Universidade de São Paulo
Programa Interunidades de Pós-Graduação em Bioinformática

NATÁLIA BAPTISTA CRUZ

**Análise transcritômica de amostras humanas naturalmente
infectadas por Virus Chikungunya**

São Paulo
2019

NATÁLIA BAPTISTA CRUZ

Análise transcriptômica de amostras humanas naturalmente infectadas por Virus Chikungunya

Dissertação apresentada à Banca Examinadora específica do Programa de Pós-graduação Interunidades em Bioinformática para obtenção do título de Mestre em Ciências pelo Instituto de Matemática e Estatística da Universidade de São Paulo.

Orientador: Prof. Dr. Helder Takashi Imoto Nakaya

São Paulo

2019

FICHA CATALOGRÁFICA

C957 Cruz, Natália Baptista
Análise transcritômica de amostras humanas naturalmente infectadas por vírus Chikungunya / Natália Baptista Cruz, orientador Helder Takashi Imoto Nakaya. São Paulo, 2019.
88 f.

Dissertação (Mestrado) - Programa Interunidades de Pós-Graduação em Bioinformática, Universidade de São Paulo

1. Biologia de Sistemas. 2. Bioinformática. 3. Vírus Chikungunya. 4. Infecção. 5. Genes. I. Nakaya, Helder Takashi Imoto, orientador. II. Universidade de São Paulo. III. Título.

Elaborada pelo Serviço de Informação e Biblioteca Carlos Benjamin de Lyra do IME-USP, pela bibliotecária Eliana Mara Martins Ramalho CRB-8/4819

*À minha família e Matheus por todo o apoio prestado e incentivo durante essa
(e todas as fases)*

Agradecimentos

Agradeço aos meus pais por todo o suporte e ajuda oferecido em todos os momentos de minha vida, imprescindível para que tudo isso fosse possível.

Agradeço ao Matheus Calil por toda paciência e amor que me oferece em todos os momentos.

Agradeço a Viviane Schuch, André Nicolau, André Martins, Felipe Martins, Lucas Cardozo, Mariana Pereira, Patrícia Gonzalez, Thiago Hirata, Mindy Muñoz, Lucas Fleig, Diógenes Lima, César Prada, Leandro Jimenez, Deney Araujo, Alysson Urbanski, Tiago Lubiana Alves, Diogo Matos, Bruna Garbes, Fabio Pohl por toda a ajuda durante esses 2 anos, seja na questão técnica quanto com uma conversa cercada de café. Vocês com certeza tornaram esta jornada mais leve e com muito aprendizado.

Agradeço a Alessandra Schanoski pela oportunidade de um trabalho em conjunto.

Agradeço a Prof. Helder Nakaya que acreditou no meu potencial e permitiu que eu me desenvolvesse em seu laboratório.

Agradeço a Fundação Bill e Melinda Gates pelo apoio financeiro que garantiu que este projeto fosse realizado.

À Universidade de São Paulo, pela oportunidade de realizar esta pós-graduação.

Aos pacientes utilizados neste trabalho pois sem o auxílio deles esta pesquisa não existiria.

Powerful avalanches begin with small shifts

(Pamela Walsh)

Resumo

Baptista-Cruz, Natalia. Análise transcriptômica de amostras humanas naturalmente infectadas por Chikungunya Virus. 2019. Dissertação (Mestrado em Ciências) - Universidade de São Paulo, São Paulo, 2019

A Febre de Chikungunya é uma doença sistêmica causada por arbovírus e estima-se que afete cerca de 1 milhão de pessoas anualmente. Os principais sintomas associados com esta doença são febre e poliartralgia, podendo esta última assumir um caráter crônico em cerca de metade dos casos. Devido aos inúmeros surtos que ocorreram nos últimos 50 anos, diversos estudos tiveram como objetivo determinar os mecanismos de replicação do vírus e da resposta imune. Mesmo assim, pouco se sabe sobre quais moléculas possibilitam o processo de infecção. Já foi demonstrada a importância de interferons, principalmente de tipo I, citocinas e quimiocinas na diminuição da replicação viral. Além disso, há uma preferência do vírus por tipos celulares específicos como células endoteliais e epiteliais. Porém, estudos mostram informações contraditórias referentes ao papel de células mononucleares do sangue periférico (PBMC), principalmente com relação a monócitos e células B e T. Diante deste contexto, o tratamento utilizado atualmente é direcionado apenas ao alívio de sintomas uma vez que não existem drogas ou vacinas licenciadas específicas para esta doença. Logo, o objetivo deste trabalho é estudar as modificações transcricionais que ocorrem em amostras humanas durante o processo de infecção pelo vírus de Chikungunya de modo a esclarecer os mecanismos utilizados pelo sistema imune em resposta a infecção. Além disso, este estudo tem como finalidade apontar possíveis diferenças transcricionais entre amostras crônicas e não-crônicas.

Palavras-chaves: biologia de sistemas, bioinformática, transcritoma, Chikungunya, infecção viral, genes diferencialmente expressos, cronicidade.

Abstract

Baptista-Cruz, Natalia. Transcriptional analysis of human samples naturally infected by Chikungunya Virus. 2019. Dissertation (Master of Science) - University of São Paulo, São Paulo, 2019

The Chikungunya Fever is a systemic disease transmitted by arboviruses and is estimated to affect 1 million people annually. The main symptoms associated with this disease are fever and polyarthralgia, which can develop to chronic features in about half of the cases. Due to outbreaks that occurred in the last 50 years, many studies had the goal of determining the mechanisms of virus replication and immune response. Nevertheless, it is still poorly understood which molecules enable the occurrence of the infection process. It has already been shown the importance of interferons, mainly type I, cytokines and chemokines in restricting the viral replication. In addition, the Chikungunya virus shows a preference for specific cell types such as endothelial and epithelial cells. However, studies display contradictory information regarding the role of peripheral blood mononuclear cells (PBMC), mainly in relation to monocytes and B and T cells. Given this context, the treatment currently used is directed only to alleviate the symptoms since there are no specific licensed drugs or vaccines for this disease. Therefore, the objective of this work was to study the transcriptional modifications that occur in humans during the process of Chikungunya virus infection in order to clarify the mechanisms used by the immune system. In addition, it aims to point out possible transcriptional differences between sample from the Chronic and Non-Chronic acute phase of the infection.

Keywords: system biology, bioinformatics, transcriptome, Chikungunya, differentially expressed genes, chronicity.

Lista de Figuras

Figura 1 - Países em que casos locais de transmissão de Chikungunya foram reportados	15
Figura 2 - Fluxograma representando as etapas do projeto	29
Figura 3 - Perfil das amostras utilizadas neste estudo	41
Figura 4 - Dados clínicos e diagnósticos das amostras	42
Figura 5 - Análise Não Supervisionada realizada nas amostras infectadas com CHIKV	43
Figura 6 - Análise de Enriquecimento do Conjunto de Genes associada à viremia	44
Figura 7 - Assinatura gênica associada positivamente com viremia	46
Figura 8 - Associação dos genes EIF com viremia	48
Figura 9 - Perfil gênico das amostras infectadas por CHIKV	50
Figura 10 - Genes consistentemente diferencialmente expressos entre amostras de CHIKV	52
Figura 11 - Análise de Enriquecimento Funcional focado nas vias do BTM	54
Figura 12 - Alterações no transcrito de amostras Crônicas e / ou Não-Crônicas em comparação com indivíduos saudáveis	55
Figura 13 - Diferenças e/ou semelhanças na expressão gênica entre de amostras crônicas e/ou não-crônicas	56
Figura 14 - Módulos de genes de co-expressão obtidos através de CEMiTool	58
Figura 15 - Análises adicionais resultantes de CEMiTool no módulo 8	59
Figura 16 - Meta-análise de expressão gênica diferencial	60
Figura 17 - Comparação dos classificadores treinados	63
Figura 18 - Matriz de Confusão nos Dados de Teste utilizando o modelo de Rede Neural	64

Lista de Tabelas

Tabela 1 - Genes com maior associação com variável resposta pelo método de regressão de Lasso

62

Lista de Abreviaturas e siglas

BTM	<i>Blood Transcripcional Modules</i>
CHIKV	Vírus Chikungunya
CT	<i>Cycle Threshold</i>
DEGs	Genes Diferencialmente Expressos
DNA	Ácido desóxirribonucleico
ECSA	Leste-Centro-Sul Africano
eIFs	Fatores Eucarióticos Iniciadores de Tradução
ELISA	<i>Enzyme-Linked Immunosorbent Assay</i>
GEO	<i>Gene Expression Omnibus</i>
GSEA	<i>Gene Set Enrichment Analysis</i>
IFN	Interferon
IgG	Imunoglobulina G
IgM	Imunoglobulina M
ISG	<i>Interferon-Stimulated Genes</i>
KNN	<i>K-Nearest Neighbors</i>
MHC	Complexo Principal de Histocompatibilidade
NES	Escore de Enriquecimento Normalizado
NIAID	Instituto Nacional de Alergia e Doenças Infecciosas dos Estados Unidos
pb	pares de bases
PBMC	<i>Peripheral blood mononuclear cell</i>
PRRs	Receptores de reconhecimento de padrões
PCR	Reação em cadeia da polimerase
PPI	Interação Proteína-Proteína
PVCA	Análise dos Componentes Variantes Principais
qRT-PCR	<i>Reverse transcription polymerase chain reaction quantitative real time</i>

RA	Artrite Reumatóide
RLRs	receptores do tipo RIG-I
RNA	Ácido Ribonucléico
RNA-seq	Sequenciamento de RNA
SVM	<i>Support Vector Machine</i>
TLRs	Receptores do tipo Toll
TMM	<i>Trimmed Mean of M-Value</i>
XGBoost	<i>eXtreme Gradient Boosting</i>

Sumário

1 INTRODUÇÃO	14
1.1 FEBRE DE CHIKUNGUNYA E SUAS PROBLEMÁTICAS	14
1.2 VÍRUS CHIKUNGUNYA	16
1.3 MANIFESTAÇÕES CLÍNICAS	17
1.3.1 Cronicidade	18
1.4 INFECÇÃO PELO VÍRUS CHIKUNGUNYA	18
1.5 RESPOSTA IMUNE INDUZIDA PELA INFECÇÃO POR CHIKV	19
1.5.1 Sistema imune inato	19
1.5.2 Sistema imune adaptativo	21
1.6 DIAGNÓSTICO	22
1.7 TRATAMENTO E PREVENÇÃO	22
1.8 BIOLOGIA DE SISTEMAS	24
1.9 ANÁLISE DE TRANSCRITOMA	24
1.9.1 Alterações transcricionais causadas pela infecção por CHIKV	25
2 OBJETIVOS GERAIS	27
2.1 OBJETIVO PRINCIPAL	27
2.2 OBJETIVOS ESPECÍFICOS	27
3 MATERIAIS E MÉTODOS	28
3.1 DADOS UTILIZADOS	28
3.1.1 Amostras humanas infectadas por CHIKV	29
3.1.1.1 Informações dos pacientes	29
3.1.1.2. Diagnóstico	30
3.1.1.3. Detecção de possível artralgia crônica	31
3.1.1.4 Preparação de biblioteca de transcrito e sequenciamento	31
3.1.2. Estudos Provenientes do GEO	31
3.2. PROCESSAMENTO DAS AMOSTRAS DE RNA-SEQ	32
3.2.1. Controle de Qualidade	32
3.2.2. Alinhamento de sequências e Quantificação de Genes	32
3.2.3. Variabilidade nos dados	33
3.2.4. Análise de Expressão Diferencial	33
3.2.4.1. Genes diferencialmente expressos entre amostras Crônicas e Não-Crônicas	34
3.2.5. Correlação Genes-Viremia	34
3.2.6. Análise de Redes de Co-Expressão	35

3.2.7. Meta-análise	36
3.2.8. Network Analysis	37
3.2.8. Feature Selection e Machine Learning	37
4. Resultados	40
4.1. INFORMAÇÃO CLÍNICA E DIAGNÓSTICO DE AMOSTRAS INFECTADAS POR CHIKV	40
4.2 O IMPACTO DE INFECÇÃO POR CHIKV NO TRANSCRITOMA DO SANGUE	42
4.3. NÚMERO MAIOR DE GENES DIFERENCIALMENTE EXPRESSOS NOS PRIMEIROS DIAS DA INFECÇÃO	49
4.4. ASSINATURA GÊNICA EM RESPOSTA A INFECÇÃO POR CHIKV	53
4.5. POTENCIAIS ALVOS DE ARTRALGIA CRÔNICA E NÃO-CRÔNICAS	54
4.6. DIFERENTES DURAÇÕES DE SINTOMAS REVELA PERFIS GÊNICOS DISTINTOS	57
4.7. PERFIL GÊNICO COMUM ENTRE FEBRE DE CHIKUNGUNYA, ARTRITE E DENGUE	59
4.7. FEATURE SELECTION	61
5. Discussão	65
6. Conclusão	68
Referências	69
Apêndices	86

1 INTRODUÇÃO

1.1 FEBRE DE CHIKUNGUNYA E SUAS PROBLEMÁTICAS

Febre de Chikungunya é uma doença tropical sistêmica debilitante transmitida por mosquitos e está associada principalmente a sintomas como febre e poliartralgia. Apesar de sua taxa de mortalidade ser baixa (inferior a 0,1%), as preocupações relacionadas a esta doença estão relacionadas à proporção de casos sintomáticos, incidência de transmissão vertical (transmissão de mãe para feto) observada em alguns estudos (GÉRARDIN *et al.*, 2008) e a possibilidade de cronicidade da doença que, em casos raros, pode durar até seis anos (JAVELLE *et al.*, 2015).

Durante os últimos 50 anos houve diversos surtos, principalmente na Ásia e África, com intervalos irregulares de 2 a 20 anos (POWERS; LOGUE, 2007). Um dos mais graves e recentes ocorreu na Índia em 2015 e afetou aproximadamente 1,4 milhões de pessoas (RAVI, 2006), seguido por um surto na Ilha Reunião, uma ilha no Oceano Índico, afetando cerca de um terço da população local (ENSERINK, 2006). Devido a turistas infectados em regiões endêmicas, a doença se alastrou para outras zonas do globo e estima-se que ocorram 1 milhão de casos anualmente (WEAVER *et al.*, 2012).

Atualmente a doença já foi identificada em mais de 110 países (Fig. 1) e em 2008 foi listada como um patógeno de prioridade C pelo Instituto Nacional de Alergia e Doenças Infecciosas dos Estados Unidos (NIAID). Esta categoria é destinada a agentes patogênicos emergentes que poderiam ser projetados para disseminação em massa no futuro em razão da presença em diversos locais e do potencial para grande impacto na saúde devido a altas taxas de morbidade, e possível mortalidade (POWERS; LOGUE, 2007). Os principais motivos apontados para a pandemia observada são a globalização, propagação dos vetores e ausência de formas de prevenção e tratamento específicas, como vacina.

Figura 1 - Países em que casos locais de transmissão de Chikungunya foram reportados.

Fonte: *Centers for Disease Control and Prevention*¹

De acordo com dados do Ministério da Saúde (SECRETARIA DE VIGILÂNCIA EM SAÚDE, 2018), entre janeiro e outubro de 2018 foram registrados mais de 80.010 possíveis casos de febre de Chikungunya no Brasil, sendo confirmados mais de 59.584. Assim sendo, a doença representa um grave problema econômico e de saúde pública.

Uma das razões dos diversos surtos ocorridos de Febre de Chikungunya mencionados anteriormente é a inexistência de tratamentos ou vacinas específicas para esta doença. Ao invés disso, as intervenções realizadas são focadas em tratar apenas os sintomas, o que pode tornar o processo de cura mais demorado e mais custoso, a depender dos medicamentos utilizados. A isso, soma-se outros custos indiretos como dias perdidos de trabalho e diminuição de produtividade das pessoas infectadas. BLOCH (2016) estima que entre 2013-2015, a doença ocasionou um custo de 185 bilhões de dólares de maneira direta e indireta nas Américas.

¹ Disponível em <<https://www.cdc.gov/chikungunya/geo/index.html>>. Acesso em: 12 de junho de 2018

1.2 VÍRUS CHIKUNGUNYA

Junto com outros vírus responsáveis por doenças similares como Febre de Mayaro e Febre do rio Ross, o vírus Chikungunya (CHIKV) pertence à família *Togaviridae* e ao gênero *Alphavirus*. Seu genoma é composto por uma fita simples de RNA de sentido positivo e possui tamanho estimado de 11.8 kpb (pares de bases), sendo responsável por codificar quatro proteínas não-estruturais (nsP1-4) e cinco proteínas estruturais (C, E1, E2, E3 e 6K) (PORTA *et al.*, 2015). Suas proteínas não estruturais são necessárias para a replicação do vírus enquanto que as estruturais, que incluem o capsídeo e proteínas do envelope viral, viabilizam o processo de endocitose e entrada do vírus na célula hospedeira.

Foram identificadas três linhagens de CHIKV, com genotipagem e características distintas agrupadas em diferentes grupos ou *clusters*. Amostras do vírus que causou o surto que ocorreu entre 2004 e 2006 no Oceano Índico formam o *cluster* Leste-Centro-Sul Africano (ECSA), além dos *clusters* Oeste Africano e Asiático que já haviam sido identificados anteriormente (SCHUFFENECKER *et al.*, 2006). No território brasileiro é possível observar casos de infecção provenientes das linhagens Asiática e ECSA (NUNES *et al.*, 2015).

O vírus possui dois ciclos de transmissão distintos: ciclo enzoótico/silvático e urbano (PIALOUX *et al.*, 2007). O primeiro consiste na ação de diversas espécies de mosquitos do gênero *Aedes* agindo como vetores e primatas não-humanos como hospedeiros amplificadores. Este ciclo normalmente está associado com o continente Africano e uma transmissão para humanos normalmente só ocorre no caso em que as pessoas residam nos arredores dos hospedeiros amplificadores.

Já o ciclo de transmissão urbano apresenta um comportamento distinto e só se estabeleceu quando o vírus foi introduzido em outras regiões do globo. Inicialmente seu principal vetor era o mosquito *Aedes aegypti*, porém mutações na linhagem ECSA, como em E1, proteína viral que medeia a entrada viral em pH baixo, possibilitou a adição de mais um vetor de transmissão, o *Aedes albopictus*

(SCHUFFENECKER *et al.*, 2006). Esta incorporação influenciou diretamente o aumento da propagação da doença devido a presença deste vetor em um território maior e maior resistência de seus ovos a condições climáticas adversas (SCHUFFENECKER *et al.*, 2006). Interessante notar que esta mutação não influenciou o processo de replicação viral em *Aedes aegypti* (TSETSARKIN *et al.*, 2007).

1.3 MANIFESTAÇÕES CLÍNICAS

Febre de Chikungunya possui um período de incubação de 2-4 dias seguido de um início de sintomas abrupto com duração de cerca de 4-7 dias, caracterizando assim a fase aguda da doença (KRUTIKOV; MANSON, 2016). As principais manifestações relatadas pelos pacientes consistem em febre alta (maior que 39°C), mialgia, dor de cabeça, erupção cutânea e artralgia. As dores e inchaços nas articulações normalmente são descritas como intensas, afetando principalmente pulsos, tornozelos, mãos e pés (WEAVER, 2014). Imagens radiológicas apresentam padrões normais e marcadores biológicos de inflamação são normais ou moderadamente elevados. Ao contrário de doenças similares, como Dengue e Zika, a proporção de casos assintomáticos é baixa: cerca de apenas 15% (BROUARD *et al.*, 2008).

A doença normalmente não apresenta um risco de óbito, porém existe o agravante com pacientes idosos ou com fatores de risco independentes da doença, como desordens cardiovasculares, respiratórias ou até mesmo diabetes (JAVELLE *et al.*, 2008). Outro grupo de risco já descrito é de neonatos e crianças pequenas (LEWTHWAITE *et al.*, 2009).

Apesar de poucos estudos abordarem características clínicas específicas apresentadas em crianças infectadas, é possível notar diferenças no quadro clínico. Febre alta está associada em ambos os casos, porém em pacientes infantis, já foi descrito o acompanhamento de convulsões com duração de cerca de 3-5 dias

(SEBASTIAN *et al.*, 2009). Além disso, a incidência de artralgia e artrite é considerada menor em crianças.

1.3.1 Cronicidade

Um dos aspectos mais problemáticos da doença é em relação a severidade da fase aguda e a possibilidade de artralgia crônica. As manifestações sintomáticas desta fase são similares às relatadas na fase aguda da doença, incluindo as principais articulações afetadas. Esta fase pode persistir por anos e pode estar associada com artrite e artrite reumatóide (RA), em até 50% dos pacientes infectados por CHIKV (SCHILTE *et al.*, 2013).

Alguns fatores são descritos como sendo capazes de aumentar a possibilidade do desenvolvimento deste aspecto crônico, como persistência viral, pré-disposição genética, indução de doença auto-imune ou até mesmo lesão tecidual causada pelo vírus (GOUPIL *et al.*, 2016).

Além disso, um estudo também fez uma associação de que carga viral alta encontrada em amostras de sangue durante a fase aguda da doença não corresponde a níveis detectáveis de RNA viral em células sanguíneas (SOURISSEAU *et al.*, 2007), mostrando assim uma não correlação da severidade da doença em relação a carga viral detectada.

1.4 INFECÇÃO PELO VÍRUS CHIKUNGUNYA

Apesar de ser uma doença amplamente estudada devido aos recentes surtos, os mecanismos de replicação do vírus e resposta da célula hospedeira ainda permanecem pouco compreendidos.

Após a inoculação do vírus através da picada do mosquito vetor, sabe-se que sua ligação aos receptores da célula hospedeira, principalmente em fibroblastos na derme, ocorre através da proteína E2 (VAN DUIJL-RICHTER *et al.*, 2015). Apesar de não haver um consenso sobre quais moléculas possibilitam este processo, há alguns indícios do papel de algumas moléculas. Alguns autores identificaram prohibitina

como uma proteína de ligação para CHIKV expressas em células da microglia (WINTACHAI *et al.*, 2012) enquanto outros estudos sugerem que proteínas de choque térmico podem ter um papel crucial na nesta ligação (DAS *et al.*, 2014). Um estudo mais recente identificou *Mxra8* como um receptor para múltiplos alphavírus que causam doenças com caráter artritogênico (ZHANG *et al.*, 2018)

Em seguida, ocorre o processo de endocitose, resultando em um ambiente de pH baixo e consequente mudança conformacional da célula hospedeira. Deste modo, é viabilizada a fusão do envelope viral e da membrana do endossomo em um processo mediado pela proteína E1, acarretando na liberação do nucleocapsídeo e genoma viral ao citoplasma (SOURISSEAU *et al.*, 2007). Alguns autores descrevem que o processo de endocitose deve-se principalmente a mediação por clatrina (LEE *et al.*, 2013), porém outros propõem que deve-se a uma via dependente de Eps15, uma proteína também envolvida neste processo (BERNARD *et al.*, 2010).

Dentro da célula hospedeira, o nucleocapsídeo fornece o genoma viral no citosol para tradução, de modo a codificar as proteínas não-estruturais. Porém esta replicação local aparenta ser pequena e limitada no tempo, com os vírus sendo transportados através da corrente sanguínea para outros locais, onde este processo se torna mais intenso e a presença viral mais persistente, como fígado, músculo, órgãos linfáticos (BORDI *et al.*, 2015). Estudos realizados com cultura de tecidos humanos mostraram que CHIKV é capaz de se replicar em diferentes tipos celulares aderentes, como células endoteliais e epiteliais, fibroblastos, macrófagos (SOURISSEAU *et al.*, 2007) e células satélites musculares (OZDEN, 2007).

1.5 RESPOSTA IMUNE INDUZIDA PELA INFECÇÃO POR CHIKV

1.5.1 Sistema imune inato

Diante da infecção viral, o sistema imune inato é a primeira linha de defesa do organismo a ser ativada de modo a limitar a propagação, replicação e disseminação do vírus. Neste momento, há uma alta concentração de interferons, acompanhada

por um aumento de produção de citocinas pró-inflamatórias e quimiocinas, como CXCL9, CXCL10, CCL2, IL-6, IL-12 e MCP-1 (NG *et al.*, 2009).

Interferons (IFN) são proteínas que prejudicam a replicação de patógenos virais e estimulam a atividade de defesa de outras células, possuindo assim um papel central no controle da gravidade de infecções virais, principalmente em relação a alphavírus (RYMAN *et al.*, 2007). São divididos em três classes, sendo que Interferons do Tipo I, ao desencadear a via de sinalização JAK/STAT, induzem a expressão de genes estimulados por interferon (ISGs). Apesar de não serem muito bem caracterizados, sabe-se que possuem um papel crucial na defesa do hospedeiro devido a sua capacidade de sintetizar inúmeras proteínas antivirais (SCHOGGINS; RICE, 2012). No contexto de CHIKV, apenas um ISG já foi descrito como envolvido no controle viral. Células HeLA transfectadas com OAS3 se mostraram mais resistente à replicação do vírus de Chikungunya (BREHIN *et al.*, 2009), apesar de não se saber o mecanismo pelo qual isso ocorre. Além disso, em estudos realizados com ratos que não apresentavam o receptor de IFN do tipo I, mostrou-se uma maior suscetibilidade a uma forma mais grave da doença (GARDNER *et al.*, 2012). Ademais, quando interferons do tipo I são adicionados em células humanas antes da infecção, a replicação *in vitro* de alphavírus responsáveis por doenças como Febre de Chikungunya, Encefalite Equina do Oeste e Febre de Sindbis é suprimida (GARDNER *et al.*, 2010; FROS *et al.*, 2010).

A produção desta resposta através de interferons é desencadeada por receptores de reconhecimento de padrões (PRRs) que detectam padrões moleculares de patógenos ao entrarem em contato com o meio celular. Estudos demonstram que fibroblastos infectados por CHIKV são capazes de secretar IFN do tipo I (GIFFORD; HELLER. 1963), principalmente devido a uma via dependente de uma proteína de sinalização antiviral mitocondrial denominada Cardif (SCHILTE *et al.*, 2010). Receptores do tipo Toll (TLRs) e receptores do tipo RIG-I (RLRs) representam duas classes de PRRs que possuem um papel crucial na resposta imunológica diante da infecção por CHIKV. TLRs referem-se a uma família de

proteínas transmembranas pertencentes ao sistema imunológico inato que detectam padrões moleculares de patógenos e são conhecidas por estarem envolvidas na imunidade antiviral ao ativarem a expressão de citocinas e genes relacionados à resposta imunológica (AKIRA; TAKEDA, 2004). Já RLRs são importantes complexos intracelulares que também são capazes de reconhecer fragmentos do genoma do patógeno no citosol e agem como sensores de agentes patogênicos (Schilte *et al.*, 2010); RIG-I e MDA-5 são exemplos destes complexos.

1.5.2 Sistema imune adaptativo

Durante o processo de replicação e formação de novas proteínas virais, algumas são degradadas em pequenos peptídeos e capturadas pelo complexo principal de histocompatibilidade (MHC), o qual as apresenta aos linfócitos T. A resposta do sistema imune adaptativo inicia-se através da ativação e proliferação de células T CD8+ logo nos estágios iniciais da doença, cuja função principal é indução da morte celular das células infectadas a partir do reconhecimento destes peptídeos específicos apresentados (WAUQUIER *et al.*, 2010). Estágios mais avançados da fase aguda da doença são caracterizados principalmente pela produção de receptores como IL-1R α e IL-2R α , aliada a uma resposta de células T CD4+, a qual expressa Th1, Th2, Th17, células relacionadas a quimiocinas e citocinas. Alguns estudos apontam que células T CD4+ podem ser o principal mediador de inflamação no inchaço das articulações e danos nos tecidos em modelo de rato, estando diretamente relacionada com a cronicidade de alguma forma (TEO *et al.*, 2012).

Um outro ponto importante refere-se a contradição na literatura, principalmente quando se trata do papel de células mononucleares do sangue periférico (PBMC) durante a infecção. Enquanto um estudo realizado *in vitro* detectou que estas células (principalmente células B e T e monócitos) não eram suscetíveis a infecção por CHIKV (SOURISSEAU *et al.*, 2007), já foi mostrado que antígenos do vírus foram detectados em monócitos *in vitro* após uma inoculação com uma carga viral alta (HER *et al.*, 2010). Dessa forma é necessário realizar uma melhor análise principalmente em relação aos monócitos avaliando se estes são os

verdadeiros alvos do vírus de Chikungunya ou se estes antígenos virais foram encontrados nestas células devido ao processo de fagocitose realizado pelas mesmas.

1.6 DIAGNÓSTICO

O diagnóstico de Chikungunya é baseado principalmente nos dados clínicos, normalmente devido a associação de febre e artralgia em pacientes residentes em áreas endêmicas (STAIKOWSKY *et al.*, 2009). Um diagnóstico mais definitivo é realizado através da detecção do vírus por meio de um exame sorológico (de sangue), onde uma análise de PCR quantitativo em tempo real é realizada durante a fase virêmica da doença.

Outra alternativa utilizada para o diagnóstico é a quantificação de anticorpos IgM e IgG, proteínas produzidas pelo sistema imune cuja função é ligar-se ao antígeno (no caso o vírus Chikungunya) para que uma resposta seja desencadeada. IgM é o primeiro anticorpo produzido e é detectável desde o quinto dia da viremia até 3 meses após a infecção, geralmente estando relacionado com a fase aguda da doença (WEAVER; LECUIT, 2015). Já IgG está relacionado com a imunidade a longo-prazo, indicando se houve infecções anteriores.

Apesar da existência destes métodos, ainda não é possível determinar previamente se o paciente desenvolverá uma cronicidade. O que alguns estudos foram capazes de mostrar é que existe uma correlação entre a atividade da doença e os níveis de IgM e IgG (SCHILTE *et al.*, 2013).

1.7 TRATAMENTO E PREVENÇÃO

Atualmente não existem drogas licenciadas com o objetivo de prevenção ou melhora do quadro clínico do paciente já infectado. Neste caso, o tratamento é direcionado apenas ao alívio dos sintomas, através de medicamentos não-esteróides como antitérmicos ou analgésicos para alívio de dor nas articulações. Já se foi testado o uso de esteróides, porém não mostrou melhoras significativas (TAUBITZ

et al., 2007). No caso de artralgia crônica, anti-inflamatórios não esteróides e imunossupressores são utilizados, apesar de sua eficácia ainda não ter sido demonstrada em testes clínicos (KRUTIKOV; MANSON, 2016).

Vacinação já se mostrou uma opção preventiva e terapêutica apropriada em várias infecções virais em humanos, inclusive mostrando indicadores positivos em doenças com sintomas similares como Dengue (WORLD HEALTH ORGANIZATION, 2016). De acordo com experimentos realizados em modelos animais, o organismo aparenta desenvolver uma imunidade de longa duração após infecção prévia de CHIKV, até mesmo uma proteção parcial contra outros alphavírus (EDELMAN *et al.*, 2000).

Questões de conduta dos testes clínicos em humanos, custo de dose e detalhes do processo da infecção permanecem em aberto. Algumas das abordagens que revelaram algum sucesso entre as vacinas em teste são: utilização do vírus inativo completo, vacinas vivas atenuadas ou até mesmo proteínas recombinantes e partículas semelhantes ao vírus (AKAHATA *et al.*, 2010). Um dos exemplos de projeto focado no desenvolvimento de vacina fez uso de vírus inativo completo que, apesar da eficácia comprovada em seroconversão, acabou sendo abandonado devido a principalmente altos custos de manufaturação (HARRISON *et al.*, 1971). Um dos projetos mais recentes consiste na modificação de elementos de RNA que permitem a iniciação do processo de tradução independente de cap, sendo capaz de mostrar resultados promissores de segurança e eficácia em diversos modelos clínicos (PLANTE *et al.*, 2011).

Diante deste cenário, o controle da doença é exclusivamente baseado no combate aos mosquitos vetores (EISEN *et al.*, 2009). Algumas práticas utilizadas são focadas na diminuição de objetos que atuam como criadouros, estimular o uso de repelente, redução da população de mosquitos através de larvicidas ou até mesmo mediante libertação de vetores transgênicos. Porém estas medidas mostraram pouco eficiência, mostrando a necessidade de candidatos a vacinas para um melhor controle da doença.

1.8 BIOLOGIA DE SISTEMAS

Uma das maneiras de se estudar as interações e o comportamento dos vários componentes do sistema biológico em um contexto complexo como a Febre de Chikungunya, é através da Biologia de Sistemas. Trata-se de uma abordagem interdisciplinar que procura descrever as interações complexas realizadas em um sistema biológico, elucidando assim novas regras capazes de prever sua resposta a possíveis perturbações. Para que esse objetivo fosse concluído, é necessário que haja a integração de dados de tecnologias de larga escala de diferentes moléculas envolvidas nos processos biológicos (MINORSKY et al., 2003). Através desta integração, é realizada a inferência de possíveis redes que possam explicar o dinamismo e o funcionamento dos diversos mecanismos moleculares observados. Este tipo de análise já foi utilizado em diversos contextos e permitindo a descoberta de novos biomarcadores para doenças (WANG et al., 2010), estratificação de pacientes de acordo com seu perfil genético (SORLIE et al., 2001) e descoberta de novas drogas (BUTCHER; BERG; KUNKEL, 2004).

1.9 ANÁLISE DE TRANSCRITOMA

Transcritoma trata-se do conjunto de todos os transcritos de RNA expressos na célula em um estágio específico de desenvolvimento ou diante a uma resposta fisiológica específica. Com a utilização desta abordagem, torna-se possível decifrar quais elementos do genoma, codificantes ou não, são expressos e/ou como são regulados. Após o surgimento da tecnologia de sequenciamento de nova geração (NGS), o custo para gerar estes dados diminuiu significativamente em relação aos métodos anteriores (MUIR et al, 2016).

Atualmente, uma das maneiras mais utilizadas de se realizar análise de transcritoma é através da abordagem do RNA-Seq, o que revolucionou a maneira como os cientistas lidam com estes dados e proporcionou grande sucesso na descoberta de novas drogas e desenvolvimento clínico (WANG; GERSTEIN;

SNYDER, 2009). Ao contrário das técnicas anteriores baseadas em microarray, RNA-Seq evita algumas limitações técnicas como variação do desempenho da sonda e hibridização cruzada (ZHAO *et al.*, 2014). Esta técnica é capaz de medir os níveis de expressão de milhares de genes simultaneamente através de seu sequenciamento, fornecendo informações sobre vias funcionais e redes regulatórias que podem estar envolvidas no contexto estudado, e até mesmo sobre a ocorrência de splicing alternativo e novos transcritos.

1.9.1 Alterações transcricionais causadas pela infecção por CHIKV

Diversos estudos já analisaram os perfis de expressão gênica de pacientes infectados com Chikungunya, apesar de a maioria dos estudos utilizarem abordagens *in vitro* e modelos animais (ABDELNABI *et al.*, 2018; Nakaya *et al.*, 2012).

Wilson *et al.* (2017) identificou que a maioria dos genes que obtiveram uma regulação positiva no estudo de RNA-seq realizados com amostras de camundongos estavam associados com inflamação e já haviam sido descritos em estudos anteriores. MX1, CXCL10, CCL2, IRF1, IRF5 são alguns exemplos obtidos desta lista.

Michlmayr *et al.* (2018) realizou um estudo com amostras de criança naturalmente infectadas através da picada do mosquito vetor e foi possível observar uma regulação positiva de genes como TNFSF15, TNFSF10 E TNFSF13B, IL10, além de genes que são componentes conhecidos de vias de imunidade inata anotadas no KEGG (Ogata *et al.*, 1999). Genes que atuam como sensores celulares para RNA viral como RIG-I e MDA, genes relacionados com a via de sinalização JAK/STAT e receptores do tipo Toll como TLR1, TLR2, TLR3, TLR5 e TLR7/8 também apresentaram uma regulação positiva.

Nakaya *et al.* (2012) realizou um estudo de *microarray* com amostras de camundongos e foi possível observar que genes associados à ativação de macrófagos, ativação e movimentação de fagócitos e resposta de IFN do tipo I foram regulados positivamente nos grupos de infecção. Além disso, genes que

apresentaram uma regulação negativa estavam associados principalmente a tecido muscular.

Apesar destes estudos, ainda é necessário uma melhor compreensão dos mecanismos utilizados pela infecção por CHIKV, principalmente em um estudo com amostras humanas naturalmente infectadas.

2 OBJETIVOS GERAIS

2.1 OBJETIVO PRINCIPAL

Estudar modificações transcricionais em amostras humanas infectadas pelo vírus Chikungunya por meio da abordagem de biologia de sistemas de modo a elucidar os mecanismos ocorridos durante a infecção.

2.2 OBJETIVOS ESPECÍFICOS

1. Identificar genes diferencialmente expressos durante a infecção de CHIKV
2. Identificar vias gênicas que podem caracterizar a doença
3. Identificar possíveis alterações entre amostras crônicas e não-crônicas da doença
4. Identificar as diferentes assinaturas moleculares existentes entre contextos semelhantes a Chikungunya

3 MATERIAIS E MÉTODOS

3.1 DADOS UTILIZADOS

Os dados iniciais utilizados neste projeto consiste de amostras coletadas de pacientes infectados por Chikungunya e indivíduos saudáveis. Com o intuito de comparar assinaturas moleculares com contextos similares, foram utilizados dois grupos de pacientes adicionais disponibilizados no banco de dados *Gene Expression Omnibus* (GEO) com números de identificação GSE51808 e GSE93272. Um fluxograma geral das análises realizadas neste trabalho pode ser observado na Figura 2.

Figura 2 - Fluxograma representando as etapas do projeto. Arquivos estão representados pelas formas em azul claro e processos em verde

Fonte: Elaborada pelo autor

3.1.1 Amostras humanas infectadas por CHIKV

3.1.1.1 Informações dos pacientes

Amostras de sangue provenientes de humanos foram coletadas nos estados de Sergipe e São Paulo. No total, foram adquiridas 39 amostras de pacientes infectados e 20 de indivíduos saudáveis utilizados como controle nas análises. Todos os participantes com a infecção receberam atendimento em unidades básicas

de saúde reportando sintomas característicos de infecção por arbovírus. As amostras foram coletadas com o objetivo de ser realizado um diagnóstico sorológico, seguido por um formulário de consentimento e um questionário clínico e social-demográfico. Além disso, sangue periférico total foi coletado entre os dias 1 e 20 após o início dos sintomas, como relatado nos questionários, e armazenado de maneira apropriada para as análises subsequentes. Como relatado em um estudo prévio realizado com estas amostras, os indivíduos infectados foram muito provavelmente infectados pela linhagem ECSA, abordado previamente (CUNHA *et al.*, 2017).

Importante ressaltar que este estudo já foi aprovado previamente pelos Comitês de Ética das Universidades ligadas à coleta das amostras (Protocolos: 54937216.5.0000.5467 e 54835916.2.0000.5546 referentes à Universidade de São Paulo e Universidade de Sergipe respectivamente).

3.1.1.2. Diagnóstico

A extração de ácido nucleico destas amostras foi realizada utilizando o *QIAamp Viral RNA Mini Kit* (Qiagen, Valencia, CA, USA) de acordo com as instruções do fabricante. Uma análise de PCR Quantitativo em tempo real foi realizada nas amostras, utilizando o *SuperScript III Platinum One-Step qRT-PCR kit* (Invitrogen), com o intuito de realizar o diagnóstico de Chikungunya, Zika e Dengue. Todas as amostras positivas para CHIKV foram negativas para as outras doenças enquanto que as amostras controle apresentaram resultado negativo para todas as doenças testadas.

O diagnóstico nas amostras também foi realizado a partir do aspecto sorológico, usando o kit comercial semi-quantitativo ELISA *Euroimmun* de acordo com as instruções do fabricante. As amostras de soro foram diluídas em uma microplaca cujos antígenos específicos são utilizados como fase sólida. Caso a amostra já tenha tido contato com o antígeno testado, anticorpos específicos serão ligados e posteriormente detectados com uma solução que proverá cor, sendo esta proporcional à concentração de anticorpos (YOLKEN, 1982). Neste projeto, o

diagnóstico sorológico foi utilizado como forma de detectar anticorpos IgM e IgG específicos para CHIKV, IgG específicos para Zika e IgG específicos para Dengue. Valores inferiores a 0.8 foram considerados negativos, entre 0.8 e 1.1 como incertos e maior que 1.1 como positivos.

3.1.1.3. Acompanhamento de possível artralgia crônica

Um aspecto importante desta doença é a possível cronicidade que pode se desenvolver. Após a confirmação do diagnóstico foi possível acompanhar o progresso da doença de 13 pacientes que aceitaram ser voluntários para a pesquisa. Durante 2 anos, os mesmos foram monitorados e só foram considerados crônicos aqueles que ainda apresentassem dor associada, ou não com edema, nas regiões articular ou periarticular após 3 meses da fase aguda da doença. Por fim, dentre as amostras, 6 pacientes foram classificados como pacientes crônicos e 7 como não-crônicos. O acompanhamento dos demais pacientes não foi possível devido principalmente pela perda de contato.

3.1.1.4 Preparação de biblioteca de transcrito e sequenciamento

Amostras de RNA total foram purificadas com *Tempus™ Spin RNA Isolation Kit* ou com *Ribo Pure™ blood isolation kit* (ambas da empresa Invitrogen) dependendo da maneira como o material foi mantido em RNAlater. Em seguida, as etapas de remoção de RNA ribossomal, RNA mensageiro codificador de globina, fragmentação de RNA, geração de DNA complementar, ligação de adaptador e amplificação por PCR foram realizados utilizando RNA total de cadeia TruSeq com kit de preparação de amostra de globina ribo-zero da empresa Illumina. Logo após, as amostras foram sequenciadas em um Illumina HiSeq 1500 platform (Illumina, San Diego, CA).

3.1.2. Estudos Provenientes do GEO

Como mencionado anteriormente, foram selecionados dois estudos de contextos similares a Chikungunya do banco de dados GEO. O estudo GSE93272

consiste de uma análise de diferenças transcricionais entre pacientes que sofrem de Artrite Reumatóide e pacientes saudáveis (TASAKI *et al.*, 2018). Apenas os pacientes pertencentes ao primeiro lote foram utilizados nas análises, totalizando o uso de 30 amostras saudáveis e 30 amostras de pacientes que apresentavam a doença.

Já o estudo GSE51808 possui como objetivo analisar a resposta do sistema imune mediante infecção por Dengue em pacientes residentes em Bangkok (KWISSA *et al.*, 2014). Foram utilizados 47 pacientes infectados deste estudo, cujo fase da doença variava entre fase aguda ou convalescente, e 9 amostras saudáveis utilizadas como controle.

Para ambos os estudos, foram utilizados os dados de expressão diferencial resultantes das análises realizadas pelos próprios autores dos artigos, nos quais os dados foram retirados.

3.2. PROCESSAMENTO DAS AMOSTRAS DE RNA-SEQ

3.2.1. Controle de Qualidade

Após o alinhamento realizado nas amostras de pessoas infectadas com Chikungunya, os fragmentos *paired-end* foram pré-processados de forma a garantir o controle de qualidade dos dados. O software *Trimmomatic*, versão 0.36 (BOLGER; LOHSE; USADEL, 2014) foi utilizado com o objetivo de remover adaptadores, bases nas extremidades com pontuação média de qualidade abaixo de 25 (Phred +33) e fragmentos com comprimento menor que 35 pb.

3.2.2. Alinhamento de sequências e Quantificação de Genes

Os fragmentos *paired-end* que passaram pelo controle de qualidade foram mapeadas no genoma de referência *Homo Sapiens* versão GRCh38.p10 *build* 38 utilizando o programa *TopHat2* (KIM *et al.*, 2013), um mapeador genérico que implementa uma busca heurística para lidar com splice-junctions. Para este alinhamento, foram considerados os seguintes parâmetros: tamanho mínimo de

intron de 30 pb, número de *mismatches* por fragmento de 3pb, número de lacunas (*gaps*) por fragmento de 3pb, sensibilidade alta, tamanho máximo de inserção de 3pb, máximo de distância dos fragmentos de 200pb e somente fragmentos concordantes unicamente mapeados (que corresponde a cerca de 92% do total de fragmentos inicial) foram utilizados nas análises seguintes.

Em relação a quantificação dos genes presentes nos fragmentos de cada amostra, a ferramenta *featureCounts* do pacote *Rsubread* do *Bioconductor* foi utilizada (LIAO; SMYTH; SHI, 2014). O número total de *counts* foi obtido a partir do mapeamento dos fragmentos de RNA-Seq a regiões do genoma conhecidas por estarem relacionadas a cada gene específico.

3.2.3. Variabilidade nos dados

Como mencionado anteriormente, vários fatores possuem a capacidade de inserir viés nas sequências resultantes, ocasionando em uma perda na confiabilidade daqueles dados. Uma maneira de se estimar os principais efeitos responsáveis por essa variabilidade experimental é conhecida como Análise dos Componentes Variantes Principais (PVCA). Seu objetivo é reduzir a dimensão dos dados, porém mantendo sua variação original, e tentar ajustar um modelo linear misto com fatores de interesse agindo como efeitos aleatórios de modo a determinar qual seu peso na variabilidade total. Esta análise não-supervisionada foi realizada utilizando o pacote de mesmo nome proveniente do *Bioconductor* (BUSHEL, 2019).

3.2.4 Análise de Expressão Diferencial

A detecção de genes diferencialmente expressos (DEGs) ocorre quando é observada uma diferença entre o número de *counts* ou nível de expressão quando comparadas duas condições experimentais diferentes, como em amostras infectadas e amostras saudáveis. A ferramenta utilizada neste projeto para esta análise foi o pacote *edgeR* do *Bioconductor*. Esta metodologia faz uso de um modelo de Poisson para levar em consideração a variabilidade que as amostras podem possuir tanto no

aspecto biológico quanto no técnico, além de um método Empírico de Bayes para moderar o grau de dispersão entre os transcritos (ROBINSON; MCCARTHY; SMYTH, 2009). Sendo assim, a partir do número de counts dos genes, o método de normalização utilizado pelo pacote, *Trimmed Mean of M-Values* (TMM), foi utilizado com o objetivo de diminuir estes fatores. Deste modo, genes diferencialmente expressos puderam ser identificados para cada comparação entre amostra infectada e todas as saudáveis seguindo dois critérios principalmente: (i) p-valor ajustado inferior a 0.05 e (ii) valor absoluto de *fold-change* igual ou superior a 2.

3.2.4.1. Genes diferencialmente expressos entre amostras Crônicas e Não-Crônicas

A amostra deste trabalho é uma amostra clínica constituída de 39 adultos infectados com Chikungunya, provenientes dos estados de São Paulo e Sergipe, sendo que 6 ainda apresentavam manifestações de dor nas articulações após 3 meses da fase aguda da doença, assumindo um caráter crônico. Por outro lado, 7 pacientes não apresentaram sintomas após o mesmo período, sendo classificados como não-crônicos. Como mencionado anteriormente, não foi possível obter os dados para o restante dos pacientes.

Para que um gene fosse considerado diferencialmente expresso dentre as sub-classificações da doença, era necessários que apresentassem uma regulação consistente (seja *up* ou *down*) em pelo menos 70% das comparações entre cada amostras que compunham o subconjunto e todas as amostras saudáveis. Em outras palavras, era necessário que o gene apresentasse uma expressão diferencial consistente em pelo menos 3 e 4 comparações no grupo dos Crônicos e Não-Crônicos respectivamente.

3.2.5. Correlação Genes-Viremia

De modo a identificar possíveis associações entre expressão de genes e valor de CT invertido (*Cycle Threshold*, que corresponde a quantificação relativa de DNA viral na amostra), o teste de correlação de Pearson foi utilizado. Apenas as

associações com $|R|$ superior a 0.5 e p-valor ajustado inferior a 0.01 foram consideradas.

Em seguida, estas correlações foram utilizados na análise de enriquecimento funcional. Esta análise é necessária pois sabe-se que geralmente um conjunto de genes é responsável por definir fenótipos característicos da maior parte das doenças, ao invés de apenas um gene como era considerado antigamente. Este tipo de análise foi realizada através do método conhecido como *Gene Set Enrichment Analysis* (GSEA). Nesta análise os genes são ordenados de modo a priorizá-los e, neste caso, o método utilizado foram os valores de correlação calculados anteriormente. Além disso, é necessário um conjunto de módulos de genes co-expressos ou vias gênicas de modo a constatar se este enriquecimento funcional ocorre em um contexto específico. Para tanto, foram utilizados os *Blood Transcriptional Modules* (BTM), um conjunto de módulos relacionados ao estudo de vacina, obtido através de uma integração em larga escala de redes de transcrito de sangue humano publicamente disponíveis (LI *et al.*, 2014). GSEA então calcula a escore de enriquecimento normalizado (NES) para estes conjuntos baseado na distribuição na lista ranqueada dos genes que os compõem. Esta análise foi realizada com o pacote R *fgsea*.

3.2.6. Análise de Redes de Co-Expressão

Redes de genes co-expressos podem ser utilizadas como uma forma de se fazer novas associações entre genes e processos biológicos, priorizar candidatos a alvo de vacinas e identificação de genes regulatórios. Devido ao fato de a análise apresentar correlações entre os genes, ou seja, que há uma ativação simultânea entre eles, isto pode indicar que estes participam das mesmas vias gênicas e são propensos a ser reguladores das diferenças fenotípicas que a doença apresenta. O aspecto regulatório pode ser melhor investigado com a integração de outros dados como interações proteína-proteína, interações entre fatores de transcrição e enriquecimento de vias (GLASS *et al.*, 2013).

Neste projeto, esta análise foi realizada com o pacote do R *CEMiTool* (RUSSO *et al.*, 2018) nos dados normalizados TMM das amostras infectadas com Chikungunya e amostras saudáveis. Dentre as possíveis análises que podem ser realizadas com esta ferramenta, o pacote é capaz de identificar módulos de gene co-expressos de uma maneira intuitiva e automática, além de filtrar os genes que não apresentam uma expressão similar aos demais e realizar uma análise de enriquecimento dos módulos encontrados. Além disso, é possível avaliar genes que estão mais representados por uma via específica ou se mostrando alterados em apenas um subconjunto de amostras. Deste modo, é possível a identificação de genes que podem agir em conjunto e/ou regulados de maneira similar e identificar os principais reguladores, promovendo assim uma melhor compreensão da fisiopatologia da infecção por Chikungunya.

3.2.7. Meta-análise

Meta-análise é uma abordagem estatística que combina estudos individuais e independentes (GLASS *et al.*, 1976) sob a mesma hipótese e pode determinar caso um resultado ocorra no contexto geral.

Esta projeto considerou a comparação entre cada amostra humana infectada por Chikungunya e todas as amostras saudáveis como um estudo independente. Para se determinar os genes que possuem um comportamento consistente, a ferramenta *MetaVolcano* foi utilizada. Trata-se de uma abordagem em que é realizada uma meta-análise seguindo os critérios definidos pelo usuário, além de permitir uma melhor visualização dos resultados.

Genes foram considerados diferencialmente expressos seguindo os mesmos critérios mencionados anteriormente: (i) p-valor ajustado inferior a 0.05 e (ii) valor absoluto de *fold-change* igual ou superior a 2. No entanto, para serem considerados DEGs consistentes, era necessário que apresentassem uma regulação consistente (seja *up* ou *down*) em pelo menos 70% das amostras, assim como foi realizado para as amostras crônicas e não-crônicas.

3.2.8. Network Analysis

Para uma melhor compreensão dos padrões de determinada doença num aspecto sistêmico, uma das etapas envolvidas é a construção e análise de redes com os genes considerados interessantes. Uma das maneiras de realizar este tipo de análise é através de ferramentas como *NetworkAnalyst* (XIA; GILL; HANCOCK, 2015), um *framework online* que permite a visualização de redes a partir de um banco de dados abrangente de interação proteína-proteína (PPI) de alta qualidade baseado em *InnateDB* (BREUER *et al.*, 2013). Com o intuito de se melhorar a visualização, os softwares *Cytoscape* (SHANNON, 2013) e *Gephi* também foram utilizados.

3.2.8. Feature Selection e Machine Learning

Tecnologias de sequenciamento de nova geração permitem mensurar a expressão de milhares de genes para cada amostra utilizada no estudo, resultando em uma grande quantidade de dados mas também com ruído, como já foi mencionado anteriormente. Porém, tratando-se de doenças e processos biológicos no geral, apenas alguns genes são realmente diretamente envolvidos.

Atualmente há uma gama de métodos com o intuito de identificar o menor conjunto de variáveis necessárias para determinar a variável resposta, de modo a reduzir a alta dimensionalidade, redundância e ruído (KIRA *et al.*, 1992). Diversas teorias estatísticas e classificadores são utilizados para este propósito (SAEYS *et al.*, 2007).

Com o uso do pacote R *glmnet*, o método utilizado neste projeto foi a regressão de Lasso. Este é um método supervisionado de aprendizagem de máquina que impõem uma restrição na soma dos valores absolutos dos parâmetros do modelo, fazendo com que os coeficientes de regressão de algumas variáveis tendam a zero. A partir desta restrição, é possível identificar as variáveis com uma maior associação com a variável alvo. Além disso, em dados de expressão gênica é comum se deparar com variáveis altamente correlacionadas. A maneira com que a

regressão de Lasso lida com esta situação é que, ao notar que existem tais variáveis, Lasso tende a selecionar apenas uma delas de modo a diminuir a redundância (ZOU; HASTIE, 2005). Este método já foi utilizado em alguns trabalhos em bioinformática mostrando que é possível atingir um bom resultado classificatório como em (GHOSH; CHINNAIYAN, 2005).

Para determinar se as variáveis selecionadas pela regressão de Lasso eram suficientes para prever a febre de Chikungunya, foi realizada uma construção com diversos modelos como rede neural, *eXtreme Gradient Boosting* (XGBoost) , *Support Vector Machine* (SVM), *Random Forest* e *K-Nearest Neighbors* (KNN) através do pacote *caret* (KUHN, 2008).

O primeiro modelo treinado baseia-se na estrutura e comportamento do cérebro humano e pode ser treinado para reconhecer padrões complexos através da minimização de erros pela aprendizagem (BISHOP, 2013). Este modelo já foi utilizado em casos de classificação de diagnóstico de cânceres específicos baseado em assinaturas gênicas (KHAN et al., 2001).

Outro modelo treinado foi XGBoost, um algoritmo de aprendizado para *boosting* de árvore, no qual baseia-se numa sequência de modelos com baixo poder preditivo com o intuito de formar um classificador com alto poder preditivo. (CHEN; GUESTRIN, 2016). Já foi utilizado em casos de classificação de diagnóstico de cânceres específicos baseado em assinaturas gênicas (LI et al., 2017).

Já SVM trata-se de um algoritmo de aprendizagem supervisionado que procura definir o melhor hiperplano possível capaz de separar as diferentes classes, sendo que já foi utilizado em casos de classificação de amostras de câncer de cólon *versus* amostras saudáveis baseado em assinaturas gênicas (VANITHA; DEVARAJ; VENKATESULU, 2015).

Random Forest trata-se de um método de *machine learning* baseado em árvore altamente adaptativo e capaz de explorar correlações e interações entre os atributos. Já foi utilizado em diferentes contextos como interações proteína-proteína

(SIKIC; TOMIC; VLAHOVICEK, 2009) e predição de resposta de linhagens celulares cancerígenas a determinada droga (RIDDICK *et al.*, 2011).

Por último, KNN consiste de um método de classificação que estima a qual classe uma amostra pertence baseada em sua semelhança com as demais. Este método já foi utilizado em pacientes com neuroblastoma (PARRY *et al.*, 2010)

Com o intuito de estabelecer o mais apropriado para este contexto, os genes que mostraram um comportamento consistente na meta-análise foram divididos entre dados de treino (70%) e dados de teste (30%). Estes classificadores foram então treinados com os dados de treino de modo a calcular sua acurácia em predizer se a amostra estava infectada com febre de Chikungunya, utilizando somente as variáveis selecionadas pela regressão de Lasso. Como os métodos foram treinados utilizando *cross-validation*, era necessário comparar os resultados nos termos de reamostragem utilizados nos folds. Deste modo, a função *resamples* do pacote R *caret* foi utilizada de modo a estimar a área da curva ROC, sensibilidade (taxa de verdadeiros positivos) e especificidade (taxa de verdadeiros negativos) de cada método.

4. Resultados

4.1. INFORMAÇÃO CLÍNICA E DIAGNÓSTICO DE AMOSTRAS INFECTADAS POR CHIKV

Em uma área endêmica para Chikungunya, Dengue e Zika (CUNHA; TRINTA, 2017) no nordeste brasileiro, foi coletado sangue total de 39 adultos com sintomas característicos de infecção por arbovírus. Além disso, sangue total de 20 indivíduos saudáveis, provenientes de área endêmicas ou não, que não apresentaram nenhum dos sintomas relacionados à infecção viral, foi coletado para compor o grupo controle (Fig. 3A).

A maioria dos indivíduos infectados (n=22) reportou que o sintomas iniciaram no mesmo dia ou no dia anterior a coleta de sangue, indicando um estágio precoce da doença (Fig. 3B). Outros 8 indivíduos relataram que o sintomas iniciaram 2 dias antes da coleta, enquanto que em outros 7 iniciou entre 3 e 4 dias. Apenas duas amostras relataram dor por 20 dias ou mais (Fig. 3B).

Figura 3 - Perfil das amostras utilizadas neste estudo. (A) Número de indivíduos saudáveis e infectados utilizados nas análises. (B) Distribuição das amostras infectadas de acordo com os dias após o início dos sintomas que as amostras de sangue foram coletadas. O status da cronicidade da artralgia a longo prazo está identificada de acordo com o esquema de cores: crônica (laranja), não-crônica (verde) e indeterminada (cinza).

Fonte: Elaborada pelo autor

Foi realizado então uma análise ELISA no soro dos 39 indivíduos infectados e verificada a presença de anticorpos IgG específicos para Zika e Dengue (Fig. 4A). Apenas em um dos indivíduos infectados foram detectados anticorpos IgG e IgM específicos para Chikungunya. Um ensaio de RT-PCR foi feito para Chikungunya, Zika e Dengue em todas as amostras infectadas, confirmando a presença de RNA viral e a ausência de infecção aguda por Zika e Dengue em todos. Nenhuma relação foi encontrada entre a duração dos sintomas e a presença de anticorpos. Além disso, os sintomas mais frequentemente relatados foram febre, dor de cabeça, mialgia e artralgia (Fig. 4B). É possível ressaltar que há uma carga viral maior em indivíduos que reportaram estar em um estágio inicial da doença (Fig. 4C).

Figura 4 - Dados clínicos e diagnósticos das amostras. (A) Diagnóstico sorológico para CHIKV, DENV e ZIKV, sendo que a escala de cores é proporcional a imunidade apresentada. (B) Gráfico de rosca indicando os sintomas mais prevalentes relatados pelos pacientes infectados. (C) Viremia (CT invertido) de CHIKV nos subgrupos determinados pelo dia após o início do sintoma que as amostras foram coletadas

Fonte: Elaborada pelo autor

4.2 O IMPACTO DE INFECÇÃO POR CHIKV NO TRANSCRIPTOMA DO SANGUE

Para entender as mudanças moleculares que ocorrem durante a fase aguda da infecção por CHIKV, foi quantificado o transcriptoma de 39 indivíduos infectados e 20 indivíduos saudáveis através da abordagem de *RNA-seq*. Análises iniciais mostraram que a maior parte da variação entre os indivíduos era explicada por efeitos residuais indefinidos (Fig. 5A). Apesar de o status da infecção contribuir para a variância observada, mesmo que de maneira leve, indivíduos não se agruparam pela duração de seus sintomas nem pelo status da infecção (Fig. 5B).

Figura 5 - Análise Não Supervisionada realizada nas amostras infectadas com CHIKV. (A) Histograma que representa os efeitos ordenados de acordo com sua contribuição para a variância da amostra medida por PVCA. (B) Análise de componentes principais não supervisionados (PCA) dos 59 pacientes, incluindo amostras saudáveis e amostras infectadas pelo CHIKV de acordo com seus dias de sintomas, representados pela PCA bidimensional.

Fonte: Elaborada pelo autor

Baseado em uma análise de correlação entre a expressão gênica e a quantidade de RNA viral no sangue periférico dos indivíduos infectados, quase 3500 genes foram detectados possuindo uma associação com RNA viral, sendo que a maioria (>85%) apresentavam níveis de expressão com uma correlação negativa (Fig. 6A).

Figura 6 - Análise de Enriquecimento do Conjunto de Genes associada à viremia. (A) Representação de genes cuja expressão foi correlacionada com viremia (CT invertido) e número de genes. (B) GSEA utilizando os valores de correlação como ranks e BTMs como conjuntos de genes. As estrelas correspondem aos conjuntos de genes usados para construir as redes. (C) Redes Mínimas construídas com os conjuntos de genes negativamente correlacionados (nós azuis) utilizando a ferramenta NetworkAnalyst e visualização de gráficos com a ferramenta Cytoscape. Os nós cinzentos foram adicionados pelo NetworkAnalyst para conectar a rede. Nós cinzas que não estavam segurando a rede foram removidos manualmente.

Fonte: Elaborada pelo autor

Uma análise utilizando o método *Gene Set Enrichment Analysis* (GSEA) foi realizada com essa lista de genes revelando vias gênicas que poderiam estar relacionadas com o progresso da doença. Entre os módulos positivamente enriquecidos estão algumas vias relacionadas à células imunes inatas (células dendríticas e neutrófilos), resposta anti-viral, inflamação e receptores do tipo Toll (Fig. 6B e 12). Por outro lado, vias enriquecidas negativamente estão associadas a resposta imune adaptativa e são compostas por genes relacionados com proliferação celular, ativação e diferenciação de linfócitos B e T, além de quimiocinas e citocinas (Fig. 6B)

Os genes relacionados com as vias enriquecidas que possuíam um score NES negativo foram utilizados para a construção de uma rede mínima utilizando dados de interações proteína-proteína (PPI). Alguns genes interessantes como CCL5, CD8A, CD8B, CD3D, CD19, CCR7, CAMK4, CD79A, CD79B, CD27, TCF7 aparecem nesta rede resultante (Fig. 6C). É importante ressaltar que a maior parte destes genes está relacionado ao recrutamento de células B e linfócitos T efetores de memória (CRAWFORD *et al.*, 2011), assim como na indução e ativação de seu status de ser um linfócito de memória (DANILO *et al.*, 2018) .

Este mesmo processo foi realizado para as vias positivamente enriquecidas, como mostrado na figura 7. Genes como IRF7, OAS3, IFIT2, IFIT3, CXCL11, SERPING1 são alguns dos exemplos mostrados como relacionados com resposta imune inata a infecções virais (CHIRATHAWORN *et al.*, 2013), podendo ser considerados marcadores. Além disso, alguns são genes estimulados por interferon (ISGs) (SCHOGGINS *et al.*, 2011), como esperado.

Figura 7 - Assinatura gênica associada positivamente com viremia. Rede construída a partir dos genes que apresentaram uma correlação positiva (nós vermelhos), baseada na integração com dados de PPI, da mesma forma como observada na figura 4C. Nós cinzas foram adicionados automaticamente pela ferramenta NetworkAnalyst para manter a rede conectada. Posteriormente, os nós que não estavam influenciando a rede foram removidos manualmente.

Fonte: Elaborada pelo autor

A expressão de vários fatores eucarióticos iniciadores de tradução (eIFs) apresentaram correlação negativa com a quantidade de RNA viral (Fig. 8A), indicando um possível papel direto na replicação viral como observado em outros trabalhos (WANG *et al.*, 2014). Os eIFs são proteínas importantes que controlam o início do processo de tradução, passo fundamental na síntese de proteínas (WALSH, 2006). Por essa razão, a reconstrução de redes que envolvem estes genes tornou-se um passo essencial (Fig. 8B). É possível observar que apenas o gene EIF2AK2, também conhecido por PKR, possui uma correlação positiva entre sua

expressão e a viremia observada no sangue das amostras infectadas. Sabe-se que esta proteína pode ser ativada ao se ligar com RNA de cadeia dupla derivado de vírus como já foi observado em estudo anterior (GAINEY *et al.*, 2007). Essa quinase proteína específica para serina / treonina dependente de dsRNA induzida por IFN inibe a replicação viral ao fosforilar EIF2S1 (ZHANG *et al.*, 2018) e interrompe a síntese de proteína celular e viral.

Figura 8 - Associação dos genes EIF com viremia. (A) Representação da correlação dos genes EIF com viremia (Ct invertido). (B) Rede construída com genes EIF correlacionados com viremia e os genes associados a eles, integrados com PPIs. A ferramenta NetworkAnalyst.ca (InnateDB PPIs; First Order) foi usada para construir a rede e o programa Gephi para visualização. A cor dos nós representa se a correlação é positiva ou negativa com vermelho e azul, respectivamente.

Fonte: Elaborada pelo autor

O EIF4B pode ser induzido quando o mTORC1 é inibido pela Rapamicina e foi correlacionado com um aumento nas cargas virais do CHIKV em modelos

animais, fibroblastos humanos e fibroblastos primários de camundongos, bem como células HeLa (JOUBERT *et al.*, 2015). Em nossos resultados, é possível observar um número alto de genes eIF negativamente correlacionados com alta carga viral do CHIKV.

4.3. NÚMERO MAIOR DE GENES DIFERENCIALMENTE EXPRESSOS NOS PRIMEIROS DIAS DA INFECÇÃO

Os pacientes cujas amostras foram coletadas nos primeiros dias após o início dos sintomas apresentaram mais DEGs detectados pela análise de RNA-seq e uma média maior de carga viral em circulação (Fig. 9A).

Figura 9 - Perfil gênico das amostras infectadas por CHIKV. O perfil de cada paciente infectado foi comparado com amostras saudáveis para mostrar o número de genes regulados positivamente (vermelho) e negativamente (azul), conforme indicado na figura. O gráfico de barras à direita indica a viremia em cada paciente separadamente

Fonte: Elaborada pelo autor

Diante deste contexto, foi utilizada outra abordagem para comparar o perfil genético de cada paciente infectado com o grupo controle de modo a descobrir genes consistentemente diferencialmente expressos ao comparar múltiplos estudos, independente da viremia. Cerca de 70% dos pacientes infectados pelo CHIKV apresentaram regulação negativa (down) de 123 genes enquanto que 382 genes apresentavam regulação positiva (up) (Fig. 10A). APOBEC3A, IFI44, OAS3 são

alguns dos genes regulados positivamente mais representados e NT5E (também conhecido como CD73) PTGS2, SNORD3C, EEF1A1P13 estão entre os menos regulados entre os pacientes. A família APOBEC3A é formada por enzimas que são consideradas importantes componentes do sistema imune. Estas já foram descritas como tendo a capacidade de controlar a replicação do HIV (SHEEHY et al., 2002), porém não haviam sido estudado seu possível papel no contexto de infecção por CHIKV. Os genes APOBEC3 também são considerados ISGs (MUßIL et al., 2011) e, portanto, era esperado que a infecção pelo CHIKV fosse capaz de induzir este gene, como foi observado. Os genes SNORD3C, (pertencente a uma classe de RNAs não-codificantes que orientam o processo pós-transcricional de outros RNA não-codificantes) e EEF1A1P13 (um pseudogene), foram pouco caracterizados por outros estudos. Curiosamente, o NT5E / CD73 é uma ecto-5'- nucleotidase que já foi descrita como importante molécula para recuperar a barreira endotelial durante a dispersão de infecção de Dengue do tipo 2 quando ambos os IFN tipo I e o CD73 estão regulados positivamente (PATKAR; GIAYA; LIBRATY, 2013).

Figura 10 - Genes consistentemente diferencialmente expressos entre amostras de CHIKV. (A) A análise do Meta Volcano mostrou genes consistentemente diferencialmente expressos em pelo menos 70% de comparações entre amostras infectadas e amostras saudáveis. O eixo x representa o número de expressões diferenciais consistentes, enquanto o eixo y corresponde ao número de estudos em que os genes são classificados como DEG. (B) Representação gráfica da relação putativa entre alguns dos genes que foram classificados como consistentemente supra-regulados em (A).

Fonte: Elaborada pelo autor

Os resultados obtidos através da análise de RNA-Seq permitiram estabelecer uma possível importante assinatura gênica do hospedeiro. Na Fig. 10B podemos observar o aumento da expressão de vários genes em resposta à viremia. Os genes STAT1, STAT2, IRF9, IRF1 e IRF7 são capazes de induzir a expressão de genes relacionados a ISGs, sendo potencialmente importantes efetores antivirais. Além disso, a infecção pelo CHIKV levou a uma regulação positiva de vários genes pró-inflamatórios, como genes relacionados a IκBs, bem como a seus genes de degradação proteossomal, importantes para a ativação do NFκB e produção de citocinas pró-inflamatórias. Finalmente, também foi possível detectar genes relacionados à ativação do inflamassomo NLRP3 que poderiam levar à autofagia (SQSTM1), apoptose de células infectadas e também à produção de outras citocinas (CARD17) em resposta à infecção (Fig. 10B).

4.4. ASSINATURA GÊNICA EM RESPOSTA A INFECÇÃO POR CHIKV

Uma Análise de Enriquecimento de Conjuntos Genéticos (GSEA) (SUBRAMANIAN *et al.*, 2005) foi realizada em genes que mostraram uma diferença de expressão quando comparadas amostras infectadas e controle para identificar vias transcricionais consistentemente alteradas associadas à resposta à infecção por CHIKV. Similarmente aos resultados apresentados na Figura 6 em relação à correlação de viremia e assinaturas gênicas, podemos observar que as vias reguladas positivamente foram mais relacionadas à ativação da imunidade inata e sua resposta a infecções virais envolvendo células dendríticas e ativação de monócitos. Por outro lado, as vias reguladas negativamente envolviam a ativação e diferenciação da resposta imune adaptativa, como a diferenciação e ativação de células T e B (incluindo células plasmáticas) (Fig. 11).

Figura 11 - Análise de Enriquecimento Funcional focado nas vias do BTM. GSEA foi realizado para cada amostra infectada contra todos os controles para recuperar um resultado das vias do BTM mais consistentemente alteradas. Os BTMs foram usados como conjuntos de genes e seus respectivos resultados de log-fold-change como rank. Cada coluna representa os resultados da comparação de GSEA de cada amostra infectada. Os nomes das vias gênicas são mostrados no lado direito do heatmap. Os perfis das vias foram ordenados pela média do valor de NES em todos os pacientes.

Fonte: Elaborada pelo autor

4.5. POTENCIAIS ALVOS DE ARTRALGIA CRÔNICA E NÃO-CRÔNICAS

Para estudar as diferenças no perfil de expressão gênica entre amostras Crônicas e Não-Crônicas, várias abordagens foram utilizadas. Como mencionado anteriormente, apenas foi possível manter contato com alguns pacientes para continuar seu acompanhamento. Ao comparar cada subgrupo de infecção por CHIKV com amostras saudáveis, podemos observar que não há diferença estatisticamente significativa na viremia entre elas (Fig. 12A).

Figura 12 - Alterações no transcrito de amostras Crônicas e / ou Não-Crônicas em comparação com indivíduos saudáveis. (a) Comparação da viremia (CT invertido) entre amostras crônicas e não-crônicas. (b) Número de genes consistentemente diferencialmente expressos em cada subgrupo da doença. O eixo x representa o número de pacientes CHIKV usados em cada análise e o eixo y o número de DEGs. O ponto marcado no gráfico representa o número de DEGs em pelo menos 70% das amostras em cada grupo

Fonte: Elaborada pelo autor

Em seguida, procuramos pesquisar a sobreposição de DEGs entre os dois grupos de pacientes. Um total de 1.262 genes consistentemente diferencialmente expressos foi identificado no subconjunto crônico, enquanto o não crônico apresentou 1.862 DEGs (Fig. 13B) com alguma sobreposição entre os grupos.

Figura 13 - Diferenças e/ou semelhanças na expressão gênica entre de amostras crônicas e/ou não-crônicas. (A) Representação de genes diferencialmente expressos (DEGs) em pelo menos 70% dos pacientes crônicos e /ou não-crônicos em comparação com indivíduos saudáveis. (B) Genes que se sobrepunham em (A) mas apresentam um comportamento inverso dependendo do grupo. (C) É possível notar uma correlação positiva e forte entre os log2 fold-change dos dois grupos, apesar de a diferença de expressão apresentada pelo subgrupo crônico ser mais suave.

Fonte: Elaborada pelo autor

Além disso, alguns genes apresentaram comportamento invertido entre essas subclassificações da infecção pelo CHIKV. O HLA-DRB5 que poderia estar relacionado à apresentação de peptídeos antigênicos e NAMPTP1 (pseudogene) são alguns exemplos de genes que são supra-regulados na maioria das amostras não-crônicas e regulados negativamente na maioria das amostras crônicas (Fig. 13B). Por outro lado, DDX3Y, EIF1AY e LINC00278 são apenas alguns exemplos de genes que são supra-regulados na maioria das amostras crônicas e regulados negativamente na maioria das amostras Não-Crônicas. Curiosamente, o DDX3Y é uma RNA helicase que foi descrita como um importante efector da replicação e propagação do vírus do herpes (EL KASMI *et al.*, 2017), mas não havia sido descrita no contexto do CHIKV. O EIF1AY codifica a proteína EIF1A da família de fatores de iniciação da tradução eucariótica que mostramos na figura 7. EIF1A é um membro

importante dessa família em termos de regulador da tradução de mRNA viral em células infectadas revisadas em DEVER (2002).

Como havia muitos genes que apresentaram uma expressão diferencial consistente em ambas as subclassificações, era interessante analisar se havia diferenças entre eles. Foi possível observar que existe uma correlação alta e positiva entre a média da expressão gênica entre os grupos. No entanto, a mudança de *fold* no conjunto crônico parece ser mais moderada em comparação com a não-crônico, possível ser observado na relação entre a linha vermelha e a verde (Fig. 13C). Isso pode ser explicado pela menor viremia média nesse grupo (Fig. 12A)

4.6. DIFERENTES DURAÇÕES DE SINTOMAS REVELA PERFIS GÊNICOS DISTINTOS

Uma análise da rede de co-expressão gênica identificou 8 módulos que apresentaram perfis de expressão (Fig. 14A) e atividades de módulo (Fig. 14B) diferentes para cada classe de amostras.

O módulo M8, que possui atividades diferentes em cada grupo, é enriquecido para monócitos e neutrófilos (Fig. 15A), que são conhecidos por estarem relacionados com Febre de Chikungunya (LONG; HEISE, 2015). Entre as células imunes, os monócitos são a subpopulação mais infectada e os macrófagos ativados são as principais células infiltrantes nos tecidos infectados (LABADIE *et al.*, 2010), provavelmente devido à alta produção de MCP-1 durante a fase aguda da doença (RUIZ SILVA *et al.*, 2016).

Figura 14 - Módulos de genes de co-expressão obtidos através de CEMiTool. (A) Perfil de expressão dos 8 módulos resultantes do CEMiTool. As cores representam os dias após o início dos sintomas de cada amostra e a linha preta representa a expressão média de todos os genes dentro do módulo. (B) GSEA mostrando a atividade do módulo para cada subgrupo. O tamanho do círculo representa o valor absoluto da pontuação do NES, assim como o esquema de cores.

Fonte: Elaborada pelo autor

A ferramenta CEMiTool também foi capaz de integrar dados de interação proteína-proteína à análise de co-expressão. Entre os hubs do módulo M8, existem genes como o DDX58 (também conhecido como RIG-1) que já foi ligado à doença (LONG; HEISE, 2015) como um inibidor de replicação de vírus em um modelo animal (SCHILTE *et al.*, 2010) e bebês (Michlmayr *et al.*, 2018) (Fig. 15B).

Figura 15 - Análises adicionais resultantes de CEMiTool no módulo 8. (A) *Over Representation Analysis* do módulo M8 usando conjuntos de genes das vias BTM. As vias gênicas são ordenados por significância, conforme indicado no eixo x. (B) Rede de genes do módulo M8, sendo os dez genes mais conectados (*hubs*) representados como nós e suas interações como arestas. O tamanho do nó representa seu grau de conectividade.

Fonte: Elaborada pelo autor

4.7. PERFIL GÊNICO COMUM ENTRE FEBRE DE CHIKUNGUNYA, ARTRITE E DENGUE

Com o objetivo de comparar assinaturas de gênicas encontradas em nossos resultados de amostras infectadas pelo CHIKV com outros contextos semelhantes, era interessante identificar possíveis genes diferencialmente expressos entre eles. Como mencionado anteriormente, foi utilizado um estudo relacionado a artrite e outro de infecção por vírus relacionado ao CHIKV recuperado do banco de dados GEO. Um total de 949, 632 e 302 genes com regulação positiva foram identificados exclusivamente nos estudos de Artrite, Dengue e CHIKV, respectivamente. Apenas 7

genes regulados positivamente foram encontrados concomitante em todos os estudos: OAS1, C1QB, ANKRD22, IRF7, CXCL10, IFI6 e IFIT3 (Fig. 16A).

Figura 16 - Meta-análise de expressão gênica diferencial. (a) Genes supra-regulados em amostras CHIKV e/ou Dengue e/ou Artrite. O tamanho da barra representa a quantidade de genes que foram detectados como supra-regulados apenas nos grupos utilizados em cada análise (indicado como um círculo cinza no gráfico abaixo). À esquerda, o número total de genes regulados positivamente somente detectados em cada subgrupo é mostrado a no canto inferior esquerdo. Os círculos cinzentos estão conectados se os genes regulados positivamente coincidirem entre os subgrupos comparados. (B) A mesma abordagem que em (A), mas para os genes regulados negativamente.

Fonte: Elaborada pelo autor

Curiosamente, alguns desses genes são ISGs e também estão relacionados à imunidade inata à resposta antiviral, apesar do estudo de Artrite utilizado não envolver amostras infectadas por vírus. Mais de 300 genes foram exclusivamente supra-regulados apenas nas amostras infectadas por CHIKV, tais como o inflamassoma NLRP3, genes relacionados com o fator de transcrição relacionado a inflamação como NFkB e genes que pertencem à resposta Th2 (IL131RA, IL4I1) ou são inibidores da atividade da IL-1 (IL1RN) (FERRETTI *et al.*, 2017) (Fig. 16).

Não houve genes que foram consistentemente regulados negativamente entre os 3 grupos. No entanto, IL7R, ESYT1, PCYOX1 e LRRN3 foram regulados negativamente exclusivamente nas amostras de CHIKV e Artrite (Fig. 16B). A

citocina IL-7 é crucial para a sobrevivência de células T virgens e relacionado a memória, que são importantes efetores contra vários patógenos, incluindo infecções virais. O receptor de IL-7R é expresso na superfície destas células e foi mostrado, em um modelo animal, que previne a cronicidade induzida pelo vírus da Coriomeningite Linfocítica devido ao aumento de resposta de células T CD8 e a prevenção de sua exaustão (NANJAPPA; KIM; SURESH, 2011).

A partir dos genes que eram exclusivamente regulados negativamente em CHIKV, CX3CR1 parece ser interessante e de acordo com nossos resultados que mostraram que vários efetores de células T / diferenciação de genes foram regulados negativamente em infecção CHIKV. Além disso, a regulação negativa de BCL-2 poderia ser um mecanismo de resposta do hospedeiro para induzir a apoptose de células infectadas.

4.7. FEATURE SELECTION

Com o objetivo de determinar os genes que melhor distinguem amostras infectadas e controle, foram utilizados os 505 genes diferencialmente expressos obtidos da análise com a ferramenta *MetaVolcano* para uma análise de seleção de variáveis com o método de regressão de Lasso. A Tabela 1 mostra os genes cujo peso foi diferente de 0 na regressão de Lasso proposta, utilizando dados de treino, mostrando assim uma associação em algum grau com a distinção entre as amostras. As variáveis que possuem um coeficiente negativo estão associadas a amostras Controle, enquanto que aquelas que possuem um coeficiente positivo estão associadas a amostras infectadas pelo vírus Chikungunya.

Tabela 1 - Genes com maior associação com variável resposta pelo método de regressão de Lasso

Gene	Coeficiente Lasso
LINC02397	-0.02
PI3	-0.01
PRAL	0.03

Fonte: Elaborada pelo autor

Como pode-se observar, apenas 3 genes demonstraram uma associação com o status da infecção, de acordo com o método de Lasso. O gene LINC02397 é um RNA não-codificante longo que não é muito descrito na literatura, porém é apontado como diferencialmente expresso e significativamente correlacionado com um aumento na taxa de sobrevida de pacientes com melanoma (SUN et al., 2019). Por apresentar um coeficiente negativo na regressão de Lasso, pode significar que esta variável possui uma correlação negativa com a variável resposta. Logo, uma maior expressão deste gene pode ser um indicativo de que a amostra trata-se de uma amostra Controle. Esta mesma relação pode ser observado na análise realizada com a ferramenta *MetaVolcano*, ao se observar que este gene apresentava uma regulação negativa consistente na maior parte das amostras infectadas.

Já PI3 é um dos elementos essenciais para o funcionamento de PI3K/Akt, via envolvida em diversos processos celulares como proliferação celular, apoptose e imunidade antiviral. Além disso, este gene já foi descrito na literatura como sendo crítico no controle de infecção por vírus do Oeste do Nilo ao induzir resposta por IFN-I (WANG et al., 2017). Por outro lado, há vírus que já desenvolveram mecanismos para manipular esta via de forma a assegurar uma replicação viral com maior sucesso, apesar de estimular uma resposta do sistema imune simultaneamente. Este mesmo comportamento é descrito como controverso no contexto de infecção por vírus de Chikungunya. Há evidências que CHIKV afeta esta

via de maneira tênue (THAA et al, 2015) enquanto há trabalhos que indicam que há uma importante regulação do vírus através da modulação desta via (DAS et al., 2014). De acordo com o coeficiente da regressão de Lasso, este gene possui uma correlação negativa com a variável resposta. Logo, uma maior expressão deste gene pode ser um indicativo de que a amostra trata-se de uma amostra saudável. Este comportamento demonstra uma maior importância de PI3 para uma resposta antiviral bem-sucedida do em relação a uma possível replicação mais intensa.

O único gene que possui um coeficiente positivo, impactando positivamente na variável resposta é o RNA não-codificador PRAL. Há indícios na literatura de uma correlação positiva da expressão deste RNA com a expressão de P53, inclusive com um trabalho em que é mostrado que agiu como um supressor tumoral em pacientes com câncer de pulmão (SU et al., 2017).

De modo a determinar se estes genes seriam realmente capazes de prever febre de Chikungunya, houve a construção de diversos modelos. Para determinar o mais apropriado para este contexto, era necessário comparar os resultados de *resampling*. Na figura 16 é apresentado todos os modelos utilizados junto de suas métricas de ROC, especificidade e sensibilidade para os dados de treino.

Figura 17 - Comparação dos classificadores treinados

Fonte: Elaborada pelo autor

É possível observar que o modelo de rede neural apresenta uma acurácia de 1 para os dados de treino, sendo o melhor modelo. Em seguida, foi calculada a acurácia com os dados de teste. Conforme é possível observar na Figura 18, o modelo proposto errou em classificar apenas uma amostra, resultando em uma acurácia de 94 %.

Figura 18: Matriz de Confusão nos Dados de Teste utilizando o modelo de Rede Neural

Fonte: Elaborada pelo autor

5. Discussão

Apesar de haver alguns estudos utilizando-se de amostras infectadas de forma natural pelo vírus de Chikungunya, sua grande maioria é focado na análise a nível de proteína. Recentemente, um estudo utilizando a abordagem de biologia de sistemas elucidou alguns mecanismos essenciais utilizados pelo sistema imunológico diante uma infecção de CHIKV, porém com amostras de crianças (Michlmayr *et al.*, 2018). Como alguns estudos já apontaram, existem diferenças no processo da infecção entre amostras de pacientes adultas e neonatos, inclusive devido a este último grupo ser de risco. Uma outra diferença importante é de que a probabilidade do desenvolvimento da cronicidade da doença em pacientes mais novos é muito menor (cerca de 5-11%). Por esse motivo, este trabalho torna-se necessário pois faltam amostras clínicas de pacientes que desenvolveram de forma natural a doença para que os processos sejam elucidados neste grupo da mesma forma.

De forma similar a outros estudos presentes na literatura, foi possível identificar genes associados com mecanismos utilizados na resposta imune diante uma infecção viral e muitos deles são ISG. Além disso, genes que apresentaram uma maior regulação positiva estavam relacionados com neutrófilos, células dendríticas e monócitos. Estes últimos vêm sendo apontados como importantes mediadores inflamatórios e reguladores da resposta imune inata diante de infecção de diferentes arbovírus (KWISSA *et al.*, 2014).

Além disso, pôde-se observar que módulos de genes regulados negativamente durante a fase aguda da infecção por CHIKV envolvem assinaturas moleculares de células B e T, principalmente envolvendo ativação e diferenciação de linfócitos B e/ou T. É intrigante observar que alguns dos genes que poderiam agir de forma a melhorar a resposta imunológica adaptativa são regulados negativamente no início da doença. Enquanto que células T CD8⁺ estão mais relacionados à fase

aguda, células CD4⁺ foram associadas ao desenvolvimento da artralgia crônica, sendo recrutadas para tecidos infectados como mediadores inflamatórios (CHALAN *et al.*, 2013). Todavia, células B foram pouco estudadas no contexto de Febre de Chikungunya, porém estudos relataram o desenvolvimento de uma resposta robusta de anticorpos com altas taxas de anticorpos neutralizantes com atividade antiviral importante.

Alguns autores correlacionaram as cargas virais com o aumento de alguns mediadores inflamatórios que poderiam ser responsáveis pela cronicidade, mas isso não foi mostrado claramente (DUPUIS-MAGUIRAGA *et al.*, 2012). Além disso, foi possível observar um grupo de genes que apresentavam uma regulação invertida quando comparadas amostras crônicas e não-crônicas. Entre estes genes é importante ressaltar EIF1AY, codificador da proteína EIF1A que pertence a família dos fatores de iniciação da tradução eucariótica. Estes fatores são importantes moléculas para a produção de poliproteína viral e sua replicação bem-sucedida através do uso de mecanismos de síntese proteica do hospedeiro, incluindo alguns eIFs (WALSH; MATHEWS; MOHR, 2012). Interessante notar que o único eIF que foi correlacionado positivamente em relação a carga viral foi o EIF2AK2 (ou PKR), enquanto que os demais apresentaram uma correlação negativa. De acordo com nossos resultados, a PKR é considerada importante para a produção de IFN tipo I, também supra-regulada em nossos dados, em resposta a infecções virais (GARCÍA; MEURS; ESTEBAN, 2007), porém ainda não havia sido descrito em estudos envolvendo pacientes infectados pelo CHIKV. Além disso, a correlação negativa e forte entre a viremia e a maioria dos genes membros da família eIF observada nos dados, não havia sido descrita anteriormente em infecções por CHIKV. É possível que esta correlação negativa da maior parte dos membros da família eIF seja um importante mecanismo de defesa do hospedeiro contra a replicação do vírus.

Ao se comparar o perfil genético dos pacientes independente da viremia apresentada, foi possível observar que haviam genes 505 consistentemente diferencialmente expressos, dentre eles a família APOBEC3A. Já descrita como

sendo capaz de controlar a replicação de HIV (KRISKO *et al.*, 2013), esta família é considerada importante componente no sistema imune porém ainda não havia sido descrito no contexto de febre de Chikungunya. Além disso, foi possível detectar diferentes ISGs e genes relacionados à ativação do inflamassomo que poderiam levar a uma resposta imune mais intensa.

Em relação a comparação entre amostras naturalmente infectados crônicas e não-crônicas, este estudo pode ser considerado um dos pioneiros. E, apesar do número de amostras ser considerado baixo, foi possível encontrar resultados interessantes. Nos dados é possível observar que o desenvolvimento da cronicidade não está relacionado a cargas virais mais elevadas na circulação, porém é possível inferir que a viremia mais prolongada possa influenciar.

A partir de comparações entre assinaturas gênicas entre Febre de Chikungunya, Dengue e Artrite foi possível obter percepções sobre o processo de patogênese pelo CHIKV. Apesar de a infecção por CHIKV possuir a regulação positiva e negativa exclusiva de 300 e 107 genes, respectivamente, é possível observar que há a sobreposição em relação aos outros dois contextos. Dados anteriores mostraram uma forte sobreposição entre as assinaturas gênicas da artrite reumatóide e um modelo animal de infecção / artrite por CHIKV (Nakaya *et al.*, 2012), neste estudo foi possível observar este comportamento com mais detalhes.

Por último, através da análise de seleção de variáveis dos genes consistentemente diferencialmente expressos, foi possível determinar que apenas 3 genes demonstraram uma associação com o status da infecção: LINC02397, PI3 e PRAL. Dentre esses genes, apenas PI3 já havia sido descrito no contexto de Chikungunya por ser um importante elemento de uma via de sinalização envolvida em diversos processos celulares. Os restantes são RNA não-codificantes que são pouco descritos na literatura, porém há indícios de que influenciem no processo infeccioso. Além disso, foi possível realizar a construção de um modelo de rede neural capaz de prever com acurácia de 94% quais são as amostras infectadas e saudáveis somente a partir destes genes.

6. Conclusão

Neste trabalho foi utilizada uma abordagem de biologia de sistemas de forma a tentar elucidar alguns dos processos que podem estar relacionados com a infecção por CHIKV, além de tentar encontrar assinaturas gênicas que pudessem diferenciar amostras Crônicas e Não-Crônicas. Mostramos que alguns genes possuem a expressão alterada consistentemente em amostras de sangue de pacientes infectados, tanto ao compararmos todas as amostras como um único grupo quanto ao separarmos as amostras quanto a sua sub-classificação. Além disso, foi possível observar a importância de ISG na resposta imune diante a infecção além de que vias gênicas associadas com neutrófilos, células dendríticas e monócitos possam ser interessantes.

Em relação a diferença entre a cronicidade das amostras, foi possível observar que a carga viral elevada na corrente sanguínea não era suficientemente necessária para explicar esta característica. Porém, um achado interessante foi em relação a correlação predominantemente negativa de genes eIF em relação a carga viral, nunca antes abordada.

Além disso, foi possível observar a possibilidade de construção de um modelo capaz de prever com uma acurácia de 94% se a amostra estava infectada com o vírus Chikungunya ou não somente a partir dos dados de expressão de 3 genes.

Ainda se torna necessário que novos experimentos sejam realizados de forma a validar estes achados, porém eles se mostram interessantes e pioneiros. Por fim, acreditamos que este trabalho serviu para revelar novas perspectivas sobre os mecanismos regulatórios da infecção por Chikungunya.

Referências

ABDELNABI, R. et al. Antiviral treatment efficiently inhibits chikungunya virus infection in the joints of mice during the acute but not during the chronic phase of the infection. *Antiviral Research*, v. 149, p. 113-117, 2018.

AKAHATA, W. et al. A virus-like particle vaccine for epidemic Chikungunya virus protects nonhuman primates against infection. *Nature Medicine*, v. 16, n. 3, p. 334-338, 2010.

AKIRA, S.; TAKEDA, K. Toll-like receptor signalling. *Nature Reviews Immunology*, v. 4, n. 7, p. 499-511, 2004.

BERNARD, E. et al. Endocytosis of Chikungunya Virus into Mammalian Cells: Role of Clathrin and Early Endosomal Compartments. *PLoS ONE*, v. 5, n. 7, p. e11479, 2010.

BISHOP, C. Neural networks for pattern recognition. Tradução . Oxford: Oxford University Press, 2013.

BLOCH, D. The Cost And Burden Of Chikungunya In The Americas. Mestrado—[s.l.] Yale, 2016.

BOLGER, A.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30, n. 15, p. 2114-2120, 2014.

BORDI, L. et al. Chikungunya and Its Interaction With the Host Cell. *Current Tropical Medicine Reports*, v. 2, n. 1, p. 22-29, 2015.

BRÉHIN, A. et al. The large form of human 2',5'-Oligoadenylate Synthetase (OAS3) exerts antiviral effect against Chikungunya virus. *Virology*, v. 384, n. 1, p. 216-222, 2009.

BREIMAN, L. *Machine Learning*, v. 45, n. 1, p. 5-32, 2001.

BREUER, K. et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Research*, v. 41, n. D1, p. D1228-D1233, 2012.

BROUARD, C. et al. Estimated risk of Chikungunya viremic blood donation during an epidemic on Reunion Island in the Indian Ocean, 2005 to 2007. *Transfusion*, v. 48, n. 7, p. 1333-1341, 2008.

BUSHEL, P. *pvca: Principal Variance Component Analysis (PVCA)*, 2019.

BUTCHER, E.; BERG, E.; KUNKEL, E. Systems biology in drug discovery. *Nature Biotechnology*, v. 22, n. 10, p. 1253-1259, 2004.

CAPONE, M.; MATTHEW, J. Fc Receptor-Like Proteins in Pathophysiology of B-cell Disorder. *Journal of Clinical & Cellular Immunology*, v. 7, n. 3, 2016.

CHALAN, P. et al. Circulating CD4+CD161+ T Lymphocytes Are Increased in Seropositive Arthralgia Patients but Decreased in Patients with Newly Diagnosed Rheumatoid Arthritis. *PLoS ONE*, v. 8, n. 11, p. e79370, 2013.

CHEN, T.; GUESTRIN, C. XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 2016.

CHIRATHAWORN, C. et al. Cytokine levels in patients with chikungunya virus infection. *Asian Pacific Journal of Tropical Medicine*, v. 6, n. 8, p. 631-634, 2013.

CRAWFORD, A. et al. A Role for the Chemokine RANTES in Regulating CD8 T Cell Responses during Chronic Viral Infection. *PLoS Pathogens*, v. 7, n. 7, p. e1002098, 2011.

CUNHA, M. et al. Outbreak of chikungunya virus in a vulnerable population of Sergipe, Brazil—A molecular and serological survey. *Journal of Clinical Virology*, v. 97, p. 44-49, 2017.

CUNHA, R.; TRINTA, K. Chikungunya virus: clinical aspects and treatment - A Review. *Memórias do Instituto Oswaldo Cruz*, v. 112, n. 8, p. 523-531, 2017.

DANILO, M. et al. Suppression of Tcf1 by Inflammatory Cytokines Facilitates Effector CD8 T Cell Differentiation. *Cell Reports*, v. 22, n. 8, p. 2107-2117, 2018.

DAS, I. et al. Heat Shock Protein 90 Positively Regulates Chikungunya Virus Replication by Stabilizing Viral Non-Structural Protein nsP2 during Infection. *PLoS ONE*, v. 9, n. 6, p. e100531, 2014.

DEVER, T. Gene-Specific Regulation by General Translation Factors. *Cell*, v. 108, n. 4, p. 545-556, 2002.

DUPUIS-MAGUIRAGA, L. et al. Chikungunya Disease: Infection-Associated Markers from the Acute to the Chronic Phase of Arbovirus-Induced Arthralgia. *PLoS Neglected Tropical Diseases*, v. 6, n. 3, p. e1446, 2012.

EDELMAN, R. et al. Phase II safety and immunogenicity study of live chikungunya virus vaccine TSI-GSD-218. *The American Journal of Tropical Medicine and Hygiene*, v. 62, n. 6, p. 681-685, 2000.

EISEN, L. et al. Proactive Vector Control Strategies and Improved Monitoring and Evaluation Practices for Dengue Prevention. *Journal of Medical Entomology*, v. 46, n. 6, p. 1245-1255, 2009.

EL KASMI, I. et al. Extended Synaptotagmin 1 Interacts with Herpes Simplex Virus 1 Glycoprotein M and Negatively Modulates Virus-Induced Membrane Fusion. *Journal of Virology*, v. 92, n. 1, 2017.

ENSERINK, M. INFECTIOUS DISEASES: Massive Outbreak Draws Fresh Attention to Little-Known Virus. *Science*, v. 311, n. 5764, p. 1085a-1085a, 2006.

FERRETTI, M. et al. Neutralization of endogenous IL-1 receptor antagonist exacerbates and prolongs inflammation in rabbit immune colitis. *Journal of Clinical Investigation*, v. 94, n. 1, p. 449-453, 1994.

FROS, J. et al. Chikungunya Virus Nonstructural Protein 2 Inhibits Type I/II Interferon-Stimulated JAK-STAT Signaling. *Journal of Virology*, v. 84, n. 20, p. 10877-10887, 2010.

GAINEY, M. et al. Paramyxovirus-Induced Shutoff of Host and Viral Protein Synthesis: Role of the P and V Proteins in Limiting PKR Activation. *Journal of Virology*, v. 82, n. 2, p. 828-839, 2007.

GARCÍA, M.; MEURS, E.; ESTEBAN, M. The dsRNA protein kinase PKR: Virus and cell control. *Biochimie*, v. 89, n. 6-7, p. 799-811, 2007.

GARDNER, C. et al. Interferon-alpha/beta deficiency greatly exacerbates arthritogenic disease in mice infected with wild-type chikungunya virus but not with the cell culture-adapted live-attenuated 181/25 vaccine candidate. *Virology*, v. 425, n. 2, p. 103-112, 2012.

GARDNER, J. et al. Chikungunya Virus Arthritis in Adult Wild-Type Mice. *Journal of Virology*, v. 84, n. 16, p. 8021-8032, 2010.

GÉRARDIN, P. et al. Multidisciplinary Prospective Study of Mother-to-Child Chikungunya Virus Infections on the Island of La Réunion. *PLoS Medicine*, v. 5, n. 3, p. e60, 2008.

GERLACH, C. et al. The Chemokine Receptor CX3CR1 Defines Three Antigen-Experienced CD8 T Cell Subsets with Distinct Roles in Immune Surveillance and Homeostasis. *Immunity*, v. 45, n. 6, p. 1270-1284, 2016.

GHOSH, D.; CHINNAIYAN, A. Classification and Selection of Biomarkers in Genomic Data Using LASSO. *Journal of Biomedicine and Biotechnology*, v. 2005, n. 2, p. 147-154, 2005.

GIFFORD, G.; HELLER, E. Effect of Actinomycin D on Interferon Production by 'Active' and 'Inactive' Chikungunya Virus in Chick Cells. *Nature*, v. 200, n. 4901, p. 50-51, 1963.

GLASS, G. Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, v. 5, n. 10, p. 3-8, 1976.

GLASS, K. et al. Passing Messages between Biological Networks to Refine Predicted Interactions. PLoS ONE, v. 8, n. 5, p. e64832, 2013.

GOUPIL, B. et al. Novel Lesions of Bones and Joints Associated with Chikungunya Virus Infection in Two Mouse Models of Disease: New Insights into Disease Pathogenesis. PLOS ONE, v. 11, n. 5, p. e0155243, 2016.

HARRISON, V. et al. Production and Evaluation of a Formalin-Killed Chikungunya Vaccine, 1971.

HER, Z. et al. Active Infection of Human Blood Monocytes by Chikungunya Virus Triggers an Innate Immune Response. The Journal of Immunology, v. 184, n. 10, p. 5903-5913, 2010.

JAVELLE, E. et al. Inflammation of the external ear in acute chikungunya infection: Experience from the outbreak in Johor Bahru, Malaysia, 2008. Journal of Clinical Virology, v. 59, n. 4, p. 270-273, 2014.

JAVELLE, E. et al. Specific Management of Post-Chikungunya Rheumatic Disorders: A Retrospective Study of 159 Cases in Reunion Island from 2006-2012. PLOS Neglected Tropical Diseases, v. 9, n. 3, p. e0003603, 2015.

JORJANI, H. et al. An updated human snoRNAome. Nucleic Acids Research, v. 44, n. 11, p. 5068-5082, 2016.

JOUBERT, P. et al. Inhibition of mTORC1 Enhances the Translation of Chikungunya Proteins via the Activation of the MnK/eIF4E Pathway. PLOS Pathogens, v. 11, n. 8, p. e1005091, 2015.

KHAN, J. et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, v. 7, n. 6, p. 673-679, 2001.

KRISKO, J. et al. HIV Restriction by APOBEC3 in Humanized Mice. *PLoS Pathogens*, v. 9, n. 3, p. e1003242, 2013.

KIM, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, v. 14, n. 4, p. R36, 2013.

KIRA, K.; A. RENDELL, L. The feature selection problem: traditional methods and a new algorithm, Ninth National Conference on Artificial Intelligence. San Jose, California: 2019.

KRUTIKOV, M.; MANSON, J. Chikungunya Virus Infection: An Update on Joint Manifestations and Management. *Rambam Maimonides Medical Journal*, v. 7, n. 4, p. e0033, 2016.

KRUTIKOV, M.; MANSON, J. Chikungunya Virus Infection: An Update on Joint Manifestations and Management. *Rambam Maimonides Medical Journal*, v. 7, n. 4, p. e0033, 2016.

KUHN, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, v. 28, n. 5, 2008.

KWISSA, M. et al. Dengue Virus Infection Induces Expansion of a CD14⁺CD16⁺ Monocyte Population that Stimulates Plasmablast Differentiation. *Cell Host & Microbe*, v. 16, n. 1, p. 115-127, 2014.

LABADIE, K. et al. Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages. *Journal of Clinical Investigation*, v. 120, n. 3, p. 894-906, 2010.

LANGMEAD, B.; SALZBERG, S. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, v. 9, n. 4, p. 357-359, 2012.

LEE, R. et al. Mosquito Cellular Factors and Functions in Mediating the Infectious entry of Chikungunya Virus. *PLoS Neglected Tropical Diseases*, v. 7, n. 2, p. e2050, 2013.

LEWTHWAITE, P. et al. Chikungunya Virus and Central Nervous System Infections in Children, India. *Emerging Infectious Diseases*, v. 15, n. 2, p. 329-331, 2009.

LI, S. et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nature Immunology*, v. 15, n. 2, p. 195-204, 2013.

LI, Y. et al. A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data. *BMC Genomics*, v. 18, n. 1, 2017.

LIAO, Y.; SMYTH, G.; SHI, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. 2018.

LONG, K.; HEISE, M. Protective and Pathogenic Responses to Chikungunya Virus Infection. *Current Tropical Medicine Reports*, v. 2, n. 1, p. 13-21, 2015.

MICHLMAYR, D. et al. Comprehensive innate immune profiling of chikungunya virus infection in pediatric cases. *Molecular Systems Biology*, v. 14, n. 8, p. e7862, 2018.

MINORSKY, P et al. Achieving the in Silico Plant. Systems Biology and the Future of Plant Biological Research. PLANT PHYSIOLOGY, v. 132, n. 2, p. 404-409, 2003.

MUIR, P. et al. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biology, v. 17, n. 1, 2016.

MUßIL, B. et al. Increased APOBEC3G and APOBEC3F expression is associated with low viral load and prolonged survival in simian immunodeficiency virus infected rhesus monkeys. Retrovirology, v. 8, n. 1, p. 77, 2011.

NAKAYA, H. et al. Gene profiling of Chikungunya virus arthritis in a mouse model reveals significant overlap with rheumatoid arthritis. Arthritis & Rheumatism, v. 64, n. 11, p. 3553-3563, 2012.

NUNES, M. et al. Emergence and potential for spread of Chikungunya virus in Brazil. BMC Medicine, v. 13, n. 1, 2015.

NANJAPPA, S.; KIM, E.; SURESH, M. Immunotherapeutic effects of IL-7 during a chronic viral infection in mice. Blood, v. 117, n. 19, p. 5123-5132, 2011.

NG, L. et al. IL-1 β , IL-6, and RANTES as Biomarkers of Chikungunya Severity. PLoS ONE, v. 4, n. 1, p. e4261, 2009.

OGATA, H. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, v. 27, n. 1, p. 29-34, 1999.

OZDEN, S. et al. Human Muscle Satellite Cells as Targets of Chikungunya Virus Infection. PLoS ONE, v. 2, n. 6, p. e527, 2007.

PARRY, R. et al. k-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction. *The Pharmacogenomics Journal*, v. 10, n. 4, p. 292-309, 2010.

PATKAR, C.; GIAYA, K.; LIBRATY, D. Dengue Virus Type 2 Modulates Endothelial Barrier Function through CD73. *The American Journal of Tropical Medicine and Hygiene*, v. 88, n. 1, p. 89-94, 2013.

PIALOUX, G. et al. Chikungunya, an epidemic arbovirosis. *The Lancet Infectious Diseases*, v. 7, n. 5, p. 319-327, 2007.

PLANTE, K. et al. Novel Chikungunya Vaccine Candidate with an IRES-Based Attenuation and Host Range Alteration Mechanism. *PLoS Pathogens*, v. 7, n. 7, p. e1002142, 2011.

PORTA, J. et al. Structural Studies of Chikungunya Virus-Like Particles Complexed with Human Antibodies: Neutralization and Cell-to-Cell Transmission. *Journal of Virology*, v. 90, n. 3, p. 1169-1177, 2015.

POWERS, A.; LOGUE, C. Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *Journal of General Virology*, v. 88, n. 9, p. 2363-2377, 2007.

RAVI, V. Re-emergence of Chikungunya virus in India. *Indian Journal of Medical Microbiology*, v. 24, n. 2, p. 83, 2006.

RIDDICK, G. et al. Predicting in vitro drug sensitivity using Random Forests. *Bioinformatics*, v. 27, n. 2, p. 220-224, 2010.

ROBINSON, M.; MCCARTHY, D.; SMYTH, G. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, v. 26, n. 1, p. 139-140, 2009.

RUIZ SILVA, M. et al. Mechanism and role of MCP-1 upregulation upon chikungunya virus infection in human peripheral blood mononuclear cells. *Scientific Reports*, v. 6, n. 1, 2016.

RUSSO, P. et al. CEco-expression MiTool: a Bioconductor package for performing comprehensive modular analyses. *BMC Bioinformatics*, v. 19, n. 1, 2018.

RYMAN, K. et al. Non-pathogenic Sindbis virus causes hemorrhagic fever in the absence of alpha/beta and gamma interferons. *Virology*, v. 368, n. 2, p. 273-285, 2007.

SAEYS, Y.; INZA, I.; LARRANAGA, P. A review of feature selection techniques in bioinformatics. *Bioinformatics*, v. 23, n. 19, p. 2507-2517, 2007.

SANFILIPPO, C. et al. SERPING1 mRNA overexpression in monocytes from HIV+ patients. *Inflammation Research*, v. 66, n. 12, p. 1107-1116, 2017.

SCHILTE, C. et al. Chikungunya Virus-associated Long-term Arthralgia: A 36-month Prospective Longitudinal Study. *PLoS Neglected Tropical Diseases*, v. 7, n. 3, p. e2137, 2013.

SCHILTE, C. et al. Type I IFN controls chikungunya virus via its action on nonhematopoietic cells. *The Journal of Experimental Medicine*, v. 207, n. 2, p. 429-442, 2010.

SCHOGGINS, J. et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, v. 472, n. 7344, p. 481-485, 2011.

SCHOGGINS, J.; RICE, C. Interferon-stimulated genes and their antiviral effector functions. *Current Opinion in Virology*, v. 1, n. 6, p. 519-525, 2011.

SCHUFFENECKER, I. et al. Genome Microevolution of Chikungunya Viruses Causing the Indian Ocean Outbreak. *PLoS Medicine*, v. 3, n. 7, p. e263, 2006.

SEBASTIAN, M.; LODHA, R.; KABRA, S. Chikungunya infection in children. *The Indian Journal of Pediatrics*, v. 76, n. 2, p. 185-189, 2009.

SECRETARIA DE VIGILÂNCIA EM SAÚDE | MINISTÉRIO DA SAÚDE. Monitoramento dos casos de dengue, febre de chikungunya e doença aguda pelo vírus Zika até a Semana Epidemiológica 34 de 2018. [s.l: s.n.].

SHANNON, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, v. 13, n. 11, p. 2498-2504, 2003.

SHEEHY, A. et al. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, v. 418, n. 6898, p. 646-650, 2002.

ŠIKIĆ, M.; TOMIĆ, S.; VLAHOVIČEK, K. Prediction of Protein–Protein Interaction Sites in Sequences and 3D Structures by Random Forests. *PLoS Computational Biology*, v. 5, n. 1, p. e1000278, 2009.

SOOD, V. et al. ATF3 negatively regulates cellular antiviral signaling and autophagy in the absence of type I interferons. *Scientific Reports*, v. 7, n. 1, 2017.

SORLIE, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, v. 98, n. 19, p. 10869-10874, 2001.

SOURISSEAU, M. et al. Characterization of Reemerging Chikungunya Virus. *PLoS Pathogens*, v. 3, n. 6, p. e89, 2007.

STAIKOWSKY, F. et al. Prospective Study of Chikungunya Virus Acute Infection in the Island of La Réunion during the 2005–2006 Outbreak. *PLoS ONE*, v. 4, n. 10, p. e7603, 2009.

SU, P. et al. P53 Regulation-Association Long Non-Coding RNA (LncRNA PRAL) Inhibits Cell Proliferation by Regulation of P53 in Human Lung Cancer. *Medical Science Monitor*, v. 23, p. 1751-1758, 2017.

SUBRAMANIAN, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, v. 102, n. 43, p. 15545-15550, 2005.

SUN, L. et al. Identification of Long Non-coding and Messenger RNAs Differentially Expressed Between Primary and Metastatic Melanoma. *Frontiers in Genetics*, v. 10, 2019.

TASAKI, S. et al. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nature Communications*, v. 9, n. 1, 2018.

TAUB, M.; CORRADA BRAVO, H.; IRIZARRY, R. Overcoming bias and systematic errors in next generation sequencing data. *Genome Medicine*, v. 2, n. 12, p. 87, 2010.

TAUBITZ, W. et al. Chikungunya Fever in Travelers: Clinical Presentation and Course. *Clinical Infectious Diseases*, v. 45, n. 1, p. e1-e4, 2007.

TEO, T. et al. A Pathogenic Role for CD4⁺ T Cells during Chikungunya Virus Infection in Mice. *The Journal of Immunology*, v. 190, n. 1, p. 259-269, 2012.

THAA, B. et al. Differential Phosphatidylinositol-3-Kinase-Akt-mTOR Activation by Semliki Forest and Chikungunya Viruses Is Dependent on nsP3 and Connected to Replication Complex Internalization. *Journal of Virology*, v. 89, n. 22, p. 11420-11437, 2015.

TSETSARKIN, K. et al. A Single Mutation in Chikungunya Virus Affects Vector Specificity and Epidemic Potential. *PLoS Pathogens*, v. 3, n. 12, p. e201, 2007.

VAN DUIJL-RICHTER, M. et al. Early Events in Chikungunya Virus Infection—From Virus Cell Binding to Membrane Fusion. *Viruses*, v. 7, n. 7, p. 3647-3674, 2015.

VANITHA, C.; DEVARAJ, D.; VENKATESULU, M. Gene Expression Data Classification Using Support Vector Machine and Mutual Information-based Gene Selection. *Procedia Computer Science*, v. 47, p. 13-21, 2015.

WALSH, D. Assembly of an active translation initiation factor complex by a viral protein. *Genes & Development*, v. 20, n. 4, p. 461-472, 2006.

WALSH, D.; MATHEWS, M.; MOHR, I. Tinkering with Translation: Protein Synthesis in Virus-Infected Cells. Cold Spring Harbor Perspectives in Biology, v. 5, n. 1, p. a012351-a012351, 2012.

WANG, K. et al. Systems Biology and the Discovery of Diagnostic Biomarkers. Disease Markers, v. 28, n. 4, p. 199-207, 2010.

WANG, L. et al. An essential role of PI3K in the control of West Nile virus infection. Scientific Reports, v. 7, n. 1, 2017.

WANG, Z. et al. The Cap-Binding Translation Initiation Factor, eIF4E, Binds a Pseudoknot in a Viral Cap-Independent Translation Element. Structure, v. 19, n. 6, p. 868-880, 2011.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics, v. 10, n. 1, p. 57-63, 2009.

WAUQUIER, N. et al. The Acute Phase of Chikungunya Virus Infection in Humans Is Associated With Strong Innate Immunity and T CD8 Cell Activation. The Journal of Infectious Diseases, v. 204, n. 1, p. 115-123, 2010.

WEAVER, S. Arrival of Chikungunya Virus in the New World: Prospects for Spread and Impact on Public Health. PLoS Neglected Tropical Diseases, v. 8, n. 6, p. e2921, 2014.

WEAVER, S. et al. Chikungunya virus and prospects for a vaccine. Expert Review of Vaccines, v. 11, n. 9, p. 1087-1101, 2012.

WEAVER, S.; LECUIT, M. Chikungunya Virus and the Global Spread of a Mosquito-Borne Disease. *New England Journal of Medicine*, v. 372, n. 13, p. 1231-1239, 2015.

WILSON, J. et al. RNA-Seq analysis of chikungunya virus infection and identification of granzyme A as a major promoter of arthritic inflammation. *PLOS Pathogens*, v. 13, n. 2, p. e1006155, 2017.

WINTACHAI, P. et al. Identification of prohibitin as a Chikungunya virus receptor protein. *Journal of Medical Virology*, v. 84, n. 11, p. 1757-1770, 2012

WORLD HEALTH ORGANIZATION. Dengue vaccine: WHO position paper, July 2016 – recommendations. *Vaccine*, v. 35, n. 9, p. 1200-1201, 2017.

YOLKEN RH. Enzyme immunoassays for the detection of infectious antigens in body fluids: current limitations and future prospects. *Rev Infect Dis*, p. 35–68, 1982

XIA, J.; GILL, E.; HANCOCK, R. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature Protocols*, v. 10, n. 6, p. 823-844, 2015.

YANG, I. Use of External Controls in Microarray Experiments. *Methods in Enzymology*, p. 50-63, 2006.

ZHANG, X. et al. Endoplasmic reticulum stress induced by tunicamycin and thapsigargin protects against transient ischemic brain injury. *Autophagy*, v. 10, n. 10, p. 1801-1813, 2014.

ZHANG, RONG, KIM, ARTHUR S.FOX, JULIE M. et al. Mxra8 is a receptor for multiple arthritogenic alphaviruses. *Nature*, v. 557, n. 7706, p. 570-574, 2018.

ZHAO, S. et al. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLoS ONE, v. 9, n. 1, p. e78644, 2014.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), v. 67, n. 2, p. 301-320, 2005.

APÊNDICE A - ARTIGO PUBLICADO

Parte dos resultados foram recentemente aceitos para publicação na revista *PLOS Pathogens*. Na figura em anexo é possível observar o artigo em pré-impressão na plataforma *bioRxiv*.

Apêndice digital: *screenshot* da plataforma *bioRxiv* mostrando o artigo publicado em pré-impressão

The screenshot displays the bioRxiv preprint server interface. At the top, the bioRxiv logo is accompanied by the text 'THE PREPRINT SERVER FOR BIOLOGY'. Navigation links include HOME, ABOUT, SUBMIT, ALERTS / RSS, and CHANNELS. A search bar is present with a magnifying glass icon and the text 'Advanced Search'.

The main content area features the title 'Systems Analysis of Subjects Acutely Infected with Chikungunya Virus' and the authors: Alessandra Soares-Schanoski, Natalia Baptista-Cruz, Luiza A Castro-Jorge, Renan VH Carvalho, Clomar A Santos, Nancy Ros, Ursula Oliveira, Danuza D Costa, Cecilia LS Santos, Marielton P Cunha, Maria LS Oliveira, Juliana C Alves, Regina ALC Ocea, Danielle R Ribeiro, Andre NA Goncalves, Patricia Gonzalez, Andreas Suhrbier, Paolo MA Zanotto, Inacio J Azevedo, Dario S Zamboni, Roque P Almeida, Paulo L Ho, Jorge Kalil, Milton Y Nishiyama-Jr, and Helder I Nakaya. The DOI is provided as <https://doi.org/10.1101/531921>. A note states: 'This article is a preprint and has not been peer-reviewed [what does this mean?]'.

On the right side, there are options to 'Download PDF', 'Email', 'Share', and 'Citation Tools'. Social media sharing buttons for 'Tweet' and 'Curtir 2' are also visible. The 'Subject Area' is listed as 'Immunology'.

The 'Abstract' section is highlighted, showing the following text: 'The largest ever recorded epidemic of the chikungunya virus (CHIKV) began in 2004 and affected four continents. Acute symptomatic infections are typically associated with the onset of fever and often debilitating polyarthralgia/polyarthritis. In this study, a systems biology approach was used to analyze the blood transcriptomes of adults acutely infected with CHIKV. Gene signatures that were associated with viral RNA amounts and to the onset of symptoms were identified. Among those genes, the putative role of Eukaryotic Initiation Factor (eIF) family genes and apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC3A) in the CHIKV replication process were displayed. We further compared these signatures with those induced by dengue virus infection and rheumatoid arthritis. Finally, we demonstrated that CHIKV infection in mice induced IL-1 beta production in a mechanism highly dependent on the inflammasome NLRP3 activation. The findings provided valuable insights into the virus-host interactions during the acute phase and could be useful in the investigation of new and effective'.

On the far right, a 'Subject Areas' sidebar lists various categories: All Articles, Animal Behavior and Cognition, Biochemistry, Bioengineering, Bioinformatics, Biophysics, Cancer Biology, Cell Biology, Clinical Trials, Developmental Biology, Ecology, Epidemiology, and Evolutionary Biology.

APÊNDICE B - VERSÕES DOS PROGRAMAS E PACOTES UTILIZADOS

Programa / pacote	Versão
Trimmomatic	0.36
Bowtie	2.2.5
Genoma de referência Homo Sapiens	GRCh38.p10 build 38
Rsubread	1.34.2
pVCA	1.22.0
edgeR	3.24.3
fgsea	1.10.0
CEMiTool	1.8.0
MetaVolcano	0.99.5
glmnet	2.0.18
caret	6.0.84
randomForest	4.6.14
xgboost	0.82.1