

Universidade de São Paulo  
Instituto de Medicina Tropical  
Faculdade de Medicina

Bárbara Brito de Souza Pereira

**Análise sistêmica de tecido cerebral em pacientes com morte associada ao  
vírus febre amarela**

São Paulo  
2023

Bárbara Brito de Souza Pereira

**Análise sistêmica de tecido cerebral em pacientes com morte associada ao  
vírus febre amarela**

**Versão original**

Tese apresentada à Faculdade de Medicina da  
Universidade de São Paulo para obtenção do título de  
Doutor em Ciências

Programa de Medicina Tropical

Área de concentração: Doenças Tropicais e Saúde  
Internacional

Orientador: Prof. Dr. Paolo Marinho de Andrade  
Zanotto

São Paulo

2023

**Dados Internacionais de Catalogação na Publicação (CIP)**

Preparada pela Biblioteca da  
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Pereira, Bárbara Brito de Souza  
Análise sistêmica de tecido cerebral em  
pacientes com morte associada ao vírus febre amarela  
/ Bárbara Brito de Souza Pereira. -- São Paulo,  
2023.  
Tese(doutorado)--Faculdade de Medicina da  
Universidade de São Paulo.  
Programa de Medicina Tropical. Área de  
Concentração: Doenças Tropicais e Saúde Internacional.  
Orientador: Paulo Marinho de Andrade Zanotto.

Descritores: 1.Arbovírus 2.Vírus da febre amarela  
3.Febre amarela 4.Neutropismo 5.Neuroinvasividade  
6.Sistema nervoso central 7.RNA-seq

USP/FM/DBD-379/23

Responsável: Erinalva da Conceição Batista, CRB-8 6755

**Nome:** PEREIRA, Bárbara Brito de Souza

**Título:** Análise sistêmica de tecido cerebral em pacientes com morte associada ao vírus Febre Amarela

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para a obtenção do título de Doutor em Ciências.

Aprovado em: \_\_\_ / \_\_\_ / \_\_\_\_\_

**Banca Examinadora**

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_



## RESUMO

Pereira BBS. Análise sistêmica de tecido cerebral em pacientes com morte associada ao vírus febre amarela [tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2023.

A doença causada pelo YFV (do inglês, *Yellow Fever Virus*) é chamada de Febre Amarela (do inglês, *Yellow Fever-YF*), transmitida para hospedeiros vertebrados pela picada de mosquitos hematófagos portadores do vírus. A patogênese da doença pode dividir-se em dois padrões distintos de infecção: (i) viscerotrópico e (ii) neurotrópico. No Brasil, entre 2017 e 2018 foram confirmados no estado de São Paulo, 555 casos em humanos, e 203 casos de óbito. Destes, 85 foram acompanhados pelo Departamento de Patologia da Faculdade de Medicina, da Universidade de São Paulo, e também pelo Laboratório de Evolução Molecular e Bioinformática (ICB/USP), e realizada autópsia e coletas de tecidos para investigação molecular com o intuito de verificar a presença do vírus no cérebro, e também nos outros tecidos. Aqui, o presente investigou aspectos biológicos ligados à acessibilidade, neuroinvasividade e respostas diferenciais associadas aos diferentes tipos de vírus febre amarela (YFV-17DD e YFV-silvático) utilizando dados moleculares, patológicos e expressão gênica diferencial obtida pela técnica de RNAseq. Para alcançar os objetivos propostos, o RNAseq foi realizado a partir de 13 amostras de tecido cerebral, obtidos de 10 pacientes que foram a óbito com suspeita de febre amarela, e 3 pacientes que fazem parte do grupos considerados controles. Os grupos foram classificados como YFV-silvático com vírus detectado no cérebro, YFV-silvático sem detecção viral no cérebro, YFV-17DD com vírus detectado no cérebro, YFV-17DD sem detecção viral no cérebro. A partir disso, níveis de expressão gênica em cada grupo foram analisados de forma a elucidar a neuropatogenia do YFV, buscando compreender a neuroinvasividade e neuroinfecção das diferentes linhagens. Ainda não está claro na literatura se as síndromes neurológicas associadas ao YFV são devido à infecção direta por YFV das células cerebrais, ou são secundárias à fenômenos imunomoduladores e como conclusão de nossos achados, foi possível hipotetizar que: (i) as partículas virais de YFV ultrapassaram a BHE e atingiram o tecido cerebral levando ao dano neurológico; (ii) as partículas de YFV atingiram a BHE, infectaram as células endoteliais, contudo não a ultrapassam chegando ao cérebro; (iii) as partículas virais de YFV não atingem o SNC e os danos neurológicos são efeitos secundários da infecção sistêmica.

Palavras-chave: Arbovírus. Vírus da febre amarela. Febre amarela. Neurotropismo. Neuroinvasividade. Sistema nervoso central. RNA-seq.

## ABSTRACT

Pereira.BBS. Systemic analysis of brain tissue in patients with death associated with yellow fever virus [thesis]. São Paulo: “Faculdade de Medicina, Universidade de São Paulo”; 2023.

The YFV (Yellow Fever Virus) disease is called Yellow Fever (YF), and it is transmitted to vertebrate hosts by the bite of hematophagous mosquitoes carrying the virus. The pathogenesis of the disease can be divided into two distinct patterns of infection: (i) viscerotropic and (ii) neurotropic. In Brazil, according to the Ministry of Health, from July 2017 to June 2018, 555 human cases were confirmed, and 203 lethal cases, in the state of São Paulo. Of these, 85 cases were followed up by the Department of Pathology at the Faculty of Medicine, University of São Paulo, and by the Laboratory of Molecular Evolution and Bioinformatics (ICB/USP), with autopsies being carried out on all patients and tissue collections for molecular investigation, to verify the presence of the virus in the brain, and also in other tissues. Here, we present new scientific research on the study of biological aspects related to accessibility, neuroinvasiveness and differential responses associated with different types of yellow fever virus (YFV-17DD and YFV-sylvatic), using molecular, pathological and differential gene expression obtained by RNAseq technique. To achieve the proposed objectives, RNAseq was performed on 13 samples of brain tissue, obtained from 10 patients who died with suspected yellow fever, and 3 patients who considered as control group. The groups were classified as sylvatic YFV with virus detected in the brain, sylvatic YFV without viral detection in the brain, vaccine YFV with virus detected in the brain and, vaccine YFV without viral detection in the brain. Gene expression levels in each group were analyzed to elucidate the neuropathogenesis of YFV, seeking to understand the neuroinvasiveness and neuroinfection of different strains. It is still unclear in the literature whether the neurological syndromes associated with YFV are due to direct YFV infection of brain cells, or are secondary to immunomodulatory phenomena. As a conclusion of our findings, it was possible to hypothesize that: (i) YFV viral particles crossed the BBB and reached the brain leading to neurological damage; (ii) the YFV particles reached the BBB, infected the endothelial cells, but did not go beyond it to reach the brain and; (iii) YFV viral particles do not reach the CNS and neurological damage is most-likely a secondary effect of the systemic infection.

Keywords: Arbovirus. Yellow fever virus. Yellow fever. Neurotropism. Neuroinvasiveness. Central nervous system. RNA-seq.

## AGRADECIMENTOS

Sou muito grata ao meu orientador Prof. Paolo M. de A. Zanotto, pela orientação, apoio e inspiração ao longo da jornada de pesquisa que culminou na conclusão da minha tese. Desde o início das discussões sobre o tema da pesquisa até as inúmeras conversas que tivemos para refinar minhas ideias, sua experiência e conhecimento enriqueceram cada aspecto do meu trabalho. Suas orientações críticas e insights foram fundamentais para moldar a direção da minha pesquisa e para aprofundar minha compreensão do assunto. Gostaria de destacar especialmente a sua paciência, que desempenhou um papel crucial na minha jornada durante estes 5 anos. No momento em que a maternidade se tornou novamente parte importante da minha vida, sua compreensão e apoio foram um verdadeiro presente. Sua gentileza e parceria durante esse período me permitiram abraçar tanto os desafios da maternidade quanto às demandas da pesquisa, tudo de uma maneira equilibrada. Suas palavras de incentivo e o espaço que ele me proporcionou para crescer como mãe e pesquisadora fizeram toda a diferença. Seu comprometimento e crença em mim como mãe e pesquisadora inspiraram-me a ver o melhor em ambos os papéis que desempenho. Sua atitude encorajadora incentivou-me a superar obstáculos e a acreditar que posso alcançar minhas metas, independentemente das circunstâncias. A conclusão desta tese não teria sido possível sem seu valioso apoio e orientação. Anseio em compartilhar os resultados desta pesquisa com a comunidade acadêmica, sabendo que sua influência está profundamente enraizada em cada página. OBRIGADA!

Gostaria de expressar minha sincera gratidão a todos os alunos e servidores que estiveram presentes ao longo da minha trajetória acadêmica no laboratório de Evolução Molecular e Bioinformática (LEMB), bem como aqueles que estenderam sua influência além desse ambiente. Suas contribuições foram inestimáveis para o meu crescimento e desenvolvimento como pesquisadora. Especialmente, quero agradecer à Carla Torres Braconi, Shahab Zaki Pour, Rafaella Sayuri Yoshino, Aureliano Guedes e Irina Yuri Kawashima. Sua colaboração, troca de ideias e apoio foram fundamentais para o progresso da minha pesquisa. Cada um de vocês deixou uma marca única e valiosa em minha jornada acadêmica.

Gostaria de expressar minha sincera gratidão à minha amiga e notável neurocientista, Ethiane Segabinazi. Sua dedicação incansável e contribuições inestimáveis para a minha tese foram fundamentais para o seu sucesso. Ethiane, seu trabalho árduo e dedicação à pesquisa no campo da neurociência são verdadeiramente admiráveis. Seus achados foram um alicerce sólido para esta tese, enriquecendo nosso entendimento sobre o funcionamento do

cérebro. É importante reconhecer e celebrar o trabalho árduo dos cientistas que ampliam nossos horizontes de conhecimento, e você certamente é um exemplo brilhante desse compromisso.

Também desejo estender meu agradecimento a todos os membros do corpo docente da Universidade de São Paulo que contribuíram para minha formação ao longo deste percurso acadêmico. Suas aulas, discussões e contribuições em sala de aula enriqueceram minha compreensão do mundo acadêmico e me motivaram a buscar sempre mais conhecimento e excelência.

Agradeço também o suporte estrutural proporcionado pelo Instituto de Ciências Biomédicas, Departamento de Microbiologia, do Centro de Facilidades de Apoio à Pesquisa (CEFAP-USP) e do Laboratório de Genética da Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo (ESALQ-USP) os quais foram essenciais para o desfecho deste trabalho.

Minha gratidão estende-se aos colaboradores envolvidos neste trabalho. Especialmente gostaria de mostrar meu apreço ao Prof. Dr. Amaro N. Duarte-Neto, que fez um trabalho espetacular no processamento das autópsias de todos os pacientes que morreram com suspeita clínica associada ao vírus da Febre Amarela entre 2017-2019. Sua dedicação e expertise foram fundamentais para o sucesso deste trabalho e para a obtenção de resultados significativos. A qualidade do seu trabalho impactou positivamente, não apenas o projeto em si, mas também a compreensão mais ampla da comunidade sobre a Febre Amarela e suas implicações em geral. Seu esforço e dedicação serão lembrados como uma parte integral do sucesso deste trabalho. Espero que saiba o quanto sua contribuição é valorizada e apreciada por mim e por todos os envolvidos neste projeto.

Agradeço aos meus pais Nilson e Eliana, e padrasto Marcelo, pelos seus encorajamentos, palavras de incentivo e compreensão que foram um pilar essencial em momentos desafiadores.

Às minhas filhas Júlia e Elisa, vocês são a minha fonte constante de alegria e amor. A maternidade pode ser uma jornada desafiadora, mas vocês tornam tudo valioso. Suas personalidades únicas, suas perguntas curiosas e suas descobertas diárias trazem uma nova perspectiva à minha vida. Cada dia ao lado de vocês, é uma oportunidade única para aprender com vocês a importância da perseverança, da paciência e do amor incondicional. Vocês me ensinaram a ver o mundo com olhos renovados e a encontrar beleza nas pequenas coisas. Vocês são minhas maiores conquistas e meu maior orgulho. Obrigada por serem a razão pela qual sorrio mesmo nos momentos mais desafiadores. Eu amo vocês mais do que as palavras podem expressar.

Por fim, a minha gratidão especial àquele com quem compartilhei meus genes, Marielton dos Passos Cunha, virologista, bioinformata, meu marido e pai das minhas filhas. Sua paixão pelo conhecimento e pela pesquisa é uma fonte constante de inspiração para mim. A conclusão deste projeto é uma realização compartilhada. Saiba que, sem você, nada disso seria possível. Sua influência está presente em cada página escrita, em cada desafio superado e em cada vitória alcançada. Você é parte integral desta conquista. Obrigada!

Ao longo desse período, importantes apoios financeiros aconteceram para possibilitar o desfecho deste trabalho. Registro aqui agradecimento ao fomento fornecido pela: (i) Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, (ii) Conselho Nacional de Desenvolvimento Científico e Tecnológico (Processo n° 441105/2016-5); Fundação de Amparo à Pesquisa do Estado de São Paulo (Processo n° 2017/23281-6); e Fiocruz/Pasteur/USP (Processo n° 314502).

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	17
<b>1.1. Contextualização</b> .....	17
<b>1.2. Justificativa</b> .....	19
<b>1.3. Estrutura</b> .....	21
<b>2. REVISÃO BIBLIOGRÁFICA</b> .....	23
<b>2.1. Arbovírus</b> .....	23
<b>2.1.1. <i>Flavivirus</i></b> .....	23
<b>2.3. Rotas de acessibilidade dos <i>Flavivirus</i> ao SNC</b> .....	34
<b>2.3. Análise de expressão gênica</b> .....	37
<b>2.4. Biologia de sistemas em Virologia</b> .....	38
<b>2.4.1. Análises de dados biológicos explorando análise de redes</b> .....	39
<b>3. OBJETIVOS</b> .....	42
<b>3.1. Objetivo geral</b> .....	42
<b>3.1. Objetivos específicos</b> .....	42
<b>4. MATERIAIS E MÉTODOS</b> .....	44
<b>4.1. Aspectos éticos</b> .....	44
<b>5.2. Material de estudo</b> .....	44
<b>4.2. Definição de caso</b> .....	45
<b>4.3. Caracterização molecular</b> .....	45
<b>4.4. Imunohistoquímica</b> .....	46
<b>4.5. RNA-seq</b> .....	46
<b>4.5.1. Desenho experimental e hipóteses consideradas no estudo</b> .....	47
<b>4.5.2. Preparação do RNA (Purificação e quantificação)</b> .....	49
<b>4.5.3. Bioanalyzer</b> .....	49
<b>4.5.3. Preparo das bibliotecas para RNAseq</b> .....	49
<b>4.5.4. Quantificação e normalização das bibliotecas montadas</b> .....	50

4.5.5. Sequenciamento.....	51
4.6. Análise de dados.....	51
4.6.1. Pré-processamento de dados, análise de qualidade das sequências e remoção de adaptadores.....	51
4.6.2. Mapeamento e quantificação da expressão gênica.....	52
4.6.3. Avaliação da variabilidade de dados.....	52
4.6.4. Análise de expressão diferencial via DESeq2.....	52
4.6.5. Análise de enriquecimento.....	53
5. RESULTADOS.....	57
5.1 Caracterização molecular.....	57
5.2. Análise de qualidade (pré-sequenciamento).....	59
5.2.1. Análise da integridade do RNA total.....	59
5.3. Análise de qualidade pós sequenciamento.....	60
5.3.1. Trimagem.....	60
5.4. Análise de dados RNAseq.....	61
5.4.1. Pré processamento.....	61
5.4.2. Análises de expressão gênica diferencial e ontológica.....	67
6. DISCUSSÃO.....	128
7. Limitações do trabalho.....	137
8. Perspectivas Futuras.....	138
Anexos.....	148

## Lista de figuras

<b>Figura 1</b> - Organização genômica do vírus Febre Amarela .....	24
<b>Figura 2</b> - Casos de Febre Amarela confirmados e letalidade no Brasil.....	26
<b>Figura 3</b> - Árvore filogenética representando os vírus do gênero Flavivirus .....	35
<b>Figura 4</b> - Representação esquemática dos mecanismos de entrada de Flavivirus no SNC via BHE.....	36
<b>Figura 5</b> - Etapas necessárias para um experimento padrão de RNAseq.....	47
<b>Figura 6</b> - Fluxograma adaptado para preparo de bibliotecas de RNAseq .....	50
<b>Figura 7</b> - Caracterização dos valores de Cts identificados considerando os cérebros positivos e negativos, para os diferentes tecidos.....	57
<b>Figura 8</b> - Caracterização molecular das amostras estudadas.....	58
<b>Figura 9</b> - Comparação entre os tempos post-mortem e valores de RIN.....	60
<b>Figura 10</b> - Caracterização da qualidade dos índices de qualidade para todas as bases das sequências .....	61
<b>Figura 11</b> - Gráfico de densidade com contagem bruta de expressão por amostra.....	62
<b>Figura 12</b> - Valores de contagem de genes expressos, antes e depois da normalização.....	63
<b>Figura 13</b> - Análise de componentes principais (PCA) do dados de expressão gênica das amostras .....	64
<b>Figura 14</b> - Gráfico de pareto.....	65
<b>Figura 15</b> - Análise exploratória dos dados de expressão gênica considerando diferentes combinações de PCs.....	66
<b>Figura 16</b> - Fluxograma para interpretação da análise de dados.....	68
<b>Figura 17</b> - Grupos de genes caracterizados no RNAseq (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral).....	70
<b>Figura 18</b> – Comparação dos genes diferencialmente expressos (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral).....	71
<b>Figura 19</b> - Distribuição de valores de p (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral).....	72
<b>Figura 20</b> - Volcano plot mostrando a expressão gênica diferencial (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral).....	73
<b>Figura 21</b> - Vias enriquecidas (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral).....	75



<b>Figura 22-</b> Vias enriquecidas pelo GSEA (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral) .....	76
<b>Figura 23</b> - Rede Gene-concept (pacientes negativos para a detecção molecular do vírus no tecido cerebral) .....	78
<b>Figura 24</b> - Rede Gene-concept (pacientes positivos para a detecção molecular do vírus no tecido cerebral) .....	79
<b>Figura 25</b> - Classificação funcional dos genes para o grupo de pacientes com negatividade para detecção do vírus no cérebro .....	81
<b>Figura 26</b> - Classificação funcional dos genes para o grupo de pacientes com positividade para detecção do vírus no cérebro .....	82
<b>Figura 27</b> - Rede de Similaridade ponderada (YFV não foi detectado no cérebro).....	84
<b>Figura 28</b> - Rede de Similaridade ponderada (YFV foi detectado no cérebro).....	85
<b>Figura 29</b> - Grupos de genes caracterizados no RNAseq (detecção molecular do YFV no tecido cerebral).....	87
<b>Figura 30</b> – Comparação dos genes diferencialmente expressos (detecção molecular do YFV no tecido cerebral) .....	88
<b>Figura 31</b> - Distribuição de valores de p .....	89
<b>Figura 32</b> - Volvano plot mostrando a expressão gênica diferencial para o YFV-silvático e YFV-17DD .....	90
<b>Figura 33</b> - Diferentes vias enriquecidas dos genes interclasses.....	92
<b>Figura 34</b> - Principais vias enriquecidas pelo GSEA dos genes interclasses .....	93
<b>Figura 35</b> - Rede Gene-concept da condição em que foram analisados os genes interclasses	94
<b>Figura 36</b> - Classificação funcional dos genes para o grupo de genes interclasses.....	95
<b>Figura 37</b> - Vias enriquecidas para os grupos de pacientes com YFV-silvático detectado por RT-qPCR do RNA viral no tecido cerebral.....	97
<b>Figura 38</b> - Vias enriquecidas para os grupos de pacientes com YFV-17DD detectado por RT-qPCR do RNA viral no tecido cerebral .....	98
<b>Figura 39</b> - Principais vias enriquecidas pelo GSEA para o YFV-silvático e YFV-17DD.....	99
<b>Figura 40</b> - Rede Gene-concept para o o YFV-silvático .....	101
<b>Figura 41</b> - Classificação funcional dos genes diferencialmente expressos na condição A, em que o YFV-silvático foi detectado no cérebro.....	102
<b>Figura 42</b> - Rede de Similaridade ponderada para os genes relacionados à condição A, em que o YFV-silvático foi detectado no cérebro.....	103
<b>Figura 43</b> - Rede <i>Gene-concept</i> (YFV-17DD foi detectado no cérebro) .....	104

<b>Figura 44</b> - Classificação funcional dos genes diferencialmente expressos na condição B, em que o YFV-17DD foi detectado no cérebro .....	105
<b>Figura 45</b> - Rede de Similaridade ponderada (condição B), em que o YFV-17DD foi detectado no cérebro.....	106
<b>Figura 46</b> - Grupos de genes caracterizados no RNAseq (pacientes negativos para a detecção molecular do YFV no tecido cerebral).....	108
<b>Figura 47</b> – Comparação dos genes diferencialmente expressos (YFV não detectado no tecido cerebral) .....	109
<b>Figura 48</b> - Distribuição de valores de p (YFV não detectado no tecido cerebral) .....	110
<b>Figura 49</b> - Volcano plot mostrando a expressão gênica diferencial (YFV não detectado no tecido cerebral).....	111
<b>Figura 50</b> - Diferentes vias enriquecidas dos genes localizados na intersecção os pacientes em que o YFV-silvático não foi detectado no cérebro e para os pacientes em que o YFV-17DD não foi detectado no cérebro .....	113
<b>Figura 51</b> - Principais vias significativas enriquecidas pelo GSEA (utilizando os genes presentes na intersecção).....	114
<b>Figura 52</b> - Rede Gene-concept da condição em que foram analisados os genes presentes na intersecção entre os pacientes em que o YFV-silvático não foi detectado no cérebro e para os pacientes em que o YFV-17DD não foi detectado no cérebro .....	115
<b>Figura 53</b> - Classificação funcional dos genes presentes na intersecção entre os pacientes em que o YFV-silvático não foi detectado no cérebro e para pacientes em que o YFV-17DD não foi detectado no cérebro .....	116
<b>Figura 54</b> - Diferentes vias enriquecidas em condições distintas (YFV-silvático e YFV-17DD negativos no tecido cerebral) .....	119
<b>Figura 55</b> - Vias enriquecidas pelo GSEA (YFV-silvático e YFV-17DD negativos no tecido cerebral) .....	120
<b>Figura 56</b> - Rede Gene-concept da condição em que foram analisados os genes diferencialmente expressos na condição A, em que o YFV-silvático não foi detectado no cérebro.....	121
<b>Figura 57</b> - Classificação funcional dos genes diferencialmente expressos na condição A, em que o YFV-silvático não foi detectado no cérebro .....	122
<b>Figura 58</b> - Rede de Similaridade ponderada atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição A, em que o YFV-silvático não foi detectado no cérebro.....	123

<b>Figura 59</b> - Rede Gene-concept da condição em que o YFV-17DD não foi detectado no cérebro .....	124
<b>Figura 60</b> - Classificação funcional dos genes diferencialmente expressos na condição B, em que o YFV-17DD não foi detectado no cérebro.....	125
<b>Figura 61</b> - Rede de Similaridade ponderada, com pesos atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição em que o YFV-17DD não foi detectado no cérebro.....	126

## Lista de tabelas

<b>Tabela 1</b> - Caracterização da amostragem utilizada no estudo. Os pacientes estão distribuídos em cinco grupos, que representam: 1 = YFV-Silvático (Cérebro positivo); 2 = YFV-Silvático (Cérebro negativo); 3 = YFV-17DD (Cérebro positivo); 4 = YFV-17DD (Cérebro negativo); 5 = controles. As informações clínicas e laboratoriais associadas aos grupos também estão contidas na tabela.....	48
<b>Tabela 2</b> - Dados de tamanho médio da biblioteca de cada amostra e quantidade dada por PCR, para cálculo de montagem de pools.....	51

# *Capítulo 1*

# 1. INTRODUÇÃO

## 1.1. Contextualização

Os arbovírus, vírus transmitidos por artrópodes (derivado do inglês, *arthropod borne virus*), formam um grupo não taxonômico compostos em sua grande maioria por vírus de RNA, que são mantidos de forma cíclica na natureza replicando em hospedeiros vertebrados susceptíveis (também definidos como hospedeiros amplificadores) e artrópodes hematófagos (vetores). Estes agentes virais, são responsáveis por causarem doenças infecciosas classificadas como emergentes e reemergentes em diferentes áreas do mundo, especialmente em regiões tropicais, subtropicais, estendendo abrangência para as altas latitudes em todo hemisfério Norte, da eurásia à América do Norte (como por exemplo, o *Flavivirus* da encefalite transmitida por ácaros *Tick Borne Encephalitis Virus* - TBEV) (1). Dentro do grupo dos arbovírus, está o gênero *Flavivirus* constitui um grupo amplo e diverso, sendo o vírus Febre Amarela (do inglês, *Yellow Fever virus* - YFV) o vírus protótipo para o gênero. A doença causada pelo YFV é chamada de Febre Amarela (do inglês, *Yellow Fever* - YF), transmitida para hospedeiros vertebrados pela picada de mosquitos hematófagos portadores do vírus. A sua incidência estimada é 200 mil casos por ano, com 30 mil mortes em áreas endêmicas, que incluem África e Américas. Apesar da disponibilidade de uma vacina eficaz contra o YFV, a circulação do vírus permanece endêmica nessas regiões. O principal ciclo de transmissão e que assegura a permanência da circulação viral nas Américas, é o ciclo silvático, onde os primatas não-humanos (PNH) atuam como reservatório primário do YFV, e quando picados por mosquitos silvestres competentes, transmitem o vírus para outros PNH. Os humanos quando susceptíveis também são infectados pelo vírus através de picadas vetores silvestres infectados com o vírus, principalmente quando não são vacinados e trabalham ou visitam áreas tropicais e endêmicas, e assim, podem desenvolver a doença.

A vacina de febre amarela (YF-17D) é considerada uma das mais eficazes e seguras, administrada nos últimos 65 anos em mais de 600 milhões de pessoas no mundo. No Brasil ela é produzida desde 1937 pela Bio-Manguinhos (Fundação Oswaldo Cruz - FIOCRUZ) a partir do isolado viral atenuado YFV-17D, o qual é cultivado em ovos embrionados de galinhas. Uma única dose desta vacina induz uma resposta imune, incluindo ativação de linfócitos T citotóxicos, linfócitos T *helper* do Tipo 1 e 2 e formação de anticorpos neutralizantes e protetores com longa duração frente à doença. Até 1941, nenhum caso de reação adversa à vacina havia sido relatado. Contudo, em julho deste mesmo ano descobriu-se que após

episódios de vacinação, uma doença invasiva e disseminada poderia ocorrer em alguns indivíduos saudáveis, com alto índice de letalidade e associada ao vírus vacinal. Foram notificados 192 casos de encefalite em uma população recém vacinada, mostrando que 0,5% do total de vacinados apresentaram complicações neurológicas após a vacinação.

O espectro clínico da Febre Amarela (do inglês, *Yellow Fever* - YF) pode variar desde uma apresentação subclínica à grave doença sistêmica apresentando-se, em sua forma clássica, como: febre, graves danos ao fígado e aos rins, complicações no miocárdio e por fim, febre hemorrágica de elevada letalidade em 20-50% dos casos sintomáticos. A patogênese da doença pode dividir-se em dois padrões distintos de infecção: (i) viscerotrópico (capacidade de infectar diversos órgãos, incluindo fígado, baço, coração, rim, pulmão, entre outros) e (ii) neurotrópico (capacidade do vírus infectar o parênquima cerebral causando um quadro de encefalite).

Os primeiros relatos de danos neurológicos foram descritos por *Lasnet* em 1927 com pacientes previamente infectados com YFV na África (2), contudo 90 anos após essa preocupação associada ao quadro neurotrópico associado ao vírus, nenhum outro estudo tratou desse tema com profundidade. Neste período, os trabalhos científicos explorando esse tema dedicam-se a poucos relatos na literatura mostrando a neuroinvasividade do YFV em casos que com presença de sintomas neurológicos em pacientes pós-vacinais. Outros arbovírus do gênero *Flavivirus* como o vírus Dengue, Zika vírus, vírus *West Nile*, entre outros, são conhecidos por causar doenças neurológicas graves em humanos. Uma grande questão científica acerca da acessibilidade desses vírus ao sistema nervoso central (SNC), mostrando como estes vírus atravessam a barreira hematoencefálica ainda é muito discutida. Alguns arbovírus infectam diretamente os neurônios após a entrada ao SNC, contudo esta informação ainda não foi relatada na literatura quanto ao YFV, podendo desta forma ser discutido um possível dano neural secundário devido a inflamação no tecido cerebral, por citocinas pró inflamatórias, assim como ocorre no vírus da imunodeficiência humana (do inglês, *Human Immunodeficiency virus* – HIV).

Segundo dados do *World Health Organization* (WHO), entre Julho de 2017 a Junho de 2018, 1376 casos de YF foram confirmados nos estados do Rio de Janeiro, São Paulo e Minas Gerais. Destes, 483 vieram a óbito (letalidade de 35.1%). Diante deste cenário, em parceria com pesquisadores do Departamento de Patologia do Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo (HCFMUSP) (Prof. Dr. Paulo Hilário Nascimento Saldiva & Prof. Dr. Amaro Nunes Duarte-Neto), o Laboratório de Evolução Molecular e Bioinformática do Instituto de Ciências Biomédicas da Universidade de São Paulo (LEMB-

ICB), participou de um esforço conjunto no enfrentamento do surto de YFV na cidade de São Paulo. Diversos tecidos foram coletados de 85 pacientes que foram a óbito com suspeita de causa associada ao YFV. Dos 77 pacientes que foram positivos para o YFV, oito tiveram o vírus vacinal 17DD (aqui, definiremos essa situação como YFV-17DD) detectado por RT-qPCR em pelo menos um dos órgãos coletados, enquanto o restante entre os positivos, foram classificados como positivos para o YFV-silvático (aqui, definiremos essa situação como YFV-silvático). Outros trabalhos científicos foram publicados a partir deste projeto, explorando aspectos ligados à epidemiologia e evolução molecular do vírus (3,4), o impacto de transplante de fígado na YF (5), e propondo novas estratégias de autópsia para doenças infecciosas (6). Aqui, o presente estudo trata-se de uma nova proposta científica de investigação para estudar aspectos biológicos ligados à acessibilidade, neuroinvasividade e respostas diferenciais associadas aos diferentes tipos de vírus (YFV-17DD e YFV-silvático) utilizando dados moleculares, patológicos e expressão gênica diferencial obtida pela técnica de RNAseq.

## 1.2. Justificativa

Entre os arbovírus, a família *Flaviviridae* apresenta pelo menos 70 vírus responsáveis por uma variedade de doenças em humanos e animais. Em geral, esses vírus são agentes causadores de doenças infecciosas emergentes e reemergentes que causam morbidade e mortalidade em regiões endêmicas, tornando-se um importante problema em Saúde Pública à nível global. Na América do Sul, já foi relatada a circulação de três *Flavivirus* importantes: o dengue vírus (DENV), o vírus Zika (ZIKV) e o YFV. Como trata-se de uma zoonose, erradicar a YF é muito difícil, e de maneira recorrente surtos e epidemias são reportados em PNH e humanos.

Dados da Organização Mundial da Saúde (OMS) apontam para a ocorrência de 200 mil casos de YF com 30 mil mortes associadas à doença anualmente, embora as notificações oficiais sejam menores ([www.who.int/csr](http://www.who.int/csr)). Durante as epidemias, em populações não vacinadas as taxas de letalidade podem exceder 50% e não existe nenhum tratamento específico para a doença, além de cuidados de suporte (7). O Brasil – principalmente a bacia amazônica – possui a maior área de atividade viral para o YFV do mundo (8–10), fatores que aumentam a preocupação dos serviços de saúde quando consideramos o elevado crescimento demográfico nestas regiões junto com a baixa adesão às campanhas de vacinação atual. Combinados, estes fatores tornam os moradores de centros urbanos também como vulneráveis à infecção associada ao vírus. Além disso, a resistência do *Aedes aegypti* aos inseticidas, a dificuldade operacional



para desenvolver ações de vigilância e combate ao vetor aumentam a chance da re-emergência da febre amarela urbana. Em contrapartida, a vacina anti-YF é produzida utilizando a linhagem atenuada 17D do YFV e representa a vacina de maior sucesso desenvolvida até os dias de hoje. É classificada como possuidora de características biológicas consideradas seguras, induz uma alta taxa de soroconversão, e possui uma eficiência entre 90% e 98%. Porém, mesmo com seu sucesso, são observados casos de mortalidade induzidos pela administração da vacina, variando de 1 em 250.000 até 1 em 500.000 episódios de vacinação. Nestes casos, após 2 a 5 dias a vacina pode induzir o desenvolvimento de febre alta, mal-estar, mialgia, seguindo por icterícia, instabilidade cardiovascular, hemorragia e insuficiência renal e respiratória. As causas e os mecanismos imunológicos que resultam a esta séria reação adversa ainda não são desconhecidos pela comunidade médica e científica.

No Brasil, segundo o Ministério da Saúde, de Julho 2017 a Junho 2018 foram confirmados a região sudeste foi a mais afetada com 99.9% dos casos de YF e o estado de São Paulo, com 555 casos confirmados em humanos, e 203 casos de óbito. Desses 203 casos, entre Dezembro de 2017 a Abril de 2018, 85 casos de pacientes que progrediram a óbito foram acompanhados pelo Departamento de Patologia da Faculdade de Medicina, Universidade de São Paulo, e também pelo Laboratório de Evolução Molecular e Bioinformática (ICB/USP), sendo realizado a autópsia de todos os pacientes e coletas de tecidos para investigação molecular com o intuito de verificar a presença do vírus no cérebro, e também nos outros tecidos. Sendo assim, acreditamos que as análises de dados provenientes da integração de dados moleculares, patológicos e de RNA-seq aqui propostas servirão de base para traçar os efeitos da acessibilidade, replicação viral em tecido cerebral e entender melhor o funcionamento do estado neuroinflamatório causado pela YF. Acreditamos que estudar a reação sistêmica do hospedeiro na presença da infecção causada pelos vírus vacinal e silvático durante uma epidemia ocorrida no Estado de São Paulo é de extrema importância para a elucidação dos mecanismos envolvidos na neuropatologia associada ao YFV-silvático e YFV-17DD, com esses resultados podendo ter implicações em pacientes com sequelas neurológicas e/ou em pacientes com altas taxas de sobrevida e recidiva após infecção aguda. Acreditamos que este estudo é inédito quanto a identificação e o enriquecimento de genes expressos em tecido cerebral *post mortem*, infectados com as linhagens do YFV-silvático e YFV-17DD. As implicações dos conhecimentos aqui gerados, poderão ser importantes e passos fundamentais para identificar drogas e alvos terapêuticos para serem estudados em condições como silenciamento gênico induzidas por RNA de interferência. Desta forma, este trabalho pretende,

via análise de transcriptoma comparativo, obter um perfil de interações entre genes que nos permita tirar conclusões sobre a neuropatogenia do YFV.

### **1.3. Estrutura**

Esta tese foi estruturada da seguinte maneira: No *Capítulo 1*, foi feita uma breve introdução com (i) a contextualização da doença febre amarela e sua neuropatogenia; (ii) justificativa baseada na necessidade de análise de expressão gênica integrada aos dados moleculares e patológicos para estudo da resposta sistêmica à interação vírus-hospedeiro entre diferentes linhagens do YFV (YFV-silvático e YFV-17DD), buscando marcadores que justifiquem a acessibilidade ou o bloqueio do YFV no SNC; (iii) estrutura mostrando como o trabalho foi estruturado, servindo de guia ao leitor e (iv) limitações. No *Capítulo 2*, foi realizada uma ampla revisão bibliográfica que serviu de embasamento para a realização do estudo. No *Capítulo 3*, foram inseridos os objetivos do estudo. No *Capítulo 4*, foi descrito os materiais e métodos utilizados para atingir os objetivos. No *Capítulo 5*, apresento os resultados obtidos a partir dos materiais e métodos utilizados. E, no *Capítulo 6*, discutimos os resultados encontrados e as considerações finais do trabalho.

## *Capítulo 2*

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1. Arbovírus

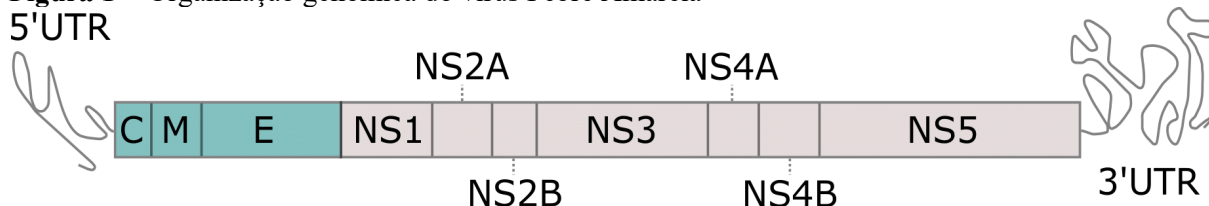
Os arbovírus (do inglês, *Arthropod-borne virus*) constituem um amplo grupo de vírus que apresentam em geral, genoma de RNA e são mantidos de forma cíclica na natureza replicando e circulando em hospedeiros vertebrados susceptíveis à infecção e artrópodes hematófagos, que atuam como vetores (11). Os vetores mais importantes na transmissão das arboviroses são os mosquitos, dentre eles: *Culex* spp., *Aedes* spp., e outros mosquitos dos gêneros *Haemagogus* e *Sabethes*. Além disso, outros artrópodes como carrapatos, moscas da areia “sandflies” (*Phlebotomus* spp.) e mosquitos pólvora “gnats” (*Culicoides* spp.) também são vetores importantes das arboviroses (12). Diferenças nas características morfológicas, antigênicas, bioquímicas e genéticas são utilizadas para classificar os arbovírus em família, gênero, subgrupos, espécie, subtipos em uma ordem crescente na taxonomia viral para cada grupo.

#### 2.1.1. *Flavivirus*

O gênero *Flavivirus* (família *Flaviviridae*) possui uma classificação complexa e pouco compreendida. A grande maioria dos patógenos naturais desse grupo de vírus são chamados de arbovírus (13). No entanto, também foram identificados *Flavivirus* sem vetor conhecido (do inglês, *no known vector* - NKV), ou que infectam apenas insetos (do inglês, *insect-specific flaviviruses*, ISFVs) (14).

Esses vírus são caracterizados por possuir uma partícula viral com simetria icosaédrica esférica, envolvida por um envelope viral composto por uma bicamada lipídica, com cerca de 50 nanômetros (nm) de diâmetro. Sua morfologia icosaédrica é conferida pelo arranjo das proteínas de superfície E e M (15). O genoma dos vírus pertencentes a este gênero, é constituído por uma molécula de RNA linear com polaridade positiva (+*ssRNA*), com aproximadamente 11 kilobases de comprimento, organizado em duas regiões não traduzidas (5' e 3' UTR) flanqueadas por uma única região de leitura aberta (RLA) que codifica um único polipeptídeo contendo aproximadamente 3400 aminoácidos. As proteínas codificadas pela região RLA são divididas em estruturais [proteína do capsídeo (C), proteína de membrana (prM), envelope (E)] e não estruturais (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) (Figura 1) (16,17).

**Figura 1** - Organização genômica do vírus Febre Amarela



A região de leitura aberta codifica três proteínas estruturais (C, M e E) e sete proteínas não estruturais (NS1-N5), flanqueada pelas extremidades 5' e 3' não codificantes.

Fonte: Figura adaptada e inspirada de Guzman *et al.* 2010 (17).

### 2.1.1.1. Vírus Febre Amarela

O vírus Febre Amarela (do inglês, *Yellow Fever virus* - YFV) é um *Flavivirus* pertencente ao gênero *Flavivirus* (família *Flaviviridae*). Constitui um grupo diverso de arbovírus que exibem relações filogenéticas e características biológicas semelhantes na interação vírus-hospedeiros (14). O YFV é a espécie viral protótipo do gênero *Flavivirus* (ICTV Virus Taxonomy Profile: *Flaviviridae*). Este vírus possui apenas um sorotipo e é classificado em quatro genótipos: (i) África Oriental, (ii) África Ocidental, (iii) América do Sul I e (iv) América do Sul II (18,19).

### 2.1.1.1.2. Ciclo de transmissão

O ciclo de manutenção e transmissão do YFV em condições naturais ocorre principalmente através da picada de mosquitos naturalmente infectados. Segundo Centros de Controle e Prevenção de Doenças dos EUA (do inglês, *Centers for Disease control and prevention* - CDC), existem dois ciclos de transmissão da febre amarela: (i) o ciclo silvestre; e (ii) o ciclo urbano. No ciclo silvestre, o vírus circula entre mosquitos e primatas não humanos (PNH) nas florestas tropicais e subtropicais do continente Americano e Africano. Os mosquitos infectados, geralmente dos gêneros *Haemagogus* e *Sabethes*, picam os PNH transmitindo o vírus. Os PNH atuam como hospedeiros amplificadores, mantendo o vírus em circulação na floresta. Na presença de PNH infectados com o YFV, mosquitos não infectados, durante o ato de hematofagia, tornam-se portadores do vírus fornecendo um ambiente celular permissivo à replicação viral. O ciclo urbano envolve a transmissão do YFV entre humanos e mosquitos em áreas urbanas ou semiurbanas. Nesse caso, o mosquito *Aedes aegypti* é o principal vetor. Quando uma pessoa infectada entra em uma área com uma população suscetível de mosquitos *Aedes aegypti*, esses mosquitos podem picar a pessoa e tornarem-se infectados com o vírus. Os

mosquitos infectados podem, então, transmitir o vírus para outras pessoas saudáveis através de suas picadas.

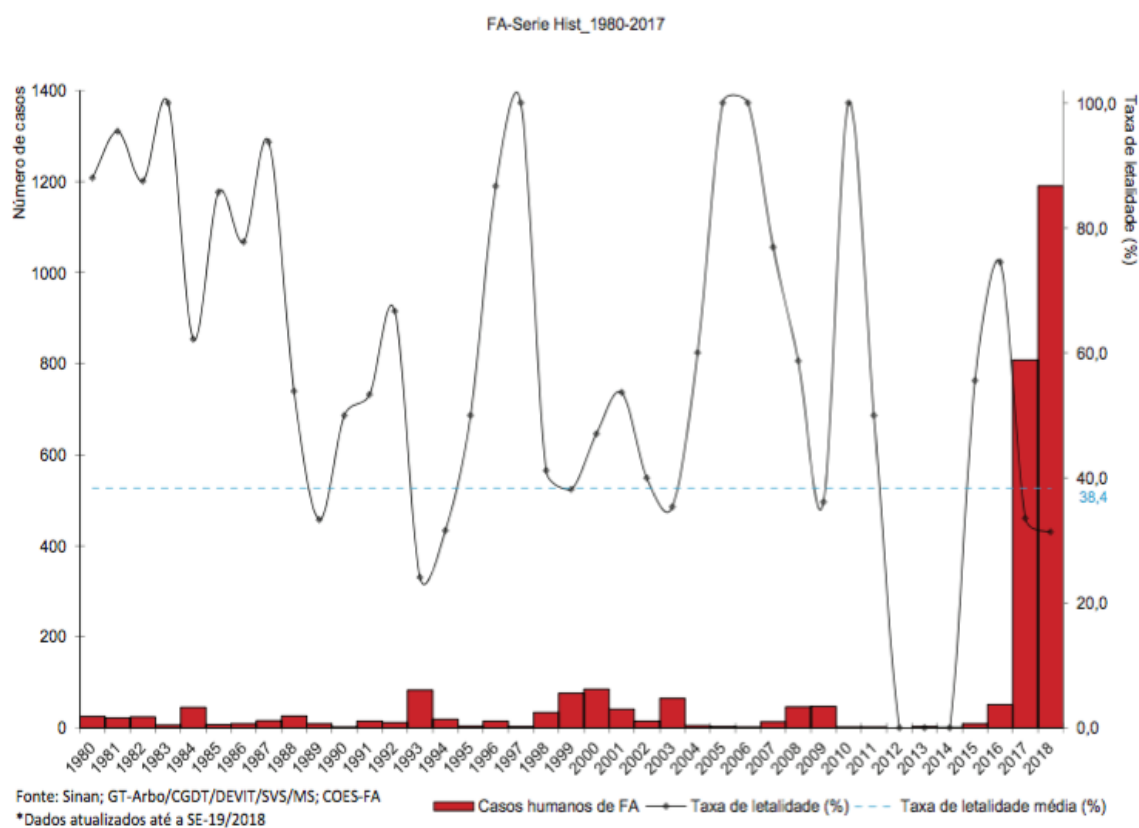
### **2.1.1.1.3. Epidemiologia**

O YFV é atualmente endêmico em 34 países na África e 13 na América do Sul e anualmente, e de acordo com a OMS, ocorrem cerca de 200 mil casos e 30 mil mortes com desfecho associado à YF, embora os números oficiais de notificações sejam menores ([www.who.int/csr/](http://www.who.int/csr/)). Destes, 90% ocorrem na África, uma vez que na América do Sul a taxa de transmissão é menor devido a alta cobertura vacinal que ocorre primordialmente devido ao sucesso das campanhas de vacinação (20).

Em epidemias que afetam comunidades não vacinadas, a mortalidade pode ultrapassar 50%, e não há nenhum tratamento direcionado para a doença, sendo apenas possível oferecer cuidados de apoio (7). Desde 1980, a YF vem reemergindo na África e na América do Sul, totalizando 18.735 casos reportados com 4.522 mortos entre 1987 a 1991 (21). Nos últimos 20 anos, a ocorrência da YF foi registrada em 9 países na América do Sul, que incluem Bolívia, Brasil, Colômbia, Equador, Peru, Venezuela, Guiana Francesa, Paraguai e Argentina (22).

Segundo a Organização Pan-Americana da Saúde (OPAS), entre Janeiro de 2016 e 13 de Março de 2018, sete países reportaram casos de YF, incluindo Bolívia, Brasil, Colômbia, Equador, Guiana Francesa, Peru, e Suriname (22). Preocupantemente, neste período o número de casos humanos causados por transmissão epizootica foi o maior observado em décadas nas Américas (3,4,23). No Brasil, segundo o Ministério da Saúde, entre Julho de 2017 a Junho 2018 foram confirmados 1.376 casos humanos de febre amarela com 434 casos evoluindo para o óbito (letalidade de 35.5%). Os estados mais afetados em ordem decrescente são: Minas Gerais (520 casos e 177 óbitos), São Paulo (516 casos e 163 óbitos), Rio de Janeiro (223 casos e 73 óbitos), Espírito Santo (6 casos e um óbito) e Distrito Federal (um caso fatal) (<http://portalms.saude.gov.br/>) (Figura 2).

**Figura 2 - Casos de Febre Amarela confirmados e letalidade no Brasil**



O gráfico demonstra o número de casos confirmados e a taxa de letalidade desde o ano de 1980 até maio de 2018.

Fonte: Figura adaptada e inspirada de <http://portalms.saude.gov.br/>.

#### 2.1.1.1.4. Diagnóstico clínico e laboratorial para YF

O diagnóstico da YF ou do YFV fundamenta-se na identificação de marcadores clínicos característicos para a doença, bem como na detecção de elementos presentes na partícula viral e/ou sinais da resposta imunológica, os quais se desenvolvem durante ou após o curso da doença (24–26). A primeira abordagem para orientar o diagnóstico para YF após o contato do paciente suspeito com a equipe médica é o levantamento da história clínica detalhada, exame físico aprofundado, descartar outras doenças febris endêmicas da região, e a realização de exames laboratoriais de apoio, quando disponíveis. Na fase inicial da doença, as características clínicas são compartilhadas por outras doenças infecciosas, principalmente as causadas por outros vírus, e o médico deve estar atento às características específicas da febre amarela, como a evolução rápida da doença para formas graves com a possibilidade de observar icterícia, que surge após o comprometimento das funções hepáticas (7).

No período de infecção aguda para a doença, o diagnóstico pode ser realizado através de métodos que objetivam identificar componentes da partícula viral, como a identificação de antígenos virais no soro, que pode ser realizada por meio de técnicas como o ELISA (do inglês, *Enzyme-Linked Immunosorbent Assay* - ELISA) ou a PCR (do inglês, *Polymerase Chain Reaction* - PCR) (26–29). Esses métodos possibilitam a detecção direta de componentes do vírus, permitindo diagnósticos precisos. O isolamento do vírus pode ser realizado a partir de amostras de sangue ou outros fluidos corporais, que tenham o vírus com integridade e quantidade suficiente para estabelecer uma infecção produtiva em um sistema celular ou animal (25). A técnica de PCR, e a sua variação em tempo real destaca-se como um método altamente sensível para a detecção do RNA viral. Isso viabiliza diagnósticos ágeis e altamente específicos, facilitando a identificação da presença do vírus, possibilitando até mesmo, quantificar a quantidade de moléculas de RNA viral presentes na amostra (27). Recentemente, o diagnóstico molecular utilizando amostras de urina revelou-se como uma estratégia útil para diagnóstico do vírus durante e após a fase aguda da doença (30,31).

Após alguns dias da infecção, os exames sorológicos, como ELISA e ensaios de neutralização, são empregados para verificar a presença de anticorpos específicos contra o YFV à partir do soro do paciente (32). A detecção de anticorpos imunoglobulinas da classe M (IgM) indica infecção recente, enquanto a identificação de imunoglobulinas da classe G (IgG) aponta para exposição prévia ao vírus, podendo essa exposição ser derivada de uma resposta à vacina, ou então derivada como uma resposta natural em exposições prévias ao vírus (33,34). A combinação de abordagens clínicas e laboratoriais oferece uma perspectiva abrangente para a detecção e confirmação da infecção causada pelo YFV.

#### **2.1.1.1.4. Patogênese viral**

O YFV provoca uma enfermidade viral infecciosa, que é transmitido aos seres humanos por meio da picada de insetos hematófagos. A gama de manifestações clínicas da doença associada a esse vírus pode variar desde uma forma subclínica até uma enfermidade sistêmica severa (25,35). Na sua forma clássica, apresenta-se causando sintomas como febre, cefaléia, mialgia, dores nas articulações, fadiga, náuseas e vômitos, enquanto que, estes sintomas podem acentuar-se para manifestações graves da doença, que incluem icterícia, hemorragias, insuficiência hepática e renal, complicações no miocárdio e alterações mentais. Nos pacientes sintomáticos, a YF caracteriza-se por uma elevada letalidade, 20-50 % dos casos (35).



A patogênese da doença causada pelo YFV pode dividir-se em dois padrões distintos de infecção: (i) viscerotrópico (capacidade de infectar diversos órgãos, incluindo fígado, baço, coração, rim, pulmão, entre outros) e (ii) neurotrópico (capacidade do vírus infectar o parênquima cerebral e causar encefalite) (36). Os primeiros relatos de danos neurológicos foram descritos por *Lasnet* 1927 com pacientes previamente infectados com YFV na África (2). Após a apresentação destes dados na Conferência Africana de Febre Amarela em 1929, *Jean Laigret*, então diretor do Instituto Pasteur, sugeriu que o sistema nervoso dos pacientes deveriam também ser testados para investigar a presença de vírus (2).

O conhecimento atual indica que, o YFV é detectado em diversos tecidos, sugerindo seu amplo espalhamento em nível sistêmico em casos fatais (3,37). Duarte-Neto *et.al.* propôs em seu estudo um modelo de patogênese e espalhamento do YFV após a inoculação dos vírus na pele pela picada do mosquito hematófago (37). Seguindo a inoculação nas células da derme, os vírus são transportados por células dendríticas da pele para os linfonodos para a apresentação de antígenos, o que caracteriza a viremia primária (surgimento dos sintomas virais inespecíficos). Somente após esta fase inicial o vírus passa a replicar em órgãos internos como fígado, baço, rins, coração, pulmão e cérebro, onde ocorre a viremia secundária, com passível detecção do RNA por reação em cadeia da polimerase quantitativa pós transcrição reversa (RT-qPCR). A fase toxêmica, ou segunda fase da doença ocorre geralmente do 6 ao 14 dia de doença e é caracterizada por hepatite, nefropatia, coagulopatia, vasculopatia e hemorragia.

#### **2.1.1.1.4. Vacina para YFV**

Em junho de 1927, teve início o processo de desenvolvimento da vacina YFV-17D a partir do isolamento do vírus silvático. Utilizando o sangue de um indivíduo africano com sintomas febris leves, esse sangue foi inoculado em um macaco *Rhesus*, que desenvolveu, pela primeira vez em ambiente laboratorial, os sintomas característicos da febre amarela (38,39). Esse vírus foi designado como linhagem “*Asibi*”, nome derivado do paciente do qual a amostra foi obtida. Ao mesmo tempo, em Dakar, o Instituto Pasteur de Dakar isolou o vírus silvático viscerotrópico francês (do inglês, *French Vicerotropic Virus* - FVV), a partir de uma amostra do sangue de um paciente, o qual provocou uma forma grave de febre amarela em um macaco *Rhesus* (40). *Max Theiler*, em 1930, descobriu que o FVV após passagens repetidas em cérebros de camundongos, levaria a uma redução sucessiva da patogenicidade do vírus em macacos e foi devido a estes experimentos preliminares, além de testes em cultura de células que a vacina

de vírus vivo atenuado 17D foi concebida em 1936 e deu à Theiler o primeiro prêmio Nobel relacionado à uma vacina de vírus (41).

Todas as vacinas são fabricadas a partir de extratos de embriões de galinha infectados, seguindo padrões idênticos (24,35) e suas sub-linhagens da vacina original são empregadas hoje para a produção das vacinas a 17DD no Brasil e 17D-204 em outros países, refletindo distintas passagens. Não são observadas diferenças significativas em termos de segurança ou imunogenicidade entre as vacinas 17D. A vacina da febre amarela 17D é amplamente empregada para prevenir a doença em viajantes, imunização de rotina em áreas endêmicas para bebês a partir de 9 meses, campanhas de atualização na África e como resposta emergencial durante surtos. Anualmente, entre vinte a sessenta milhões de doses são distribuídas.

De acordo com as diretrizes apresentadas pelo CDC, a vacina da febre amarela é recomendada para populações que vivem em áreas endêmicas do vírus YFV ou que têm planos de deslocamento para essas regiões. A vacina contra a YF, conhecida como YF-17D, é amplamente reconhecida por sua eficácia e segurança, tendo sido administrada a mais de 600 milhões de pessoas em todo o mundo nos últimos 65 anos (42–44).

No Brasil, essa vacina é produzida desde 1937 pelo Instituto de Tecnologia em Imunobiológicos (Bio-Manguinhos – Fiocruz), utilizando o isolado viral atenuado YFV-17DD, que é cultivado em ovos embrionados de galinhas. Uma única dose da vacina é capaz de induzir uma resposta imune efetiva, ativando linfócitos T citotóxicos, linfócitos T helper tipo 1 e 2, além de gerar anticorpos protetores de longa duração (42,45,46). Até o ano de 1941, não haviam sido relatados casos de reações adversas à vacina. No entanto, em Julho desse mesmo ano, foi descoberto que o vírus da febre amarela também poderia causar doença invasiva e disseminada em indivíduos saudáveis, resultando em um alto índice de letalidade (47). Foram notificados 192 casos de encefalite em uma população recém-vacinada, indicando que 0,5% dos vacinados apresentaram complicações neurológicas (48).

Um estudo de metanálise realizado com dados entre 2007 e 2013 relatou eventos adversos da vacina YF-17D, notificados ao Sistema Norte Americano de Notificações de Eventos Adversos. Esses eventos foram categorizados por sexo e idade, sendo classificados como anafiláticos (0,8 em 100 mil vacinações), sintomas neurológicos associados (0,4 em 100 mil vacinações) e doença viscerotrópica associada (0,3 em 100 mil vacinações). Essa análise permitiu concluir que aproximadamente 90% dos eventos adversos pós-vacinação não são considerados graves, sustentando o perfil de segurança da vacina (49,50).

### 2.1.1.1.5. Neuropatogênese viral

As infecções de células do sistema nervoso central (SNC) podem ser causadas por uma variedade de agentes biológicos, e entre os vírus, destacam-se os: (i) herpesvírus, (ii) retrovírus, (iii) arenavírus, (iv) picornavírus, e (v) *Flavivirus* (51). O modo de entrada desses vírus, sua replicação e especificidade pelo SNC continua a ser um assunto amplamente discutido, uma vez que relações entre proteínas do envelope viral e receptores celulares desempenham um papel essencial na determinação do tropismo celular e tecidual (52). Os processos infecciosos causados pelos *Flavivirus* são iniciados em nível sistêmico, com replicação prévia desses vírus em outras células e tecidos, e a partir desta amplificação prévia, esses vírus podem chegar no SNC e estabelecer um processo infeccioso nesse tecido, sendo este processo dependente da interação entre proteínas estruturais do vírus com componentes moleculares presentes na membrana plasmática celular e que atuam como receptores virais, e também de componentes moleculares presentes no meio intracelular atuam no processo de permissividade, responsável pela replicação viral eficiente.

O cérebro é um órgão de considerável complexidade biológica e molecular, constituído por uma diversidade de tipos e populações celulares que desempenham atividades biológicas específicas considerando suas regiões distintas, visando a adequada operacionalidade do sistema nervoso (53,54). Em condições homeostáticas, o substrato cerebral se constitui primordialmente de dois tipos celulares proeminentes: (i) os neurônios e (ii) as células gliais (54). Considerando a proporção de cada população celular no cérebro, Herculano-Houzel *et al.* (2009) mostraram a existência consistente de uma proporção de células gliais para neurônios no cérebro humano próxima a 1:1, onde o número total de células é cerca de 170,68 bilhões, das quais 86,1 bilhões são neurônios e 84,6 bilhões são células gliais (53), com diferentes proporções de células gliais para neurônios considerando as regiões distintas do cérebro. Os neurônios são as células especializadas na transmissão de sinais elétricos e químicos no sistema nervoso, acumulando atividades ligadas à recepção, processamento e transmissão de informações. Além dos neurônios, existem as células gliais, que são células não neuronais que têm diversas funções de suporte e manutenção essenciais para o funcionamento saudável dos neurônios e para o tecido cerebral. Os principais tipos de células gliais são: (i) os astrócitos, que regulam o ambiente iônico, o metabolismo energético e fornecem suporte estrutural aos neurônios, e atuam na formação da barreira hematoencefálica, protegendo o cérebro de substâncias prejudiciais; (ii) os oligodendrócitos, responsáveis pela criação da mielina, uma substância que isola os axônios dos neurônios e aumenta a velocidade dos impulsos elétricos;

(iii) a microglia, que funciona como um componente vital do sistema imunológico no cérebro, com atividades na resposta a lesões, infecções e inflamações cerebrais; (iv) células ependimárias, que são encontradas nos ventrículos cerebrais, produzem o líquido cefalorraquidiano, que envolve o cérebro e também a medula espinhal, oferecendo proteção e suporte mecânico; e (v) as células radiais, que participam ativamente no desenvolvimento cerebral, orientando a migração de neurônios durante o processo embrionário (55,56). O fato do cérebro ser composto por uma grande diversidade de células, torna a análise de expressão gênica desafiadora e interessante em nível sistêmico. Os vírus de RNA desencadeiam uma grande ativação de células da glia, incluindo astrócitos que podem acarretar em respostas imunológicas inatas e adaptativas. Estas respostas podem limitar a replicação viral, ou aumentar a neuroinflamação por repostas inflamatórias das células da glia que em sua maioria estão associadas a danos neurológicos (57).

Nos últimos 30 anos, poucos estudos foram desenvolvidos com o objetivo de estudar a neuroinvasividade do YFV, embora existam casos notificados que indicam a presença de sintomas neurológicos em pacientes pós-vacinais (58,59). Nos últimos anos, o principal arbovírus estudado dentro desta temática de neuropatogênese viral foi o vírus Zika (ZIKV), sendo o vírus com o mais completo conjunto de informações biológicas acerca deste processo. A neuropatogênese do ZIKV é um tema complexo, compreendendo os mecanismos biológicos envolvidos na interação entre o vírus e o sistema nervoso humano, e atualmente, sabe-se que este é o agente viral responsável por efeitos teratogênicos sérios como a microcefalia (60–62) e a síndrome de *Guillain-Barré* (63). Em ambos os casos, o vírus é capaz de atravessar barreiras importantes que envolvem sítios responsáveis por manter a integridade e o funcionamento adequado do feto e do cérebro, como a barreira placentária e/ou barreira hematoencefálica. O vírus perturba o desenvolvimento neural em fetos e causa desequilíbrios neurológicos em adultos (64). O ZIKV pode afetar o desenvolvimento neural fetal de diversas formas. Ao infectar células neurais e progenitoras neurais em desenvolvimento, o vírus pode comprometer a proliferação e migração dessas células, resultando em interrupções no desenvolvimento cerebral normal, o que pode culminar no desenvolvimento de microcefalia (65,66). Além disso, a infecção causada pelo ZIKV pode induzir a apoptose, causar inflamação no sistema nervoso e interferir na função mitocondrial, contribuindo para danos neurais e complicações neurológicas. A propagação do vírus para o sistema nervoso central é uma preocupação, pois o vírus pode chegar ao cérebro por meio de diferentes rotas, incluindo o sangue e através dos neurônios periféricos. A resposta imunológica desempenha um papel fundamental na

neuropatogênese, podendo ter efeitos tanto protetores quanto prejudiciais para o sistema nervoso (67).

#### **2.1.1.1.6. Vacina para YFV e sua relação com o neurotropismo viral**

Em 1927, após o desenvolvimento da vacina YFV-17DD a partir do isolamento do vírus “*Asibi*” (68), em Dakar, o Instituto Pasteur isolou o vírus silvático viscerotrópico francês (do inglês, *French Vicerotropic Virus - FVV*), que deu origem a vacina neurotrópica francesa (do inglês, *French Neurotropic Vaccine - FNV*). Theiler, em 1930 realizou estudos em camundongos com a linhagem FVV intracerebral e comprovou as propriedades neurotrópicas do FVV através da observação de infecção de células neuronais, encefalite viral e paralisia nestes animais (69). A vacina FNV foi descontinuada devido a sua ação neurotrópica, porém a observação sobre atenuação viral após passagens contínuas em animais, sugerida por Theiler foi utilizada para o desenvolvimento da vacina YFV-17D (68).

A vacina de vírus atenuado YFV-17D é considerada a de maior sucesso desenvolvida até então (70). No entanto, relatos de eventos adversos são notificados em baixa frequência e de maneira contínua desde o início de administração da vacina (48). Conforme os relatórios de sistema de notificação de eventos adversos a vacinas (VAERS), a frequência de morte associada a reação adversa à vacina varia de 1:250 mil a 1:500 mil doses administradas (49). Em 2001, sugeriu-se que casos fatais pós-vacinação são provenientes de uma doença sistêmica grave associada à replicação do vírus vacinal, sendo chamada doença viscerotrópica associada à vacina (do inglês, *associated vicerotropic disease - AVD*) a qual se assemelha a doença a qual foi projetada para prevenir (71). Além disso, existe a doença neurotrópica associada à vacina (do inglês, *associated neurotropic disease - AND*). A encefalite associada à vacina foi descrita pela primeira vez em 1950, quando foi relatado casos de encefalites em crianças que foram vacinadas com a vacina FNV (a qual foi descontinuada) (72) e as manifestações clínicas incluídas da doença são meningoencefalite, síndrome de *Guillain-Barré* e encefalomielite aguda disseminada (73).

Embora bastante estudado no início do século XX, poucos trabalhos foram realizados com o YFV relacionando dados neurológicos em pacientes infectados pelo vírus silvático.

## **2.2. Ação na neuroinflamação e encefalopatia hepática**

A encefalopatia hepática (do inglês, *hepatic encephalopathy* - HE) é uma das complicações mais importantes relacionadas à insuficiência hepática aguda (do inglês, *acute liver failure* - ALF), que pode estar presente na febre amarela (74). Esta síndrome já é conhecida há milhares de anos, desde *Hipócrates* (460 – 370 a.C.) e *Celsium* (25 a.C. – 50 d.C.), os quais foram os primeiros a descrever distúrbios comportamentais associados à insuficiência hepática (75), passando pelo grande prêmio nobel em 1890, em que a HE foi amplamente estudada mostrando o efeito da alta concentração de amônia no cérebro (76). Contudo, nos dias atuais, a etiologia da doença ainda é muito discutida e não totalmente elucidada.

A principal teoria que diz respeito à etiologia da HE, envolve insuficiência hepática aguda e crônica, resultando em uma inflamação sistêmica e hiperamonemia (77–79). Altas concentrações de amônia são tóxicas, e o fígado tem o papel de regular a concentração sistêmica deste, por meio do ciclo da uréia. Com a disfunção do fígado, níveis altos de amônia arterial chegam ao SNC e ultrapassam barreira hematoencefálica (BHE) (do inglês, *blood brain barrier* - BBB), atingindo o cérebro. As únicas células responsáveis por metabolizar amônia neste órgão são os astrócitos, transformando amônia em glutamina (80). Contudo, o acúmulo de lactato e glutamina, ativam a microglia, e conseqüentemente induzem a produção de citocinas pró-inflamatórias, levando ao quadro de neuroinflamação (81). Estudos recentes *in vivo* e *in vitro* mostram que a constante exposição do cérebro à citocinas pró inflamatórias, levam a um aumento da permeabilidade da barreira hematoencefálica, e isso, pode constituir uma possível rota para a entrada de patógenos no SNC (82,83).

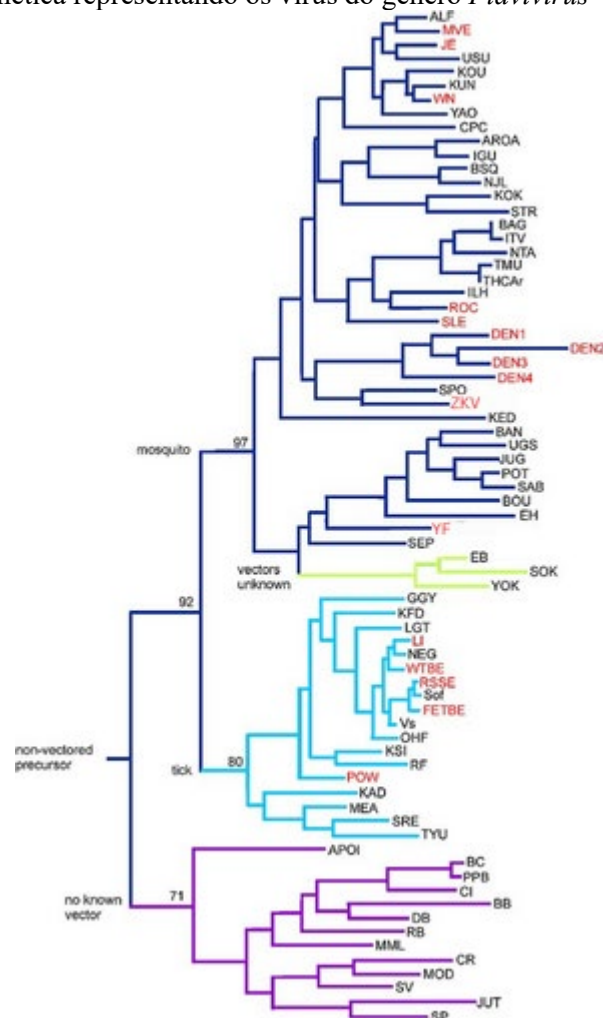
A BHE é uma barreira física e fisiológica que separa a circulação periférica, do SNC (84–86). Em condições fisiológicas, ela impede que componentes neurotóxicos periféricos tenham acesso ao cérebro, além de constituir um ambiente seletivo, regulando o transporte de moléculas para dentro e fora do SNC, controlando desta forma a composição química do microambiente neuronal (87).

Na inflamação sistêmica, a sua principal função é proteger os neurônios centrais concomitantemente a sinalização para o cérebro, que permite assim, uma resposta protetora na presença da doença (84). A BHE é composta por células endoteliais, complexos juncionais e envoltas por pericitos e prolongamentos astrocíticos (85). Os complexos juncionais são compostos pelas denominadas *tight junctions* (TJ), principais estruturas responsáveis por restringir o acesso ao SNC através do endotélio cerebral. Portanto, para a manutenção da homeostase cerebral e estabilidade da função e estrutura da BHE, um conjunto endotélio e TJ intactos são necessários.

### 2.3. Rotas de acessibilidade dos *Flavivirus* ao SNC

Os *Flavivirus*, mais prevalentes associados à síndromes neurológicas são: o vírus da encefalite japonesa (do inglês, *Japanese encephalitis virus* - JEV), o vírus do Nilo Ocidental (do inglês, *West Nile virus* - WNV), o vírus Zika (ZIKV), o vírus Dengue (DENV), o *tick-borne encephalitis virus* (TBEV) e o YFV (88). O desenvolvimento de síndromes neurológicas ou encefalopatia depende da capacidade do vírus atravessar a BHE e chegar ao SNC, processo conhecido como neuroinvasividade, além de infectar células neurais, processo conhecido como neurovirulência. Um fato interessante, é que ambas habilidades mostram-se dispersas na árvore filogenética entre vários membros do gênero *Flavivirus* (Figura 3).

**Figura 3** - Árvore filogenética representando os vírus do gênero *Flavivirus*



A árvore filogenética resalta em vermelho, aqueles vírus com capacidade de neuroinvasividade e neurovirulência.

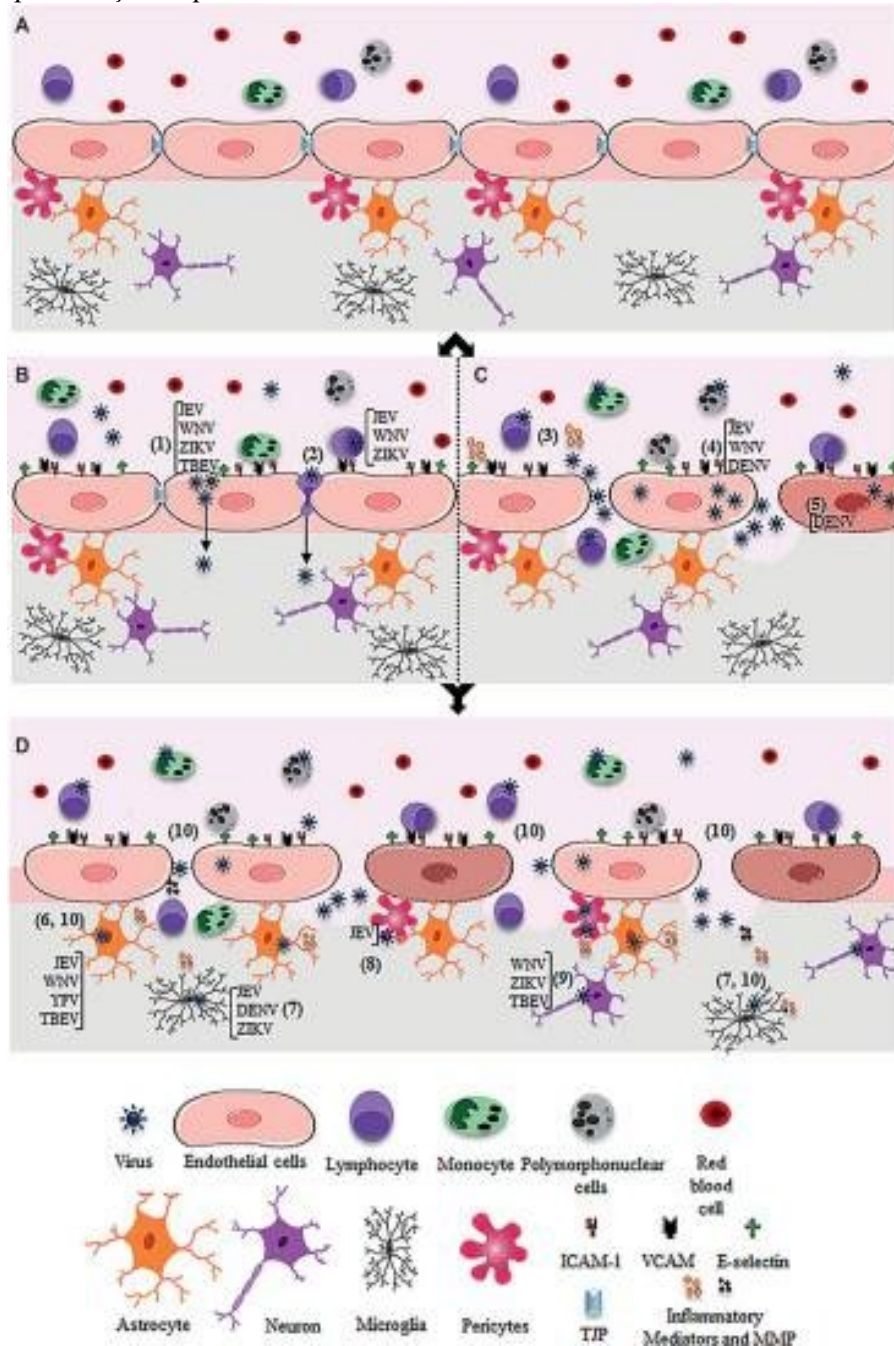
Fonte: Figura adaptada e inspirada de Sips *et al.* 2012 (89).

Os *Flavivirus* podem chegar ao SNC por diferentes rotas, sendo elas via sistema nervoso periférico, transporte axonal ou via hematogênica. O YFV atinge o SNC via rota hematogênica (90). Após contato com a BHE, o YFV pode infectar células endoteliais e atravessar a barreira endotelial como vírus livre de células, ou devido a regulação negativa da expressão de proteínas das *tight junctions* (TJ) que promovem a ruptura da barreira e entrada do vírus (Figura 4). Concomitantemente, a inflamação sistêmica ativa células imunes que liberam citocinas que também afetam a permeabilidade endotelial. Após a entrada no cérebro, o YFV infecta primordialmente astócitos que induzem mais liberação de mediadores inflamatórios que interferem diretamente no espaçamento das TJ, afrouxando ainda mais o espaço entre elas permitindo além da entrada de vírus livre de células, a entrada de leucócitos periféricos infectados via mecanismo denominado “Cavalo de Tróia”, em que uma ruptura na



estrutura da BHE ou alterações na TJ permitem a migração paracelular de células que em condições fisiológicas normais estariam restritas (88).

**Figura 4** - Representação esquemática dos mecanismos de entrada de *Flavivirus* no SNC via BHE



(A) Barreira hematoencefálica intacta, composta por células endoteliais aderidas pelas proteínas das TJ, associadas a pericitos, astrócitos e microglia. (B) Flavivirus atingem a BHE por via hematogênica e podem atravessar a BHE como vírus livres de células e via cavalo de tróia. (C) Quando o vírus atinge o cérebro, o YFV infecta astrócitos que passam a liberar citocinas que influenciam na expressão negativa de proteínas das TJ, aumentando o fluxo de vírus para dentro do cérebro.

Fonte: Figura adaptada e inspirada de Mustafá *et al.* 2019 (88).

### 2.3. Análise de expressão gênica

A análise de expressão gênica é uma das principais áreas da biologia molecular, que permite estudar quais genes estão ativos e em que níveis em um determinado momento ou condição. Historicamente, essa área passou por várias fases de desenvolvimento e inovação, permitindo grandes contribuições sobre os processos biológicos. Seu início, ainda na década de 1970 foi com a técnica de hibridização *in situ*, que permite identificar sequências de RNA específicas em tecidos (91–93). Já na década seguinte, a introdução do *Northern Blot*, uma técnica que permitia a separação e detecção de RNA por tamanho, tornou possível quantificar a expressão de genes específicos em diferentes condições, porém, tendo como principal limitação a quantidade de genes analisados (94–97). Na década de 1990, o advento da reação em cadeia da polimerase (PCR) revolucionou a análise de expressão gênica, e posteriormente o desenvolvimento de *microarranjos* de DNA também permitiu a análise simultânea (agora) de milhares de genes em um único experimento (98,99).

Desde o início da década de 2000, as tecnologias de sequenciamento de nova geração (do inglês, *Next Generation Sequencing* - NGS) revolucionaram a área de análise de expressão gênica. O RNA-Seq, uma técnica baseada em NGS, permite a quantificação e a identificação de todos os transcritos de RNA em uma amostra, incluindo aqueles não previamente conhecidos. Isso inclusive, abriu portas para descobertas de novos genes e isoformas. A sua aplicação para o estudo de doenças infecciosas, abriu novas oportunidades para melhorar a aplicabilidade diagnóstica por meio da busca por marcadores preditivos para doenças, prognóstica e terapêutica em várias doenças (100). O sequenciamento de RNAs, por meio do RNA-Seq, é uma ferramenta poderosa para análise do perfil de transcritos presente em uma determinada condição, sendo esse perfil obtido a partir dos dados de sequenciamento com elevada cobertura (101,102). O seu funcionamento baseia-se na utilização de leituras (*reads*) que são derivados do sequenciamento dos RNAs presentes em uma amostra, e após filtros de qualidade, esses são alinhadas a um genoma de referência ou transcritos de referência, ou então, montadas *de novo* (sem a sequência genômica como molde), e então, produz-se um mapa de transcrição em escala de genoma que consiste na estrutura de transcrição e no nível de expressão correspondente a cada gene (101,103).

Atualmente, a geração desses dados é realizada utilizando uma plataforma de alto rendimento, como nas plataformas de sequenciamento Illumina. Esse processo gera milhões de *reads* curtos (25 a 300 pb) recuperados de ambas extremidades dos fragmentos de cDNA, e

conhecidas como *paired-end reads*. Os dados brutos consistem em uma longa lista de sequências curtas, contendo os índices de qualidade associados.

Desde a década de 2010, a análise de expressão gênica começou a se integrar com outras áreas, como genômica funcional, transcriptômica, análise de rede, entre outras áreas (104,105). Também, novas abordagens de bioinformática passaram a permitir a interpretação de dados de expressão gênica em contextos biológicos mais amplos, revelando interações entre genes e vias regulatórias, explorando principalmente as análises de rede (106). Atualmente, as novidades no campo da análise de expressão gênica continuam a evoluir, e as principais novidades estão concentradas na utilização de tecnologias de sequenciamento de terceira geração para a geração dos dados, como o sequenciamento de *reads* longos, que poderá ajudar principalmente a entender de maneira mais detalhada transcritos complexos e a resolução de estruturas tridimensionais do genoma em relação à sua expressão (107). Outra novidade emergente, é a transcriptômica espacial, que visa analisar a expressão gênica em relação à localização espacial dentro de tecidos biológicos complexos (108–111). Por meio de tecnologias até mesmo clássicas, como hibridização *in situ* e abordagens baseadas em sequenciamento de RNA de alta resolução espacial, é possível mapear a atividade gênica em nível celular, revelando padrões de expressão em diferentes regiões do tecido. Isso oferece contribuições valiosas sobre o desenvolvimento de órgãos, a heterogeneidade celular e as bases moleculares de doenças, com amplas aplicações em pesquisa biológica e biomédica (108–111).

#### **2.4. Biologia de sistemas em Virologia**

Denis Noble, biólogo britânico e professor da Universidade de Oxford em um dos seus livros chamado “*The music of life*”, define biologia de sistemas como: “*Systems biology... is about putting together rather than taking apart, integration rather than reduction. It requires that we develop ways of thinking about integration that are as rigorous as our reductionist programmes, but different.... It means changing our philosophy, in the full sense of the term*”. (112). A definição dada pelo *National Institutes of Health*, diz que biologia de sistemas é uma abordagem da pesquisa biomédica que busca compreender e caracterizar de maneira holística, organismos biológicos em todos os seus níveis, ou seja, em nível de organismo, tecido ou célula, juntando todas as “peças” do sistema.

O termo “virologia de sistemas” é um novo termo aplicado à integração da aplicação de técnicas moleculares de alto rendimento (NGS) e ferramentas computacionais com a intenção de estudar aspectos de doenças virais. Abordagens integradas de sistemas, incluindo

experimentos *in vitro*, transcriptômica, metabolômica, proteômica, junto com abordagens computacionais e modelos matemáticos podem ser aplicadas para estudar as complexas interações vírus-hospedeiro gerando informações relevantes para possibilitar elucidar a biologia da doença e conseqüentemente fornecer um repertório fundamentado para aplicação em medicina preventiva de combate à infecções virais (113).

#### 2.4.1. Análises de dados biológicos explorando análise de redes

A teoria de grafos consiste na utilização de abordagens matemáticas de redes, o qual representa interações funcionais em um sistema (114,115). Esta teoria foi dada em 1736 na Rússia, pelo matemático Suiço *Leonard Euler*. A cidade russa de Königsberg, é atravessada pelo Rio Pregel e possui duas ilhas conectadas entre si por sete pontes. Euler, propôs um modelo matemático de como caminhar pela cidade atravessando cada ponte uma única vez e esse modelo ficou conhecido como “*Königsberg bridge problem*”, em que foi adaptado e transformado o mapa da cidade com as margens do rio e ilhas em nós e as pontes em conectores da rede (116). Por mais que *Euler* não tenha encontrado solução para o problema, a sua metodologia de análise combinatória foi um evento histórico para o ramo matemático e da postulação da teoria de grafos. Hoje, a teoria do grafos é aplicada de diversas formas multidisciplinares, e em biologia de sistemas serve para compor a análise de dados ômicos de forma a estudar o grau de conectividade entre genes, compostos, proteínas, fornecendo informações acerca da atividade biológica de entidades que podem apresentar-se como funcionalmente relacionadas (117,118). Também, essa abordagem pode ser utilizada para explorar o grau de centralidade, portanto maior interação epistática, que caracteriza uma molécula com maior influência sobre a rede.

Para entender como ocorre o relacionamento de diversas entidades biológicas em um determinado sistema, é necessário entender como se organiza um grafo. Um grafo é definido por uma dupla de  $G(V, E)$ , onde  $V(G)$  é um conjunto de objetos denominados vértices e  $E(G)$  é um conjunto de pares não ordenados de  $V$ , chamados de arestas (do inglês, *edges*) em que os nós ou vértices estão conectados por arestas ( $G$ ). As definições e conceitos da Teoria dos Grafos podem ser mais efetivamente compreendidos por meio de representações visuais, onde cada ponto no plano representa um vértice e cada linha que conecta um par de pontos representa uma aresta. Em bioinformática, as redes ponderadas amplamente usadas são aquelas cujas representações possuem pesos ou atributos, e portanto, são mais complexas e informativas (119).

Assim, as propriedades das redes são empregadas em análises de dados biológicos a fim de obter informações sobre as diversas possíveis interações entre seus componentes. A estrutura das redes possui como característica o grau de nó, ou conectividade. Isso, por sua vez, depende da interação entre nós através de conectores. A avaliação da densidade de uma rede implica em avaliar o grau de conectividade. Em alguns modelos de rede, os nós tendem a formar grupos ou módulos. A modularidade é um conceito que descreve padrões de conectividade, nos quais os seus componentes individuais estão agrupados em conjuntos altamente interconectados. Em termos gerais, a modularidade é um princípio de união que ocorre em diferentes tipos de elementos e conexões naturalmente estabelecidas no contexto biológico, como na interação entre indivíduos da mesma espécie. Em um nível molecular, podemos observar, por exemplo, elementos (como genes) que desempenham funções em um mesmo processo biológico (119).

A análise da centralidade permite a avaliação dos elementos mais significativos em uma rede de interação. Esse conceito tem como objetivo identificar elementos de relevância na configuração global da rede, considerando todas as suas estruturas, como centros de alta conectividade (*hubs*), módulos e a análise de intermediação (*betweenness*). A avaliação topológica da rede pode ser realizada por meio do grau de nó, que representa o número de conexões diretas entre um nó e outros nós. Nós que apresentam uma maior quantidade de conexões, que a média geral, são chamados de *hubs*. Nós classificadas como *party hubs* são aquelas altamente ligadas dentro do seu próprio módulo, com alto grau de *betweenness* e consideradas componentes funcionais da rede. E, os *date-hubs*, aquelas que se ligam em diferentes módulos, e portanto, desempenhando uma função abrangente na rede como um todo (120). Assim, o *betweenness* de um conector é um parâmetro definido como a contagem de caminhos mais curtos que atravessam um único nó, representando a interconexão existente entre eles. Por exemplo, um processo biológico dado do enriquecimento de genes diferencialmente expressos, com valor alto de *betweenness centrality* apresentaria uma capacidade elevada de interação com outro processo biológico ou *cluster*. Um gene, ou processo biológico com esta característica é chamado de *bottleneck* ou gargalo (120).

## *Capítulo 3*

### 3. OBJETIVOS

#### 3.1. Objetivo geral

Avaliar a interação vírus-hospedeiro no tecido cerebral em casos fatais de Febre Amarela, descrevendo a resposta sistêmica do hospedeiro: (i) no que diz ao controle de acessibilidade do YFV ao tecido cerebral independente da linhagem viral, (ii) à perturbação na expressão gênica de genes codificantes de proteínas quando comparados pacientes positivos para o YFV-silvático com o YFV-17DD no tecido cerebral, e (iii) à diferença de resposta sistêmica associada aos pacientes positivos para o YFV-silvático e para o YFV-17DD, porém com RT-qPCR negativos para investigar padrões de resistência à permissividade viral ao tecido cerebral associados à cada linhagem viral.

#### 3.1. Objetivos específicos

(i) Caracterizar por métodos moleculares (RT-qPCR) a positividade e a quantidade de RNA do YFV identificado do tecido cerebral de pacientes com suspeita clínica de morte associada ao YFV no surto de 2017-2018 ocorrido no estado de São Paulo;

(ii) Analisar o padrão de alterações patológicas para descrever alterações macroscópicas e microscópicas resultantes do processo infeccioso no tecido cerebral;

(iii) Analisar por imunohistoquímica a distribuição espacial de positividade do YFV por meio da identificação de seus genes estruturais no tecido cerebral infectado com o YFV-silvático e YFV-17DD;

(iv) Sequenciar os RNAs codificantes nos pacientes positivos para o YFV (considerando as diferentes classes) para identificar diferenças quantitativas dos mRNAs no tecido cerebral proveniente de pacientes com morte associada ao YFV;

(v) Realizar análise de ontologia gênica com genes diferencialmente expressos em diferentes condições;

(vi) Integrar os dados moleculares, expressão gênica e patológicos para compreender de maneira holística, a resposta do hospedeiro durante a infecção causada pelo YFV.

## *Capítulo 4*



## 4. MATERIAIS E MÉTODOS

### 4.1. Aspectos éticos

O estudo foi submetido e aprovado pelo Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (CEP-HC-FMUSP) com o número do CAPPesq: #426.643. Este, trata-se de um estudo retrospectivo, descritivo, utilizando como fonte de dados os prontuários, laudos de exames laboratoriais, radiológicos, bem como cortes histológicos de casos autopsiados, suspeitos ou confirmados como Febre Amarela, pelo Departamento de Patologia da FMUSP. Todos os pacientes encaminhados para autópsia preencheram critérios clínicos e epidemiológicos, e foram submetidos à confirmação diagnóstica laboratorial molecular e patológica. Os pacientes com o óbito confirmado, foram encaminhados para o exame *post mortem* e as autópsias foram realizadas após a assinatura do termo de livre consentimento esclarecido (TLCE) pelos familiares do paciente.

### 5.2. Material de estudo

As amostras foram coletadas durante a epidemia causada pelo YFV em São Paulo, entre 2017-2018 em pacientes suspeitos de morte associada ao YFV. Para a amostragem completa, foi realizada a coleta de diferentes tecidos (coração, pulmão, cérebro, rim, baço, pâncreas, e fígado), e aqui no presente estudo, consideramos os resultados moleculares obtidos para os tecidos, mas para a análise de expressão gênica diferencial, consideramos apenas as amostras clínicas do tecido cerebral, coletado à partir do lobo frontal. Todas as autópsias foram realizadas pelo Serviço de Verificação de Óbitos da Capital da Universidade de São Paulo (SVOC-USP). A coleta das amostras foi realizada seguindo o protocolo de *Letulle* (ou *en masse*) ou *Virchow* (órgãos removidos individualmente) e sistematicamente examinados. Para os ensaios moleculares, foram coletadas amostras de tecidos medindo 1.0 cm<sup>3</sup> para diagnóstico molecular, armazenadas em tubo Falcon de 15 ml. Após a realização da coleta dos espécimes biológicos destinados ao ensaio molecular e também à caracterização do padrão de expressão gênica, todas as amostras foram transportadas ao Laboratório de Evolução Molecular e Bioinformática do Instituto de Ciências Biomédicas da Universidade de São Paulo (LEMB/ICB/USP) aclimatizadas à -80°C até a realização dos ensaios moleculares.

## 4.2. Definição de caso

A definição de caso suspeito de infecção causada pelo YFV foi estabelecida pelo Ministério da Saúde e pelo Departamento de Saúde do Estado de São Paulo e incluiu: pacientes com febre alta de início súbito associada a icterícia e/ou hemorragia, que viviam ou haviam visitado áreas com epizootias de febre amarela (isto é, áreas com relatos de casos de YFV em PNH ou de isolamento do YFV em vetores), independentemente do status vacinal para febre amarela, durante os 15 dias precedentes. Os casos confirmados apresentavam clínica compatível e confirmação laboratorial por pelo menos um dos seguintes métodos: (i) sorologia positiva para IgM (MAC-ELISA); (ii) detecção de RNA viral por RT-qPCR em amostras de sangue; (iii) isolamento de vírus; (iv) histopatologia compatível com hepatite por YFV com antígeno detectável em tecidos pela técnica de imunoistoquímica. Todos os casos receberam o diagnóstico laboratorial definitivo para YFV pelo Instituto Adolfo Lutz (IAL), um Laboratório Nacional em Saúde Pública e Laboratório de Referência Macroregional pelo Ministério da Saúde brasileiro, com sede em São Paulo. Em todos os casos, o vírus das Hepatites A (HVA), B (HBV), C (HVC), citomegalovírus (CMV), Herpes (HSV), Dengue (DENV), Chikungunya (CHIKV) e Imunodeficiência Humana 1 (HIV-1), Leptospirose e outras doenças não infecciosas com etiologias para hepatite aguda foram excluídos por métodos de diagnóstico clínico, epidemiológicos (incluindo dados demográficos, condições médicas preexistentes, sinais e sintomas clínicos e acompanhamento intra-hospitalar até a morte) e características laboratoriais foram coletados dos prontuários médicos. Pacientes com testes negativos tanto para YFV quanto para outras doenças infecciosas, foram incluídos no grupo controle para estudo de expressão gênica do hospedeiro.

## 4.3. Caracterização molecular

A extração do RNA foi realizada utilizando o reagente TRIzol (Life Technologies, Grand Island, NY, EUA) seguindo o protocolo recomendado pelo fabricante. A triagem molecular para Febre Amarela foi realizada com o uso do reagente RT-PCR One-Step AgPath-ID (Ambion, Austin, TX, EUA), com o uso de *primers* e sondas específicos para o YFV (considerando tanto o YFV-silvático quanto o YFV-17DD) (RP-YFV - 5'-GTCRRTTCTCTGCTAATCGCTCA-3'; FP-YFV - 5'-ATTGAGGTGYATTGGTCTGC-3'; e a sonda P-YFV - FAM-5'-AGTTGCTAGGCAATAAA-3') (28). Para identificação dos casos positivos para YFV-17DD, e caracterizados como resposta adversa à vacina, foram utilizados

*primers* e sondas específicos para o vírus vacinal (YF-NS5\_F - 5'-GCACGGATGTAACAGACTGAAGA-3'; YF-NS5\_R - 5'-CCAGGCCGAACCTGTCAT-3'; e a sonda YF-NS5Probe - 5'-FAM-CGACTGTGTGGTCCGGCCCCATC-3') (121). A reação de RT-qPCR foi realizada com a etapa de transcrição reversa a 45°C por 10 min, ativação enzimática a 95°C por 10 min, seguida de 40 ciclos de amplificação a 95°C por 15 segundos e 60°C por 45 segundos para hibridação e extensão, no equipamento de ABI7500 (Thermo Fisher Scientific, Waltham, MA, EUA). À partir dos dados obtidos com o RT-qPCR, extraímos a condição associada à positividade (YFV detectado no tecido cerebral ou YFV não detectado no tecido cerebral) e os valores de Ct, que refletem a quantificação do RNA viral em cada amostra pelo valor de threshold cycle (do inglês, *Cycle threshold (Ct) value*).

#### 4.4. Imunohistoquímica

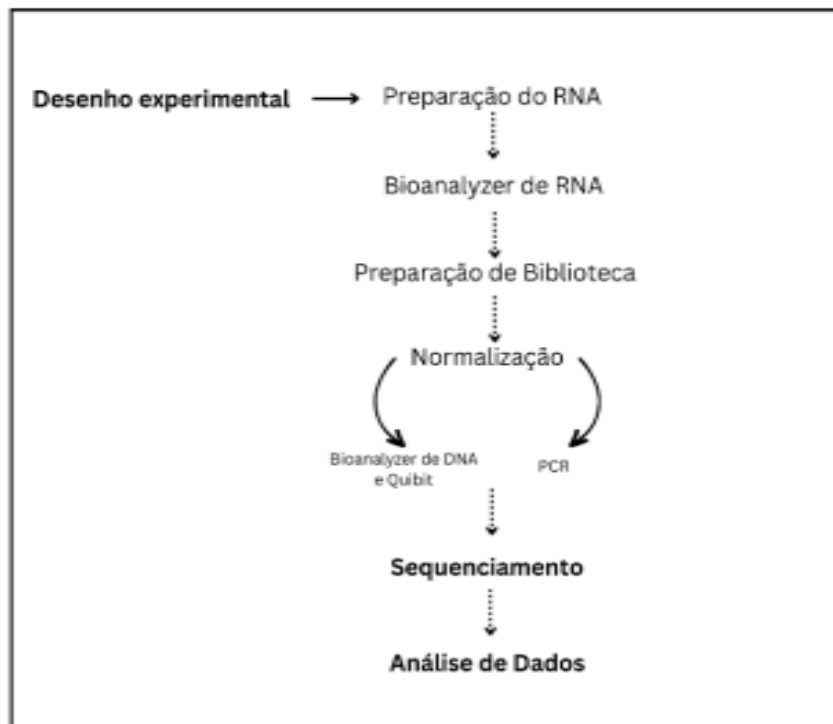
Para a análise de imunohistoquímica (IHC, do inglês, *immunohistochemistry*), foi utilizado o anticorpo primário fluido Ascítico de Camundongo anti-YFV (imunoascite), específico para o vírus (policlonal anti-febre amarela), fornecido pelo Institut Pasteur de Dakar. Este anticorpo foi originalmente padronizado e utilizado para a realização de ensaio imunoenzimático (ELISA) e validado pelo departamento de patologia da USP na diluição otimizada de 1:20.000. Para validação de anticorpos, amostras de fígado de casos confirmados de Febre Amarela foram usadas como controles positivos e casos negativos de FA, incluindo amostras de biópsia e autópsia com diferentes doenças infecciosas (doenças virais e bacterianas) e esteato-hepatite como controles negativos. As células foram coradas e contadas em 30 campos de alta potência ( $\times 400$ ) e foram feitas comparações entre pacientes com FA e controles. Avaliamos a presença de angiopoietina-2 (Santa Cruz Biotechnology Cat# sc-74403, RRID:AB\_1118956, diluição 1:50) semiquantitativamente por IHC, como 0 (ausente), 1+ (<10%), 2+ (11% -50%) e 3+ (>50%), de acordo com o número de vasos corados no corte histológico. Em todas as análises de IHC, usamos Kit de Detecção de IHC HRP/DAB Específico para Camundongos e Coelhos - Micropolímero (ABCAM, AB\_236466), cromogênio diaminobenzidina (DAB) e contracoloração de hematoxilina de *Harris*.

#### 4.5. RNA-seq

O RNA-Seq, por meio do sequenciamento dos RNAs totais ou pelo sequenciamento dos mRNAs, é uma abordagem desenvolvida para traçar o perfil de expressão gênica, usando

tecnologias de sequenciamento de alta profundidade. Aqui neste estudo, o RNA-seq consistiu nas etapas descritas na Figura 5.

**Figura 5** - Etapas necessárias para um experimento padrão de RNAseq



As etapas entre o desenho experimental e o sequenciamento, são realizadas em bancada, enquanto a análise de dados é realizada utilizando recursos de bioinformática.

#### 4.5.1. Desenho experimental e hipóteses consideradas no estudo

Para alcançar os objetivos propostos, o RNAseq foi realizado a partir de 13 amostras de tecido cerebral, obtidos de 10 pacientes que foram a óbito com suspeita de febre amarela, e 3 pacientes que fazem parte do grupos considerados controles (sendo casos de óbito não relacionado à YF) (Tabela 1).

**Tabela 1** - Caracterização da amostragem utilizada no estudo. Os pacientes estão distribuídos em cinco grupos, que representam: 1 = YFV-Silvático (Cérebro positivo); 2 = YFV-Silvático (Cérebro negativo); 3 = YFV-17DD (Cérebro positivo); 4 = YFV-17DD (Cérebro negativo); 5 = controles. As informações clínicas e laboratoriais associadas aos grupos também estão contidas na tabela.

Grupo	ID do Paciente	Idade (anos)	Sexo biológico	Diagnóstico (RT-qPCR)	YFV	Valor de Ct no cérebro
1	<b>YFV029</b>	48	F	Positivo	Silvático	26,02
1	<b>YFV031</b>	25	M	Positivo	Silvático	31,32
	Média da idade	37				
2	<b>YFV020</b>	51	M	Positivo	Silvático	Não detectado
2	<b>YFV049</b>	34	M	Positivo	Silvático	Não detectado
2	<b>YFV036</b>	27	F	Positivo	Silvático	Não detectado
	Média da idade	34				
3	<b>YFV012</b>	49	M	Positivo	17DD	25,78
3	<b>YFV057</b>	66	F	Positivo	17DD	24,20
	Média da idade	57,5				
4	<b>YFV037</b>	19	M	Positivo	17DD	Não detectado
4	<b>YFV050</b>	37	M	Positivo	17DD	Não detectado
4	<b>YFV051</b>	32	M	Positivo	17DD	Não detectado
	Média da idade	32				
5	<b>YFV006</b>	50	M	Negativo	Controle	Negativo
5	<b>YFV071</b>	82	F	Negativo	Controle	Negativo
5	<b>YFV083</b>	43	M	Negativo	Controle	Negativo
5	Média da idade	50				

A combinação de grupos foi realizada com o objetivo de testar as seguintes hipóteses:

(i)  $H_0$ : não existe diferença de expressão gênica entre pacientes com detecção do YFV no cérebro quando comparados àqueles em que não houve detecção do vírus.

$H_1$ : existe diferença de expressão gênica entre pacientes que tiverem o vírus da febre amarela detectado no cérebro e aqueles em que não houve detecção do vírus no mesmo tecido.

Para isso, os grupos 1  $\cup$  3 foram agrupados (YFV detectado no cérebro) e 2  $\cup$  4 (YFV não detectado no cérebro), e a análise foi realizada sempre mantendo o grupo controle (grupo 5), e representa o grupo de pacientes contendo expressão gênica basal e não relacionada

com a Febre Amarela. Assim, a primeira análise foi feita comparando valores de expressão gênica em TPM (do inglês, *transcript per million*) do grupo 1 U 3 e do grupo 2 U 4.

(ii) *H<sub>0</sub>*: não há diferença de expressão gênica entre pacientes que tiveram o vírus da febre amarela silvático e vacinal detectado no cérebro.

*H<sub>i</sub>*: existe diferença de expressão gênica entre pacientes que tiveram o YFV-silvático e YFV-17DD detectado no cérebro.

Para essa investigação, a análise foi realizada comparando valores de expressão gênica em TPM, do grupo 1 *versus* grupo 5 e grupo 3 *versus* grupo 5.

(iii) *H<sub>0</sub>*: não existe diferença de expressão gênica entre pacientes que não tiveram o YFV-silvático e YFV-17DD detectado no cérebro.

*H<sub>i</sub>*: existe diferença de expressão gênica entre pacientes que não tiveram o YFV-silvático e YFV-17DD detectado no cérebro.

Para esta, a análise foi realizada comparando valores de expressão gênica em TPM, do grupo 2 e grupo 4 *versus* grupo 5.

#### **4.5.2. Preparação do RNA (Purificação e quantificação)**

O RNA previamente extraído, foi purificado com DNase I e concentrado usando o kit RNA Clean and Concentrator TM-5 (Zymo Research, Irvine, CA, EUA) seguindo as instruções do fabricante. O RNA viral foi quantificado utilizando o equipamento Qubit® 3.0 Fluorometer (Invitrogen, Carlsbad, Califórnia, EUA).

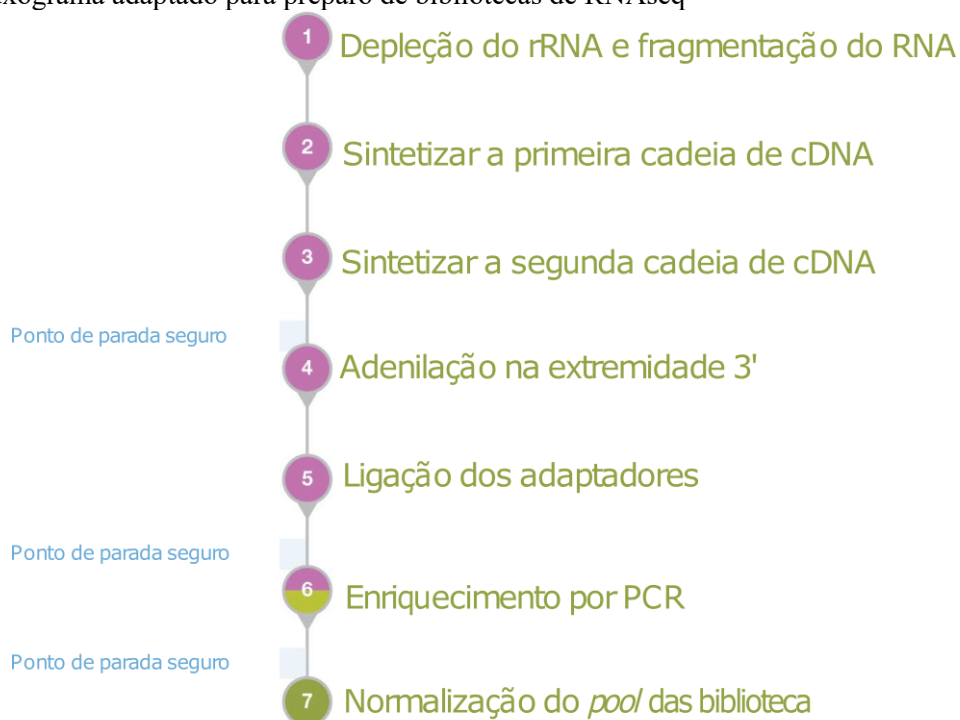
#### **4.5.3. Bioanalyzer**

A qualidade do RNA foi avaliada usando o equipamento Bioanalyzer da Agilent Bioanalyzer e posteriormente a amostra foi armazenada à - 80 °C até sua utilização. Para amostras de RNA degradadas, o ajuste do tempo de fragmentação ajuda a evitar a fragmentação excessiva das amostras de RNA.

#### **4.5.3. Preparo das bibliotecas para RNAseq**

O preparo das bibliotecas foi realizado utilizando o kit TruSeq Stranded Total RNA HT Sample Prep Kit (Illumina, San Diego, CA) seguindo as recomendações do fabricante (Figura 6).

**Figura 6** - Fluxograma adaptado para preparo de bibliotecas de RNAseq



Fluxograma adaptado do manual de referência da Illumina para preparo de bibliotecas de RNAseq que passa pelos seguintes processos: (i) depleção do rRNA do hospedeiro utilizando oligonucleotídeos alvo específicos combinados com beads de rRNA ribo-zero que depletam rRNA citoplasmático e mitocondrial. (ii) síntese de cDNA a partir do RNA total livre de rRNA. Os fragmentos de RNA livre são clivados e copiados para primeira fita de cDNA usando a enzima transcriptase reversa e PRIMERS. (iii) síntese da fita complementar de DNA a partir do cDNA, utilizando DNA polimerase I. (iv) fragmentação e adenilação das extremidades 3', em que um único nucleotídeo de adenila é adicionado na extremidade 3' para evitar a formação de quimera durante a fase de ligação de adaptadores. (v) ligação de adaptadores. (vi) PCR para enriquecimento específico de moléculas com os adaptadores (vii) Normalização e preparo de *pools*.

Fonte: Figura adaptada e inspirada de *Illumina*.

#### 4.5.4. Quantificação e normalização das bibliotecas montadas

Após o preparo das bibliotecas, estas foram quantificadas utilizando dois métodos principais: (i) Qubit 3.0 Fluorometer (Thermo Fisher), (ii) PCR em tempo real (NEBNext® Ultra™ II DNA Library Prep Kit for Illumina). Para a normalização das bibliotecas montadas, foi utilizado a quantificação obtida pelo método de qPCR e os tamanhos médios das bibliotecas, que foram obtidos por meio da análise do DNA das bibliotecas no Bioanalyzer. A partir dos dados da quantidade, tamanho médio dos DNAs presentes nas bibliotecas, estes valores foram utilizados na montagem do *pool* considerando um valor final de 4 Nm (Tabela X). Ao final,

foram transferidos 5 uL de cada biblioteca normalizada para a montagem do pool contendo todas as bibliotecas.

**Tabela 2** - Dados de tamanho médio da biblioteca de cada amostra e quantidade dada por PCR, para cálculo de montagem de pools

<b>Amostra</b>	<b>Tamanho médio da biblioteca (pb)</b>	<b>PCR (nM/uL)</b>	<b>Volume da Biblioteca (em 4Nm) (µL)</b>	<b>Volume do Buffer (µL)</b>
C29	560	49,83	0,80	9,20
C31	316	191,00	2,09	97,91
C20	322	123,40	3,24	9,48
C49	307	127,40	3,14	96,86
C36	318	83,32	4,48	95,52
C12	310	76,67	0,52	9,48
C57	360	107,60	3,72	96,28
C37	303	120,40	3,32	96,68
C50	340	84,86	4,71	95,29
C51	418	40,18	1,00	9,00
C06	479	42,48	0,94	9,06
C83	319	20,40	1,96	8,04
C71	309	57,94	0,69	9,31

#### 4.5.5. Sequenciamento

O sequenciamento foi realizado utilizando a plataforma HiSeq2500 para DNA com sequências *paired-end* no módulo High Output (2 x 100pb) (Catálogo #PE-401-4001 e #FC-401-4003), no Laboratório de Genética da Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo (ESALQ USP).

#### 4.6. Análise de dados

##### 4.6.1. Pré-processamento de dados, análise de qualidade das sequências e remoção de adaptadores



Os dados brutos baixados do BaseSpace foram filtrados removendo as leituras restantes correspondentes aos adaptadores utilizados na montagem das bibliotecas, e leituras com N (bases cujas informações não podem ser determinadas corretamente). Leituras curtas, bases e leituras de baixa qualidade, também foram removidas usando o programa Trimmomatic v.0.36 de acordo com a demanda de qualidade da amostra (LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36) (122). A visualização deste controle de qualidade, antes e após a trimagem, foi realizada utilizando o programa FastQC v.0.11.8 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

#### **4.6.2. Mapeamento e quantificação da expressão gênica**

O pseudo-alinhamento foi realizado utilizando o transcriptoma de referência humano (GRCh38.14), obtido do *Genome Reference Consortium*, encontrado no banco de dados público *ENSEMBL* ([ensembl.org](http://ensembl.org)), usando o programa Kallisto versão 0.46.1 (Near-optimal probabilistic RNA-seq quantification) para realizar a quantificação de transcritos, e produziu uma lista de transcritos compatíveis para cada leitura, evitando o alinhamento da base individual e diminuindo o tempo de processamento. Ao final do processamento, o programa apresenta os dados de quantificação para o nível de expressão de cada transcrito e apresenta-os como transcritos por milhão (TPM).

#### **4.6.3. Avaliação da variabilidade de dados**

A avaliação da variabilidade dos dados foi estudada utilizando o método de PCA, do inglês, *Principal Component Analysis*. Por meio deste método de análise não supervisionada, fica possível visualizar possíveis agrupamentos e amostras que comportam-se como *outliers*, por meio da observação das distâncias relativas entre amostras no plano de componentes principais. Assim, os agrupamentos de amostras podem refletir padrões de expressão semelhantes entre grupos (ou classes) consideradas no estudo. O PCA foi obtido por meio da função `prcomp` no RStudio versão 2021.09.1.

#### **4.6.4. Análise de expressão diferencial via DESeq2**

A partir dos dados quantificados, foi utilizado o pacote do R, DESeq2 versão 1.41.6 (123) para realizar a expressão diferencial de genes entre os grupos considerados no estudo. O

DESeq2 utiliza o teste estatístico de *Wald* para avaliar se existe diferença entre a média de expressão gênica entre dois grupos, ou seja, para verificar se a hipótese nula é verdadeira. Se o valor de  $p$  for pequeno, rejeitamos a hipótese nula e declaramos que há evidências contra o  $H_0$  (ou seja, o gene é expresso diferencialmente). Aqui, fixamos arbitrariamente o limite de significância ( $\alpha = 0,05$ ), onde valores de  $p$  inferiores a 0,05 indicam os genes diferencialmente expressos, significando que existe 5% de chance para um gene expresso de forma não diferencial apresentar este tipo de diferença de expressão. No entanto, falsos positivos podem ser gerados a partir de variâncias de valores próximos de zero. Desta forma, o Deseq2 foi desenvolvido com a intenção de lidar com a alta dimensionalidade de dados reduzindo falsos positivos.

A partir da obtenção dos dados de expressão gênica diferencial, o pacote *biomaRt* foi utilizado para conectar-se aos bancos de dados ENSEMBL, mapear IDs dos transcrição para IDs de genes e obter as análises funcionais a partir das listas de genes. Foram quantificadas e classificadas a abundância de cada biotipo gênico para cada comparação, para fim de tomada de decisão sobre qual classe deveria ser incluída nas análises posteriores. Considerando o objetivo principal do trabalho apresentado anteriormente, à partir desta etapa, foi considerado para a análise subsequentes somente os genes codificadores de proteínas, e descartamos os genes não codificadores, pseudogenes, RNA longos ou com *stop codon* prematuros. Após a exclusão de genes não codificadores de proteínas, dentro da tabela de contagem, existiram múltiplos IDs mapeados para o mesmo identificador de gene (ou seja, para um mesmo identificador do gene, existem possui vários valores de expressão). O banco de dados ENSEMBL forneceu diferentes IDs mapeados em cromossomos distintos: um para o cromossomo principal e outro para regiões haplotípicas, e desta forma, foram filtrados os genes referentes apenas aos cromossomos principais (1:22, X,Y).

Para a definição de gene diferencialmente expresso, foram considerados aqueles com  $p < 0,05$ , e  $\text{Log}_2(\text{FC}) \leq -1$  e  $\text{Log}_2(\text{FC}) \geq 1$ . Assim, para um gene ser considerado *down-regulated*, a sua expressão precisa apresentar-se com valores de expressão inferiores a  $\text{Log}_2(\text{FC}) - 1$ . Por outro lado, para ser considerado *up-regulated*, a sua expressão apresenta-se com valores acima de  $\text{Log}_2(\text{FC}) 1$ . E, para ambos os casos, o valor de  $p$  precisa apresentar-se inferior a 0,05.

#### 4.6.5. Análise de enriquecimento

Para obter informações sobre as mudanças em nível funcional entre pacientes afetados pelo YFV (YFV-17DD e YFV-Silvático), com presença ou não do vírus no tecido cerebral, foram investigados genes que compartilham funções biológicas comuns via um método analítico chamado *gene set enrichment analysis* (GSEA) (124). Para análise dos resultados do enriquecimento obtidas por meio do GSEA, foi utilizado o pacote *Clusterprofiler* versão 3.16.1 por meio do Rstudio versão 2021.9.1.

O resultado inicial da análise de enriquecimento é a pontuação de enriquecimento, ou ES (*enrichment score*). O método GSEA inicialmente divide os genes em subconjuntos de genes superexpressos ou *up-regulated*, e ranqueia os genes que estão super representados (*overrepresentation* - conjunto de genes que compartilham mais ou menos genes com vias observadas no banco de dados de referência) e tendem a aparecer na parte superior da lista, ou seja, mais superexpressos, ou na parte inferior (mais *down-regulated*). O grau de enriquecimento foi então estabelecido pela pontuação de enriquecimento normalizada (*normalized enrichment score*, NES). Assim, um ES positivo indica que o conjunto de genes está no topo da lista, enquanto um ES negativo indica que o conjunto de genes está no final da lista. O ES normalizado, NES, é utilizado para analisar os resultados de enriquecimento. GSEA, determina o NES pela fórmula abaixo:

$$\text{NES} = \text{ES atual} / \text{média ES (todos ES do data)}$$

Por utilizar a média, as pontuações de enriquecimento normalizado (NES), podem ser utilizadas para comparar as análises entre conjunto de genes.

#### 4.6.5.1. Redes gênicas e mapas de enriquecimento

Para compor as redes gênicas e mapas de enriquecimento, foram utilizadas métricas de similaridade de *Jaccard* e coeficiente de sobreposição. Nas redes gênicas, o grau de sobreposição é visualizado pela distância entre os nós e definido pelo coeficiente de sobreposição de *Szymkiewicz-Simpson*, calculado pelo valor do tamanho da intersecção do conjunto A e conjunto B, sobre o tamanho do menor conjunto entre A e B, onde:

$$oc(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}$$

onde, *oc* representa *Overlap Coefficient*, e  $A = a$ ,  $B = b$ .

Para os mapas de enriquecimento, além das distâncias calculadas pelo coeficiente de sobreposição de *Szymkiewicz-Simpson*, a espessura da borda é proporcional ao grau de sobreposição entre os conjuntos de genes, calculada pelo coeficiente de *Jaccard*, valor dado pelo tamanho da intersecção do conjunto A e do conjunto B (número de elementos comuns), sobre o tamanho da união do conjunto A e do conjunto B (número de elementos únicos), dados pela fórmula abaixo:

$$js(A, B) = \frac{|A \cap B|}{|A \cup B|},$$

onde, *js* representa *Jaccard Similarity*, e  $A = a$ ,  $B = b$ .

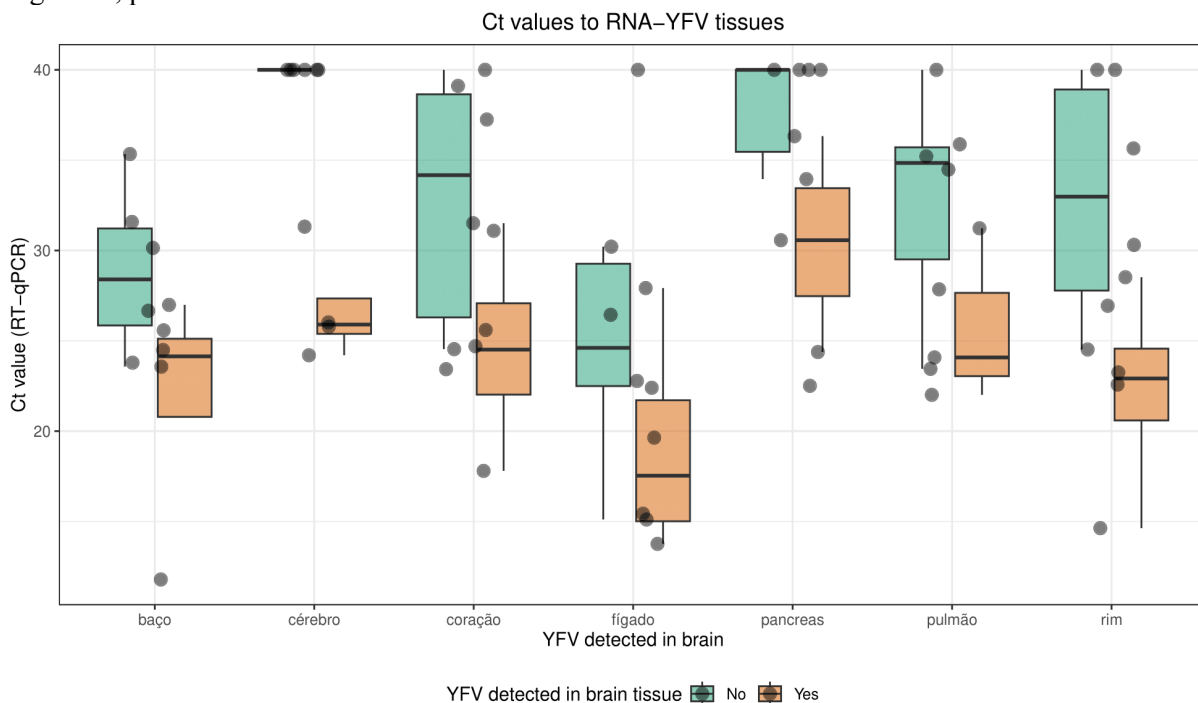
## *Capítulo 5*

## 5. RESULTADOS

### 5.1 Caracterização molecular

A caracterização molecular por RT-qPCR para detecção do RNA viral (RNA<sub>v</sub>) foi realizada à partir de sete tecidos (coração, pulmão, cérebro, rim, baço, pâncreas e fígado) coletados de 85 pacientes (n = 85) que tiveram óbito com suspeita de associação ao YFV. Setenta e sete (89,53%) pacientes foram positivos para febre amarela (detecção do vírus em pelo menos um órgão). Em geral, pacientes que tiveram o YFV detectado no cérebro, tiveram uma menor amplitude nos valores Cts em outros órgãos, porém, com esses valores menores. Os dados sugerem que a infecção do SNC pelo YFV pode estar condicionada a uma carga viral nos outros órgãos, sendo isso, a provável explicação para explicar o processo de passagem pelas barreiras de acesso ao SNC (Figura 7).

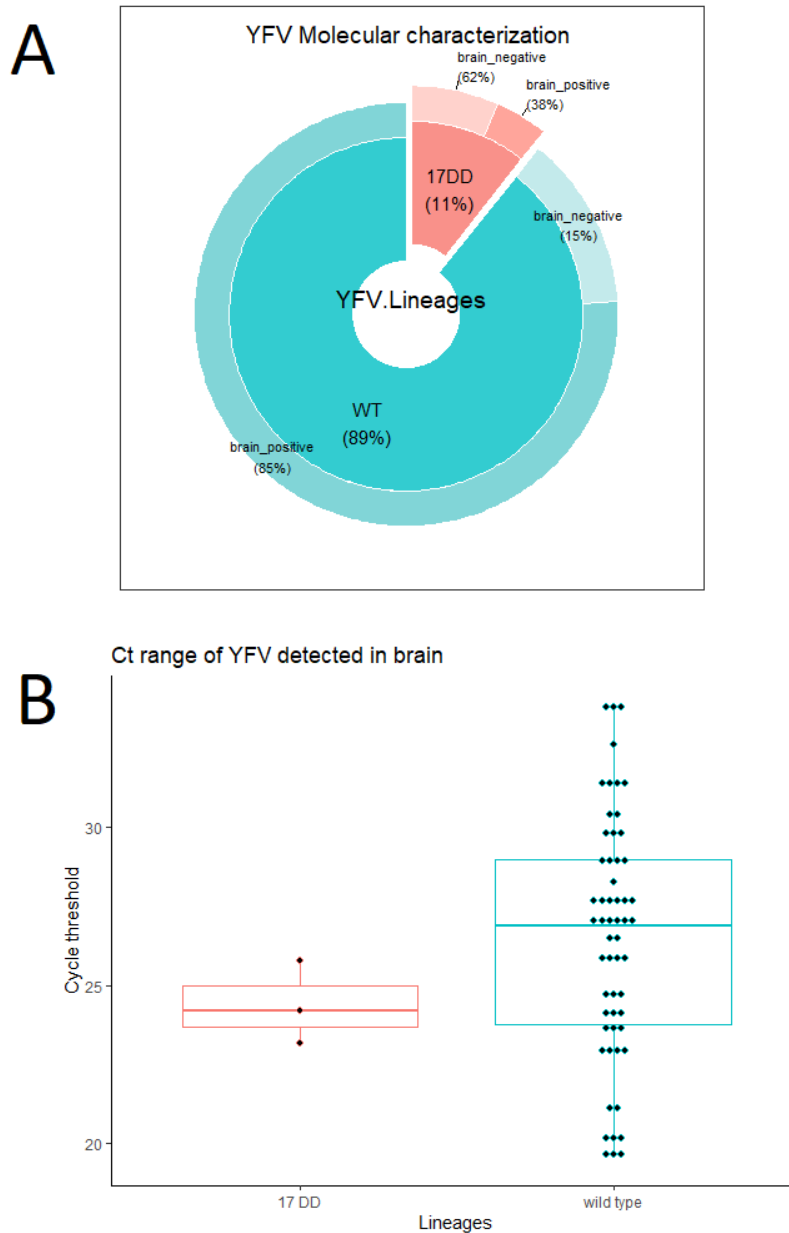
**Figura7** - Caracterização dos valores de Cts identificados considerando os cérebros positivos e negativos, para os diferentes tecidos



Valores de Cts identificados em diferentes tecidos de 77 pacientes que tiveram óbito com suspeita de associação ao YFV e detecção do vírus em pelo menos um órgão. Foi considerada a detecção ou não do vírus FA no tecido cerebral para a comparação entre tecidos. Boxplots representam o 75º percentil, mediana, 25º percentil e os bigodes se estendem ao valor mais alto e mais baixo considerando, dentro do intervalo 1,5 x intervalo interquartil (IQ).

Dos 77 pacientes, 69 foram positivos para o vírus silvático (89,61%), com 85 % de positividade no tecido cerebral. Os outros 8 pacientes foram positivos para o vírus vacinal (10,38%), com 38% de positividade no tecido cerebral (Figura 8A). Dos pacientes com vírus detectado no cérebro, 57 foram positivos para detecção do vírus silvático (95%), e 3 o vírus vacinal (5%) e a variação de Ct maior na linhagem silvática (Figura 8A-B).

**Figura 8** - Caracterização molecular das amostras estudadas



(A) Descrição do padrão de positividade, considerando os pacientes e o tecido alvo do estudo. (B) Valores de Cts identificados no cérebro, considerando os diferentes tipos de vírus analisados (YFV-17DD e YFV-Wild-type). Boxplots representam o 75º percentil, mediana, 25º percentil e os bigodes se estendem ao valor mais alto e mais baixo considerando, dentro do intervalo 1,5 x intervalo interquartil (IQ).

## 5.2. Análise de qualidade (pré-sequenciamento)

### 5.2.1. Análise da integridade do RNA total

Os resultados obtidos a partir do bioanalyzer foram utilizados para descrever o padrão de qualidade da amostra estudada. O perfil obtido à partir dos RNAs totais foram determinados, e utilizados para estipular o tempo de fragmentação da amostra durante o preparo da biblioteca, facilidade permitida devido ao kit escolhido para esta etapa do desenvolvimento do estudo. Devido à sua natureza molecular, o RNA é mais suscetível à degradação do que o DNA, assim, sabemos que a qualidade das moléculas de RNA extraídas podem ter impacto nos resultados do experimento RNA-seq. O número que representa a integridade do RNA, calculado à partir das intensidades correspondentes aos RNAs ribossomais, e aqui chamado de RIN (do inglês, *RNA Integrity Number*) é baseado em um intervalo de numeração de 1 a 10, sendo 1 atribuído ao RNA mais degradado e 10 é correspondente ao mais intacto. Um RIN de 6 é normalmente considerado por alguns autores como um "threshold" no cérebro humano pós-morte, embora esses números não sejam consistentes na literatura e nos repositórios cerebrais (REF) (*Limited predictability of postmortem human brain tissue quality by RNA integrity numbers* Kai-C. Sonntag, George Tejada, Sivan Subburaju, Sabina Berretta, Francine M. Benes, and Tsung-Ung W. Woo). Observamos que, a correlação entre o tempo *post-mortem* e o valor de RIN correspondente às amostras utilizadas no experimento de RNAseq foi considerada positiva, porém fraca, sugerindo que o tempo *post-mortem* não é determinante para preservar a integridade do RNA total (Figura 9).



**Figura 9** - Comparação entre os tempos post-mortem e valores de RIN

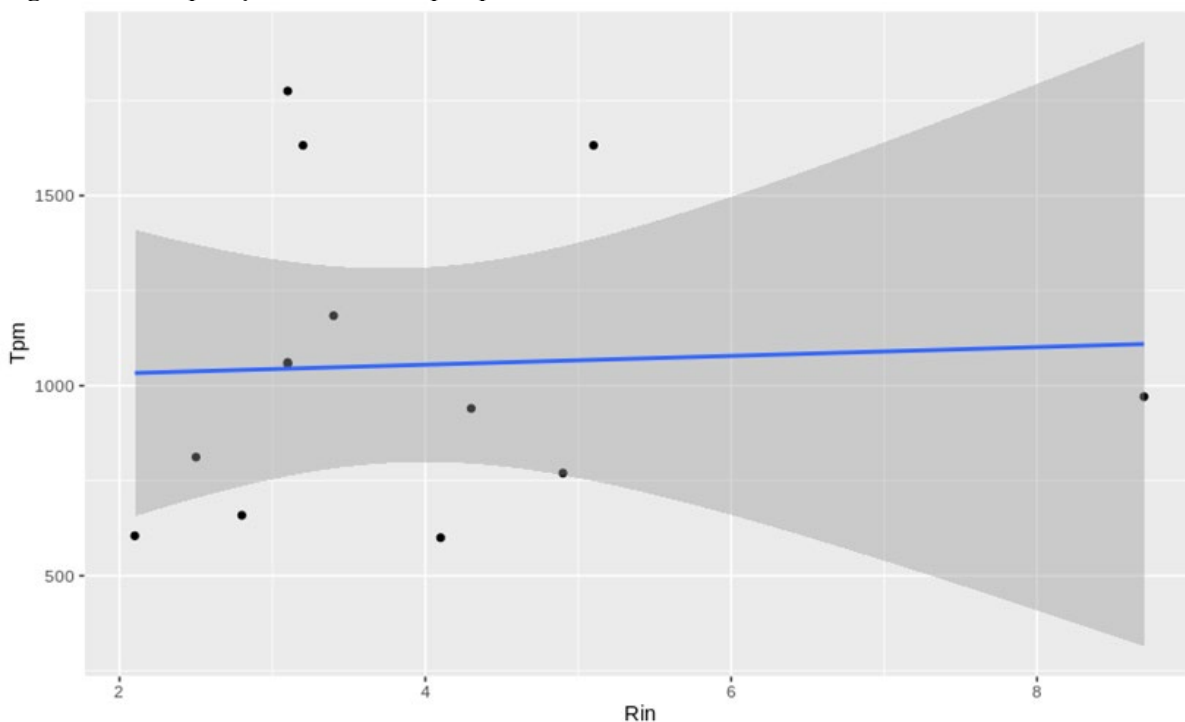


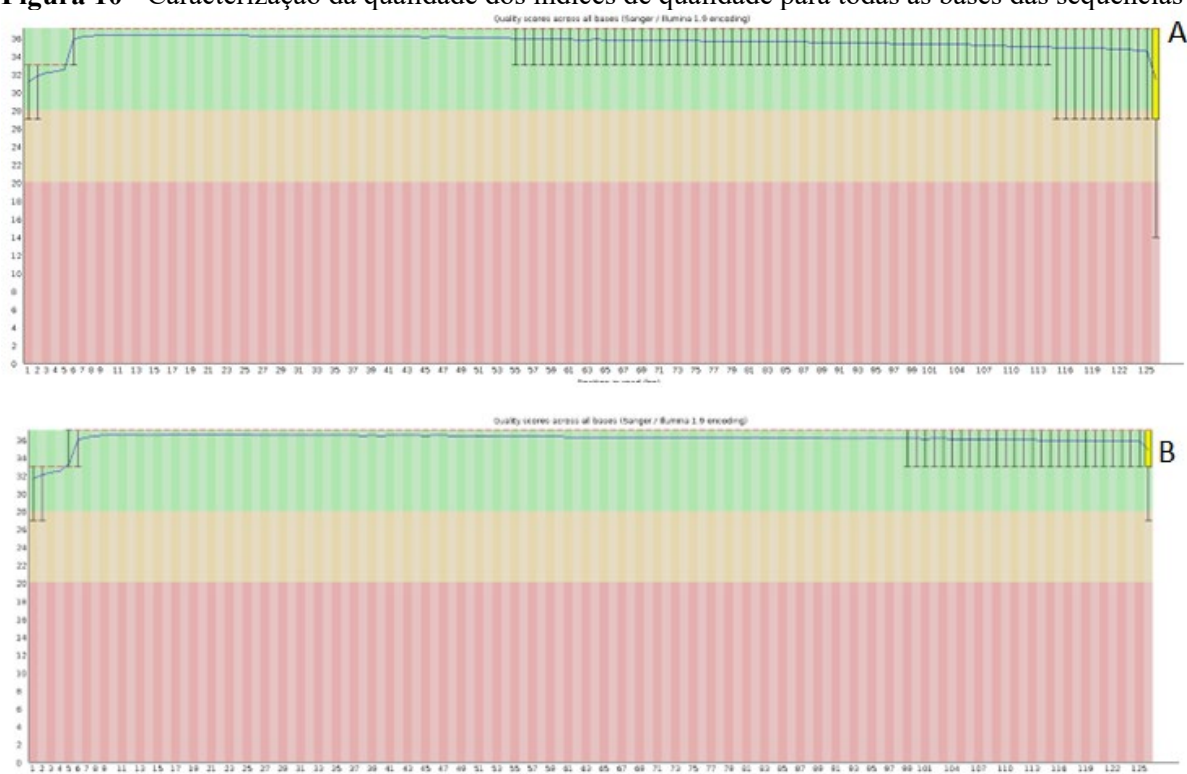
Gráfico mostrando tempo pós morte (TPM) *versus* RIN, mostrando que não houve correlação entre as variáveis. Os resultados indicam como é a associação entre o tempo pós morte (TPM) *versus* RIN de cada amostra, sugerindo que não existe correlação entre TPM e RIN nas amostras coletadas para o estudo.

### 5.3. Análise de qualidade pós sequenciamento

#### 5.3.1. Trimagem

Os erros de sequenciamento mais comuns, estão relacionados à baixa qualidade, e a contaminação pelos adaptadores. Assim, após o sequenciamento das amostras faz-se necessário realizar o controle de qualidade dos *reads*. Índices de qualidade Phred (Q), é uma escala logarítmica relacionada com as probabilidades de erro de chamada de base (P). Após algumas testagens, foi decidido que o melhor parâmetro seria o uso de Phred 30 com LEADING:20 TRAILING:20 MINLEN:36. Como é possível observar no gráfico (a) os *reads* já possuíam qualidade superior ao esperado. Sendo assim, os números de *reads* observados antes e depois da trimagem não obtiveram diferença significativa (Figura 10A-B). Com essa informação foi possível decidir pular a etapa de trimagem, uma vez que é um passo facultativo de acordo com a demanda de cada amostra.

**Figura 10** - Caracterização da qualidade dos índices de qualidade para todas as bases das sequências



(A) Gráfico com indicador de qualidade dos *reads* antes da trimagem (B) Gráfico com indicador de qualidade dos *reads* depois da trimagem. O eixo y mostra as pontuações de qualidade. Quanto maior a pontuação, melhor qualidade de base. O plano de fundo do gráfico divide o eixo y em bases de muito boa qualidade (verde), qualidade razoável (laranja) e qualidade ruim (vermelho).

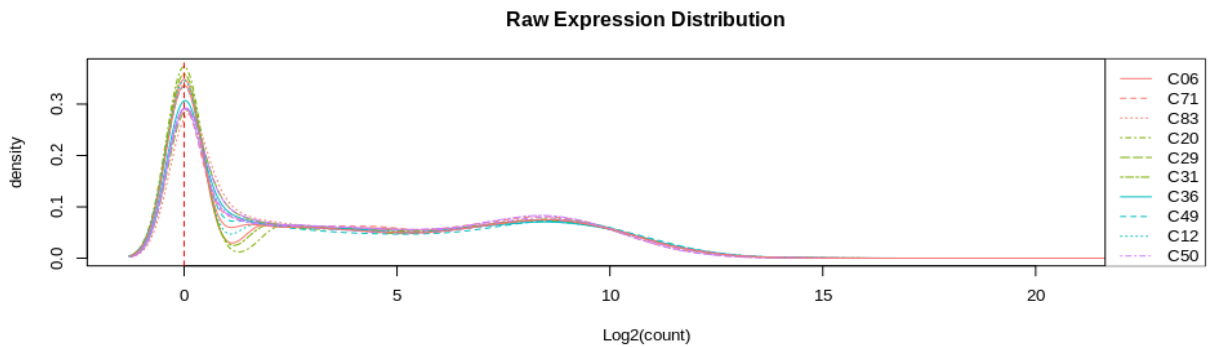
## 5.4. Análise de dados RNAseq

### 5.4.1. Pré processamento

#### 5.4.1.1. Filtragem de genes pouco expressos

A quantificação do nível de expressão de cada transcrito foi realizada pelo método de pseudo-alinhamento e representados em transcritos por milhão (TPM). Também, foi feita a filtragem de genes pouco expressos, os quais iriam acarretar em ruído na análise. Inicialmente, devido ao grande número de transcritos com expressão igual a zero (Figura 11), foram excluídas todas as linhas cuja soma das colunas era igual a zero (onde cada coluna corresponde a uma amostra). Assim, antes da filtragem haviam 39339 genes no total, após a filtragem a contagem foi de 34030.

**Figura 11** - Gráfico de densidade com contagem bruta de expressão por amostra

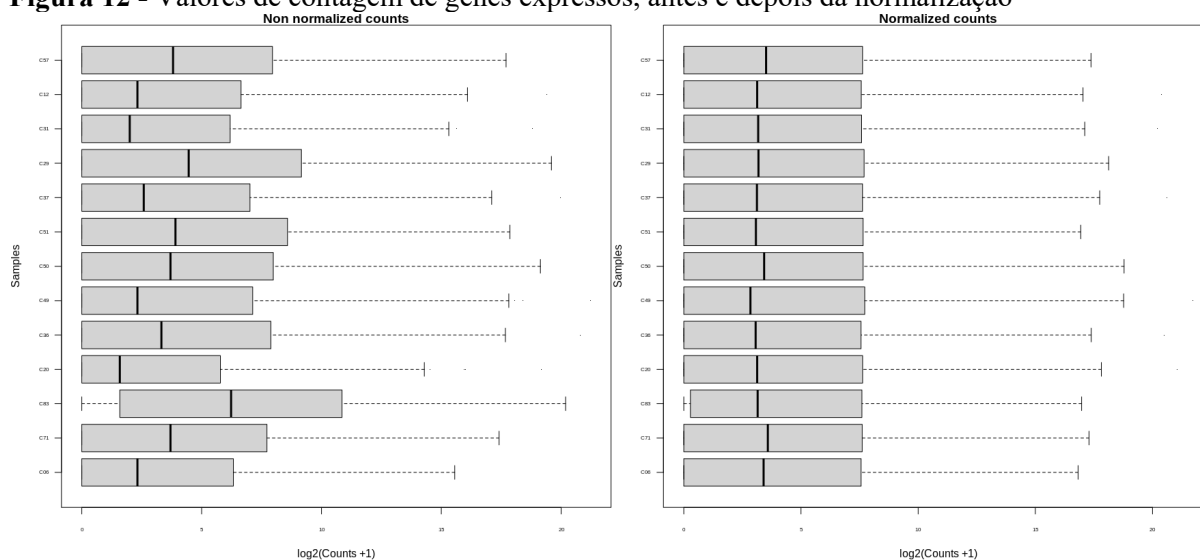


A contagem no eixo y foi dada em  $\log_2(\text{valor})$ .

#### 6.4.4.2. Normalização para viés de composição

Na etapa de normalização foram considerados fatores que interferem no número de leituras mapeadas para um gene sendo eles: (i) profundidade da sequência, (ii) comprimento do gene, e (iii) também composição da população de RNAm que está sendo amostrada. A contagem da profundidade de sequenciamento foi necessária para a comparação da expressão gênica entre as amostras, já que uma amostra pode ter maior expressão de um gene devido a sua maior profundidade (quantidade de *reads* totais obtidas) e não por uma expressão maior em si. Levando em consideração o fator “comprimento”, um determinado gene pode produzir uma maior quantidade de *reads* quando realizado o sequenciamento pelo fato de ser mais longo, ou então, uma menor expressão pode ser obtida por ser mais curto, fato que justifica a necessidade de uma normalização. Ainda, alguns genes altamente expressos podem ter uma fração muito elevada na proporção de leituras em uma amostra e, portanto, as contagens para todos os outros genes são reprimidas. Como resultado, em comparação com uma amostra em que as leituras são distribuídas de maneira mais uniforme, esses genes reprimidos parecem ter uma expressão mais baixa, o que pode dar origem a muitos genes diferencialmente expressos falsos positivos. Para normalizar a profundidade de sequenciamento, comprimento de sequência e a composição do RNA, o DESeq2 usa o método da mediana das proporções, em que as contagens são divididas por fatores de tamanho específicos da amostra determinados pela proporção da mediana de contagens de genes em relação à média geométrica por gene. Valores de contagem de genes expressos, antes e depois da normalização foram mostrados na Figura 12.

**Figura 12 - Valores de contagem de genes expressos, antes e depois da normalização**

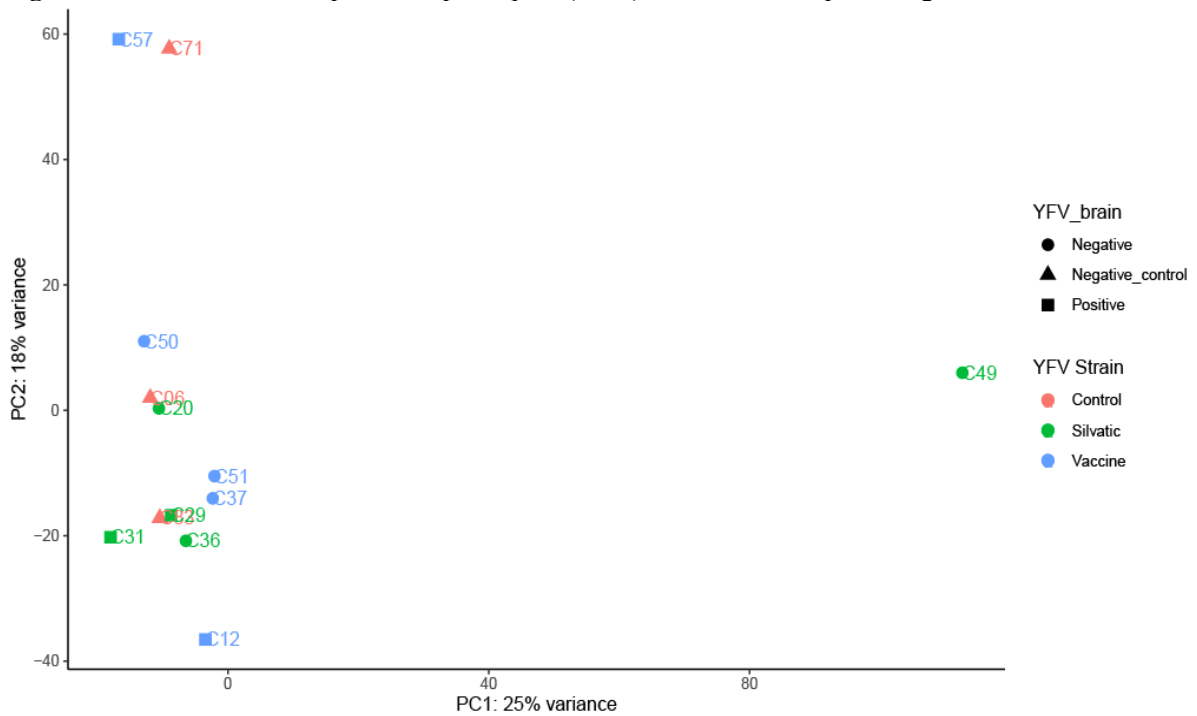


Boxplot de contagem de *reads* por amostra antes (A) da etapa de normalização e após a normalização (B). Boxplots representam o 75º percentil, mediana, 25º percentil e os bigodes se estendem ao valor mais alto e mais baixo considerando, dentro do intervalo 1,5 x intervalo interquartil (IQ).

#### 5.4.4.4. Análise de componentes principais

Os dados de expressão gênica obtidos, filtrados e normalizados foram utilizados para uma análise exploratória inicial, em que dados de alta dimensionalidade foram transformados pelo método de análise de componentes principais), de modo que o primeiro componente principal (PC1) representa a maior fonte de variação entre os dados, o segundo PC (PC2) é aquele que possui segunda maior fonte de variação, e assim sucessivamente. Utilizando os dois primeiros PCs, observamos a ausência de *clusters* bem definidos. Porém, observamos que a variância dos dois primeiros PCs apresentam frequência relativa baixa (PC1 apresenta 25% de variância e PC2 18%), e portanto, existe muita informação não representada por esses dois primeiros PCs (Figura 13).

**Figura 13** - Análise de componentes principais (PCA) do dados de expressão gênica das amostras



Análise de componentes principais (PCA) do dados de expressão gênica das amostras, separadas, organizadas em função das classes consideradas no estudo, que foram caracterizadas nos ensaios moleculares. As classes consideradas são: YFV-silvático, YFV-17DD e controles, onde as cores distintas representam as classes distintas. Os diferentes símbolos observados tratam-se da caracterização molecular considerando o tecido cerebral por meio da detecção do YFV.

Para representar melhor a variância dos dados, exploramos os valores da variância total explicada por cada PC. No total, foram necessários 13 PCs para mostrar o total de variação acumulada nos dados. O método de elbow é útil para determinar quantos PCs são necessários para capturar a maior parte da variação nos dados. O valor de elbow foi determinado no PC3, indicando que, portanto, faz-se necessário no mínimo 3 PCs para determinar a maioria da variação dos dados. Para explicar 95% dos dados seria necessário no mínimo 10 PCs (Figura 14).

**Figura 14** - Gráfico de pareto

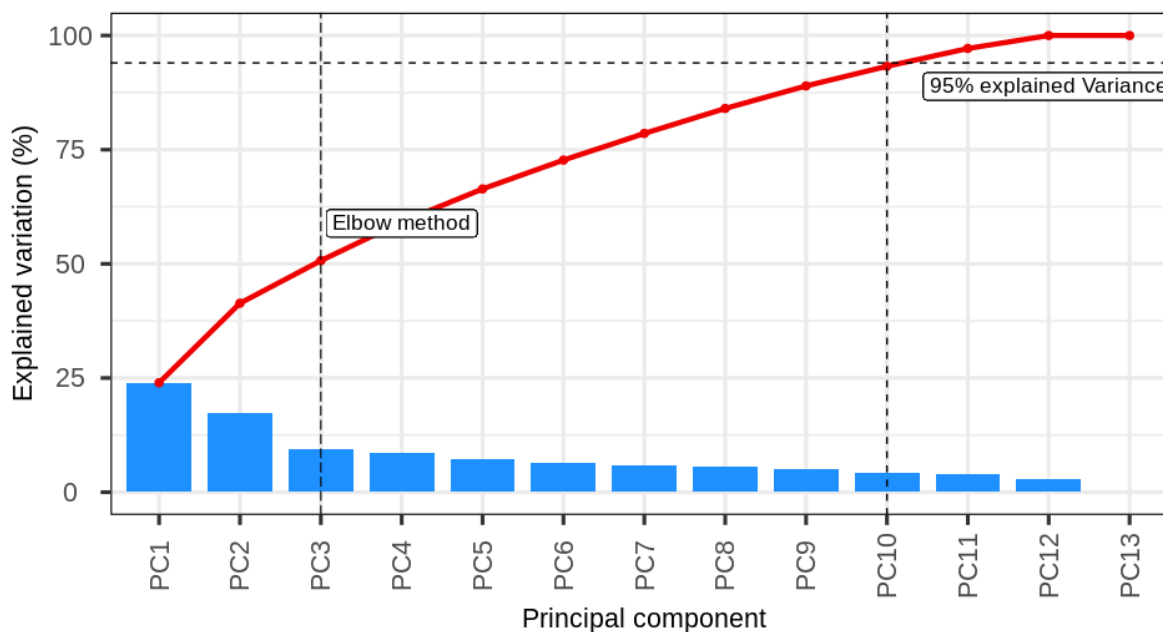
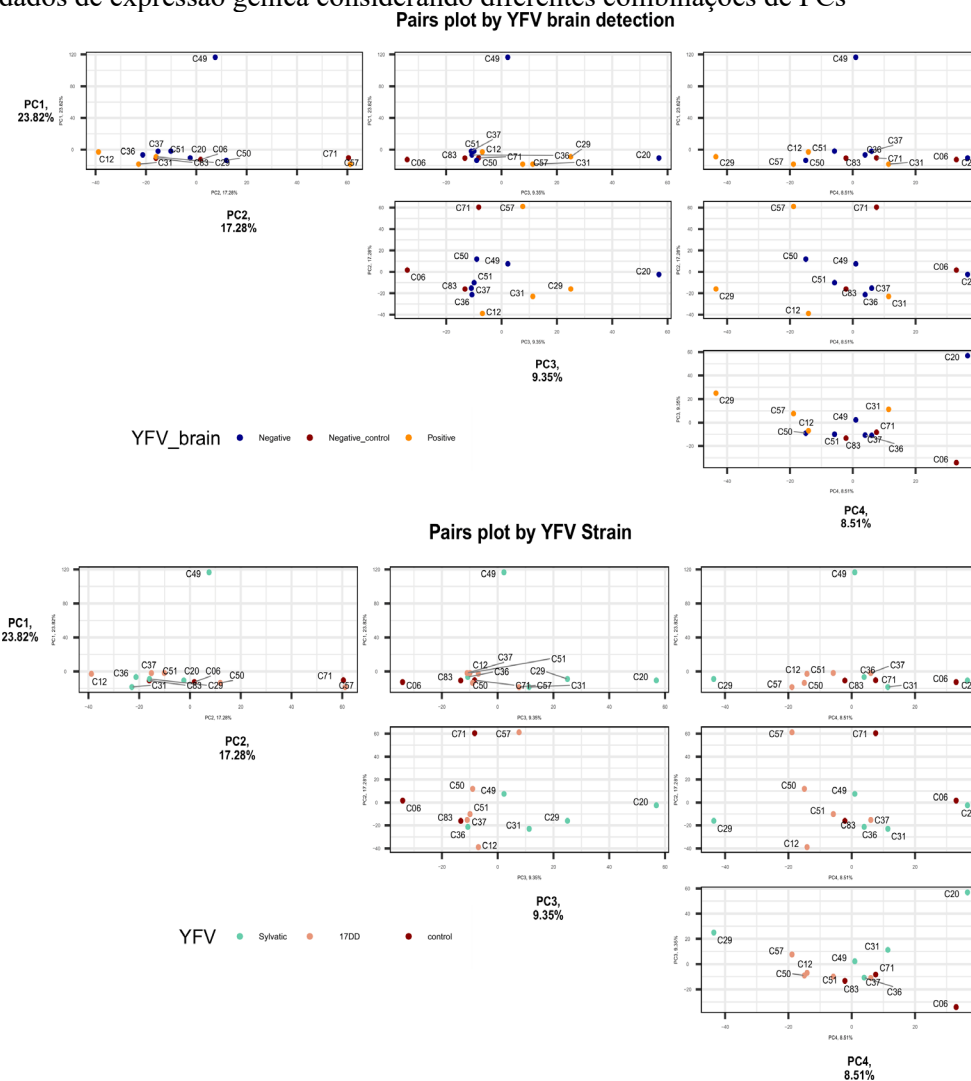


Gráfico de pareto, mostrando a redução variável do PCA no conjunto de dados no qual a variância explicada é expressa como uma função do número de componentes principais. O cut-off no PC3 foi traçado pelo método de elbow, enquanto o cut-off no PC10 foi traçado mostrando a explicação de 95% dos dados.

Com a finalidade de observar melhor os padrões existentes considerando outros PCs não abordados em análises tradicionais envolvendo PCA, consideramos adicionar outras combinações de PCs. Assim, esta combinação mostrou a existência de grupos organizados em função das classes, tanto quando observadas a presença ou ausência do vírus no cérebro (Figura 15A), quando considerada linhagem do YFV (Figura 15B), o que indica a existência de padrões organizados de maneira complexa, natureza existente em dados biológicos envolvendo por exemplo, análises da expressão gênica partindo de RNAseq (Figura 15A-B), mostrando-se efetivo quando considerado PC2 e PC3.

**Figura 15 - Análise exploratória dos dados de expressão gênica considerando diferentes combinações de PCs**



Observa-se que no gráfico considerando PC2 (17.28%) e PC3 (9.35%) existe uma pequena formação de *clusters* de amostras referentes às classes consideradas no estudo. As classes consideradas são: YFV-silvático (verde), YFV-17DD (salmão), YFV detectado no cérebro (amarelo) e YFV não detectado no cérebro (azul), além dos controles (vermelho), onde as cores distintas representam as classes distintas..

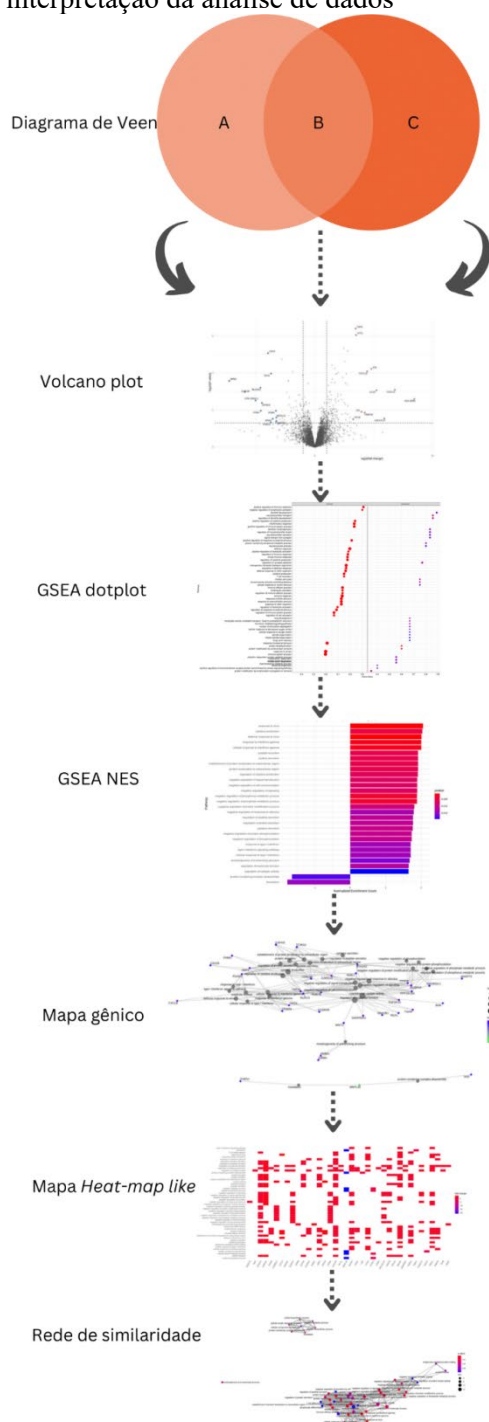
#### 5.4.2. Análises de expressão gênica diferencial e ontológica

Para a análise de expressão gênica diferencial, foram anotados 34030 genes, com seus respectivos *p-values*, *padj*, *log2FoldChange*, *BaseMean* (média das contagens normalizadas de todas as amostras, normalizando para profundidade de sequenciamento) e *gene ids*. Os resultados de ontologia foram analisados de maneira gráfica e utilizados para indicar informações relacionadas à função de cada gene, ou um conjunto deles, dentro de uma via ontológica, ou termo ontológico. Considerando-se a premissa de que se muitos genes associados a um determinado processo biológico são diferencialmente expressos em uma determinada condição, diferentes maneiras de visualização podem ser propostas para determinar quais processos biológicos distintos são importantes para determinar as alterações fenotípicas observadas nas condições estudadas.

Para melhor entendimento da visualização dos dados e consequente interpretação, sugiro o acompanhamento seguindo o fluxograma abaixo, nas análises mostradas à seguir (Figura 16).



**Figura 16** - Fluxograma para interpretação da análise de dados



A partir dos genes selecionados exclusivamente para cada condição (A e B) e genes presentes na intersecção entre grupos A e B são plotadas as seguintes etapas analíticas: (i) volcano plots para visualização de genes *up-regulated* e *down-regulated*; (ii) GSEA dotplots com as principais vias enriquecidas *up-regulated* e *down-regulate*; (iii) Gráfico de barras com GSEA *normalized enrichment score*, mostrando as vias que obtiveram maior e menor NES segundo o GSEA; (iv) Mapa gênico formado por meio de cálculo de coeficiente de similaridade, com 20 principais vias e seus respectivos gene; (v) Mapa *Heat-map like*, mostrando tendências de centralidade de vias e genes e; (vi) Rede de similaridade com principais vias enriquecidas formada por meio de cálculo de coeficiente de similaridade e coeficiente de *Jaccard*.

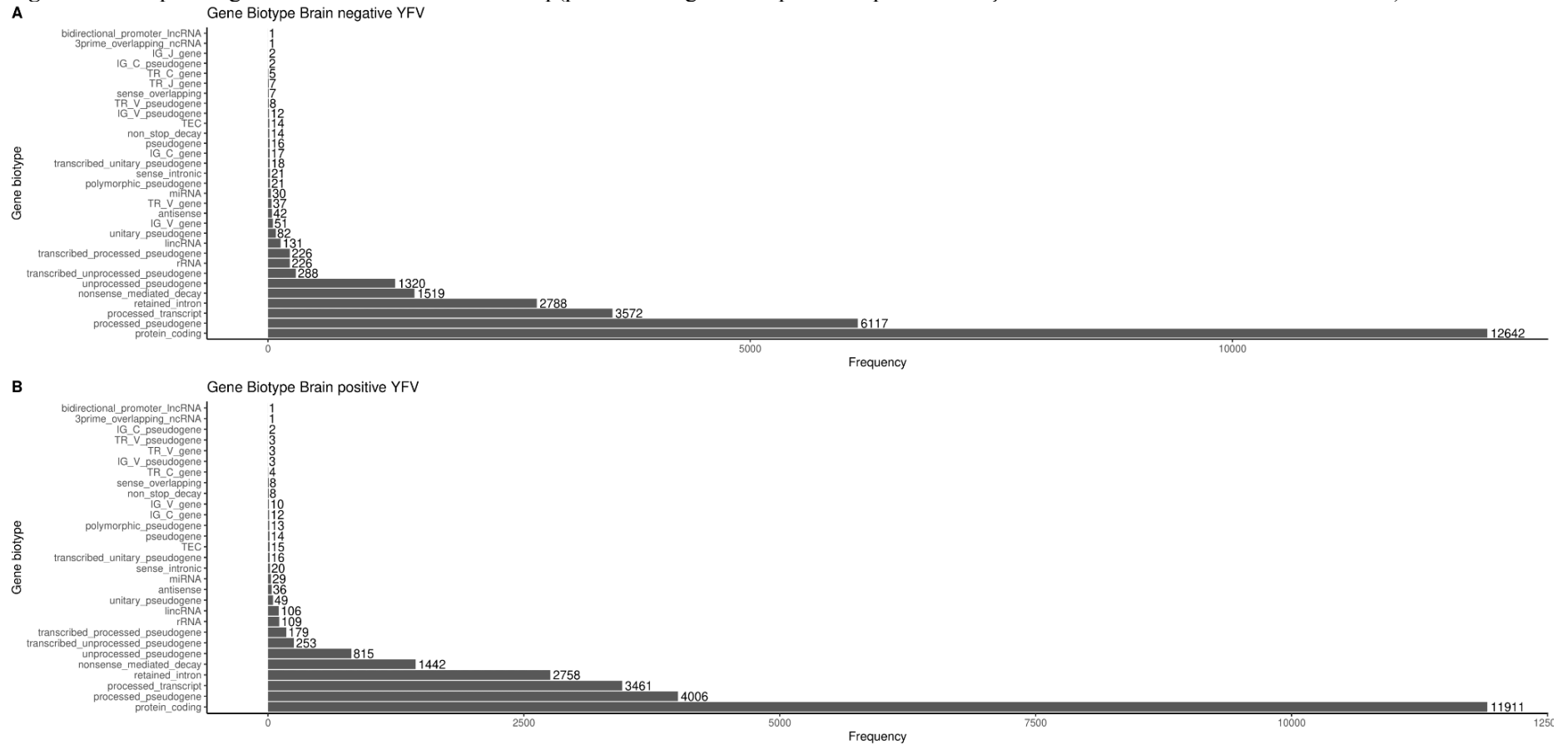
#### **5.4.4.2.1. Expressão diferencial entre grupo de pacientes com vírus da febre amarela detectado no cérebro e grupo em que o vírus não foi detectado no SNC**

A primeira análise considerada, refere-se à comparação envolvendo os pacientes com vírus da febre amarela detectado no cérebro e o grupo em que o vírus não foi detectado, sem nenhuma ponderação considerando a linhagem viral. A hipótese nula aqui considerada, é que não existe diferença entre as duas classes, uma vez que todos os pacientes tratam-se de casos fatais associados ao YFV. A hipótese alternativa a ser testada refere-se à observação de diferença no perfil de expressão gênica entre as duas classes (pacientes que tiveram o vírus da febre amarela detectado no cérebro e aqueles em que não houve detecção do vírus no mesmo tecido). Os grupos 1 U 3 (classe com a detecção no cérebro) e os grupos 2 U 4 (classe onde o YFV não foi detectado no cérebro) tiveram os valores de expressão gênica comparados ao grupo controle (grupo 5), aqui considerados como um grupo de pacientes contendo expressão basal em condições fisiológicas, e assim, obtivemos os valores de expressão gênica diferencial.

##### **5.4.2.1.1. Expressão gênica diferencial**

A análise da expressão gênica diferencial considerando a condição experimental comparada ao grupo controle indicou que a maioria dos genes diferencialmente expressos (do inglês, *differentially expressed genes* - DEGs) compreendem genes codificadores de proteína. Para o grupo de pacientes com ausência da detecção do YFV no cérebro (Condição A), foram identificados 12642 genes codificadores de proteínas expressos no tecido cerebral, e destes, 501 genes foram considerados como diferencialmente expressos ( $p < 0,05$ ) quando comparados ao grupo controle. No grupo de pacientes com detecção do YFV no cérebro (Condição B), foram caracterizados 11911 genes como sendo expressos, e destes, 402 genes ( $p < 0,05$ ) foram considerados como diferencialmente expressos (Figura 17).

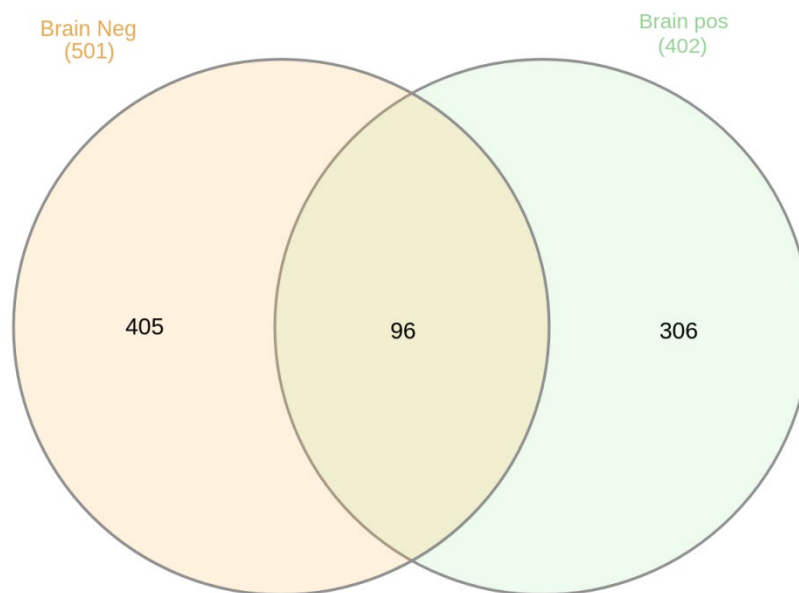
**Figura 17-** Grupos de genes caracterizados no RNAseq (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral)



Caracterização do perfil dos genes expressos considerando os grupos de pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral. (A) Grupo de pacientes sem detecção do YFV no cérebro comparado ao grupo controle (Condição A). (B) Grupo de pacientes com detecção do YFV no cérebro comparado ao controle (Condição B). O eixo x representa a frequência absoluta de genes expressos, e o eixo y representa a classificação dos genes expressos atribuídos aos diferentes biotipos gênicos.

As DEGs considerando o grupo de pacientes positivos e negativos para a detecção do YFV no tecido cerebral permitiu identificar os genes exclusivamente relacionados a cada condição. Assim, para a condição A, 405 genes são exclusivos desta condição, enquanto para a condição B, 306 genes foram considerados exclusivos. Noventa e seis DEGs são caracterizados como pertencentes a ambas as condições, representados como a intersecção das duas condições (Figura 18).

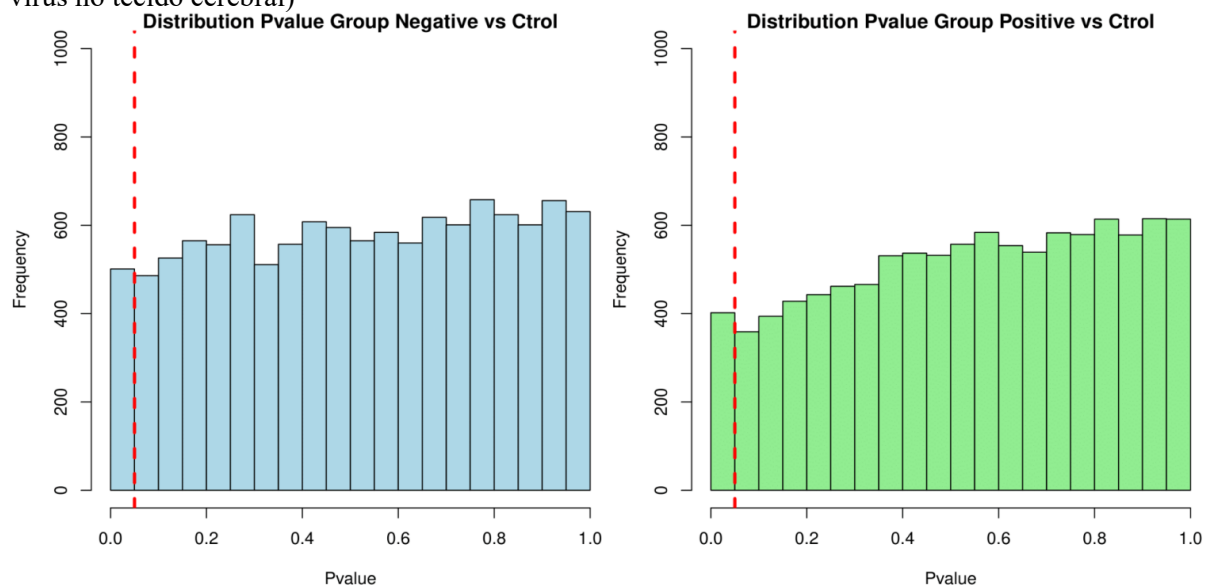
**Figura 18** – Comparação dos genes diferencialmente expressos (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral)



Comparação dos genes diferencialmente expressos considerando os grupos de pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral. O Diagrama de *Venn* representa genes exclusivos de cada classe, e também, aqueles compartilhados entre classes. (A) Em laranja, condição A - pacientes sem detecção do YFV no cérebro comparados ao controle. (B), em verde, condição B - pacientes com detecção do YFV no cérebro comparados ao controle. A intersecção de classes indica os DEGs comuns encontrados em ambas as classes consideradas nesta análise mostrando genes interclasses.

Histogramas da distribuição de valores de  $p$  mostram o padrão de expressão de cada grupo estudado (Figura 19). Os dois grupos possuem um padrão de distribuição similar, com genes diferencialmente expressos ( $p < 0,05$ ) ocupando uma pequena parcela do total de genes naquela condição, ou seja uma pequena porcentagem de hipóteses não é nula.

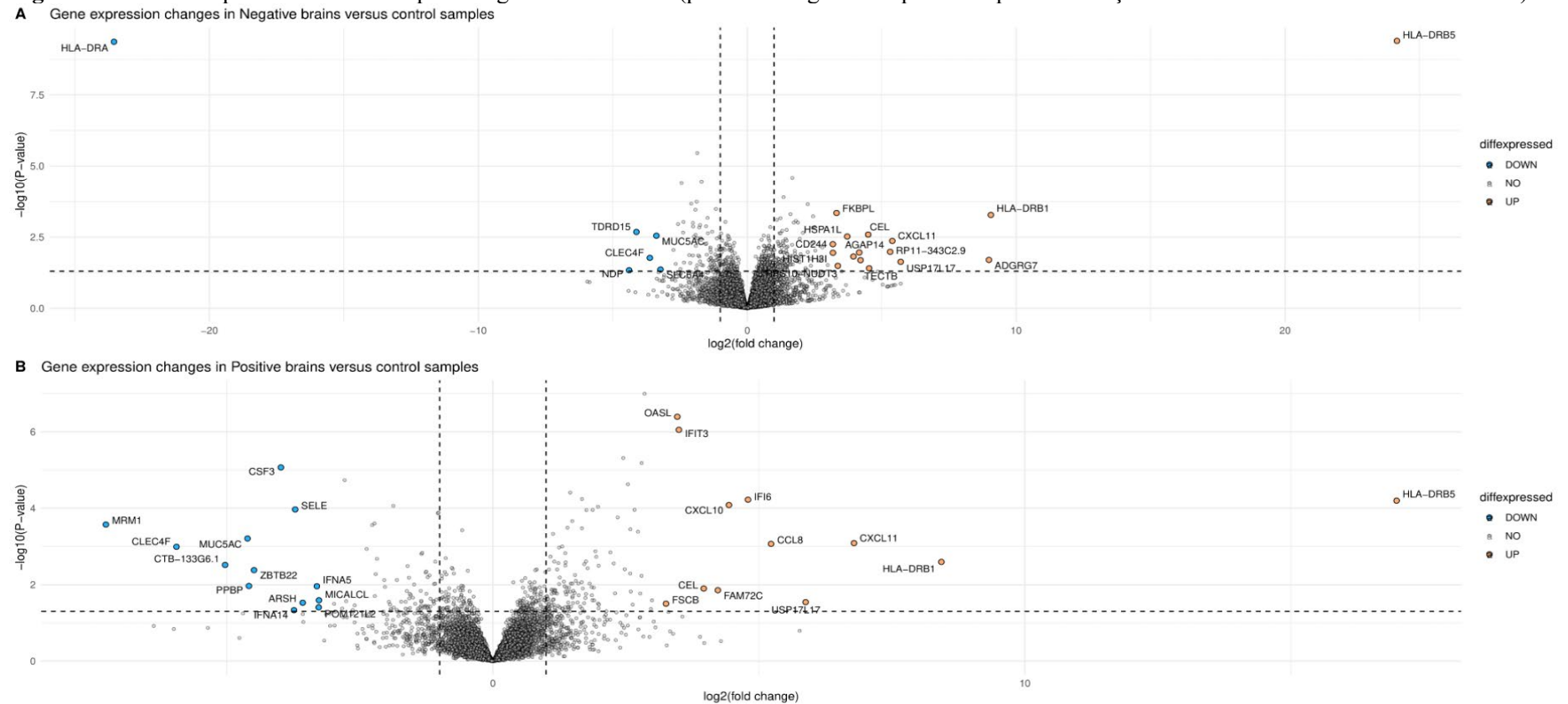
**Figura 19** - Distribuição de valores de p (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral)



(A) Considerando os pacientes em que não houve detecção do YFV no cérebro, e grupo (B) em que houve detecção do YFV no cérebro. A linha pontilhada vermelha indica o threshold dos genes considerados diferencialmente expressos com valor de  $p < 0,05$ .

Com o objetivo de combinar o resultado da análise estatística para determinação dos genes diferencialmente expressos com as informações quantitativas obtidas para o nível de expressão e a magnitude da diferença em cada gene, utilizamos o volcano plot para mostrar visualmente os DEGs que destacam-se, tanto do lado dos *up-regulated* quanto dos *down-regulated* (Figura 20). Assim, foram destacados os genes com os maiores valores de FoldChange (ou diferença de expressão), que podem fornecer indícios dos mecanismos envolvidos na acessibilidade do YFV no tecido cerebral. Na condição A, o gene *HLA-DRA* ganhou destaque por ser 20 vezes ( $\log_2$ ) menos expresso nesta condição do que o grupo controle. Já o gene *HLA-DRB5* foi destaque com FC maior que 20 ( $\log_2$ ) e foi comum em ambas condições.

**Figura 20 - Volcano plot mostrando a expressão gênica diferencial (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral)**

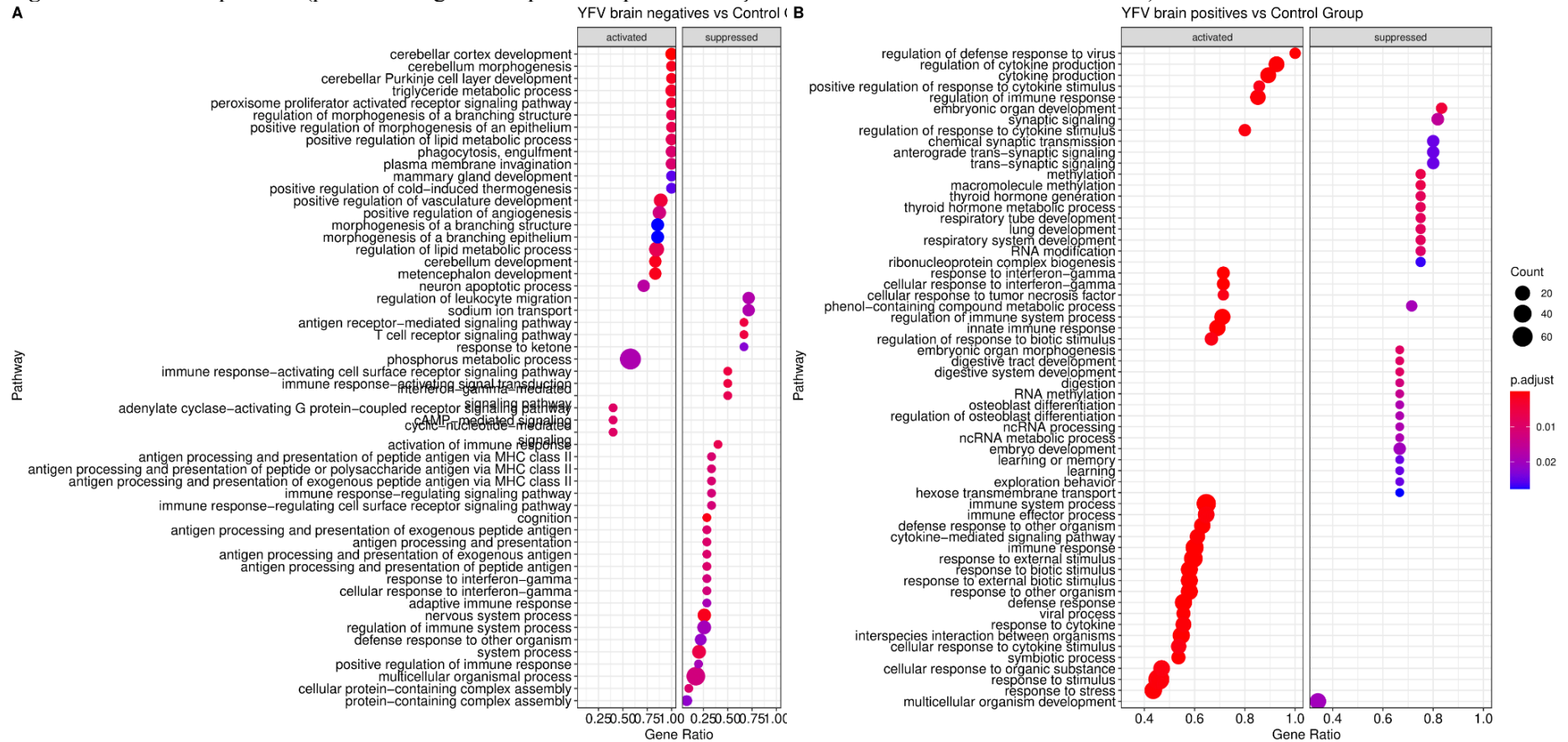


Expressão gênica diferencial considerando as diferentes condições. (A) Volcano plot representando os genes expressos na condição A - pacientes sem detecção de YFV no cérebro em comparação ao grupo controle. (B) Volcano plot representando os genes expressos na condição B - pacientes com detecção de YFV no cérebro em comparação ao grupo controle. O eixo y representa o valor negativo de log10 do p-value para cada gene [ $-\log_{10}(p \text{ values})$ ], e traçado um *threshold* onde o valor de p é 0,05, mostrando valores significativos acima do tracejado. No eixo x, foi representado o log2 do valor de *fold-change* (FC) para cada gene, indicando quais grupo de genes estão *up-regulated* ou *down-regulated* em comparação com o grupo controle. O FC é calculado pelo valor de expressão do gene na condição B / valor da expressão do gene na condição A. O *threshold* de 1 e -1 Log2FC foi estabelecido. Os genes representados em azul são os que satisfizeram a condição  $p < 0,05$  &  $FC \leq -3$  (*down-regulated*). Os genes que satisfizeram a condição  $p < 0,05$  &  $FC \geq 3$ , estão *up-regulated* e são representados em laranja.

#### 5.4.2.1.2. Análise de ontologia gênica

A partir das DEGs obtidas considerando as diferentes condições/classes, utilizamos-as para recuperar e inferir as principais vias (do inglês, *pathways*) de processos biológicos que possuem maior quantidade de genes associados a elas por meio da análise de ontologia gênica. Na condição A, onde os pacientes não tiveram o YFV detectado no ensaio molecular no tecido cerebral, alguns *pathways* ganham destaque, como a ativação da via de metabolismo de fósforo, onde há uma grande quantidade de genes envolvidos. Por outro lado, algumas das vias estão envolvidas com o sistema imunológico, como aquelas relacionadas a apresentação de antígeno via classe de MHC II, regulação da resposta imune adaptativa, resposta a interferon gama, estão *down-regulated*. Um gene de HLA superexpresso indica possivelmente a apresentação de antígeno por células como macrófagos e linfócitos, já no tecido nervoso, a microglia e o astrócito são as responsáveis por apresentar antígeno. Nesta condição, este gene está *down-regulated*, o que acarreta em níveis de citocinas mais baixos, menor inflamação do tecido cerebral, e conseqüentemente possível baixa permeabilidade da barreira hematoencefálica e recrutamento de células imunes da periferia para o parênquima cerebral, as quais poderiam estar transportando o YFV por mecanismo de cavalo de tróia. Um achado interessante nessa condição, é a superexpressão da via de regulação da morfogênese do epitélio ramificado, e desenvolvimento de vasculatura presente provavelmente proveniente no plexo coróide. Em contrapartida, na condição B, em que o vírus da febre amarela foi detectado no cérebro dos pacientes, foi possível observar uma grande quantidade de *pathways* ativados que estão relacionadas à resposta imune, em processos que envolvem produção de citocinas, resposta de defesa à vírus, resposta à interferon-gama e resposta à estresse celular (Figura 21).

**Figura 21-** Vias enriquecidas (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral)

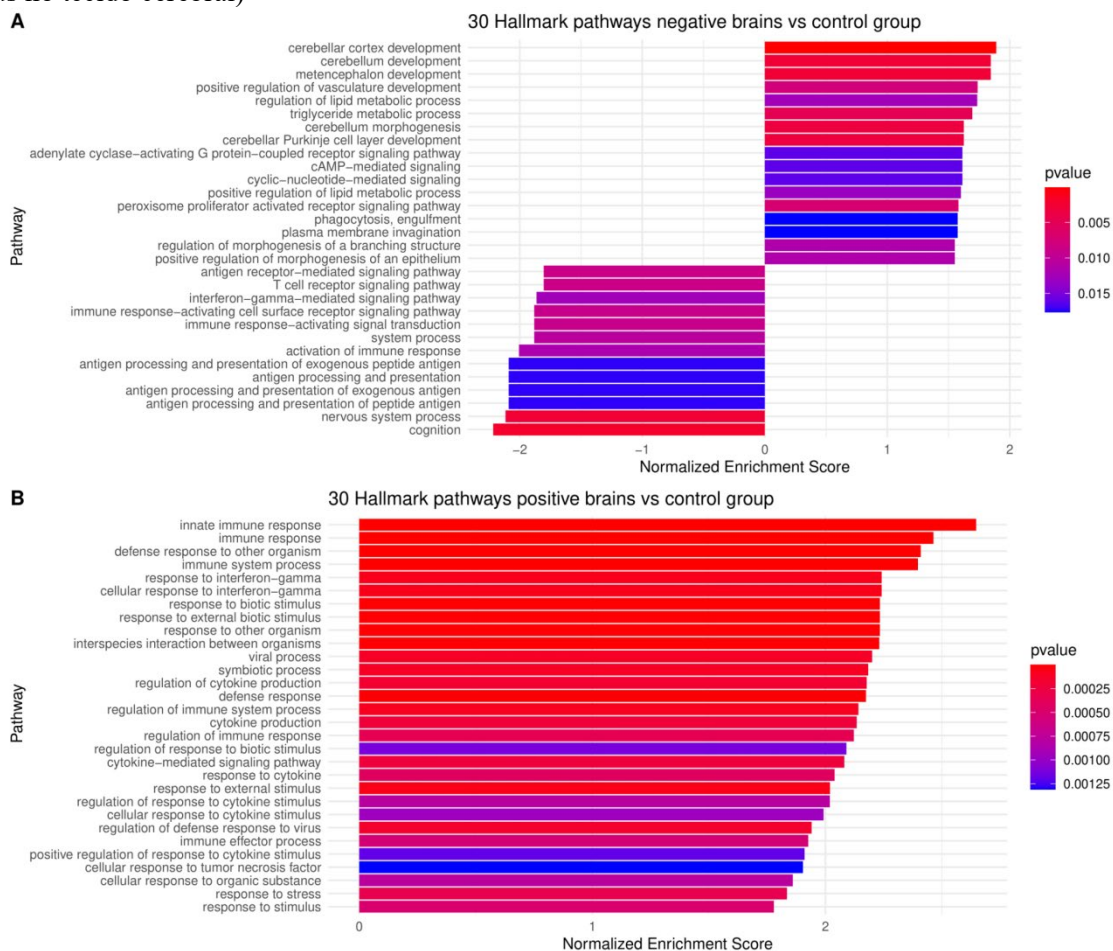


Vias enriquecidas para os grupos de pacientes com YFV detectado ou não detectado por RT-qPCR do RNA viral no tecido cerebral. (A) Vias super expressas e *down-regulated* na condição A (pacientes sem detecção do YFV no cérebro *versus* controle). (B) Vias super expressas e *down-regulated* na condição B (pacientes com detecção de FA no cérebro *versus* controle). No eixo x, o “gene ratio” representa a porcentagem do total de DEGs na via. O gráfico visualiza valores em duas dimensões: cor e tamanho. O gradiente de cor, representa a significância de valor de *p*-ajustado pelo método de correção de Bonferroni. Já o tamanho, representa a quantidade de genes dentro de cada de via.



Para análise, o GSEA dividiu genes em um subconjunto daqueles hiperegulados e outro *down-regulated*. Em seguida, nesta lista foram ranqueados aqueles genes que tendem a aparecer na parte superior da lista (mais hiperegulados) ou na parte inferior da lista (*down-regulated*) (Figura 22). O grau de enriquecimento foi estabelecido pelo NES (*normalized enrichment score*). Na Figura 22, são então mostradas as principais vias enriquecidas baseadas no ranqueamento por NES. Na condição A, em que o vírus da febre amarela não foi detectado no tecido cerebral, as principais vias que estão superexpressas tendem a estar relacionadas ao desenvolvimento e funcionamento homeostático do cérebro, como desenvolvimento do córtex cerebelar, mesencéfalo e sinalização mediada por cAMP. Nesta mesma condição as vias que obtiveram um maior *score*, porém estão *down-regulated*, estão envolvidas na ativação do sistema imune e apresentação de antígeno.

**Figura 22-** Vias enriquecidas pelo GSEA (pacientes negativos e positivos para a detecção molecular do vírus no tecido cerebral)

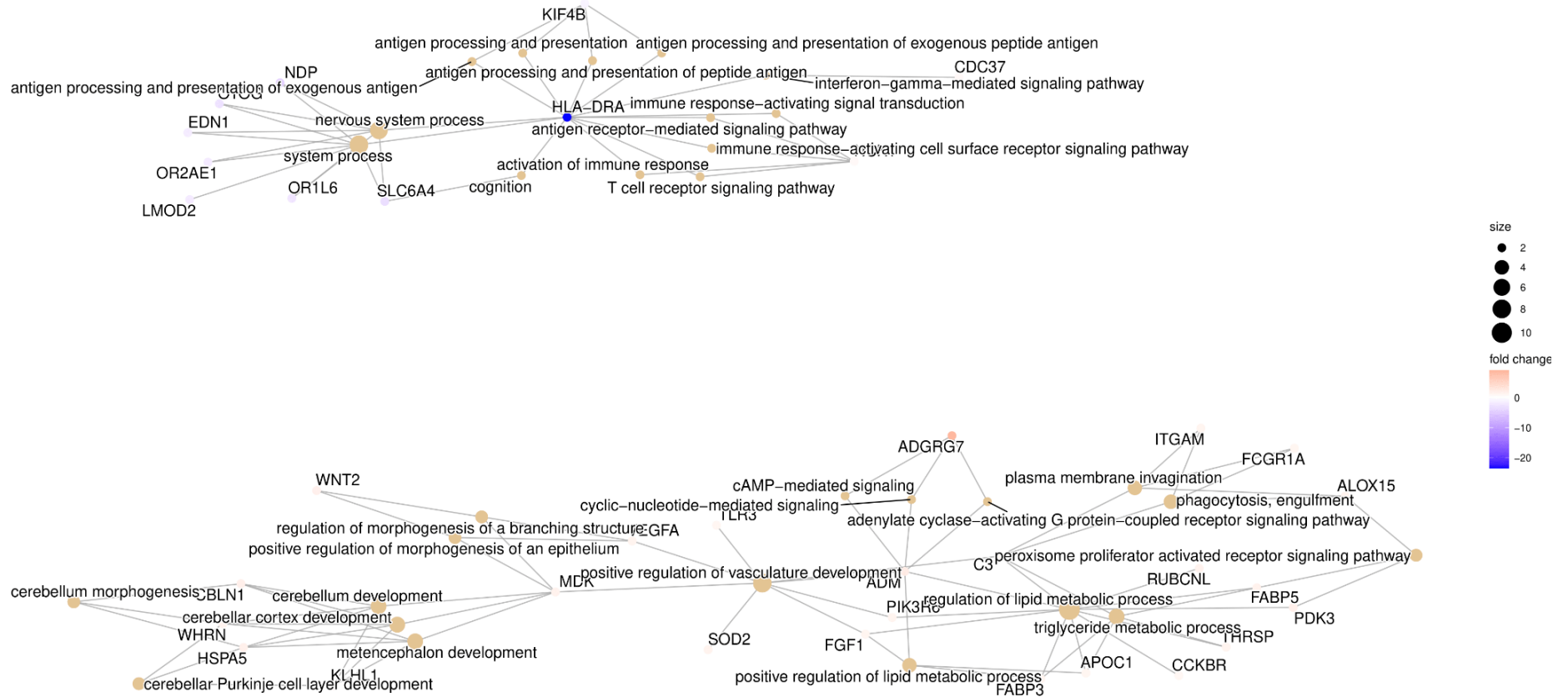


Vias enriquecidas pelo GSEA, que foram significativas ( $p < 0,05$ ), e seu valor respectivo valor de NES. Um valor de NES significativo positivo, indica que os membros do conjunto de genes tendem a aparecer no topo do transcriptoma e um NES significativo negativo, indica o oposto. Os valores de  $p$  são destacados pela variação de cores. (A) Principais vias em destaque da condição A, em que YFV não foi

detectado no cérebro. (B) Principais vias em destaque da condição B, em que YFV foi detectado no cérebro.

Os resultados obtidos com a análise em rede (gráfico *gene-concept network*), mostrou que para os pacientes da condição A, em que não houve vírus da febre amarela detectado no tecido cerebral, o gene *HLA-DRA* (também destacado anteriormente no volcano plot) (Figura 23), apresenta-se como central na rede, com ligação à funções associadas à resposta imunológica do hospedeiro frente à FA. Para a condição B em que o vírus da febre amarela foi detectado no tecido cerebral, ficou evidenciado dois genes centrais a processos biológicos importantes na caracterização da condição estudada (Figura 24). Diferente da análise obtida apenas pelo volcano plot (Figura 20), aqui é possível obter noção de conectividade entre termos e consequente centralidade de genes, a partir de como a rede está estruturada. Foi possível, portanto, observar a superexpressão do gene *CCL8*, central em termos de resposta imunológica e expressão de quimiocinas que apresentam papel importante no tráfico leucocitário e IF16, gene responsável pela produção do tipo I de interferon conhecido por regular a resposta antiviral.

**Figura 23 - Rede *Gene-concept* (pacientes negativos para a detecção molecular do vírus no tecido cerebral)**



Rede *Gene-concept*, gerado pelo ClusterProfiler, da condição A, em que não houve vírus da febre amarela detectado no tecido cerebral. Genes que são associados aos 20 termos principais ranqueados pelo GSEA, foram destacados na figura. Os nós representam os termos ontológicos, e a distância entre eles está diretamente relacionada à proporção de sobreposição de genes entre os termos. O tamanho do nó está relacionado ao número de genes associados a cada termo e as cores destacam os valores de FC.



A rede baseada em termos ontológicos e seus respectivos genes, pode ser de difícil interpretação quando se quer mostrar um grande número de termos significativos. Assim, a representação “*heatmap-like*” pode simplificar a visualização dos dados e facilitar a identificação de padrões de expressão (Figura 25). Na Figura 25, foi retratado o padrão de expressão de genes da condição A, em que o vírus não foi detectado no cérebro. Nela foi possível observar, assim como visto anteriormente na rede gênica (Figura X), a centralidade do gene *HLA-DRA*, por fazer parte das principais vias que caracterizam essa condição.

Por outro lado, quando retratado o padrão de expressão de genes da condição B, em que o vírus foi detectado no cérebro, não é possível observar um padrão de expressão claro, o que justifica-se pelo elevado número de genes presentes nas vias e assim o mapa “*heatmap-like*” deixa claro genes chaves para discutir a entrada do vírus da febre amarela no sistema nervoso, incluindo genes como, *CCL8* e *IFI16*, que são aqueles com maior expressão e genes tais como, *CASP1*, *CCL3L3*, *DDX58*, *IFI30*, *IFIH1*, *IFIT1*, *IFIT2*, *IFIT3*, *IKBKE*, *NMI*, *OASL*, *PARPA9*, *PSMB8*, *STAT1*, *TREM2*, *TREX1*, *TRIM21* e *TRIM34* mostram-se, também, centrais na caracterização das principais vias nesta condição (Figura 26).







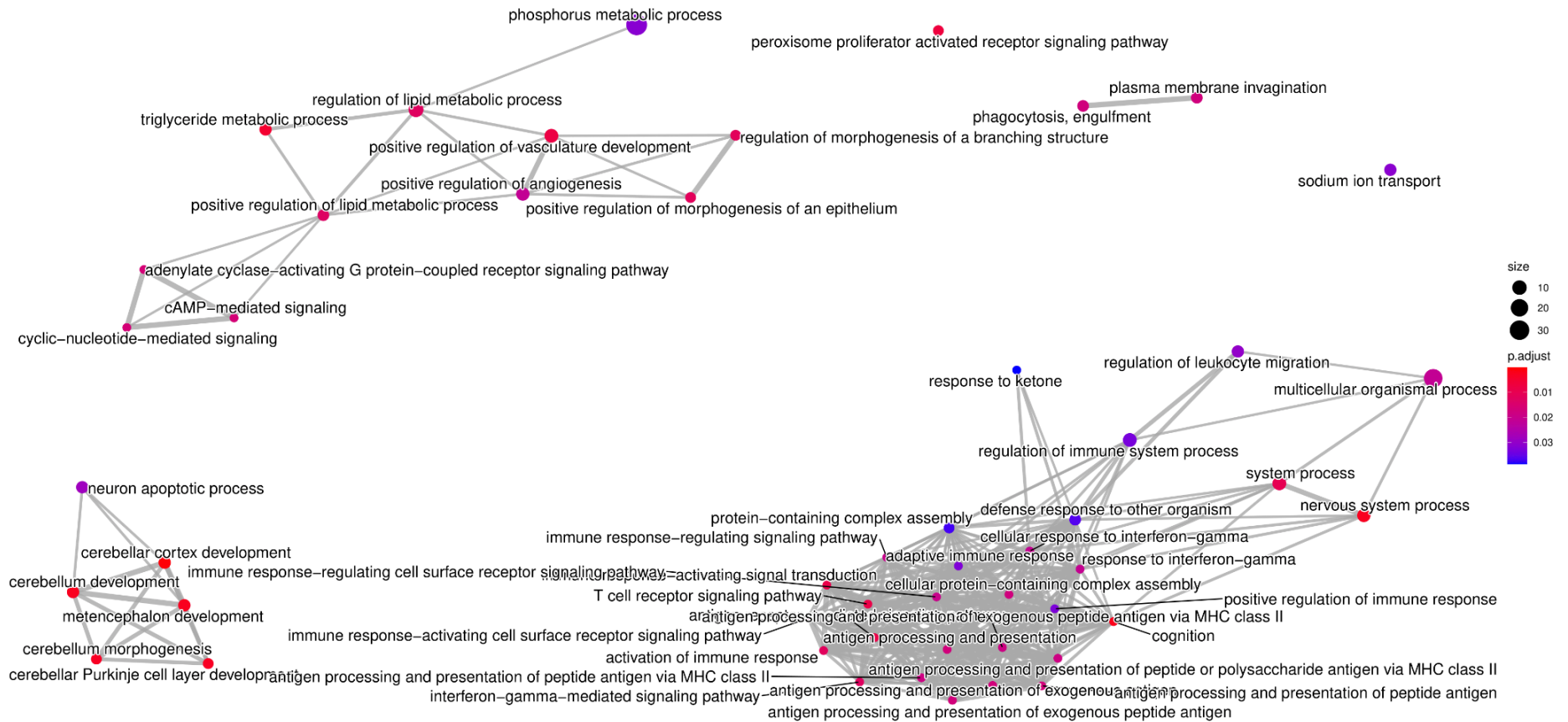
Para facilitar a interpretação dos dados de RNAseq, foi feito um mapa de enriquecimento utilizando métodos de visualização de resultados de enriquecimento de conjunto de genes baseados em rede. Métodos de sobreposição de *Jaccard* e *Szymkiewicz-Simpson* foram utilizados para organizar a rede quanto ao nível de similaridade entre os conjuntos de genes dentro de cada termo ontológico. Os nós são arranjados de maneira a agrupar-se em *clusters* e *sub-clusters* a depender da pontuação de sobreposição estabelecida pelos coeficientes. Os agrupamentos (*clusters* ou *sub-clusters*) podem ser relacionados a uma função biológica principal.

O mapa de enriquecimento dado por uma rede de similaridade ponderada relacionada a genes diferencialmente expressos na condição A, em que o YFV não foi detectado no cérebro, mostrou que, é possível observar 3 *clusters*, sendo o principal deles, ou seja, aquele com maior quantidade de vias incluídas e genes sobrepostos, relacionado a resposta imunológica adaptativa do hospedeiro e apresentação de antígeno via MHC tipo I. No entanto, como visto anteriormente, as vias relacionadas a este cluster estão *down-regulated*. Genes com menor coeficiente de sobreposição são encontrados no *cluster* relacionado ao metabolismo celular e morfogênese de tecido epitelial ramificado. O terceiro *cluster*, aquele com menor quantidade de vias relacionadas a termos ontológicos, está principalmente relacionado ao desenvolvimento da parte morfológica do encéfalo (Figura 27).

Por outro lado, o mapa de enriquecimento retratou a rede de similaridade ponderada relacionada a genes diferencialmente expressos na condição B, em que o YFV foi detectado no cérebro. Neste foi possível observar a formação de um único grande cluster em que houve sobreposição de todos os genes diferencialmente expressos nesta condição (Figura 28). A função biológica atribuída ao *cluster* está diretamente relacionada à resposta imune inata do hospedeiro frente à infecção causada pelo YFV, com a presença de vias que sugerem a tempestade de citocinas responsivas à inflamação do cérebro.

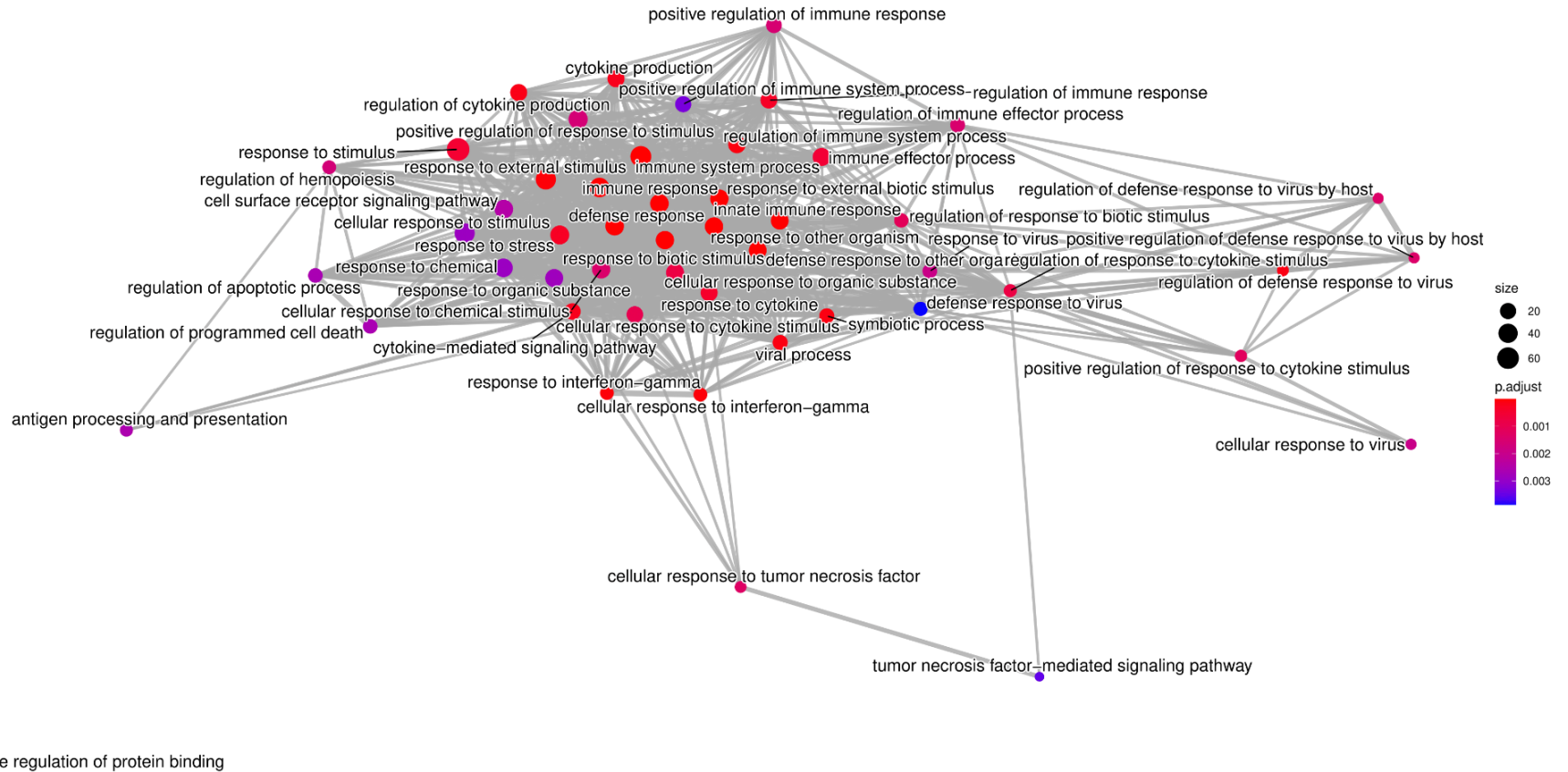


**Figura 27 - Rede de Similaridade ponderada (YFV não foi detectado no cérebro)**



Rede de Similaridade ponderada, com pesos (grau de sobreposição de genes, dentro dos termos ontológicos) atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição A, em que o YFV não foi detectado no cérebro. Os nós representam um conjunto de genes, com seu respectivo termo de processo biológico. O tamanho do nó representa o número de genes no conjunto de genes, já a espessura da borda é proporcional ao grau de sobreposição entre os conjuntos de genes, calculada pelo coeficiente de *Jaccard*. As distâncias de ligações entre nós, ou seja, as arestas, representam o grau sobreposição entre o conjunto de genes definida pela coeficiente de *Szymkiewicz-Simpson*. A variação da cor do nó está relacionada à variação do valor de *p* ajustado.

**Figura 28 - Rede de Similaridade ponderada (YFV foi detectado no cérebro)**



Rede de Similaridade ponderada, com pesos (grau de sobreposição de genes, dentro dos termos ontológicos) atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição B, em que o YFV foi detectado no cérebro. Os nós representam um conjunto de genes, com seu respectivo termo de processo biológico. O tamanho do nó representa o número de genes no conjunto de genes, já a espessura da aresta é proporcional ao grau de sobreposição entre os conjuntos de genes, calculada pelo coeficiente de *Jaccard*. As distâncias de ligações entre nós, ou seja, as arestas, representam o grau de sobreposição entre o conjunto de genes definida pela coeficiente de *Szymkiewicz-Simpson*. A variação da cor do nó está relacionada à variação do valor de *p* ajustado.

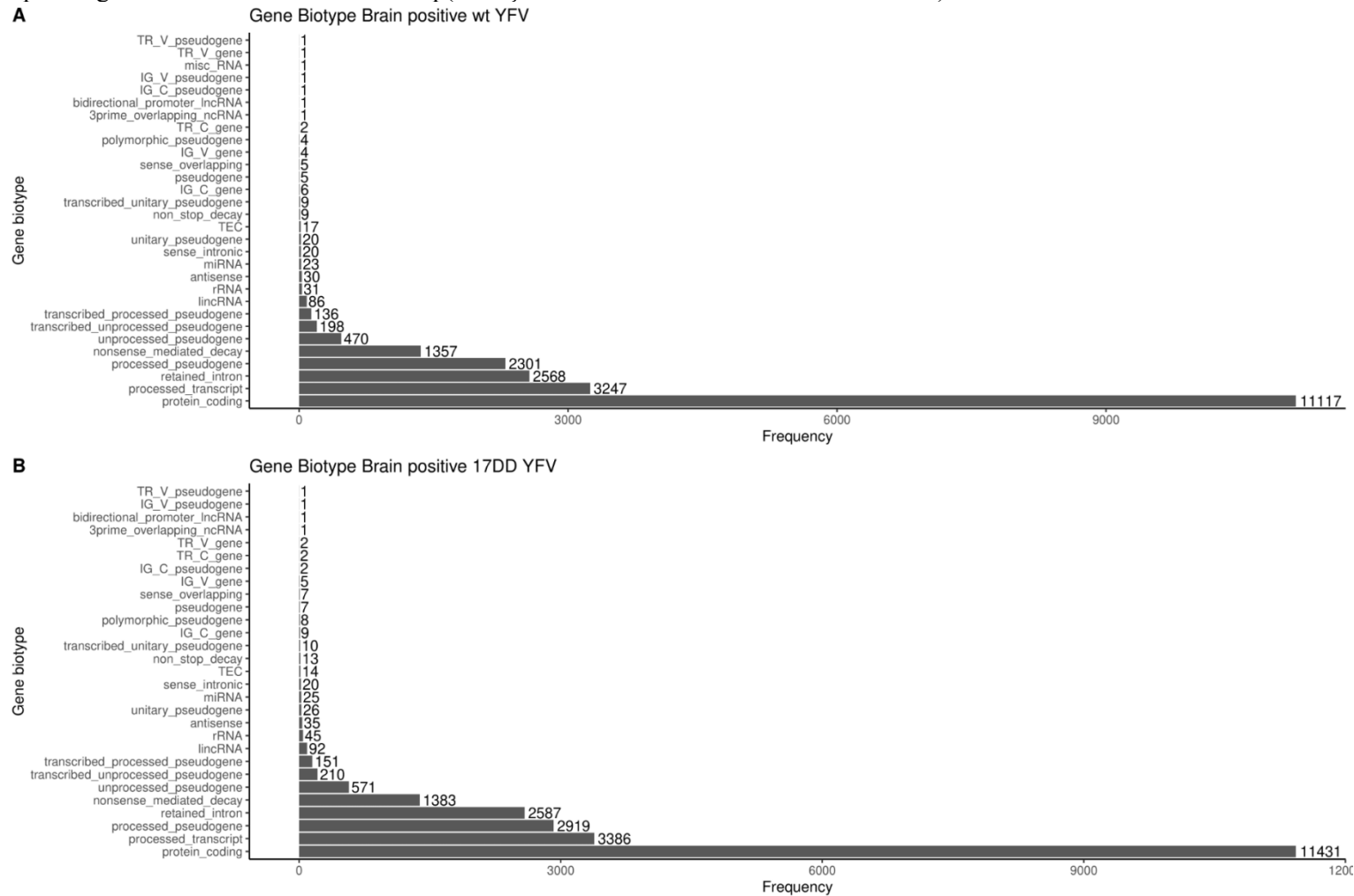
#### **5.4.2.2. Expressão diferencial entre grupo de pacientes com vírus da febre amarela silvático detectado no cérebro e pacientes com vírus da febre amarela vacinal detectado no cérebro**

A segunda hipótese alternativa a ser testada refere-se à diferença de expressão gênica entre pacientes que tiveram o vírus da febre amarela silvático e vacinal detectado no cérebro. Para esta análise, valores de expressão gênica do grupo 1 (YFV-silvático detectado no cérebro) e grupo 3 (YFV-17DD detectado no cérebro) foram confrontadas com o grupo 5 (controle).

##### **5.4.2.2.1 Expressão diferencial**

Foram quantificadas e classificadas a abundância de cada biotipo gênico para cada comparação, para fim de tomada de decisão sobre qual classe deveria ser incluída nas análises posteriores (Figura 29). Foram identificados 11117 genes codificadores de proteínas na comparação A (pacientes com o YFV-silvático detectado no cérebro), com 619 genes significativos (DEGs), com  $p < 0,05$ . Foram 11431 genes codificadores de proteínas comparação B (pacientes com o YFV vacinal detectado no cérebro), com 255 genes diferencialmente expressos ( $p < 0,05$ ).

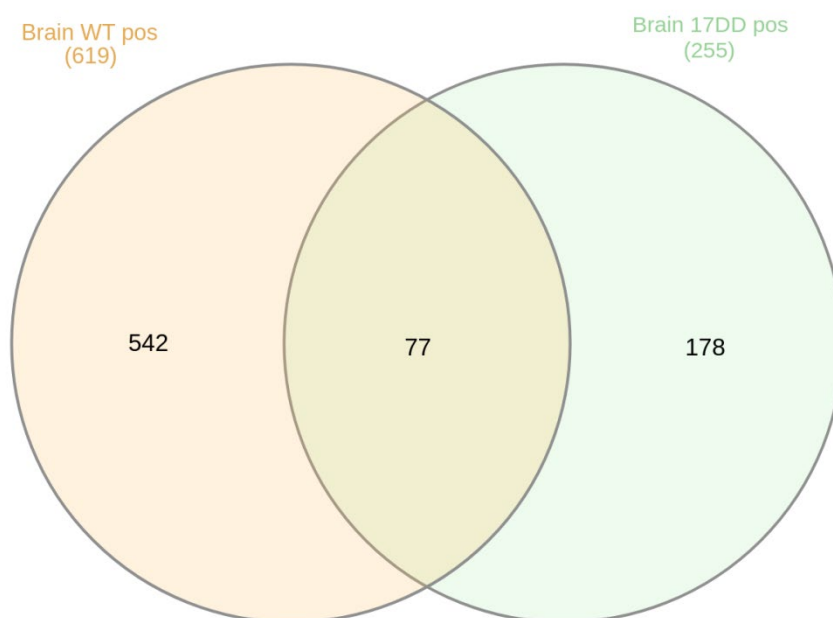
**Figura 29** - Grupos de genes caracterizados no RNAseq (detecção molecular do YFV no tecido cerebral)



Caracterização do perfil dos genes expressos considerando os grupos de pacientes positivos para a detecção molecular do vírus no tecido cerebral, da linhagem silvática e vacinal. (A) Condição A - pacientes com detecção de FA silvático no cérebro *versus* controle. (B) Condição B - pacientes com detecção de FA vacinal no cérebro *versus* controle.

Genes os quais são exclusivos de cada classe foram selecionados de forma que o enriquecimento gênico foi realizado apenas para este conjunto de genes. Na condição A, 542 genes são exclusivos, enquanto na condição B, 178 genes. Na intersecção foram encontrados 77 genes, classificados como pertencentes a ambas as condições (Figura 30).

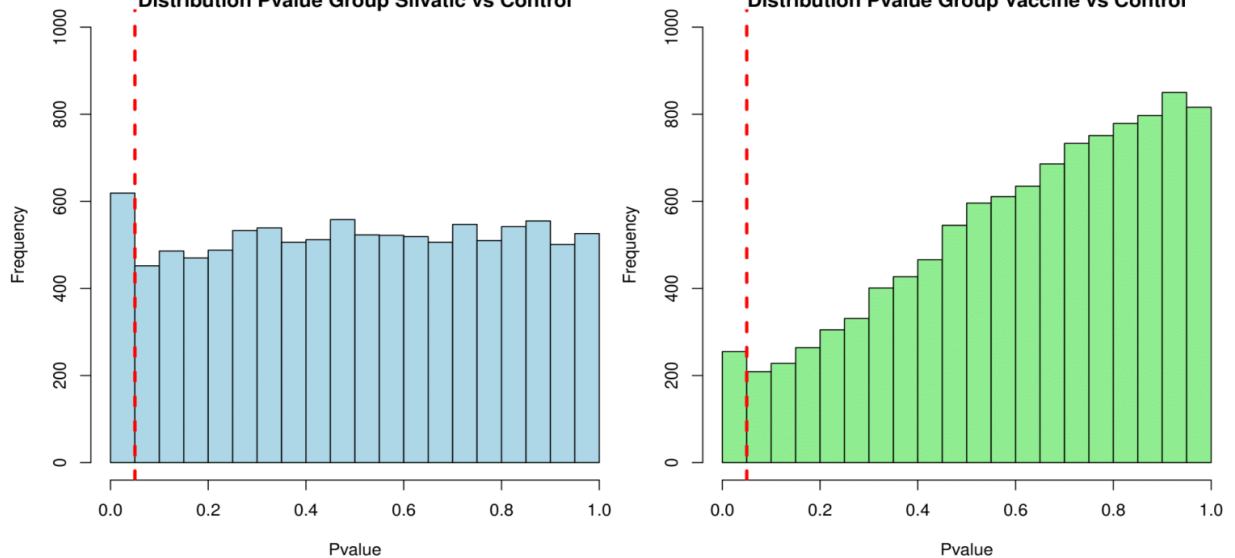
**Figura 30** – Comparação dos genes diferencialmente expressos (detecção molecular do YFV no tecido cerebral)



Comparação dos genes diferencialmente expressos considerando os grupos de pacientes positivos com YFV-silvático (WT-pos) e vacinal (17DD-pos) detectado no tecido cerebral. Em laranja, condição A - pacientes com o YFV-silvático detectado no cérebro *versus* controle. Em verde, condição B - pacientes com o YFV vacinal detectado no cérebro *versus* controle. A intersecção de classes indica os DEGs comuns encontrados em ambas as classes consideradas nesta análise mostrando genes interclasses.

Histogramas da distribuição de valores de  $p$  mostram o padrão de expressão de cada grupo estudado (Figura 31). Os dois grupos possuem um padrão de distribuição bem diferentes, mostrando um comportamento distinto de expressão gênica. Na condição A, em que o vírus silvático foi detectado no cérebro, por mais que exista uma grande proporção de genes que aceitem a hipótese nula ( $p > 0,05$ ), uma curva em “L” indica maior frequência de genes com  $p$  significativos em relação aos demais valores de  $p$ . Já na condição B a frequência de genes que aceitam a hipótese nula é muito maior, podendo assim sugerir uma maior agressividade da linhagem silvática em relação à vacina.

**Figura 31** - Distribuição de valores de  $p$   
Distribution Pvalue Group Silvatic vs Control



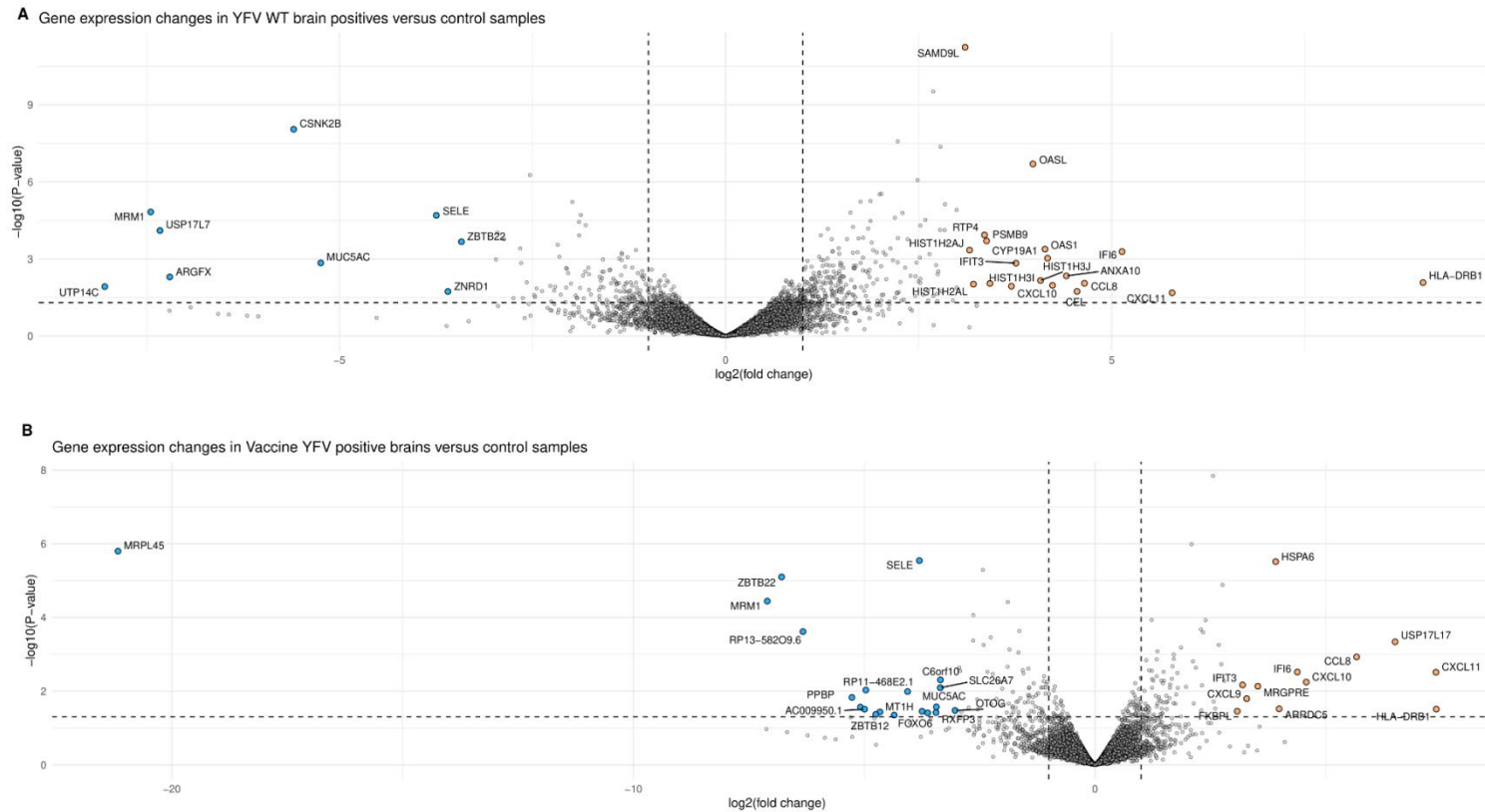
Distribuição de valores de  $p$  para condição (A) em que houve detecção do YFV-silvático no cérebro, e grupo (B) em que houve detecção do YFV-17DD no cérebro. A linha pontilhada vermelha indica o *threshold* dos genes considerados diferencialmente expressos com valor de  $p < 0,05$ .

De forma a hipotetizar quais genes estão envolvidos na entrada do vírus silvático e vacinal no tecido cerebral, foram comparados os dois grupos em busca de assinaturas referentes a cada linhagem.

Foram destacados alguns genes com maior FoldChange (ou diferença de expressão) e com valor de  $p$  significativo, nos volcano plots (Figura 32). Na condição A, em que foi detectado o YFV-silvático no cérebro, alguns genes *down-regulated* mostraram-se relevantes como: *UTP14C*, *MRM1*, *USP17L7*, *ARGFX*, *CSNK2B*, *MUC5AC*, *SELE*, *ZBTB22* e *ZNDRI*, porém não houve nenhum com posição de destaque. O gene *HLA-DRB1*, foi superexpresso nesta condição cerca de 4000 vezes mais do que no grupo controle ( $FC = \log_2(12)$ ).

Já na condição B, em que foi detectado o YFV-17DD no cérebro, o gene *MRPL45* foi down-regulado mais de 4 milhões de vezes em relação ao grupo controle ( $FC = \log_2(22)$ ) e ganha destaque perante os outros genes com  $\log_2(FC)$  em torno de  $\log_2(2-5)$ .

**Figura 32 - Volcano plot** mostrando a expressão gênica diferencial para o YFV-silvático e YFV-17DD



(A) *Volcano plot* representando os genes expressos na condição A - pacientes com detecção de YFV-silvático no cérebro em comparação ao grupo controle. (B) *Volcano plot* representando os genes expressos na condição B - pacientes com detecção de YFV-17DD no cérebro em comparação ao grupo controle. O eixo y representa o valor negativo de log10 do p-value para cada gene [ $-\log_{10}(p \text{ values})$ ], e traçado um *threshold* onde o valor de p é 0,05, mostrando valores significativos acima do tracejado. No eixo x, foi representado o log2 do valor de *fold-change* (FC) para cada gene, indicando quais grupo de genes estão *up-regulated* ou *down-regulated* em comparação com o grupo controle. O FC é calculado pelo valor de expressão do gene na condição B/valor da expressão do gene na condição A. O *threshold* de 1 e -1 Log2FC foi estabelecido. Os genes representados em azul são os que satisfizeram a condição  $p < 0,05$  &  $FC \leq -3$  (*down-regulated*). Os genes que satisfizeram a condição  $p < 0,05$  &  $FC \geq 3$ , estão *up-regulated* e são representados em laranja.

#### 5.4.2.2 Análise de ontologia

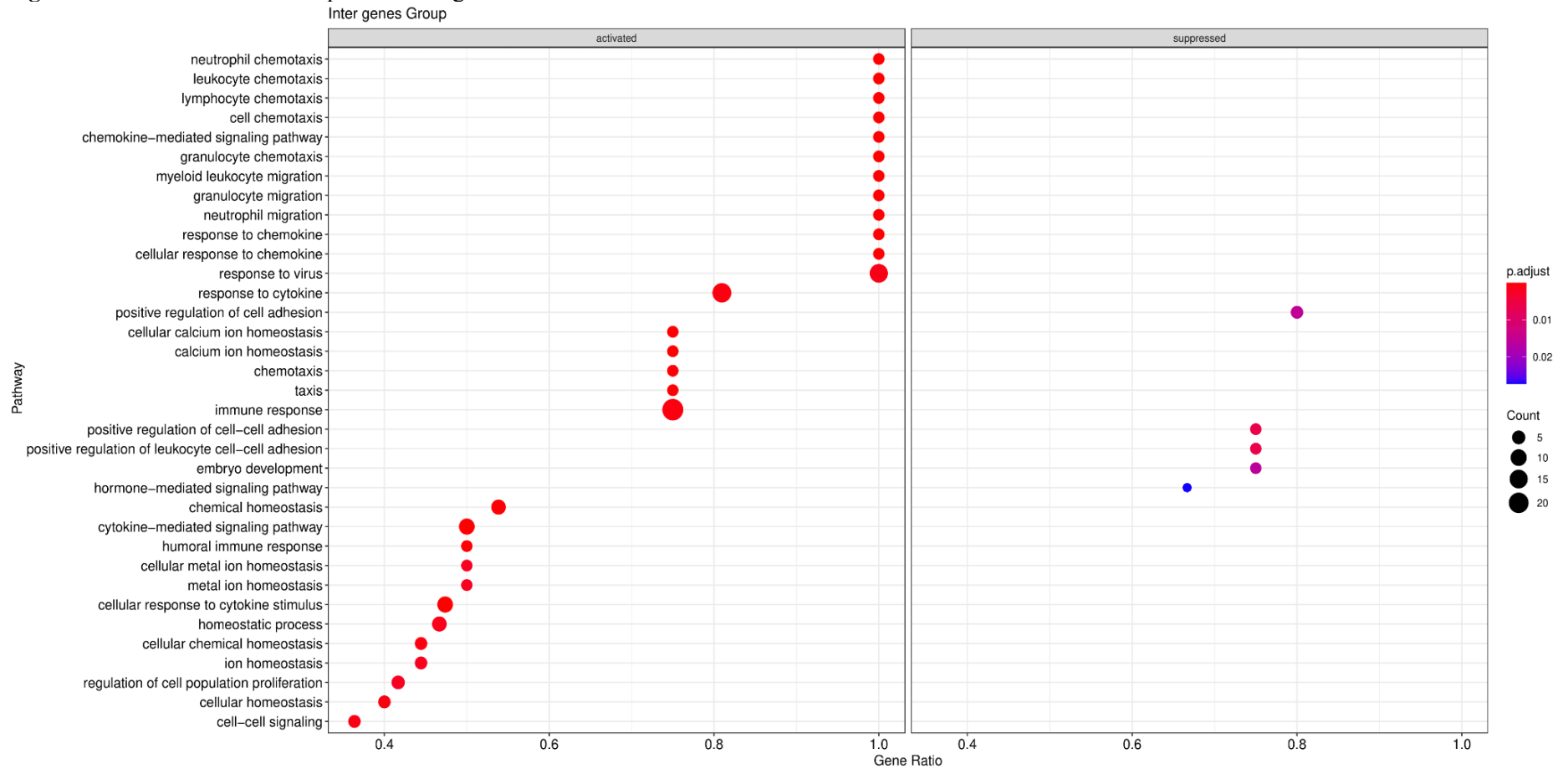
Genes compartilhados entre as duas condições estudadas, foram enriquecidos buscando inferir quais seriam os processos comuns às duas linhagens que estariam relacionados à detecção do YFV no cérebro. Foram, portanto, enriquecidos 74 genes interclasses e analisados na Figura 33. De forma geral, foi possível observar grande quantidade de vias superexpressas associadas à resposta antiviral, produção de citocinas, quimiocinas, recrutamento de leucócitos, linfócitos e neutrófilos.

O grau de enriquecimento foi estabelecido pelo NES (*normalized enrichment score*), e todas as vias com valor de  $p$  significativa foram superexpressas e tendem a estar relacionadas a uma resposta imune exacerbada, com recrutamento de citocinas pró-inflamatórias, leucócitos, linfócitos e neutrófilos. A presença do vírus no tecido cerebral, pode estar levando a apoptose de neurônios, que conseqüentemente tende a regular vias que reestabelecem a homeostase celular (via super expressa de homeostase de íon de cálcio). A superexpressão de vias relacionadas a proliferação celular pode ser relacionada à proliferação de astrócitos, que frente a estados patológicos, como infecções, podem fazer hiperplasia ou hipertrofia e apresentar um estado reativo, com produção de citocinas e tentativa de conter "lesão", processo este denominado astrogliose reativa. (Figura 34).

Os resultados obtidos com a análise em rede (gráfico *gene-concept network*), mostram destaque ao gene *CXCL11* que é superexpresso, e em tese, devido à encefalite viral por FA, é responsável principalmente por expressar quimiocinas regulatórias e regular a migração de células T para o local da inflamação (Figura 35). A centralidade deste gene, nas principais funções biológicas características desta condição, torna este gene extremamente relevante para a elucidação da doença FA no cérebro. Outro gene com destaque foi gene de *HLA* com o haplótipo *DRB1*, que é superexpresso, porém não desempenha papel central nas principais vias significativas destacadas. Esta informação se torna ainda mais relevante, e de fácil acesso, quando observada a Figura 36, um *heatmap-like* de enriquecimento dos genes interclasses que mostra uma via central, com muitos dos principais genes inclusos nela. A via de resposta imunológica ativa é esperada nesta condição em que o cérebro está inflamado. Como ressaltado anteriormente, genes classificados como principais ganham destaque por terem papel central na elucidação de funções biológicas comuns às duas linhagens de FA. Estes genes, portanto, devem fazer parte da maioria das vias enriquecidas que foram significativas. Nesta condição, os genes *CCL-8*, *CXCL-10* e *CXCL-11* são principais e fazem também parte da principal via biológica de resposta imunológica.

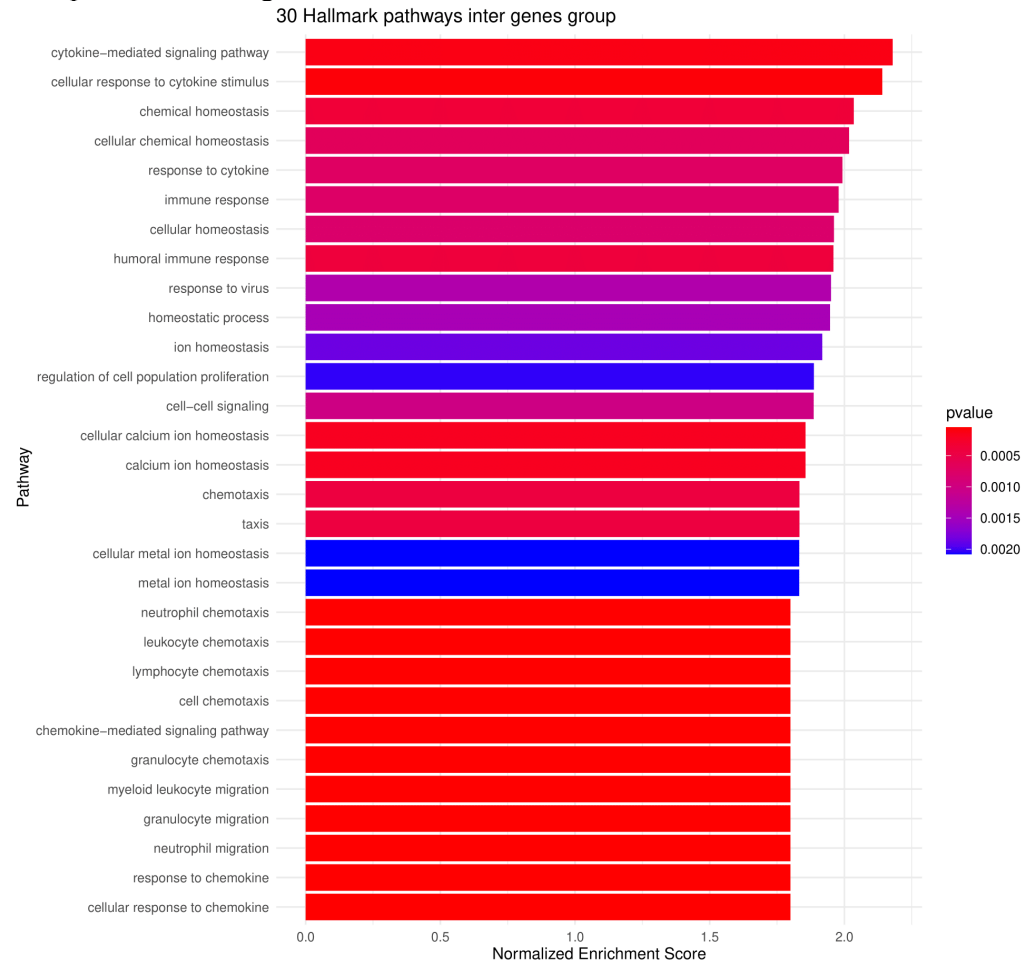


**Figura 33 - Diferentes vias enriquecidas dos genes interclasses**



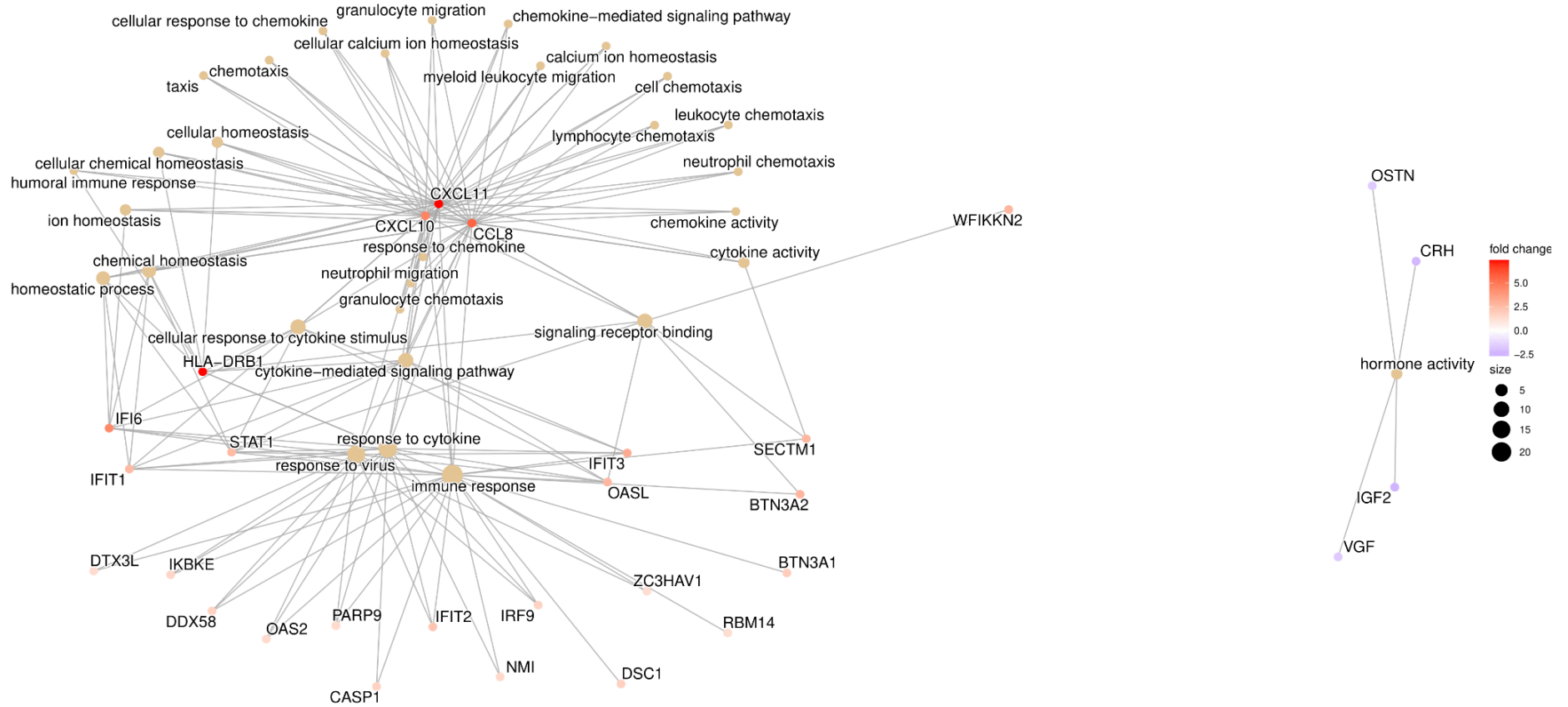
No eixo x, o “*gene ratio*” representa a porcentagem do total de DEGs na via dada. O gráfico visualiza valores em duas dimensões: cor e tamanho. A cor, representa a significância de valor de *p* ajustado pelo método de correção de Bonferroni (*p* valores brutos\* número de comparações). Já o tamanho, representa a quantidade de genes dentro de cada de via. Ao lado esquerdo do gráfico estão as vias superexpressas nesta condição, enquanto ao lado direito as vias *down-regulated*.

**Figura 34 - Principais vias enriquecidas pelo GSEA dos genes interclasses**



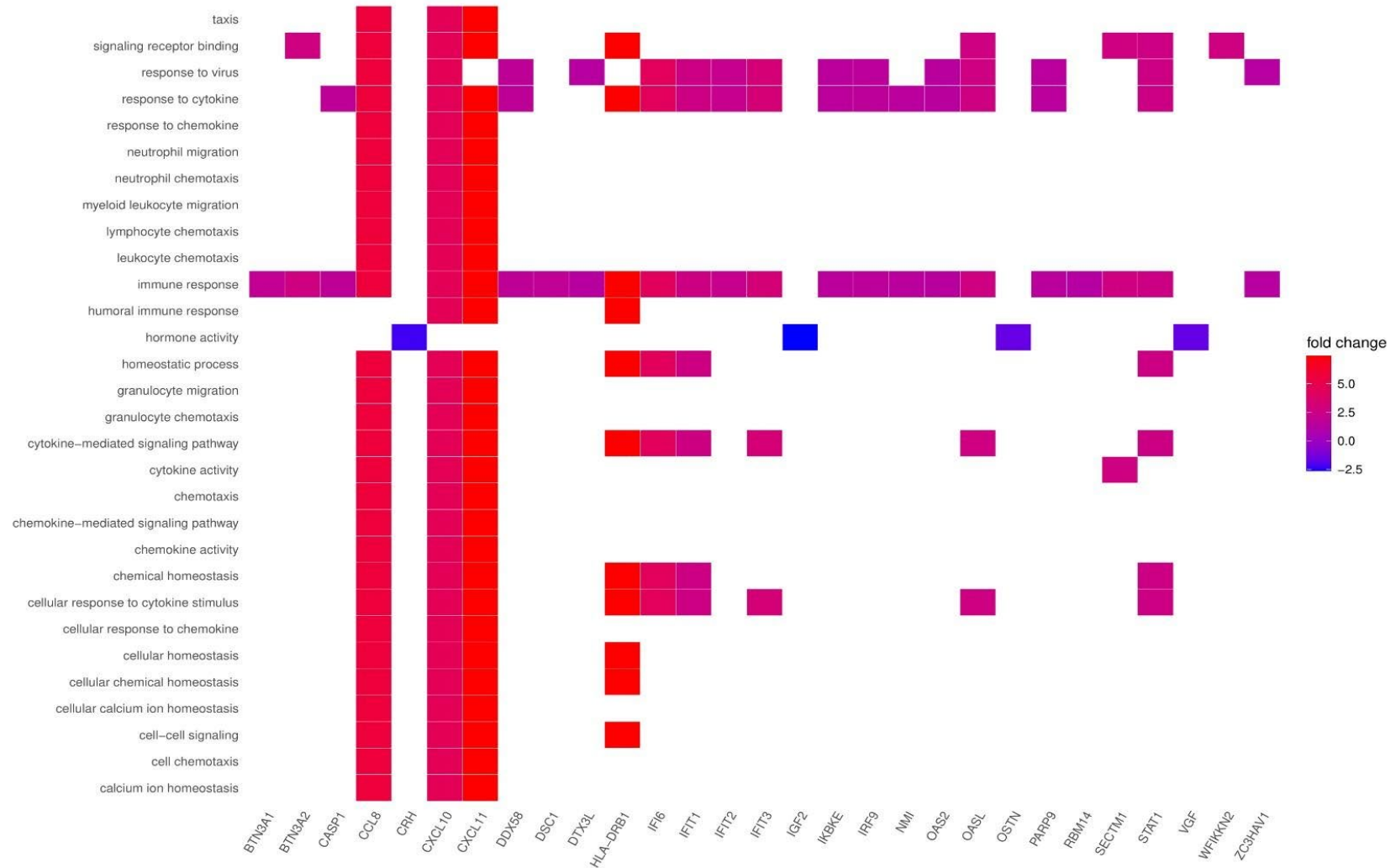
30 principais vias enriquecidas pelo GSEA, que foram significativas ( $p < 0,05$ ), e seu valor respectivo valor de NES. Um valor de NES significativo positivo, indica que os membros do conjunto de genes tendem a aparecer no topo do transcriptoma e estão super expressos em relação à condição controle. Os valores  $p$  são destacados pela variação de cores.

**Figura 35 - Rede Gene-concept da condição em que foram analisados os genes interclasses**



Genes que são associados aos 20 termos principais ranqueados pelo GSEA, foram destacados na figura. Os nós representam os termos ontológicos, e a distância entre eles está diretamente relacionada à proporção de sobreposição de genes entre os termos. O tamanho do nó está relacionado ao número de genes associados a cada termo e as cores destacam os valores de FC.

**Figura 36 - Classificação funcional dos genes para o grupo de genes interclasses**



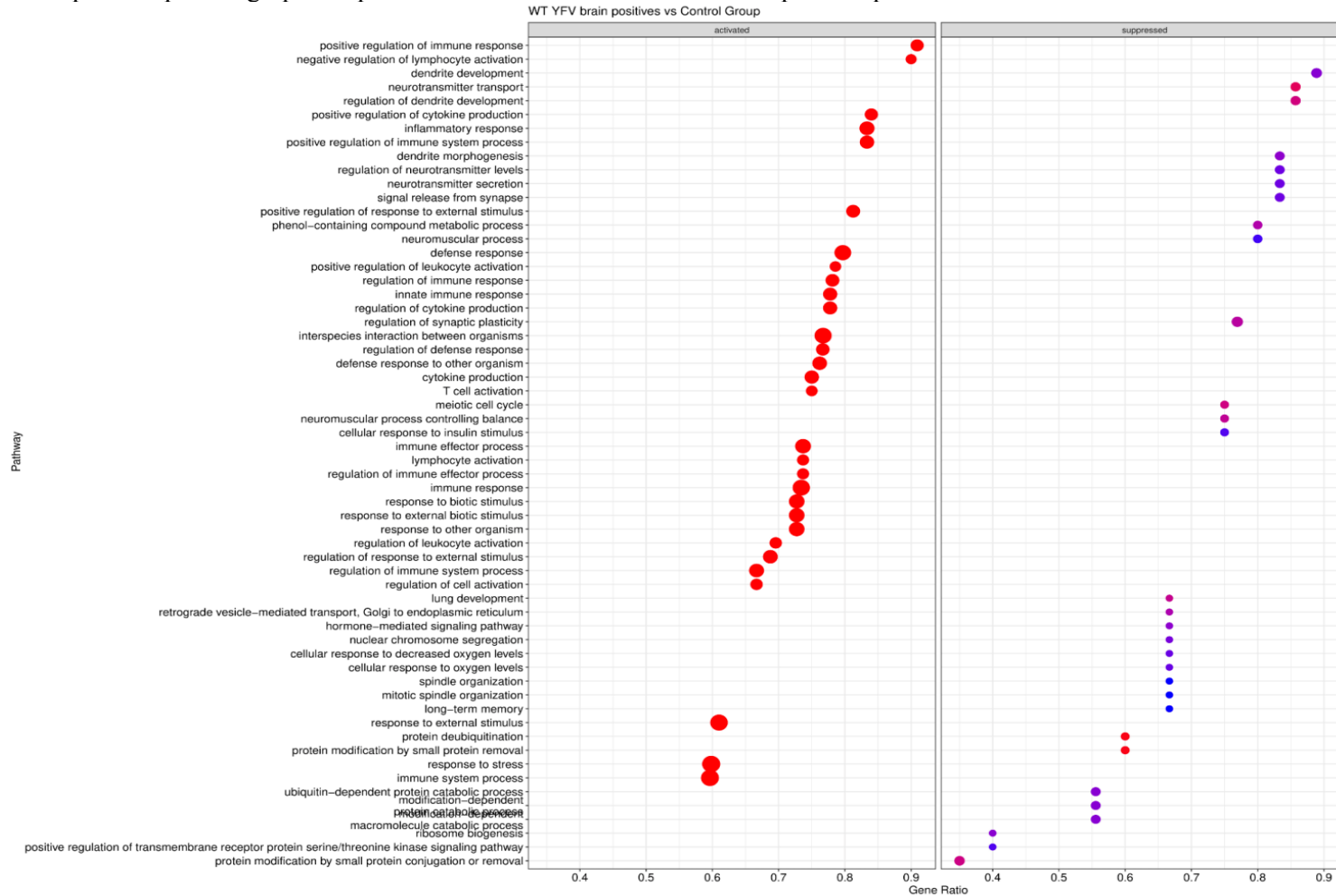
No eixo y, estão as principais vias enriquecidas relativas aos genes interclasses. No eixo x, os genes que fazem parte dessas vias, mostrando um padrão de expressão gênica característico do que é comum nas linhagens virais de febre amarela. A variação de cor indica variação do FC.

As principais vias de processos biológicos possuem maior quantidade de genes associados a elas. Na condição A, em que o vírus da febre amarela silvático foi detectado no cérebro, fica claro um padrão de resposta imunológica adaptativa envolvendo produção de citocinas que irão atuar na ativação e recrutamento de linfócitos e leucócitos e também resposta inata frente a infecção (Figura 37). Além de recrutar células do sistema imune, as citocinas também ativam astrócitos e microglia no SNC, que vão agir como mecanismo de defesa do tecido no contexto da infecção. No entanto, destaca-se a *down-regulation* de genes que estão em vias relacionadas à neurotransmissão, como: secreção de neurotransmissores, liberação de sinais nas sinapses, plasticidade sináptica e morfogênese dos dendritos neuronais (Figura 37). Estes resultados, podem estar possivelmente relacionados a sequelas neurológicas observadas em pacientes com sobrevivência pós infecção por febre amarela da linhagem silvática.

As principais vias de resposta à presença do vírus vacinal no tecido cerebral parecem ser predominantemente provenientes de uma resposta imunológica inata, com produção de interferon-gama, componente chave do sistema imune contra infecções virais e consequente diferenciação de linfócitos e células T (Figura 38).

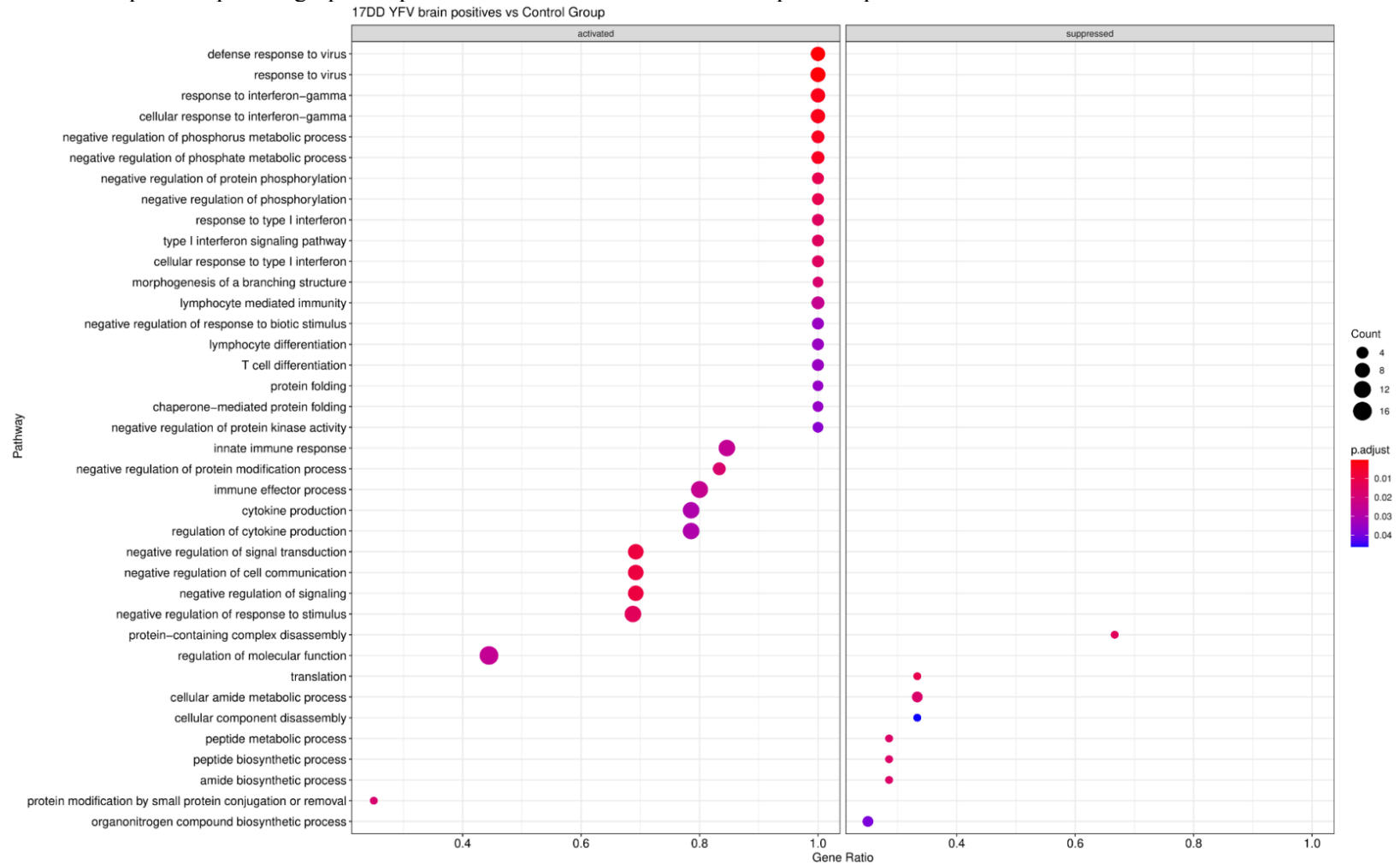
As principais vias selecionadas baseadas no NES, mostram que na condição A (Figura 39A), em que foi detectado o YFV-silvático no cérebro. Todas as vias ranqueadas que foram significativas estão superexpressas e, assim como visto anteriormente, mostram um padrão de resposta imunológica adaptativa ativa. Na condição B, em que o YFV-17DD foi detectado no cérebro, encontrou-se genes vinculados a vias superexpressas de produção de interferon do tipo 1, contribuindo para a primeira linha de defesa antiviral e também de interferon-gamma, que atua como molécula de citocina ativando células fagocíticas, linfócitos T e B (Figura 39B).

**Figura 37 - Vias enriquecidas para os grupos de pacientes com YFV-silvático detectado por RT-qPCR do RNA viral no tecido cerebral**



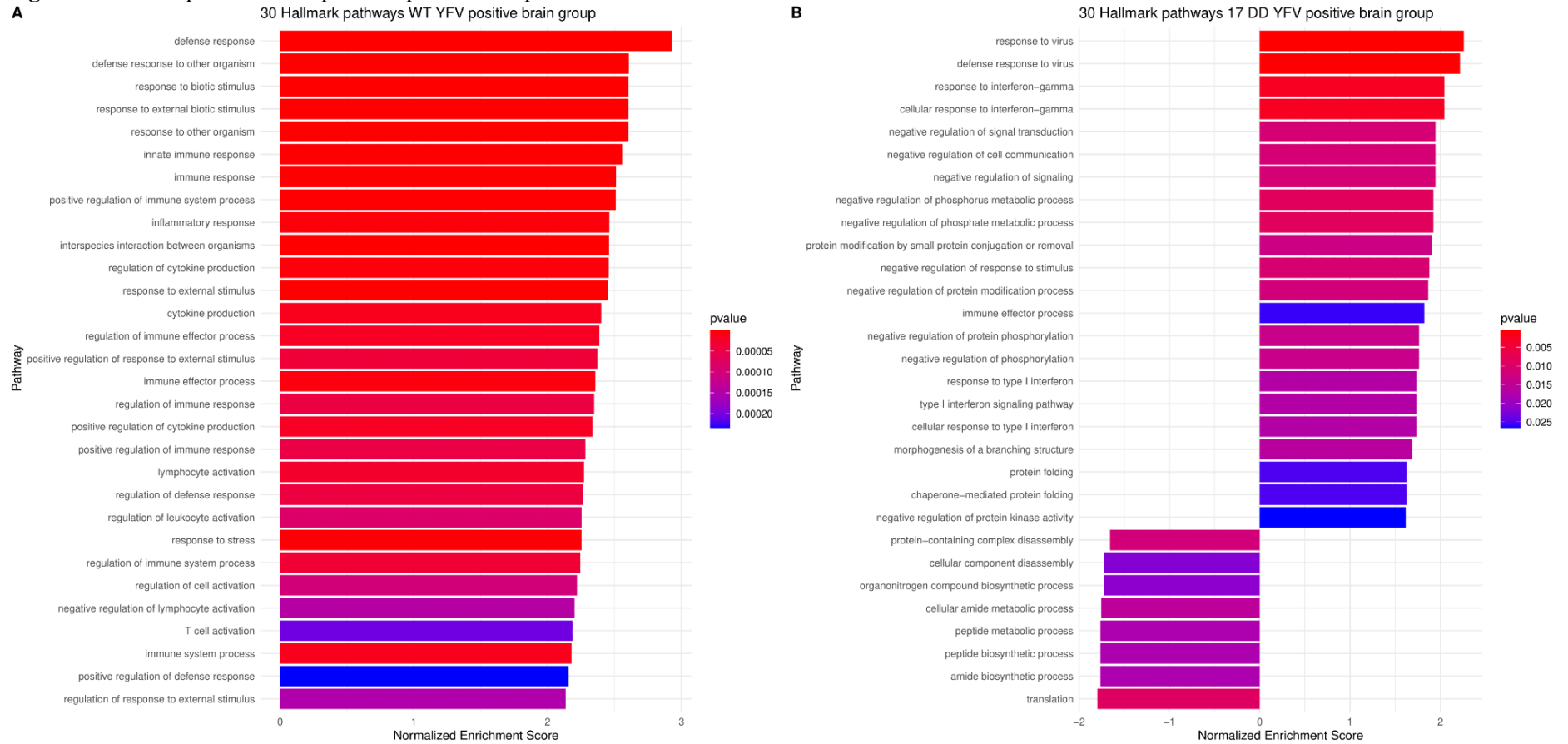
No eixo x, o “gene ratio” representa a porcentagem do total de DEGs na via dada. O gráfico visualiza valores em duas dimensões: cor e tamanho. A cor, representa a significância de valor de  $p$  ajustado pelo método de correção de Bonferroni ( $p$  valores brutos\* número de comparações). Já o tamanho, representa a quantidade de genes dentro de cada de via. Ao lado esquerdo do gráfico estão as vias superexpressas nesta condição, enquanto ao lado direito as vias *down-regulated*.

**Figura 38** - Vias enriquecidas para os grupos de pacientes com YFV-17DD detectado por RT-qPCR do RNA viral no tecido cerebral



No eixo x, o “gene ratio” representa a porcentagem do total de DEGs na via dada. O gráfico visualiza valores em duas dimensões: cor e tamanho. A cor, representa a significância de valor de  $p$  ajustado pelo método de correção de Bonferroni ( $p$  valores brutos\* número de comparações). Já o tamanho, representa a quantidade de genes dentro de cada de via. Ao lado esquerdo do gráfico estão as vias superexpressas nesta condição, enquanto ao lado direito as vias *down-regulated*.

**Figura 39 - Principais vias enriquecidas pelo GSEA para o YFV-silvático e YFV-17DD**



Principais vias enriquecidas pelo GSEA, que foram significativas ( $p < 0,05$ ), e seu valor respectivo valor de NES. Um valor de NES significativo positivo, indica que os membros do conjunto de genes tendem a aparecer no topo do transcriptoma e estão superexpressos em relação à condição controle. Os valores de  $p$  são destacados pela variação de cores. (A) Destaque as 30 principais vias mais significativas na condição A, em que o YFV-silvático foi detectado no cérebro. (B) Destaque as 30 principais vias mais significativas na condição B, em que o YFV vacinal foi detectado no cérebro.



A rede gênica (gráfico *gene-concept network*) obtida da análise de enriquecimento dos genes diferencialmente expressos da condição A, em que o YFV-silvático foi detectado no cérebro, mostrou uma única rede, com 3 *sub-clusters*. Em um deles, as suas vias estão relacionadas a resposta imunológica adaptativa, com ativação de linfócitos, leucócitos e produção de citocinas pró-inflamatórias. Em outro *sub-cluster*, a resposta inata é característica da resposta de defesa do hospedeiro a organismos externos. O último *sub-cluster*, envolve genes associados a resposta à estresse. Destacam-se entre os *clusters*, muitos genes centrais associados às vias (Figura 40). Contudo, ao sobrepor a análise desta figura à análise da Figura 41, em que foi apresentado o mapa *heatmap-like* nesta condição, fica claro quais são os principais genes (incluídos na maioria das vias destacadas - eixo y), que fazem parte também das principais vias (que incluem a maioria dos genes - eixo x). As principais vias são de resposta imunológica e de resposta a estresse. Já os principais genes são, *AIF1*, *ANXA1*, *C3AR1*, *CD80*, *CD84*, *DHX58*, *FCER1G*, *IFI16*, *MNDA*, *SUCNR1*, *TLR6*, *TREM2*, *VSIG4*.

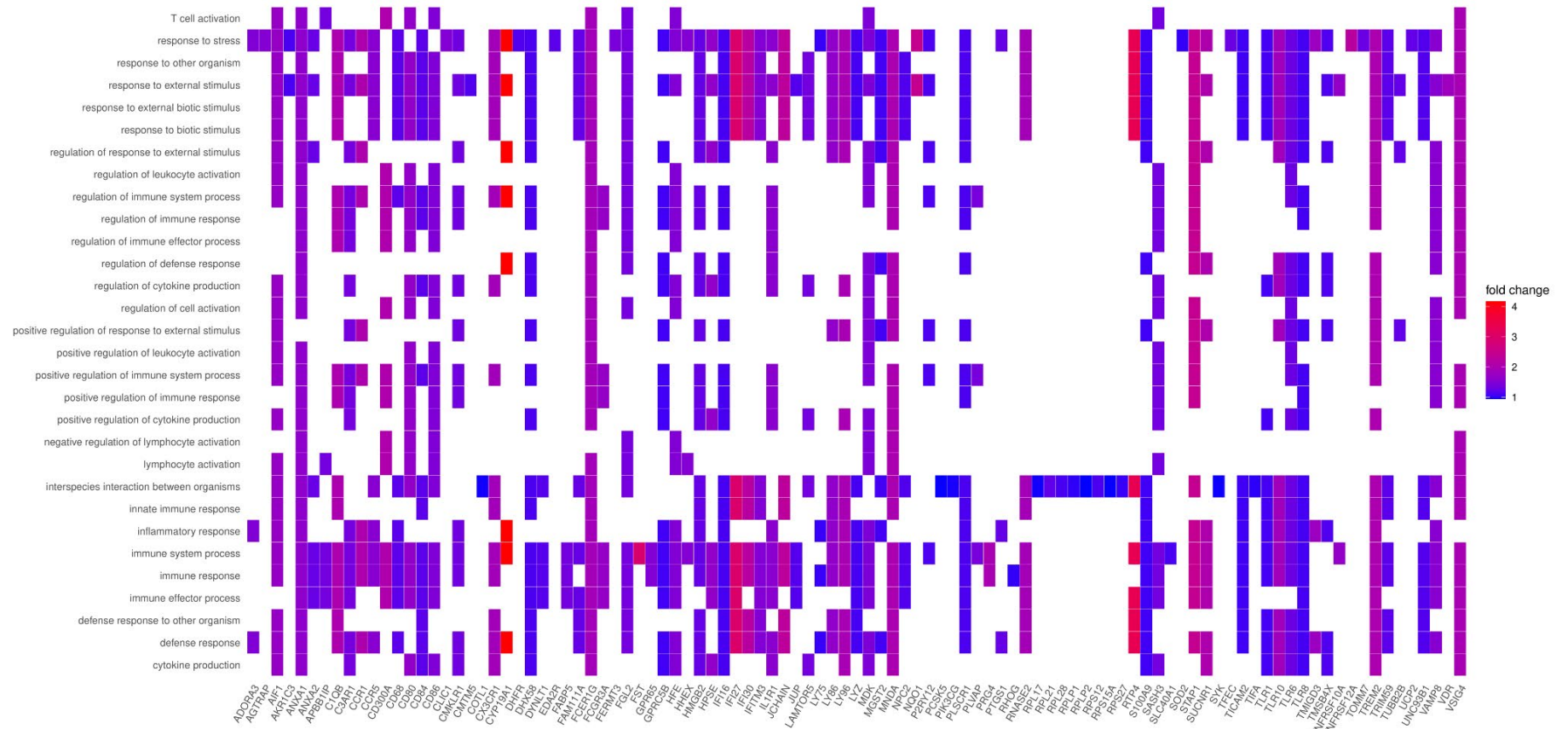
O mapa de enriquecimento dado por uma rede de similaridade ponderada relacionada a genes diferencialmente expressos na condição A, em que o YFV-silvático foi detectado no cérebro mostra um único grande *cluster* com função biológica principal de regulação de resposta imunológica, com produção de citocinas perante a uma resposta inflamatória (Figura 42).

A rede gênica que foi obtida via análise de enriquecimento de genes diferencialmente expressos da condição B, em que o YFV-17DD foi detectado no cérebro, mostra três *clusters*, sendo que a principal deles possui genes vinculados a resposta imune inata, com produção de interferon-gama. O gene *MRPL45*, é destaque em um *cluster* em que processos de metabolismo celular são ressaltados (Figura 43).

Diferente do observado na condição A, tanto a rede gênica, quanto o mapa *heatmap-like* não deixam claro vias principais, com genes principais relacionados a elas (Figura 44). Contudo, ressalta, a expressão de  $\log_2(20)$  vezes menos do gene *MRPL45*, o qual expressa proteínas do ribossomo mitocondrial. O mapa de enriquecimento de genes relacionado a condição B, mostra dois grandes *clusters*, sendo o principal deles, ou seja, aquele com maior quantidade de vias incluídas e genes sobrepostos, relacionado a resposta imunológica e transdução de sinal intracelular. Já o outro *cluster* menor, tem como função biológica principal funções de metabolismo celular (Figura 45).

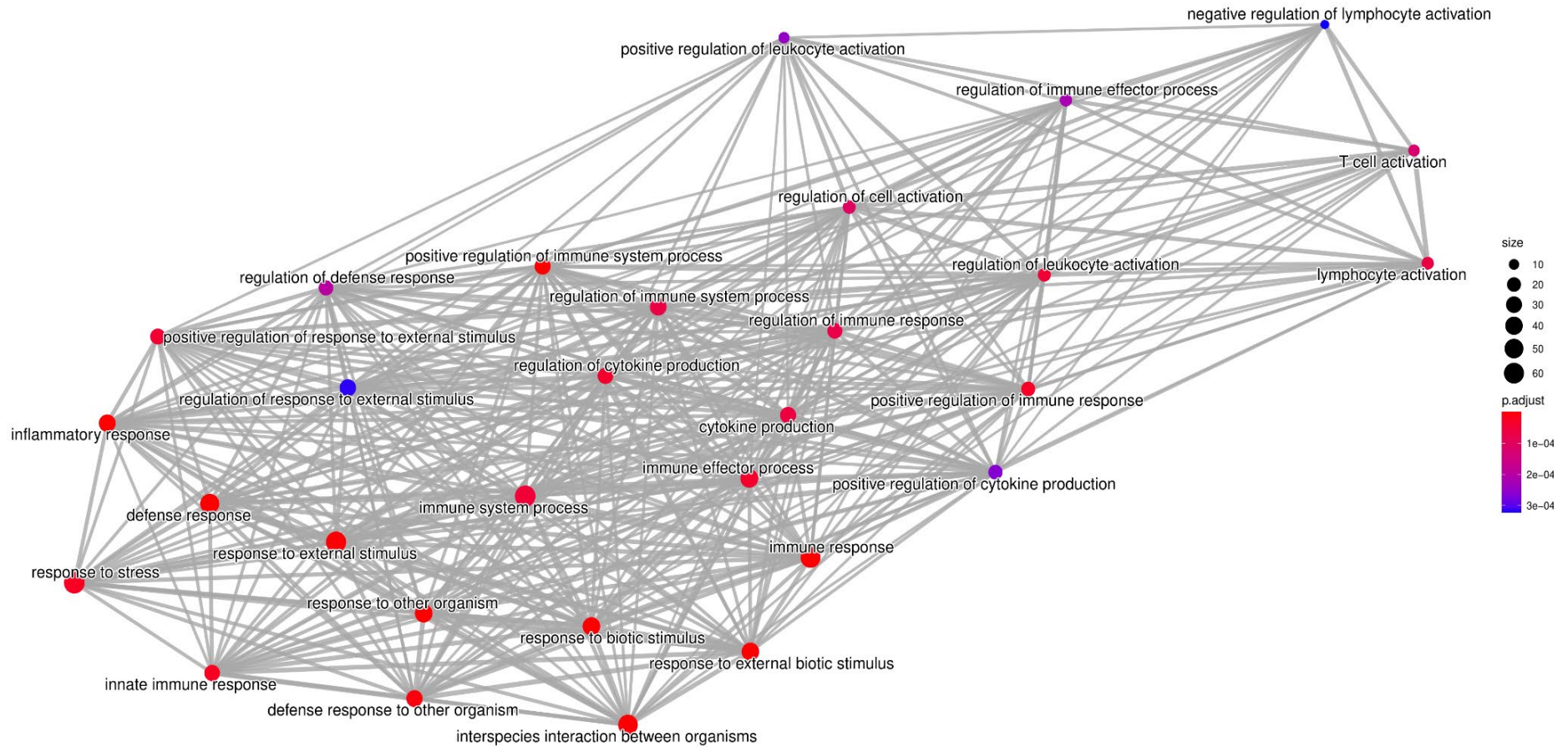


**Figura 41 - Classificação funcional dos genes diferencialmente expressos na condição A, em que o YFV-silvático foi detectado no cérebro**



No eixo y, estão as principais vias enriquecidas relativas aos genes interclasses. No eixo x, os genes que fazem parte dessas vias, mostrando um padrão de expressão gênica característico do que é comum nas linhagens virais de febre amarela. A variação de cor indica variação do FC.

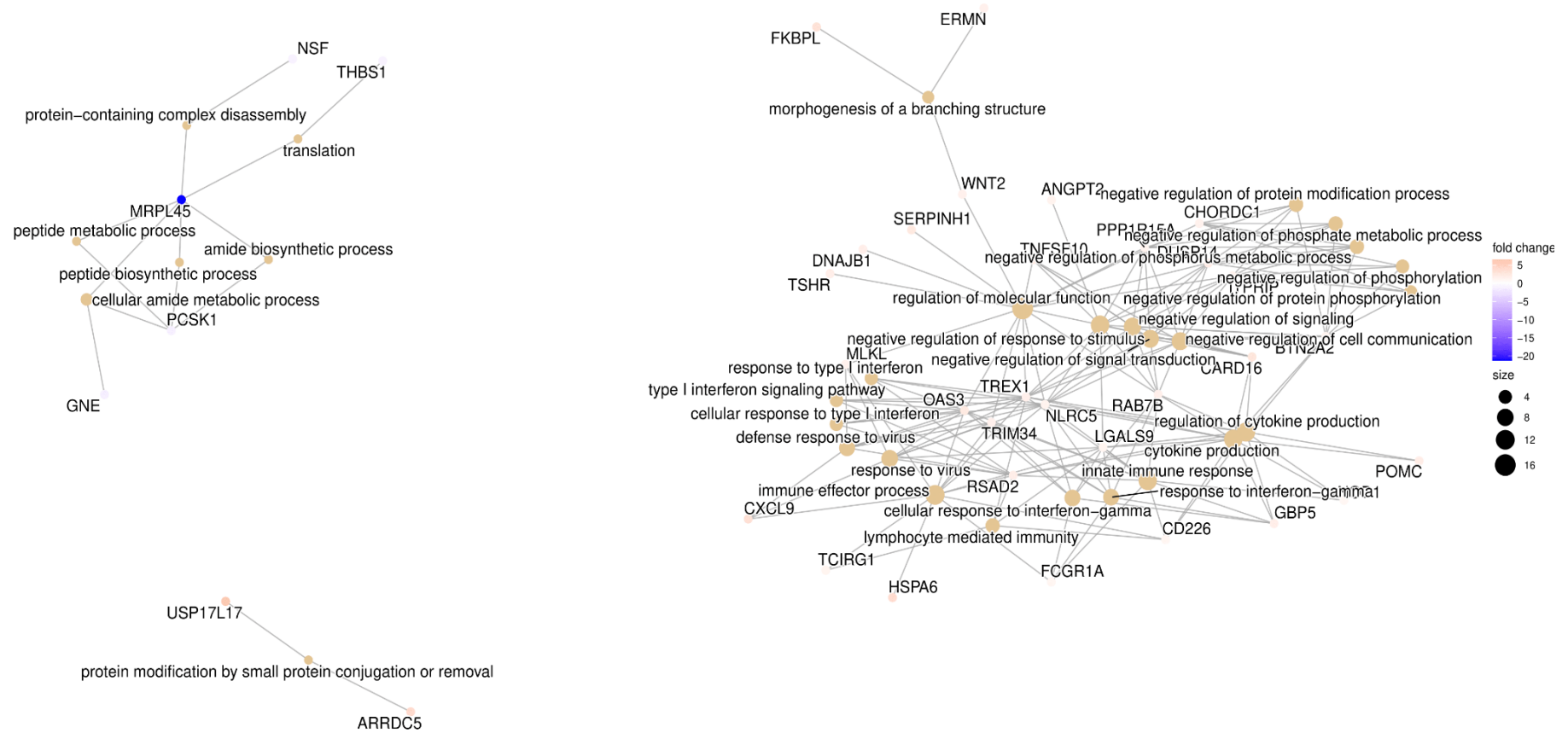
**Figura 42** - Rede de Similaridade ponderada para os genes relacionados à condição A, em que o YFV-silvático foi detectado no cérebro



Rede de Similaridade ponderada, com pesos (grau de sobreposição de genes, dentro dos termos ontológicos) atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição A, em que o YFV-silvático foi detectado no cérebro. Os nós representam um conjunto de genes, com seu respectivo termo de processo biológico. O tamanho do nó representa o número de genes no conjunto de genes, já a espessura da aresta é proporcional ao grau de sobreposição entre os conjuntos de genes, calculada pelo coeficiente de Jaccard. As distâncias de ligações entre nós, ou seja, as arestas, representam o grau de sobreposição entre o conjunto de genes definida pela coeficiente de Szymkiewicz-Simpson. A variação da cor do nó está relacionada a variação do valor de  $p$  ajustado.

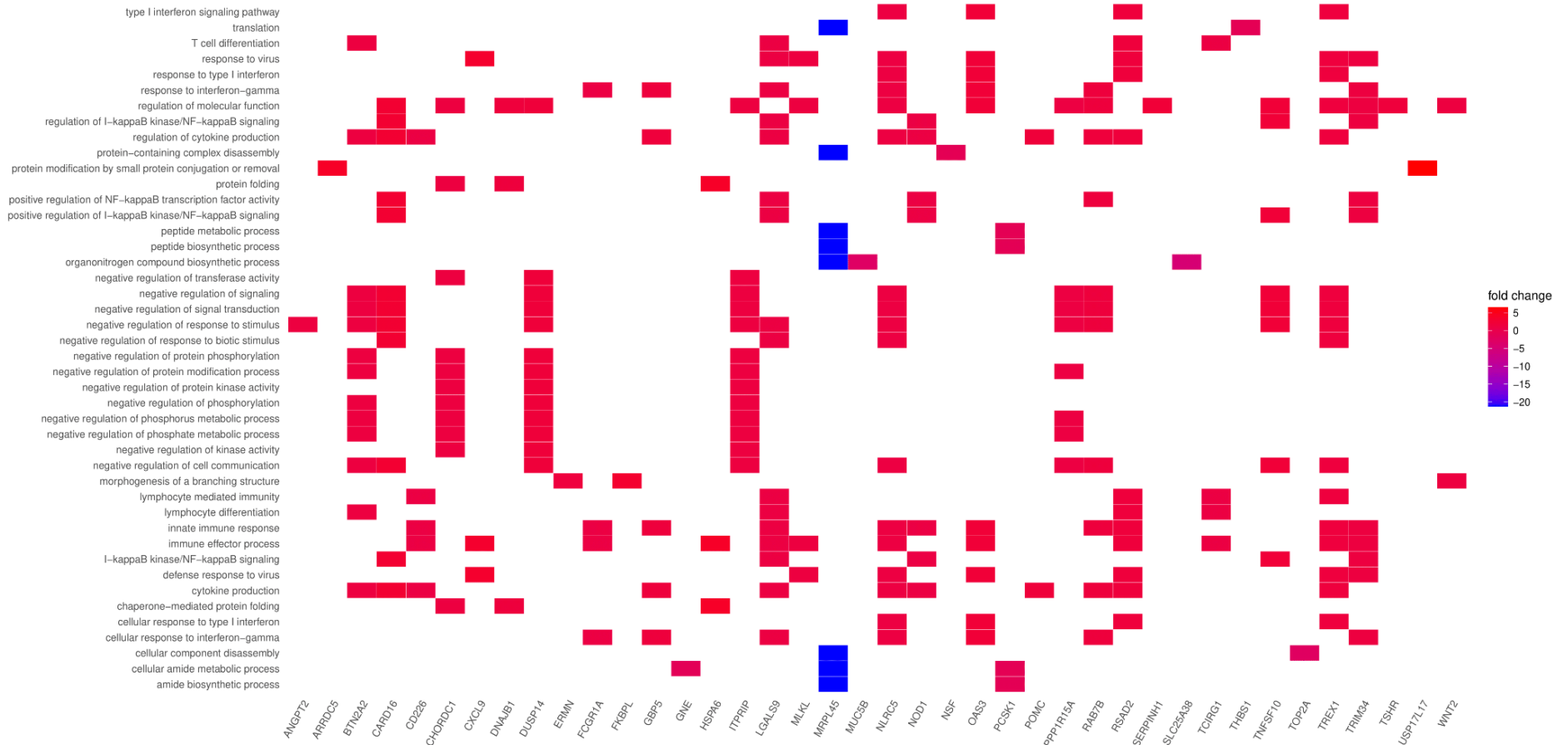


**Figura 43 - Rede Gene-concept (YFV-17DD foi detectado no cérebro)**



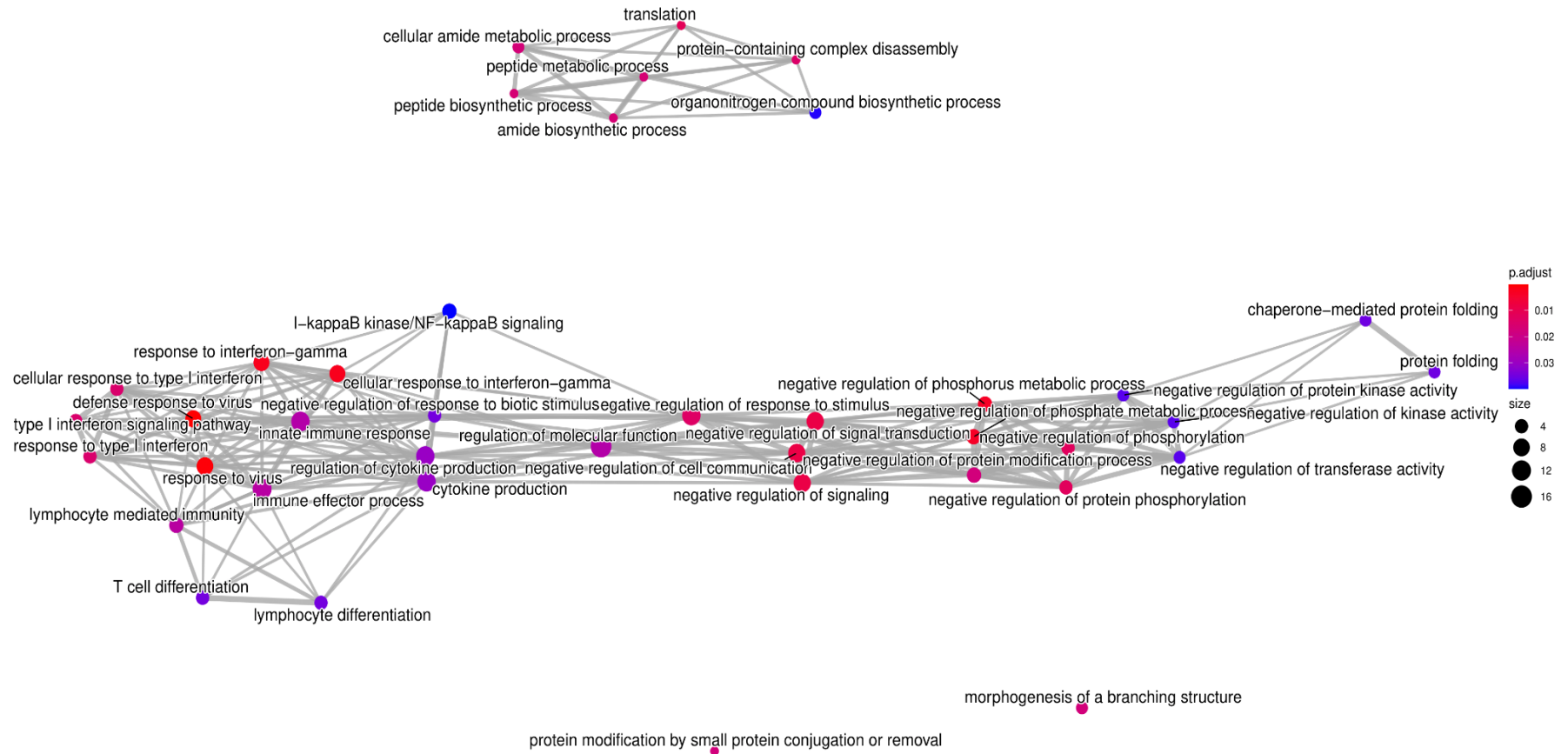
Rede *Gene-concept* da condição em que foram analisados os genes diferencialmente expressos na condição B, em que o YFV-17DD foi detectado no cérebro. Genes que são associados aos 20 termos principais ranqueados pelo GSEA, foram destacados na figura. Os nós representam os termos ontológicos, e a distância entre eles está diretamente relacionada à proporção de sobreposição de genes entre os termos. O tamanho do nó está relacionado ao número de genes associados a cada termo e as cores destacam os valores de FC.

**Figura 44 - Classificação funcional dos genes diferencialmente expressos na condição B, em que o YFV-17DD foi detectado no cérebro**



No eixo y, estão as principais vias enriquecidas relativas aos genes interclasses. No eixo x, os genes que fazem parte dessas vias, mostrando um padrão de expressão gênica característico do que é comum nas linhagens virais de febre amarela. A variação de cor indica variação do FC.

**Figura 45 - Rede de Similaridade ponderada (condição B), em que o YFV-17DD foi detectado no cérebro**



Rede de Similaridade ponderada, com pesos (grau de sobreposição de genes, dentro dos termos ontológicos) atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição B, em que o YFV-17DD foi detectado no cérebro. Os nós representam um conjunto de genes, com seu respectivo termo de processo biológico. O tamanho do nó representa o número de genes no conjunto de genes, já a espessura da borda é proporcional ao grau de sobreposição entre os conjuntos de genes, calculada pelo coeficiente de Jaccard. As distâncias de ligações entre nós, ou seja, as arestas, representam o grau de sobreposição entre o conjunto de genes definida pela coeficiente de Szymkiewicz-Simpson. A variação da cor do nó está relacionada à variação do valor de  $p$  ajustado.

### **5.4.2.3. Expressão diferencial entre grupo de pacientes que não tiveram o vírus da febre amarela silvático detectado no cérebro e pacientes que não tiveram o vírus da febre amarela vacinal detectado no cérebro**

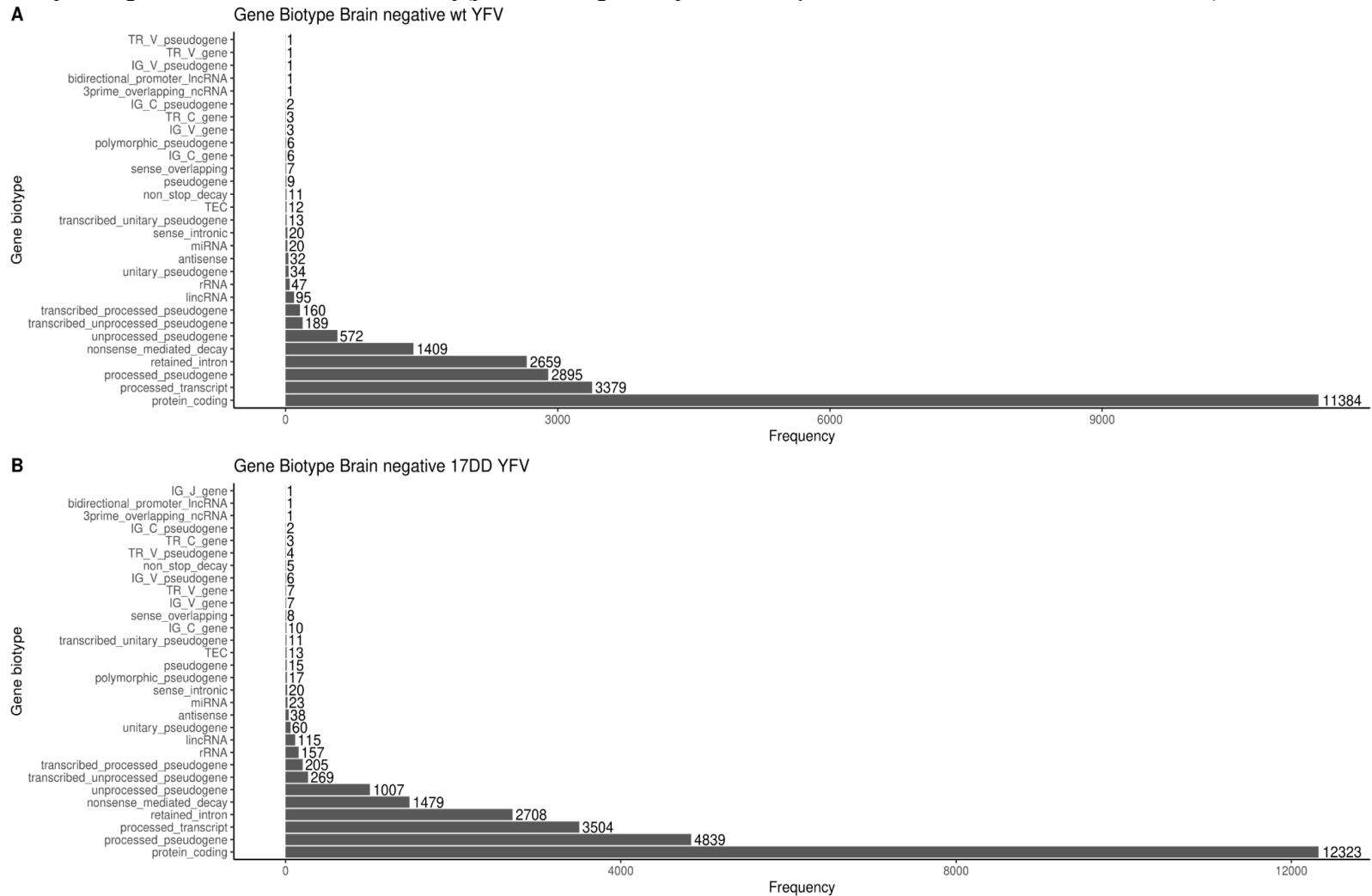
A terceira hipótese alternativa a ser testada refere-se à diferença de expressão gênica entre pacientes que não tiveram o vírus da febre amarela silvático e vacinal detectado no cérebro, porém eram positivos para estes vírus em outro órgão. Para esta análise, valores de expressão gênica do grupo 2 (YFV-silvático não detectado no cérebro) e grupo 4 (YFV-17DD não detectado no cérebro) foram confrontadas com o grupo 5 (controle).

#### **5.4.2.3.1. Expressão diferencial**

Foram quantificadas e classificadas a abundância de cada biotipo gênico para cada comparação, para fim de tomada de decisão sobre quais classes deveria ser incluída nas análises posteriores (Figura 46). Foram quantificados 11384 genes codificadores de proteínas na condição A (pacientes com YFV-silvático não detectado no cérebro), com 815 genes diferencialmente expressos significativos ( $p < 0,05$ ). Na condição B, 12323 genes codificadores de proteínas B foram quantificados (pacientes com YFV-17DD não detectado no cérebro), com 309 genes diferencialmente expressos significativos ( $p < 0,05$ ).



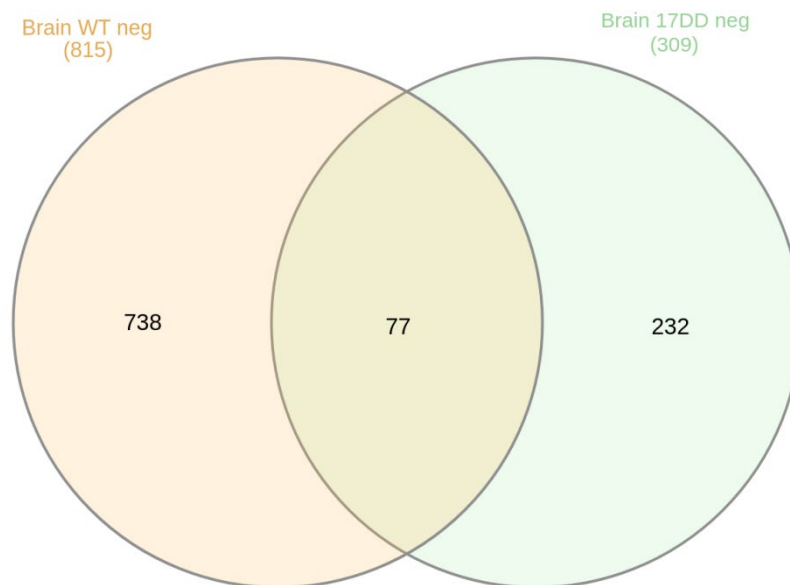
**Figura 46 - Grupos de genes caracterizados no RNAseq (pacientes negativos para a detecção molecular do YFV no tecido cerebral)**



Caracterização do perfil dos genes expressos considerando os grupos de pacientes negativos para a detecção molecular do YFV no tecido cerebral, da linhagem silvática e vacinal. (A) Condição A - pacientes sem detecção YFV-silvático no cérebro *versus* controle. (B) Condição B - pacientes sem detecção de YFV-17DD no cérebro *versus* controle.

Genes exclusivos de cada classe foram selecionados de forma que o enriquecimento gênico foi realizado apenas para este conjunto de genes. Na condição A, 738 genes são exclusivos, enquanto na condição B, 232 genes. Na intersecção foram encontrados 77 genes, pertencentes a ambas condições (Figura 47).

**Figura 47** – Comparação dos genes diferencialmente expressos (YFV não detectado no tecido cerebral)

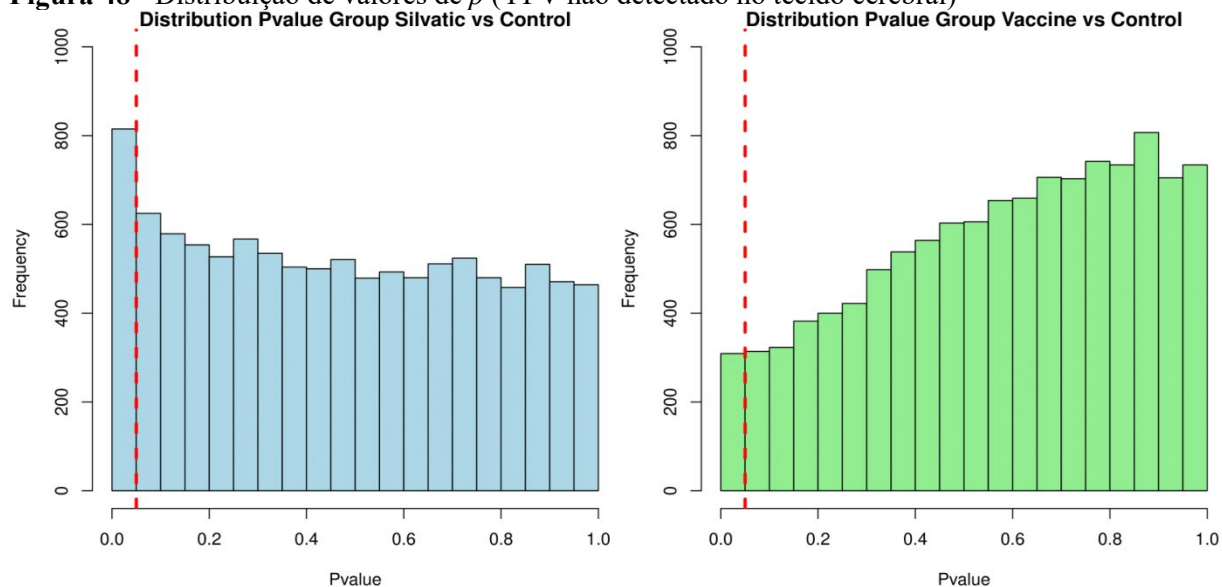


Comparação dos genes diferencialmente expressos considerando os grupos de pacientes com YFV-silvático (WT-neg) e vacinal (17DD-neg) não detectado no tecido cerebral. Em laranja, condição A - pacientes em que o YFV-silvático não foi detectado no cérebro *versus* controle. Em verde, condição B - pacientes em que o YFV vacinal não detectado no cérebro *versus* controle. E Intersecção de classes mostrando genes interclasses.

A distribuição de valores de  $p$  mostram o padrão de expressão de cada grupo estudado (Figura 48). Um padrão de distribuição de valor de  $p$  praticamente oposto é observado quando comparado às duas condições, mostrando desta forma um comportamento distinto de expressão gênica. Na condição A, em que o YFV-silvático não foi detectado no cérebro, porém foi detectado em pelo menos um outro órgão, é visto um padrão similar do observado na expressão de genes na condição em que esta mesma linhagem foi detectada no tecido cerebral. Uma curva em “L” indica maior frequência de genes com valores de  $p$  significativos em relação aos demais valores de  $p$ , por mais que exista uma grande proporção de genes que aceitem a hipótese nula ( $p < 0,05$ ).

Já na condição B a frequência de genes que aceitam a hipótese nula é muito maior, podendo inferir que mesmo quando o YFV não foi detectado no tecido cerebral, genes expressos nestas condições sugerem uma maior agressividade da linhagem silvática em relação à vacinal.

**Figura 48** - Distribuição de valores de  $p$  (YFV não detectado no tecido cerebral)



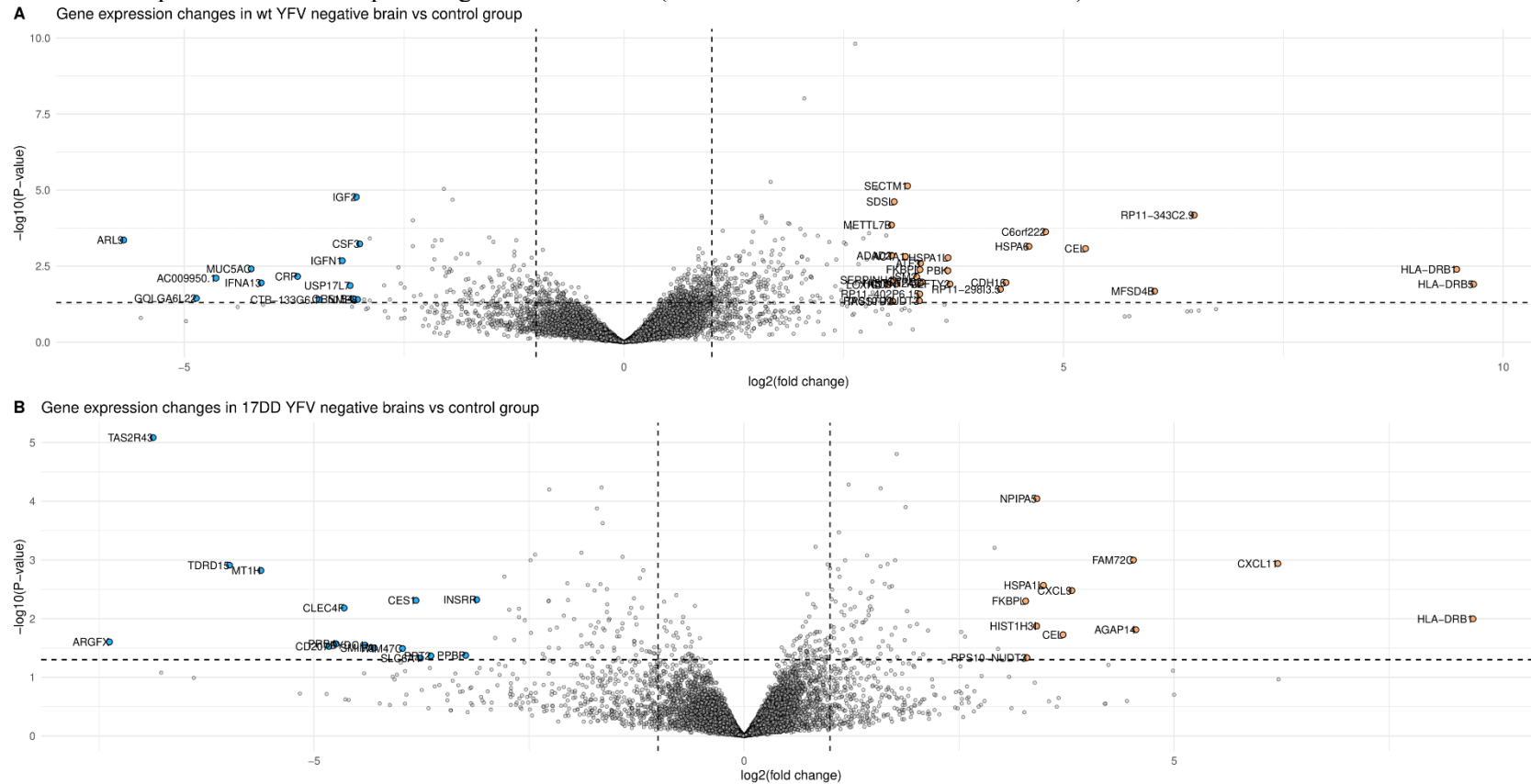
Distribuição de valores de  $p$  para condição (A) em que o YFV-silvático não foi detectado no cérebro, e grupo (B) em que o YFV-17DD não foi detectado no cérebro. A linha pontilhada vermelha indica o *threshold* dos genes considerados diferencialmente expressos com valor de  $p < 0,05$ .

A fim de hipotetizar quais genes estão envolvidos no bloqueio de entrada do vírus silvático e vacinal no tecido cerebral, foram comparadas as duas condições e também realizada a busca de assinaturas referente a cada linhagem. Nos *volcano plots*, foram destacados alguns genes com maior *FoldChange* (ou diferença de expressão) e com valor de  $p$  significativo (Figura 49A-B).

Na condição A, em que YFV-silvático não foi detectado no cérebro, alguns genes *down-regulated* mostraram-se relevantes como: *IGF2*, *CSF3*, *IGFN3*, *USPL17L7*, *CTB-133G*, *IFNA13*, *CRP*, *MUC5AC*, *AC009950.3*, *GOLGA6L22*, *ARL9*, porém não houve nenhum com posição de destaque. Em contrapartida, os genes *HLA-DRB1*, e *HLA-DRB5* foram superexpressos nesta condição cerca de 500 vezes mais do que no grupo controle.

Já na condição B, em que não houve a detecção do YFV-17DD no cérebro, tantos os genes *down-regulated*, quanto os superexpressos, tiveram uma variação de  $\log_2(1,5 - 3,5)$  FC entre os genes, com relação ao grupo controle, com destaque aos genes *TAS2R43*, *ARFX*, *TDRD15* e *MT1H* (*down-regulated*) e *FAM72O*, *AGAP14*, *CXCL11* e *HLA-DRB1* (*up-regulated*).

**Figura 49** - Volcano plot mostrando a expressão gênica diferencial (YFV não detectado no tecido cerebral)



Representação da expressão diferencial de genes em diferentes condições. (A) *Volcano plot* representando os genes expressos na condição A - pacientes sem detecção de YFV-silvático no cérebro em comparação ao grupo controle. (B) *Volcano plot* representando os genes expressos na condição B - pacientes sem detecção de YFV-17DD no cérebro em comparação ao grupo controle. O eixo y representa o valor negativo de log10 do p-value para cada gene [ $-\log_{10}(p\text{-values})$ ], e traçado um *threshold* onde o valor de p é 0,05, mostrando valores significativos acima do tracejado. No eixo x, foi representado o log2 do valor de *fold-change* (FC) para cada gene, indicando quais grupo de genes estão *up-regulated* ou *down-regulated* em comparação com o grupo controle. O FC é calculado pelo valor de expressão do gene na condição B / valor da expressão do gene na condição A. O *threshold* de 1 e -1 Log2FC foi estabelecido. Os genes representados em azul são os que satisfizeram a condição  $p < 0,05$  &  $FC \leq -3$  (*down-regulated*). Os genes que satisfizeram a condição  $p < 0,05$  &  $FC \geq 3$ , estão *up-regulated* e são representados em laranja.

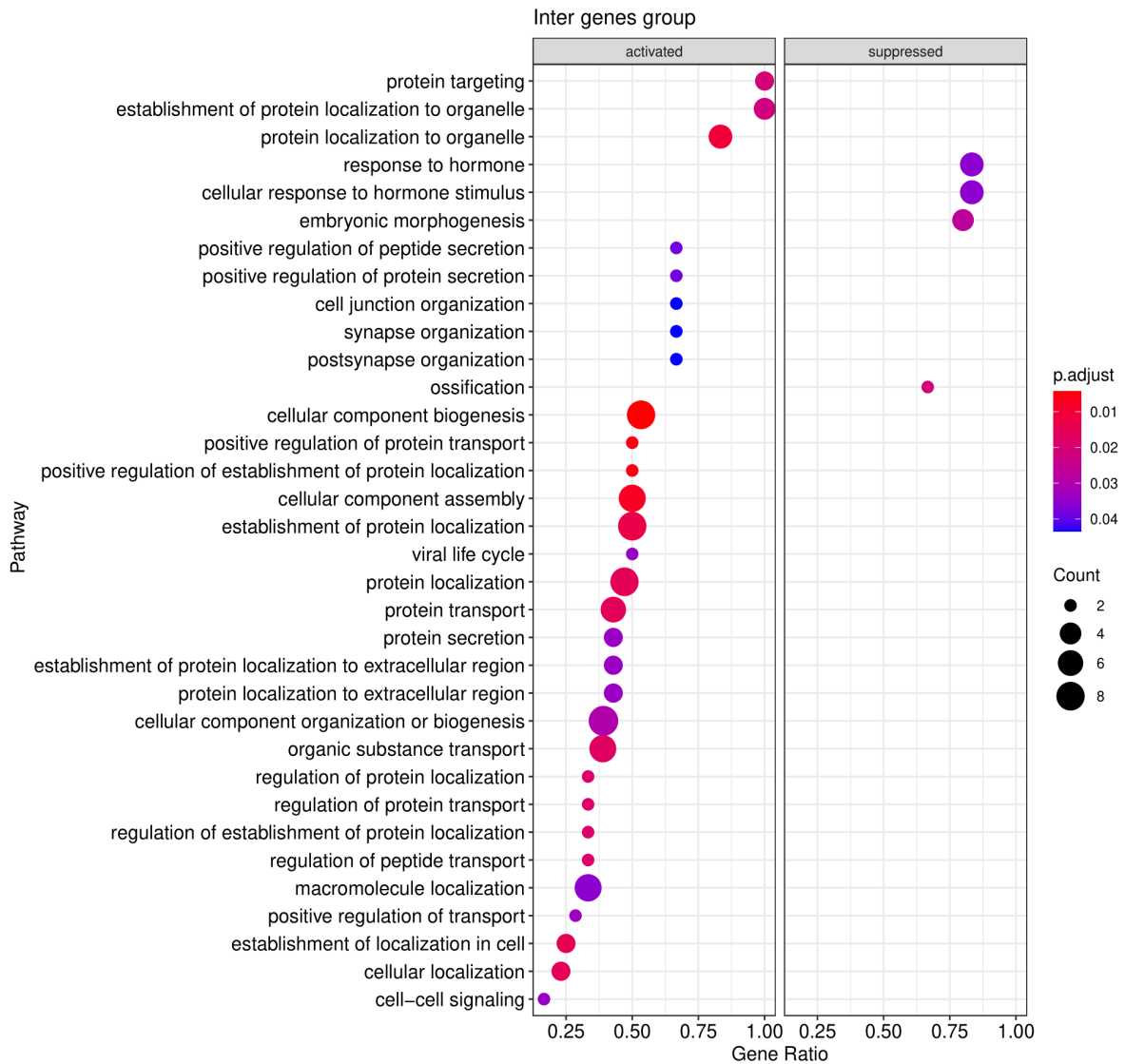
#### 5.4.2.3.2. Análise de ontologia

Para inferir quais seriam os processos comuns às duas linhagens de YFV que estariam relacionados ao possível bloqueio da entrada do vírus no cérebro, as principais vias (*pathways*) de processos biológicos referentes ao enriquecimento dos genes retirados da intersecção dos grupos, ou seja, aqueles genes compartilhados entre as duas condições estudadas, foram enriquecidos 77 genes pertencentes à classe interclasses. De forma geral, a grande maioria das vias que foram superexpressas estão relacionadas ao transporte e secreção de proteínas, além da organização de junções celulares, o que nos permite hipotetizar o papel destes processos biológicos na regulação da transferência de substratos para dentro e fora do SNC. As células endoteliais dos capilares sanguíneos do encéfalo estão unidas por complexos protéicos chamados “*tight junctions*” ou junções de oclusão, formando a barreira hematoencefálica, a qual bloqueia a passagem de substâncias hidrofílicas e grandes moléculas do sangue para o parênquima do SNC e *vice-versa*. Além das junções de oclusão, o revestimento dos capilares sanguíneos pelos pericitos e pés-astrocíticos também são importantes para controle de fluxo de substâncias pela barreira hematoencefálica (Figura 50).

A via mais relevante apresentada no gráfico de enriquecimento com valores de NES, foi a de biogênese de componentes celulares, que segundo a definição do gene ontology diz respeito à síntese de macromoléculas envolvidas em componentes celulares (Figura 51). Junto com outros achados relevantes, vias relacionadas à transporte e secreção de proteínas, além da biossíntese de macromoléculas (*e.g.*, proteínas), formam um arcabouço importante para a hipotetização do papel da barreira hematoencefálica no bloqueio do vírus da febre amarela.

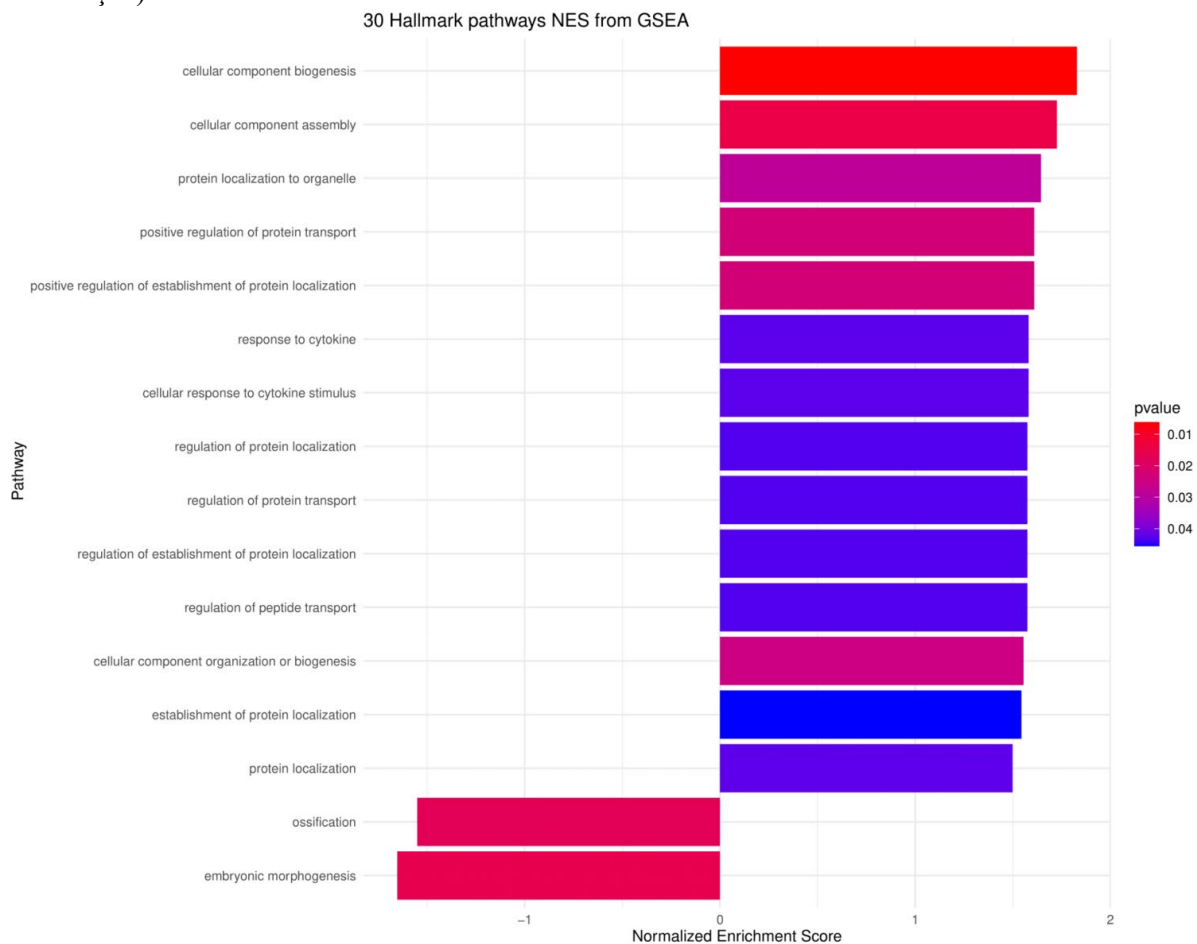
Tanto no mapa de genes (Figura 52), quanto no *heatmap-like* (Figura 53) o gene *HLA-DRBI*, que desempenha um papel crítico no sistema imunológico, mostrou-se como central às principais vias correspondentes a processos biológicos vinculados a condição de possível bloqueio da entrada do YFV no cérebro. Além disso ele é superexpresso nessa condição cerca de  $\log_2(10)$  vezes mais que no controle. O gene *HSPAIL*, que codifica uma proteína da família *heat shock* e é expressa perante a situação de estresse, também é central entre vias.

**Figura 50** - Diferentes vias enriquecidas dos genes localizados na intersecção os pacientes em que o YFV-silvático não foi detectado no cérebro e para os pacientes em que o YFV-17DD não foi detectado no cérebro



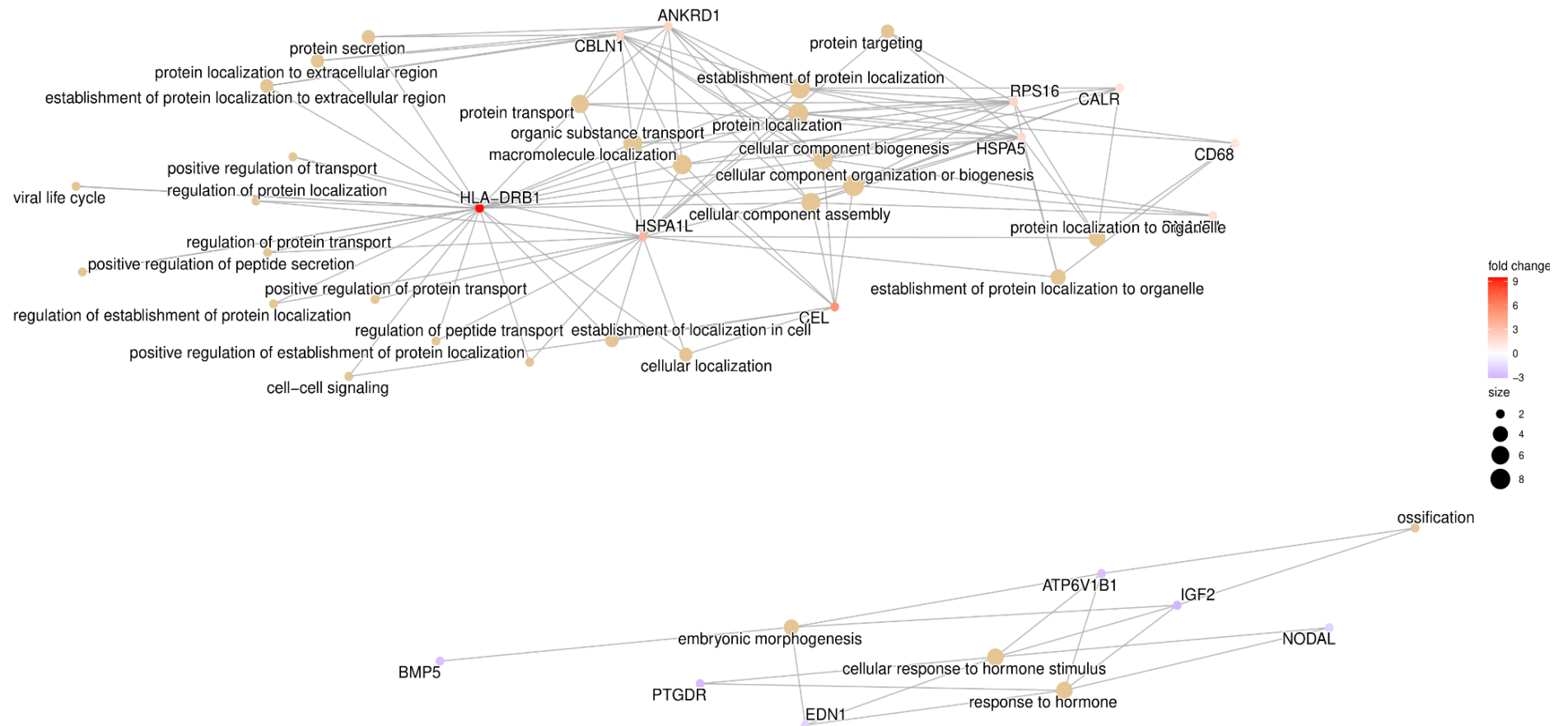
Diferentes vias enriquecidas dos genes localizados na intersecção entre grupos A - pacientes em que o YFV-silvático não foi detectado no cérebro *versus* controle e grupo B - pacientes em que o YFV-17DD não foi detectado no cérebro *versus* controle. No eixo x, o “gene ratio” representa a porcentagem do total de DEGs na via dada. O gráfico visualiza valores em duas dimensões: cor e tamanho. A cor, representa a significância de valor de  $p$  ajustado pelo método de correção de Bonferroni ( $p$  valores brutos\* número de comparações). Já o tamanho, representa a quantidade de genes dentro de cada de via. Ao lado esquerdo do gráfico estão as vias superexpressas nesta condição, enquanto ao lado direito as vias *down-regulated*.

**Figura 51** - Principais vias significativas enriquecidas pelo GSEA (utilizando os genes presentes na intersecção)



Principais vias significativas enriquecidas pelo GSEA, ( $p < 0,05$ ), e seu valor respectivo valor de NES. Os genes que foram enriquecidos estão presentes na intersecção entre grupos A - pacientes em que o YFV-silvático não foi detectado no cérebro *versus* controle e grupo B - pacientes em que o YFV-17DD não foi detectado no cérebro *versus* controle. As principais vias enriquecidas pelo GSEA, que foram significativas ( $p < 0,05$ ), e seu valor respectivo valor de NES, foram apresentadas. Um valor de NES significativo positivo, indica que os membros do conjunto de genes tendem a aparecer no topo do transcriptoma e estão superexpressos em relação à condição controle. Os valores  $p$  são destacados pela variação de cores.

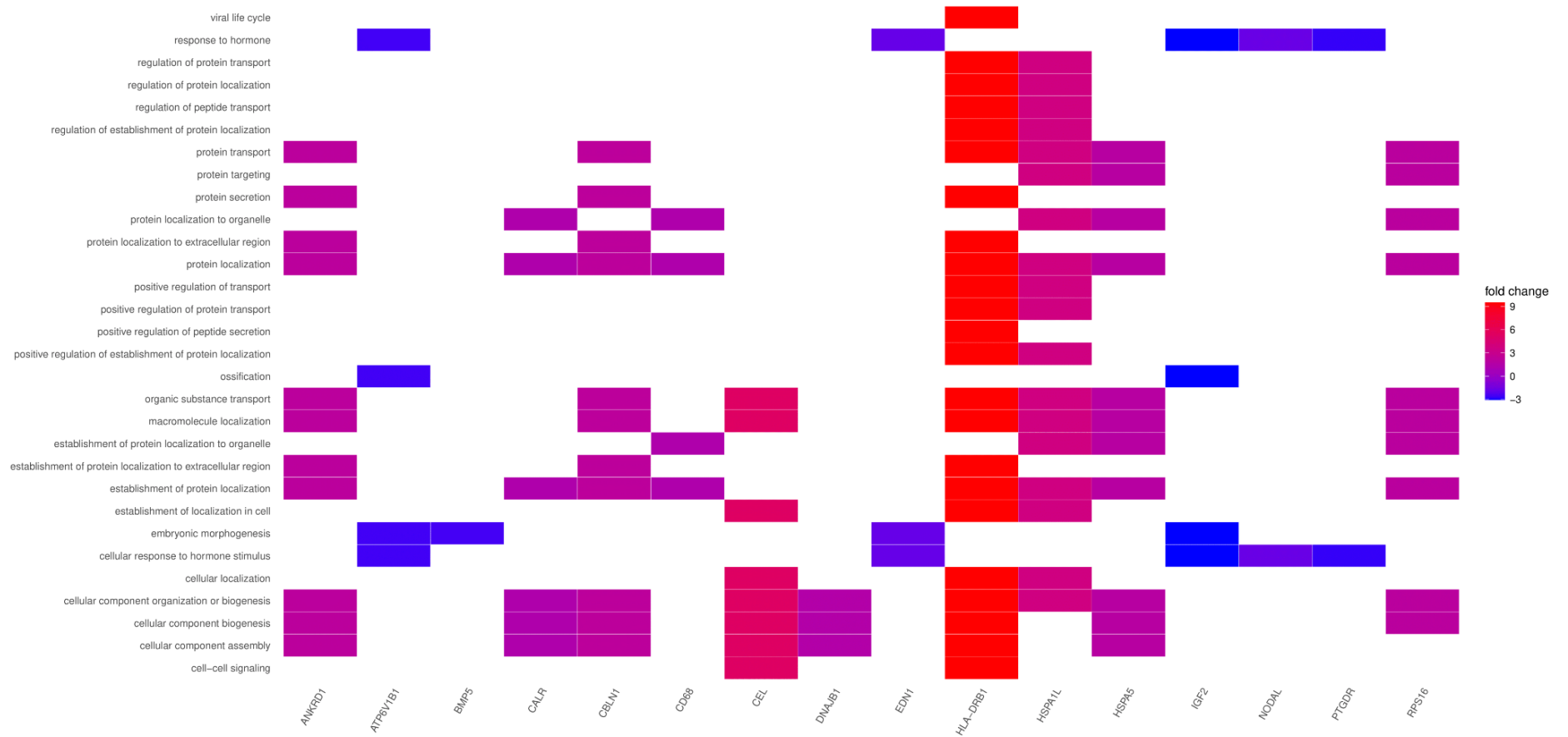
**Figura 52** - Rede Gene-concept da condição em que foram analisados os genes presentes na intersecção entre os pacientes em que o YFV-silvático não foi detectado no cérebro e para os pacientes em que o YFV-17DD não foi detectado no cérebro



Rede *Gene-concept* da condição em que foram analisados os genes presentes na intersecção entre grupos A - pacientes em que o YFV-silvático não foi detectado no cérebro *versus* controle e grupo B - pacientes em que o YFV vacinal não foi detectado no cérebro *versus* controle. Genes que são associados aos 20 termos principais ranqueados pelo GSEA, foram destacados na figura. Os nós representam os termos ontológicos, e a distância entre eles está diretamente relacionada à proporção de sobreposição de genes entre os termos. O tamanho do nó está relacionado ao número de genes associados a cada termo e as cores destacam os valores de FC.



**Figura 53** - Classificação funcional dos genes presentes na intersecção entre os pacientes em que o YFV-silvático não foi detectado no cérebro e para pacientes em que o YFV-17DD não foi detectado no cérebro



No eixo *y*, estão as principais vias enriquecidas relativas aos genes interclasses. No eixo *x*, os genes que fazem parte dessas vias, mostrando um padrão de expressão gênica característico do possível bloqueio do YFV em ambas linhagens virais de febre amarela. A variação de cor indica variação do FC.

Os *pathways* enriquecidos de genes pertencentes a classes em que o YFV-silvático e vacinal não foram detectado no cérebro, porém foram detectados em pelo menos um órgão (Figura 54A-B) de maneira a elucidar se há mecanismos envolvidos no bloqueio da passagem do YFV para o sistema nervoso exclusivo de cada linhagem viral.

Na condição A, em que o vírus silvático não foi detectado no tecido cerebral, as principais vias enriquecidas mostram que mesmo com a ausência do vírus genes de resposta a estímulos externos são ativados, além também de mecanismos de defesa e resposta a estresse, possivelmente indicando inflamação no cérebro. Em contrapartida, na condição B, em que o YFV-17DD não foi detectado no cérebro, ressalta-se vias de transporte e regulação de cálcio intracelular, o qual é importante para a integridade das “*tight junctions*” que formam a barreira hematoencefálica e bloqueiam passagem de substâncias e agentes infecciosos para dentro do SNC.

O gráfico com valores de NES, mostram que na condição A, as vias mais significativas são aquelas vinculadas à diferenciação de células musculares lisas vasculares, componentes críticos da unidade neurovascular e importantes para regulação do fluxo sanguíneo cerebral e integridade da barreira hematoencefálica. Na condição B, a via com maior NES foi a de resposta imunológica inata. Contudo a presença da maioria das vias voltadas a transporte e regulação do cálcio continua intrigante (Figura 55).

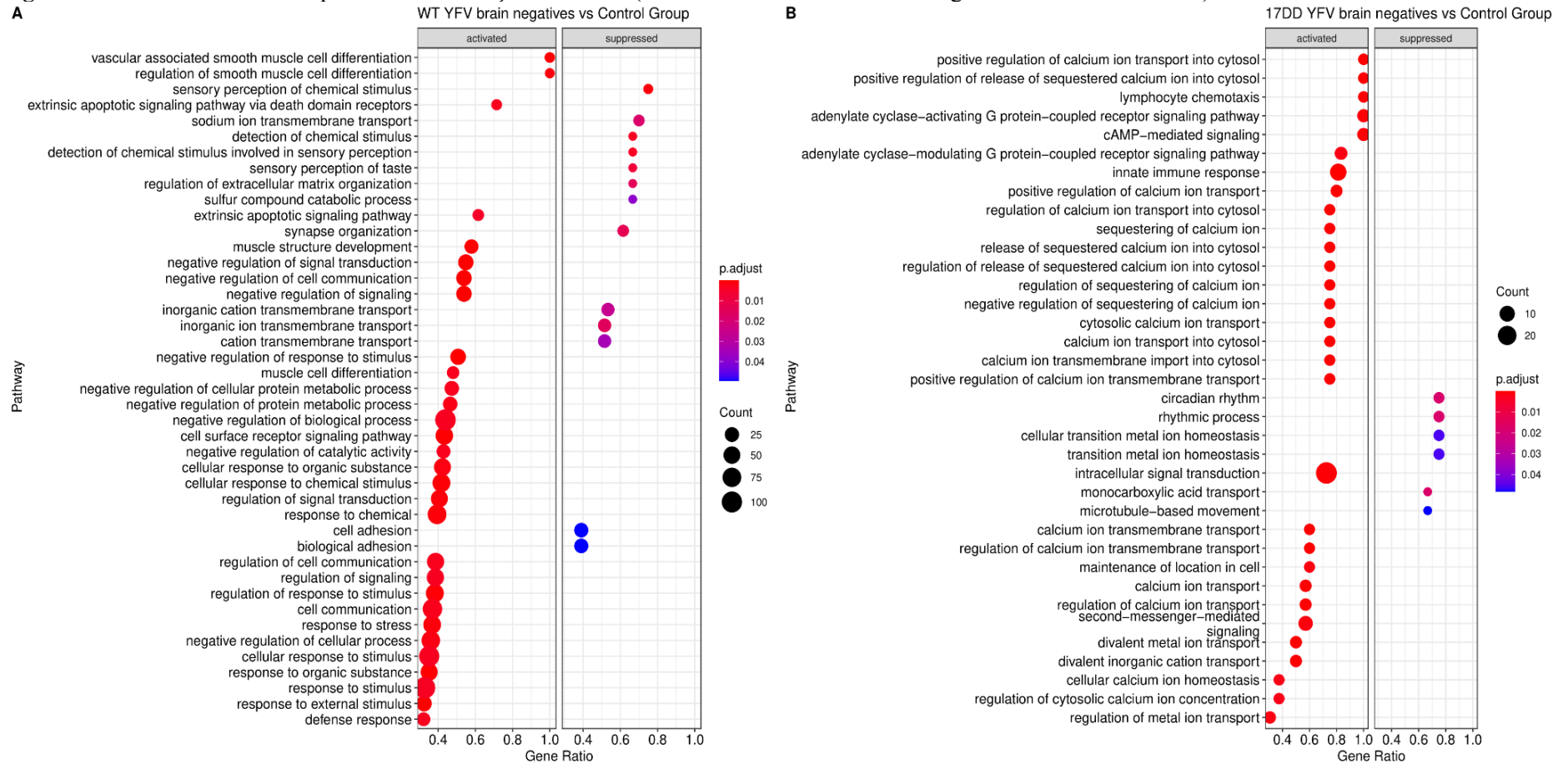
Ao estudar o mapa de genes (Figura 56), junto com o *heatmap-like* (Figura 57), muitas vias se mostram centrais ao perfil de expressão gênica da condição A, principalmente aquelas relacionadas à resposta a estímulos. Contudo, nenhum gene se relaciona de maneira central às essas principais vias. As figuras ressaltam a superexpressão do gene *HLA-DRB5*, porém ele não é chave na definição do perfil gênico desta condição.

Diferente do observado na condição A, a análise do mapa de genes (Figura 59) e do *heatmap-like* (Figura 60) dos genes enriquecidos referente a condição B, em que o YFV-17DD não foi detectado no cérebro, mostrou uma via principal de destaque com 3 principais genes associados a elas. A via de transdução de sinal intracelular pode estar vinculada a sinalização de cálcio, o qual é utilizado para conduzir processos intracelulares. E os principais genes são o *CXCL9*, *CXCL10*, *CXCL11*, quimiocinas que desempenham papel crítico no controle de entrada de leucócitos no cérebro, perante neuroinflamação.

O mapa de enriquecimento da condição A, em que o YFV-silvático não foi detectado no cérebro mostra a formação de 3 clusters (Figura 58), sendo o principal deles, ou seja, aquele com maior quantidade de vias incluídas e genes sobrepostos, dividido em dois sub-clusters ligados por via de genes expressos em células musculares em desenvolvimento. A

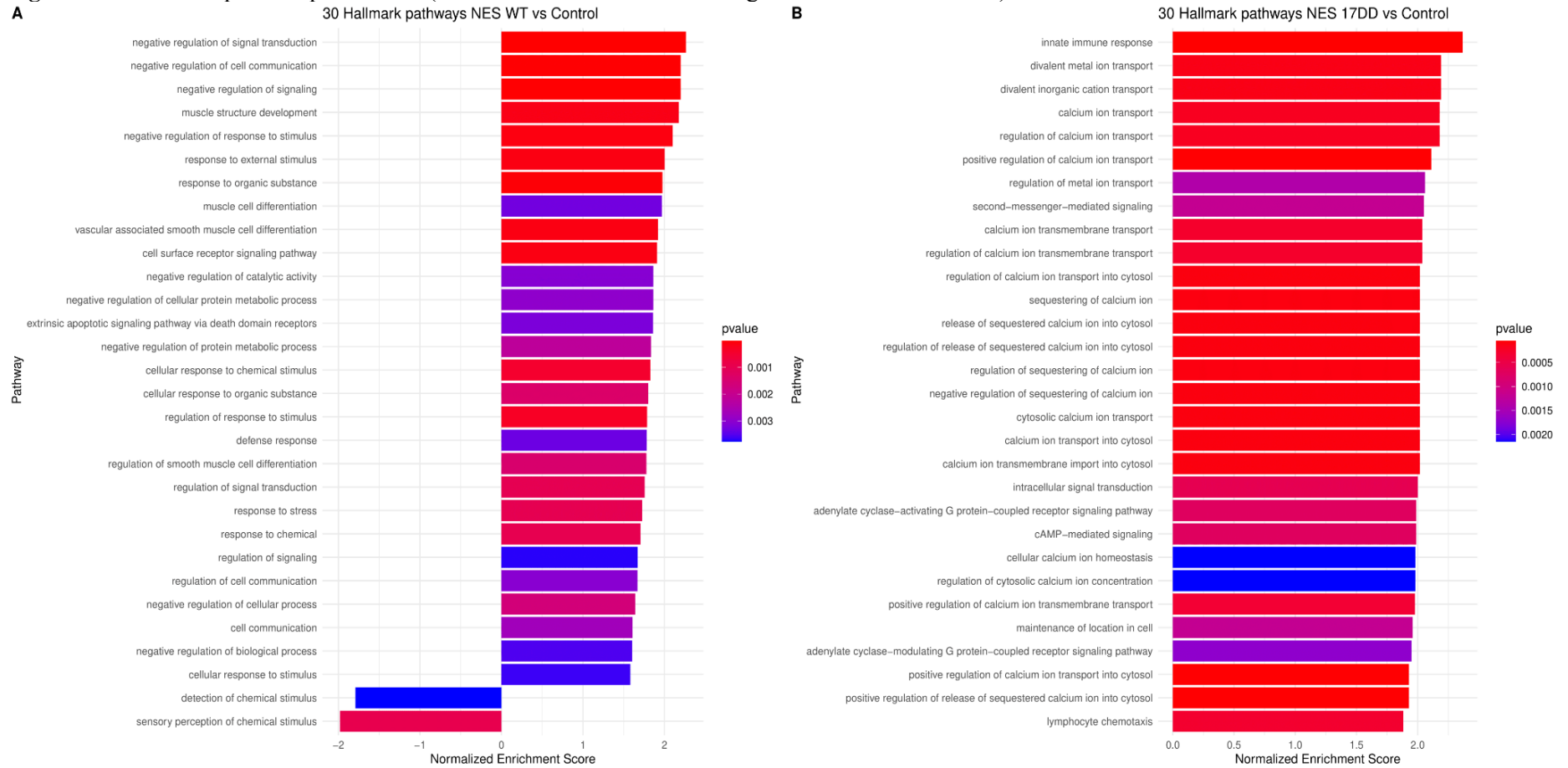
principal relação entre genes é visto entre vias de resposta à sinalização celular perante estresse celular devido a possível inflamação no tecido cerebral. Os outros dois clusters menores estão relacionados a quimiotaxia, migração celular, vias apoptóticas e percepção sensorial (Figura 58). No mapa de enriquecimento da rede de similaridade ponderada relacionada a genes diferencialmente expressos na condição B, em que o YFV-17DD não foi detectado no cérebro foi possível observar a formação de um único grande cluster, dividido em dois subclusters (Figura 61). O menor, é relacionado a resposta imunológica e genes que expressam interferon-gama. Já o principal deles, com maior sobreposição de genes diferencialmente expressos nesta condição, tem como principal função biológica atribuída a manutenção da homeostase celular por íons de cálcio.

**Figura 54 - Diferentes vias enriquecidas em condições distintas (YFV-silvático e YFV-17DD negativos no tecido cerebral)**



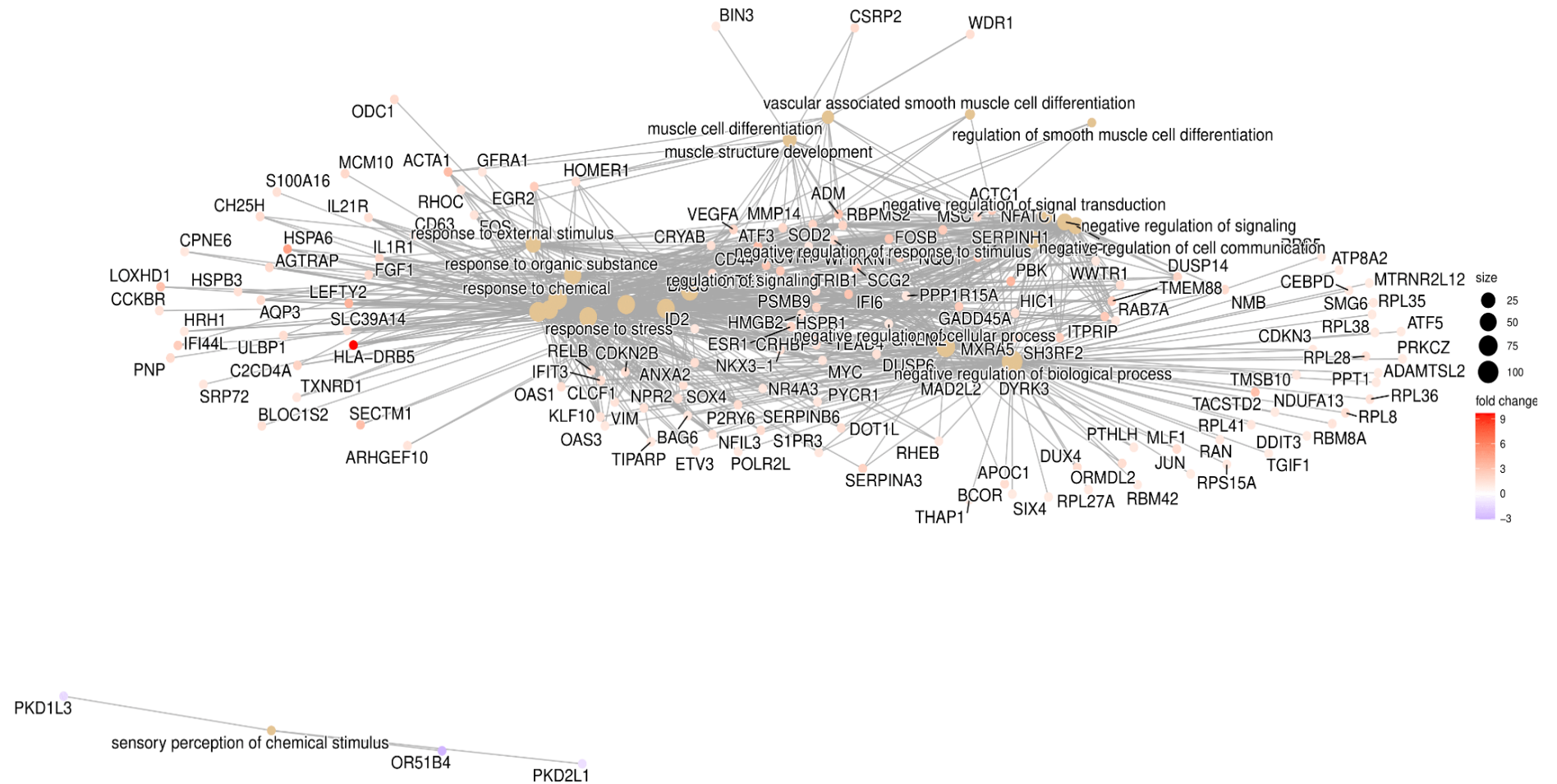
No eixo x, o “gene ratio” representa a porcentagem do total de DEGs na via dada. O gráfico apresenta valores em duas dimensões: cor e tamanho. A cor, representa a significância de valor de  $p$  ajustado pelo método de correção de Bonferroni ( $p$ -valores brutos\* número de comparações). Já o tamanho, representa a quantidade de genes dentro de cada de via. (A) Vias super expressas e *down-regulated* na condição A - pacientes sem detecção do YFV-silvático no cérebro *versus* controle. (B) Vias superexpressas e *down-regulated* na condição B - pacientes sem detecção do YFV-17DD no cérebro *versus* controle.

**Figura 55 - Vias enriquecidas pelo GSEA (YFV-silvático e YFV-17DD negativos no tecido cerebral)**



Vias enriquecidas pelo GSEA, que foram significativas ( $p < 0,05$ ), e seu valor respectivo valor de NES. Um valor de NES significativo positivo, indica que os membros do conjunto de genes tendem a aparecer no topo do transcriptoma e um NES significativo negativo, indica o oposto. Os valores de  $p$  são destacados pela variação de cores. (A) Principais vias em destaque da condição A, em que YFV-silvático (YFV-WT) não foi detectado no cérebro. (B) Principais vias em destaque da condição B, em que YFV-17DD (YFV-17DD) não foi detectado no cérebro.

**Figura 56** - Rede *Gene-concept* da condição em que foram analisados os genes diferencialmente expressos na condição A, em que o YFV-silvático não foi detectado no cérebro



Genes que são associados aos 20 termos principais ranqueados pelo GSEA, foram destacados na figura. Os nós representam os termos ontológicos, e a distância entre eles está diretamente relacionada à proporção de sobreposição de genes entre os termos. O tamanho do nó está relacionado ao número de genes associados a cada termo e as cores destacam os valores de FC.

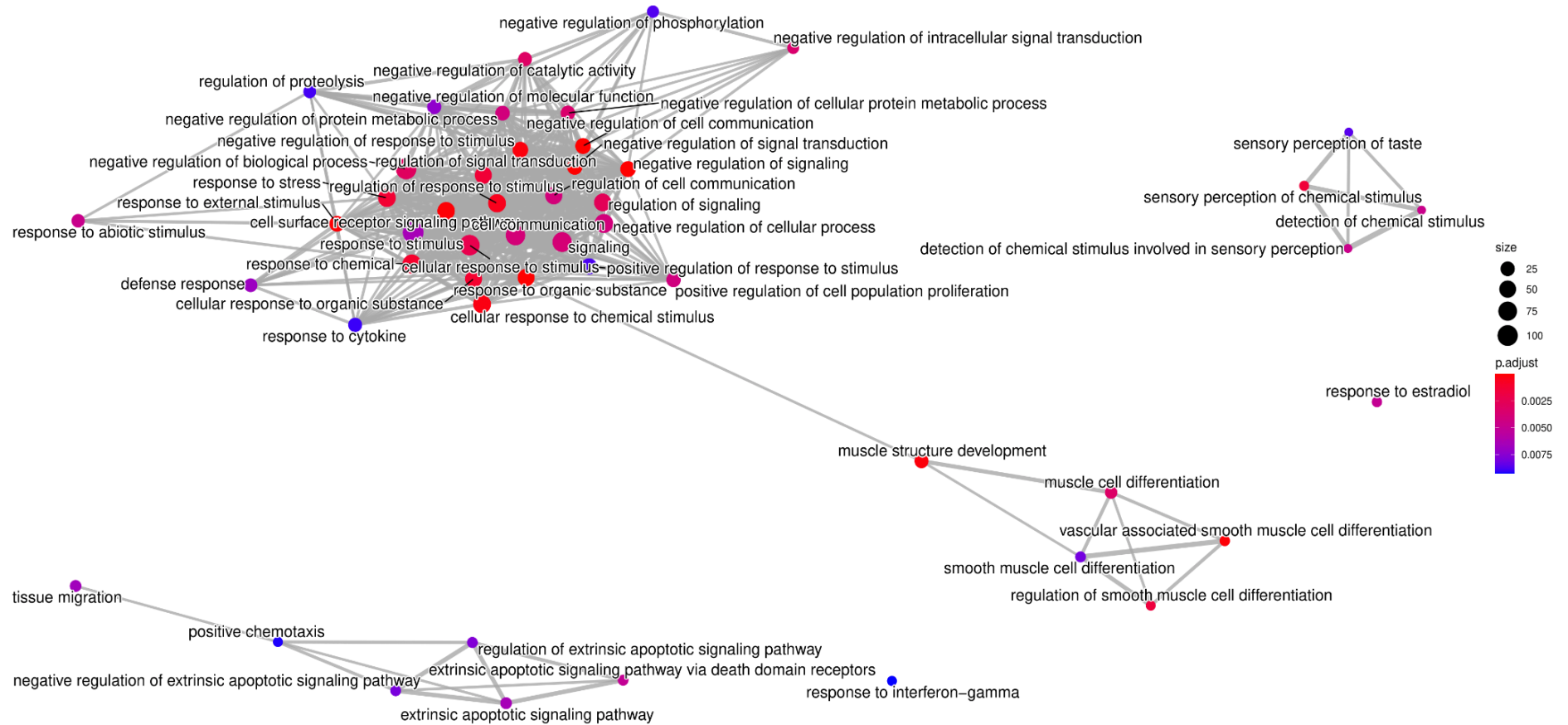


**Figura 57 - Classificação funcional dos genes diferencialmente expressos na condição A, em que o YFV-silvático não foi detectado no cérebro**



No eixo *y*, estão as principais vias enriquecidas relativas aos genes interclasses. No eixo *x*, os genes que fazem parte dessas vias, mostrando um padrão de expressão gênica característico comum nas linhagens virais de febre amarela. A variação de cor indica variação do FC.

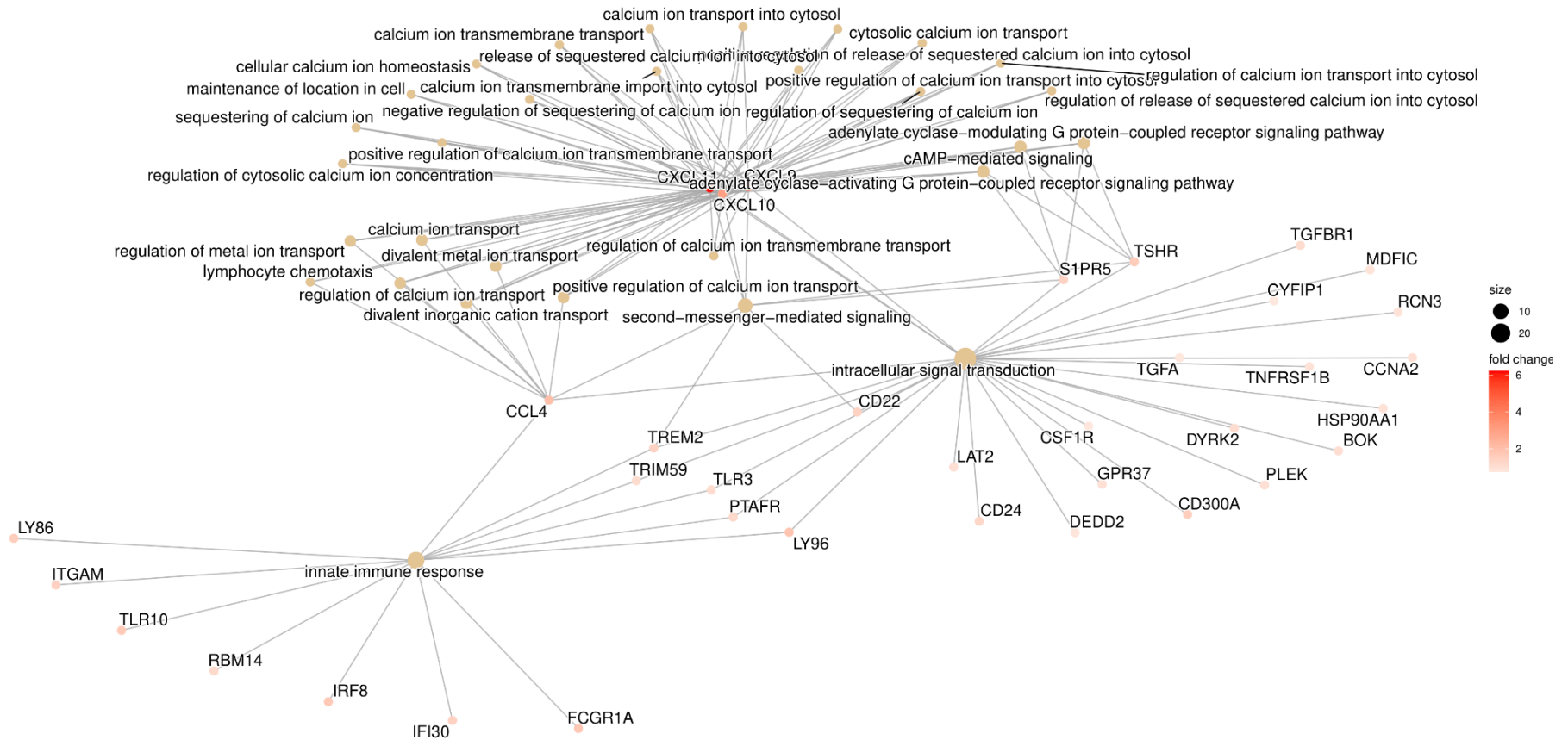
**Figura 58** - Rede de Similaridade ponderada atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição A, em que o YFV-silvático não foi detectado no cérebro



Rede de Similaridade ponderada, com pesos (grau de sobreposição de genes, dentro dos termos ontológicos) atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição A, em que o YFV-silvático não foi detectado no cérebro. Os nós representam um conjunto de genes, com seu respectivo termo de processo biológico. O tamanho do nó representa o número de genes no conjunto de genes, já a espessura da borda é proporcional ao grau de sobreposição entre os conjuntos de genes, calculada pelo coeficiente de Jaccard. As distâncias de ligações entre nós, ou seja, as arestas, representam o grau de sobreposição entre o conjunto de genes definida pela coeficiente de Szymkiewicz-Simpson. A variação da cor do nó está relacionada à variação do valor de *p*-ajustado.

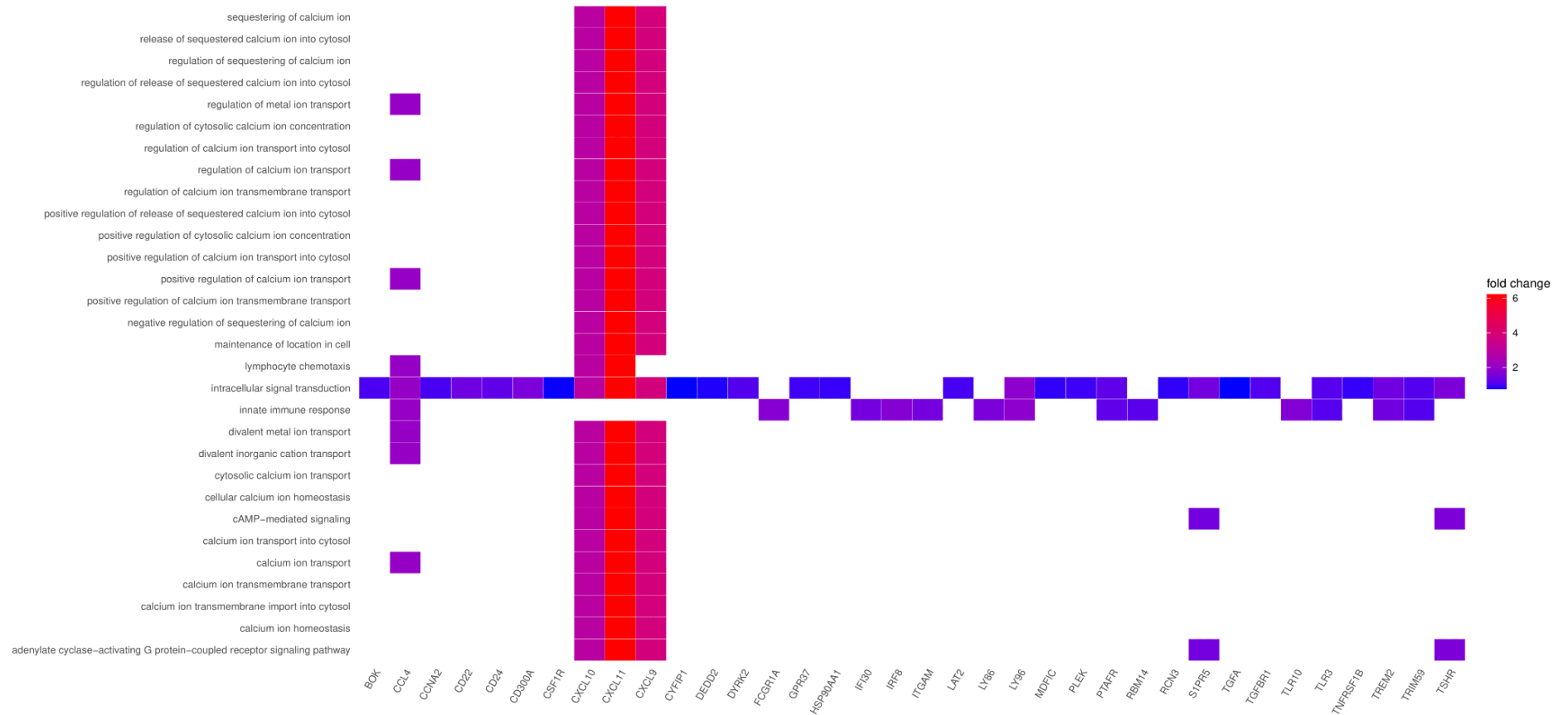


**Figura 59 - Rede Gene-concept da condição em que o YFV-17DD não foi detectado no cérebro**



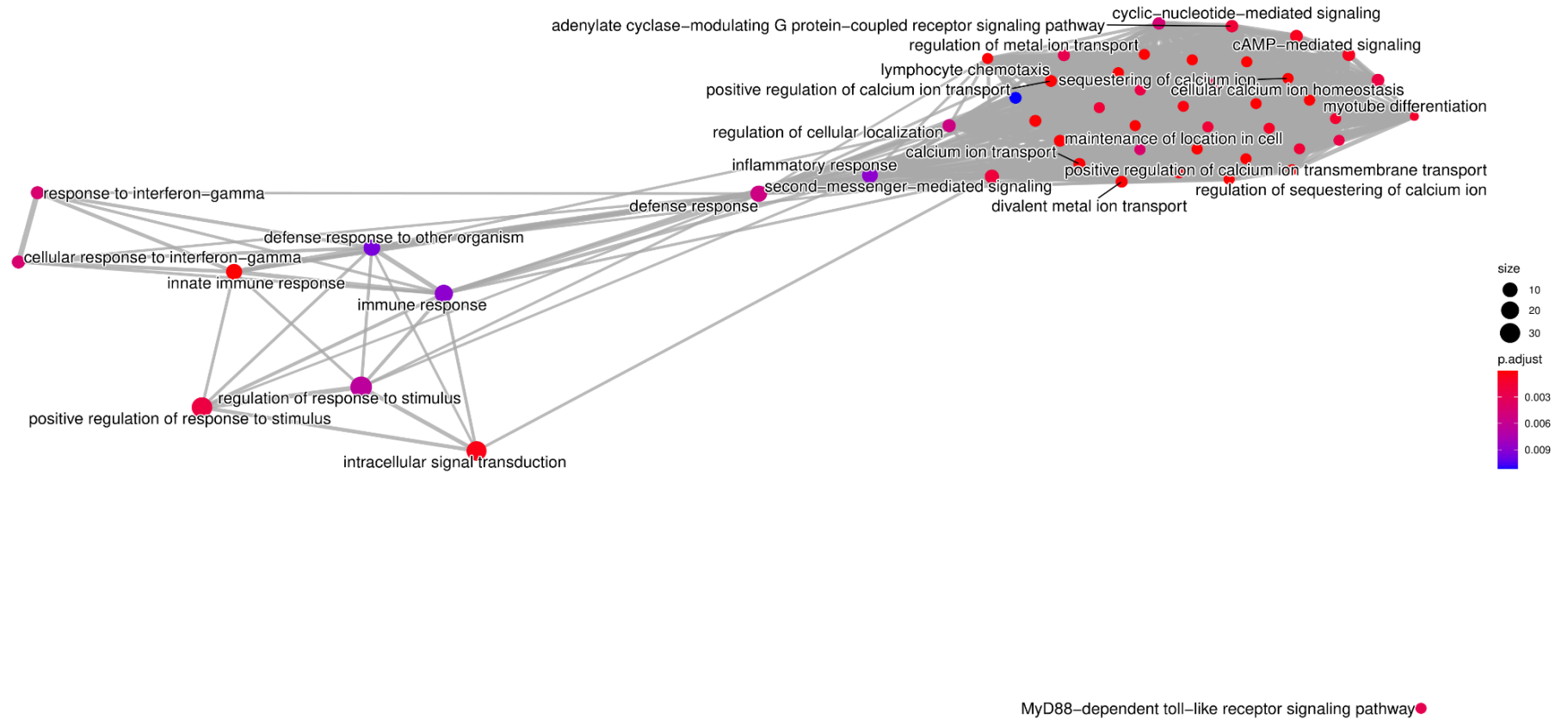
Genes que são associados aos 20 termos principais ranqueados pelo GSEA, foram destacados na figura. Os nós representam os termos ontológicos, e a distância entre eles está diretamente relacionada à proporção de sobreposição de genes entre os termos. O tamanho do nó está relacionado ao número de genes associados a cada termo e as cores destacam os valores de FC.

**Figura 60 - Classificação funcional dos genes diferencialmente expressos na condição B, em que o YFV-17DD não foi detectado no cérebro**



No eixo y, estão as principais vias enriquecidas relativas aos genes interclasses. No eixo x, os genes que fazem parte dessas vias, mostrando um padrão de expressão gênica característico do que é comum nas linhagens virais de febre amarela. A variação de cor indica variação do FC.

**Figura 61** - Rede de Similaridade ponderada, com pesos atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição em que o YFV-17DD não foi detectado no cérebro



Rede de Similaridade ponderada, com pesos (grau de sobreposição de genes, dentro dos termos ontológicos) atribuídos durante o processo de agrupamento de enriquecimento de genes relacionados à expressão diferencial da condição B, em que o YFV-17DD não foi detectado no cérebro. Os nós representam um conjunto de genes, com seu respectivo termo de processo biológico. O tamanho do nó representa o número de genes no conjunto de genes, já a espessura da aresta é proporcional ao grau de sobreposição entre os conjuntos de genes, calculada pelo coeficiente de Jaccard. As distâncias de ligações entre nós, ou seja, as arestas, representam o grau de sobreposição entre o conjunto de genes definida pelo coeficiente de Szymkiewicz-Simpson. A variação da cor do nó está relacionada à variação do valor de *p*-ajustado.

## *Capítulo 6*

## 6. DISCUSSÃO

O gênero *Flavivirus* compreende um grupo de vírus de RNA, transmitidos por vetores artrópodes, que causam um amplo espectro de doenças potencialmente graves, que incluem as febres hemorrágicas (125), choque (126), encefalites (127), anormalidades congênitas (128), que podem culminar em casos fatais (3,129). Este amplo espectro clínico para apresentação da doença que podem ser causadas por diferentes *Flavivirus*, refletem principalmente o potencial tropismo celular característico de cada espécie viral, além da capacidade de antagonizar a imunidade do hospedeiro (130). Quando infectados por *Flavivirus*, os hospedeiros vertebrados, como os humanos podem enquadrar-se em duas grandes categorias de doenças graves, que incluem as viscerotrópicas e as neurotrópicas (131). Este estudo teve como objetivo investigar a capacidade de neuroinvasão do SNC pelo YFV-silvático e YFV-17DD (vacinal), buscando elucidar a diferença de expressão gênica quando comparados grupos em que estas linhagens virais foram detectadas no cérebro de pacientes que apresentaram como desfecho o óbito com suspeita associação ao YFV, e também, comparando o grupo dos pacientes que apresentaram neuroinvasão ao grupo em que não houve neuroinvasão.

Para realização do estudo, tecido cerebral *post-mortem* (lobo frontal) foi coletado de treze pacientes, sendo três deles controles (onde o óbito não foi associado à doenças infecciosas). Estudos de expressão gênica em tecidos *post-mortem* são ferramentas muito úteis para entender a patogênese de doenças infecciosas neurotrópicas, contudo a obtenção de RNA de qualidade é um grande desafio técnico que pode afetar significativamente a qualidade dos dados gerados para análises de expressão gênica utilizando transcriptoma. Em materiais cerebrais *post-mortem*, geralmente é assumido que a determinação de RIN (*RNA integrity number*), é um indicador para a caracterizar a qualidade/integridade dos RNAs presentes na amostra. Contudo, diversos estudos sugerem que o RIN não reflete necessariamente a disponibilidade de mRNAs intactos e que esse parâmetro é considerado potencialmente enganoso, principalmente quando usado como uma medida da qualidade geral do tecido cerebral *post-mortem* (132,133). Segundo Sonntag *et al.* 2016 (133), ao analisar o tecido cerebral humano *post-mortem*, o RIN reflete uma diminuição na quantidade total de mRNA, em vez da integridade, sugerindo que é uma medida incompleta da qualidade geral dos RNAs presentes no cérebro (133). Em contraste com dados publicados, nos quais a degradação do RNA era controlada em condições de laboratório (134), a degradação do RNA no tecido cerebral *post-mortem* não é controlada, o que pode ter sido um fator contribuinte para a variabilidade de RIN observada no estudo.

A reprodução dos dados ômicos, com multivariáveis (como por exemplo, utilizando os genes) em forma de PCA, fornece informações sobre a estrutura geral do conjunto de dados analisados e são usados para entender padrões de agrupamento entre amostras. Mais do que clusterização por linhagem viral, a clusterização por positividade no tecido cerebral foi também observada (Figura 15). Quando observado o PC1 (com 25% de variância) *versus* PC2 (com 18% de variância), mesmo não havendo clusterização, observa-se que a amostra C49 (YFV com vírus silvático detectado em pelo menos um órgão, porém negativo no tecido cerebral) está como *outlier*. Devido à falência hepática de todos os pacientes incluídos no grupo com YFV detectado em pelo menos um órgão, o cérebro passa a responder alterando a expressão de genes para proteínas que desempenham papel fundamental para funcionamento e homeostase do SNC. Contudo, o paciente C49 foi o único a ter uma grave herniação cerebral proveniente de um edema causado pela encefalopatia hepática devido a complicações da insuficiência hepática aguda. O aumento da pressão intracraniana se deve ao edema cerebral citotóxico (inchaço astrocítico) causado pelo acúmulo de glutamina, a qual não é eficientemente convertida a partir da alta concentração de amônia cerebral. A catálise da reação que converte amônia em glutamina é feita pela enzima glutamina sintetase, primordialmente expressa nos astrócitos (135). Assim, essa diferença de expressão justifica-se por genes alterados vinculados à herniação, ou então, por ser mais um dos quadros atípicos para desfecho associado à YF.

Para a investigação de doença associada ao YFV-17DD, amostras coletadas foram caracterizadas por histopatologia (infiltrados virais condizentes com doença por YF) e por técnica molecular (detecção do YFV-17DD por RT-qPCR), pois ambos os critérios devem ser preenchidos para conclusão de caso. O paciente C37, um paciente de 19 anos, sem comorbidades, foi vacinado e teve óbito associado a uma apendicite aguda perfurada dois dias após ser vacinado. O RNA viral correspondente ao YFV-17DD foi identificado por RT-qPCR em 5 dos 7 órgãos estudados, sendo eles: coração (Ct = 37,25), pulmão (Ct = 35,88), rim (Ct = 35,65), baço (Ct = 26,66), fígado (Ct = 22,78). Embora nenhum achado histopatológico tenha sido encontrado, classificando-o, portanto, como sem critérios para doença vacinal, o vírus vacinal foi detectado em 5 órgãos. No entanto, YFV não foi detectado no tecido cerebral, o que torna este paciente, importante para estudar a acessibilidade ao SNC, neurotropismo e neuroinvasividade. Além disso, muitos estudos atuais (136–139) mostram uma prevalência marcante de casos de apendicite aguda após a vacinação (140,141), embora nenhum estudo tenha sido relatado com a febre amarela.

*Max Theiler*, ganhou o prêmio Nobel de Medicina e Fisiologia em 1951 pelo desenvolvimento da vacina anti-YF (vacina de vírus vivo atenuado) (68). Porém, ainda em

1937, ele mostrou a existência de um neurotropismo significativo do YFV em camundongos e macacos, o que constituiu um desafio para o desenvolvimento de plataformas vacinais para humanos (142). Desde então, muitos estudos mostraram o efeito da neurovirulência do YFV-17D, associados a efeitos adversos vacinais (59,143–145). Contudo, nenhum estudo foi publicado ainda sobre o efeito da linhagem silvática no cérebro de pacientes infectados com YFV, e também sobre quais fatores estão relacionados ao neurotropismo, neuroinfectividade e neurovirulência. Portanto, o ineditismo deste nosso estudo, é de suma importância para a descrição das primeiras informações acerca da neuropatogenicidade na YF, não somente da linhagem silvática, mas também como um padrão novo de reação adversa vacinal ao vírus YFV-17DD. Decifrar o papel de diferentes genes neste contexto, permite postular, a partir de vias que foram enriquecidas, assinaturas preditoras de neuroinvasão e de respostas compartilhadas e diferenciais observadas para as duas linhagens do YFV.

Visto que este estudo foi realizado observando um *status* pontual da doença, em que indivíduos com uma infecção sistêmica por YF tiveram o SNC atingido ou não pelo vírus no momento da sua morte, foi possível investigar cenários da patogenicidade do YF, porém de forma muito pontual em relação à linha do tempo da infecção. Inicialmente, com a intenção de observar apenas a influência da detecção ou não do YFV (independente da linhagem), em tecido cerebral, os grupos 1 U 3 (com detecção do YFV no cérebro) e os grupos 2 U 4 (onde YFV não foi detectado no cérebro) tiveram os valores de expressão gênica comparados ao grupo controle (grupo 5), aqui considerados como um grupo de pacientes contendo expressão basal em condições fisiológicas, e assim, foi obtido os valores de expressão gênica diferencial. Uma quantidade similar de genes diferencialmente expressos em ambas as condições mostram que a presença causada pelo YFV no tecido cerebral não está associada a grandes diferenças numéricas em nível de expressão gênica, quando comparado à ausência do YFV no cérebro.

Durante a coevolução vírus-hospedeiro, os *Flavivirus* desenvolveram estratégias notáveis para a sua manutenção e sobrevivência dentro da célula hospedeira, e dentre elas, um mecanismo evolutivo de evasão imunológica que limita as respostas imunes do hospedeiro em favor da replicação viral (130). Os *Flavivirus* possuem proteínas imunomoduladoras que desempenham este papel (130). Após a inoculação a partir da picada de um mosquito infectado, o YFV passa a replicar nos tecidos locais, em seguida migra para os gânglios linfáticos, replicando em monócitos, resultando na chamada viremia primária, e subsequentemente, o vírus é distribuído para tecidos periféricos pelo sistema circulatório (5). Para completar o ciclo de biomagnificação e replicação viral intra-hospedeiro, esses vírus utilizam os mecanismos de evasão da resposta imunológica para que o sistema imune, apresente tolerância à amplificação

periférica, e conseqüentemente, possa atingir a circulação e atravessar as barreiras importantes para a fisiologia e defesa de agentes externos, como a BHE. Devido ao recorte temporal do estudo, os pacientes considerados negativos para YFV, não tiveram o vírus detectado no cérebro. Contudo, por mais que não haja em nossos resultados uma correlação direta entre tempo de doença e detecção do vírus no cérebro, sabemos que possivelmente, o vírus nestes pacientes ainda não teria tido tempo de alcançar/romper a BHE.

Para investigar os mecanismos moleculares envolvidos na acessibilidade do vírus ao tecido cerebral, um dos principais genes destacados nos pacientes que conseguiram limitar essa acessibilidade (YFV-negativos no tecido cerebral) foi o gene *HLA-DRA*, caracterizado como *down-regulated* com cerca aproximadamente  $\log_2(20)$  menos expresso. Cabe frisar que o gene *HLA-DRA* desempenha um papel crucial na resposta imunológica adaptativa, permitindo ao sistema imunológico reconhecer e reagir de forma coordenada a ameaças como patógenos invasores. Ele está ligado à apresentação de antígenos às células T auxiliares, essenciais na coordenação da resposta imune adaptativa, ativando outras células como células B e T citotóxicas para combater infecções. O gene *HLA-DRA down-regulated* pode impactar a capacidade das células apresentadoras de antígenos em alertar eficazmente as células T auxiliares sobre a presença de patógenos, potencialmente explorada por patógenos para evadir o sistema imunológico e prejudicar uma resposta eficaz. Ademais, a ligação de IFN- $\alpha/\beta$  ao receptor de IFN- $\alpha/\beta$  na superfície das células infectadas desencadeia a ativação da via de expressão gênica que estabelece mecanismos antivirais no hospedeiro. A *down-regulation* de vias de IFN- $\alpha$  nesta condição também enfatiza o potencial efeito neuroevasor do YFV. Em outro trabalho desenvolvido pelo nosso grupo de pesquisa, dados de transcriptoma de fígado *post-mortem* também mostram inúmeras vias imunológicas *down-regulated*, principalmente quando comparados ao grupo do YFV-silvático *versus* controle, onde ativação de leucócitos, ativação de resposta imune adaptativa, e ativação da produção de IFN, são algumas das principais vias classificadas como *down-regulated* (*dados não publicados*). Um estudo analisando o soro de pacientes com YFV revelou a ocorrência de um aumento significativo de citocinas, indicando gravidade na doença sistêmica. Os pacientes mais graves distinguem-se dos casos leves principalmente devido à flutuações no número de neutrófilos, que é atribuída à supressão da resposta imune ativa (146).

Neste contexto, para o grupo com YFV detectado no cérebro, os nossos resultados corroboram com resultados esperados e encontrados na literatura, indicando que a presença de vias super expressas de produção de citocinas, podem estar relacionadas ao agravamento da doença no cérebro. Sob uma perspectiva histórica, a tempestade de citocinas, foi bem elucidada



durante o estudo abordando a infecção causada pelo vírus Influenza, e ficou conhecida como *influenza-like syndrome*, que de forma geral era encontrada após infecções sistêmicas (147,148). Rothwell *et al.* descreve no capítulo do livro “*Immune response in nervous systems*”, que as citocinas possuem efeitos relevantes na função do sistema nervoso quando sob condição patológica (149). Em nossos dados, a produção de citocinas pelo sistema imunológico inato ocorre em resposta ao YFV, quando este, consegue atravessar a BHE. No entanto, a presença exacerbada destas proteínas pode estar ligada a danos severos no SNC, como consequente prejuízo das funções neurais. Infelizmente, não houve o acompanhamento prospectivo de nenhum paciente com sobrevida para realizar uma avaliação de possíveis sequelas neurológicas associadas à infecção causada pelo YFV.

A BHE, composta por células endoteliais dos capilares cerebrais que estão interligadas pelas *tight junctions*, e suportadas por astrócitos e pericitos é apenas uma das barreiras entre o sistema periférico e o SNC. O endotélio do plexo-coróide não possui a função de barreira em si, contudo um conjunto de células epitélio-endoteliais formam *tight junctions*, que inibem a difusão paracelular de moléculas solúveis em água no líquido cefalorraquidiano, estabelecendo a denominada barreira hemato-líquido cefalorraquidiano (BHLC) (150). Alguns vírus, como Zika (ZIKV), vírus do oeste do Nilo (WNV), e o vírus da encefalite japonesa (JEV), são conhecidos por também utilizar a BHLC para atingir o SNC, uma vez que antígenos virais foram encontrados em células do plexo coróide em experimentos *in vitro* (151). Os resultados de expressão gênica observados em nosso estudo, no grupo em que o YFV não foi detectado no tecido cerebral, mostram a superexpressão de genes relacionados a vias de morfogênese do epitélio ramificado, sugerindo que em resposta à infecção, genes foram expressos de forma a aumentar a proteção do SNC, diminuindo a passagem de patógenos via BHLC. No entanto, ainda não há estudos suficientes que mostrem o papel da BHLC na neuroinvasão viral. Porém, estudos atuais, a exemplo do desenvolvido por Marshall *et. al.*, que utilizou organóides infectados com SARS-CoV-2 para estudar patogênese e neurotropismo podem servir como base para o estudo de neuroinvasão para os arbovírus (152).

A análise de dados de expressão gênica entre pacientes positivos (vírus detectado por RT-qPCR) e controles, foi realizada na busca de assinaturas gênicas referente à infecção e replicação das linhagens silvática e vacinal do YFV no tecido cerebral. Quando observados os histogramas de valores de *p* para as duas linhagens, os dois grupos apresentaram assinaturas distintas. Enquanto a curva em formato de “colina” do grupo YFV-17DD *versus* controle, ressalta uma pequena quantidade do total de genes como diferencialmente expressos e que rejeitam a hipótese nula ( $p < 0,05$ ), a curva em formato “L”, o grupo YFV-silvático *versus*

controle mostra que uma grande quantidade de genes rejeitam a hipótese nula e se enquadram como diferencialmente expressos com  $p < 0,05$ . Ainda, para o vírus silvático, este possui mais de duas vezes a quantidade DEGs em relação ao vírus vacinal.

A obrigatoriedade do uso de uma célula hospedeira, faz com que vírus tenha que fazer uso eficaz da maquinaria celular para o sucesso na biomagnificação e formação de progênie viral explorando picos adaptativos. Durante as infecções virais, as mitocôndrias são também afetadas pelo vírus, e possivelmente influenciadas por alterações fisiológicas, as quais podem ser um dos componentes chave para a sua persistência viral no organismo hospedeiro, como já foi descrito para o vírus da hepatite B (HBV) e hepatite C (HCV) (153,154). Esta persistência deve-se principalmente à teoria de que interações vírus-mitocôndria promovem alterações nos mecanismos de sinalização antiviral associados às mitocôndrias (do inglês, *mitochondrial antiviral signaling* - MAVS). Estudos mostram que para o HCV, mitofusinas (proteínas de membrana mitocondrial) interagem com o MAVS e iniciam uma atividade antiviral eficaz (155), no entanto, o bloqueio destas leva ao decaimento das respostas antivirais, podendo assim sugerir que a integridade mitocondrial e de suas funções é essencial para a manutenção da imunidade inata perante uma infecção viral (156). Crucialmente, em nosso estudo, o gene *MRPL45*, um gene que codifica proteínas ribossomais da mitocôndria, encontrou-se aproximadamente 4 milhões de vezes menos expresso na condição em que o vírus vacinal foi detectado no cérebro em relação à condição controle. Isso nos possibilita hipotetizar que a *down-regulation* de um gene mitocondrial poderia influenciar diretamente no MAVS, fazendo com que a imunidade inata do hospedeiro seja prejudicada, beneficiando desta forma, a replicação do YFV. Ademais, este achado implica numa possível redução de atividade mitocondrial, que pode ter um impacto na progressão da neuropatogênese e em desfechos desfavoráveis.

Neste sentido, vale mencionar um estudo utilizando o ZIKV em cultura de células neurais humanas, mostrou que a infecção por ZIKV interrompeu a dinâmica mitocondrial, alterando sua função nestas estruturas, e que o silenciamento de genes envolvidos na fissão das mitocôndrias, influenciou positivamente na sobrevivência das células neuronais, revelando uma possível abordagem terapêutica promissora envolvendo o controle da desregulação mitocondrial após infecção por ZIKV. O estudo conclui que a desregulação mitocondrial contribui para a morte celular mediada por ZIKV em células neuronais (157). Seguindo este raciocínio, um possível dano neurológico devido à morte neuronal poderia ter ocorrido no grupo em que o vírus vacinal foi detectado no cérebro e que a atividade mitocondrial poderia estar desregulada, dado o padrão de expressão gênica observado.

O denominado “privilegio imunológico” do SNC é definido pelo controle do tráfego de células imunes para dentro do SNC, o qual é altamente regulado de forma a evitar respostas inflamatórias agudas *in situ*, já que o cérebro possui baixa capacidade regenerativa. Contudo, ainda é permitido um baixo nível de imunovigilância de células T para manter sob controle pequenas inflamações (84). Estudos mostram que em resposta à uma inflamação sistêmica, o endotélio cerebral componente da BHE, estimula o tráfico linfocitário e de células *natural killer*, neutrófilos, e monócitos (84,158), respostas estas também observadas em nosso estudo, em que vias enriquecidas de genes comuns à condição positiva no cérebro (detecção do vírus YFV-silvático e vacinal por RT-qPCR) mostram a superexpressão de vias envolvidas na quimiotaxia de células imunes. Neste sentido cabe observar que o mecanismo de astrogliose reativa, é característico de distúrbios neurológicos, incluindo infecções do SNC. Foi definido por Moulson *et. al.*, que a astrogliose reativa é um processo pelo qual os astrócitos tornam-se reativos em resposta a lesões e inflamações, que são tipicamente descritas como hipertrofia com expressão aumentada da proteína ácida fibrilar glial (do inglês, *glial fibrillary acidic protein - GFAP*). Estas mudanças incluem alterações na regulação transcricional, mudança bioquímica, morfológica e/ou fisiológica (159) e permite que os astrócitos realizem papéis adicionais quando o SNC encontra-se em estado inflamatório, a exemplo de suporte metabólico para neurônios que estão vulneráveis, regulação da BHE, imunomodulação e remodelamento sináptico. Assim como observado em nosso estudo em que vias de cálcio foram superexpressas quando o YFV foi detectado no tecido cerebral, estudos mostram que as vias alteradas de sinalização de cálcio nos astrócitos reativos podem refletir diretamente o consumo local de energia em resposta à tecidos neuronais danificados, sendo assim considerado um marcador de inflamação local, uma vez que o fenótipo reativo dos astrócitos frequentemente é encontrado em doenças do SNC (160,161).

Genes centrais como *CXCL10*, podem estar diretamente relacionados à infiltração leucocitária, diminuição da expressão de proteínas importantes para o funcionamento das *tight junctions* e consequente aumento da permeabilidade da BHE (162). Outro gene que pode ser considerado central e marcador de inflamação, também foi encontrado nesta condição, é o gene *CCL8*, além de estar incluso em quase todas as principais vias expressas nesta condição, foi superexpresso e é responsável pela quimiotaxia de monócitos para dentro do SNC. A infiltração de monócitos é tida como marcador de inflamação no SNC, incluindo por infecções virais. Estas células não especializadas, migram para o SNC, e se diferenciam em células dendríticas e macrófagos, e passam a desempenhar funções antivirais efetoras, como apresentação de antígenos, estimulação de células T e produção de citocinas pró-inflamatórias (163).

De uma visão sistêmica, é possível afirmar que em ambas as condições estudadas (YFV-silvático e vacinal detectado no cérebro) houve a ativação de resposta imune inata e adaptativa em resposta ao vírus, com ativação de células T, principalmente linfócitos TCD4+ auxiliares que são diferenciados em células Th1 responsáveis pela expressão de interferon gamma IFN- $\gamma$ . Posso aqui acrescentar, uma diferença observada se dá em relação ao metabolismo de amônia, já que no grupo pacientes que tiveram o vírus vacinal detectado no cérebro, vias de processamento de glutamina estão *down-regulated*, podendo assim perturbar a homeostase cerebral em consequência da hiperamonemia e decréscimo da expressão de glutamina sintetase, responsável pela metabolização de amônia (164). Além disso, a função mitocondrial perturbada (*down-regulation* do gene *MRPL-45*) é resultado direto de citotoxicidade induzida por amônia (165).

É interessante notar que em outros tipos de vacinas, como as plataformas baseadas em mRNA e adenovírus, desenvolvidas para contra o vírus SARS-CoV-2, lesão hepática devido a exacerbação autoimune e aumento do nível sistêmico de amônia foram relatados em casos de reações adversas devido a hepatite autoime. Isto permite hipotetizar que além do hepatotropismo clássico da YF, semelhanças moleculares entre proteínas virais e proteínas específicas do fígado, podem levar à um sistema imunológico reativo e a consequente destruição de células hepáticas (166–169).

Da mesma forma que foi observado no grupo em que o YFV-17DD e silvático foram detectados no cérebro, uma maior quantidade de DEGs (aproximadamente 3 vezes mais) foram encontradas no grupo de pacientes em que YFV-silvático não foi detectado no cérebro, porém foi detectado em pelo menos um órgão e caracterizando o paciente como positivo. Portanto, observando a superexpressão de vias vinculadas ao transporte e secreção de proteínas no grupo de genes comuns à classe de YFV-silvático e vacinal não detectado no cérebro, é possível hipotetizar que, devido à infecção sistêmica de pacientes que foram à óbito por YF, existe uma alteração na regulação da transferência de proteínas para dentro e fora do SNC. Isso, pode estar associado a um mecanismo de proteção do tecido cerebral contra eventuais danos, dada a infiltração viral. Sabe-se que as funções da BHE são altamente reguladas quando o organismo encontra-se sob infecção sistêmica e esta alteração pode ser classificada como disruptiva (presença ou ausência de ruptura física da BHE devido à alterações em nível histológico, com alterações nas *tight junctions*) e não disruptiva (como por exemplo, mudança no tráfego através da BHE e portanto nas quantidades de moléculas atuando no cérebro). Se não houver alteração disruptiva da BHE, não significa que o SNC não possa estar sendo alterado e sofrendo alterações neurológicas, uma vez que as alterações não disruptivas também podem

levar à danos (84), sugerindo assim, que alterações compensatórias na BHE perante uma inflamação sistêmica, previne o afrouxamento das *tight junctions*, porém facilita o aumento da permeabilidade de proteínas pró-inflamatórias para dentro do SNC (84).

Ainda não está claro na literatura se as síndromes neurológicas associadas ao YFV são devido à infecção direta por YFV das células cerebrais, ou então secundárias à fenômenos imunomoduladores. É interessante pensar que mesmo na ausência de componentes virais, danos à BHE podem ocorrer simplesmente mediados por efeitos inflamatórios da infecção sistêmica, sem infecção direta de neurônios ao redor do endotélio cerebral. Em um estudo realizado com uma coorte de pacientes que foram à óbito devido à COVID-19, ensaios de imunofluorescência mostraram lesões à BHE, mesmo sem a detecção do vírus neste tecido, ressaltando que se a infecção sistêmica for suficientemente grave, pode causar alterações não disruptivas associadas a perturbações na BHE (170). Junto com outros estudos publicados recentemente, pode-se concluir que mesmo quando relatados sintomas neurológicos *pre-mortem*, este achado não é resultado de infecção viral direta do parênquima cerebral, mas provavelmente influenciado pela infecção sistêmica (171).

Wenzel *et. al.* também mostrou que o SARS-CoV-2 pode afetar o SNC diretamente pela infecção das células epiteliais da BHE, contudo nenhum vírus foi detectado no tecido cerebral neste estudo (172). O autor ainda hipotetiza que a rara detecção de SARS-CoV-2 no tecido cerebral poderia ser explicada pela rápida indução de morte de células endoteliais infectadas, assim como ocorre em macrófagos pulmonares infectados por este vírus, que sofrem morte celular devido um estado inflamatório prévio devido à liberação de citocinas locais. O achado mais interessante deste estudo, no entanto, é que a morte celular induzida pela replicação do SARS-CoV-2 nos macrófagos infectados, impediu a replicação viral no tecido pulmonar e conseqüente a propagação viral (173), e o mesmo foi demonstrado quanto à infecção do endotélio cerebral (174). Em princípio, pode-se hipotetizar algo semelhante quando pensado na infecção do YFV, uma vez que pacientes que não tiveram a detecção do YFV no cérebro tiveram o vírus detectado por imunohistoquímica nas células do endotélio. Como efeito da morte celular, o vírus perde o seu local de replicação, sendo esta uma das possíveis estratégias antivirais do hospedeiro. Acho relevante aqui enfatizar que em nosso estudo, achados histopatológicos no grupo de pacientes que não tiveram o YFV-silvático e vacinal detectado no tecido cerebral, ressaltam possíveis danos à BHE devido à infiltração inflamatória perivascular que poderiam levar a danos neurológicos.

Quando observada as vias do grupo com YFV-silvático e vacinal detectado no cérebro, a via de secreção e transporte de íons de cálcio está super expressa. Isto sugere, como

já enfatizado anteriormente, um estado inflamatório devido a astrogliose reativa. De maneira similar, e indo de encontro com as hipóteses aqui apresentadas de que a ausência do vírus não é indicador de ausência de neuroinflamação, muitas vias de regulação de transporte e secreção de íons de cálcio também estão expressas na condição em que o YFV-17DD não foi detectado no cérebro mostrando que mesmo na ausência do vírus, a exacerbação da sinalização de cálcio devido à astrócitos reativos à YF sistêmica são indicadores de gravidade. Assim como descrito por Galea *et al.* (84) mesmo na ausência de neuroinvasão de patógenos, a inflamação sistêmica estimula a passagem de leucócitos para o cérebro, resultados estes observados com a super expressão de genes que expressam citocinas que induzem o tráfico leucocitário, *CXCL9*, *CXCL10* e *CXCL11*.

Pacientes com sobrevida e que tiveram sintomas neurológicos persistentes após a YF, como dificuldade cognitiva, fadiga e disfunção autônoma, sugerem um quadro de síndrome pós-viral associada ao SNC. A doença neurológica associada à vacina da febre amarela (do inglês, *YF vaccine-associated neurologic disease -YEL-AND*) é um evento adverso raro, e potencialmente grave já bem descrito na literatura, com uma incidência no Brasil que varia de 0.2- 0.9 casos/100.000 doses (59,175), contudo nada foi encontrado na literatura referente à síndrome neurológica associada à infecção pelo vírus silvático.

Como conclusão de nossos achados até aqui descritos, fica possível sugerir que em casos de síndrome neurológica, pacientes que foram infectados com o YFV-silvático ou tiveram reação adversa à vacina (YFV-17DD) podem apresentar os seguintes atributos: (i) as partículas virais de YFV podem ultrapassar BHE e atingir o tecido cerebral; (ii) as partículas de YFV atingiram a BHE, infectaram as células endoteliais, contudo, podem não ultrapassar e não chegaram ao cérebro; (iii) quando as partículas virais de YFV não atingem o SNC, existe perturbação na expressão gênica que pode estar associada aos danos neurológicos secundários da infecção sistêmica. Em casos que o YFV foi detectado no cérebro, uma resposta imunológica contra o vírus foi destaque com produção de citocinas pró-inflamatórias, e migração de células imunes. No entanto, naqueles casos em que o YFV não foi detectado no cérebro, indícios de inflamação como a exacerbação de vias relacionadas a secreção e transporte de proteínas e íons de cálcio para dentro do SNC fundamentam nossa conclusão de que síndromes neurológicas associadas ao YFV podem ser secundárias a efeitos da infecção sistêmica.

## **7. Limitações do trabalho**

O trabalho aqui apresentado foi desenvolvido ao longo de 5 anos e durante essa jornada houveram alguns fatores limitantes do nosso trabalho. Inicialmente cabe dizer que por se tratar de um estudo realizado em cérebro humano *post-mortem*, a disponibilidade de amostra que satisfizesse os critérios de inclusão nos grupos delineados para o estudo era escassa, principalmente amostras de casos caracterizados como suspeita de reação adversa vacinal. Pacientes controle que não foram à óbito por nenhuma doença infecciosa também foram difíceis de encontrar e principalmente de conseguir consentimento da família para inclusão na pesquisa.

As amostras foram coletadas de dezembro de 2017 à setembro de 2018, levando cerca de um ano para que conseguíssemos a inclusão de todos os pacientes com critérios de inclusão válidos. No entanto, para que um experimento de RNA-seq seja feito sem nenhum viés, os grupos deveriam ter aproximadamente a mesma média de idade, sexo e não possuir comorbidades que poderiam alterar o resultado da expressão gênica. Contudo, isso não foi possível devido à escassez de amostras. Devido à degradação do RNA, o tempo *post-mortem* (TPM), também foi um fator importante para a inclusão de casos e muitas vezes não tivemos como optar por aqueles com menor TPM, mesmo que não tenhamos encontrado correlação entre TPM e RIN. Contornamos este problema, utilizando um *kit* que permite trabalhar com RNAs fragmentados/com algum nível de degradação. Ademais, não obtivemos informações relevantes quanto ao hospedeiro em relação à vacinação prévia, o que nos permitiria avaliar estes casos de uma forma mais completa. Creio que os dois principais fatores limitante de nosso estudo estejam relacionados à: (i) o estudo corresponde à um recorte temporal da doença como um todo, não permitindo assim obter resultados totalmente conclusivos quanto a neuroinvasividade do YFV no SNC; (ii) mesmo sendo uma técnica muito sensível, resultados negativos de RT-qPCR não necessariamente garantem que o vírus não conseguir invadir o SNC e chegar ao cérebro, e sendo portanto, sugestivo para isso.

No entanto mesmo considerando os *caveat* acima citados, nosso estudo colaborativo foi longo, detalhado, exploratório, desafiador e inovador em inúmeras frentes. Acredito que conseguimos atingir uma grande parte dos objetivos esperados e ter trazido informação meritória, que acrescenta um marco ao entendimento da neuropatologia da febre amarela.

## 8. Perspectivas Futuras

Com este trabalho foi possível realizar uma vasta leitura da neuropatologia do YFV, via análise de expressão gênica de pacientes com o vírus da linhagem silvática e vacinal

detectados em tecido cerebral *post-mortem*. A análise das principais vias que foram superexpressas e *down-regulated* associadas à genes em diferentes grupos de pacientes, nos permitiu fazer algumas hipóteses referente à neuropatologia do YFV. Contudo, muito ainda deve ser feito para que as hipóteses aventadas fiquem melhor embasadas ou sejam refutadas.

Inicialmente para futura publicação entendo que os principais genes relevantes em cada condição devam ser validados por RT-qPCR. Devido ao alto custo e como alternativa ao método de *single-cell* RNAseq (sc-RNAseq), gostaria de propor o uso do *bulk*-RNAseq para realização do método de deconvolução para compreender a composição celular e possíveis padrão de reguladores associados às condições estudadas. Com esta análise poderemos inferir se houve migração de células imunes, perda neuronal, entre outros. Experimentos *in vitro* utilizando o YFV-silvático e vacinal para infectar células neurais, ou organoides cerebrais seriam muito importantes para investigar a infectividade, o tropismo e a cinética de replicação do YFV. Em nosso estudo por meio da análise do corte histológico foi possível observar que o YFV não infecta neurônios, e possivelmente a morte neuronal se dá devido a efeitos secundários da infecção de outras células do SNC, assim como acontece com o HIV (176). No entanto, experimentos *in vitro*, seriam de extrema importância para validar tais achados.

Genes principais aqui citados, também poderiam servir de base para experimentos de silenciamento gênico em camundongos em busca de melhor entendimento da neuropatogenicidade do YFV. Não foi possível somente com os dados obtidos aqui, concluir se o YFV não foi detectado no cérebro devido ao tempo de doença do paciente, já que não houve diferença significativa de tempo de infecção entre os grupos estudados, assim, seria de extrema valia estudar em que momento o vírus adentra o sistema nervoso analisando camundongos infectados com YFV modificado com a inserção de um gene luminescente e monitoramento diário coletando o cérebro para verificação de carga viral e imunofluorescência. É possível, também, injetar corantes (*Blue Evans*) no sistema circulatório de camundongos para ajudar a entender a dinâmica da BHE perante a uma infecção viral sistêmica.

## Referências

1. de PM, Gao GF, Gritsun T, Marin MS, Jiang WR, Venugopal K, et al. An Arbovirus Cline across the Northern Hemisphere. *Virology*. 1995;210(1).
2. Lasnet. African Conference of the Yellow Fever. 1929. 1–298 p.
3. Cunha M dos P, Duarte-Neto AN, Pour SZ, Ortiz-Baez AS, Černý J, Pereira BB de S, et al. Origin of the São Paulo Yellow Fever epidemic of 2017–2018 revealed through molecular epidemiological analysis of fatal cases. *Sci Rep*. 2019;9(1):1–10.
4. Cunha MDP, Duarte-Neto AN, Pour SZ, Pereira BB de S, Ho YL, Perondi B, et al. Phylogeographic patterns of the yellow fever virus around the metropolitan region of São Paulo, Brazil, 2016–2019. *PLoS Negl Trop Dis*. 2022;16(9).



5. Duarte-Neto AN, Cunha M dos P, Marcilio I, Song ATW, de Martino RB, Ho Y, et al. Yellow Fever and Orthotopic Liver Transplantation: new insights from the autopsy room for an old but reemerging disease. *Histopathology*. 2019;(455):1–11.
6. Duarte-neto AN, Monteiro RA de A, Johnsson J, Cunha M dos P, Pour SZ, Saraiva AC, et al. Ultrasound-guided minimally invasive autopsy as a tool for rapid post-mortem diagnosis in the 2018 São Paulo yellow fever epidemic: Correlation with conventional autopsy. *PLoS Negl Trop Dis*. 2019;13(7):e0007625.
7. Vasconcelos PF da C. Yellow Fever. *Rev Soc Bras Med Trop*. 2003;36(2):275–93.
8. Kaul RB, Evans M V., Murdock CC, Drake JM. Spatio-temporal spillover risk of yellow fever in Brazil. *Parasites and Vectors*. 2018;11(1).
9. Mondet B. [Yellow fever epidemiology in Brazil]. *Bull Soc Pathol Exot*. 2001;94(3).
10. Rezende IM de, Sacchetto L, Munhoz de Mello É, Alves PA, Iani FC de M, Adelino TÊR, et al. Persistence of Yellow fever virus outside the Amazon Basin, causing epidemics in Southeast Brazil, from 2016 to 2018. *PLoS Negl Trop Dis*. 2018;12(6):1–12.
11. CASALS J. Arboviruses: Incorporation in a General System of Virus Classification. In: *Comparative Virology*. 1971.
12. Madewell ZJ. Arboviruses and their vectors. Vol. 113, *Southern Medical Journal*. 2020.
13. Kuno G, Chang G-JJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the Genus *Flavivirus*. *J Virol*. 1998;72(1).
14. Moureau G, Cook S, Lemey P, Nougairede A, Forrester NL, Khasnatinov M, et al. New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences. *PLoS One*. 2015;10(2):e0117849.
15. Zhang W, Chipman PR, Corver J, Johnson PR, Zhang Y, Mukhopadhyay S, et al. Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. *Nat Struct Biol*. 2003;10(11):907–12.
16. Chambers TJ, Hahn CS, Galler R, Rice CM. Flavivirus Genome Organization, Expression, and Replication. *Annu Rev Microbiol*. 1990;44:649–88.
17. Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, et al. Dengue: a continuing global threat. *Nat Publ Gr* [Internet]. 2010;8(12):S7–16. Available from: <http://dx.doi.org/10.1038/nrmicro2460>
18. Bryant JE, Holmes EC, Barrett ADT. Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. *PLoS Pathog* [Internet]. 2007;3(5):e75. Available from: <http://dx.plos.org/10.1371/journal.ppat.0030075>
19. Nunes MRT, Palacios G, Cardoso JF, Martins LC, Sousa-Junior EC, Lima CPS de, et al. Genomic and phylogenetic characterization of Brazilian Yellow Fever virus strains. *J Virol*. 2012;1–33.
20. Barnett ED. Yellow fever: Epidemiology and prevention. *Clin Infect Dis*. 2007;44(6).
21. Robertson SE, Hull BP, Tomori O, Bele O, LeDuc JW, Esteves K. Yellow fever: A decade of reemergence. *JAMA*. 1996;276(14).
22. Torres Sara Gabriela L, Caicedo Byron Enrique B, Torres Oscar Patricio N, Ortíz Luis Agustín C. Yellow Fever: A Reemerging Disease in South America, 2000–2016. *ESPOCH Congr Ecuadorian J STEAM*. 2022;
23. Figueiredo P de O, Stoffella-Dutra AG, Costa GB, de Oliveira JS, Amaral CD, Santos JD, et al. Re-emergence of yellow fever in Brazil during 2016–2019: Challenges, lessons learned, and perspectives. Vol. 12, *Viruses*. 2020.
24. Monath TP, Vasconcelos PFC. Yellow fever. *J Clin Virol* [Internet]. 2015;64:160–73. Available from: <http://dx.doi.org/10.1016/j.jcv.2014.08.030>
25. Xavier AR, Freitas GS, Santos CF, Januzzi WA, Lacerda GS, Carvalho ERM, et al. Yellow fever: Laboratorial diagnosis and clinical manifestations. *J Bras Patol e Med Lab*. 2018;54(5).
26. Houghton-Triviño N, Montaña D, Castellanos JE. Dengue-yellow fever sera cross-reactivity; challenges for diagnosis. *Rev Salud Publica*. 2008;10(2).
27. Domingo C, Patel P, Yillah J, Weidmann M, Méndez JA, Nakouné ER, et al. Advanced yellow fever virus genome detection in point-of-care facilities and reference laboratories. *J Clin Microbiol*. 2012;50(12):4054–60.
28. Diallo D, Sall AA, Diagne CT, Faye O, Hanley KA, Buenemann M, et al. Patterns of a sylvatic

- yellow fever virus amplification in southeastern Senegal, 2010. *Am J Trop Med Hyg.* 2014;90(6):1003–13.
29. Sow A, Loucoubar C, Diallo D, Faye O, Ndiaye Y, Senghor CS, et al. Concurrent malaria and arbovirus infections in Kedougou, southeastern Senegal. *Malar J.* 2016;15(1):47.
  30. Barbosa CM, Paola N Di, Cunha MP, Rodrigues-jesus MJ, Araujo DB, Silveira VB, et al. Yellow Fever Virus RNA in Urine and Semen of Convalescent Patient, Brazil. *Emerg Infect Dis.* 2018;24(1):176–8.
  31. de Rezende IM, Oliveira GFG, Costa TA, Khan A, Pereira LS, Santos TA, et al. Yellow Fever Molecular Diagnosis Using Urine Specimens during Acute and Convalescent Phases of the Disease. *J Clin Microbiol.* 2022;60(8).
  32. Dia M, Bob NS, Talla C, Dupressoir A, Escadafal C, Thiam MS, et al. Performance assessment and validation of a plaque reduction neutralization test (PRNT) in support to yellow fever diagnostic and vaccine clinical trials. *J Med Virol.* 2023;95(4).
  33. Basile AJ, Goodman C, Horiuchi K, Laven J, Panella AJ, Kosoy O, et al. Development and validation of an ELISA kit (YF MAC-HD) to detect IgM to yellow fever virus. *J Virol Methods.* 2015;225.
  34. Ofosu-Appiah LH, Amelor DK, Ayensu B, Akyereko E, Rabiwu NI, Opare D, et al. An evaluation of the diagnostic performance characteristics of the Yellow Fever IgM immunochromatographic rapid diagnostic test kit from SD Biosensor in Ghana. *PLoS One.* 2022;17(1 January).
  35. Monath TP. Yellow fever: an update. *Lancet Infect Dis.* 2001;1:11–20.
  36. Monath TP, Barrett ADT. Pathogenesis and Pathophysiology of Yellow Fever. *Adv Virus Res.* 2003;60:343–95.
  37. Duarte-Neto AN, Cunha MDP, Marcilio I, Song ATW, de Martino RB, Ho Y-L, et al. Yellow fever and orthotopic liver transplantation: new insights from the autopsy room for an old but re-emerging disease. *Histopathology.* 2019;75(5).
  38. Aragão H de B. Report upon some researches on Yellow Fever. *Mem Inst Oswaldo Cruz.* 1928;21:23–46.
  39. N Paul Hudson. The Pathology of Experimental Yellow Fever in the Macacus Rhesus - Comparison with the Pathology of Yellow Fever in Man. *Am J Pathol.* 1828;4(5):419–430.9.
  40. Mathis C, Sellards AW, Laigret J. Sensibilité du Macacus rhesus au virus de la fièvre jaune. *CR Acad Sci.* 1928;186:604–6.
  41. Norrby E. Yellow fever and Max Theiler: The only Nobel Prize for a virus vaccine. Vol. 204, *Journal of Experimental Medicine.* 2007.
  42. Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol.* 2009;10(1):116–25.
  43. Beck AS, Barrett ADT. Current status and future prospects of yellow fever vaccines. Vol. 14, *Expert Review of Vaccines.* 2015.
  44. Bovay A, Fuertes Marraco SA, Speiser DE. Yellow fever virus vaccination: an emblematic model to elucidate robust human immune responses. Vol. 17, *Human Vaccines and Immunotherapeutics.* 2021.
  45. Gotuzzo E, Yactayo S, Córdova E. Review article: Efficacy and duration of immunity after yellow fever vaccination: Systematic review on the need for a booster every 10 years. *Am J Trop Med Hyg.* 2013;89(3).
  46. Staples JE, Barrett ADT, Wilder-Smith A, Hombach J. Review of data and knowledge gaps regarding yellow fever vaccine-induced immunity and duration of protection. *npj Vaccines.* 2020;5(1).
  47. Vasconcelos PFC, Luna EJ, Galler R, Silva LJ, Coimbra TL, Barros VLRS, et al. Serious adverse events associated with yellow fever 17DD vaccine in Brazil: A report of two cases. *Lancet.* 2001;358(9276):91–7.
  48. Fox JP, Lennette EH, Manso C, Souza Aguiar JR. Encephalitis in man following vaccination with 17 d yellow fever virus. *Am J Epidemiol.* 1942;36(2).
  49. Lindsey NP, Schroeder BA, Miller ER, Braun MM, Hinckley AF, Marano N, et al. Adverse event reports following yellow fever vaccination. *Vaccine.* 2008;26(48):6077–82.

50. Lindsey NP, Rabe IB, Miller ER, Fischer M, Staples JE. Adverse event reports following yellow fever vaccination, 2007-13. *J Travel Med.* 2016;23(5).
51. van den Pol AN. Viral infections in the developing and mature brain. Vol. 29, *Trends in Neurosciences.* 2006.
52. Laureti M, Narayanan D, Rodriguez-Andres J, Fazakerley JK, Kedzierski L. Flavivirus Receptors: Diversity, Identity, and Cell Entry. Vol. 9, *Frontiers in Immunology.* 2018.
53. Herculano-Houzel S. The human brain in numbers: A linearly scaled-up primate brain. Vol. 3, *Frontiers in Human Neuroscience.* 2009.
54. Ungerleider LG, Haxby J V. “What” and “where” in the human brain. *Curr Opin Neurobiol.* 1994;4(2).
55. Jinnou H, Sawada M, Kawase K, Kaneko N, Herranz-Pérez V, Miyamoto T, et al. Radial Glial Fibers Promote Neuronal Migration and Functional Recovery after Neonatal Brain Injury. *Cell Stem Cell.* 2018;22(1).
56. Wang R, Sharma R, Shen X, Laughney AM, Funato K, Clark PJ, et al. Adult Human Glioblastomas Harbor Radial Glia-like Cells. *Stem Cell Reports.* 2020;14(2).
57. Furr SR, Marriott I. Viral CNS infections: Role of glial pattern recognition receptors in neuroinflammation. *Front Microbiol.* 2012;3(JUN).
58. Bayão T de S, Cupertino M do C, Gomes AP, Siqueira-Batista R. Yellow fever vaccination and neurological complications. Vol. 51, *Revista da Sociedade Brasileira de Medicina Tropical.* 2018.
59. Martins R de M, Pavão ALB, de Oliveira PMN, dos Santos PRG, Carvalho SMD, Mohrdieck R, et al. Adverse events following yellow fever immunization: Report and analysis of 67 neurological cases in Brazil. *Vaccine.* 2014;32(49).
60. Alvarado-Socarras JL, Rodriguez-Morales AJ. Etiological agents of microcephaly: Implications for diagnosis during the current Zika virus epidemic. *Ultrasound Obstet Gynecol.* 2016;47(4):525–6.
61. Oliveira DBL, Almeida FJ, Durigon EL, Mendes ÉA, Braconi CT, Marchetti I, et al. Prolonged Shedding of Zika Virus Associated with Congenital Infection. *N Engl J Med* [Internet]. 2016;375(12):1202–4. Available from: <http://www.nejm.org/doi/10.1056/NEJMc1607583>
62. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, et al. The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* [Internet]. 2016;534(7606):267–71. Available from: <http://dx.doi.org/10.1038/nature18296>
63. Araujo LM, Ferreira MLB, Nascimento OJM. Guillain-Barré syndrome associated with the Zika virus outbreak in Brazil. *Arq Neuropsiquiatr.* 2016;74(3).
64. Hammack C, Ogden SC, Madden JC, Medina A, Xu C, Phillips E, et al. Zika Virus Infection Induces DNA Damage Response in Human Neural Progenitors That Enhances Viral Replication. *J Virol.* 2019;93(20).
65. Victora CG, Schuler-Faccini L, Matijasevich A, Ribeiro E, Pessoa A, Barros FC. Microcephaly in Brazil: How to interpret reported numbers? *Lancet.* 2016;387(10019):621–4.
66. de Araújo TVB, Ximenes RA de A, Miranda-Filho D de B, Souza WV, Montarroyos UR, de Melo APL, et al. Association between microcephaly, Zika virus infection, and other risk factors in Brazil: Final report of a case-control study. *Lancet Infect Dis.* 2018;
67. Needham EJ, Helmy A, Zanier ER, Jones JL, Coles AJ, Menon DK. The immunological response to traumatic brain injury. Vol. 332, *Journal of Neuroimmunology.* 2019.
68. Theiler M, Smith HH. The Use of Yellow Fever Virus Modified By in Vitro Cultivation for Human Immunization. *J Exp Med.* 1937;65(6):787–800.
69. Theiler M. Studies on the action of Yellow Fever virus in mice. *Ann Trop Med Parasitol.* 1930;24(2):249–72.
70. Porudominsky R, Gotuzzo EH. Yellow fever vaccine and risk of developing serious adverse events: A systematic review. Vol. 42, *Revista Panamericana de Salud Publica/Pan American Journal of Public Health.* 2018.
71. Seligman SJ. Yellow fever virus vaccine-associated deaths in young women. *Emerg Infect Dis.* 2011;17(10).
72. Stones PB, Macnamara FN. Encephalitis following neurotropic yellow fever vaccine administered by scarification in nigeria: Epidemiological and laboratory studies. *Trans R Soc*

- Trop Med Hyg. 1955;49(2).
73. McMahon AW, Eidex RB, Marfin AA, Russell M, Sejvar JJ, Markoff L, et al. Neurologic disease associated with 17D-204 yellow fever vaccination: A report of 15 cases. *Vaccine*. 2007;25(10).
  74. Lemos F de O, França A, Lima Filho ACM, Florentino RM, Santos ML, Missiaggia DG, et al. Molecular Mechanism for Protection Against Liver Failure in Human Yellow Fever Infection. *Hepatol Commun*. 2020;4(5).
  75. Butterworth RF. The neurobiology of hepatic encephalopathy. Vol. 16, *Seminars in Liver Disease*. 1996.
  76. Aldridge DR, Tranah EJ, Shawcross DL. Pathogenesis of hepatic encephalopathy: Role of ammonia and systemic inflammation. Vol. 5, *Journal of Clinical and Experimental Hepatology*. 2015.
  77. Tranah TH, Vijay GKM, Ryan JM, Shawcross DL. Systemic inflammation and ammonia in hepatic encephalopathy. Vol. 28, *Metabolic Brain Disease*. 2013.
  78. Rodrigo R, Cauli O, Gomez-Pinedo U, Agusti A, Hernandez-Rabaza V, Garcia-Verdugo JM, et al. Hyperammonemia induces neuroinflammation that contributes to cognitive impairment in rats with hepatic encephalopathy. *Gastroenterology*. 2010;139(2).
  79. Jayakumar AR, Norenberg MD. Hyperammonemia in Hepatic Encephalopathy. Vol. 8, *Journal of Clinical and Experimental Hepatology*. 2018.
  80. Vidal-Cevallos P, Chávez-Tapia NC, Uribe M. Current approaches to hepatic encephalopathy. Vol. 27, *Annals of Hepatology*. 2022.
  81. Butterworth RF. The liver-brain axis in liver failure: Neuroinflammation and encephalopathy. Vol. 10, *Nature Reviews Gastroenterology and Hepatology*. 2013.
  82. Didier N, Romero IA, Créminon C, Wijkhuisen A, Grassi J, Mabondzo A. Secretion of interleukin-1 $\beta$  by astrocytes mediates endothelin-1 and tumour necrosis factor- $\alpha$  effects on human brain microvascular endothelial cell permeability. *J Neurochem*. 2003;86(1).
  83. De Vries HE, Blom-Roosemalen MCM, Van Oosten M, De Boer AG, Van Berkel TJC, Breimer DD, et al. The influence of cytokines on the integrity of the blood-brain barrier in vitro. *J Neuroimmunol*. 1996;64(1).
  84. Galea I. The blood–brain barrier in systemic infection and inflammation. Vol. 18, *Cellular and Molecular Immunology*. 2021.
  85. Abbott NJ, Patabendige AAK, Dolman DEM, Yusof SR, Begley DJ. Structure and function of the blood-brain barrier. Vol. 37, *Neurobiology of Disease*. 2010.
  86. Wong AD, Ye M, Levy AF, Rothstein JD, Bergles DE, Searson PC. The blood-brain barrier: An engineering perspective. *Frontiers in Neuroengineering*. 2013.
  87. Chen Z, Li G. Immune response and blood–brain barrier dysfunction during viral neuroinvasion. *Innate Immun*. 2021;27(2).
  88. Mustafá YM, Meuren LM, Coelho SVA, De Arruda LB. Pathways exploited by flaviviruses to counteract the blood-brain barrier and invade the central nervous system. Vol. 10, *Frontiers in Microbiology*. 2019.
  89. Sips GJ, Wilschut J, Smit JM. Neuroinvasive flavivirus infections. Vol. 22, *Reviews in Medical Virology*. 2012.
  90. Luethy LN, Erickson AK, Jesudhasan PR, Ikizler M, Dermody TS, Pfeiffer JK. Comparison of three neurotropic viruses reveals differences in viral dissemination to the central nervous system. *Virology*. 2016;487.
  91. Arrighi FE, Hsu TC, Saunders P, Saunders GF. Localization of repetitive DNA in the chromosomes of *Microtus agrestis* by means of in situ hybridization. *Chromosoma*. 1971;32(2).
  92. Veselinyová D, Mašlanková J, Kalinová K, Mičková H, Mareková M, Rabajdová M. Selected in situ hybridization methods: Principles and application. Vol. 26, *Molecules*. 2021.
  93. Jensen E. Technical review: In situ hybridization. *Anat Rec*. 2014;297(8).
  94. Wahl GM, Meinkoth JL, Kimmel AR. Northern and Southern Blots. *Methods Enzymol*. 1987;152(C).
  95. Martinho C, Lopez-Gomollon S. Detection of MicroRNAs by Northern Blot. *Methods Mol Biol*. 2023;2630.
  96. Gressner AM, Gressner OA. Northern blot. In: *Lexikon der Medizinischen Laboratoriumsdiagnostik*. 2018.

97. Rohde W, Sanger HL. Detection of complementary RNA intermediates of viroid replication by Northern blot hybridization. *Biosci Rep.* 1981;1(4).
98. Goldstein DR. Analyzing Microarray Gene Expression Data. *J Am Stat Assoc.* 2005;100(472).
99. Zvara , Kitajka K, Farag N, Pusks LG. Microarray technology. Vol. 59, *Acta Biologica Szegediensis.* 2015.
100. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. Vol. 17, *Nature Reviews Genetics.* 2016.
101. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Vol. 20, *Nature Reviews Genetics.* 2019.
102. Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Vol. 10, *Nature Reviews Genetics.* 2009.
103. Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R, et al. Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome Biol.* 2014;15(12).
104. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Vol. 14, *Bioinformatics and Biology Insights.* 2020.
105. Canzler S, Schor J, Busch W, Schubert K, Rolle-Kampczyk UE, Seitz H, et al. Prospects and challenges of multi-omics data integration in toxicology. Vol. 94, *Archives of Toxicology.* 2020.
106. Wyrick JJ, Young RA. Deciphering gene expression regulatory networks. Vol. 12, *Current Opinion in Genetics and Development.* 2002.
107. Glinos DA, Garborcauskas G, Hoffman P, Ehsan N, Jiang L, Gokden A, et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature.* 2022;608(7922).
108. Kleino I, Frolovait P, Suomi T, Elo LL. Computational solutions for spatial transcriptomics. Vol. 20, *Computational and Structural Biotechnology Journal.* 2022.
109. Ortiz C, Carle acute n M, Meletis K. Spatial Transcriptomics: Molecular Maps of the Mammalian Brain. Vol. 44, *Annual Review of Neuroscience.* 2021.
110. Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. An introduction to spatial transcriptomics for biomedical research. Vol. 14, *Genome Medicine.* 2022.
111. Park HE, Jo SH, Lee RH, Macks CP, Ku T, Park J, et al. Spatial Transcriptomics: Technical Aspects of Recent Developments and Their Applications in Neuroscience and Cancer Research. Vol. 10, *Advanced Science.* 2023.
112. Noble D. *The Music of Life: Biology Beyond the Genome.* Lavoisierfr. 2006;
113. Tan SL, Ganji G, Paeper B, Proll S, Katze MG. Systems biology and the host response to viral infection. Vol. 25, *Nature Biotechnology.* 2007.
114. Majeed A, Rauf I. Graph theory: A comprehensive survey about graph theory applications in computer science and social networks. Vol. 5, *Inventions.* 2020.
115. Koutrouli M, Karatzas E, Paez-Espino D, Pavlopoulos GA. A Guide to Conquer the Biological Network Era Using Graph Theory. Vol. 8, *Frontiers in Bioengineering and Biotechnology.* 2020.
116. Euler L. Leonhard Euler and the Koenigsberg Bridges. *Sci Am.* 1953;189(1).
117. Zhang P, Itan Y. Biological network approaches and applications in rare disease studies. Vol. 10, *Genes.* 2019.
118. Thomas JP, Modos D, Korcsmaros T, Brooks-Warburton J. Network Biology Approaches to Achieve Precision Medicine in Inflammatory Bowel Disease. Vol. 12, *Frontiers in Genetics.* 2021.
119. Randi M, Novi M, Plavi D. Milestones in graphical bioinformatics. Vol. 113, *International Journal of Quantum Chemistry.* 2013.
120. Gursoy A, Keskin O, Nussinov R. Topological properties of protein interaction networks from a structural perspective. In: *Biochemical Society Transactions.* 2008.
121. Avelino-Silva VI, Leal FE, Sabino EC, Nishiya AS, Da Silva Freire M, Blumm F, et al. Yellow fever vaccine viremia following ablative BM suppression in AML. *Bone Marrow Transplant* [Internet]. 2013;48(7):1008–9. Available from: <http://dx.doi.org/10.1038/bmt.2012.277>
122. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
123. Love MI, Anders S, Huber W. Differential analysis of count data - the DESeq2 package. Vol. 15, *Genome Biology.* 2014.

124. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43).
125. Bailey AL, Diamond MS. Hepatopathology of flaviviruses. Vol. 77, *Journal of Hepatology*. 2022.
126. Rajapakse S. Dengue shock. *Journal of emergencies, trauma, and shock. J Emerg Trauma Shock*. 2011;4(1).
127. Turtle L, Griffiths MJ, Solomon T. Encephalitis caused by flaviviruses. Vol. 105, *QJM*. 2012.
128. Howard-Jones AR, Pham D, Sparks R, Maddocks S, Dwyer DE, Kok J, et al. Arthropod-Borne Flaviviruses in Pregnancy. Vol. 11, *Microorganisms*. 2023.
129. LaBeaud AD, Bashir F, King CH. Measuring the burden of arboviral diseases: The spectrum of morbidity and mortality from four prevalent infections. *Popul Health Metr*. 2011;9.
130. Ye J, Zhu B, Fu ZF, Chen H, Cao S. Immune evasion strategies of flaviviruses. Vol. 31, *Vaccine*. 2013.
131. Pierson TC, Diamond MS. The continued threat of emerging flaviviruses. Vol. 5, *Nature Microbiology*. 2020.
132. Gallego Romero I, Pai AA, Tung J, Gilad Y. RNA-seq: Impact of RNA degradation on transcript quantification. *BMC Biol*. 2014;12.
133. Sonntag KC, Tejada G, Subburaju S, Berretta S, Benes FM, Woo TUW. Limited predictability of postmortem human brain tissue quality by RNA integrity numbers. *J Neurochem*. 2016;
134. Heng Y, Dubbelaar ML, Marie SKN, Boddeke EWGM, Eggen BJL. The effects of postmortem delay on mouse and human microglia gene expression. *Glia*. 2021;69(4).
135. Desjardins P, Bélanger M, Butterworth RF. Alterations in expression of genes coding for key astrocytic proteins in acute liver failure. In: *Journal of Neuroscience Research*. 2001.
136. Duffy J, Johnsen P, Ferris M, Miller M, Leighton K, McGilvray M, et al. Safety of a meningococcal group B vaccine used in response to two university outbreaks. *J Am Coll Heal*. 2017;65(6).
137. Oganessian A, Schäfer M, Lesh C. Acute appendicitis following the COVID-19 vaccine. *J Surg Case Reports*. 2022;2022(6).
138. Quint E, Kukeev I, Hazan I, Grupel D, Dukhno O, Osyntsov A, et al. Clinical characteristics of SARS-CoV-2 vaccine-related acute appendicitis. *Can J Surg*. 2023;66(3).
139. Reisinger KS, Block SL, Lazcano-Ponce E, Samakoses R, Esser MT, Erick J, et al. Safety and persistent immunogenicity of a quadrivalent human papillomavirus types 6, 11, 16, 18 L1 virus-like particle vaccine in preadolescents and adolescents: A randomized controlled trial. *Pediatr Infect Dis J*. 2007;26(3).
140. Kildegaard H, Ladebo L, Andersen JH, Jensen PB, Rasmussen L, Damkier P, et al. Risk of Appendicitis after mRNA COVID-19 Vaccination in a Danish Population. Vol. 182, *JAMA Internal Medicine*. 2022.
141. Mitchell J, Yue QY. Appendicitis as a possible safety signal for the COVID-19 vaccines. *Vaccine X*. 2021;9.
142. Theiler M, Smith HH. The Use of Yellow Fever Virus Modified By in Vitro Cultivation for Human Immunization. *J Exp Med [Internet]*. 1937;65(6):787–800. Available from: <http://www.jem.org/cgi/doi/10.1084/jem.65.6.787>
143. Liu T, Chambers TJ. Yellow Fever Virus Encephalitis: Properties of the Brain-Associated T-Cell Response during Virus Clearance in Normal and Gamma Interferon-Deficient Mice and Requirement for CD4 + Lymphocytes . *J Virol*. 2001;75(5).
144. de Andrade Gandolfi F, Estofolete CF, Wakai MC, Negri AF, Barcelos MD, Vasilakis N, et al. Yellow Fever Vaccine-Related Neurotropic Disease in Brazil Following Immunization with 17DD. *Vaccines*. 2023;11(2).
145. Guimard T, Minjolle S, Polard E, Fily F, Zeller H, Michelet C, et al. Short report: Incidence of yellow fever vaccine-associated neurotropic disease. *Am J Trop Med Hyg*. 2009;81(6).
146. Pelletier A-N, Tomazella M, de Carvalho K, Nicolau A, Marmoratto M, Silveira C, et al. Yellow fever disease severity is driven by an acute cytokine storm modulated by an interplay between the human gut microbiome and the metabolome. *medRxiv*. 2021.
147. Gu Y, Zuo X, Zhang S, Ouyang Z, Jiang S, Wang F, et al. The mechanism behind influenza virus

- cytokine storm. Vol. 13, *Viruses*. 2021.
148. Liu Q, Zhou YH, Yang ZQ. The cytokine storm of severe influenza and development of immunomodulatory therapy. Vol. 13, *Cellular and Molecular Immunology*. 2016.
  149. Allan SM, Rothwell NJ. Cytokines and acute neurodegeneration. Vol. 2, *Nature Reviews Neuroscience*. 2001.
  150. Schwerk C, Tenenbaum T, Kim KS, Schrotten H. The choroid plexus—a multi-role player during infectious diseases of the CNS. Vol. 9, *Frontiers in Cellular Neuroscience*. 2015.
  151. Marshall EM, Koopmans MPG, Rockx B. A Journey to the Central Nervous System: Routes of Flaviviral Neuroinvasion in Human Disease. Vol. 14, *Viruses*. 2022.
  152. Pellegrini L, Albecka A, Mallery DL, Kellner MJ, Paul D, Carter AP, et al. SARS-CoV-2 Infects the Brain Choroid Plexus and Disrupts the Blood-CSF Barrier in Human Brain Organoids. *Cell Stem Cell*. 2020;27(6).
  153. Kim SJ, Syed GH, Khan M, Chiu WW, Sohail MA, Gish RG, et al. Hepatitis C virus triggers mitochondrial fission and attenuates apoptosis to promote viral persistence. *Proc Natl Acad Sci U S A*. 2014;111(17).
  154. Javed F, Manzoor S. HCV non-structural NS4A protein of genotype 3a induces mitochondria mediated death by activating Bax and the caspase cascade. *Microb Pathog*. 2018;124.
  155. Koshiba T, Yasukawa K, Yanagi Y, Kawabata SI. Mitochondrial membrane potential is required for MAVS-mediated antiviral signaling. *Sci Signal*. 2011;4(158).
  156. Elesela S, Lukacs NW. Role of mitochondria in viral infections. Vol. 11, *Life*. 2021.
  157. Yang S, Gorshkov K, Lee EM, Xu M, Cheng YS, Sun N, et al. Zika Virus-Induced Neuronal Apoptosis via Increased Mitochondrial Fragmentation. *Front Microbiol*. 2020;11.
  158. Banks WA, Niehoff ML, Ponzio NM, Erickson MA, Zalzman SS. Pharmacokinetics and modeling of immune cell trafficking: quantifying differential influences of target tissues versus lymphocytes in SJL and lipopolysaccharide-treated mice. *J Neuroinflammation*. 2012;9.
  159. Moulson AJ, Squair JW, Franklin RJM, Tetzlaff W, Assinck P. Diversity of Reactive Astrogliosis in CNS Pathology: Heterogeneity or Plasticity? Vol. 15, *Frontiers in Cellular Neuroscience*. 2021.
  160. Pekny M, Pekna M, Messing A, Steinhäuser C, Lee JM, Parpura V, et al. Astrocytes: a central element in neurological diseases. Vol. 131, *Acta Neuropathologica*. 2016.
  161. Shigetomi E, Saito K, Sano F, Koizumi SC. Aberrant calcium signals in reactive astrocytes: A key process in neurological disorders. *Int J Mol Sci*. 2019;20(4).
  162. Chai Q, She R, Huang Y, Fu ZF. Expression of Neuronal CXCL10 Induced by Rabies Virus Infection Initiates Infiltration of Inflammatory Cells, Production of Chemokines and Cytokines, and Enhancement of Blood-Brain Barrier Permeability. *J Virol*. 2015;89(1).
  163. Terry RL, Getts DR, Deffrasnes C, van Vreden C, Campbell IL, King NJC. Inflammatory monocytes and the pathogenesis of viral encephalitis. Vol. 9, *Journal of Neuroinflammation*. 2012.
  164. Suárez I, Bodega G, Fernández B. Glutamine synthetase in brain: Effect of ammonia. *Neurochem Int*. 2002;41(2–3).
  165. Niknahad H, Jamshidzadeh A, Heidari R, Zarei M, Ommati MM. Ammonia-induced mitochondrial dysfunction and energy metabolism disturbances in isolated brain and liver mitochondria, and the effect of taurine administration: Relevance to hepatic encephalopathy treatment. *Clin Exp Hepatol*. 2017;3(3).
  166. Bril F, Al Diffalha S, Dean M, Fettig DM. Autoimmune hepatitis developing after coronavirus disease 2019 (COVID-19) vaccine: Causality or casualty? Vol. 75, *Journal of Hepatology*. 2021.
  167. Shroff H, Satapathy SK, Crawford JM, Todd NJ, VanWagner LB. Liver injury following SARS-CoV-2 vaccination: A multicenter case series. Vol. 76, *Journal of Hepatology*. 2022.
  168. Garrido I, Lopes S, Simões MS, Liberal R, Lopes J, Carneiro F, et al. Autoimmune hepatitis after COVID-19 vaccine – more than a coincidence. *J Autoimmun*. 2021;125.
  169. Efe C, Kulkarni A V., Terziroli Beretta-Piccoli B, Magro B, Stättermayer A, Cengiz M, et al. Liver injury after SARS-CoV-2 vaccination: Features of immune-mediated hepatitis, role of corticosteroid therapy and outcome. *Hepatology*. 2022;76(6).
  170. Lee M-H, Perl DP, Nair G, Li W, Maric D, Murray H, et al. Microvascular Injury in the Brains of Patients with Covid-19. *N Engl J Med*. 2021;384(5).

171. Thakur KT, Miller EH, Glendinning MD, Al-Dalahmah O, Banu MA, Boehme AK, et al. COVID-19 neuropathology at Columbia University Irving Medical Center/New York Presbyterian Hospital. *Brain*. 2021;144(9).
172. Wenzel J, Schwaninger M. How COVID-19 affects microvessels in the brain. Vol. 145, *Brain*. 2022.
173. Sefik E, Qu R, Junqueira C, Kaffe E, Mirza H, Zhao J, et al. Inflammasome activation in infected macrophages drives COVID-19 pathology. *Nature*. 2022;606(7914).
174. Wenzel J, Lampe J, Müller-Fielitz H, Schuster R, Zille M, Müller K, et al. The SARS-CoV-2 main protease M<sub>pro</sub> causes microvascular brain pathology by cleaving NEMO in brain endothelial cells. *Nat Neurosci*. 2021;24(11).
175. Camacho LAB, da Silva Freire M, da Luz Fernandes Leal M, de Aguiar SG, Pereira do Nascimento J, Iguchi T, et al. Immunogenicity of WHO-17D and Brazilian 17DD yellow fever vaccines: A randomized trial. *Rev Saude Publica*. 2004;38(5).
176. Xu Y, Kulkosky J, Acheampong E, Nunnari G, Sullivan J, Pomerantz RJ. HIV-1-mediated apoptosis of neuronal cells: Proximal molecular mechanisms of HIV-1-induced encephalopathy. *Proc Natl Acad Sci U S A*. 2004;101(18).



## **Anexos**

OPEN

# Origin of the São Paulo Yellow Fever epidemic of 2017–2018 revealed through molecular epidemiological analysis of fatal cases

Marielson dos Passos Cunha<sup>1\*</sup>, Amaro Nunes Duarte-Neto<sup>2</sup>, Shahab Zaki Pour<sup>1</sup>, Ayda Susana Ortiz-Baez<sup>1,9</sup>, Jiří Černý<sup>3</sup>, Bárbara Brito de Souza Pereira<sup>1</sup>, Carla Torres Braconi<sup>1</sup>, Yeh-Li Ho<sup>4</sup>, Beatriz Perondi<sup>5</sup>, Jaques Sztajnbock<sup>6</sup>, Venancio Avancini Ferreira Alves<sup>2</sup>, Marisa Dolhnikoff<sup>2</sup>, Edward C. Holmes<sup>7</sup>, Paulo Hilário Nascimento Saldiva<sup>2</sup> & Paulo Marinho de Andrade Zanotto<sup>1,8\*</sup>

The largest outbreak of yellow fever of the 21<sup>st</sup> century in the Americas began in 2016, with intense circulation in the southeastern states of Brazil, particularly in sylvatic environments near densely populated areas including the metropolitan region of São Paulo city (MRSP) during 2017–2018. Herein, we describe the origin and molecular epidemiology of yellow fever virus (YFV) during this outbreak inferred from 36 full genome sequences taken from individuals who died following infection with zoonotic YFV. Our analysis revealed that these deaths were due to three genetic variants of sylvatic YFV that belong the South American I genotype and that were related to viruses previously isolated in 2017 from other locations in Brazil (Minas Gerais, Espírito Santo, Bahia and Rio de Janeiro states). Each variant represented an independent virus introduction into the MRSP. Phylogeographic and geospatial analyses suggested that the virus moved around the peri-urban area without detectable human-to-human transmission, and towards the Atlantic rain forest causing human spill-over in nearby cities, yet in the absence of sustained viral transmission in the urban environment.

Yellow fever virus (YFV) is an enveloped virus of the family *Flaviviridae* (genus *Flavivirus*) with a single stranded, positive-sense RNA genome of approximately 11 kb that encodes a single polyprotein cleaved into three structural (capsid (C), membrane (M) and envelope (E)) and seven non-structural proteins (NS1-NS5)<sup>1</sup>. The virus comprises a single serotype with four genotypes: (i) East Africa, (ii) West Africa, (iii) South American I and (iv) South American II<sup>2,3</sup> that may have diverged around several thousand years before present<sup>4,5</sup> with a possible origin in the African continent<sup>5,6</sup>. Historical evidence points to a YFV introduction in the Americas around the 17<sup>th</sup> century, possibly due to the slave trade<sup>3,5-7</sup>. After its introduction, YFV established both urban and sylvatic cycles<sup>7,8</sup>, and several urban outbreaks have been reported in Brazil since the 17<sup>th</sup> century<sup>9</sup>. The circulation of YFV in the urban cycle in the American continent was initially mitigated by curbing the infestation of *Aedes aegypti* and later with the advent of an effective vaccine in the early 20<sup>th</sup> century<sup>7,10</sup>, with considerable success. As a consequence, the last urban outbreak of YFV was officially reported in 1942 in Brazil<sup>9</sup>. After the reintroduction of *A. aegypti* in the 1970s<sup>11,12</sup> the virus remained, until recently, largely in sylvatic environments in the Americas, infecting

<sup>1</sup>Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of São Paulo, São Paulo, Brazil. <sup>2</sup>Pathology Department, Clinical Hospital, Faculty of Medicine, University of São Paulo, São Paulo, Brazil. <sup>3</sup>Faculty of Tropical AgriSciences, Czech University of Life Sciences in Prague, Prague, Czech Republic. <sup>4</sup>Intensive Care Unit, Division of Clinical Infectious and Parasitic Diseases, Clinical Hospital, Faculty of Medicine, University of São Paulo, São Paulo, Brazil. <sup>5</sup>Yellow Fever Crisis Committee, Clinical Hospital, Faculty of Medicine, University of São Paulo, São Paulo, Brazil. <sup>6</sup>Institute of Infectology Emílio Ribas, São Paulo, Brazil. <sup>7</sup>School of Life and Environmental Sciences & Sydney Medical School, The University of Sydney, Sydney, Australia. <sup>8</sup>Scientific Platform Pasteur - USP, São Paulo, Brazil. <sup>9</sup>Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life & Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia. \*email: [marieltondospassos@gmail.com](mailto:marieltondospassos@gmail.com); [pzanotto@usp.br](mailto:pzanotto@usp.br)

non-human primates (NHPs) with sporadic cases in susceptible human hosts. The main vectors of YFV in the sylvatic cycle are mosquitoes of the genera *Haemagogus* and *Sabethes*<sup>13,14</sup>.

In 2014, intense enzootic activity of YFV was detected in Mato Grosso do Sul and Goiás states that adjoin the Amazon region of Brazil<sup>15,16</sup>. YFV carried by infected monkeys kept moving in a general southeasterly direction, and in 2016 cases were reported in Minas Gerais, reaching epidemic proportions in 2017, during which cases were also reported in the states of Rio de Janeiro, Espírito Santo and Bahia<sup>17–21</sup>.

Between January 2016 and January 2018, seven countries and regions of the Americas reported cases of yellow fever in their territories (Bolivia, Brazil, Colombia, Ecuador, French Guiana, Peru and Suriname), with the highest indices in Brazil. In early 2018 an unusually large increase in the number of confirmed cases was observed in the state of São Paulo<sup>22</sup>. A peak of notified human cases was reached in January 2018<sup>23</sup>. This was the largest outbreak registered in 21<sup>st</sup> century in the most populated state of Brazil, including the densely populated metropolitan region of São Paulo city (MRSP), which is the largest conurbation in the southern hemisphere with around 23 million inhabitants. Until 2018, vaccination was not generally recommended in MRSP because YFV had been absent in recent decades. Hence, most of the population in the area was susceptible to YFV and autochthonous cases were reported<sup>24,25</sup>. Due to the outbreak in São Paulo, vaccination campaigns were initiated for resident populations, starting in northern peri-urban settings bordering forested and rural areas, where cases of YFV were previously reported. Subsequently, vaccination was extended to the whole urban population as well as to all inhabitants of the São Paulo state as the epidemic expanded<sup>22</sup>. As this is the first time in the 21<sup>st</sup> century that cases of YFV have appeared in the MRSP, we sought to characterize the circulating viruses and establish their origin by studying their evolution and phylogeography based on samples taken from patients who died during the 2017–2018 outbreak.

## Material and Methods

**Ethical statement.** The human autopsies analyzed in this study were performed after obtaining informed consent of the family members and following the protocol approved by the research ethics committee of the Clinical Hospital of the University of São Paulo School of Medicine (HCFMUSP) (CAPPesq #426.643). All the methods were performed in accordance with the relevant guidelines and regulations of the ethics committee of the HCFMUSP following the approval CAPPesq #426.643. All participating families were asked to sign a free and informed consent form, authorizing the autopsy and all experiments performed with the collected tissues. All laboratory procedures listed below were performed in a biosafety level (BSL)-2 laboratory, in accordance with the Brazilian standards of the Ministry of Health for Biological Agents Risk Classification<sup>26</sup>.

**Patients and samples.** Overall, we analyzed 81 patients 67 of whom were confirmed to have died following YFV infection. We successfully acquired 36 genome sequences from the 67 yellow fever deaths, with the remaining samples being of insufficient quality to obtain YFV genomes at the necessary coverage. The suspected case definition of YFV infection was established by the Brazilian Ministry of Health and the Health Department of São Paulo State and included patients with sudden onset high fever associated with jaundice and/or hemorrhage who had lived or had visited areas with YFV epizootics (*i.e.*, clusters of infections in non-human primates (NHP) or isolation of YFV in vectors), regardless of the vaccine status for YFV, during the preceding 15 days. Confirmed cases had compatible clinical presentation and laboratory confirmation by at least one of the following methods: (i) serum IgM positive (MAC-ELISA); (ii) detection of YFV-RNA by qRT-PCR in blood samples; (iii) virus isolation; (iv) histopathology compatible with YFV hepatitis with detectable antigen in tissues by immunohistochemistry technique. All cases received the definitive laboratorial diagnosis of YFV by the Adolfo Lutz Institute (IAL), the State Reference Laboratory. Previous exposure or co-infection by Hepatitis A virus (HVA), B (HBV), C (HVC), Cytomegalovirus (CMV), Herpes virus (HSV), Dengue virus (DENV), Chikungunya virus (CHIKV), Human Immunodeficiency virus type 1 (HIV-1), leptospirosis and other non-infectious diseases etiologies for acute hepatitis were accessed and cases were excluded following clinical diagnostic methods. Epidemiological, clinical (including demographic data, preexisting medical conditions, clinical signs and symptoms and in-hospital follow-up until death) and other laboratory features were collected from the medical charts.

**Autopsy protocol and tissue processing.** The Service of Verification of Deaths of the Capital - USP investigated deaths due to yellow fever from December/2017 to April/2018. Autopsies were performed following the Letulle technique, where all the organs were removed *en masse* (one block), requiring dissection organ by organ to exam them individually. Briefly, the dissection was performed in the following organs: (i) heart; (ii) lung; (iii) brain; (iv) kidney; (v) spleen; (vi) pancreas; and (vii) liver.

**Molecular characterization.** Nucleic acid extraction from all collected tissues was performed using the TRIzol<sup>®</sup> reagent (Life Technologies, Carlsbad, CA, USA) and carried out according to the manufacturer's instructions. Molecular detection of YFV was performed with the use of the AgPath-ID One-Step RT-PCR Reagents (Ambion, Austin, TX, USA) with specific primers/probe previously described<sup>27</sup>. To identify cases of adverse vaccine response (*i.e.*, fatal cases associated with the vaccine virus) we used specific primers/probe specific for the vaccine virus<sup>28</sup>. qRT-PCR reactions consisted of a step of reverse transcription at 45 °C for 10 min, enzyme activation at 95 °C for 10 min, and 40 cycles at 95 °C for 15 s and 60 °C for 45 s for hybridization and extension using the ABI7500 equipment (Thermo Fisher Scientific, Waltham, MA, USA).

**Sequencing and viral genome assembly.** Based on the RNA viral concentration, total RNA were extracted from the liver tissues using the TRIzol<sup>®</sup> reagent (Life Technologies, Carlsbad, CA, USA). Subsequently, the RNA was purified with DNase I and concentrated using the RNA Clean and Concentrator<sup>™-5</sup> kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. The paired-end RNA libraries were constructed and validated using the TruSeq Stranded Total RNA HT sample prep kit (Illumina, San Diego, CA,

USA). Sequencing was done at the Core Facility for Scientific Research – University of São Paulo (CEFAP-USP/ GENIAL) using the Illumina NextSeq platform. Each sample was barcoded individually, which allowed separation of reads for each patient. Short unpaired reads and low-quality bases and reads were removed using Trimmomatic version 0.36 (LEADING:20 TRAILING:20 SLIDINGWINDOW:4:25 MINLEN:36)<sup>29</sup>. Consensus genomes were assembled with paired-end reads using Bowtie2 v.2.3.4.3<sup>30</sup> using default parameters.

**Data sets.** All full genomic sequences available from YFV that contained information on location and date of isolation were recovered from the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/genbank/>) website. Sequences were aligned to our 36 new YFV genomes (Supplementary Table 1) using Clustal Omega v.1.2.4<sup>31</sup>. A list of the sequences used is available in Supplementary Table 2. Recombinant sequences were screened using all algorithms implemented in RDP4 program (RDP, GENECONV, BootScan, MaxChi, Chimaera, Siscan and 3Seq) using the default settings<sup>32</sup>. No evidence for recombination was detected. Sequences containing long contiguous stretches of undefined nucleotides were excluded. A final alignment of complete genome sequences was manually inspected and edited using the program AliView v.1.18<sup>33</sup>. After preliminary phylogenetic analyses, the master alignment comprising 135 full-length, curated sequences encoding the complete viral polyprotein (dataset-1) (Supplementary Table 2) was subdivided into two data sets for further analysis: (i) a data set containing 98 genomes of the SA1 and SA2 genotypes from the Americas (dataset-2); and (ii) 74 sequences from 2017 and 2018 sampled from the states of Minas Gerais, Espírito Santo, Bahia, Rio de Janeiro and São Paulo (dataset-3) (Supplementary Tables 2 and 3). All alignments are available in the Supplementary Data and on GitHub (<https://github.com/MarieltonCunha/ViralDiversity/>).

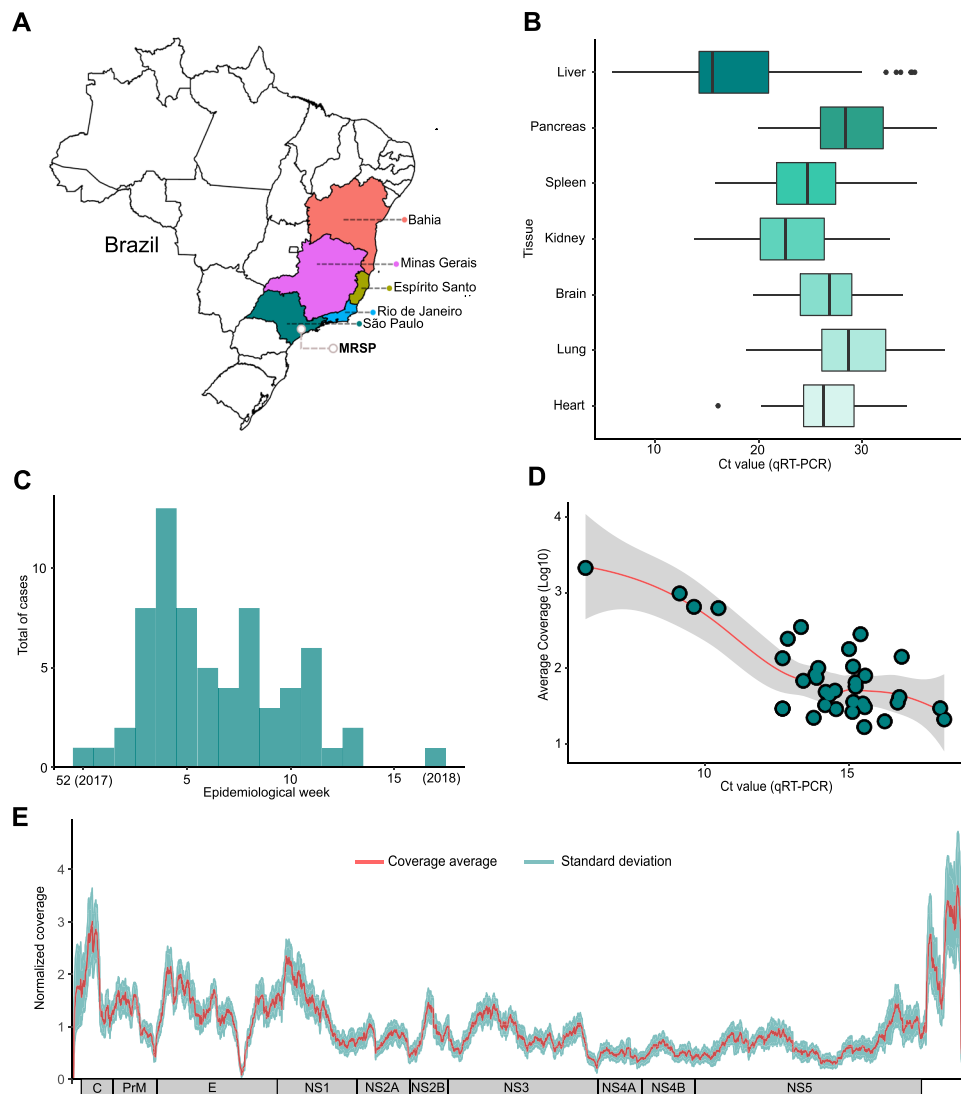
**Phylogenetic analysis.** Phylogenetic trees of YFV based on full-length, curated coding sequences for all the data sets were estimated using the Maximum Likelihood (ML) method implemented in IQ-TREE 1.5.5<sup>34</sup> with automatic model selection by ModelFinder and using the Bayesian Information Criterion (BIC)<sup>35</sup>. The robustness of the groupings observed was assessed using 1,000 non-parametric bootstrap replicates. ML and Bayesian maximum clade credibility (MCC) trees (see below) were visualized and plotted using FigTree v.1.4.3<sup>36</sup>. All taxon labels for sequences used in this work are presented in the format: genotype/accession number/strain name/local of isolation/date of isolation. We explored the temporal signal (*i.e.*, molecular clock structure) and quality of our data set using TempEst v.1.5.1<sup>37</sup>.

**Phylodynamics and phylogeographic analysis.** The spatio-temporal evolution of YFV spread was inferred within a Bayesian framework as implemented in BEAST v.1.10.1<sup>38</sup>. An initial descriptive summary of the demographic history of YFV was approximated using the Bayesian SkyGrid coalescent model<sup>39</sup> and revealed no significant variation in genetic diversity (a marker of population size) during the period of our analysis. Based on previous estimates of evolutionary dynamics of related YFV<sup>17,40</sup>, we tested uncorrelated relaxed molecular clocks assuming a log-normal distribution, in combination with constant size, exponential and logistic growth demographic models (Supplementary Tables 4 and 5). Phylogeographic patterns and parameters were estimated using the Bayesian inference through Markov chain Monte Carlo (MCMC) run for 50 million states, sampling every 5,000 states with a 10% burn-in. Convergence and the effective sample size (ESS) > 200 were examined using Tracer v.1.7.1<sup>41</sup>. Likewise, the maximum clade credibility tree (MCC) was visualized and edited in FigTree v.1.4.3<sup>36</sup>. We recorded the time to the most recent common ancestor (tMRCA) and their 95% Bayesian credible intervals (HPD) for the MCC tree. To calculate the log marginal likelihood for molecular clock and demographic model selection, we used the path sampling (PS) and the stepping-stone (SS) sampling approaches by running 100 path steps of 1 million iterations each<sup>42</sup>. The spatiotemporal spread of YFV was visualized and plotted with SPREAD3<sup>43</sup>. XML input files for BEAST are available in the Supplementary Data and on GitHub (<https://github.com/MarieltonCunha/ViralDiversity/>).

**Geopositioning of samples.** To analyze the geographical proximity among fatal human and NHP cases we calculated the spatial distances between all cases using available geolocation information. We geolocated only those fatal human and NHP YFV cases that occurred in the MRSP (47.0–46.2 S, 23.9–23.1 W), using the available data on patient residence and day of death. NHP cases were included only for those where coordinates for the place of where carcasses were found was available. For fatal NHP cases, the date the carcass was found was assumed to be the day of death, although death may have taken place a few days before. Distances between the human and NHP fatal YFV cases were calculated based on the available coordinates. Geographic pairwise distance matrices among all YFV cases (in kilometers) were clustered using the neighbor joining algorithm available in the PHYLIP v.3.695 package<sup>44</sup>, this enabled us to produce a dendrogram based on geolocation information.

## Results

**Epidemiological surveillance of YFV in São Paulo, 2017–2018.** From January to August 17, 2018, the State of São Paulo reported 3028 suspected cases of yellow fever, 537 (17.7%) of which were confirmed, with 498 (92.7%) autochthonous cases and 35 (6.5%) imported from other states<sup>45</sup>. Of the 498 autochthonous cases, 176 died, resulting in a mortality frequency of 35.4%<sup>45</sup>. Despite the magnitude of the outbreak in São Paulo, little is known about the epidemiological, genetic and evolutionary characteristics of the virus circulating in the state. Accordingly, among all patients who died with suspected YFV infection between December 2017 and April 2018, we focused on 81 cases identified through the service of verification of deaths of the capital - USP (SVOC-USP) in the city of São Paulo (Fig. 1A). Our qRT-PCR results indicated that 67/81 (82.7%) individuals had been infected by YFV, while five were shown by qRT-PCR to only carry the vaccine strain YFV-17DD alone, suggesting that their death was associated with an adverse response to the vaccine as previously reported<sup>46–48</sup>, and nine were negative for YFV infection in all tissues tested (Fig. 1B,C). All 67 confirmed YFV deaths were due to complications of fulminant yellow fever hepatitis, with hepatic encephalopathy, severe coagulopathy, bleeding (mainly

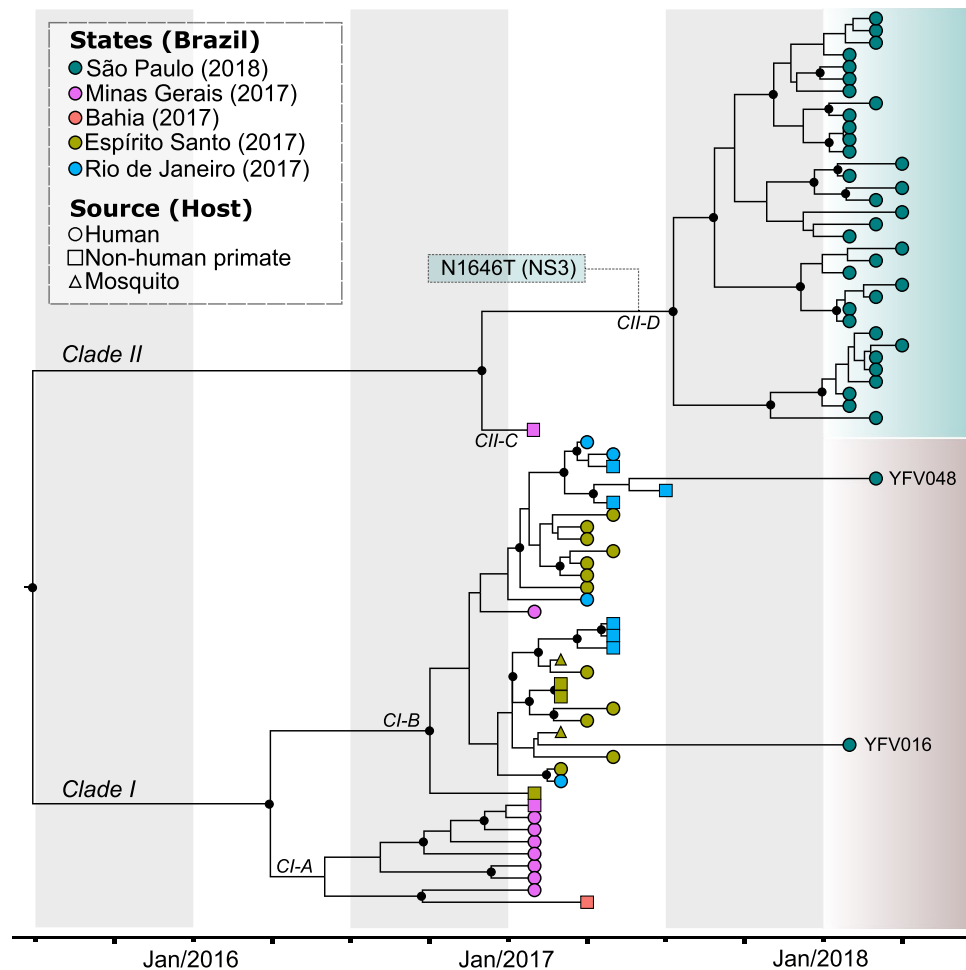


**Figure 1.** The current outbreak of yellow fever virus in Brazil (2016–2018). **(A)** Brazilian states with YFV cases recorded and sequenced in humans, non-human primates (NHP) and mosquitoes between 2017–2018. A grey circle marks the metropolitan region of São Paulo (MRSP). **(B)** Cycle threshold according to each of the 7 tissues analyzed for positive patients. Boxplots represent the 75th percentile, median, 25th percentile and the whiskers extend to the highest and lowest value in the 1.5x interquartile range. The different colors represent the different tissues analyzed. **(C)** Total cases recorded represented sylvatic cases of YFV (qRT-PCR positive cases) during the epidemiological weeks covered by the study (week 52 of 2017 to week 17 of 2018). **(D)** Relationship between the average coverage and the Ct values obtained for each sequenced sample. The data indicate that we obtained the expected direct inverse relationship between Ct and coverage parameters, as indicated by the trend line. **(E)** Combined coverage (normalized by the sample average) along all 36 sequenced YFV genomes generated in this study.

gastrointestinal, pulmonary and/or cerebral hemorrhages), renal dysfunction and secondary infections. We were able to successfully sequence the full YFV genome from 36 of these patient samples.

All of our cases were sampled in 17 localities in the São Paulo state, from which 16 localities had fatal cases due to YFV (Supplementary Table 6). Our molecular diagnostics indicated a peak of cases during the first epidemiological weeks of 2018, particularly at the end of January, coinciding with official cases notifications data (Fig. 1C). The median age of people with confirmed infection was 49.12 years (range 16–87) and were mainly male (82.09–55/67).

**Genomic surveillance.** Because detailed spatio-temporal resolution of viral evolution often relies on a few nucleotide differences among otherwise closely related viruses, complete genomes with high coverage for each base position are a prerequisite for robust inference. Therefore, to select the appropriate clinical specimens for viral sequencing, we analyzed cycle threshold (Ct) data from qRT-PCR from viral RNA in seven distinct tissues/organs (heart, lung, brain, kidney, spleen, pancreas and liver) to choose samples with the lowest possible Cts.



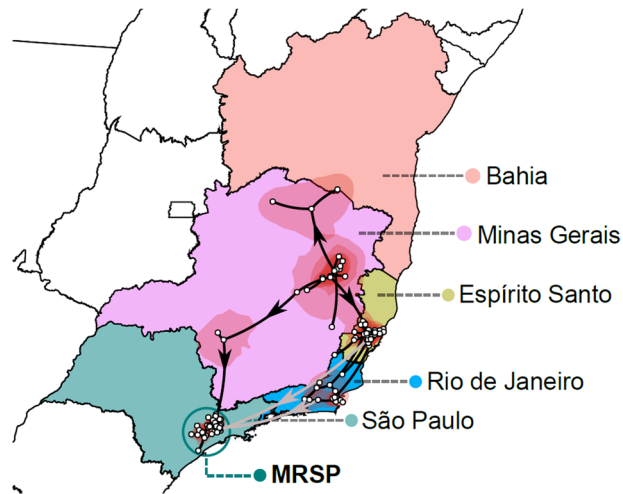
**Figure 2.** Time-stamped, MCC tree of YFV South American genotype I in Brazil recovered under the logistic-lognormal demographic model. The different colours indicate samples from different locations. The black circles represent posterior support upper than 0.7. The single synapomorphic change observed in Clade II [N1646T (NS3)] is shown in the box over the branch leading to Clade II-D. The three distinct introductions in the metropolitan region of São Paulo (MRSP) are shown (See also Fig. 3).

In general, all tissues had normally distributed Ct values, with the exception of the liver, which had a moderately asymmetrical distribution and a deviation to lower Ct values, and hence generally inferior to other tissues (Fig. 1B). In total, we obtained 36 complete YFV genomes from the 67 positive patients (Fig. 1D,E). All sequences of the current outbreak belonged to the South American I genotype (Supplementary Fig. 1), and were related with sequences previously isolated in neighboring states in 2017 (Fig. 2) with no evidence of recombination. Based on the phylogenetic analysis, we could infer at least three distinct introductions of YFV in the MRSP: (i) A major clade (34 genomes) in the northwest of the MRSP coming from Minas Gerais due to NHP movement, and likely emerging between April 2017–October 2017 (95% HPD; mean - July 2017) (Fig. 3 and Supplementary Tables 4 and 5), (ii) one virus lineage from a case from Espírito Santo (Patient 16), and (iii) one from a case from Rio de Janeiro (Patient 48) (Fig. 2). Importantly, our patient's records indicated the two single introductions were due to people visiting enzootic locations in these states and did not appear to have caused detectable additional cases in the MRSP.

**Origin of the 2016–ongoing Yellow Fever virus outbreak.** Phylogenetic (Supplementary Figs. 1 and 2) and phylogeographic (Figs. 2, 3 and Supplementary Tables 4 and 5) analyses of samples from the 2017–2018 YFV outbreak allowed us to reveal the origin and spread of YFV in the Southeast and Northeast region of Brazil. In particular, there was evidence of two distinct zoonotic clades (Clade I and II) that likely separated in Minas Gerais (location posterior support of 0.8) between November 2013–June 2016 (95% HPD; mean date of June 2015). The mean rate of Clade I and II migration during the whole sampled period 2017 to 2018 was approximately 3.3 km/day (95% HPD = 2.25–4.37 km/day) with a mean evolutionary rate of  $9.85 \times 10^{-4}$  nucleotide substitutions per site, per year (subs/site/year) (95% HPD =  $6.52 \times 10^{-4}$  –  $1.35 \times 10^{-3}$  subs/site/year). We now describe these two clades in more detail.

**Clade I.** Clade I divided into two smaller clades (CI-A and CI-B) in 2016 (95% HPD of divergence time = July 2015 – September 2016) and likely in Minas Gerais (location posterior support of 0.82) (Fig. 2). CI-A then





**Figure 3.** Highest posterior probability migration paths for the YFV Clades I and II from 2016 to 2018 towards the metropolitan region of São Paulo (MRSP), based on the analysis of 74 complete genomes. Although the sample size is small such that inferences should be made with caution, three distinct introductions in the MRSP are shown and strongly supported. The spatiotemporal spread was visualized with SPREAD3.

diversified and moved and into peri-urban and forested regions in the state of Minas Gerais, causing an outbreak after January 2017, then moving onto Bahia. In contrast, Clade CI-B likely diversified in the forest region in the border between Minas Gerais and Espírito Santo, also in 2016, and then moved to Espírito Santo and Rio de Janeiro, causing in both states an outbreak during the first part of 2017. Two YFV patients who died in 2018 and resided in the MRSP had visited Espírito Santo (Patient 16) and Rio de Janeiro (patient 48). Fittingly, the virus phylogeny showed that their posthumous viral samples were nested among isolated viruses from the areas they visited (Fig. 2). These results indicated that CI-B was circulating until early 2018.

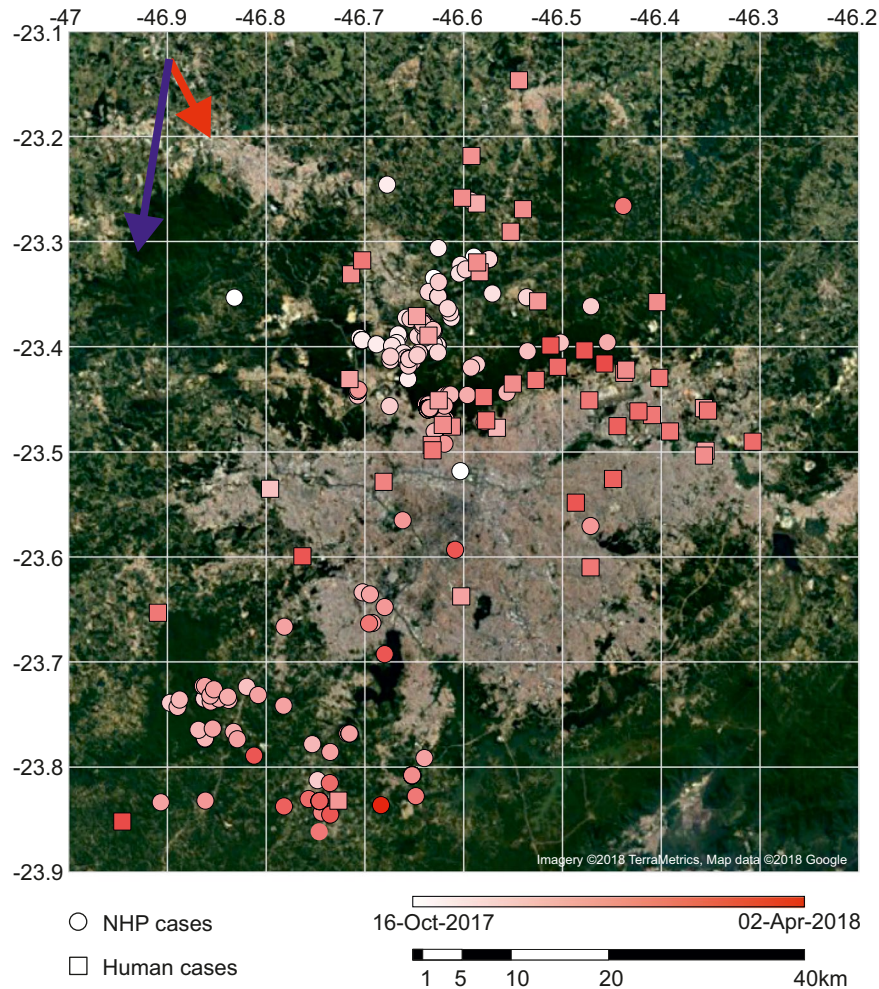
**Clade II.** This clade caused the majority of the deaths in the MRSP (Fig. 2). It diverged into Clades CII-C and CII-D in the state of Minas Gerais, with a location posterior support of 0.87, near the border with São Paulo between June 2016 - January 2017 (95% HPD; mean - December 2016) (Fig. 3). Subsequently, CII-D moved towards the MRSP, causing epizootics beginning between April 2017–October 2017 (95% HPD; mean - July 2017) (Supplementary Fig. 3) in forest parks (Horto Florestal and Cantareira State Park) that form a belt around the Northern part of the MRSP (Fig. 4). It is noteworthy that our inferred dates correspond well with the reported official cases of YFV cases in NHP and humans (Fig. 4). It is also notable that CII-D is also defined by a unique synapomorphic substitution (N1646T) in the NS3 gene that is not present in CII-C and Clade I viruses (Fig. 2).

**Geopositioning analysis.** In total, 230 NHP carcasses were collected in the MRSP. Of these, 136 were members of the genus *Alouatta* (howler monkeys), 14 were *Callithrix* genus (marmosets), and five were *Cebus* genus (capuchin monkeys). The species identity of the remaining 75 carcasses were not determined (Fig. 4) (data provided by the Adolfo Lutz Institute). Analysis of spatio-temporal data showed that the YFV outbreak progressed in different directions in humans and NHPs (Figs. 4 and 5). While the outbreak in NHPs had a tendency to move in a south-southwest direction, in humans the outbreaks in a southeast direction (Figs. 4 and 5).

Several geographically well-defined clusters can be observed in the dendrogram inferred from the pairwise geographic distances matrix among all YFV cases (Fig. 5). Two areas of intense epizootics were inferred in the north and southwest forested areas around the MRSP. We also inferred a large cluster of cases of NHP and humans in the northern region, Cantareira and Horto Florestal State parks, spreading to the nearby towns of Mairiporã and Guarulhos, where most of the human and NHP cases were reported. Another cluster represents NHP cases from the southwestern of the MRSP, around Cotia, where the second most affected NHP population was present. Hence, the most striking finding of this analysis was that most human cases occurred close to both the NHP cases and the forested belt around the MRSP.

## Discussion

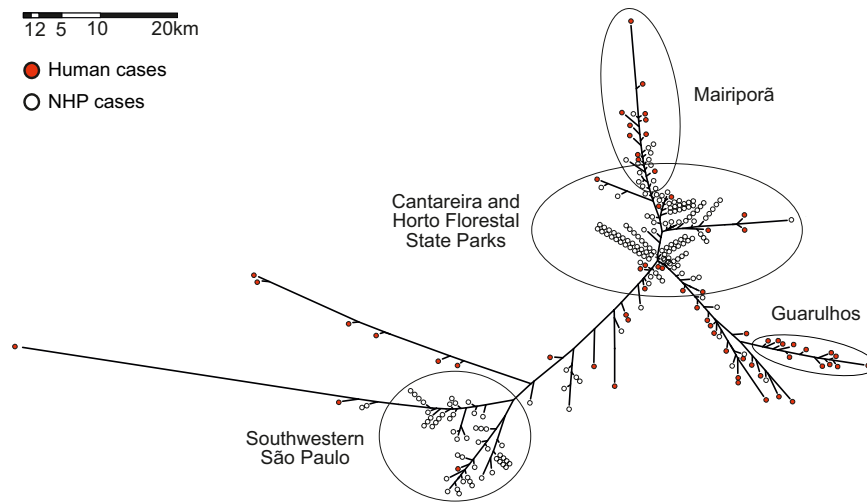
We describe the outbreak of YFV in the MRSP, Brazil, in 2016–2018, particularly its origin and how the virus diversified and moved around the largest conurbation in the southern hemisphere carried by NHP, killing 176 people during 2018 in the process<sup>45</sup>. All the isolates from São Paulo belonged to the South American I genotype and formed a single monophyletic group along with viruses (comprising Clades I and II) that also circulated in 2016–2017 in the states of Minas Gerais, Bahia, Rio de Janeiro and Espírito Santo<sup>17,19,40</sup>. Several synapomorphic mutational changes in different genes were previously reported by our group<sup>19</sup>, and here we report a synapomorphy (N1646T) in the protease NS3 gene shared by all CII-D. The mean evolutionary rate for all the YFV sequences of the Brazilian outbreak (2017–2018) was  $9.85 \times 10^{-4}$  subs/site/year, and hence compatible with those previously estimated for YFV and for other flaviviruses<sup>4,5,49</sup>.



**Figure 4.** Spatial distribution of YFV deaths through time in non-human primates (NHP) and humans. Arrows indicate the general trend of movement around the metropolitan region of São Paulo (MRSP) estimated from distance matrices (see Fig. 5). The earliest cases in NHP are shown in the north, and later in the south and northeast of the MRSP. Most human cases are near sites with reported deaths of NPH, confirmed to be caused by YFV. The outbreak appears to have been confined mostly near the forested belt around the MRSP, contrasting with the almost empty, heavily urbanized center. Cardinal points are aligned according to the main axis of the page, (e.g., top being north, etc.). The figure was created by plotting the coordinates of reported cases to a satellite image available from Google Maps ([google.com/maps](http://google.com/maps)) as background.

The current Brazilian outbreak began in the state of Minas Gerais in June 2015, with all viruses sampled from 2017 belonging to a single monophyletic group that diverged into two main clades (Clade I and II), and indicative of a single introduction of the virus in the region. These observations are supported by other molecular epidemiological studies conducted in Brazil from 2016<sup>17–19,40,50</sup>. Although previous evolutionary studies point to an origin of the virus in Venezuela<sup>40</sup>, epidemiological monitoring carried out by the Brazilian Ministry of Health suggest a likely origin of the Brazilian outbreak in 2014, with confirmed epizootics in the transitional area between the Amazon and the Cerrado biomes (with most of the confirmed cases occurring in the states of Goiás and Mato Grosso do Sul)<sup>15,16,21</sup>. This region was the probable link between the Amazon basin and the state of Minas Gerais, located in southeastern Brazil. It is likely that the numbers of human cases in this region were not high due to the vaccine coverage there<sup>51</sup>. The viral invasion into southeast Brazil, associated with the rapid spatial spread of the virus (estimated here at a mean of rate 3.3 km/day), caused the virus to circulate in important fragments of the Atlantic Forest near the peri-urban areas of the main Brazilian megacities (notably São Paulo and Rio de Janeiro), and led to a marked increase in the number of cases during the outbreak. In the MRSP, the virus (Clade CII-D) was introduced, maintained and spread in the sylvatic transmission cycle, with occasional cases of infection in humans between April 2017 and October 2017, with the interstate border between São Paulo and Minas Gerais as the route of introduction. In São Paulo state, the routes of viral dispersion included only interconnected forested, corridors linked to peri-urban regions. The patients studied here were mainly unvaccinated adult males that had contact with the sylvatic environment or lived nearby. No autochthonous cases were documented in the central region of the city of São Paulo. Importantly, the MRSP cases reduced in numbers as the populations of NHP collapsed and with vaccination campaigns in areas classified as at risk<sup>52</sup>.





**Figure 5.** Neighbor joining tree calculated from pairwise geolocation distances among all the non-human primates and human cases available from the metropolitan region of São Paulo (MRSP).

The introduction and establishment of the YFV Clade II-D in the state of São Paulo can be further explained by environmental factors, including: (i) mosquitoes of the genus *Haemagogus* are abundant in the forested areas of the state of São Paulo<sup>53,54</sup> and were the primary vectors in the YFV outbreak occurred in Brazil, 2016–2018<sup>55</sup>; (ii) NHPs are found in areas of the Atlantic Forest and are susceptible and responsible for the maintenance of the virus in the sylvatic cycle<sup>56–58</sup>, and (iii) the regions affected by the current outbreak had low vaccine coverage<sup>51</sup>. Our findings support previous work indicating that the outbreak of 2016–2018 (sampled in the states of Minas Gerais, Bahia, Espírito Santo and Rio de Janeiro), occurred in a sylvatic environment with occasional infections in humans<sup>17</sup>.

Importantly, we also recorded two introductions of YFV Clade I-B detected in patients who travelled to Espírito Santo and Rio de Janeiro - both states that experienced significant circulation of this virus lineage in 2018. In both these states an increase in the number of YFV notifications was reported in 2017 across successive epidemic periods, showcasing the maintenance of epizootic YFV. In addition, we highlighted the extent of viral movement, such as observed in cases imported from Brazil by other countries<sup>59</sup>, largely facilitated by rapid human movement such as those resulting from air travel<sup>60</sup>.

In contrast to other arboviruses in Brazil such as dengue virus, in which continuous reintroductions are responsible for keeping the virus circulating in the urban cycle<sup>61–63</sup>, YFV is dependent on epizootics to cause cases in humans. The South American I genotype belongs to a “modern lineage”, that has been circulating in America since 1995 and that perhaps originated in Trinidad and Tobago<sup>40</sup>. It is believed that from there the virus spread to South American countries, especially Venezuela and Brazil<sup>40</sup>, carried mainly by NHP and sylvatic mosquitoes, moving along forested corridors and perhaps promoted by a series of interlocked epizootics involving the exchange of viruses among infected and susceptible individuals<sup>64,65</sup>. Epizootics among social animals, such as New World arboreal primates, may be reduced by self-exclusion of infected individuals<sup>66</sup>. For instance, it is in theory possible that social avoidance, changes in group size, group isolation and several other behaviors may have evolved due to reduce pathogen transmission. Nevertheless, in the case of vector-borne diseases any isolation mechanism is efficient only at distances that minimize transmission<sup>66</sup>. Howlers were the most affected monkey species in the forested belt around the MRSP<sup>52</sup>. As in several other previous YFV epizootics<sup>64</sup>, the high overall fatality rate in howlers led to almost the complete extinction of these monkeys in most areas around Sao Paulo<sup>52</sup>.

It has been assumed that the decline in the numbers of howler monkeys and the severe reduction of several species of NHP from around the MRSP had a significant effect on ending the outbreak. Although perhaps due to poor sampling of monkeys in that locality, it is possible that Clade II-D could have caused a limited number of human-to-human transmission cases, as suggested by a cluster of human cases in Guarulhos (Fig. 5). Critically, however, a key factor that differentiates the current outbreaks of YFV in the Americas and Africa is that there is no clear evidence for urban cycles of YFV in the Americas has been observed since the first half of the 20<sup>th</sup> Century. A possible, although untested, explanation is that the former *A. aegypti* colonizing the Americas was from Africa (Senegalese strain), while the *A. aegypti* reintroduced in the early 1970's is Asiatic, where no urban spread of YFV is observed<sup>67</sup>.

Received: 22 April 2019; Accepted: 11 December 2019;

Published online: 31 December 2019

## References

- Chambers, T. J., Hahn, C. S., Galler, R. & Rice, C. M. Flavivirus Genome Organization, Expression, and Replication. *Annu. Rev.* **44**, 649–688 (1990).
- Barrett, A. D. T. & Higgs, S. Yellow Fever: A Disease that Has Yet to be Conquered. *Annu. Rev. Entomol.* **52**, 209–229 (2007).
- Monath, T. P. & Vasconcelos, P. F. C. Yellow fever. *J. Clin. Virol.* **64**, 160–173 (2015).

4. Zanotto, P. M. *et al.* Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc. Natl. Acad. Sci.* **93**, 548–553 (1996).
5. Bryant, J. E., Holmes, E. C. & Barrett, A. D. T. Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. *PLoS Pathog.* **3**, e75 (2007).
6. Nunes, M. R. T. *et al.* Genomic and phylogenetic characterization of Brazilian Yellow Fever virus strains. *J. Virol.*, 1–33, <https://doi.org/10.1128/JVI.00565-12> (2012).
7. Vasconcelos, P. F. & da, C. Yellow Fever. *Rev. Soc. Bras. Med. Trop.* **36**, 275–293 (2003).
8. Soper, F. L. *et al.* Yellow Fever without *Aedes Aegypti*. Study of a rural epidemic in the Valle Do Chanaan, Espirito Santo, Brazil, 1932. *Am. J. Hyg.* **18**, 555–587 (1933).
9. Consoli, R. A. G. B. & Oliveira, R. L. *Principais mosquitos de importância sanitária no Brasil.* (1994).
10. Theiler, M. & Smith, H. H. The Use of Yellow Fever Virus Modified By *in Vitro* Cultivation for Human Immunization. *J. Exp. Med.* **65**, 787–800 (1937).
11. Kraemer, M. U. G. *et al.* The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. Albopictus*. *Elife* **4**, 1–18 (2015).
12. Braga, I. A. & Valle, D. *Aedes aegypti*: History of Control in Brazil. *Epidemiol. e Serviços Saúde* **16**, 113–118 (2007).
13. Rodaniche, Ede & Galindo, P. Isolation of Yellow Fever virus from *Haemagogus mesodentatus*, *H. equinus* and *Sabethes chloropterus* captured in Guatemala in 1956. *Am. Soc. Trop. Med. Hyg.* **6**, 232–237 (1957).
14. Rodaniche, E., de, Galindo, P. & Johnson, C. M. Isolation of Yellow Fever Virus from *Haemagogus Lucifer*, *H. Equinus*, *H. Spegazzinii Falco*, *Sabethes Chloropterus* and *Anopheles Neivai* Captured in Panama in the Fall of 1956. *Am. Soc. Trop. Med. Hyg.* **6**, 681–685 (1957).
15. Possas, C. *et al.* Yellow fever outbreak in Brazil: the puzzle of rapid viral spread and challenges for immunisation. *Mem. Inst. Oswaldo Cruz* **113**, e180278 (2018).
16. da Saúde, Ministério Reemergência da Febre Amarela Silvestre no Brasil, 2014/2015: situação epidemiológica e a importância da vacinação preventiva e da vigilância intensificada no período sazonal. *Bol. Epidemiológico* **46**, 1–10 (2015).
17. Faria, N. R. *et al.* Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science (80-)*. **7115**, 1–12 (2018).
18. Rezende, I. Mde *et al.* Persistence of Yellow fever virus outside the Amazon Basin, causing epidemics in Southeast Brazil, from 2016 to 2018. *PLoS Negl. Trop. Dis.* **12**, 1–12 (2018).
19. Barbosa, C. M. *et al.* Yellow Fever Virus RNA in Urine and Semen of Convalescent Patient, Brazil. *Emerg. Infect. Dis.* **24**, 176–178 (2018).
20. Figueiredo, P. O. *et al.* Detection and Molecular Characterization of Yellow Fever Virus, 2017, Brazil. *Ecohealth* **14**, 1–7 (2018).
21. Delatorre, E. *et al.* Distinct YFV lineages co-circulated in the Central-Western and Southeastern Brazilian regions from 2015 to 2018. *Front. Microbiol.* **10**, 1–12 (2018).
22. da Saúde, Ministério Monitoramento do Período Sazonal da Febre Amarela Brasil-2017/2018. *Bol. Epidemiológico* **27**, 1–14 (2018).
23. Saúde, S. & do E. da. Boletim epidemiológico Febre Amarela -28/12/2018. *Bol. Epidemiológico* 1–9 (2018).
24. da Saúde, M. Monitoramento do Período Sazonal da Febre Amarela Brasil-2017/2018. *Bol. Epidemiológico* **27**, 1–14 (2018).
25. Moreno, E. S. & Barata, RdeC. B. Municipalities of higher vulnerability to sylvatic Yellow Fever occurrence in the São Paulo state, Brazil. *Rev. Inst. Med. Trop. Sao Paulo* **53**, 335–339 (2011).
26. da Saúde, M. Classificação de Risco dos Agentes Biológicos. *Ministério da Saúde* 1–50 Available at, [http://bvsmis.saude.gov.br/bvsmis/publicacoes/classificacao\\_risco\\_agentes\\_biologicos\\_3ed.pdf](http://bvsmis.saude.gov.br/bvsmis/publicacoes/classificacao_risco_agentes_biologicos_3ed.pdf), (Accessed: 2nd September 2019) (2017).
27. Diallo, D. *et al.* Patterns of a sylvatic yellow fever virus amplification in southeastern Senegal, 2010. *Am. J. Trop. Med. Hyg.* **90**, 1003–1013 (2014).
28. Avelino-Silva, V. I. *et al.* Yellow fever vaccine viremia following ablative BM suppression in AML. *Bone Marrow Transplant.* **48**, 1008–1009 (2013).
29. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
30. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
31. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
32. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, 1–5 (2015).
33. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
34. Nguyen, L.-T., Schmidt, H. A., Haeseler, Avon & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
35. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermini, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
36. Rambaut, A. & Drummond, A. *FigTree* (2009).
37. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
38. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, 1–5 (2018).
39. Gill, M. S. *et al.* Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
40. Mir, D. *et al.* Phylogenetics of Yellow Fever Virus in the Americas: New insights into the origin of the 2017 Brazilian outbreak. *Sci. Rep.* **7**, 1–9 (2017).
41. Rambaut, A. & Drummond, A. J. Tracer v1.4. Available from, <http://beast.bio.ed.ac.uk/Tracer>, Available from, <http://beast.bio.ed.ac.uk/Tracer> (2007).
42. Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A. & Lemey, P. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol. Biol. Evol.* **30**, 239–243 (2013).
43. Bielejec, F. *et al.* Spread3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Mol. Biol. Evol.* **33**, 2167–2169 (2016).
44. Retief, J. D. Phylogenetic analysis using PHYLIP. In *Bioinformatics methods and protocols* **132**, 243–258 (2000).
45. da Saúde, M. Boletim epidemiológico - Febre Amarela - 17/08/2018. *Bol. Epidemiológico* 1–10 (2018).
46. Vasconcelos, P. F. C. *et al.* Serious adverse events associated with yellow fever 17DD vaccine in Brazil: A report of two cases. *Lancet* **358**, 91–97 (2001).
47. Lindsey, N. P. *et al.* Adverse event reports following yellow fever vaccination. *Vaccine* **26**, 6077–6082 (2008).
48. Whittembury, A. *et al.* Viscerotropic disease following yellow fever vaccination in Peru. *Vaccine* **27**, 5974–5981 (2009).
49. Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. *J. Mol. Evol.* **54**, 156–165 (2002).
50. Moreira-Soto, A. *et al.* Evidence for multiple sylvatic transmission cycles during the 2016–2017 yellow fever virus outbreak, Brazil. *Clin. Microbiol. Infect.* **24**, 1019.e1–1019.e4 (2018).
51. Shearer, F. M. *et al.* Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet Glob. Heal.* **6**, e270–e278 (2018).
52. Fioravanti, C. *O alarme dos macacos.* *Pesquisa Fapesp* 263 (2018).

53. Camargo-Neves, V. L. F. *et al.* Entomological investigation of a sylvatic yellow fever area in São Paulo State, Brazil. *Cad. Saude Publica* **21**, 1278–1286 (2005).
54. Mucci, L. F. *et al.* *Haemagogus leucocelaenus* and Other Mosquitoes Potentially Associated With Sylvatic Yellow Fever In Cantareira State Park In the São Paulo Metropolitan Area, Brazil. *J. Am. Mosq. Control Assoc.* **32**, 329–332 (2016).
55. Abreu, F. V. S. *et al.* *Haemagogus leucocelaenus* and *Haemagogus janthinomys* are the primary vectors in the major yellow fever outbreak in Brazil, 2016–2018. *Emerg. Microbes Infect.* **8**, 218–231 (2019).
56. Chiarello, A. G. Home Range of the Brown Howler Monkey, *Alouatta fusca*, in a Forest Fragment of Southeastern Brazil. *Folia Primatol.* **60**, 173–175 (1993).
57. Pinto, L. P. S., Costa, C. M. R., Strier, K. B. & Fonseca, G. A. B. Habitat, Density and Group Size of Primates in a Brazilian Tropical Forest. *Folia Primatol.* **61**, 135–143 (1993).
58. Bonvicino, C. R. & Viana, M. C. Genetic Diversity of *Alouatta* (Primates) from Brazilian Atlantic Forest. *J. Primatol.* **04**, 2 (2015).
59. Hamer, D. H. *et al.* Fatal Yellow Fever in Travelers to Brazil, 2018. *Morb. Mortal. Wkly. Rep.* **67**, 340–341 (2018).
60. Nunes, M. R. T. *et al.* Air Travel Is Associated with Intracontinental Spread of Dengue Virus Serotypes 1–3 in Brazil. *PLoS Negl. Trop. Dis.* **8** (2014).
61. Cunha, M. *et al.* Phylodynamics of DENV-1 reveals the spatiotemporal co-circulation of two distinct lineages in 2013 and multiple introductions of dengue virus in Goiás, Brazil. *Infect. Genet. Evol.* **43**, 130–134 (2016).
62. Ortiz-baez, A. S. *et al.* Origin, tempo, and mode of the spread of DENV-4 Genotype IIB across the state of São Paulo, Brazil during the 2012–2013 outbreak. *Mem. Inst. Oswaldo Cruz* **114**, 1–8 (2019).
63. Nunes, M. R. T. *et al.* Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med.* **13** (2015).
64. Vargas-Mendez, O. & Elton, N. W. Naturally acquired Yellow Fever in wild monkeys of Costa Rica. *Am. J. Trop. Med. Hyg.* **2**, 850–863 (1952).
65. Vasconcelos, P. F. C. *et al.* Genetic Divergence and Dispersal of Yellow Fever Virus, Brazil. **10** (2004).
66. Loehle, C. Social Barriers to Pathogen Transmission in Wild Animal Populations. *Ecology* **76**, 326–335 (1995).
67. Powell, J. R., Gloria-Soria, A. & Kotsakiozi, P. Recent History of *Aedes aegypti*: Vector Genomics and Epidemiology Records. *Bioscience* **68**, 854–860 (2018).

## Acknowledgements

We thank Renato de Souza and Fernando L. Macedo from the Adolfo Lutz Institute for sharing dates and places of confirmed cases of non-human primate (NHP) deaths. We thank the Core Facility for Scientific Research – University of São Paulo (CEFAP-USP/GENIAL) for excellent technical assistance. PMAZ was funded by the Brazilian National Council of Scientific and Technological Development (CNPq) (process no. 441105/2016-5), by the Fiocruz/Pasteur/Aucani-FUSP (process no. 314502) and by the São Paulo Research Foundation (FAPESP) (process no. 2017/23281-6). PHNS is funded by the FAPESP (process no. 2013/21728-2). ECH is funded by an ARC Australian Laureate Fellowship (FL170100022). CTB is funded by the CNPq (process no. 405691/2018-1). MPC and ASOB received FAPESP grants: no. 2016/08204-2 and 2013/25434-3, respectively. CTB received a CAPES fellowship grant no. 1796450. JC received Czech Academy of Sciences grant no. MSM200961703.

## Author contributions

Conceptualization: M.d.P.C., A.N.D.N., J.C., P.H.N.S., E.C.H. and P.M.d.A.Z. Medical care for patients during hospitalization: Y.L.H. and J.S. Responsible by the Yellow Fever crisis committee – HCFMUSP: B.P. Autopsies: A.N.D.N., V.A.F.A., M.D. and P.H.N.S. Sample collection: M.d.P.C., and S.Z.P. Molecular analysis: M.d.P.C., S.Z.P., B.B.S.P. and C.T.B. Bioinformatics analysis: M.d.P.C., A.S.O.B. and J.C. Data curation: M.d.P.C., A.N.D.N., A.S.O.B., J.C., E.C.H. and P.M.d.A.Z. Funding acquisition: C.T.B., P.H.S.N., E.C.H. and P.M.d.A.Z. Project administration: M.d.P.C., and P.M.d.A.Z. Manuscript preparation and revision – M.d.P.C., E.C.H. and P.M.d.A.Z. All authors revised, edited and approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-56650-1>.

**Correspondence** and requests for materials should be addressed to M.d.P.C. or P.M.d.A.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

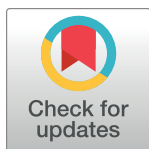
## RESEARCH ARTICLE

## Phylogeographic patterns of the yellow fever virus around the metropolitan region of São Paulo, Brazil, 2016–2019

Marielton dos Passos Cunha<sup>1\*</sup>, Amaro Nunes Duarte-Neto<sup>2</sup>, Shahab Zaki Pour<sup>1</sup>, Bárbara Brito de Souza Pereira<sup>1</sup>, Yeh-Li Ho<sup>3</sup>, Beatriz Perondi<sup>4</sup>, Jaques Sztajnbock<sup>5</sup>, Venancio Avancini Ferreira Alves<sup>2</sup>, Luiz Fernando Ferraz da Silva<sup>2,6</sup>, Marisa Dolhnikoff<sup>2</sup>, Paulo Hilário Nascimento Saldiva<sup>2</sup>, Paolo Marinho de Andrade Zanutto<sup>1</sup>

**1** Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of São Paulo, São Paulo, Brazil, **2** Pathology Department, Faculty of Medicine, University of São Paulo, São Paulo, Brazil, **3** Intensive Care Unit, Division of Clinical Infectious and Parasitic Diseases, Clinical Hospital, Faculty of Medicine, University of São Paulo, São Paulo, Brazil, **4** Yellow Fever Crisis Committee, Clinical Hospital, Faculty of Medicine, University of São Paulo, São Paulo, Brazil, **5** Institute of Infectology Emilio Ribas, São Paulo, Brazil, **6** Service of Verification of Deaths of the Capital—University of São Paulo, São Paulo, Brazil

\* [marieltondospassos@gmail.com](mailto:marieltondospassos@gmail.com)



## OPEN ACCESS

**Citation:** Cunha MdP, Duarte-Neto AN, Pour SZ, Pereira BBdS, Ho Y-L, Perondi B, et al. (2022) Phylogeographic patterns of the yellow fever virus around the metropolitan region of São Paulo, Brazil, 2016–2019. *PLoS Negl Trop Dis* 16(9): e0010705. <https://doi.org/10.1371/journal.pntd.0010705>

**Editor:** Gregory Gromowski, WRAIR, UNITED STATES

**Received:** May 13, 2021

**Accepted:** July 28, 2022

**Published:** September 23, 2022

**Copyright:** © 2022 Cunha et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The new sequences here characterized were deposited in GenBank under the accession numbers MW308134 and MW308135.

**Funding:** PMAZ was supported by the Brazilian National Council of Scientific and Technological Development (CNPq) (grant 441105/2016-5), by the São Paulo Research Foundation (FAPESP) (grant 2017/23281-6), and by the Fiocruz/Pasteur/USP (grant 314502). PHNS was supported by the

## Abstract

From 2016 to 2019, the largest outbreak caused by the Yellow Fever virus (YFV) in the 21<sup>st</sup> century in the Americas occurred in southeastern Brazil. A sylvatic cycle of transmission was reported near densely populated areas, such as the large metropolitan area of the city of São Paulo. Here, we describe the origin, spread, and movement of the YFV throughout the state of São Paulo. Whole-genome sequences were obtained from tissues of two patients who died due to severe yellow fever, during 2018–2019. Molecular analysis indicated that all analyzed tissues were positive for YFV RNA, with the liver being the organ with the highest amount of viral RNA. Sequence analysis indicates that genomes belonged to the South American genotype I and were grouped in the epidemic clade II, which includes sequences from the states of Goiás, Minas Gerais, and São Paulo of previous years. The analysis of viral dispersion indicates that the outbreak originated in Goiás at the end of 2014 and reached the state of São Paulo through the state of Minas Gerais after 2016. When the virus reached near the urban area, it spread towards both the east and south regions of the state, not establishing an urban transmission cycle in the metropolitan region of São Paulo. The virus that moved towards the east met with YFV coming from the south of the state of Rio de Janeiro, and the YFV that was carried to the south reached the Brazilian states located in the south region of the country.

## Author summary

The Yellow Fever virus (YFV) has caused serious public health issues in Africa and the Americas. YFV infections can vary from asymptomatic to severe clinical conditions,

CNPq (grant 304987/2017-4), by the FAPESP (grant 2013/17159-2), by the FunderDOI: 10.13039/501100001807, by the FunderDOI: 10.13039/100000865 and by the Bill and Melinda Gates Foundation (grant INV-002396). MD was supported by the CNPq (grant 316485/2021-7). MPC received a FAPESP fellowship (grant 2016/08204-2). The funders had no role in study design, data collection, and analysis, the decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

which may culminate in death. In Brazil, the main viral origin of YFV outbreaks is the states located in the Amazon Basin. In the present study, we detected YFV RNA in several tissues, including the heart, lung, brain, kidney, spleen, pancreas, liver, and testicle in two human cases. The YFV characterized in the present study belong to the South American genotype I, clade II, and they are associated with viral spread throughout the state of São Paulo after intense viral circulation in the metropolitan region in 2018. Future investigations are necessary for a better understanding of the complex ecological factors involved in viral dispersion from the Amazon Basin.

## Introduction

The Yellow Fever virus (YFV) is an enveloped arbovirus of the family *Flaviviridae*, genus *Flavivirus*, with a single-stranded, positive-sense RNA virus of approximately 11 kb genome encoding a single polyprotein that is cleaved into three structural, including capsid (C), membrane (M) and envelope (E), and seven non-structural (NS) proteins named NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 [1]. Based on the genetic characteristics, the YFV comprises a single serotype with four genotypes: (i) East Africa, (ii) West Africa, (iii) South American I (SA-I), and (iv) South American II (SA-II) [2–4] that may have diverged hundreds of years ago [2,5] following a potential West African origin [2,6]. The infection is naturally initiated by mosquitoes. At the endothelial system, virions are carried by dendritic cells to regional lymph nodes for antigen presentation. This process of onset of infection is followed by the viraemic phase when the virus also migrates to multiple organs. In some cases, a toxæmic phase follows.

YFV is maintained in sylvatic cycles of transmission involving non-human primates (NHP) and sylvatic mosquito species such as *Haemagogus* spp. and *Sabethes* spp. in the Americas, and *Aedes africanus* in Africa. Urban and intermediate cycles of transmission involving humans and *Aedes* spp. have also been reported in Africa [7]. The main urban vector of YFV in Africa, the *Aedes aegypti* mosquito, also transmits Dengue (DENV), Chikungunya (CHIKV), and Zika (ZIKV) viruses to humans in tropical and subtropical areas [8–10]. In both cycles, clinical manifestations range from asymptomatic to fatal conditions [11]. In Brazil, the South American genotype I circulating in a sylvatic environment in the Amazon Basin, including countries such as Brazil, Peru, Bolivia, Colombia, Ecuador, Venezuela, French Guiana, Suriname, and Guyana and is the main group involved in yellow fever epidemics [12,13].

Recently, YFV reemerged in southeast Brazil with intense circulation in the four states of the region, including São Paulo, Rio de Janeiro, Minas Gerais, and Espírito Santo, followed by cases in the northeast and south regions of the country [14–19]. In Brazil, yellow fever (YF) cases are being related to forested environments, and more recently reaching peri-urban areas at the city-forest interface, as the metropolitan region of São Paulo (MRSP) [18,20]. Despite the proximity to urban areas, only sylvatic cycles of transmission have been reported for the last decades in the country [18].

From January to August 17, 2018, 537 fatal cases were reported in the state of São Paulo [21]. Since the beginning of the 21<sup>st</sup> century, YFV enzootic areas have expanded from the north and central regions to the east coast of Brazil. The advance of yellow fever in the east, which is the most populated region of the country, has led the Ministry of Health to recommend nationwide vaccination for yellow fever [22]. While eradication is not feasible due to the sylvatic cycles of transmission, large-scale vaccination coverage is believed to provide considerable protection against a potential re-urbanization of YFV transmission [23].

Since 2016, an epidemic and several epizootics of yellow fever affecting mainly howler monkey populations, have occurred in regions with low vaccination coverage [18,24]. Here, we



described the whole-genome sequencing of two yellow fever cases that occurred in late 2018 and early 2019, in two different locations of the state of São Paulo. The main objective of the present study was to analyze and compare YFV sequences from recent cases to sequences available on GenBank since 2015 (all from clade II of the outbreak) in the states of Goiás, Minas Gerais, and São Paulo, to better estimate the routes of introduction, dispersion, and escape of the virus during the entire period of the outbreak.

## Material and methods

### Ethical statement

The human autopsies reported in this study were performed after obtaining the written consent of the family members and following the protocol approved by the research ethics committee of the Clinical Hospital of the University of São Paulo School of Medicine (HCFMUSP) (CAAE protocol number: 18781813.2.0000.0068). All the methods were performed following the relevant guidelines and regulations of the ethics committee of the HCFMUSP following the approved protocol. All laboratory procedures listed below were performed in a biosafety level 2 laboratory (BSL-2), under the Brazilian standards of the Ministry of Health for Biological Agents Risk Classification [25].

### Patients and samples

From November/2018 to February/2019, we investigated three patients suspected of yellow fever, all of which were later diagnosed as fatal cases of YFV infection. The suspected case definition of yellow fever was established by the Brazilian Ministry of Health in consonance with the Health Department of the state of São Paulo. Suspected yellow fever cases are defined as illness presenting sudden onset high fever associated with jaundice and/or haemorrhage who had lived or had visited areas with evidence of YFV circulation (*i.e.*, epizootics in NHP or isolation of YFV from vectors), regardless of the vaccine status, during the preceding 15 days.

Confirmed cases had compatible clinical presentation and laboratory confirmation by at least one of the following methods: (i) Seropositive by Immunoglobulin M antibody capture enzyme-linked immunosorbent assay (MAC-ELISA); (ii) detection of YFV RNA by Real-time quantitative reverse transcription-polymerase chain reaction (RT-qPCR) in blood samples; (iii) virus isolation; (iv) histopathology compatible with yellow fever hepatitis with detectable antigen in tissues by immunohistochemistry technique. All cases received the definitive laboratory diagnosis of yellow fever by the Adolfo Lutz Institute, the state Reference Laboratory. Previous exposure or co-infection by Hepatitis A virus (HVA), B (HBV), C (HCV), Cytomegalovirus (CMV), Herpes virus (HSV), DENV, CHIKV, Human Immunodeficiency virus type 1 (HIV-1), *Leptospira*, and other non-infectious diseases aetiologies for acute hepatitis were accessed and cases were excluded following clinical diagnostic methods.

Epidemiological data, including demographic aspects, clinical data, including pre-existing medical conditions, clinical signs and symptoms, and in-hospital follow-up until death, as well as laboratory data, including diagnostic tests, were collected from all patients.

### Autopsy protocol and tissue processing

The Service of Verification of Deaths of the Capital—USP (SVOC-USP) investigated deaths due to yellow fever from November/2018 to February/2019. Autopsies were performed following the *Letulle* technique, where all the organs were removed *en masse* (one block), requiring dissection organ by organ to examine them individually. Briefly, the dissection to all the

patients were performed in the following organs: (i) heart; (ii) lung; (iii) brain; (iv) kidney; (v) spleen; (vi) pancreas; (vii) liver; and (viii) testicle.

### Molecular characterization

Nucleic acid extraction from all collected tissues using 50 mg of the material was performed using the TRIzol reagent (Life Technologies, Carlsbad, CA, USA) and carried out according to the manufacturer's instructions. Molecular detection of YFV was performed by AgPath-ID One-Step RT-PCR Reagents (Ambion, Austin, TX, USA) with specific primers/probe, as previously described [26]. To identify cases of rare adverse events from yellow fever vaccination we used specific primers/probe for the detection of the attenuated viral strain 17D, the vaccine virus [27]. To quantify the genomic viral concentration in each tissue, we used a double-stranded DNA fragment containing the amplification region (Exxtend, SP, Brazil), which from several known molecules, was diluted and used to estimate the viral RNA concentration by linear regression. RT-qPCR reactions were run in ABI7500 equipment (Thermo Fisher Scientific, Waltham, MA, USA). To visualize the spatial location of the positive patients, we used R software, geobr package, (MIT license <https://ipeagit.github.io/geobr/>) [28] to create the maps.

### Sequencing and viral genome assembly

Based on the RNA viral concentration and in our previous protocol to obtain viral complete genomes from human tissues [18,29], positive samples had the total RNA re-extracted by the abovementioned procedure. RNA samples were then purified and concentrated with DNase I and RNA Clean and Concentrator -5 kit (Zymo Research, Irvine, CA, USA). The paired-end RNA libraries were constructed and validated using the TruSeq Stranded Total RNA HT sample prep kit (Illumina, San Diego, CA, USA). Sequencing was done at the Core Facility for Scientific Research—University of São Paulo (CEFAP-USP/GENIAL) using the Illumina MiSeq platform. Each sample was barcoded individually, which allowed the separation of reads for each patient. Short unpaired reads and low-quality bases and reads were removed using Trimmomatic version 0.39 (LEADING:20 TRAILING:20 SLIDINGWINDOW:4:25 MINLEN:36) [30]. Consensus genomes were assembled with paired-end reads using Bowtie2 version 2.4.1 using default parameters [31]. To access the coverage and generate the consensus genome (Fig A in S1 Text), we used UGENE v.33.0 [32].

### Data sets

All the complete genomic sequences available for YFV (until August 20, 2020), and their associated information, such as location and year of detection were recovered from the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/genbank/>) website in GenBank format. All sequences were converted into FASTA format, and to avoid duplicated and low-quality sequences with many unknown nucleotides (N), we set an exclusion threshold for sequences that had more than 1% of “N” concerning their total length ([https://biopython.org/wiki/Sequence\\_Cleaner](https://biopython.org/wiki/Sequence_Cleaner)). These sequences were grouped with the sequences isolated from patients in São Paulo 2018–2019, the FASTA sequences were aligned using Clustal Omega [33] and recombinant sequences were screened using all algorithms implemented in RDP5 Beta 5.05 program (RDP, GENECONV, MaxChi, BootScan, and Siscan) using the standard settings [34]. The alignment of recombinant free sequences was manually inspected and edited using the program AliView v.1.18 [35].

To determine which genotypes and lineages circulated in the state of São Paulo during the outbreak, two datasets were constructed: (i) the first (dataset-1) included all the sequences deposited at GenBank (before August—2020), with the sequences determined in our study

( $n = 264$  sequences) (Table A in [S1 Text](#)); (ii) the second (dataset-2), based in the phylogenetic analysis of the dataset-1, included all sequences deposited at GenBank (before August—2020) belonging to the genotype SA-I, clade II of the outbreak (sequences isolated from 2015 in the states of Goiás, Minas Gerais, and São Paulo states) ( $n = 91$  sequences) (Table B in [S1 Text](#)), according to our previous publication [18].

### Phylogenetic analysis

Phylogenetic inference was performed with all the curated datasets. The phylogenetic tree was reconstructed based on the full-length, curated coding sequences using the Maximum Likelihood (ML) method implemented in IQ-TREE 1.5.5 [36]. The robustness of the groupings observed was assessed using an ultrafast bootstrap approximation (UFboot) during 1,000 replicates. The ML tree was visualized and plotted using FigTree v.1.4.3 [37]. Taxon labels for sequences used in this work had the format: genotype/accession number/local of isolation/date of isolation.

### Phylogeographic analysis

All the sequences characterized as clade II were used to reconstruct the viral origin and to trace the movement of the C-II sequences (dataset-2). We explored the temporal signal (*i.e.*, molecular clock structure) and the quality of our data set using TempEst v.1.5.3 [38]. The spatiotemporal spread was reconstructed under a Bayesian framework implemented in BEAST v.1.10.4 [39], using the general time-reversible model incorporating invariant sites with gamma-distributed rate variation substitution model (GTR+I+G), as described by the Bayesian information criterion (BIC) using jModelTest 2.1.10 [40]. Based on previous estimates of evolutionary dynamics of related YFV [17,41], we tested uncorrelated relaxed molecular clocks assuming a log-normal distribution, in combination with non-parametric population growth models: (i) the standard Bayesian skyline plot (BSP; 10 groups) [42], (ii) the Bayesian skyride plot [43], and (iii) the Bayesian skygrid model [42] (Tables C and D in [S1 Text](#)). Phylogeography patterns and parameters were estimated by running a Markov Chain Monte Carlo (MCMC) for 200 million states and sampling every 200,000 states with 10% burn-in. The effective sample size (ESS) and convergence were examined with Tracer v.1.7.1 [44]. The maximum clade credibility (MCC) tree was also visualized with FigTree v.1.4.3 [37]. To calculate the log marginal likelihood for demographic model selection, we used the path sampling (PS) [45,46] and the stepping-stone (SS) sampling [47] approaches by running 100 path steps of 1 million iterations each. The spatiotemporal spread was visualised with spread3 v.0.9.7.1 [48].

### Data sequences

The new sequences here characterized were deposited in GenBank under the accession numbers MW308134 and MW308135.

## Results

### YFV epidemiological surveillance in São Paulo, 2016–2019

YFV emerged in the state of São Paulo and was first recognized in early 2016, through the detection of two human cases. The monitoring of epizootics revealed an ongoing occurrence of yellow fever in NHP. In 2016 and 2017, the number of NHP cases reported was higher than in humans. Since the end of 2017, YFV was reported in an extensive area where there was no recommendation for vaccination, mainly in the MRSP. In late 2018, a specific case was observed in the Atlantic Forest in the eastern region of the state, near the border with the state



of Rio de Janeiro. In early 2019, some cases were observed in the south of the state, close to the border with the state of Paraná, and in the last two years of the epidemic (2018 and 2019), the number of human cases has suppressed the number of reported NHP cases (Fig 1A–1C).

From October 2018 to February 2019, three patients suspected of fatal yellow fever entered the SVOC-USP in the city of São Paulo. All patients had organs that tested positive for RT-qPCR. Two patients (patient 084 and patient 085) were confirmed positive for sylvatic YFV, and one of them tested positive for the vaccine strain YFV-17DD.

### Clinical and demographic features associated with the sylvatic YF cases

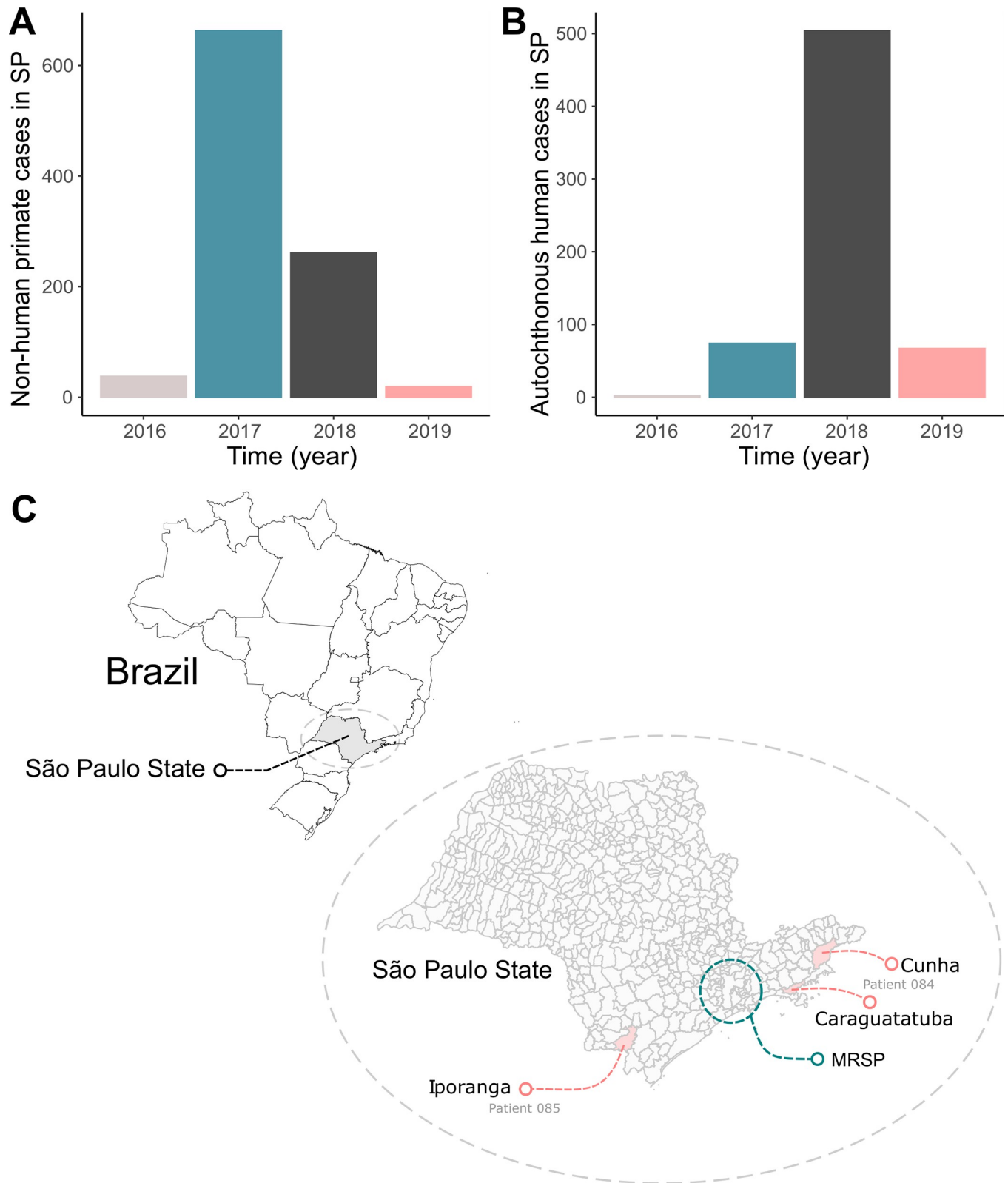
The first case (patient 084) was a 26-year-old man, from the city of Cunha, a peasant, without any medical condition. He travelled one week before symptoms onset to Caraguatatuba, a region of Atlantic Forest, to work on a farm. He had no YF vaccination. His disease started on October 20<sup>th</sup>, 2018, presenting asthenia, generalized myalgia, intense headache, fever, oliguria, nausea, vomits, abdominal pain, and developed icterus in the following days. He was transferred to São Paulo to the intensive care unit of the Infectology Institute “Emílio Ribas”, on the 7<sup>th</sup> day. The initial clinical hypothesis was: YF, leptospirosis, dengue, or Brazilian spotted fever. At admission, he was in a severe clinical condition, with torpor, bradycardia, and uremic, and was put under mechanical ventilation. During hospitalization fulminant hepatitis was diagnosed, with gastrointestinal haemorrhages, convulsions, and acute kidney injury, needing dialysis and multiple hemocomponents transfusion. He died on the 9<sup>th</sup> day of disease onset. The diagnosis of YF was done by a positive anti-YFV IgM in serum (collected on the 7<sup>th</sup> day); detectable YFV RNA in the blood by RT-PCR (7<sup>th</sup> day), and autopsy findings. Based on the demographic history that preceded this patient’s fatal condition, we decided to consider the probable site of infection as Caraguatatuba in the viral spread reconstruction.

The second case (patient 085) was a 58-year-old male, peasant, resident of Iporanga, close to areas with confirmed cases of YF epizootics and other human cases. He had a medical history of high blood pressure and was treated for prostate cancer. His symptoms started on January 15<sup>th</sup>, 2019, with intense headache, fever, malaise, nausea, vomiting, and jaundice. The patient was transferred to IIER ICU with clinical suspicion of yellow fever. He evolved with fulminant hepatitis with coma and multiple organ dysfunction, treated with mechanical ventilation, hemocomponents transfusion, antibiotics, and other intensive measures. The patient died on the 5<sup>th</sup> day of illness. The diagnosis of YF was done by a positive anti-YFV IgM in serum (collected on the 5<sup>th</sup> day); detectable YFV RNA in the blood by RT-PCR (4<sup>th</sup> day), and autopsy findings.

Both cases had mid-zonal hepatitis, with steatotic and apoptotic hepatocytes, sinusoidal congestion and Kupffer cells hyperplasia; diffuse gastrointestinal haemorrhages; pulmonary haemorrhages, lymphoid hypoplasia, and acute tubular necrosis. Liver tissues presented high levels of viral RNA when compared to other tissues (Fig 2).

### Phylogenetic analysis

Based on our previous work that showed the efficiency in recovering quality viral YFV sequences from liver autopsy material of patients with clinical outcomes of death [18], we submitted the total RNA obtained from the patients’ liver to total genomic RNA sequencing. In both, quality sequences were obtained, one with the complete coding region, and the other almost complete, having 20 contiguous bp that were not sequenced, that was assumed as “N” from the 6315–6334 site of the coding sequence (Fig A in S1 Text). However, we chose to keep this sequence in our analysis, since no nucleotide substitution in that region was observed in these positions in all clade II sequences (n = 91 sequences), which were assumed as possibly non-informative sites.

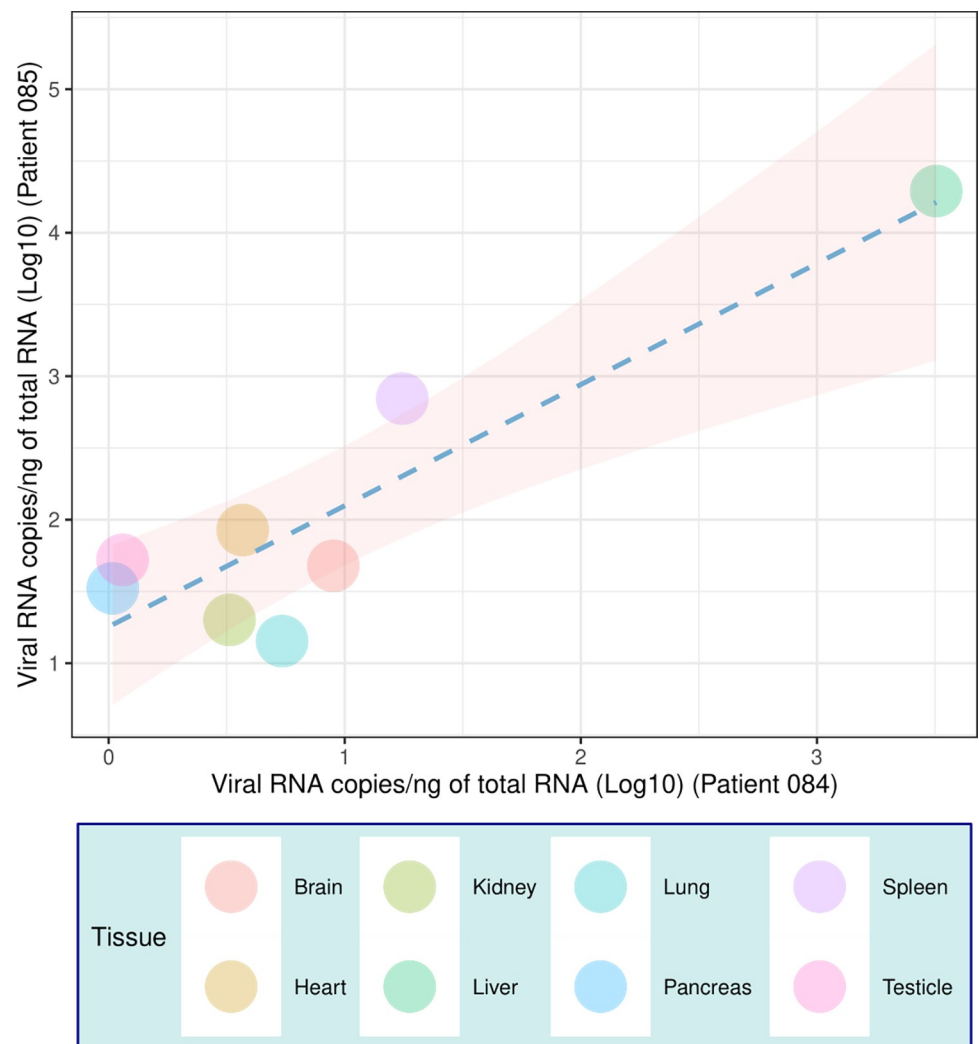


**Fig 1. Epidemiological situation of the YFV outbreak in São Paulo, 2016–2019.** (A) The number of YFV positive NHPs in São Paulo, from 2016–2019. (B) The number of YFV-positive human cases in São Paulo, from 2016–2019. (C) Location of the two fatal human cases studied in the transition from 2018 to 2019. We used R software, geobr package, (MIT license <https://ipeagit.github.io/geobr/>) to create the maps produced in this Fig to visualize the spatial location.

<https://doi.org/10.1371/journal.pntd.0010705.g001>

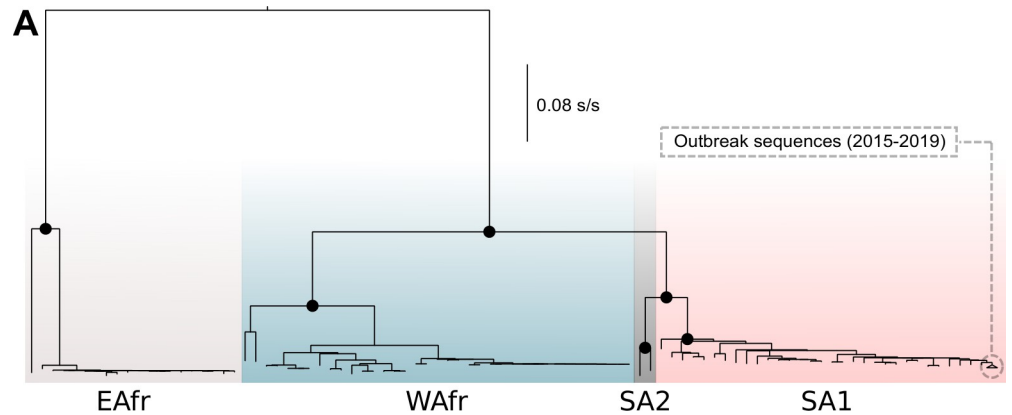
## Phylogenetic characterization

The phylogenetic characterization of the two viruses sequenced here, compared to sequences previously characterized and deposited in the GenBank indicated that both sequences grouped within the South American genotype I, within the group of sequences associated with the outbreak identified in Brazil between 2016–2019, including sequences of midwest, southeast and northeast Brazil (Fig 3). The coding region of the two genomes sequenced here was grouped into a multi-Fasta file with another dataset of 89 non-redundant genomes recovered from GenBank ( $n = 91$  sequences), previously filtered to exclude incomplete genomes, sequences with low quality, and sequences without information on location and date of collection. This alignment, with 91 sequences, was generated producing a file with 10,236 nucleotide sites, with 10,073 constant and invariant (constant or ambiguous constant) sites (98.41% of all sites). Among the variant sites, 60 sites were considered informative under parsimony and 202 sites had distinct site patterns. A root-to-tips analysis showed that the virus accumulated genetic diversity over time ( $r = 0.55$ ) (Fig 4).



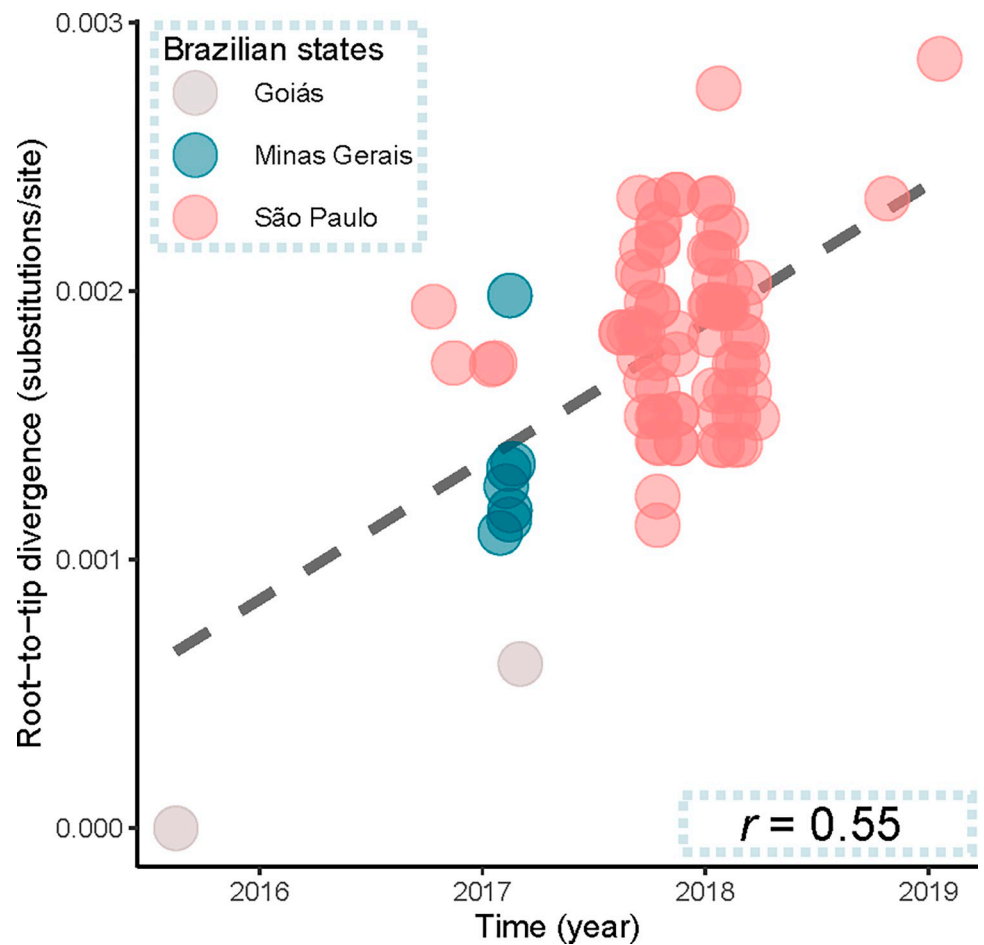
**Fig 2. Molecular characterization of the sequenced yellow fever cases, showing the number of YFV RNA copies according to each of the 8 tissues analyzed for positive YFV-sylvatic patients.** The different colors represent the different tissues analyzed. The results indicate that both patients present a positive correlation for the quantification of genetic material.

<https://doi.org/10.1371/journal.pntd.0010705.g002>



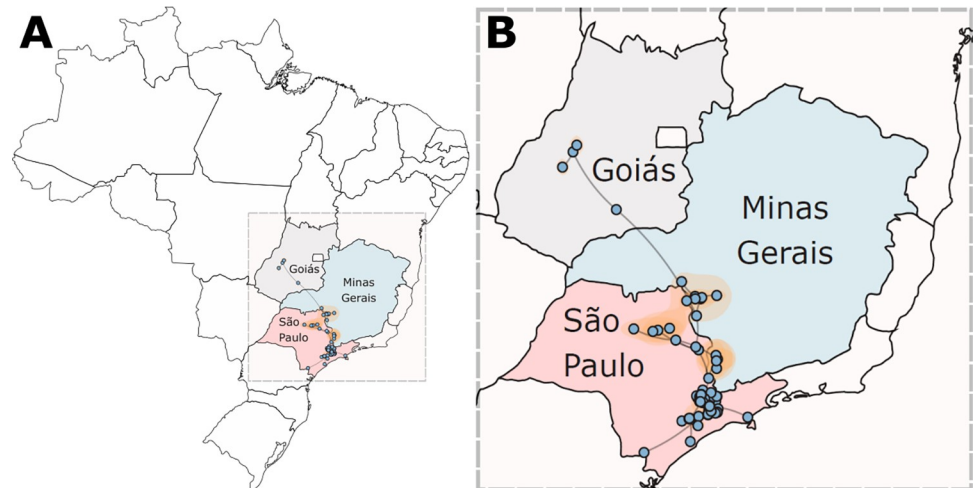
**Fig 3. Maximum likelihood phylogenetic trees for YFV based on full-length polyprotein sequences (n = 264) indicate that the sequences associated with the outbreak in Brazil between 2016–2019 grouped within the genotype South American I.** The tree is midpoint-rooted and the black circles on the main nodes represent bootstrap support values greater than 0.7. The distinct colors represent distinct genotypes: (i) EAfr—East African; (ii) WAfr—West African; (iii) SA1—South American I; (iv) SA2—South American II.

<https://doi.org/10.1371/journal.pntd.0010705.g003>



**Fig 4. Correlation between genetic divergence and sampling time was obtained by a root-to-tips analysis using the YFV sequences.** Root-to-tips distances were inferred using a maximum likelihood phylogeny for the genotype South American I, Clade II, (dataset-2).

<https://doi.org/10.1371/journal.pntd.0010705.g004>



**Fig 5. Highest posterior probability migration paths for the YFV Clade II from 2015 to 2019 towards and beyond the MRSP, based on the analysis of 91 complete genomes (dataset 2).** (A) Migration on the full map of Brazil, and (B) migration being viewed up close. The spatiotemporal spread was visualized with *spreaD3* v.0.9.7.1, using a GeoJSON file, (MIT license <https://www.kaggle.com/datasets/moiseslima/majson?resource=download>) to visualize the spatial YFV spread.

<https://doi.org/10.1371/journal.pntd.0010705.g005>

### Phylogeographic characterization

The comparison between some of the different models of viral propagation and diffusion (Tables C and D in *S1 Text*) allowed us to estimate the pattern of viral dispersion with greater accuracy. According to the inferential results (Fig 5), the outbreak began in the state of Goiás in December/2014 (95% HPD = February/2014 – August/2015), and subsequently reached the state of Minas Gerais in September/2015 (95% HPD = November/2014 – May/2016), also reaching the state of São Paulo. In São Paulo, the virus was introduced a single time in February/2016 (95% HPD = September/2015 – July/2016), and since then it has spread to several locations throughout the state reaching back to the state of Minas Gerais after 2016.

### Discussion

Since the last urban cases of yellow fever reported in the Americas in the early 1950s, YFV has been maintained in forested areas through sylvatic cycles of transmission. The largest yellow fever outbreak in the 21<sup>st</sup> century in the Americas began in 2016 passing through all states of southeast Brazil reaching the south and northeast regions of the country [14,18,19,49]. To reach the state of Minas Gerais, the virus possibly was carried across the state of Goiás, located in the west-central region, as the bridge from its natural maintenance area, the Amazon basin. Goiás is located in a dry savannah area (cerrado), which is a biome physically linked to the Amazon and Atlantic Forest [49,50]. In recommended areas for YF vaccination, the virus circulated silently causing epizootics in NHP. In areas with low vaccination coverage, YFV affected human populations causing a high number of deaths from 2017 to 2018 [18,22].

In the MRSP, the virus began its circulation in 2017, reaching the region from the south Minas Gerais, moving slowly and causing infections in NHP and humans along with its spread [18,24]. Despite circulating in forested regions close to densely inhabited areas in the MRSP during 2017–2018, YFV only infected susceptible people who had reported exposure to forested environments [15,16,18]. In the following epidemiological period (2018–2019), no cases were reported in NHP or humans in the MRSP. This can be explained by two factors: (i) YFV collapsed many populations of NHP in the region and (ii) the human population adhered to

mass vaccination during the outbreak. During 2018–2019, the YF cases were reported mainly in the Vale do Ribeira (south), but also in the east of the state of São Paulo. In the case of the patient who lived in the city of Cunha, he visited a neighboring city (Caraguatatuba) at a time compatible with the viral incubation period, which is about 3 to 6 days before the onset of symptoms [51,52], which is located in the same region in a distance of 76 km away.

Multiple organs were positive for the presence of YFV RNA, as we have shown recently for other cases [15,16,18]. The liver was the organ with a high viral load in both patients. These findings reinforce the well-established evidence that the liver tissue is a replication site of YFV and the notion of multiple tissue compartmentalization in severe cases of infection. In some severe and non-fatal cases, hepatic tissue infection is seen months after the clinical outcome of the disease, causing late hepatitis, with positive molecular detection and evidence of lesions in pathological analyses [53].

In experimental models using NHP, the genome and viral antigens are found in multiple organs with a greater amount of virus, when compared to serum, between one to 30 days post-infection. Interestingly, neutralizing antibodies are found after the seventh day of infection, in consonance with the reduction of viremia and subsequently reduction of virus levels in tissues [54]. Our findings shown in Fig 2 support a model for YFV infection in humans based on autopsy and molecular findings, characterized by hepatitis, nephropathy, coagulopathy, and haemorrhagic vasculopathy that can lead to a fatal condition [16].

The viral sequences recovered from both patients reported here are grouped within the South American genotype I, alongside sequences previously characterized in previous years in São Paulo and neighboring states [14,18,49]. Within the group of outbreak sequences, they were grouped within clade II, which are sequences corresponding to viral isolates that circulated in the states of São Paulo, Minas Gerais, and Goiás [18,49].

The pattern of viral spread observed in the transition from 2018 to 2019 suggested that the densely urbanized area of the MRSP acted as a barrier to viral dispersion, deflecting the virus to spread to the east and south of the state. When comparing the transmission network of the 2015–2019 outbreak with other well-noted outbreaks from the past, we noticed that a very similar pattern of viral movement was observed in Brazil during outbreaks between 1932 and 1942 [50,55,56]. During those outbreaks, YFV reached the state of São Paulo from the Amazon Basin, also through the state of Goiás. Possibly, the maintenance of the viral spread pattern is associated with biological variables associated with virus-host interaction, including vectors and NHP, and geo-environmental variables, such as altitude, annual rainfall, and temperature [57]. As occurred between 1932–1942, YFV reached the MRSP and spread to the east and south regions of the state. Two possible hypotheses can explain the recent virus siege of the MRSP: (i) when the first cases of YFV were identified in NHP and human populations close to urban areas in the MRSP, the Department of Health of the state of São Paulo started mass campaigns for vaccination in peri-urban regions, causing an immunological protection belt around the densely populated area of the MRSP, and (ii) since the first half of the 20<sup>th</sup> century, outbreaks in urban environments in Brazil caused by YFV are not documented, and one of the hypothesis is the lack of vector competence of the current *Aedes aegypti* populations in Brazil. However, a recent study demonstrated that different populations of *Aedes aegypti* in Brazil are competent for urban transmission of YFV [58]. Thus, vaccination associated with the circulation of other closely related flaviviruses in urban areas, as well as other ecological and epidemiological factors is believed to be involved in the continuous absence of urban cycles of transmission of YFV in the Americas.

In 2019, the YFV circulation decreased considerably compared to previous years, but with the circulation of the virus associated with clade I in the state of Rio de Janeiro with few cases [59], and in São Paulo associated with the clade II, some cases were reported. After 2019, NHP



and human cases of YFV were reported in south Brazil, including the states of Paraná, Santa Catarina, and the Rio Grande do Sul [60–62]. Considering the limited number of YFV sequences available for better evaluation of the movement of YFV in these states, our analysis suggests that YFV that arrived in the state of Paraná moved to the southern region of the state of São Paulo, as they are border states.

## Supporting information

**S1 Text. Fig A in S1 Text.** Combined coverage (normalized by the sample average) along the two sequenced Yellow Fever virus (YFV) genomes generated in this study. The genomic position reflects the genomic organization of the YFV, which is organized as follows: a single polyprotein cleaved into three structural, including capsid (C), membrane (M), and envelope (E), and seven non-structural (NS) proteins named NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5. The polyprotein is flanked by the 5' and 3' ends, which are non-coding. **Table A in S1 Text.** Complete YFV polyprotein sequences used in the phylogenetic analysis (dataset-1) (n = 264). **Table B in S1 Text.** Complete YFV polyprotein sequences used in the phylogeographic analysis (dataset-2) (n = 91). **Table C in S1 Text.** Model comparison of the relaxed molecular clock and demographic growth models through path sampling (PS) and stepping stone (SS) methods. Bold numbers indicate the best fitting model. **Table D in S1 Text.** Comparison among continuous diffusion models for Brazilian sequences. (DOCX)

## Acknowledgments

The authors thank all physicians, residents, physiotherapists, nurses, and auxiliary nurses who provided care to the patient; all residents of the Pathology Department who participated in the autopsy procedures; all technicians in our histology laboratory; and the relatives who agreed with the autopsy. We thank the Core Facility for Scientific Research-University of Sao Paulo (CEFAP-USP/GENIAL) for technical assistance.

## Author Contributions

**Conceptualization:** Marielton dos Passos Cunha, Amaro Nunes Duarte-Neto, Marisa Dolhnikoff, Paulo Hilário Nascimento Saldiva, Paolo Marinho de Andrade Zanotto.

**Data curation:** Marielton dos Passos Cunha, Amaro Nunes Duarte-Neto, Shahab Zaki Pour, Paolo Marinho de Andrade Zanotto.

**Formal analysis:** Marielton dos Passos Cunha, Amaro Nunes Duarte-Neto, Shahab Zaki Pour, Bárbara Brito de Souza Pereira, Yeh-Li Ho, Beatriz Perondi, Jaques Sztajn bok, Paulo Hilário Nascimento Saldiva, Paolo Marinho de Andrade Zanotto.

**Funding acquisition:** Marielton dos Passos Cunha, Marisa Dolhnikoff, Paulo Hilário Nascimento Saldiva, Paolo Marinho de Andrade Zanotto.

**Investigation:** Marielton dos Passos Cunha, Amaro Nunes Duarte-Neto, Paolo Marinho de Andrade Zanotto.

**Methodology:** Marielton dos Passos Cunha, Amaro Nunes Duarte-Neto, Shahab Zaki Pour, Bárbara Brito de Souza Pereira, Paulo Hilário Nascimento Saldiva, Paolo Marinho de Andrade Zanotto.

**Project administration:** Paolo Marinho de Andrade Zanotto.

**Resources:** Paulo Hilário Nascimento Saldiva, Paolo Marinho de Andrade Zanotto.

**Supervision:** Venancio Avancini Ferreira Alves, Luiz Fernando Ferraz da Silva, Marisa Dolhnikoff, Paulo Hilário Nascimento Saldiva, Paolo Marinho de Andrade Zanotto.

**Visualization:** Marielton dos Passos Cunha, Amaro Nunes Duarte-Neto, Paolo Marinho de Andrade Zanotto.

**Writing – original draft:** Marielton dos Passos Cunha, Amaro Nunes Duarte-Neto.

**Writing – review & editing:** Marielton dos Passos Cunha, Amaro Nunes Duarte-Neto, Shahab Zaki Pour, Bárbara Brito de Souza Pereira, Yeh-Li Ho, Beatriz Perondi, Jaques Sztajnbock, Venancio Avancini Ferreira Alves, Luiz Fernando Ferraz da Silva, Marisa Dolhnikoff, Paulo Hilário Nascimento Saldiva, Paolo Marinho de Andrade Zanotto.

## References

1. Chambers TJ, Hahn CS, Galler R, Rice CM. Flavivirus Genome Organization, Expression, and Replication. *Annu Rev.* 1990; 44: 649–688.
2. Bryant JE, Holmes EC, Barrett ADT. Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. *PLoS Pathog.* 2007; 3: e75. <https://doi.org/10.1371/journal.ppat.0030075> PMID: 17511518
3. Monath TP, Vasconcelos PFC. Yellow fever. *J Clin Virol. Elsevier B.V.*; 2015; 64: 160–173. <https://doi.org/10.1016/j.jcv.2014.08.030> PMID: 25453327
4. Barrett ADT, Higgs S. Yellow Fever: A Disease that Has Yet to be Conquered. *Annu Rev Entomol.* 2007; 52: 209–229. <https://doi.org/10.1146/annurev.ento.52.110405.091454> PMID: 16913829
5. Zanotto PM de A, Gould EA, Gao GF, Harveyt PH, Holmestt EC. Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci.* 1996; 93: 548–553. <https://doi.org/10.1073/pnas.93.2.548> PMID: 8570593
6. Nunes MRT, Palacios G, Cardoso JF, Martins LC, Sousa-Junior EC, Lima CPS de, et al. Genomic and phylogenetic characterization of Brazilian Yellow Fever virus strains. *J Virol.* 2012; 1–33. <https://doi.org/10.1128/JVI.00565-12> PMID: 23015713
7. Abreu FVS de, Ribeiro IP, Ferreira-de-Brito A, Santos AAC dos, Miranda RM de, Bonelly I de S, et al. *Haemagogus leucocelaenus* and *Haemagogus janthinomys* are the primary vectors in the major yellow fever outbreak in Brazil, 2016–2018. *Emerg Microbes Infect.* Taylor & Francis; 2019; 8: 218–231. <https://doi.org/10.1080/22221751.2019.1568180> PMID: 30866775
8. Salazar MI, Richardson JH, Sánchez-Vargas I, Olson KE, Beaty BJ. Dengue virus type 2: replication and tropisms in orally infected *Aedes aegypti* mosquitoes. *BMC Microbiol.* 2007; 7: 9. <https://doi.org/10.1186/1471-2180-7-9> PMID: 17263893
9. Costa-da-Silva AL, Ioshino RS, Petersen V, Lima AF, Cunha MDP, Wiley MR, et al. First report of naturally infected *Aedes aegypti* with chikungunya virus genotype ECSA in the Americas. *PLoS Negl Trop Dis.* 2017; 11. <https://doi.org/10.1371/journal.pntd.0005630> PMID: 28614394
10. Romo H, Kenney JL, Blitvich BJ, Brault AC. Restriction of Zika virus infection and transmission in *Aedes aegypti* mediated by an insect-specific flavivirus. *Emerg Microbes Infect.* Springer US; 2018; 7: 1–13. <https://doi.org/10.1038/s41426-018-0180-4> PMID: 30429457
11. Beasley DWC, McAuley AJ, Bente DA. Yellow fever virus: Genetic and phenotypic diversity and implications for detection, prevention and therapy. *Antiviral Res. Elsevier B.V.*; 2015; 115: 48–70. <https://doi.org/10.1016/j.antiviral.2014.12.010> PMID: 25545072
12. Fernandes NCCA, Guerra JM, Díaz-Delgado J, Cunha MS, Saad L del C, Iglezias SD, et al. Differential yellow fever susceptibility in new world nonhuman primates, comparison with humans, and implications for surveillance. *Emerg Infect Dis. Centers for Disease Control and Prevention (CDC);* 2021; 27: 47–56. <https://doi.org/10.3201/eid2701.191220> PMID: 33350931
13. Vasconcelos PFC, Bryant JE, Travassos APA, Tesh RB, Rodrigues SG, Barrett ADT. Genetic Divergence and Dispersal of Yellow Fever Virus, Brazil. 2004; 10. <https://doi.org/10.3201/eid1009.040197> PMID: 15498159
14. Barbosa CM, Paola N Di, Cunha MP, Rodrigues-jesus MJ, Araujo DB, Silveira VB, et al. Yellow Fever Virus RNA in Urine and Semen of Convalescent Patient, Brazil. *Emerg Infect Dis.* 2018; 24: 176–178.
15. Duarte-neto AN, Monteiro RA de A, Johnsson J, Cunha M dos P, Pour SZ, Saraiva AC, et al. Ultrasound-guided minimally invasive autopsy as a tool for rapid post-mortem diagnosis in the 2018 São



- Paulo yellow fever epidemic: Correlation with conventional autopsy. *PLoS Negl Trop Dis*. 2019; 13: e0007625.
16. Duarte-Neto AN, Cunha M dos P, Marcilio I, Song ATW, de Martino RB, Ho Y, et al. Yellow Fever and Orthotopic Liver Transplantation: new insights from the autopsy room for an old but reemerging disease. *Histopathology*. 2019; 1–11. <https://doi.org/10.1111/his.13904> PMID: 31087672
  17. Faria NR, Kraemer MUG, Hill S, Jesus JG de, Aguiar RS de, Iani FCM, et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science (80-)*. 2018; 7115: 1–12. <https://doi.org/10.1126/science.aat7115> PMID: 30139911
  18. Cunha M dos P, Duarte-Neto AN, Pour SZ, Ortiz-Baez AS, Černý J, Pereira BB de S, et al. Origin of the São Paulo Yellow Fever epidemic of 2017–2018 revealed through molecular epidemiological analysis of fatal cases. *Sci Rep*. 2019; 9: 1–10. <https://doi.org/10.1038/s41598-019-56650-1> PMID: 31892699
  19. Goes de Jesus J, Gräf T, Giovanetti M, Mares-Guia MA, Xavier J, Lima Maia M, et al. Yellow fever transmission in non-human primates, Bahia, Northeastern Brazil. *PLoS Negl Trop Dis*. 2020; 14: e0008405. <https://doi.org/10.1371/journal.pntd.0008405> PMID: 32780745
  20. Lacerda AB, del Castillo Saad L, Ikefuti PV, Pinter A, Chiaravalloti-Neto F. Diffusion of sylvatic yellow fever in the state of São Paulo, Brazil. *Sci Reports* 2021 111. Nature Publishing Group; 2021; 11: 1–11. <https://doi.org/10.1038/s41598-021-95539-w> PMID: 34381111
  21. Ministério da Saúde. Boletim epidemiológico—Febre Amarela—17/08/2018. *Boletim Epidemiológico*. 2018; 1–10. Available: <http://portalms.saude.gov.br/saude-de-a-z/febre-maculosa/situacao-epidemiologica>
  22. Romano APM, Costa ZGA, Ramos DG, Andrade MA, de Jayme VS, de Almeida MAB, et al. Yellow Fever Outbreaks in Unvaccinated Populations, Brazil, 2008–2009. *PLoS Negl Trop Dis*. 2014; 8: e2740. <https://doi.org/10.1371/journal.pntd.0002740> PMID: 24625634
  23. Shearer FM, Longbottom J, Browne AJ, Pigott DM, Brady OJ, Kraemer MUG, et al. Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet Glob Heal*. 2018; 6: e270–e278. [https://doi.org/10.1016/S2214-109X\(18\)30024-X](https://doi.org/10.1016/S2214-109X(18)30024-X) PMID: 29398634
  24. Hill SC, Souza R, Thézé J, Claro I, Aguiar RS, Abade L, et al. Genomic surveillance of yellow fever virus epizootic in São Paulo, Brazil, 2016–2018. *PLoS Pathog*. 2020; 16: 2016–2018. <https://doi.org/10.1371/JOURNAL.PPAT.1008699> PMID: 32764827
  25. Ministério da Saúde. Classificação de Risco dos Agentes Biológicos. In: Ministério da Saúde [Internet]. 2017 [cited 2 Sep 2019] pp. 1–50. Available from: [http://bvsmms.saude.gov.br/bvsv/publicacoes/classificacao\\_risco\\_agentes\\_biologicos\\_3ed.pdf](http://bvsmms.saude.gov.br/bvsv/publicacoes/classificacao_risco_agentes_biologicos_3ed.pdf).
  26. Domingo C, Patel P, Yillah J, Weidmann M, Méndez JA, Nakouné ER, et al. Advanced yellow fever virus genome detection in point-of-care facilities and reference laboratories. *J Clin Microbiol*. 2012; 50: 4054–4060. <https://doi.org/10.1128/JCM.01799-12> PMID: 23052311
  27. Avelino-Silva VI, Leal FE, Sabino EC, Nishiya AS, Da Silva Freire M, Blumm F, et al. Yellow fever vaccine viremia following ablative BM suppression in AML. *Bone Marrow Transplant*. Nature Publishing Group; 2013; 48: 1008–1009. <https://doi.org/10.1038/bmt.2012.277> PMID: 23334273
  28. Pereira RH, Gonçalves CN. geobr: Loads Shapefiles of Official Spatial Data Sets of Brazil. GitHub repository; 2019.
  29. Cunha M dos P, Duarte-Neto AN, Pour SZ, Hajjar LA, Frassetto FP, Dolhnikoff M, et al. Systemic dengue infection associated with a new dengue virus type 2 introduction in Brazil—a case report. *BMC Infect Dis*. BioMed Central Ltd; 2021; 21: 1–6. <https://doi.org/10.1186/S12879-021-05959-2/FIGURES/4>
  30. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
  31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9: 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
  32. Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, et al. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*. 2012; 28: 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091> PMID: 22368248
  33. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23: 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036
  34. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol*. 2015; 1: 1–5. <https://doi.org/10.1093/ve/vev003> PMID: 27774277
  35. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014; 30: 3276–3278. <https://doi.org/10.1093/bioinformatics/btu531> PMID: 25095880

36. Nguyen L-T, Schmidt HA, Haeseler A von, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol.* 2015; 32: 268–274. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430
37. Rambaut A. FigTree Version 1.4.4. 2018.
38. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2016; 2: vew007. <https://doi.org/10.1093/ve/vew007> PMID: 27774300
39. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 2018; 4: 1–5. <https://doi.org/10.1093/vey016> PMID: 29942656
40. Darriba D, Taboada GL, Doallo R, Posada D. JModelTest 2: More models, new heuristics and parallel computing. *Nat Methods.* Nature Publishing Group; 2012; 9: 772. <https://doi.org/10.1038/nmeth.2109> PMID: 22847109
41. Mir D, Delatorre E, Bonaldo M, Lourenço-De-Oliveira R, Vicente AC, Bello G. Phylodynamics of Yellow Fever Virus in the Americas: New insights into the origin of the 2017 Brazilian outbreak. *Sci Rep.* 2017; 7: 1–9. <https://doi.org/10.1038/s41598-017-07873-7> PMID: 28785067
42. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol Biol Evol.* 2013; 30: 713–724. <https://doi.org/10.1093/molbev/mss265> PMID: 23180580
43. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol.* 2008; 25: 1459–1471. <https://doi.org/10.1093/molbev/msn090> PMID: 18408232
44. Rambaut A, Drummond AJ. Tracer v1.4. Available from <http://beast.bio.ed.ac.uk/Tracer>. Available from <http://beast.bio.ed.ac.uk/Tracer>. 2007.
45. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol.* 2013; 30: 239–243. <https://doi.org/10.1093/molbev/mss243> PMID: 23090976
46. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol.* 2012; 29: 2157–2167. <https://doi.org/10.1093/molbev/mss084> PMID: 22403239
47. Baele G, Lemey P, Suchard MA. Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty. *Syst Biol.* 2016; 65: 250–264. <https://doi.org/10.1093/sysbio/syv083> PMID: 26526428
48. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. Spred3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Mol Biol Evol.* 2016; 33: 2167–2169. <https://doi.org/10.1093/molbev/msw082> PMID: 27189542
49. Delatorre E, Abreu FVS, Ribeiro IP, Gómez MM, Santos AAC, Ferreira-De-Brito A, et al. Distinct YFV Lineages Co-circulated in the Central-Western and Southeastern Brazilian Regions from 2015 to 2018. *Front Microbiol.* 2019; 10: 1–12. <https://doi.org/10.3389/fmicb.2019.01079> PMID: 31178835
50. Possas C, Lourenço-de-Oliveira R, Tauil PL, Pinheiro F de P, Pissinatti A, Cunha RV da, et al. Yellow fever outbreak in Brazil: the puzzle of rapid viral spread and challenges for immunisation. *Mem Inst Oswaldo Cruz.* 2018; 113: e180278. <https://doi.org/10.1590/0074-02760180278> PMID: 30427974
51. CDC. Yellow Fever [Internet]. [cited 22 Jun 2022]. Available from: <https://www.cdc.gov/yellowfever/index.html>.
52. Rudolph KE, Lessler J, Moloney RM, Kmush B, Cummings DAT. Incubation Periods of Mosquito-Borne Viral Infections: A Systematic Review. *Am J Trop Med Hyg.* 2014; 90: 882–891. <https://doi.org/10.4269/ajtmh.13-0403> PMID: 24639305
53. Casadio L, Nastro AC, Malta FM, Araujo J, Silva JB, Salomao J, et al. Late-Onset Relapsing Hepatitis Associated with Yellow Fever. *new engl J Med.* 2020; 2059–2061. <https://doi.org/10.1056/NEJMc1913036> PMID: 32433844
54. Ferreira MS, Soares P, Júnior B, Cerqueira VD, Riet G, Rivero C, et al. Experimental yellow fever virus infection in the squirrel monkey (*Saimiri* spp.) I: gross anatomical and histopathological findings in organs at necropsy. *Mem Inst Oswaldo Cruz.* 2020; 115: 1–8. <https://doi.org/10.1590/0074-02760190501> PMID: 33174908
55. Costa ZGA, Romano APM, Elkhoury ANM, Flannery B. Evolução histórica da vigilância epidemiológica e do controle da febre amarela no Brasil. *Rev Pan-Amazônica Saúde.* 2011; 2: 11–26. <https://doi.org/10.5123/s2176-62232011000100002>
56. Fioravanti CH. O Combate à febre amarela no Estado de São Paulo: História, Desafios e Inovações. 2018.

57. Hamrick PN, Aldighieri S, Machado G, Leonel DG, Vilca LM, Uriona S, et al. Geographic patterns and environmental factors associated with human yellow fever presence in the Americas. *PLoS Negl Trop Dis*. 2017; 11: 1–27. <https://doi.org/10.1371/journal.pntd.0005897> PMID: 28886023
58. Di Couto-Lima, Madec Y Bersot MI, Campos SS Motta MDA, Dos Santos FB, et al. Potential risk of re-emergence of urban transmission of Yellow Fever virus in Brazil facilitated by competent *Aedes* populations. *Sci Reports* 2017 71. Nature Publishing Group; 2017; 7: 1–12. <https://doi.org/10.1038/s41598-017-05186-3> PMID: 28687779
59. Giovanetti M, Mendonça MCL, Fonseca V, Mares-Guia MA, Fabri A, Xavier J, et al. Yellow Fever Virus Reemergence and Spread in Southeast Brazil, 2016–2019. *J Virol*. 2020; 94: e01623–19.
60. Secretaria de Estado da Saúde—Santa Catarina. Aumento do número de epizootias em Primatas Não Humanos (PNH) no ano de 2020. *Nota Alerta 001/2020/DIVE/SUV/SES*. 2020; 1–2.
61. Secretaria da Saúde—Paraná. Monitoramento da Situação Epidemiológica da Febre Amarela no Paraná. *Inf epidemiológico n° 019/2020*. 2020; 1–7.
62. Andrade M de S, Campos FS, Campos AAS, Abreu FVS, Melo FL, Sevá A da P, et al. Real-Time Genomic Surveillance during the 2021 Re-Emergence of the Yellow Fever Virus in Rio Grande do Sul State, Brazil. *Viruses* 2021, Vol 13, Page 1976. Multidisciplinary Digital Publishing Institute; 2021; 13: 1976. <https://doi.org/10.3390/v13101976> PMID: 34696408