*. 2020; 21(23), 9032. <https://doi.org/10.3390/ijms21239032>

Borao, S., Ayté, J., & Hümmer, S. **Evolution of the Early Spliceosomal Complex-From Constitutive to Regulated Splicing.** *International journal of molecular sciences*. 2021; 22(22), 12444. <https://doi.org/10.3390/ijms222212444>

Bruchez, A., Sha, K., Johnson, J., Chen, L., Stefani, C., McConnell, H., Gaucherand, L., Prins, R., Matreyek, K. A., Hume, A. J., Mühlberger, E., Schmidt, E. V., Olinger, G. G., Stuart, L. M., & Lacy-Hulbert, A. **MHC class II transactivator CIITA induces cell resistance to Ebola virus and SARS-like coronaviruses.** *Science (New York, N.Y.)*. 2020; 370(6513), 241–247. <https://doi.org/10.1126/science.abb3753>

Carlson M. *org.Hs.eg.db: Genome wide annotation for Human*. 2019; R package version 3.8.2.

Carvalho, T., Krammer, F. & Iwasaki, A. **The first 12 months of COVID-19: a timeline of immunological insights.** *Nat Rev Immunol.* 2021; **21**, 245–256. <https://doi.org/10.1038/s41577-021-00522-1>

Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R. **Features, Evaluation, and Treatment of Coronavirus (COVID-19).** In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 32150360.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. **Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.** *BMC Bioinformatics.* 2013; *128(14)*.

Christofi, T., Zaravinos, A. **RNA editing in the forefront of epitranscriptomics and human health.** *J Transl Med.* 2019; 17, 319. <https://doi.org/10.1186/s12967-019-2071-4>

Codo, A. C., Davanzo, G. G., Monteiro, L. B., de Souza, G. F., Muraro, S. P., Virgilio-da-Silva, J. V., Prodonoff, J. S., Carregari, V. C., de Biagi Junior, C. A. O., Crunfli, F., Jimenez Restrepo, J. L., Vendramini, P. H., Reis-de-Oliveira, G., Bispo Dos Santos, K., Toledo-Teixeira, D. A., Parise, P. L., Martini, M. C., Marques, R. E., Carmo, H. R., Borin, A., ... Moraes-Vieira, P. M. **Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1 α /Glycolysis-Dependent Axis.** *Cell metabolism.* 2020; *32(3)*, 437–446.e5. <https://doi.org/10.1016/j.cmet.2020.07.007>

Crooke, P. S., 3rd, Tossberg, J. T., Porter, K. P., & Aune, T. M. **Cutting Edge: Reduced Adenosine-to-Inosine Editing of Endogenous Alu RNAs in Severe COVID-19 Disease.** *Journal of immunology (Baltimore, Md. : 1950).* 2021; *206(8)*, 1691–1696. <https://doi.org/10.4049/jimmunol.2001428>

Cuddleston, W. H., Li, J., Fan, X., Kozenkov, A., Lalli, M., Khalique, S., Dracheva, S., Mukamel, E. A., & Breen, M. S. **Cellular and genetic drivers of RNA editing variation in**

the human brain. *Nature communications.* 2022; 13(1), 2997.
<https://doi.org/10.1038/s41467-022-30531-0>

Deininger P. **Alu elements: know the SINEs.** *Genome biology.* 2011; 12(12), 236.
<https://doi.org/10.1186/gb-2011-12-12-236>

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics (Oxford, England).* 2013; 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics (Oxford, England).* 2016; 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>

Fabregat A, Korninger F, Viteri G, Sidiropoulos K, Marin-Garcia P, Ping P, Wu G, Stein L, D'Eustachio P, Hermjakob H. **Reactome graph database: Efficient accessto complex pathway data.** *PLoS Comput Biol.* 2018; Jan 29;14(1):e1005968. doi: 10.1371/journal.pcbi.1005968. eCollection 2018 Jan. [PubMed PMID: 29377902](https://pubmed.ncbi.nlm.nih.gov/29377902/)

Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, D'Eustachio P, Stein L, Hermjakob H. **Reactome pathway analysis: a high-performance in-memory approach.** *BMC Bioinformatics.* 2017 Mar 2; 18(1):142. doi: 10.1186/s12859-017-1559-2. [PubMed PMID: 28249561](https://pubmed.ncbi.nlm.nih.gov/28249561/)

Fabregat A, Sidiropoulos K, Viteri G, Marin-Garcia P, Ping P, Stein L, D'Eustachio P, Hermjakob H. **Reactome diagram viewer: data structures and strategies to boost performance.** *Bioinformatics.* 2018 Apr 1; 34(7):1208-1214. doi: 10.1093/bioinformatics/btx752. [PubMed PMID: 29186351](https://pubmed.ncbi.nlm.nih.gov/29186351/)

Feng Zhang, Yulan Lu, Sijia Yan, Qinghe Xing, Weidong Tian; **SPRINT: an SNP-free toolkit for identifying RNA editing sites**. *Bioinformatics*. 2017; Volume 33, Issue 22, 15. Pages 3538–3548, <https://doi.org/10.1093/bioinformatics/btx473>

Ferré, Q., Charbonnier, G., Sadouni, N., Lopez, F., Kermezli, Y., Spicuglia, S., Capponi, C., Ghattas, B., & Puthier, D. **OLOGRAM: Determining significance of total overlap length between genomic regions sets**. *Bioinformatics (Oxford, England)*, btz810. 2019. <https://doi.org/10.1093/bioinformatics/btz810>

Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., ... Flicek, P. **GENCODE reference annotation for the human and mouse genomes**. *Nucleic acids research*. 2019; 47(D1), D766–D773. <https://doi.org/10.1093/nar/gky955>

Freitas, R. S., Crum, T. F., & Parvatiyar, K. **SARS-CoV-2 Spike Antagonizes Innate Antiviral Immunity by Targeting Interferon Regulatory Factor 3**. *Frontiers in cellular and infection microbiology*. 2022; 11, 789462. <https://doi.org/10.3389/fcimb.2021.789462>

GENCODE. **Human: Statistics about the current GENCODE Release (version 43)**. Disponível em: <https://www.ensembl.org/Homo_sapiens/Info/Annotation>. Acesso em: 4 fev. 2023.

GISAID. **hCoV-19 data sharing via GISAID**. Disponível em: <<https://gisaid.org/>>. Acesso em: 9 fev. 2023.

Gong X, Khan A, Wani MY, Ahmad A, Duse A. **COVID-19: A state of art on immunological responses, mutations, and treatment modalities in riposte.** *J Infect Public Health.* 2023 Feb; 16(2):233-249. doi: 10.1016/j.jiph.2022.12.019. Epub 2022 Dec 29. PMID: 36603376; PMCID: PMC9798670.

Gong, X., Khan, A., Wani, M. Y., Ahmad, A., & Duse, A. **COVID-19: A state of art on immunological responses, mutations, and treatment modalities in riposte.** *Journal of infection and public health.* 2023; 16(2), 233–249. <https://doi.org/10.1016/j.jiph.2022.12.019>

Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. **ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis.** *Mol Cell Proteomics.* 2020 Sep 9; doi: 10.1074/mcp. [PubMed PMID: 32907876](https://pubmed.ncbi.nlm.nih.gov/32907876/)

Harb J, Debs N, Rima M, Wu Y, Cao Z, Kovacic H, Fajloun Z, Sabatier J-M. **SARS-CoV-2, COVID-19, and Reproduction: Effects on Fertility, Pregnancy, and Neonatal Life.** *Biomedicines.* 2022; 10(8):1775. <https://doi.org/10.3390/biomedicines10081775>

Hu, T., Chitnis, N., Monos, D., & Dinh, A. **Next-generation sequencing technologies: An overview.** *Human immunology.* 2021; 82(11), 801–811. <https://doi.org/10.1016/j.humimm.2021.02.012>

Hu, X., Zou, Q., Yao, L. *et al.* **Survey of the binding preferences of RNA-binding proteins to RNA editing events.** *Genome Biol.* 2022; 23, 169. <https://doi.org/10.1186/s13059-022-02741-8>

Jafarzadeh, A., Nemati, M., & Jafarzadeh, S. **Contribution of STAT3 to the pathogenesis of COVID-19.** *Microbial pathogenesis.* 2021; 154, 104836. <https://doi.org/10.1016/j.micpath.2021.104836>

Jain, M., Jantsch, M. F., & Licht, K. **The Editor's I on Disease Development.** *Trends in genetics.* 2019; *TIG*, 35(12), 903–913. <https://doi.org/10.1016/j.tig.2019.09.004>

Jake R. Conway, Alexander Lex, Nils Gehlenborg. **UpSetR: An R Package For The Visualization Of Intersecting Sets And Their Properties.** *Bioinformatics.* 2017; 33(18): 2938-2940, doi:10.1093/bioinformatics/btx364

Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P. **The reactome pathway knowledgebase.** *Nucleic Acids Res.* 2020 Jan 8; 48(D1):D498-D503. doi: 10.1093/nar/gkz1031. [PubMed PMID: 31691815](https://pubmed.ncbi.nlm.nih.gov/31691815/).

Katz, K., Shutov, O., Lapoint, R., Kimelman, M., Brister, J. R., & O'Sullivan, C. **The Sequence Read Archive: a decade more of explosive growth.** *Nucleic acids research.* 2022; 50(D1), D387–D390. <https://doi.org/10.1093/nar/gkab1053>

Knoll, R., Schultze, J. L., & Schulte-Schrepping, J. **Monocytes and Macrophages in COVID-19.** *Frontiers in immunology.* 2021; 12, 720109. <https://doi.org/10.3389/fimmu.2021.720109>

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.** *Nucleic Acids Research.* 2016; [gkw377](https://pubmed.ncbi.nlm.nih.gov/27187011/) .

Kurkowiak, M., Arcimowicz, Ł., Chruściel, E., Urban-Wójciuk, Z., Papak, I., Keegan, L., O'Connell, M., Kowalski, J., Hupp, T., & Marek-Trzonkowska, N. **The effects of RNA editing in cancer tissue at different stages in carcinogenesis.** *RNA biology.* 2021; 18(11), 1524–1539. <https://doi.org/10.1080/15476286.2021.1877024>

Lauxmann, M. A., Santucci, N. E., & Autrán-Gómez, A. M. **The SARS-CoV-2 Coronavirus and the COVID-19 Outbreak.** *International braz j urol.* 2020; : official journal of the

Brazilian Society of Urology, 46(suppl.1), 6–18.
<https://doi.org/10.1590/S1677-5538.IBJU.2020.S101>

Li, L., G. Qian, Y. Zuo, Y. Yuan, Q. Cheng, T. Guo, J. Liu, C. Liu, L. Zhang, and H. Zheng. **Ubiquitin-dependent turnover of adenosine deaminase acting on RNA 1 (ADAR1) is required for efficient antiviral activity of type I interferon.** *J. Biol. Chem.* 2016; 291: 24974–24985.

Li, Q., Gloudemans, M. J., Geisinger, J. M., Fan, B., Aguet, F., Sun, T., Ramaswami, G., Li, Y. I., Ma, J. B., Pritchard, J. K., Montgomery, S. B., & Li, J. B. **RNA editing underlies genetic risk of common inflammatory diseases.** *Nature.* 2022; 608(7923), 569–577.
<https://doi.org/10.1038/s41586-022-05052-x>

Li, T., Yang, X., Li, W., Song, J., Li, Z., Zhu, X., Wu, X., & Liu, Y. **ADAR1 Stimulation by IFN- α Downregulates the Expression of MAVS via RNA Editing to Regulate the Anti-HBV Response.** *Molecular therapy. the journal of the American Society of Gene Therapy.* 2021; 29(3), 1335–1348. <https://doi.org/10.1016/j.ymthe.2020.11.031>

Li, Y., Duche, A., Sayer, M. R., Roosan, D., Khalafalla, F. G., Ostrom, R. S., Totonchy, J., & Roosan, M. R. **SARS-CoV-2 early infection signature identified potential key infection mechanisms and drug targets.** *BMC genomics.* 2021; 22(1), 125.
<https://doi.org/10.1186/s12864-021-07433-4>

Liu, H., Golji, J., Brodeur, L.K. *et al.* **Tumor-derived IFN triggers chronic pathway agonism and sensitivity to ADAR loss.** *Nat Med.* 2019; 25, 95–102.
<https://doi.org/10.1038/s41591-018-0302-5>

Liu, Y., & Rocklöv, J. **The effective reproductive number of the Omicron variant of SARS-CoV-2 is several times relative to Delta.** *Journal of travel medicine.* 2022; 29(3), taac037. <https://doi.org/10.1093/jtm/taac037>

Lo Giudice, C., Tangaro, M. A., Pesole, G., & Picardi, E. **Investigating RNA editing in deep transcriptome datasets with REDIttools and REDIportal.** *Nature protocols.* 2020. 15(3), 1098–1131. <https://doi.org/10.1038/s41596-019-0279-7>

Lo Giudice, C., Silvestris, D. A., Roth, S. H., Eisenberg, E., Pesole, G., Gallo, A., & Picardi, E. **Quantifying RNA Editing in Deep Transcriptome Datasets.** *Frontiers in genetics.* 2020; 11, 194. <https://doi.org/10.3389/fgene.2020.00194>

Love, M.I., Huber, W. & Anders, S. **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol.* 2014; 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>

Lu, Y., Shen, Y., Warren, W., & Walter, R., **Next Generation Sequencing in Aquatic Models. Next Generation Sequencing - Advances, Applications and Challenges.** InTech. Jan. 2016. Crossref, doi:10.5772/61657.2.

Ma, C. P., Liu, H., Yi-Feng Chang, I., Wang, W. C., Chen, Y. T., Wu, S. M., Chen, H. W., Kuo, Y. P., Shih, C. T., Li, C. Y., & Tan, B. C. **ADAR1 promotes robust hypoxia signaling via distinct regulation of multiple HIF-1 α -inhibiting factors.** *EMBO reports.* 2019; 20(5), e47107. <https://doi.org/10.15252/embr.201847107>

Malavige, G. N., Jeewandara, C., & Ogg, G. S. **Dengue and COVID-19: two sides of the same coin.** *Journal of biomedical science.* 2022; 29(1), 48. <https://doi.org/10.1186/s12929-022-00833-y>

Malavige, G.N., Jeewandara, C. & Ogg, G.S. **Dengue and COVID-19: two sides of the same coin.** *J Biomed Sci.* 2022; 29, 48. <https://doi.org/10.1186/s12929-022-00833-y>

Marc Gillespie, Bijay Jassal, Ralf Stephan, Marija Milacic, Karen Rothfels, Andrea Senff-Ribeiro, Johannes Griss, Cristoffer Sevilla, Lisa Matthews, Chuqiao Gong, Chuan

Deng, Thawfeek Varusai, Eliot Ragueneau, Yusra Haider, Bruce May, Veronica Shamovsky, Joel Weiser, Timothy Brunson, Nasim Sanati, Liam Beckman, Xiang Shao, Antonio Fabregat, Konstantinos Sidiropoulos, Julieth Murillo, Guilherme Viteri, Justin Cook, Solomon Shorser, Gary Bader, Emek Demir, Chris Sander, Robin Haw, Guanming Wu, Lincoln Stein, Henning Hermjakob, Peter D'Eustachio. **The reactome pathway knowledgebase 2022.** *Nucleic Acids Research*. 2021; gkab1028. <https://doi.org/10.1093/nar/gkab1028>

Martina, Z., Christina, H., Le, L., Jesse, D. C., Liang, Y. C., Christian, S. M., Monaghan, A. G., Kennedy, A. A., Yesselman, J. D., Gifford, R. R., Tai, A. W., Kutluay, S. B., Li, M. L., Brewer, G., Tolbert, B. S., & Hargrove, A. E. **Amilorides inhibit SARS-CoV-2 replication in vitro by targeting RNA structures.** *bioRxiv. the preprint server for biology*. 2020; 2020.12.05.409821. <https://doi.org/10.1101/2020.12.05.409821>

Mishra, G. R., Suresh, M., Kumaran, K., Kannabiran, N., Suresh, S., Bala, P., Shivakumar, K., Anuradha, N., Reddy, R., Raghavan, T. M., Menon, S., Hanumanthu, G., Gupta, M., Upendran, S., Gupta, S., Mahesh, M., Jacob, B., Mathew, P., Chatterjee, P., Arun, K. S., ... Pandey, A. **Human protein reference database--2006 update.** *Nucleic acids research*. 2006; 34(Database issue), D411–D414. <https://doi.org/10.1093/nar/gkj141>

Montazersaheb, S., Hosseiniyan Khatibi, S. M., Hejazi, M. S., Tarhriz, V., Farjami, A., Ghasemian Sorbeni, F., Farahzadi, R., & Ghasemnejad, T. **COVID-19 infection: an overview on cytokine storm and related interventions.** *Virology journal*. 2022; 19(1), 92. <https://doi.org/10.1186/s12985-022-01814-1>

Moraes, F., & Góes, A. **A decade of human genome project conclusion: Scientific diffusion about our genome knowledge.** *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology*. 2016; 44(3), 215–223. <https://doi.org/10.1002/bmb.20952>

Muller, I.B., Meijers, S., Kampstra, P. *et al.* **Computational comparison of common event-based differential splicing tools: practical considerations for laboratory**

researchers. *BMC Bioinformatics*. 2021; 22, 347.
<https://doi.org/10.1186/s12859-021-04263-9>

Nakahama, T., & Kawahara, Y. **Deciphering the Biological Significance of ADAR1-Z-RNA Interactions.** *International journal of molecular sciences*. 2021; 22(21), 11435.
<https://doi.org/10.3390/ijms222111435>

NCBI. Tool: fastq-dump. Disponível em: <<https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?>

Park JW., Tokheim C., Shen S., Xing Y. **Identifying differential alternative splicing events from RNA sequencing data using RNASeq-MATS.** *Methods in Molecular Biology: Deep Sequencing Data Analysis*. 2013; 1038:171-179 [doi: 10.1007/978-1-62703-514-9_10](https://doi.org/10.1007/978-1-62703-514-9_10)

Park, A., & Iwasaki, A. **Type I and Type III Interferons - Induction, Signaling, Evasion, and Application to Combat COVID-19.** *Cell host & microbe*. 2020; 27(6), 870–878.
<https://doi.org/10.1016/j.chom.2020.05.008>

Park, E., Pan, Z., Zhang, Z., Lin, L., & Xing, Y. **The Expanding Landscape of Alternative Splicing Variation in Human Populations.** *American journal of human genetics*. 2018; 102(1), 11–26. <https://doi.org/10.1016/j.ajhg.2017.11.002>

Peacock, T.P., Goldhill, D.H., Zhou, J. *et al.* **The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets.** *Nat Microbiol*. 2021; 6, 899–909.
<https://doi.org/10.1038/s41564-021-00908-w>

Pecori, R., Di Giorgio, S., Paulo Lorenzo, J., & Nina Papavasiliou, F. **Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination.** *Nature reviews*. 2022; *Genetics*, 23(8), 505–518. <https://doi.org/10.1038/s41576-022-00459-8>

Pisapia, P., Pepe, F., Sgariglia, R., Nacchio, M., Russo, G., Conticelli, F., Girolami, I., Eccher, A., Bellevicine, C., Vigliar, E., Malapelle, U., & Troncone, G. **Next generation sequencing**

in cytology. *Cytopathology : official journal of the British Society for Clinical Cytology.* 2021; 32(5), 588–595. <https://doi.org/10.1111/cyt.12974>

Raghava Kurup, R., Oakes, E. K., Manning, A. C., Mukherjee, P., Vadlamani, P., & Hundley, H. A. **RNA binding by ADAR3 inhibits adenosine-to-inosine editing and promotes expression of immune response protein MAVS.** *The Journal of biological chemistry.* 2022; 298(9), 102267. <https://doi.org/10.1016/j.jbc.2022.102267>

Redondo N, Zaldívar-López S, Garrido JJ and Montoya M. **SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns.** *Front. Immunol.* 2021; 12:708264. doi: 10.3389/fimmu.2021.708264

Regino-Zamarripa, N. E., Ramírez-Martínez, G., Jiménez-Álvarez, L. A., Cruz-Lagunas, A., Gómez-García, I. A., Ignacio-Cortés, S., Márquez-García, J. E., Pacheco-Hernández, L. M., Ramírez-Noyola, J. A., Barquera, R., Mendoza-Milla, C., Luna-Rivero, C., Domínguez-Cherit, J. G., Ramírez-Rangel, R., Rodríguez-Reyna, T. S., Hernández-Cárdenas, C. M., Choreño-Parra, J. A., León-Ávila, G., & Zúñiga, J. **Differential Leukocyte Expression of *IFITM1* and *IFITM3* in Patients with Severe Pandemic Influenza A(H1N1) and COVID-19.** *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research.* 2022; 42(8), 430–443. <https://doi.org/10.1089/jir.2022.0036>

Ren, Z., Ding, T., Zuo, Z., Xu, Z., Deng, J., & Wei, Z. **Regulation of MAVS Expression and Signaling Function in the Antiviral Innate Immune Response.** *Frontiers in immunology.* 2020; 11, 1030. <https://doi.org/10.3389/fimmu.2020.01030>

Riva, G., Albano, C., Gugliesi, F., Pasquero, S., Pacheco, S. F. C., Pecorari, G., Landolfo, S., Biolatti, M., & Dell'Oste, V. **HPV Meets APOBEC: New Players in Head and Neck Cancer.** *International journal of molecular sciences.* 2021; 22(3), 1402. <https://doi.org/10.3390/ijms22031402>

Roth, S.H., Levanon, E.Y. & Eisenberg, E. **Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity.** *Nat Methods.* 2019; 16, 1131–1138. <https://doi.org/10.1038/s41592-019-0610-9>

Saberian, M., Karimi, E., Khademi, Z. *et al.* **SARS-CoV-2: phenotype, genotype, and characterization of different variants.** *Cell Mol Biol Lett.* 2022; 27, 50. <https://doi.org/10.1186/s11658-022-00352-6>

Saberian, M., Karimi, E., Khademi, Z., Movahhed, P., Safi, A., & Mehri-Ghahfarrokhi, A. **SARS-CoV-2: phenotype, genotype, and characterization of different variants.** *Cellular & molecular biology letters.* 2022; 27(1), 50. <https://doi.org/10.1186/s11658-022-00352-6>

Schaffer, A. A., & Levanon, E. Y. **ALU A-to-I RNA Editing: Millions of Sites and Many Open Questions.** *Methods in molecular biology (Clifton, N.J.).* 2021; 2181, 149–162. https://doi.org/10.1007/978-1-0716-0787-9_9

Shen S., Park JW., Huang J., Dittmar KA., Lu ZX., Zhou Q., Carstens RP., Xing Y. **MATS: A Bayesian Framework for Flexible Detection of Differential Alternative Splicing from RNA-Seq Data.** *Nucleic Acids Research.* 2012; 40(8):e61 [doi: 10.1093/nar/gkr1291](https://doi.org/10.1093/nar/gkr1291)

Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., & Xing, Y. **rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data.** *Proceedings of the National Academy of Sciences of the United States of America.* 2014; 111(51), E5593–E5601. <https://doi.org/10.1073/pnas.1419161111>

Shi, G., Kenney, A. D., Kudryashova, E., Zani, A., Zhang, L., Lai, K. K., Hall-Stoodley, L., Robinson, R. T., Kudryashov, D. S., Compton, A. A., & Yount, J. S. **Opposing activities of IFITM proteins in SARS-CoV-2 infection.** 2021; *The EMBO journal*, 40(3), e106501. <https://doi.org/10.15252/emboj.2020106501>

Sidiropoulos K, Viteri G, Sevilla C, Jupe S, Webber M, Orlic-Milacic M, Jassal B, May B, Shamovsky V, Duenas C, Rothfels K, Matthews L, Song H, Stein L, Haw R, D'Eustachio P,

Ping P, Hermjakob H, Fabregat A. **Reactome enhanced pathway visualization.** *Bioinformatics*. 2017 Nov 1;33(21):3461-3467. doi: 10.1093/bioinformatics/btx441. [PubMed PMID: 29077811](#).

Siriwardena, S. U., Chen, K., & Bhagwat, A. S. **Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases.** *Chemical reviews*. 2016; 116(20), 12688–12710. <https://doi.org/10.1021/acs.chemrev.6b00296>

Song, B., Shiromoto, Y., Minakuchi, M., & Nishikura, K. **The role of RNA editing enzyme ADAR1 in human disease.** *Wiley interdisciplinary reviews*. 2022; *RNA*, 13(1), e1665. <https://doi.org/10.1002/wrna.1665>

Tang, S. J., Shen, H., An, O., Hong, H., Li, J., Song, Y., Han, J., Tay, D. J. T., Ng, V. H. E., Bellido Molias, F., Leong, K. W., Pitcheshwar, P., Yang, H., & Chen, L. **Cis- and trans-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development.** *Nature communications*. 2020; 11(1), 799. <https://doi.org/10.1038/s41467-020-14621-5>

Tolomeo, M., Cavalli, A., & Cascio, A. **STAT1 and Its Crucial Role in the Control of Viral Infections.** *International journal of molecular sciences*. 2022; 23(8), 4095. <https://doi.org/10.3390/ijms23084095>

Tomaselli, S., Bonamassa, B., Alisi, A., Nobili, V., Locatelli, F., & Gallo, A. **ADAR enzyme and miRNA story: a nucleotide that can make the difference.** *International journal of molecular sciences*. 2013; 14(11), 22796–22816. <https://doi.org/10.3390/ijms141122796>

Troyano-Hernández P, Reinos R, Holguín Á. **Evolution of SARS-CoV-2 Envelope, Membrane, Nucleocapsid, and Spike Structural Proteins from the Beginning of the Pandemic to September 2020: A Global and Regional Approach by Epidemiological Week.** *Viruses*. 2021 Feb 4; 13(2):243. doi: 10.3390/v13020243. PMID: 33557213; PMCID: PMC7913946.

Unterman, A., Sumida, T. S., Nouri, N., Yan, X., Zhao, A. Y., Gasque, V., Schupp, J. C., Asashima, H., Liu, Y., Cosme, C., Jr, Deng, W., Chen, M., Raredon, M. S. B., Hoehn, K. B., Wang, G., Wang, Z., Deluliis, G., Ravindra, N. G., Li, N., Castaldi, C., ... Dela Cruz, C. S. **Single-cell multi-omics reveals dyssynchrony of the innate and adaptive immune system in progressive COVID-19.** *Nature communications.* 2022; 13(1), 440. <https://doi.org/10.1038/s41467-021-27716-4>

V'kovski, P., Kratzel, A., Steiner, S. *et al.* **Coronavirus biology and replication: implications for SARS-CoV-2.** *Nat Rev Microbiol.* 2021; 19, 155–170. <https://doi.org/10.1038/s41579-020-00468-6>

Venkatesan, S., Rosenthal, R., Kanu, N., McGranahan, N., Bartek, J., Quezada, S. A., Hare, J., Harris, R. S., & Swanton, C. **Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution.** *Annals of oncology : official journal of the European Society for Medical Oncology.* 2018; 29(3), 563–572. <https://doi.org/10.1093/annonc/mdy003>

Vitali, P., Scadden, A. **Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis.** *Nat Struct Mol Biol.* 2010; 17, 1043–1050. <https://doi.org/10.1038/nsmb.1864>

Wang C, Chen L, Chen Y, Jia W, Cai X, Liu Y, Ji F, Xiong P, Liang A, Liu R, Guan Y, Cheng Z, Weng Y, Wang W, Duan Y, Kuang D, Xu S, Cai H, Xia Q, Yang D, Wang MW, Yang X, Zhang J, Cheng C, Liu L, Liu Z, Liang R, Wang G, Li Z, Xia H, Xia T. **Abnormal global alternative RNA splicing in COVID-19 patients.** *PLoS Genet.* 2022 Apr 14; 18(4):e1010137. doi: 10.1371/journal.pgen.1010137. PMID: 35421082; PMCID: PMC9089920.

Wang M-Y, Zhao R, Gao L-J, Gao X-F, Wang D-P and Cao J-M. **SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development.** *Front. Cell. Infect. Microbiol.* 2020; 10:587269. doi: 10.3389/fcimb.2020.587269

Wang, H., Chen, S., Wei, J., Song, G., & Zhao, Y. **A-to-I RNA Editing in Cancer: From Evaluating the Editing Level to Exploring the Editing Effects.** *Frontiers in oncology.* 2021; 10, 632187. <https://doi.org/10.3389/fonc.2020.632187>

Wang, W., & Krug, R. M. **U6atac snRNA, the highly divergent counterpart of U6 snRNA, is the specific target that mediates inhibition of AT-AC splicing by the influenza virus NS1 protein.** *RNA (New York, N.Y.).* 1998; 4(1), 55–64

Wang, Y., Liu, J., Huang, B. O., Xu, Y. M., Li, J., Huang, L. F., Lin, J., Zhang, J., Min, Q. H., Yang, W. M., & Wang, X. Z. **Mechanism of alternative splicing and its regulation.** *Biomedical reports.* 2015; 3(2), 152–158. <https://doi.org/10.3892/br.2014.407>

Wu G, Haw R. **Functional Interaction Network Construction and Analysis for Disease Discovery.** *Methods Mol Biol.* 2017; 1558:235-253. doi: 10.1007/978-1-4939-6783-4_11. [PubMed PMID: 28150241.](https://pubmed.ncbi.nlm.nih.gov/28150241/)

Xie Z, Bailey A, Kuleshov MV, Clarke DJB., Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, Jeon M, & Ma'ayan A. **Gene set knowledge discovery with Enrichr.** *Current Protocols.* 2021; 1, e90. doi: 10.1002/cpz1.90

Xu, L. D., & Öhman, M. **ADAR1 Editing and its Role in Cancer.** *Genes*. 2018; 10(1), 12. <https://doi.org/10.3390/genes10010012>

Yang, C. C., Chen, Y. T., Chang, Y. F., Liu, H., Kuo, Y. P., Shih, C. T., Liao, W. C., Chen, H. W., Tsai, W. S., & Tan, B. C. **ADAR1-mediated 3' UTR editing and expression control of antiapoptosis genes fine-tunes cellular apoptosis response.** *Cell death & disease*. 2017; 8(5), e2833. <https://doi.org/10.1038/cddis.2017.12>

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. **clusterProfiler: an R package for comparing biological themes among gene clusters.** *Omics : a journal of integrative biology*. 2012; 16(5), 284–287. <https://doi.org/10.1089/omi.2011.0118>

Zauli, G., AlHilali, S., Al-Swailem, S., Secchiero, P., & Voltan, R. **Therapeutic potential of the MDM2 inhibitor Nutlin-3 in counteracting SARS-CoV-2 infection of the eye through p53 activation.** *Frontiers in medicine*. 2022; 9, 902713. <https://doi.org/10.3389/fmed.2022.902713>

Zhang, S., Amahong, K., Zhang, C., Li, F., Gao, J., Qiu, Y., & Zhu, F. **RNA-RNA interactions between SARS-CoV-2 and host benefit viral development and evolution during COVID-19 infection.** *Briefings in bioinformatics*. 2022; 23(1), bbab397. <https://doi.org/10.1093/bib/bbab397>

Zhang, Y., Yang, X., Cui, Y., & Zhang, X. **Suppression of RNA editing by miR-17 inhibits the stemness of melanoma stem cells.** *Molecular therapy. Nucleic acids*. 2021; 27, 439–455. <https://doi.org/10.1016/j.omtn.2021.12.021>

Zhao X, Li J, Winkler CA, An P and Guo J-T. **IFITM Genes, Variants, and Their Roles in the Control and Pathogenesis of Viral Infections.** *Front. Microbiol.* 2019; 9:3228. doi:10.3389/fmicb.2018.03228

Zhao, Y., Kuang, M., Li, J. *et al.* **SARS-CoV-2 spike protein interacts with and activates TLR41.** *Cell Res.* 2021; 31, 818–820. <https://doi.org/10.1038/s41422-021-00495-9>

Zhong, Y., Xu, F., Wu, J., Schubert, J., & Li, M. M. **Application of Next Generation Sequencing in Laboratory Medicine.** *Annals of laboratory medicine.* 2021; 41(1), 25–43. <https://doi.org/10.3343/alm.2021.41.1.25>

Zhou, Y., Liu, Y., Gupta, S. *et al.* **A comprehensive SARS-CoV-2–human protein–protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets.** *Nat Biotechnol.* 2023; 41, 128–139. <https://doi.org/10.1038/s41587-022-01474-0>

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. **Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.** *Nature communications.* 2019; 10(1), 1523. <https://doi.org/10.1038/s41467-019-09234-6>

Zmasek, C. M., Lefkowitz, E. J., Niewiadomska, A., & Scheuermann, R. H. **Genomic evolution of the Coronaviridae family.** *Virology.* 2022; 570, 123–133. <https://doi.org/10.1016/j.virol.2022.03.005>