Universidade De São Paulo Instituto de Matemática e Estatística Programa Interunidades de Pós-graduação em Bioinformática

MARIANA ARAÚJO PEREIRA

Análise transcritômica de macacos Rhesus durante infecção aguda por Zika Vírus

Dissertação de Mestrado Versão corrigida

São Paulo Data do Depósito Original na CPG: 12/06/2019

MARIANA ARAÚJO PEREIRA

Análise transcritômica de macacos Rhesus durante infecção aguda por Zika Vírus

Versão corrigida

Dissertação apresentada ao Programa Interunidades de Pós-graduação 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 I. Nakaya

São Paulo 2019 Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer

meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada

a fonte.

FICHA CATALOGRÁFICA



Elaborada pelo Serviço de Informação e Biblioteca Carlos Benjamin de Lyra do IME-USP, pela bibliotecária Maria Lucia Ribeiro CRB-8/2766

	AT	A DE DEF	ESA	63
personal de la construcción de la c			Alur	io: 95131 - 10554491 - 1 / Págin
Ata da datara da Dircar	tacta dala) Casharia) Mariana I	des Branders		
Bioinformática da Unive	risidade de São Paulo.	iujo Pereira r	io Programa: Bioin	formática, do(a) Interunidades er
Aos 12 dias do mês de apresentada para a obt	julho de 2019, no(a) IME realizou-s enção do título de Mestra intitulada	se a Defesa (II	la Dissertação do(a	s) Senhor(a) Mariana Araújo Pere
	"Análise transcritômica de mac	acos Rhesus	durante infecção p	oor Zika Virus"
Anés destanda abasta a		2	100	
para as devidas arguiçõ	es que se desenvolvem nos termos	sa a palavra regimentais	ao candidato para . Em seguida, a Co	exposição e a seguir aos examina missão Julgadora proclama o res
No	me dos Participantes da Banca	Função	Sigla da CPG	Resultado
He	Ider Takashi Imoto Nakaya	Presidente	FCF - USP	Não Votante
Se	rojo Verlovski de Almeida	Titular	ICB - USP	A PROVADA
Da	niela Santoro Rosa	Titular	UNIFESP - Extern	· APROVADA
Barris Anti-	WARA A			
Resultado Final: ATN	<u>NANDU </u>			
	Parecer da	Comissão	Julgadora •	
	-			
	10.9			20 E
Eu Marola Araula Farmi	- 110-		20 april 20	e di padre con presente a servici ca con
Eu, Marcia Araujo Ferrei Senhores(as). São Paulo	ra, aos 17 dias do mile de julho de 2	019.	_, lavrei a present	e ata, que assino juntamente con
Eu, Marcia Araujo Perrei Senhores(as). São Paulo	ra b, aos 1 dias do mar de julho de 2	019.	, lavrei a present	e ata, que assino juntamente con
Eu, Marcia Araujo Perrei Senhores(as). São Paulo	ra o, aos 12 dias do mão de julho de 2	019.	, lavrei a present	e ata, que assino juntamente con
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Adage Edison	ra o, aos 12 dias do miler de julho de 2 Luiz Durigon	019.	, lavrei a present	e ata, que assino juntamente con over Abiento
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Cargo Edison	ra b, aos 12 dias do mila de julho de 2 Luiz Durigon	019.	, lavrei a present	e ata, que assino juntamente con rollo de Almerda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison	ra o, aos 12 dias do miler de julho de 2 Luiz Durigon	019. Az	, lavrei a present	e ata, que assino juntamente con Jovski de Almerda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison	ra o, aos 12 dias do mile de julho de 2 Luiz Durigon	1019. Ma	, lavrei a present	e ata, que assino juntamente con Jovski de Almercia
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison	ra o, aos 1 dias do mile de julho de 2 Luiz Durigon a Santoro Rosa	one.	, lavrei a present	e ata, que assino juntamente con Jove Abrich Jovski de Almorda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison Daniel	ra o, aos 1 dias do mar de julho de 2 Luiz Durigon a Santoro Rosa Helder 1	1019.	, lavrei a present	e ata, que assino juntamente con Jovski de Almorda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison	ra o, aos 1 dias do mile de julho de 2 Luiz Durigon a Santoro Rosa Heider T Presidente	a kashi I mote da Comissão	, lavrei a present Sergio Ve Sargio Ve	e ata, que assino juntamente con Jovski de Almarda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison Daniel	ra o, aos 1 dias do mile de julho de 2 Luiz Durigon a Santoro Rosa Heider T Presidente ado por algum dos membros, o preenchimento	a shi Foto da Comissão de parecer é ob	, lavrei a present Sergio Ve Sergio Ve	e ata, que assino juntamente con Jovski de Almorda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison Daniel	ra o, aos 1 dias do miler de julho de 2 Luiz Durigon a Santoro Rosa Heider T Presidente ado por algum dos membros, o preenchimento pela Comissão de Pós-Graduação	a comissão de porecer é ob em 31.002	, lavrei a present Sergio Ve Sergio Ve	e ata, que assino juntamente con Jovski de Almorda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison Daniel * Oto: Se o candidato for reprov A defesa foi homologada ao título de Mestra em C	ra o, aos 1 dias do mar de julho de 2 Luiz Durigon a Santoro Rosa Heider T Presidente ado por algum dos membros, o prenchimento pela Comissão de Pós-Graduação iências obtido no Programa Bioinfo	a shi I noto da Comissão de parecer é ob em 31 0-2 rm ⁵³⁰ ca.	, lavrei a present Sergio Ve Sergio Ve Julgadora rigatório.	e ata, que assino juntamente con Jovski de Almerda
Eu, Marcia Araujo Perrei Senhores(as). São Pauk Edison Daniel * Obs: Se o candidato for reprov A defesa foi homologada ao título de Mestra em C	ra b, aos 1 dias do mile de julho de 2 Luiz Durigon Luiz Durigon a Santoro Rosa Helder 7 Presidente ado por algum dos membros, o prenchimento pela Comissão de Pós-Graduação d iências obtido no Programa Bioinfo	al shi I oto da Comissão de parecer é di em 31.03 rm31.03	, lavrei a present Sergio Ve Sergio Ve Nakaya o Nakaya o Jugadora mgatóno.	e ata, que assino juntamente con Jovski de Almorda
Eu, Marcia Araujo Perrei Senhores(as). São Pauk Edison Daniel	ra o, aos 1 dias do mile de julho de 2 Luiz Durigon a Santoro Rosa Heider T Presidente ado por algum dos membros, o prenchimento pela Comissão de Pós-Graduação iências obtido no Programa Bioinfo	a rishi I moto da Comissão de percer é de em 31 6 2 rimeter a.	, lavrei a present Sergio Ve Sergio Ve Nakaya o Julgadora mjatório. That 9 e, portan	e ata, que assino juntamente con Jovski de Almerda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison O Daniel	ra o, aos 1 dias do mile de julho de 2 Luiz Durigon a Santoro Rosa Heider T Presidente ado por algum dos mambros, o prenchimento pela Comissão de Pós-Graduação iências obtido no Programa Bioinfo Presidente da C	a teshi Imoto da Comissão o de parecer é et em <u>3103</u> rm ⁴ 20a.	, lavrei a present Sergio Ve Sergio Ve Julgadora rigatório. Tia L 9 e, portan	e ata, que assino juntamente con Jovski de Almerdia
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison Obs: Se o candidato for reprov A defesa foi homologada ao título de Mestra em C	ra o, aos 1 dias do mise de julho de 2 Luiz Durigon Luiz Durigon a Santoro Rosa Heider T Presidente ado por algum dos membros, o prenchimento pela Comissão de Pós-Graduação iências obtido no Programa Bioinfo Presidente da C	a shi I noto da Comissão de parecer é ob em 3403 rmstaca.	, lavrei a present Sergio Ve Sergio Ve Julgadora Ingatório. Tall e, portan Pós-Graduação etubal	e ata, que assino juntamente con Jovski de Almerda
Eu, Marcia Araujo Perrei Senhores(as). São Paulo Edison Obs: Se o candidato for reprov A defesa foi homologada ao título de Mestra em C	ra b, aos 1 dias do mile de julho de 2 Luiz Durigon Luiz Durigon a Santoro Rosa Heider T Presidente ado por algum dos membros, o preenchimento pela Comissão de Pós-Graduação lências obtido no Programa Bioinfo Presidente da C Prof. Dr. Ja Comissão	en Shi Froto da Comissão de parecer é di em Suca comissão de comissão de de Pós-Gradu	, lavrei a present Sergio Ve Sergio Ve Nakaya o Julgadora ngatóno. Tial Lº, e, portan Pós-Graduação etubal ação	e ata, que assino juntamente con Jovski de Almorda
Eu, Marcia Araujo Perrei Senhores(as). São Pauk Edison Obs: Se o candidato for reprov A defesa foi homologada ao título de Mestra em C	ra b, aos 1 dias do mile de julho de 2 Luiz Durigon Luiz Durigon a Santoro Rosa Heider T Presidente ado por algum dos membros, o prenchimento pela Comissão de Pós-Graduação iências obtido no Programa Bioinfo Presidente da C Presidente da C Interunidad	a ushi I moto da Comissão de percer é de em 31.6.2 rm ⁶²⁶ a. Comissão de este Carlos Se de Pós-Gradu es em Bioinfo	, lavrei a present Sergio Ve Sergio Ve Nakaya o Julgadora ngatório. Tial 9 e, portan Pós-Graduação etubal ação minática	e ata, que assino juntamente con Joviski de Almerda

Ao meu pai e gênio favorito, Ornando, e à minha mãe e favorita, Jacqueline. Ao meu bem, Hami.

Aos meus melhores amigos, principalmente ao meu muchacho Alan (*in memorian*), que sempre estiveram aqui, mesmo sem estar.

AGRADECIMENTOS

Seria injusto levar todo o crédito por este trabalho, porque sozinha eu não teria conseguido. Embora pareça clichê, muitas pessoas direta e indiretamente me ajudaram a concluir essa dissertação e a elas devo toda a minha gratidão.

Agradeço primeiramente aos meus pais, à minha família e aos meus amigos soteropolitanos por todo amor ao longo de todos esses anos. Por me apoiarem sempre, mesmo a 2000 km de distância. Não posso deixar de agradecer também por todo o carinho que me dão mesmo que eu demore de responder às mensagens, e por sempre me receberem com uma boa comida baiana a cada volta pra casa.

Ao meu orientador, Prof. Dr. Helder Nakaya, pela oportunidade, paciência e ensinamentos. Eu aprendi tanto no CSBL sob sua orientação que não sou capaz de mensurar ou agradecer à altura. Agradeço pela porta sempre aberta, confiança e a disponibilidade de conversar não só sobre o meu projeto, mas também sobre diversos outros assuntos desde divulgação científica até galos gigantes.

Aos meus amigos e colegas do laboratório: Patrícia, Natália, Fábio, Pedro, Thiago, Fernando, Mindy, Lucas, Jaqueline, Bruna, Diogo, Diógenes, Tiago, Alysson, Deney, Leandro e César. Agradeço em dobro aos pós-docs maravilhosos Vivi, Nicolau, Guilherme e Felipe pela paciência comigo. Ir ao 19, jogar *Just dance* e todos os HH tornaram tudo muito mais leve. Obrigada por me acolherem tão bem.

Ao professor Vinícius Maracajá-Coutinho pela colaboração neste projeto e por se mostrar sempre tão disponível e disposto a tirar as minhas dúvidas.

À FAPESP (processo nº 2018/10748-6, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)) e à Fundação Bill e Melinda Gates pelo financiamento, concessão da bolsa e incentivo à pesquisa. Assim como à Universidade de São Paulo, pelas experiências incríveis dentro e fora das aulas.

Por último, gostaria de agradecer imensamente (e nunca será o suficiente) ao meu maravilhoso Hami por estar comigo desde antes da minha jornada científica, aguentar os meus ataques de ansiedade quando os prazos se aproximam, me ouvir apresentando o projeto mil vezes mesmo sem entender e por não me deixar surtar.

Vocês são incríveis.

"A ciência é composta de erros que, por sua vez, são passos em direção à verdade"

- Jules Verne

PEREIRA, M. A. Análise transcritômica de macacos Rhesus durante infecção aguda por Zika Vírus. 2019. Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística, Universidade de São Paulo. São Paulo. 2019

RESUMO

Zika vírus (ZIKV) é um arbovírus que ganhou grande relevância nos últimos anos. Embora sua infecção geralmente induza sintomas leves, em alguns casos, a está associada à microcefalia em recém-nascidos e síndrome de Guillain-Barré em adultos. RNAs longos não codificadores (IncRNAs) desempenham um papel fundamental na modulação do sistema imunológico, mas pouco se sabe sobre sua função na infecção aguda por ZIKV. Portanto, o objetivo deste projeto foi estudar o transcritoma de macacos Rhesus durante a infecção e como os IncRNAs podem alterar a resposta em diferentes mamíferos infectados por ZIKV. Foram infectados quatro macacos Rhesus com a cepa HS-2015-BA-01 e o sangue foi coletado antes e depois de 1, 3, 5, 7, 10 e 14 dias de infecção. Em seguida, foi realizado o RNA-seq com amostras de sangue total para avaliar as alterações transcricionais durante a infecção. Nossas análises revelaram uma regulação da resposta imune, incluindo a identificação de IncRNAs que foram induzidos na infecção por ZIKV. Identificamos cerca de 9,210 lncRNAs, dos quais 3,246 ainda não foram anotados na base de dados ENSEMBL e 140 deles foram diferencialmente expressos em alguma comparação. Alguns genes associados à resposta imune, como STAT1, JAK1 e IFNGR2, parecem ser regulados negativamente. Análises de módulos de co-expressão e de redes revelaram as possíveis vias e genes afetados por alguns IncRNAs identificados. Os IncRNAs PROAP1, 2 e 3, ainda não depositados em bancos públicos, foram altamente correlacionados com BCL2, CYBA e MYCT1, que podem potencialmente regular a resposta pró-apoptótica. Além disso, outros conjuntos de dados (camundongos, neonatos e cultura celular) foram utilizados para avaliar a importância e a interferência dos IncRNAs na infecção. Com este estudo, encontramos IncRNAs previamente não descritos no genoma de *M. mulatta*, alguns dos quais estão correlacionados com genes importantes. Comparando todos os resultados, percebemos que alguns IncRNAs podem desempenhar papéis semelhantes, ainda que em diferentes animais. No geral, nossos resultados

sugerem que a infecção por ZIKV modifica a expressão de genes codificadores e IncRNAs, quanto à mecanismos apoptóticos de resposta contra o vírus.

Palavras-chave: ZIKV; Zika vírus; IncRNA; infecção experimental;

PEREIRA, M. A. Transcriptomic analysis of Rhesus monkeys during acute infection by Zika Virus. 2019. Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística, Universidade de São Paulo. São Paulo, 2019.

ABSTRACT

Zika Virus (ZIKV) is an arbovirus that has gained high relevance in recent years. Although ZIKV infection generally induces mild symptoms, in some cases, the infection is associated to microcephaly in newborns and Guillain-Barré syndrome in adults. Long non-coding RNAs (IncRNAs) play a key role in modulating the immune system but nothing is known about their function in acute ZIKV infection. Therefore, the aim of this project was to study the transcriptome of infected Rhesus monkeys throughout the infection and how IncRNAs can alter the response in different mammals infected by ZIKV. We had infected four Rhesus monkeys with HS-2015-BA-01 strain and collected their blood before and after 1, 3, 5, 7, 10 and 14 days of infection. We then performed RNA-seq with whole blood samples to assess the transcriptional changes during infection. Our systems biology analyses revealed a regulation of immune response during the infection including the identification of many IncRNAs that were induced upon ZIKV infection. We identified around 9.210 IncRNAs, 3.246 of which weren't annotated on the ENSEMBL database, and 140 of which were differentially expressed in some comparison. Some genes associated with immune response such as STAT1, JAK1 and IFNGR2, appear to be negatively regulated. Co-expression modules and network analyses have revealed the putative pathways and genes affected by these IncRNAs. Such as PROAP1, 2 and 3, not yet deposited in public banks, were highly correlated with BCL2, CYBA and MYCT1, that can potentially regulate pro-apoptotic response. In addition, other datasets (mice, neonates an cell culture) were used to further assess the importance and interference of IncRNAs in the infection. With this study we found many previously undescribed IncRNAs in the *M. mulatta* genome, some of which are correlated with important genes. Comparing all these results we found that some IncRNAs may play similar roles although in different animals. Overall, our results suggest that ZIKV infection modifies the expression of coding genes and IncRNAs that lead to apoptotic mechanisms against the virus.

Keyword: ZIKV; Zika vírus; IncRNA; experimental infection; mammals;

LISTA DE ILUSTRAÇÕES

Figura 1 - Estrutura e genoma do Zika vírus	18
Figura 2 - Ciclo de replicação do Zika vírus	19
Figura 3 - Países americanos com infecção autóctone por Zika vírus	21
Figura 4 - Risco de infecção em 2019 no Brasil por Zika, Dengue e Chikungunya de acordo com o índice de infestação do <i>Aedes Aegypti</i>	22
Figura 5 - Invasão viral por endocitose e alterações causadas pelo vírus	25
Figura 6 - Mecanismos de ação dos IncRNAs	27
Figura 7 - Linha do tempo de coleta cuja referência é a inoculação	31
Figura 8 - Esquema de obtenção do transcritoma dos macacos inoculados	32
Figura 9 - Pipeline geral para análise dos dados transcritômicos	34
Figura 10 - Pipeline para identificação dos IncRNAs	36
Figura 11 - Pipeline para análise de expressão diferencial	37
Figura 12 - Pipeline para análise de co-expressão e associação	40
Figura 13 - Caracterização das amostras de Macaca mulatta	42
Figura 14 - Número de transcritos identificados no genoma de Macaca mulatta	43
Figura 15 - Localização cromossômica dos transcritos de M. mulatta	44
Figura 16 - Distribuição de transcritos diferencialmente expressos	45
Figura 17 - Sinalização por Interferon e Mecanismo antiviral	46
Figura 18 - Vias de sinalização por Interferon e mecanismo antiviral	48

Figura 19 - Cinética da expressão de genes e IncRNAs do módulo 1	51
Figura 20 - Cinética da expressão de genes e IncRNAs do módulo 3	52
Figura 21 - Indução da apoptose a partir da regulação de IncRNAs	53
Figura 22 - Regulação da apoptose a partir de IncRNAs.	56

LISTA DE SIGLAS

cDNA	- DNA complementar
DAMP	- Padrões moleculares associados a dano celular
DEGs	- Genes Diferencialmente Expressos
DENV	- vírus da Dengue
DNA	- Ácido Desoxirribonucleico
FC	- Fold Change (razão entre dois valores)
GEO	- banco de dados de expressão Gene Expression Omnibus
GSEA	- Análise de enriquecimento de lista de genes
НСУ	- Vírus da Hepatite C
HEV	- Vírus da Hepatite E
HIV	- Vírus da Imunodeficiência Humana
IFN	- Interferon
IRF	- Fator regulatório de transcrição do interferon
IL-	- Interleucina
IncRNA	- RNA longo não codificante
LTc	- Linfócito T citotóxico
mRNA	- RNA mensageiro
NCR	- Região não codificadora
NF-Kb	- Fator nuclear Kappa B
NK	- Células Natural killer
NMD	- Decaimento de mRNA mediado por mutações nonsense
NS	- Proteína Não Estrutural
РАМР	- Padrões moleculares associados a patógenos

PFU -	- Unidades formadoras de placas
PROAP -	- IncRNA pro-apoptótico
PRR -	- Receptor de reconhecimento de padrões
RF -	- Random Forest
RNA	- Ácido ribonucléico
RNA-seq	- Sequenciamento de RNA de última geração
STAT -	- Transdutor de sinal e ativador de transcrição
SVM -	- Suport Vector Machine
TFN-α -	- Fator de necrose tumoral alfa
TLR -	- Receptores do tipo Toll
UCSC	- Universidade da Califórnia - Santa Cruz
WNV -	- Vírus do Oeste do Nilo
YFV	- Vírus da Febre Amarela
ZIKV	- Zika vírus

SUMÁRIO

INTRODUÇÃO	19	
ZIKA VÍRUS		
BIOLOGIA DO VÍRUS	19	
EPIDEMIOLOGIA	21	
RESPOSTA IMUNE	23	
ALTERAÇÕES TRANSCRITÔMICAS CAUSADAS POR ZIKV	26	
RNAS LONGOS NÃO CODIFICADORES	27	
TRANSCRITOMA DE ZIKV IN VITRO E IN VIVO	29	
OBJETIVOS	31	
OBJETIVO PRINCIPAL:	31	
OBJETIVOS ESPECÍFICOS:	31	
MATERIAL E MÉTODOS	32	
DADOS	32	
OUTROS MAMÍFEROS	34	
PROCESSAMENTO DAS AMOSTRAS DO RNA-SEQ	34	
CONTROLE DE QUALIDADE	35	
ALINHAMENTO DE SEQUÊNCIAS E QUANTIFICAÇÃO DE READS	36	
IDENTIFICAÇÃO DE NOVOS LNCRNAS DE RHESUS	36	
ANÁLISE DE EXPRESSÃO DIFERENCIAL	38	
ANÁLISES DE CO-EXPRESSÃO E ASSOCIAÇÃO	40	
PREDIÇÃO DA INTERAÇÃO LNCRNA-PROTEÍNA	41	
RESULTADOS E DISCUSSÃO	43	
CARACTERIZAÇÃO DAS AMOSTRAS	43	
IDENTIFICAÇÃO DE NOVOS LNCRNAS DE MACACOS	44	
ANÁLISE DE EXPRESSÃO DIFERENCIAL E ENRIQUECIMENTO	45	
ANÁLISE DE CO-EXPRESSÃO E ASSOCIAÇÃO	50	
PREDIÇÃO DA INTERAÇÃO LNCRNA-PROTEÍNA	54	
OUTROS MAMÍFEROS	55	
VISÃO GLOBAL DO PAPEL DE LNCRNAS NA INFECÇÃO	56	
CONCLUSÃO	58	
REFERÊNCIAS	59	
Apêndice I - Pacotes da linguagem R e ferramentas utilizados		
Apêndice II - Publicação	76	

INTRODUÇÃO

1. ZIKA VÍRUS

a. **BIOLOGIA DO VÍRUS**

Zika Vírus (ZIKV) é um arbovírus pertencentes à família *Flaviviridae* (GOULD e SALOMON, 2008) e transmitido principalmente por mosquitos do gênero *Aedes* (FAYE et al., 2008). O seu genoma consiste em uma molécula de RNA de fita única de sentido positivo, com duas regiões não codificantes flanqueantes (5' e 3' NCR). Possui aproximadamente 10,7 quilobases e codifica uma única poliproteína (Figura 1), que é clivada em capsídeo, proteínas da membrana, envelope e não estruturais (CHAMBERS et al., 1990; DIKHIT et al., 2016). Essa estrutura é similar à de outros flavivírus como o vírus da Dengue (DENV) e do oeste do Nilo (WNV), o que pode explicar o seu neurotropismo (KIELIAN, 2014; SIROHI et al., 2016).





A figura mostra os locais de clivagem da poliproteína para a obtenção das proteínas do capsídeo (C), de membrana (M), do envelope viral (E) e as não estruturais (NS). Abaixo da poliproteína pode-se observar a distribuição das proteínas E, M e C na estrutura do vírus. Adaptado de ViralZone (2019).

A proteína do capsídeo, com cerca de 100 aminoácidos, é responsável pelo empacotamento do genoma viral e formação do nucleocapsídeo. O envelope viral consiste em uma rede icosaédrica de associação entre as proteínas E (do envelope, E-ZIKV) e proteínas M (da membrana). As sete proteínas não estruturais (NS) são

essenciais para a replicação do genoma viral e evasão do sistema imune: NS1, NS2A e B, NS3, NS4A e B e NS5 (HASAN et al., 2018).

A entrada do vírus nas células do hospedeiro depende do conjunto de proteínas expressas no seu envelope e como elas interagem com a superfície celular (AGRELLI et al., 2019). O mecanismo mais provável de endocitose - processo pelo qual o material extracelular é transportado para o citoplasma - do ZIKV é através de clatrinas, onde após uma invaginação da membrana plasmática há modificações do envelope viral, fusão de membrana e liberação do genoma viral na região intracelular (HACKETT & CHERRY., 2018; PERSAUD et al., 2018). No citoplasma ocorre a tradução e clivagem da poliproteína, seguida de replicação, montagem, maturação e liberação da partícula viral (figura 2) (HASAN et al., 2018).



Figura 2 - Ciclo de replicação do Zika vírus

1) ZIKV se liga a receptores de superfície celular e é endocitado, mecanismo este mediado por clatrinas. 2) Devido ao baixo pH endossomal, ocorre a fusão da sua membrana com a membrana viral, resultando na liberação do RNA viral no citoplasma. 3) O genoma viral é traduzido em uma única poliproteína que é processada em proteínas estruturais e não estruturais. 4) O RNA é replicado pelas proteínas virais não-estruturais. 5) No retículo endoplasmático ocorre a montagem de partículas virais imaturas. 6) No complexo de Golgi as partículas imaturas sofrem clivagem da proteína precursora da proteína M, o que finaliza a maturação do vírus. 7) Os vírus maduros são englobados por exossomos. 8) Após exocitose o vírus está pronto para infectar outras células. Fonte: autoral, criado com Biorender e baseado em Perera et al., 2009.

b. EPIDEMIOLOGIA

ZIKV foi isolado em macacos Rhesus pela primeira vez em 1947 durante uma expedição na floresta de Zika - Uganda para estudar surtos de febre amarela (DICK, KITCHEN e HADDOW, 1952). Em 1952 houve o primeiro isolamento a partir de uma amostra humana e dois anos depois foi confirmada sua capacidade efetiva de infecção (MACNAMARA, 1954). Embora estes estudos tenham sido realizados na década de 50, relatos dessa infecção foram escassos até 2007, quando houve uma epidemia na Micronésia (LANCIOTTI et al., 2008) e nos anos seguintes foram relatados surtos em outras regiões do mundo, como Polinésia Francesa (CAO-LORMEAU et al., 2013), Ilhas Cook, Ilha de Páscoa (SANTÉ PUBLIQUE FRANCE, 2014) e Brasil (ZANLUCA et al., 2015).

Existem duas linhagens do ZIKV descritas: africana e asiática. A linhagem africana foi a primeira isolada (MR766) e é mais letal que a asiática (GARCEZ et al., 2016). Esta última, por sua vez, é a responsável pelos surtos dos últimos anos, incluindo na Polinésia Francesa e Brasil, e já está associada epidemiologicamente a doenças severas como a síndrome de Guillain-Barré e microcefalia (GONG et al., 2016).

Apesar da primeira notificação no nosso país ter ocorrido em 2015 (ZANLUCA et al., 2015), há indícios de que o vírus chegou ao território brasileiro dois anos antes, através de imigrantes haitianos e militares brasileiros recém-chegados de missões no Haiti (CAMPOS et al., 2018). Até o início de 2018 a infecção por ZIKV havia sido relatada em 84 países, sendo 48 territórios americanos, onde havia sido relatados mais de 223 mil casos cumulativos de infecção confirmados, sendo 137 mil deles no Brasil (PAHO/OMS, 2018). Na figura 3 pode ser observada a incidência de infecção autóctone pelo vírus em toda a extensão das Américas de acordo com os dados do relatório divulgado pela PAHO/OMS em 2018.



Figura 3 - Países americanos com infecção autóctone por Zika vírus

A densidade das cores está diretamente relacionada ao número de casos notificados. Pode-se observar que o Brasil é o país americano com maior número de casos relatados, seguido de países no norte da América Latina e parte da América Central. Figura autoral criada a partir dos dados disponibilizados pela OMS. "Zika Cumulative Cases". 2018.

Em 2019, segundo o último Informe de Arboviroses da Secretaria de Vigilância em Saúde, até a semana epidemiológica 24 (30/12/2018 a 15/06/2019) foram notificados 7.705 casos prováveis de infecção por Zika Vírus, com 1.788 confirmados até o momento (BRASIL, 2019). Além disso, de acordo com o Levantamento Rápido de Índices de Infestação pelo *Aedes aegypti* (LIRAa) de 2019, cerca de 994 municípios apresentaram alto índice de infestação do mosquito, com risco de surto para Zika, Dengue e Chikungunya em todo o país (figura 4) (BRASIL, 2019).

Figura 4 - Risco de infecção em 2019 no Brasil por Zika, Dengue e Chikungunya de acordo com o índice de infestação do *Aedes Aegypti*



A densidade das cores está diretamente relacionada ao risco calculado. Observa-se que a maior parte do país está em alerta de risco. Figura autoral criada a partir dos dados disponibilizados pela Secretaria de Vigilância em Saúde com os dados do LIRAa. 2019.

A disseminação do ZIKV depende da abundância e da diversidade de vetores e hospedeiros. O fato do principal transmissor ser o *Aedes aegypti* explica porque as regiões próximas aos trópicos (possuem condições ideais de vida para esse mosquito) são as mais afetadas. No Brasil os principais hospedeiros silvestres do vírus são os primatas não humanos (GUTIÉRREZ-BUGALLO et al., 2019).

c. RESPOSTA IMUNE

Embora a infecção seja assintomática na maioria dos casos, ZIKV pode causar sintomas brandos como febre alta e mialgia com duração média de 3 a 7 dias

(CHAN et al., 2016). Além disso, devido ao seu neurotropismo, ZIKV está também associado a casos da Síndrome Guillain-Barré em adultos e microcefalia em neonatos, podendo levar à morte dos indivíduos em ambas as condições (GARCIA et al., 2016).

Quando ocorre uma infecção, o sistema imunológico do hospedeiro desencadeia resposta contra o vírus, ao passo que este desenvolve mecanismos de escape das células de defesa. Para que a infecção seja bem sucedida, o patógeno precisa encontrar um meio de invadir a célula, se replicar utilizando a sua maquinaria e, ao mesmo tempo, evadir das respostas montadas contra ele. Em decorrência desses processos de invasão e replicação, mecanismos de resposta contra o patógeno ocorrem em diferentes níveis, incluindo o transcricional. Dentre os transcritos afetados existem os RNAs longos não-codificadores (IncRNAs), que podem sofrer alterações de expressão que podem favorecer tanto a infecção viral quanto o seu combate pelo sistema imunológico do hospedeiro (FORTES e MORRIS, 2016).

A resposta imunológica inata desempenha um papel importante na defesa contra o vírus. Uma defesa efetiva geralmente se inicia com o reconhecimento do vírus através de receptores. Estes receptores reconhecem padrões (PRRs, do inglês *Pattern recognition receptors*) de moléculas associadas a patógenos (PAMPs do inglês *Pathogen-associated molecular pattern*) e moléculas associadas a dano ou morte celular (DAMPs, do inglês *Damage-associated molecular pattern*) (KUMAR, KAWAI & AKIRA., 2011). De acordo com a localização, os PRRs podem ser classificados em transmembrana (como receptores de lectina tipo C e tipo *Toll*) ou citoplasmáticos (como receptores tipo NOD e tipo RIG-I) (KAWAI & AKIRA., 2009). Embora ainda em pouca quantidade, na literatura existem estudos que descreveram o envolvimento dos receptores tipo *Toll* e RIG-1 na infecção por ZIKV (ROLFE et al., 2016; KIM et al., 2018).

No estudo de Rolfe et al. (2016) com células progenitoras neurais infectadas por ZIKA (cepa asiática MR766) após 56h foi observado que os genes regulados positivamente durante a infecção eram enriquecidos para vias associadas a

23

inflamação e resposta imune mediada por diferentes receptores do tipo *Toll* (TLRs, do inglês *toll-like receptors*), *como* TLR3, TLR7/8 e TLR9. Além disso, os autores discutiram que houve a indução da expressão de genes associados à ativação do Fator Nuclear Kappa B (NF-kB), relacionando esses dois fatos (ROLFE et al., 2016). TLRs são capazes de reconhecer glicoproteínas da superfície do patógeno (TLR2 e TLR4), RNA viral de fita dupla (TLR3), fita simples (TLR7 e TLR8), dentre outras partículas imunogênicas (AKIRA, TAKEDA e KAISHO, 2001). O reconhecimento do RNA viral por TLRs resulta na produção de interferon do tipo I (IFN-I) e na ativação de NF-kB, o que leva à produção de citocinas pró-inflamatórias (KAWAI et al., 2010).

Receptores do tipo RIG-I são capazes de reconhecer o vírus no citoplasma e, ao se ligar ao RNA viral, iniciam eventos de sinalização que levam à ativação dos fatores regulatórios de transcrição do interferon IRF3 e IRF7 que, por sua vez, induzem o aumento da produção de IFN-I (TAKEUSHI & AKIRA., 2008). Em contrapartida, as proteínas virais são capazes de inibir esse tipo de estratégia. Um estudo conduzido por Kim et al (2018), após 24 horas de infecção por ZIKV em queratinócitos, demonstrou que a proteína NS1 interage com o receptor tipo RIG-I e suprime a produção de IFN-I a partir da inibição das vias de sinalização que seriam mediadas por ele. Os autores discutem que este pode ser um mecanismo de evasão da resposta imune em células epiteliais e que um agonista do receptor poderia ser utilizado como método de terapia antiviral, dado o seu potencial de suprimir a replicação do vírus (KIM et al., 2018).

As outras proteínas não estruturais também interferem nas cascatas de sinalização que levariam ao controle antiviral. As proteínas NS2A, NS4A e NS4B, por exemplo, inibem a ativação do transdutor de sinal e ativador de transcrição 1 (STAT1), bloqueando toda а cascata de sinalização por interferon (MUÑOZ-JORDAN et al., 2003). Essas duas últimas ainda estão envolvidas na desregulação de AKT, molécula central da via de PIK3, que regula a supressão da morte celular (LIANG et al., 2016). A proteínas NS5, por sua vez, recruta a Ubiguitina Ligase UBR4 para degradar STAT2, inibindo também a via de sinalização por interferon (MORRISON et al., 2013).

Figura 5 - Invasão viral e alterações causadas pelo vírus



Na imagem pode-se observar as alterações causadas pela entrada do vírus (seta vermelha) na célula e as modificações celulares em resposta a esse processo (setas contínuas escuras). As linhas tracejadas em vermelho representam alterações mediadas ou não com proteínas virais (NS) para driblar essa resposta do hospedeiro. Fonte: autoral, criado com Biorender.

d. ALTERAÇÕES TRANSCRITÔMICAS CAUSADAS POR ZIKV

Segundo Wang et al. (2009), transcritoma pode ser definido como o conjunto completo de transcritos de uma célula e a sua quantidade em um determinado estágio de desenvolvimento ou condição fisiológica. A análise transcritômica estuda como a expressão gênica muda em diferentes condições, o que é fundamental na compreensão das doenças humanas. Atualmente existem duas tecnologias predominantes no campo da transcritômica: microarranjos (ou *microarrays*), que quantificam um conjunto de sequências pré-determinadas (BROWN e BOTSTEIN, 1999), e sequenciamento de RNA (RNA-seq), que usa sequenciamento de alto rendimento para capturar todas as sequências (OZSOLAK & MILOS, 2011). Esta última começou a ganhar popularidade em 2008, quando a criação das novas tecnologias Solexa/Illumina (San Diego, CA) permitiu o sequenciamento de 10⁹

sequências, possibilitando o sequenciamento de transcritomas inteiros (MORTAZAV et al., 2008).

i. RNAS LONGOS NÃO CODIFICADORES

A partir do avanço das tecnologias de sequenciamento, foi possível estudar aqueles RNAs com níveis de expressão intrínseca mais baixos, como os chamados RNAs não codificantes (ncRNAs) (DJEBALI et al., 2012). Estes RNAs são divididos em duas grandes classes de acordo com o seu tamanho: ncRNAs curtos, quando menores que 200 nucleotídeos (nt); e ncRNAs longos (lncRNAs), quando possuem tamanho maior que 200 nucleotídeos. Hoje se sabe que esta última classe de RNA pode desempenhar papéis cruciais para o funcionamento celular como regulação de atividade enzimática, *imprinting* genômico e mediação de processos celulares como apoptose e regulação do ciclo celular, por exemplo (CIMMINO et al., 2015; WILUSZ et al., 2009; RINN & CHANG, 2012).

Quanto à localização no genoma, os IncRNAs podem ser classificados em: *intergênicos*, que estão localizados entre os genes codificadores de proteínas mas estão a pelo menos 1kb de distância de cada um; ou *intrônicos*, que como o próprio nome sugere se sobrepõem a regiões intrônicas dos genes codificadores independente da direção. Dentro desta classe os IncRNAs podem ser divididos ainda em *senso*, que se sobrepõem aos genes codificadores de proteínas e na mesma direção; *antisenso*, localizados sobrepostos aos mRNAs mas na cadeia oposta; ou *bidirecionais*, transcritos na cadeia oposta do mRNA com cerca de 1kb sobreposto (LOSKO et al., 2016).

Os mecanismos de ação dos IncRNAs são variados, podendo ocorrer no núcleo ou citoplasma das células, em *cis* ou *trans,* de forma que os IncRNAs de ação *cis* são aqueles que regulam a expressão de genes próximos aos seus locais de transcrição e os IncRNAs de ação *trans* são aqueles que podem reprimir ou ativar a expressão gênica independentemente do local do alvo (RINN & CHANG, 2012). No núcleo, a função mais descrita é a de regulação da expressão gênica por recrutamento de complexos de modificação de cromatina e alteração de histonas (WANG et al., 2011). Exemplos de IncRNAs com essa função são: XIST, importante

26

para a inativação do cromossomo X, que age em *cis* através do recrutamento do complexo repressivo *polycomb* 2 (figura 6A) (WUTZ et al., 2002; ZHAO et al., 2008) e Jpx, cuja função em *trans* é ativar a expressão de XIST através da remoção da proteína CTCF, seu repressor transcricional (figura 6B) (SUN et al., 2013). Já no citoplasma os lncRNAs podem modular a estabilidade do mRNA (figura 6C), como o BACE1-AS que aumenta a estabilidade do mRNA BACE1 (FAGHIHI et al., 2008); o controle de tradução (figura 6D), como o lincRNA-p21 que inibe a tradução dos mRNAs JUNB e CTNNB1 através do recrutamento do repressor da tradução RCK (YOON et al., 2012); ou agir como esponjas de microRNAs (figura 6E), como o lncRNA SNHG15 que "sequestra" o miR-211-3p, aumentando a proliferação celular em câncer (KONG & QIU, 2018).





Exemplos de RNAs longos não codificantes (IncRNAs) que regulam a transcrição gênica no núcleo (A e B) e no citoplasma (C, D e E). A) O IncRNA XIST induz a repressão da expressão gênica a partir do recrutamento do Complexo Repressivo *Polycomb* 2 (PRC2), que produz trimetilação da histona H3 lisina 27. B) O IncRNA JPX se liga ao repressor transcricional CTCF e inibe sua ligação ao promotor do IncRNA XIST, ativando assim a sua transcrição. C) O emparelhamento de bases entre regiões específicas do IncRNA anti-senso BACE1-AS com o mRNA BACE1 induz a sua estabilização e aumenta a expressão da proteína BACE1. D) Alguns IncRNAs podem desencadear um efeito

repressor, como o lincRNA-p21 que recruta o repressor traducional RCK que, por sua vez, inibe a tradução de proteínas como Junb e Ctnnb1. E) O efeito "esponja" é o mecanismo de ação de lncRNAs como o SNHG15, que compete com outros transcritos pela interação com microRNAs. Fonte: autoral, com base em Fatica & Bozzoni (2014).

ii. TRANSCRITOMA DE ZIKV IN VITRO E IN VIVO

Estudos *in vitro* utilizando células progenitoras neurais humanas infectadas por diferentes cepas de ZIKV, descreveram expressão diferencial de centenas de genes. Neste estudo foi observada uma redução na expressão daqueles relacionados ao ciclo celular, regulação e reparo do DNA, e aumento na expressão dos relacionados ao sistema imune, indicando aumento da inflamação e diminuição de células viáveis durante a infecção (SHAO et al., 2016; ZHANG et al., 2016; MCGRATH et al., 2017).

Utilizando neurônios periféricos humanos, foram observadas modificações na expressão de genes envolvidos na regulação apoptótica. Especificamente, houve o aumento de genes relacionados ao fator transcricional c-Jun e diminuição dos envolvidos na expressão das histonas, sendo que ambos os eventos podem ser relacionados com a morte celular das células infectadas (OH et al., 2017).

Poucos estudos transcritômicos de infecção experimental por ZIKV *in vivo* foram publicados até o presente momento (SHAO et al., 2016; WALDORF et al., 2016; AID et al., 2017; TRUS et al., 2018). O mais similar a este projeto utilizou a técnica de *microarray* em amostras de macacos Rhesus (*Macaca mulatta*) infectados por uma cepa brasileira durante 21 dias após o desafio (AID et al., 2017). O resultado foi semelhante ao já descrito *in vitro*, com expressão diferencial dos genes relacionados ao sistema imunológico. Foi observado aumento na expressão de genes associados aos mecanismos de monócitos, células *Natural Killer* e Linfócitos T auxiliares, assim como uma diminuição na expressão dos genes associados a sinalização de linfócitos T e B. Os autores discutiram que a replicação viral parece ser contida pela resposta imune inata mediada por interferon e que, somado à rápida produção de anticorpos neutralizantes a infecção é definitivamente contida.

Outro estudo, cujas amostras foram re-analisadas neste trabalho, utilizou amostras de cérebros de camundongos infectados, ainda em fase embrionária, por uma cepa mexicana (MEX1-44). Os autores constataram uma forte resposta imunológica com grande aumento da expressão da interleucina 1-beta (IL-1B) e fator de necrose tumoral alfa (TFN-α), além de um grande dano celular associado a essa desregulação inflamatória (SHAO et al., 2016). Foi observado o aumento acentuado de apoptose de células neuronais após a infecção que os autores discutiram como sendo um forte contribuinte ao desenvolvimento da microcefalia associada ao ZIKV. Nenhum destes dois estudos mencionou qualquer coisa sobre desregulação de IncRNAs ou sua associação com a infecção.

Estudos recentes demonstram que durante uma infecção viral pode haver modificações na expressão de IncRNAs e que estas, por sua vez, podem ser associadas ao curso da infecção, servindo como um marcador ou até mesmo envolvidas na regulação da resposta imunológica, positiva ou negativamente (GOMEZ et al., 2013; ZHANG et al., 2013; IMAMURA et al., 2014; CARNERO et al., 2016; QIAN et al., 2017). Alguns vírus, por exemplo, induzem a expressão do IncRNA NEAT1 (do inglês, nuclear-enriched abundant transcript), que regula positivamente a resposta imunológica antiviral. NEAT1 é capaz de inibir a produção de proteínas do vírus da imunodeficiência humana (HIV) através da repressão de elementos contendo mRNA viral (ZHANG et al., 2013), tal como aumentar a transcrição da interleucina 8 e outros genes do sistema imune após a infecção por influenza ou herpes vírus (IMAMURA et al., 2014). Carnero et al. (2016) estudaram IncRNAs estimulados pela infecção pelo vírus da hepatite C (HCV) e observaram que alguns destes (PVT1 e UCA1) estavam envolvidos na tumorigênese associada ao vírus, enquanto outros como o EGOT (do inglês eosinophil granule ontogeny transcript) promovem uma regulação negativa da atividade antiviral do hospedeiro.

Até o presente momento, nenhum estudo *in vivo* buscou identificar e descrever a associação de IncRNAs ao longo da infecção por ZIKV. A análise do transcritoma de *Macaca mulatta*, portanto, permitirá a compreensão de possíveis mecanismos de regulação gênica durante a infecção viral.

29

OBJETIVOS

OBJETIVO PRINCIPAL:

Identificar possíveis mecanismos de regulação gênica na infecção in vivo por ZIKV.

OBJETIVOS ESPECÍFICOS:

- Identificar novos IncRNAs no genoma de Macaca mulatta expressos durante a infecção;
- Analisar o perfil de expressão e identificar as possíveis vias associadas ao desenvolvimento da infecção por ZIKV em macacos;
- Identificar IncRNAs correlacionados com genes importantes para a resposta imunológica;
- Identificar IncRNAs diferencialmente expressos na infecção por ZIKV em outros mamíferos;

MATERIAL E MÉTODOS

1. DADOS

A parte experimental desse estudo foi realizada em colaboração com o grupo do Professor Dr. François Villinger da Universidade de Louisiana em Lafayette (ULL) - Estados Unidos da América. As amostras utilizadas foram provenientes de quatro macacos Rhesus (*Macaca mulatta*), com idades entre 9 e 17 anos, mantidos na ULL e cuidados de acordo com os protocolos do Comitê de Cuidados e Uso de Animais de Laboratório do país. Os animais foram infectados utilizando o volume de 1mL de uma solução em tampão salina-fosfato com 5 x 10⁷ unidades formadoras de placas (PFU, do inglês *plaque forming unit*) da cepa brasileira de Zika vírus HS-2015-BA-01 (genebank: KX520666) por via intravenosa. Os vírus utilizados foram obtidos a partir da primeira passagem de cultura da amostra clínica. As amostras de sangue total para o sequenciamento de RNA de alto rendimento (RNA-seq) foram coletadas em sete pontos distintos, sendo um antes da infecção (2 semanas antes) e os outros seis durante o seu curso (1, 3, 5, 7, 10 e 14 dias) até o total de 14 dias, como mostrado no esquema abaixo:





A primeira amostra foi coletada 14 dias antes do dia de inoculação. Após a inoculação de 5 x 10⁷ PFU foram coletados outros seis pontos até o 14º dia após a infecção.

O período estudado de infecção (14 dias) é correspondente à cinética de viremia dos flavivirus onde, após o décimo dia geralmente não se consegue mais detectar a presença do vírus no organismo (RATTERREE et al., 2004; DUYEN et al., 2011). Sangue total dos animais foi coletado nos 7 pontos (um tubo do 5º dia foi

acidentalmente quebrado antes da extração) utilizando tubos de coleta *Paxgene* (Qiagen), que garante a estabilização e conservação do RNA intracelular (RAINEN, 2002), seguida da extração do RNA total utilizando o *kit PAXgene Blood RNA* (Qiagen). A biblioteca de cDNA foi preparada utilizando *KAPA Stranded RNA-Seq Kit* (Illumina[®]) e as amostras foram sequenciadas utilizando o sequenciador *HiSeq 3000* (Illumina[®]) com a tecnologia *single-end*, seguida de avaliação do controle de qualidade utilizando o *software Sequencing Analysis Viewer* (*Illumina[®]*). Por fim, foi utilizado o *software Bcl2fastq2* (versão 2.17) para converter os arquivos gerados pelo sequenciador (formato .blc) em arquivos capazes de ser analisados pela linguagem R (formato .fastq). O esquema de obtenção das amostras pode ser conferido na figura 8. Para estes mesmos pontos também foram coletadas amostras para o acompanhamento da viremia e alterações celulares, como descrito por Silveira e colaboradores (2017).



Figura 8 - Esquema de obtenção do transcritoma dos macacos inoculados

Após a primeira coleta de sangue (2 semanas antes) e inoculação da cepa HS-2015-BA-01 nos quatro macacos, sangue foi coletado em 6 pontos após a infecção e todas as amostras seguiram para as etapas de extração de RNA. Em seguida houve a criação da biblioteca de cDNA seguida do sequenciamento das amostras. Após processamento do sequenciamento foram gerados arrquivos no formato fastq para a análise computacional.

a. OUTROS MAMÍFEROS

Embora o principal foco das análises tenha sido o conjunto de dados obtidos de macacos Rhesus, foram utilizados outros *datasets* (camundongo e humano) a fim de identificar as alterações causadas pela infecção também nestes mamíferos. O *dataset* de *Homo sapiens* (próprio do nosso laboratório), assim como o de camundongo, foi explorado até o momento apenas no que diz respeito aos mRNAs.

O dataset de humanos é constituído de 4 amostras de RNA-seq provenientes de cérebros de bebês brasileiros que vieram a óbito logo após o parto, com idade gestacional entre 29 e 41 semanas. O exame de detecção do RNA viral foi realizado durante e após a gestação, de forma que três bebês apresentaram resultado positivo para infecção por Zika e um apresentou resultado negativo, que foi utilizado como controle. Como as mães dos bebês foram naturalmente infectadas durante a gestação e permaneceram em território brasileiro, possivelmente a cepa infectante é da linhagem asiática.

Os dados provenientes de camundongo (GSE89069), de acordo com informações do GEO, foram obtidos de quatro animais com 17.5 dias de vida, sendo dois do grupo controle e dois do grupo infectado por 10⁶ PFU da cepa mexicana (de linhagem asiática) MEX1-44 (genebank: KX856011). As amostras foram coletadas três dias após a infecção experimental.

2. PROCESSAMENTO DAS AMOSTRAS DO RNA-SEQ

As análises de RNA-seq foram realizadas de acordo com o *pipeline* descrito de forma geral na figura 9, onde pode-se observar que a partir do processamento dos dados e identificação dos IncRNAs o objetivo final é a identificar o enriquecimento funcional dos genes ao longo do tempo e a associação entre IncRNAs e mRNAs. Nas figuras subsequentes (figuras 10, 11 e 12) estão detalhados os passos para a obtenção desse resultado. Todos os pacotes utilizados e suas respectivas versões estão listados no Apêndice I deste documento.



Figura 9 - Pipeline geral para análise dos dados transcritômicos

Após a obtenção dos arquivos provenientes do sequenciador foi realizada uma análise de qualidade de sequenciamento seguida de mapeamento das *reads* e contagem de transcritos. A partir dessa quantificação foram realizadas três análises independentes: Identificação de IncRNAs no genoma, análise de expressão diferencial para cada momento de infecção e análise de associação.

a. CONTROLE DE QUALIDADE

Todas as amostras passaram por uma avaliação da qualidade das sequências utilizando FastQC. Este programa verifica diferentes parâmetros de qualidade como tamanho da sequência e conteúdo de guanina-citosina (GC), por exemplo, e fornece um relatório onde estão contidas informações de problemas que podem ter sido originados no sequenciamento ou no material biológico da biblioteca (ANDREWS, 2010). A partir do relatório pode-se identificar se há contaminação e a necessidade de cortar uma parte da sequência antes de prosseguir com as análises. Um exemplo disso é a clivagem da região dos adaptadores para dados provenientes do Illumina (BOLGER et al., 2014). Seguindo as métricas utilizadas no FastQC, cada sequência deve possuir uma pontuação Phred acima de 28 (ANDREWS, 2010) para ter uma boa qualidade de sequenciamento.

Neste trabalho todas as amostras apresentaram alta qualidade, com valor médio acima de 32. Além disso, o comprimento das sequências foi de 25-50, a porcentagem de guanina-citosina variou de 48 a 52 % e o número total de sequências apresentou uma mediana de 26 milhões.

b. ALINHAMENTO DE SEQUÊNCIAS E QUANTIFICAÇÃO DE READS

As amostras obtidas de macacos Rhesus foram alinhadas ao genoma de referência MMul_8.0.1 (UCSC) utilizando a ferramenta *TopHat2*, um mapeador que alinha *reads* do RNA-Seq a grandes genomas de mamíferos usando o alinhador de ultra-alto rendimento *Bowtie* (KIM et al., 2013). Houve uma alta taxa (> 91%) de alinhamento com o genoma de referência. Em seguida foi utilizada a ferramenta *Cufflinks*, do pacote homônimo, que utiliza as coordenadas genômicas para a reconstrução dos transcritos completos a partir dos fragmentos sequenciados, permitindo a identificação de novos genes. Os transcritos de cada amostra então foram quantificados e unificados pela função *Cuffmerge*, gerando uma lista única e completa de *counts* (TRAPNELL et al., 2012).

3. IDENTIFICAÇÃO DE NOVOS LNCRNAS DE RHESUS

Tendo em vista que os IncRNAs anotados no genoma de *M. mulatta* foram, em sua grande maioria, preditos apenas por ferramentas computacionais, foi necessário fazer a reanotação do genoma para garantir que os RNAs pertencentes a este grupo fossem corretamente analisados. Para a reanotação e identificação de novos IncRNAs foi desenvolvido um *pipeline* em colaboração com o grupo do professor Dr. Vinícius Maracajá-Coutinho (Universidad de Chile), tendo como base as características dessa classe de RNAs, descrito na figura abaixo (Figura 10).

Desta forma, os transcritos com regiões codificantes e/ou que possuíam tamanho menor que 200 nucleotídeos foram removidos. Aqueles transcritos que além de \geq 200 nt possuíam número de éxons \geq 2 passaram por uma ferramenta de avaliação de potencial de codificação CPAT (do inglês *Coding-Potential Assessment Tool* - <u>http://ma-cpat.sourceforge.net/</u>), que através de um modelo de regressão logística, calcula um *score* de 0 a 1, distinguindo uma sequência de RNA codificante (>0,36) e não-codificante, com maior sensibilidade e especificidade frente às outras calculadoras já conhecidas (WANG et al., 2013). Selecionamos para as próximas etapas apenas aqueles transcritos cujo valor foi igual a 0, que em seguida foram

comparados com os dados de IncRNAs existentes para identificar os já anotados, assim como os novos.



Figura 10 - Pipeline para identificação dos IncRNAs

Para a identificação de IncRNAs no genoma foi feita a remoção de genes codificantes, transcritos pequenos e potencialmente codificantes. Em seguida foi realizada a comparação com sequências já conhecidas e, por fim, identificados aqueles IncRNAs novos. Estes IncRNAs identificados foram então reclassificados na tabela de *counts*, gerando uma nova tabela reanotada.

As quatro variáveis utilizadas pelo CPAT para avaliação do potencial de codificação são o tamanho da janela de leitura aberta (ORF, do inglês *Open Reading Frame*), a cobertura da ORF, o *score* de Fickett e o viés do uso de hexâmeros. Cada uma delas está explicada nos tópicos abaixo:

- O comprimento máximo da ORF é uma das características mais distintas entre RNA mensageiro e não codificantes, tendo em vista que ORFs longas sejam improváveis em sequências não codificantes.
- A cobertura diz respeito à proporção que aquela ORF representa do comprimento total do transcrito. Em ambos os casos um tamanho e cobertura menores são mais característicos dos RNAs não codificantes.
- O score de Fickett possibilita a distinção de mRNA e ncRNA através do efeito combinatório da composição de nucleotídeos e códons. Esse score é o resultado dos valores associados à posição da sequência e composição.
O viés de uso de hexâmeros mede o uso diferencial de hexâmeros entre sequências codificantes e não-codificantes. Para uma dada sequência de DNA, calcula-se a probabilidade da sequência sob o modelo de DNA codificador e sob o modelo de DNA não codificante, e então faz-se o logaritmo da razão dessas probabilidades.

4. ANÁLISE DE EXPRESSÃO DIFERENCIAL

Segundo Signal et al. (2016), o método mais comum de inferir a função de genes codificante e IncRNAs em um conjunto de dados é através da análise de expressão diferencial. Esta análise pode ser utilizada para identificar possíveis RNAs importantes para aquela análise, dada a diferença de expressão entre distintas condições, assim como resultar em uma lista de genes diferencialmente expressos (DEGs, do inglês *Differential Expression Analysis*) que pode ser utilizada para análise de enriquecimento funcional. O pipeline utilizado para este fim está diagramado na figura 11.



Figura 11 - Pipeline para análise de expressão diferencial

Para a análise de diferença de expressão o transcritoma de macaco foi avaliado com cada tempo após a infecção comparado com o controle. Além disso foram analisados também os *datasets* de camundongos e humanos *in vivo* comparando as amostras infectadas e não infectadas. Em todos os casos foi utilizada a ferramenta DESeq2 para análise de expressão diferencial, seguida do enriquecimento funcional dos genes utilizando a base de dados Reactome. Após a obtenção dos dados com o genoma já reanotado, foi realizada uma análise de expressão diferencial dos genes utilizando o pacote *DESeq2* no programa estatístico R. Nesta análise foram comparadas as amostras infectadas em relação às não infectadas para os *datasets* de humano e camundongos. Para a análise dos dados de Rhesus, comparou-se cada ponto de coleta ao ponto de 2 semanas antes da infecção (controle) utilizando o mesmo pacote. Esse pacote realiza a normalização dos dados para comparar as amostras com maior acurácia, calcula os níveis de expressão e realiza testes para verificar as diferenças de expressão. *DESeq2* é capaz de testar a expressão diferencial quantitativa focada na força da expressão e não apenas na mera presença desta (LOVE et al., 2014).

Para determinar um transcrito como diferencialmente expresso utilizou-se duas métricas estatísticas: p-valor ajustado usando 0 método de Benjamani-Hochberg e logaritmo do fold change, de modo que uma comparação com diferença de expressão válida deveria ter um p-valor menor que 0,05 e valor do log2 de fold change maior que 1,5 ou menor que -1,5. O p-valor é o menor nível de significância com que se rejeitaria a hipótese nula. Se o valor p ajustado for menor que o cut-off, você rejeita a hipótese nula (WRIGHT, 1992). O fold change (FC) representa a razão da expressão de cada gene na amostra teste em relação ao valor da amostra controle. Para facilitar os cálculos e comparações, são calculados em logaritmos, de forma que um log2 de FC > 1,5 ou < -1,5 representa uma diferença de valor aproximadamente três vezes maior ou menor da amostra de teste contra a controle (LOVE et al., 2014).

As listas de DEGs obtidas para cada comparação foram submetidas a uma análise de enriquecimento de vias, a fim de inferir uma importância biológica dessas alterações e encontrar as vias mais relacionadas ao processo de infecção. Para isso utilizamos o banco de dados de vias manualmente curadas chamado Reactome (FABREGAT et al., 2017), que pode ser acessado tanto na *web* quanto através do pacote *ReactomePA* disponível para download através do Bioconductor. Cada gene contido na lista de DEGs inserida como dado de entrada é mapeado com a vias existentes no banco através de uma análise de sobreposição e de topologia. Um

teste hipergeométrico é utilizando a fim de responder à pergunta: "Minha lista contém mais proteínas para a via X do que seria esperado por acaso?". Este teste produz um *score* de probabilidade, que é corrigido para taxa de falsa descoberta (FDR) usando o método de Benjamani-Hochberg. Já a análise topológica considera a conectividade dos genes nas vias onde estão contidos (FABREGAT et al., 2017). As vias foram consideradas alteradas quando o p-valor encontrado de acordo com a lista de genes inserida era menor que 0,05 em cada uma das comparações.

5. ANÁLISES DE CO-EXPRESSÃO E ASSOCIAÇÃO

A análise de associação pode fornecer informações de como os IncRNAs são regulados nas células e revelar possíveis alvos e mecanismos (SIGNAL et al., 2016). Dentre os diversos métodos existentes para medir e agrupar transcritos co-expressos, foi utilizada a ferramenta CEMiTool, que realiza a filtragem não supervisionada dos genes, identifica módulo de co-expressão e análises funcionais distintas (RUSSO et al., 2018). Em um único relatório essa ferramenta disponibiliza resultados de diferentes análises que foram obtidos através dos dados utilizados: ORA (do inglês Over-representation analysis), que pode ser utilizada para identificar as vias enriquecidas pelos genes daqueles módulos; GSEA (do inglês gene set enrichment analysis), que possibilita a associação da atividade daquele módulo em cada grupo de amostra; e análise de interação proteína-proteína (PPI, do inglês protein-protein interaction), que pode ser utilizada para identificar os genes principais (mais conectados) dos módulos (RUSSO et al., 2018). O dado de entrada da ferramenta é a tabela total de expressão normalizada (antes de fazer a análise de DEGs) e é possível definir diferentes parâmetros que serão utilizados pela ferramenta. Neste trabalho utilizamos os parâmetros padrão do pacote e o coeficiente de correlação de Spearman.

Após utilizar o CEMiTool foi realizada uma análise a parte de correlação também utilizando o coeficiente de correlação de Spearman para que fosse possível obter este valor par a par (neste caso, IncRNA - mRNA). Foi utilizado um *cut-off* de | 0,8| para classificar a correlação como significativa ou não. Desta forma, todos os IncRNAs que tinham o valor da correlação da expressão maior que 0,8 ou menor

que - 0,8 em relação a algum mRNA ao longo do tempo, foram considerados correlacionados.



Figura 12 - Pipeline para análise de co-expressão e associação.

Para a análise de co-expressão foi utilizada a ferramenta CEMiTool do *Bioconductor project*. Após essa análise foi possível identificar quais IncRNAs estavam relacionados a genes codificantes e inferir a sua função ou mecanismo no qual estava envolvido a partir do enriquecimento e predição IncRNA-proteína.

6. PREDIÇÃO DA INTERAÇÃO LNCRNA-PROTEÍNA

A fim de compreender como os IncRNAs e os genes co-expressos e correlacionados (coeficiente de correlação de *Spearman* > |0,8|) com estes interagiam, foi realizada uma análise de predição da interação par a par através da ferramenta RPISeq (<u>http://pridb.gdcb.iastate.edu/RPISeq/index.html</u>). O RPISeq é um conjunto de classificadores de aprendizado de máquina para prever interações de RNA-proteína usando apenas informações de sequência (MUPIRRALA; HONAVAR; DOBBS, 2011).

As previsões do RPISeq são baseadas em classificadores *Random Forest* (RF) ou *Support Vector Machine* (SVM) treinados e testados em dois conjuntos de dados de referência de interações RNA-proteína extraídos do PRIDB, um abrangente banco de dados de complexos RNA-proteína. As probabilidades de

interação geradas pelo RPISeq variam de 0 a 1. Considera-se uma previsão positiva a partir de > 0,5, indicando que o RNA e a proteína provavelmente interagem (MUPIRRALA; HONAVAR; DOBBS, 2011).

RESULTADOS E DISCUSSÃO

1. CARACTERIZAÇÃO DAS AMOSTRAS

As quatro amostras provenientes de macacos Rhesus ao longo dos catorze dias de infecção foram processadas e analisadas para o sequenciamento e acompanhamento da carga viral. Esse acompanhamento foi realizado a partir da quantificação genômica em um sistema de PCR em tempo real (SILVEIRA et al., 2017). Apesar do inóculo de alta dose,não foram detectados sintomas clínicos ou neurológicos em nenhum dos macacos durante todo o estudo. Para facilitar a compreensão das análises cada animal recebeu uma identificação simples (A, B, C e D). Os dados de sexo, idade e viremia ao longo da infecção estão ilustrados na figura abaixo (figura 13).



Figura 13 - Caracterização das amostras de Macaca mulatta

cópias/ mL de plasma

Na imagem é possível identificar as informações de sexo, idade e viremia de cada animal a partir do seu identificador. A viremia está representada por uma escala de cores vermelha, onde quanto mais escuro (animal C, dia 3) maior a quantidade de cópias/mL de sangue e quanto mais claro, menor a quantidade. A cor branca representa o valor de zero cópias virais/mL.

Em três dos quatro animais a carga viral no plasma já estava positiva no primeiro dia após a infecção, mas mas a resolução da infecção nestes animais também foi mais rápida, não sendo possível detectar carga virêmica a partir do quinto dia. O macaco C só apresentou viremia detectável a partir do terceiro dia de infecção, e esta detecção se estendeu até o dia 7. Considerando todos os animais, o

pico de infecção ocorreu no dia 3, com uma média de 23.379 cópias por mililitro de plasma (10.870-74.945 cópias por mililitro) e a partir do dia 10 nenhum dos animais apresentou viremia detectável no sangue.

2. IDENTIFICAÇÃO DE NOVOS LNCRNAS DE MACACOS

No genoma reanotado de Macaca mulatta foram identificados 45.293 transcritos codificantes e 14.696 não codificantes (figura 14A), dos quais 9.210 eram RNAs longos não codificantes, sendo 5.964 já anotados e 3.246 inéditos (figura 14B). Embora o número de IncRNAs já anotados seja próximo de 6.000, poucos possuem função e/ou validação relatada, de forma que a maioria foi obtida a partir de anotações automáticas. O genoma Mmul 8.0.1 (GCA 000772875.3) possui em sua anotação atual (versão 102) cerca de 6.898 lncRNAs e não foi encontrado na literatura nenhum trabalho mais recente que tenha buscado a reanotação do genoma de Rhesus. Isso significa que o nosso trabalho tem a capacidade de aumentar o número de IncRNAs conhecidos neste organismo expressivamente, o que pode ser de grande valia científica. Os transcritos totais identificados apresentaram uma proporção de distribuição dentro dos cromossomos (Figura 15A) similar ao descrito na documentação do genoma referência (GCA 000772875.3), de forma que os 1 e 7 são os que apresentam maior número. Apesar da diferença em número absoluto, a proporção relativa de genes codificantes e não codificantes é similar entre todos eles (Figura 15B).



Figura 14 - Número de transcritos identificados no genoma de Macaca mulatta

(A) Distribuição de transcritos codificantes, longos e outros não codificantes. (B) Dentro do número de IncRNAs, distribuição dos transcritos já anotados e dos descobertos neste estudo.



Figura 15 - Localização cromossômica dos transcritos de M. mulatta

(A) Distribuição do número absoluto de transcritos por localização cromossômica. (B) Valor relativo (%) das classes de transcritos (codificantes, não codificantes e novos IncRNAs) dentro de cada cromossomo.

3. ANÁLISE DE EXPRESSÃO DIFERENCIAL E ENRIQUECIMENTO

Na análise individual (por animal) de expressão diferencial, encontramos 3.437 transcritos (todos os animais e todos os tempos), sendo 2.426 codificantes e 1.011 não codificantes, dos quais 854 são IncRNAs (140 já anotados e 714 novos) (Figura 16A). Entretanto, apenas 38 transcritos são consistentemente expressos em todos os animais, sendo 20 IncRNAs (Figura 16B e 16C). Pode-se observar que os animais que apresentam um maior número de DEGs são aqueles que apresentam uma viremia mais acentuada frente aos demais.

Para as análises seguintes consideramos consistentes os genes que foram encontrados diferencialmente expressos em pelo menos 75% das amostras (n=3), totalizando 2211 em todas as comparações. Os genes foram analisados em cada comparação de tempo em relação ao controle e o maior número de DEGs ocorreu no primeiro dia após a infecção (D01vs2WK). Os dias 10 e 14 após a infecção foram os que menos apresentaram perturbação (Figura 16D), o que pode estar relacionado com a resolução da infecção viral no sangue já descrito em outros flavivírus neste mesmo período após a infecção (RATTERRE et al., 2004).



Figura 16 - Distribuição de transcritos diferencialmente expressos

Distribuição de transcritos diferencialmente expressos em todos os macacos e todos os tempos, onde a cor laranja representa o número de DEGs codificantes e a azul de não codificantes. A) Número total de DEGs nas comparações do *dataset* de *Macaca mulatta*. B) DEGs codificantes por animal e suas interseções. C) IncRNAs diferencialmente expressos por animal e suas interseções. D) Número de genes diferencialmente expressos ao longo da infecção, sempre em comparação com a amostra basal.

Foi realizada uma análise de enriquecimento de vias a partir das listas de DEGs consistentes de cada tempo, na base de dados Reactome, e aquelas que foram identificadas como significativas em pelo menos uma das comparações foi avaliada quanto à expressão e enriquecimento nos demais tempos.

A via de regulação de BCR foi demonstrada como negativamente regulada no primeiro dia, voltando ao normal logo em seguida. Já a via de mecanismo antiviral teve um pico de expressão do terceiro dia e redução acentuada no quinto. A sinalização por IL-4 e IL-13 também foi alterada ao longo dos dias e, embora seja associada à resposta imunológica contra helmintos, a sua expressão diferencial já foi demonstrada em infecções virais (PALUDAN et al., 1997). Dentre estas encontramos "Mecanismo antiviral por genes estimulados por Interferon" (R-HSA-1169410) como a mais enriquecida significativamente. A fim de analisar de forma mais abrangente, buscamos também por genes contidos na via "Sinalização por Interferon" (R-HSA-913531). Nestas vias estão contidos os genes MX2, ISG15 e OAS1, regulados positivamente nas comparações dos dias 3 e 5, e STAT1, JAK1, *IFNGR2* e *EGR1* regulados negativamente em diferentes tempos (figura 17).





Na parte inferior da figura estão sinalizadas as comparações temporais e cada animal em todas as comparações (representados pelas cores rosa, verde, amarelo e turquesa). No eixo X direito estão os identificadores dos genes pertencentes à via, de forma que as verdes estão associadas à sinalização por IFN e a marrom à resposta antiviral. No eixo esquerdo uma árvore de classificação da expressão. Os genes em vermelho representam aqueles expressos positivamente em relação ao tempo controle e os expressos negativamente estão representados em azul. Os diferencialmente expressos são os de cor mais forte em ambas as condições de acordo com o *cut-off* de *foldchange* de 1.5.

Na comparação do dia 1 após a infecção, há a maior quantidade de genes com regulação negativa da expressão, incluindo o gene *IFNGR2*, que codifica uma subunidade do receptor de interferon gama. Essa redução impede a formação do complexo responsável pela maquinaria de ativação da via JAK-STAT, cujos genes também estão regulados negativamente neste dia, reduzindo a resposta inflamatória (SCHNEIDER et al., 2015). Nas comparações dos dias seguintes de infecção em relação ao controle (D03vs 2WK e D05vs2WK), a regulação positiva dos genes *MX2*, *ISG15* e *OAS1* está de acordo com o que foi encontrado por Aid et al. (2017) em um estudo de *microarray* em amostras de Rhesus infectados. O gene MX2 também é essencial na supressão de HIV *in vitro* e *in vivo* (ASMUTH et al., 2010; SCHOGGINS et al., 2011), assim como *OAS1* é mediador da degradação de RNA viral na infecção por vírus da Hepatite E (HEV) (SOORYANARAIN et al., 2017). A atividade antiviral de *ISG15* frente a alguns flavivírus como DENV e WNV foi demonstrada por Dai et al. (2011) em um estudo *in vitro*, onde a expressão desse gene foi pelo menos nove vezes superior às células controle.

Na comparação do dia 10 e do dia 14 após a infecção há um menor número de genes diferencialmente expressos consistentemente, demonstrando que o organismo está voltando ao seu controle de expressão basal. No dia 10 o único gene consistentemente negativo é *STAT1*, cuja diminuição desencadeia uma supressão na atividade de macrófagos frente ao estímulo por interferon (BROMBERG et al., 1996) e diminuição da atividade citolítica de células NK (LIANG et al., 2003).

Nos dados clínicos dos nossos animais (SILVEIRA et al., 2017), observamos uma diminuição na secreção de perforina, molécula necessária para a resposta citotóxica de células NK, também no dia 10 após a infecção. Na figura 18 esses genes estão sinalizados dentro das vias de ação, onde é possível visualizar de melhor forma como a repressão de *JAK1*, *STAT1* e *IFNGR2* leva à redução da resposta imunológica. Além disso é possível ver o envolvimento de *OAS1*, *ISG15* e proteínas *MX* tanto na sinalização quanto na migração para mecanismos antivirais.



Figura 18 - Vias de Sinalização por Interferon e Mecanismo antiviral

Em cinza estão os genes que não apareceram na nossa análise, enquanto os coloridos são nossos DEGs com o *cut-off* de *foldchange* de 1,5. Imagem adaptada de Reactome (<u>https://reactome.org/</u>).

Os IFN alfa e beta normalmente recrutam proteínas *JAK1* e *TYK2*. Após a formação do complexo, o seus sinais são transmitidos para *STAT1* e 2 que, em combinação com *IFN9*, formam o heterodímero ISGF3 que, por sua vez, promove a indução da transcrição de genes através da ligação ao elemento responsivo a Interferon (ISRE). Cascata semelhante ocorre na sinalização por IFN-gama, onde após se ligar aos receptores *IFNGR1* e *IFNGR2* e recrutar proteínas *JAK*, há a formação de um dímero de *STAT1* que, ao se ligar a elementos de sequência ativadas por IFN-gama (*GAS*) inicia a transcrição de genes responsivos à citocina de mesmo nome (PLATANIAS, 2005).

Encontramos também vias relacionadas ao ciclo de vida viral e decaimento de mRNA mediado por mutações *nonsense* (NMD, do inglês *Nonsense-mediated decay*), que compartilham grande parte dos genes descritos em cada via, como diferentes proteínas ribossomais e os genes SMG6 e DCP1A. NMD é um mecanismo de controle celular responsável pela degradação de mRNAs aberrantes (BEHM-ANSMANT & IZAURRALDE, 2006), que desempenha um papel antiviral

capaz de restringir a replicação de vírus de RNA fita positiva (BALISTRERI et al., 2017).

Nas comparações dos dias 1, 3 e 5 após a infecção, o aumento da expressão de genes que codificam proteínas ribossomais como RPLP1, somado à diminuição acentuada de expressão de SMG6 e DCP1A, sugerem um forte mecanismo de evasão viral logo na fase inicial de infecção. De acordo com Campos et al. (2017), o gene RPLP1 é essencial para a infecção por flavivirus em modelos in vitro. Na infecção por ZIKV à células humanas provenientes de hepatocarcinoma (HuH-7), a replicação viral e a capacidade de infectividade celular depende da expressão de RPLP1. Os genes SMG6 e DCP1A codificam enzimas envolvidas na degradação final de RNAs aberrantes pela via do NMD (EBERLE et al., 2009), de forma que a sua supressão resulta no aumento de RNA viral circulante (RAO et al., 2018). Vírus como o poliovírus e o vírus da síndrome reprodutiva e respiratória dos suínos (PRRSV) possuem enzimas capazes de clivar DCP1A atenuando a resposta antiviral do hospedeiro (DOUGHERTY et al., 2011; TAO et al., 2018).

5. ANÁLISE DE CO-EXPRESSÃO E ASSOCIAÇÃO

A inferência de função por associação baseia-se na ideia de que há uma probabilidade grande de transcritos co-expressos compartilharem funções ou estarem envolvidos nos mesmos processos. Com essa premissa é possível atribuir funções putativas de acordo com os transcritos co-expressos, principalmente no caso de IncRNAs. Para tanto, é necessário fazer um enriquecimento de vias com os genes relacionados a cada módulo (SIGNAL et al., 2016). Utilizando a ferramenta CEMiTool (RUSSO et al., 2018) para todo o *dataset* (sem a análise diferencial), foram identificados nove módulos de co-expressão, cujos genes de cada um destes módulos foram enriquecidos utilizando as base de dados Reactome.

Dentre os módulos, os que apresentaram maior enriquecimento significativo foram o módulo M1 - Resposta Imune Inata (p = 0,008) e módulo M3 - Regulação de BCR por CD22 (p = 0,0001), vias estas que corroboram com as encontradas na análise diferencial. A partir desses módulos buscamos por aqueles IncRNAs cuja

correlação de expressão com outros genes foi maior que 0,8 ao longo do tempo em comparação ao controle.

O módulo M1 foi enriquecido para resposta imune inata, que é importante para o combate à infecção, como citado na introdução. Neste módulo estão contidos 44 IncRNAs e distintos genes, como *BCL2*, *CYBA* e *MYCT1*, associados com a regulação da apoptose. Os principais genes e IncRNAs identificados seguem uma tendência similar de perturbação da expressão principalmente no primeiro dia após a infecção, com subsequente oscilação nos demais dias, até voltarem ao valor basal no décimo quarto dia (figura 19).

Suzuki et al. (2018), demonstraram em estudos *in vitro* que a inibição de *BCL2* pode ser um mecanismo antiviral, tendo em vista que essa supressão desencadeia o aumento da apoptose de células infectadas através do aumento de genes *BAX/BAK* e, subsequentemente, ocorre o aumento da fagocitose, que leva ao clareamento viral.

No nosso trabalho os genes *CYBA* e *MYCT1* estão contidos no mesmo módulo que *BCL2* e, possivelmente, estão envolvidos nesse mesmo mecanismo de resposta. O gene *CYBA* codifica a proteína p22^{phox}, que faz parte do complexo de NADPH oxidase fagocítica, presente na superfície de macrófagos e neutrófilos (SCHÄPPI et al., 2008). Fu et al. (2018), correlacionaram a superexpressão de *MYCT1* com a supressão tumoral e discutiram que a inibição da proliferação ocorria por apoptose. Essa superexpressão ocorreu concomitantemente com a diminuição da expressão de *BCL2*, assim como no primeiro dia após infecção do nosso estudo.

Neste mesmo módulo quatro IncRNAs ainda não anotados foram identificados com uma alta correlação com esses genes supracitados. Destes, três possuem uma correlação inversa com *MYCT1* e *BCL2* e por isso aqui foram listados como *PROAP*, tendo em vista que o seu aumento está associado a modificações pró-apoptóticas. *PROAP1* é um IncRNA de 1642 nucleotídeos sem similaridade significante com sequências de qualquer organismo contido na base de dados utilizada pelo BLAST (do inglês *Basic Local Alignment Search Tool*).





Em azul e verde estão sinalizados os genes codificantes e em tons de laranja tracejados os não codificantes. No eixo X há tempo de infecção desde o basal (controle). Pode-se observar que a tendência de expressão dos lncRNAs PROAP1, 2 e 3 é a mesma do mRNA MYCT1, tal como a expressão do lncRNA SURV1 segue tendência similar a dos mRNAs BCL2 e CYBA ao longo do tempo.

O módulo M3 foi enriquecido principalmente para as vias de regulação do receptor de células B (BCR), onde encontramos genes importantes como IL12B e RAB27B. As células B desempenham um papel crucial no combate a infecções a partir da produção de anticorpos que, por sua vez, são essenciais para a neutralização e destruição dos patógenos. O gene da IL12B codifica uma subunidade da interleucina 12, uma citocina com ampla atividade biológica, que é expressa por macrófagos e tem um importante papel para a resposta imunológica de perfil Th1 (CHEN et al., 2012). RAB27B aumenta o número de exossomas, que são utilizados pelo vírus como uma forma de escape e propagação (ANDERSON et al., 2016). Neste módulo estão contidos também os genes *BLK*, *KLHL25*, *IGLC2*, *IGLC7* e *GIMAP1* que, assim como os do módulo 1, apresentam maior perturbação nas

primeiras 24 horas de infecção. A diminuição desses genes no primeiro momento e posterior aumento pode estar associado com uma diminuição da resposta imune induzida pelo vírus, mas com recuperação da homeostase até o basal no dia 14 (figura 20).



Figura 20 - Cinética da expressão de genes e IncRNAs do módulo 3

Em azul e verde estão sinalizados os genes codificantes e em laranja tracejado o não codificante. No eixo X há tempo de infecção desde o basal (controle). Pode-se observar que os mRNAs e o lncRNA identificado possuem a mesma tendência de expressão ao longo do tempo, de forma que no dia 14 tendem a voltar aos valores de expressão do basal (controle).

A proteína quinase de linfócitos B (*BLK*) envolvida no desenvolvimento, diferenciação e sinalização de linfócitos B (LAIRD et al., 2014). *GIMAP1* é essencial para a sobrevivência de linfócitos B naive e ativados (WEBB et al., 2016). As *IGLC* são regiões constantes de cadeias leves de imunoglobulinas, que são proteínas que quando associadas à células B atuam como seu receptor (SCHROEDER et al., 2010). *KLHL25* é uma proteína adaptadora que, ao se ligar à proteína *CUL3*, possibilita a ubiquitinação e degradação da enzima ATP citrato liase (*ACLY*), o que provoca a diminuição da síntese lipídica e energia disponível para célula (ZHANG et al., 2016). O seu aumento após o dia 01, de acordo com o que está descrito na literatura, pode auxiliar na proliferação de células B (ZHANG et al., 2016). O lcnRNA identificado como diretamente correlacionado com esses genes possui 534 nucleotídeos e alta similaridade com a sequência predita computacionalmente

LOC108513237, sem referências na literatura. Esse IncRNA não nomeado também tem a sua expressão alterada pelo início da infecção, mas retorna ao seu valor basal após 14 dias da inoculação.

6. PREDIÇÃO DA INTERAÇÃO LNCRNA-PROTEÍNA

Os IncRNAs e mRNAs identificados foram avaliados quanto ao seu potencial de interação utilizando a ferramenta web RPISeq. Para tanto foi utilizada a sequência de nucleotídeos de cada um dos IncRNAs frente à sequência de aminoácidos das proteínas codificadas pelos mRNAs correlacionados. Essas sequências de aminoácidos foram obtidas na base de dados do NCBI. A partir desta análise o IncRNA PROAP1 foi o que apresentou maior probabilidade de interagir diretamente com *BCL2* (com valor igual a 0.8 utilizando *Random Forest* e 0.96 utilizando *Suport Vector Machine*) de acordo com a predição do RPISeq. *PROAP2* e *PROAP3* possuem similaridade com LOC106999565 e LOC106992852, respectivamente, mas ambos foram anotados automaticamente utilizando o preditor Gnomon, ainda sem relatos da sua expressão na literatura. Na figura 21 a possível ação desses IncRNAs na indução da apoptose está ilustrada.



Figura 21 - Indução da apoptose a partir da regulação de IncRNAs

Os IncRNAs PROAP parecem estar envolvidos com a inibição de *BCL2*, o que acaba por levar ao aumento de *BAX*, e o aumento de *MYCT1*. A modificação da expressão desses dois genes têm uma ação pró-apoptótica. Fonte: autoral, criada com Biorender (https://biorender.com/).

7. OUTROS MAMÍFEROS

A fim de tentar avaliar como os IncRNAs são alterados durante a infecção, os dados de transcritoma de outros estudos foram analisados com foco nos IncRNAs, o que não havia sido feito pelos seus autores originais. Para este fim utilizamos um *dataset* de infecção *in vivo* em camundongos e de bebês humanos.

Dos transcritos analisados do *dataset* de camundongos infectados, 285 apresentaram diferença de expressão, todos com regulação positiva em relação ao grupo não infectado. Destes, 10 correspondem a RNAs longos não codificantes, dos quais metade também aparecem como DEGs em um estudo de infecção por WNV em cérebro de camundongos (LIM et al., 2017), embora não tenham função descrita na literatura. Dentre os nossos DEGs podemos citar o IncRNA *XIST* (log2FC de 10,12 e p ajustado igual a 0,0085), cuja superexpressão já foi relacionada à redução da proliferação celular, migração e aumento de apoptose em células cancerígenas a partir da inibição do microRNA anti apoptótico miR-21-5p, que quando expresso aumenta a expressão de *BCL2* (XU et al., 2014; ZHANG & XIA, 2017). Embora pouco explorado em infecções virais, a partir da sua função pró-apoptótica já conhecida, o aumento de XIST pode ter sido originado de modificações causadas pelo sistema imunológico dos camundongos a fim de combater o vírus, corroborando com o aumento de apoptose observada anteriormente em nossos resultados.

Dos transcritos analisados no *dataset* de bebês, 442 atenderam aos critérios de diferencialmente expressos. Dentre estes encontra-se o IncRNA H19 (log2FC de -10,65 e p ajustado de 1,71e-26), envolvido na regulação do crescimento celular (KENIRY et al., 2012). A sua regulação positiva está associada uma maior proteção ao estresse celular e melhor sobrevivência sob hipóxia, sendo comum e necessária em tecido cerebral durante o desenvolvimento fetal (PHAM et al., 1998; DENG et al., 2016). A diminuição de sua expressão pode ocasionar a redução de marcadores de pluripotência em células-tronco embrionárias, além de reduz a viabilidade celular e a angiogênese (ZEIRA et al., 2015). H19 pode também inibir a ação de miR-19b, que por sua vez tem como alvo o gene pró-apoptótico SOX6 (HAN et al., 2016).

Partindo do pressuposto levantado por um trabalho recente que mostra que o início das alterações causadas por ZIKV envolve a redução do crescimento celular e alterações na neurogênese (GARCEZ et al., 2016), uma redução de expressão do H19 tão acentuada não só corrobora com os nossos resultados em macacos como também podem indicar uma relação entre a supressão de expressão desse IncRNA e a alteração do desenvolvimento neuronal dos bebês.

8. VISÃO GLOBAL DO PAPEL DE LNCRNAS NA INFECÇÃO

A partir da união de todos esses *datasets* podemos observar que, além da alteração de resposta imune já esperada nesse tipo de infecção, os IncRNAs encontrados parecem ter um papel importante na regulação e aumento da apoptose de células infectadas, principalmente a partir da inibição direta ou indireta da expressão de *BCL2*. O mecanismo de apoptose na infecção por ZIKV já foi descrito em experimentos *in vitro* e *in vivo* cujo objetivo era avaliar a neuropatologia causada por este vírus (MINER et al., 2016; HO et al., 2017; HU et al., 2017; SLOMNICK et al., 2017), mas os mecanismos de apoptose em uma infecção sistêmica por Zika ainda não foram bem discutidos.

Interessantemente, em um desses estudos (Hu et al., 2017) - realizado com células progenitoras neurais - os autores encontraram dois IncRNAs (LINC00963 e LINC00342) anti-apoptóticos regulados negativamente após a infecção por Zika. Somado ao aumento do IncRNA PVT1, inibidor de proliferação, os autores discutem que a infecção leva a uma desregulação do ciclo celular e alterações transcricionais acentuadas. Resultados estes que corroboram com o que encontramos neste trabalho.

Estudos em outros flavivírus como dengue, oeste do nilo e vírus da encefalite japonesa já descreveram a atividade pró-apoptótica que ocorre após a infecção (DESPRES et al., 1996; PARQUET et al., 2001; HUANG et al., 2014; HUANG et al., 2016; SUZUKI et al., 2018) mas nenhum deles fez referência à presença ou importância de IncRNAs no mecanismo. Desta forma, a partir dos resultados apresentados, sugere-se que a atividade pró-apoptótica já conhecida e estudada nas infecções por flavivírus sofre interferência de RNAs longos não codificadores, principalmente pela regulação de *BCL2*, em diferentes organismos, mesmo que os IncRNAs não seja exatamente os mesmos (figura 22).



Figura 22 - Possível regulação da apoptose a partir de IncRNAs

(a) H19 foi encontrado em amostras de bebês infectado e, de acordo com a literatura, acaba tendo uma função pró-apoptótica por contribuir com a inibição de BCL2. (b) O IncRNA XIST inibe miR-21, que quando expresso aumenta a atividade de BCL2. (c.) Os IncRNAs identificados neste estudo parecem ter uma função na inibição de BLC2, mas não se sabe se é uma ação direta ou há um intermediário. (d) em células infectadas *in vitro* foi observado o enriquecimento da via UTR, que hiperativa a proteína PERK, levando à produção de proteínas pró-apoptóticas. Observamos a expressão de SNHG15, mas não há relação dele com o processo estudado na literatura.

CONCLUSÃO

A partir dos resultados da diferença de expressão, enriquecimento de vias e análise de co-expressão e associação, podemos observar que já no primeiro dia de infecção há resposta antiviral mediada pelo sistema imune inato do hospedeiro e aumento da apoptose. Após o sétimo dia de infecção parece haver uma redução de expressão diferencial, que pode estar associada a um combate efetivo da infecção, com "clareamento viral" já no último dia, onde o perfil de expressão se assemelha ao controle. Durante a resposta antiviral parece que alguns IncRNAs desempenham papéis importantes na regulação da expressão gênica, principalmente relacionada com atividade pró-apoptótica e redução da proliferação celular. Esse resultado parece ser uma resposta comum à diferentes mamíferos infectados por ZIKV, embora os IncRNAs encontrados não apresentem similaridade.

REFERÊNCIAS¹

AGRELLI, A. et al. ZIKA virus entry mechanisms in human cells. **Infection, Genetics and Evolution**, 2019.

AID, M. et al. Zika Virus Persistence in the Central Nervous System and Lymph Nodes of Rhesus Monkeys. **Cell**, v. 169, n. 4, p. 610–620.e14, 2017.

AKIRA, S.; TAKEDA, K.; KAISHO, T. Toll-like receptors: critical proteins linking innate and acquired immunity. **Nature immunology**, v. 2, n. 8, p. 675, 2001.

ANDERSON, M.R.; KASHANCHI, F.; JACOBSON, S. Exosomes in viral disease. **Neurotherapeutics**, v. 13, n. 3, p. 535-546, 2016.

ANDREWS, S. *et al.* FastQC: a quality control tool for high throughput sequence data. Website. 2010. Disponível em: <<u>https://www.bioinformatics.babraham.ac.uk</u>>.

ANFASA, F. et al. Phenotypic differences between Asian and African lineage Zika viruses in human neural progenitor cells. **MSphere**, v. 2, n. 4, p. e00292-17, 2017.

ASHBURNER, M.et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p. 25, 2000.

ASMUTH, D. et al. Safety, Tolerability, and Mechanisms of Antiretroviral Activity of Pegylated Interferon Alfa-2a in HIV-1–Monoinfected Participants: A Phase II Clinical Trial. **The Journal of Infectious Diseases**, v. 201, n. 11, p. 1686-1696, 2010.

BALISTRERI, Giuseppe; BOGNANNI, Claudia; MÜHLEMANN, Oliver. Virus escape and manipulation of cellular nonsense-mediated mRNA decay. **Viruses**, v. 9, n. 1, p. 24, 2017.

BEHM-ANSMANT, I. Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. **Genes & Development**, v. 20, n. 4, p. 391-398, 2006.

¹ Correspondente às normas da ABNT vigentes em 2019

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

BOURGON, R.; GENTLEMAN, R.; HUBER, W. Independent filtering increases detection power for high-throughput experiments. **Proceedings of the National Academy of Sciences**, v. 107, n. 21, p. 9546-9551, 2010.

BRIGGS, J. et al. Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. **Neuron**, v. 88, n. 5, p. 861-877, 2015.

BROMBERG, J.F. et al. Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma. **Proceedings** of the National Academy of Sciences, v. 93, n. 15, p. 7673-7678, 1996.

BROWN, P.O.; BOTSTEIN, D. Exploring the new world of the genome with DNA microarrays. **Nature Genetics**, v. 21, 1999.

CAMPOS, R. et al. RPLP1 and RPLP2 Are Essential Flavivirus Host Factors That Promote Early Viral Protein Accumulation. **Journal of Virology**, v. 91, n. 4, 2016.

CAMPOS, T. L. et al. Revisiting Key Entry Routes of Human Epidemic Arboviruses into the Mainland Americas through Large-Scale Phylogenomics. **International Journal of Genomics**, v. 2018, 2018.

CAO-LORMEAU, V.-M. et al. Zika virus, French polynesia, South pacific, 2013. **Emerging Infectious Diseases**, v. 20, n. 6, p. 1085–6, 2014.

CARNERO, E. et al. Long noncoding RNA EGOT negatively affects the antiviral response and favors HCV replication. **EMBO reports**, 2016.

CARPENTER, S. Long noncoding RNA: Novel links between gene expression and innate immunity. **Virus Research**, v. 212, p. 137–145, jan. 2016.

CHAMBERS, T. J. et al. Flavivirus Genome Organization, Expression, and Replication. **Annual Review of Microbiology**, v. 44, n. 1, p. 649–688, 28 out. 1990.

CHAN, J. et al. Zika fever and congenital Zika syndrome: an unexpected emerging arboviral disease. **Journal of Infection**, 2016.

CHEN, J; ZHANG, Y.; DENG, Z. Imbalanced shift of cytokine expression between T helper 1 and T helper 2 (Th1/Th2) in intestinal mucosa of patients with post-infectious irritable bowel syndrome. **BMC gastroenterology**, v. 12, n. 1, p. 91, 2012.

CIMMINO, A. et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. **Proceedings of the National Academy of Sciences**, v. 102, n. 39, p. 13944-13949, 2005.

CROFT, D. et al. Reactome: a database of reactions, pathways and biological processes. **Nucleic Acids Research**, v. 39, n. suppl_1, p. D691-D697, 2010.

DAI, J.; PAN, W.; WANG, P. ISG15 facilitates cellular antiviral response to dengue and west nile virus infection in vitro. **Virology Journal**, v. 8, n. 1, p. 468, 2011.

DARBELLAY, J. et al. Zika virus causes persistent infection in porcine conceptuses and may impair health in offspring. **EBioMedicine**, v. 25, p. 73-86, 2017.

DAS, S. et al. Tumor cell entry into the lymph node is controlled by CCL1 chemokine expressed by lymph node lymphatic sinuses. **The Journal of Experimental Medicine**, v. 210, n. 8, p. 1509-1528, 2013.

DELGADO-VEGA, A. M. et al. Fine mapping and conditional analysis identify a new mutation in the autoimmunity susceptibility gene BLK that leads to reduced half-life of the BLK protein. **Annals of the rheumatic diseases**, v. 71, n. 7, p. 1219-1226, 2012.

DENG, Y. et al. Prostacyclin-producing human mesenchymal cells target H19 IncRNA to augment endogenous progenitor function in hindlimb ischaemia. **Nature Communications**, v. 7, p. 11276, 2016.

DICK, G. W. A.; KITCHEN, S. F.; HADDOW, A. J. Zika virus (I). Isolations and serological specificity. **Transactions of the royal society of tropical medicine and hygiene**, v. 46, n. 5, p. 509-520, 1952.

DIKHIT, M. R. et al. Computational prediction and analysis of potential antigenic CTL epitopes in Zika virus: A first step towards vaccine development. **Infection, Genetics and Evolution**, v. 45, p. 187-197, 2016.

DING, Y. et al. Relationship of long noncoding RNA and viruses. **Genomics**, v. 107, n. 4, p. 150-154, 2016.

DJEBALI, S. et al. Landscape of transcription in human cells. **Nature**, v. 489, n. 7414, p. 101, 2012.

DOUGHERTY, J.; WHITE, J.; LLOYD, R. Poliovirus-Mediated Disruption of Cytoplasmic Processing Bodies. **Journal of Virology**, v. 85, n. 1, p. 64-75, 2010.

DUDLEY, D.M. et al. A rhesus macaque model of Asian-lineage Zika virus infection. **Nature Communications**, v. 7, p. 12204, 2016.

DUYEN, H. TL et al. Kinetics of plasma viremia and soluble nonstructural protein 1 concentrations in dengue: differential effects according to serotype and immune status. **Journal of Infectious Diseases**, v. 203, n. 9, p. 1292-1300, 2011.

EBERLE, A. et al. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. **Nature Structural & Molecular Biology**, v. 16, n. 1, p. 49-55, 2008.

FAGHIHI, M. A. et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β -secretase. **Nature medicine**, v. 14, n. 7, p. 723, 2008.

FATICA, A. & BOZZONI, I. Long non-coding RNAs: new players in cell differentiation and development. **Nature Reviews Genetics**, v. 15, n. 1, p. 7, 2014.

FAYE, O. et al. One-step RT-PCR for detection of Zika virus. **Journal of Clinical**, 2008.

FAYE, O. et al. Molecular evolution of Zika virus during its emergence in the 20th century. **PLoS neglected tropical diseases**, v. 8, n. 1, p. e2636, 2014.

FORTES, P.; MORRIS, K. Long noncoding RNAs in viral infections. **Virus research**, 2016.

FRIBLEY, A. M. et al. Proteasome inhibitor PS-341 induces apoptosis in cisplatin-resistant squamous cell carcinoma cells by induction of Noxa. **Journal of Biological Chemistry**, v. 281, n. 42, p. 31440-31447, 2006.

FU, S. et al. Overexpression of MYCT1 inhibits proliferation and induces apoptosis in human acute myeloid leukemia HL-60 and KG-1a cells in vitro and in vivo. **Frontiers in pharmacology**, v. 9, p. 1045, 2018.

GARCEZ, P.P. et al. Zika virus impairs growth in human neurospheres and brain organoids. **Science**, v. 352, n. 6287, p. 816-818, 2016.

GARCIA, E. et al. Zika virus infection: global update on epidemiology and potentially associated clinical manifestations. **Weekly Epidemiol**, 2016.

GOMEZ, J. et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. **Cell**, 2013.

GONG, Z.; GAO, Y; HAN, G. Zika virus: two or three lineages?. **Trends in microbiology**, v. 24, n. 7, p. 521-522, 2016.

GOULD, E.; SOLOMON, T. Pathogenic flaviviruses. **The Lancet**, v. 371, n. 9611, p. 500–509, 2008.

GUPTA, R.A. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. **Nature**, v. 464, n. 7291, p. 1071, 2010.

HACKETT, B.A.; CHERRY, S. Flavivirus internalization is regulated by a size-dependent endocytic pathway. **Proceedings of the National Academy of Sciences**, v. 115, n. 16, p. 4246-4251, 2018.

HAN, Y. et al. Downregulation of long non-coding RNA H19 promotes P19CL6 cells proliferation and inhibits apoptosis during late-stage cardiac differentiation via miR-19b-modulated Sox6. **Cell & bioscience**, v. 6, n. 1, p. 58, 2016.

HARDING, H.P. et al. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. **Molecular cell**, v. 11, n. 3, p. 619-633, 2003.

HASAN, S. S.et al. Structural biology of Zika virus and other flaviviruses. **Nature structural & molecular biology**, v. 25, n. 1, p. 13, 2018.

HELLER, M. DNA microarray technology: devices, systems, and applications. **Annual review of biomedical engineering**, 2002.

HO, C. et al. Differential neuronal susceptibility and apoptosis in congenital Zika virus infection. **Annals of neurology**, v. 82, n. 1, p. 121-127, 2017.

HU, B. et al. ZIKV infection effects changes in gene splicing, isoform composition and IncRNA expression in human neural progenitor cells. **Virology journal**, v. 14, n. 1, p. 217, 2017.

IMAMURA, K. et al. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. **Molecular cell**, 2014.

IRIZARRY, R.A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. **Biostatistics**, v. 4, n. 2, p. 249-264, 2003.

JIANG, S.H. et al. Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. **Nature Communications**, v. 10, n. 1, p. 2201, 2019.

JOSHI, R. et al. Phosphoproteomics Reveals Regulatory T Cell-Mediated DEF6 Dephosphorylation That Affects Cytokine Expression in Human Conventional T Cells. **Frontiers in Immunology**, v. 8, 2017.

KAUFFMANN, A.; GENTLEMAN, R.; HUBER, W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. **Bioinformatics**, v. 25, n. 3, p. 415-416, 2008.

KAWAI, T.; AKIRA, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. **Nature immunology**, v. 11, n. 5, p. 373, 2010.

KAWAI, T.; AKIRA, S. The roles of TLRs, RLRs and NLRs in pathogen recognition. **International immunology**, v. 21, n. 4, p. 317-337, 2009.

KENIRY, A. et al. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and lgf1r. **Nature cell biology**, v. 14, n. 7, p. 659, 2012.

KIELIAN, M. Mechanisms of virus membrane fusion proteins. **Annual review of virology**, v. 1, p. 171-189, 2014.

KUMAR, H.; KAWAI, T.; AKIRA, S. Pathogen recognition by the innate immune system. **International reviews of immunology**, v. 30, n. 1, p. 16-34, 2011.

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, 5 dez. 2013.

KIM, J. et al. Insights into ZIKV-Mediated Innate Immune Responses in Human Dermal Fibroblasts and Epidermal Keratinocytes. **Journal of Investigative Dermatology**, v. 139, n. 2, p. 391-399, 2019.

KIM, S.Y. et al. Interaction of Zika virus envelope protein with glycosaminoglycans. **Biochemistry**, v. 56, n. 8, p. 1151-1162, 2017.

KONG, Q; QIU, M. Long noncoding RNA SNHG15 promotes human breast cancer proliferation, migration and invasion by sponging miR-211-3p. **Biochemical and biophysical research communications**, v. 495, n. 2, p. 1594-1600, 2018.

LAIRD, R.M.; LAKY, K.; HAYES, S.M. Unexpected Role for the B Cell-Specific Src Family Kinase B Lymphoid Kinase in the Development of IL-17–Producing $\gamma\delta$ T Cells. **The Journal of Immunology**, v. 185, n. 11, p. 6518-6527, 2010.

LANCIOTTI, R. S. et al. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. **Emerging infectious diseases**, v. 14, n. 8, p. 1232–9, ago. 2008.

LANDER, E. et al. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860, 2001.

LEE, J. T. Epigenetic Regulation by Long Noncoding RNAs. **Science**, v. 338, n. 6113, 2012.

LIANG, S. et al. IFNα regulates NK cell cytotoxicity through STAT1 pathway. **Cytokine**, v. 23, n. 6, p. 190-199, 2003.

LIANG, Q. et al. Zika virus NS4A and NS4B proteins deregulate Akt-mTOR signaling in human fetal neural stem cells to inhibit neurogenesis and induce autophagy. **Cell stem cell**, v. 19, n. 5, p. 663-671, 2016.

LIAO, Y.; SMYTH, G. K.; SHI, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923–930, 2014.

LIM, S.M. et al. Transcriptomic Analyses Reveal Differential Gene Expression of Immune and Cell Death Pathways in the Brains of Mice Infected with West Nile Virus and Chikungunya Virus. **Frontiers in microbiology**, v. 8, p. 1556, 2017.

LIU, B. et al. A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. **Cancer cell**, v. 27, n. 3, p. 370-381, 2015.

LOSKO, M.; KOTLINOWSKI, J.; JURA, J. Long noncoding RNAs in metabolic syndrome related disorders. **Mediators of inflammation**, v. 2016, 2016.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, 2014.

MACNAMARA, F. Zika virus: a report on three cases of human infection during an epidemic of jaundice in Nigeria. **Transactions of the Royal Society of Tropical Medicine**, 1954.

MARCHESE, F.P.; RAIMONDI, I.; HUARTE, M. The multidimensional mechanisms of long noncoding RNA function. **Genome biology**, v. 18, n. 1, p. 206, 2017.

MCGRATH, E. L. et al. Differential Responses of Human Fetal Brain Neural Stem Cells to Zika Virus Infection. **Stem Cell Reports**, v. 8, n. 3, p. 715–727, mar. 2017.

MINER, J.J. et al. Zika virus infection in mice causes panuveitis with shedding of virus in tears. **Cell reports**, v. 16, n. 12, p. 3208-3218, 2016.

MONDAL, T. et al. MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA–DNA triplex structures. **Nature communications**, v. 6, p. 7743, 2015.

MORRISON, J. et al. Dengue virus co-opts UBR4 to degrade STAT2 and antagonize type I interferon signaling. **PLoS pathogens**, v. 9, n. 3, p. e1003265, 2013.

MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nature**, 2008.

MUÑOZ-JORDÁN, J.L. et al. Inhibition of interferon signaling by dengue virus. **Proceedings of the National Academy of Sciences**, v. 100, n. 24, p. 14333-14338, 2003.

MUPPIRALA, U.K.; HONAVAR, V.G.; DOBBS, D. Predicting RNA-protein interactions using only sequence information. **BMC bioinformatics**, v. 12, n. 1, p. 489, 2011.

NAKAYA, H.I. et al. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. **Genome biology**, v. 8, n. 3, p. R43, 2007.

NASIF, S.; CONTU, L.; MÜHLEMANN, O. Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. **Seminars in Cell & Developmental Biology**, v. 75, p. 78-87, 2018.

OH, Y. et al. Zika virus directly infects peripheral neurons and induces cell death. **Nature neuroscience**, v. 20, n. 9, p. 1209, 2017.

OZSOLAK, F.; MILOS, P.M. RNA sequencing: advances, challenges and opportunities. **Nature reviews genetics**, v. 12, n. 2, p. 87, 2011.

PAN AMERICAN HEALTH ORGANIZATION/WORLD HEALTH ORGANIZATION. Zika suspected and confirmed cases reported by countries and territories in the Americas Cumulative cases, 2015-2017. Washington, D.C.: Updated as of 04 January 2018.

PALUDAN, S. R. et al. Effect of IL-4 and IL-13 on IFN-γ-induced production of nitric oxide in mouse macrophages infected with herpes simplex virus type 2. **FEBS letters**, v. 414, n. 1, p. 61-64, 1997.

PERERA, M. et al. Flavivirus entry receptors: an update. **Viruses**, v. 6, n. 1, p. 69-88, 2014.

PERSAUD, Mirjana et al. Infection by Zika viruses requires the transmembrane protein AXL, endocytosis and low pH. **Virology**, v. 518, p. 301-312, 2018.

PHAM, N.V. et al. Dissociation of IGF2 and H19 imprinting in human brain. **Brain research**, v. 810, n. 1-2, p. 1-8, 1998.

PHILLIPS, K. A. et al. Why primate models matter. **American journal of primatology**, v. 76, n. 9, p. 801-827, 2014.

PINTO, A. C. et al. Application of RNA-seq to reveal the transcript profile in bacteria. **Genetics and Molecular Research**, v. 10, n. 103, p. 1707–1718, 2011.

PIRHONEN, J.; MATIKAINEN, S.; JULKUNEN, I. Regulation of Virus-Induced IL-12 and IL-23 Expression in Human Macrophages. **The Journal of Immunology**, v. 169, n. 10, p. 5673-5678, 2002.

PLATANIAS, L.C. Mechanisms of type-I-and type-II-interferon-mediated signalling. **Nature Reviews Immunology**, v. 5, n. 5, p. 375, 2005.

QIAN, C. et al. Identification of differentially expressed profiles of IncRNAs and mRNAs in ER-negative and HER-2 positive breast cancer. **Archives of Medical Science-Civilization Diseases**, v. 2, n. 1, p. 148-160, 2017.

RAINEN, L. et al. Stabilization of mRNA expression in whole blood samples. **Clinical chemistry**, v. 48, n. 11, p. 1883-1890, 2002.

RAO, S. et al. The RNA surveillance proteins UPF1, UPF2 and SMG6 affect HIV-1 reactivation at a post-transcriptional level. **Retrovirology**, v. 15, n. 1, 2018.

RATTERREE, M. S. et al. Experimental infection of rhesus macaques with West Nile virus: level and duration of viremia and kinetics of the antibody response after infection. **The Journal of infectious diseases**, v. 189, n. 4, p. 669-676, 2004.

RINN, J. L.; CHANG, H. Y. Genome Regulation by Long Noncoding RNAs. **Annual Review of Biochemistry**, v. 81, n. 1, p. 145–166, 7 jul. 2012.

ROLFE, A.J. et al. Bioinformatic analysis reveals the expression of unique transcriptomic signatures in Zika virus infected human neural stem cells. **Cell & bioscience**, v. 6, n. 1, p. 42, 2016.

RUSSO, P.S.T. et al. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. **BMC bioinformatics**, v. 19, n. 1, p. 56, 2018.

SAEINASAB, M. et al. SNHG15 is a bifunctional MYC-regulated noncoding locus encoding a IncRNA that promotes cell proliferation, invasion and drug resistance in colorectal cancer by interacting with AIF. Journal of Experimental & Clinical Cancer Research, v. 38, n. 1, p. 172, 2019.

SANTÉ PUBLIQUE FRANCE. Bulletin hebdomadaire international du 5 au 11 mars 2014. n. 442, , 2014. Disponível em: https://goo.gl/x558Fq.

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.

SCHÄPPI, M.G. et al. Hyperinflammation in chronic granulomatous disease and anti-inflammatory role of the phagocyte NADPH oxidase. In: **Seminars in immunopathology**. Springer-Verlag, 2008. p. 255-271.

SCHNEIDER, W. M.; CHEVILLOTTE, M. D.; RICE, C.M. Interferon-stimulated genes: a complex web of host defenses. **Annual review of immunology**, v. 32, p. 513-545, 2014.

SCHROEDER JR, H. W.; CAVACINI, L. Structure and function of immunoglobulins. **Journal of Allergy and Clinical Immunology**, v. 125, n. 2, p. S41-S52, 2010.

SHAO, Q. et al. Zika virus infection disrupts neurovascular development and results in postnatal microcephaly with brain damage. **Development**, v. 143, n. 22, p. 4127-4136, 2016.

SIGNAL, B.; GLOSS, B.; DINGER, M. Computational Approaches for Functional Prediction and Characterisation of Long Noncoding RNAs. **Trends in Genetics**, v. 32, n. 10, p. 620-637, 2016.

SILVEIRA, E. L. V. et al. Immune cell dynamics in rhesus macaques infected with a Brazilian strain of Zika virus. **The Journal of Immunology**, v. 199, n. 3, p. 1003-1011, 2017.

SIROHI, D. et al. The 3.8 Å resolution cryo-EM structure of Zika virus. **Science**, v. 352, n. 6284, p. 467-470, 2016.

SLAVOV, S. N. et al. Overview of Zika virus (ZIKV) infection in regards to the Brazilian epidemic. **Brazilian Journal of Medical and Biological Research**, v. 49, n. 5, 2016.

SLOMNICKI, L.P. et al. Ribosomal stress and Tp53-mediated neuronal apoptosis in response to capsid protein of the Zika virus. **Scientific reports**, v. 7, n. 1, p. 16652, 2017.

SOORYANARAIN, H. et al. ISG15 Modulates type I interferon signaling and the antiviral response during hepatitis E virus replication. **Journal of virology**, v. 91, n. 19, p. e00621-17, 2017.

SPRINGER, M. S. et al. Placental mammal diversification and the Cretaceous–Tertiary boundary. **Proceedings of the National Academy of Sciences**, v. 100, n. 3, p. 1056-1061, 2003.

SUN, S. et al. Jpx RNA activates Xist by evicting CTCF. **Cell**, v. 153, n. 7, p. 1537-1551, 2013.

SUZUKI, T. et al. Infection with flaviviruses requires BCLXL for cell survival. **PLoS pathogens**, v. 14, n. 9, p. e1007299, 2018.

TAKEUCHI, O.; AKIRA, S. MDA5/RIG-I and virus recognition. **Current opinion in immunology**, v. 20, n. 1, p. 17-22, 2008.

TAO, R. et al. Porcine Reproductive and Respiratory Syndrome Virus Nonstructural Protein 4 Cleaves Porcine DCP1a To Attenuate Its Antiviral Activity. **The Journal of Immunology**, v. 201, n. 8, p. 2345-2353, 2018.

THONGTAN, T.; PANYIM, S.; SMITH, D.R. Apoptosis in dengue virus infected liver cell lines HepG2 and Hep3B. **Journal of medical virology**, v. 72, n. 3, p. 436-444, 2004.

TIWARI, S. K. et al. Zika virus infection reprograms global transcription of host cells to allow sustained infection. **Emerging microbes & infections**, v. 6, n. 4, p. e24, 26 abr. 2017.

TRAPNELL, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature protocols**, v. 7, n. 3, p. 562, 2012.

TRUS, I. et al. Persistent Zika virus infection in porcine conceptuses is associated with elevated in utero cortisol levels. **Virulence**, v. 9, n. 1, p. 1338-1343, 2018.

ULITSKY, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nature Reviews Genetics, v. 17, n. 10, p. 601-614, 2016.

VAN DEN BROEKE, C.; JACOB, T.; FAVOREEL, H. W. Rho'ing in and out of cells: viral interactions with Rho GTPase signaling. **Small GTPases**, v. 5, n. 1, p. e28318, 2014.

WALDORF, K.A. et al. Fetal brain lesions after subcutaneous inoculation of Zika virus in a pregnant nonhuman primate. **Nature medicine**, v. 22, n. 11, p. 1256, 2016.

WANG, K.C. et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. **Nature**, v. 472, n. 7341, p. 120, 2011.

WANG, L. et al. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. **Nucleic acids research**, v. 41, n. 6, p. e74-e74, 2013.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews genetics**, 2009.

WASHIETL, S.; KELLIS, M.; GARBER, M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. **Genome research**, v. 24, n. 4, p. 616–28, 2014.

WEBB, Louise MC et al. GIMAP1 is essential for the survival of naive and activated B cells in vivo. **The Journal of Immunology**, v. 196, n. 1, p. 207-216, 2016.

WILUSZ, J. E. et al. Long noncoding RNAs: functional surprises from the RNA world. **Genes & development**, v. 23, n. 13, p. 1494-1504, 2009.

WRIGHT, S. P. Adjusted p-values for simultaneous inference. **Biometrics**, v. 48, n. 4, p. 1005-1013, 1992.

WUTZ, Anton; RASMUSSEN, Theodore P.; JAENISCH, Rudolf. Chromosomal silencing and localization are mediated by different domains of Xist RNA. **Nature genetics**, v. 30, n. 2, p. 167, 2002.

XU, L. et al. MicroRNA-21 (miR-21) regulates cellular proliferation, invasion, migration, and apoptosis by targeting PTEN, RECK and Bcl-2 in lung squamous carcinoma, Gejiu City, China. **PloS one**, v. 9, n. 8, p. e103698, 2014.

YOON, J. et al. LincRNA-p21 suppresses target mRNA translation. **Molecular cell**, v. 47, n. 4, p. 648-655, 2012.

YU, G. & HE, Q. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. **Molecular BioSystems**, v. 12, n. 2, p. 477-479, 2016.

ZANLUCA, C. et al. First report of autochthonous transmission of Zika virus in Brazil. **Memórias do Instituto Oswaldo Cruz**, v. 110, n. 4, p. 569–572, 2015.

ZEIRA, E. et al. The knockdown of H19 IncRNA reveals its regulatory role in pluripotency and tumorigenesis of human embryonic carcinoma cells. **Oncotarget**, v. 6, n. 33, p. 34691, 2015.

ZIMIN, A. V. et al. A new rhesus macaque assembly and annotation for next-generation sequencing analyses. **Biology direct**, v. 9, n. 1, p. 20, 2014.

ZHAO, J. et al. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. **Science**, v. 322, n. 5902, p. 750-756, 2008.

ZHANG, C. et al. Cullin3–KLHL25 ubiquitin ligase targets ACLY for degradation to inhibit lipid synthesis and tumor progression. **Genes & development**, v. 30, n. 17, p. 1956-1970, 2016.

ZHANG, F. et al. Molecular signatures associated with ZIKV exposure in human cortical neural progenitors. **Nucleic Acids Research**, v. 44, n. 18, p. 8610–8620, 2016.

ZHANG, Q. et al. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. **mBio**, v. 4, n. 1, 2013.

ZHANG, R.; XIA, T. Long non-coding RNA XIST regulates PDCD4 expression by interacting with miR-21-5p and inhibits osteosarcoma cell growth and metastasis. **International journal of oncology**, v. 51, n. 5, p. 1460-1470, 2017.
Apêndice I - Pacotes da linguagem R e ferramentas utilizados

Nome	Versão	Etapa
Bowtie	1.2.0	Alinhamento
CEMiTool	1.8.0	Co-expressão
CPAT	1.2.3	Reanotação
Cufflinks	2.2.1	Alinhamento
Deseq2	1.23.0	DEGs
FastQC	0.11.15	Qualidade
Inkscape	0.16	Figuras
ReactomePA	1.28.0	Enriquecimento
RPISeq	1.0	Predição proteína-IncRNA
TopHat2	2.1.1	Alinhamento

Apêndice II - Publicação

Cardozo, L.E., Russo, P.S., Gomes-Correia,B., **Araujo-Pereira, M**., Sepúlveda-Hermosilla, G., Maracaja-Coutinho, V., & Nakaya, H.I. (2019). webCEMiTool: Co-expression modular analysis made easy. *Frontiers in genetics*, *10*.

