UNIVERSIDADE DE SÃO PAULO

Faculdade de Ciências Farmacêuticas

Programa de Pós-Graduação em Farmácia

Área de Fisiopatologia

Biologia de Sistemas de RNAs Não-Codificadores Longos na Vacinação

Diógenes Saulo de Lima

Tese apresentada para obtenção do título de Doutor em Ciências

Orientador: Prof. Dr. Helder Takashi Imoto Nakaya

São Paulo

2020

UNIVERSIDADE DE SÃO PAULO

Faculdade de Ciências Farmacêuticas

Programa de Pós-Graduação em Farmácia

Área de Fisiopatologia

Biologia de Sistemas de RNAs Não-Codificadores Longos na Vacinação

Diógenes Saulo de Lima

Versão Original

Tese apresentada para obtenção do título de Doutor em Ciências

Orientador: Prof. Dr. Helder Takashi Imoto Nakaya

São Paulo

2020

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 eletronicamente pelo autor, utilizando o programa desenvolvido pela Seção Técnica de Informática do ICMC/USP e adaptado para a Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP

Bibliotecária responsável pela orientação de catalogação da publicação:
Marlene Aparecida Vieira - CRB - 8/5562

L732b

Lima, Diógenes Saulo de
Biologia de Sistemas de RNAs Não-Codificadores Longos na Vacinação / Diógenes Saulo de Lima. - São Paulo, 2020.
83 p.

Tese (doutorado) - Faculdade de Ciências Farmacêuticas da Universidade de São Paulo. Departamento de Análises Clínicas e Toxicológicas.
Orientador: Nakaya, Helder Takashi Imoto

1. lncRNAs. 2. imunologia. 3. transcritoma. 4. metanálise. 5. bioinformática. I. T. II. Nakaya, Helder Takashi Imoto, orientador.

Diógenes Saulo de Lima

Biologia de Sistemas de RNAs Não-Codificadores Longos na Vacinação

Comissão Julgadora da

Tese para obtenção do Título de Doutor

Prof. Dr. Helder Takashi Imoto Nakaya

Orientador/presidente

_____

1º examinador

_____

2º examinador

_____

3º examinador

_____

4º examinador

São Paulo, _____ de _____ de 2020.

Dedico este trabalho aos meus pais: Paulo e Terezinha.

AGRADECIMENTOS

Agradeço à minha família. Sem vocês, absolutamente nada disso seria possível. Agradeço aos meus irmãos: Paola, Pauline e Felipe, e principalmente aos meus pais: Paulo e Terezinha. Obrigado por terem forjado persistentemente as condições necessárias para que eu pudesse chegar até aqui. Agradeço também às minhas tias e tios, primos e primas, amigos e amigas. Recebi de vocês muito mais do que um dia poderei retribuir.

Agradeço ao Helder pela condução fenomenal de minha orientação. Agradeço pela pressão em alguns momentos da pós-graduação, e a compreensão em outros. Agradeço pelo tempo que você investiu em mim. Agradeço por ter me ensinado mais do que a técnica e o rigor científico, mas também por ter compartilhado comigo a sua visão científica.

Agradeço a todos os membros do CSBL, principalmente aos da primeira safra. Conhecê-los e compartilhar ideias com vocês me tornou não apenas mais capaz tecnicamente, mas também uma pessoa melhor.

Agradeço aos funcionários e professores da Faculdade de Ciências Farmacêuticas da USP. Desde 2010, no início da minha graduação, tive a oportunidade de ter acesso a um ensino de primeira classe. Se isso não fosse suficiente, ainda pude participar de projetos de extensão, estágios e iniciações científicas. Tudo isso só foi possível porque a FCF, o IQ, o ICB, o IME, o HU, a POLI e os outros institutos da USP estão cheios de pessoas sérias, capazes, esforçadas e com a noção da importância do papel que ocupam na construção de uma realidade menos bruta. Toda a minha formação acadêmica teve base nessas instituições fenomenais. Aqui não tem balbúrdia.

Agradeço a todos os professores que me ensinaram e instigaram durante minha infância e adolescência. Isso inclui os professores do SENAI, do instituto Pérola Byington e do Cursinho da Poli.

Agradeço a todos os colaboradores deste projeto. Agradeço especialmente ao professor Eduardo Lani e à Vívian pelo auxílio na interpretação dos resultados deste trabalho. Agradeço ainda aos pesquisadores que permitiram que eu colaborasse em seus projetos.

Agradeço à Raíssa. Sempre me lembrarei de nossas viagens, passeios e

conversas com muito carinho.

Agradeço à cNPQ, FAPESP e IUIS pelo apoio financeiro e formação complementar. Agradeço ao povo brasileiro, e espero sinceramente ter honrado o investimento. Sei que recebi muito mais do que poderei um dia retribuir.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Charles Monroe Schulz

RESUMO

LIMA, D. S. **Biologia de Sistemas de RNAs Não-Codificadores Longos na Vacinação**. 2020. 83f. Tese (Doutorado) – Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2020.

Entender os mecanismos responsáveis pela proteção induzida por vacinas contribui para o desenvolvimento de novas vacinas. Uma abordagem de pesquisa denominada Vacinologia de Sistemas surgiu para endereçar essa tarefa. A aplicação da Vacinologia de Sistemas gerou informações amplas relacionadas a respostas vacinais e foi aplicada no estudo de diversas vacinas. Apesar de estarem envolvidos em diversos processos imunológicos, RNAs Não-Codificadores Longos (lncRNAs) ainda não foram estudados no contexto da imunidade induzida por vacinas. Neste trabalho, fizemos a análise de mais de 2.000 amostras de transcritoma de sangue periférico, oriundas de 17 diferentes coortes vacinadas, com foco na identificação de lncRNAs potencialmente envolvidos com a resposta induzida por vacinas contra gripe e contra febre amarela. Criamos também um banco de dados online, em que todos os nossos resultados podem ser facilmente acessados. Nossos resultados indicaram que diversos lncRNAs participam de múltiplas vias imunológicas relacionadas a respostas induzidas por vacinas. Entre esses, o transcrito *FAM30A* se destaca por ter alta expressão em células B e ser correlacionado com a expressão de genes de imunoglobulina localizados no mesmo *locus* genômico. Identificamos também alterações na expressão de lncRNAs em dados de RNA-*seq* de uma coorte de crianças imunizadas com uma vacina atenuada contra gripe, o que sugere um papel de lncRNAs na resposta a diferentes vacinas. Nossos achados trazem evidências de que lncRNAs tem um papel significativo na resposta imune induzida por vacinas.

Palavras-chave: transcritoma, imunologia, lncRNAs, metanálise, coexpressão, bioinformática

ABSTRACT

LIMA, D. S. **Systems Biology of Long Non-Coding RNAs in Vaccination**. 2020. 83f. Doctoral Thesis – School of Pharmaceutical Sciences, University of São Paulo, São Paulo, 2020.

Understanding the mechanisms of vaccine-elicited protection contributes to the development of new vaccines. The emerging field of Systems Vaccinology provides detailed information on host responses to vaccination and has been successfully applied in the study of the molecular mechanisms of several vaccines. Long Non-Coding RNAs (lncRNAs) are crucially involved in multiple biological processes, but their role in vaccine-induced immunity has not been explored. We performed an analysis of over 2,000 blood transcriptome samples from 17 vaccine cohorts to identify lncRNAs potentially involved with antibody responses to influenza and yellow fever vaccines. We have created an online database where all results from these analyses can be easily accessed. We found that lncRNAs participate in distinct immunological pathways related to vaccine-elicited responses. Among them, we showed that the expression of lncRNA *FAM30A* was high in B cells and correlates with the expression of immunoglobulin genes located in its genomic vicinity. We also identified altered expression of lncRNAs in RNA-sequencing (RNA-seq) data from a cohort of children vaccinated with intranasal live attenuated influenza vaccine, suggesting a common role across several diverse vaccines. Taken together, these findings provide evidence that lncRNAs have a significant impact on immune responses induced by vaccination.

Keywords: transcriptomics, immunology, lncRNAs, meta-analysys, co-expression, bioinformatics

LISTA DE FIGURAS

Figura 1 - Exemplos de classes de lncRNAs.....23
Figura 2 - Representação esquemática da metodologia. 31
Figura 3 - Número de amostras analisadas por estudo.....40
Figura 4 - Classes de lncRNAs em plataformas reanotadas. 41
Figura 5 – Exemplos de sondas mal anotadas na plataforma GPL570.....41
Figura 6 – Intersecção de lncRNAs representados nas plataformas.....42
Figura 7 - Intersecção entre genes diferencialmente expressos em coortes imunizadas com IIVs (A) e YF-17D (B). 43
Figura 8 - Intersecção entre lncRNAs diferencialmente expressos em coortes imunizadas com IIVs (A) e YF-17D (B). 44
Figura 9 – *Volcano plots* das metanálises de expressão diferencial de lncRNAs nas coortes imunizadas com IIVs.....45
Figura 10 - *Volcano plots* das metanálises de expressão diferencial de lncRNAs nas coortes imunizadas com YF-17D. 45
Figura 11 – Alterações na expressão de lncRNAs após a vacinação. 46
Figura 12 – Metanálise de lncRNAs representativos com expressão diferencial reiterada em coortes imunizadas com IIVs.....47
Figura 13 – Metanálise de lncRNAs representativos com expressão diferencial reiterada em coortes imunizadas com YF-17D. 48
Figura 14 - *Volcano plots* das metanálises de correlação de lncRNAs com títulos de anticorpos nas coortes imunizadas com IIVs. 49
Figura 15 - *Forest plots* de lncRNAs correlacionados com títulos de anticorpos em coortes imunizadas com IIVs.....49
Figura 16 - *Volcano plots* das metanálises de correlação de lncRNAs com mRNAs próximos nas coortes imunizadas com IIVs. 50
Figura 17 - Visão geral da rede consenso de coexpressão de IIVs. 52
Figura 18 - Visão geral da rede consenso de coexpressão de YF-17D. 54
Figura 19 - Expressão de lncRNAs da comunidade CM5 das coortes vacinadas com IIVs em linhagens de células imunes. 55
Figura 20 – Expressão de *GAS6-AS1* e *GAS6* em linhagens imunes.....56

Figura 21 – Expressão de *FAM30A* correlaciona-se com a produção de anticorpos e com segmentos gênicos no *locus* IGH. 57
Figura 22 – *Volcano plot* dos resultados de RNA-*seq* de coorte de crianças vacinadas. 58
.....

LISTA DE ABREVIATURAS E SIGLAS

APC *Antigen-Presenting Cell*

API *Application Programming Interface*

BTM *Blood Transcriptional Modules*

CD *Cluster of Differentiation*

CEMiTool *Coexpression Module Identification Tool*

DEG *Differentially Expressed Gene*

ENCODE *Encyclopedia of DNA Elements*

EST *Expressed Sequence Tags*

FPKM *Fragments Per Kilobase of transcript per Million mapped reads*

GSEA *Gene Set Enrichment Analysis*

IFN Interferon

IIV *Inactivated Influenza Vaccine*

LAIV *Live Attenuated Influenza Vaccine*

lncRNA *Long Non-Coding RNA*

logFC Logaritmo do *fold change* na base 2

NES *Normalized Enrichment Score*

PBMC *Peripheral Blood Mononuclear Cells*

RNA-*seq* *RNA sequencing*

SNP *Single Nucleotide Polymorphism*

TLR *Toll-like Receptor*

WGCNA *Weighted Gene Co-Expression Network Analysis*

YF-17D *Yellow Fever 17D Vaccine*

SUMÁRIO

1 INTRODUÇÃO 17
1.1 BREVE HISTÓRIA DA VACINAÇÃO 17
1.2 BREVE INTRODUÇÃO À IMUNOLOGIA DE VACINAS 18
1.3 BIOLOGIA DE SISTEMAS E CIÊNCIAS ÔMICAS 19
1.4 VACINOLOGIA DE SISTEMAS 20
1.5 RNAS NÃO-CODIFICADORES LONGOS E O SISTEMA IMUNE 22
1.6 REANOTAÇÃO DE PLATAFORMAS DE *MICROARRAYS* 24
1.7 METANÁLISES COM DADOS DE EXPRESSÃO GÊNICA 26
1.8 REDES BIOLÓGICAS E REDES DE COEXPRESSÃO 27
1.9 JUSTIFICATIVA 29
2 OBJETIVOS 29
2.1 OBJETIVO PRINCIPAL 29
2.2 OBJETIVOS ESPECÍFICOS 30
3 MATERIAIS E MÉTODOS 30
3.1 SELEÇÃO, OBTENÇÃO E CURADORIA DOS DADOS 31
3.1.1 Dados públicos de expressão gênica 31
3.1.2 Dados de expressão gênica de células do sistema imune 32
3.1.3 RNA-*seq* de coorte imunizada com LAIV 33
3.2 REANOTAÇÃO DAS PLATAFORMAS DE *MICROARRAYS* 33
3.3 CONTROLE DE QUALIDADE E NORMALIZAÇÃO DE DADOS DE EXPRESSÃO 34
3.4 ANÁLISES 34
3.4.1 Detecção de genes diferencialmente expressos 35
3.4.2 Detecção de genes correlacionados a títulos de anticorpos 35
3.4.3 Detecção de pares genes-lncRNAs correlacionados em *cis* 35
3.4.4 Detecção de módulos de coexpressão 36
3.4.5 Estratégias de Metanálise 37
3.4.6 Enriquecimento funcional 37
4 RESULTADOS 38
4.1 DADOS DE EXPRESSÃO E REANOTAÇÃO DE PLATAFORMAS 38
4.2 EXPRESSÃO DIFERENCIAL DE GENES E LNCRNAS APÓS A VACINAÇÃO 43

4.3 CORRELAÇÃO DA EXPRESSÃO DE LNCRNAS E TÍTULOS DE ANTICORPOS 48
4.4 CORRELAÇÃO ENTRE LNCRNAS E MRNAS NA MESMA REGIÃO GENÔMICA 50
4.5 DETECÇÃO DE REDES CONSENSO DE COEXPRESSÃO 50
4.6 EXPRESSÃO DE LNCRNAS EM CÉLULAS IMUNES.....54
4.7 COMPARAÇÃO DOS RESULTADOS E *FAM30A* 56
4.8 MUDANÇAS DE EXPRESSÃO DE LNCRNAS APÓS LAIV 57
5 DISCUSSÃO 58
6 CONCLUSÕES 63
REFERÊNCIAS.....63
ANEXOS 79

1 INTRODUÇÃO

1.1 BREVE HISTÓRIA DA VACINAÇÃO

A vacinação em grande escala foi uma das intervenções mais importantes para a saúde mundial, tendo sido responsável pela erradicação da varíola e por uma grande diminuição das mortes causadas por diversas doenças infecciosas, tais como a poliomielite, o sarampo e a coqueluche (KOFF; GUST; PLOTKIN, 2014; RAPPUOLI et al., 2014a). O princípio de funcionamento das vacinas é análogo ao processo de aquisição de imunidade experimentado por indivíduos acometidos por certas doenças infecciosas. Essa constatação ancestral motivou a realização de práticas similares à vacinação na China, Índia e África bem antes do século 18 (RAPPUOLI et al., 2014b; TOGNOTTI, 2010). Esses procedimentos consistiam na exposição intencional ao conteúdo de pústulas oriundas de outras pessoas acometidas pela varíola. Apesar de um risco considerável de desenvolvimento de uma infecção grave, a variolação trazia um risco de morte menor do que a infecção natural (RIEDEL, 2005).

No final do século 18, período em que a variolação era praticada na Inglaterra, Edward Jenner notou que a infecção por varíola bovina era capaz de prover imunidade contra a varíola humana. Ao inocular pessoas com material proveniente de pústulas de vacas infectadas, Jenner observou que estes indivíduos adquiriam imunidade a um risco muito menor do que se submetidos à variolação. Esse procedimento deu origem à vacinação como conhecemos hoje. Não por acaso, a palavra “vacina” é derivada da palavra latina *vacca*. A aplicação em larga escala da vacina contra varíola foi responsável por sua erradicação no final da década de 1970 (RIEDEL, 2005).

Com o passar das décadas, vacinas contra doenças como febre tifoide, cólera, raiva, peste, coqueluche, febre amarela, gripe e tétano foram desenvolvidas. Essas primeiras vacinas consistiam em preparações contendo patógenos inativados ou atenuados. A vacina contra a febre amarela, por exemplo, contém a cepa 17D do vírus, e é usada ainda hoje com sucesso (PLOTKIN; PLOTKIN, 2018). O desenvolvimento de novas tecnologias permitiu a criação de vacinas para diversas outras doenças. Essas tecnologias envolveram avanços na cultura e inativação de patógenos, conjugação de antígenos, produção de antígenos a partir da sequência genética de patógenos e utilização de adjuvantes (DE GREGORIO; RAPPUOLI,

2014). Entretanto, mesmo com todos os avanços, o paradigma de desenvolvimento de vacinas é primariamente empírico. Isso significa que as novas formulações são desenvolvidas e testadas sem o entendimento completo dos mecanismos moleculares induzidos por estas no organismo receptor. Em outras palavras, os avanços na imunologia ocorreram de forma desacoplada aos avanços da vacinologia. Enquanto a primeira tinha foco no estudo de modelos que permitiam o entendimento em nível celular e molecular, a segunda tinha foco na questão prática da indução de proteção em organismos (D'ARGENIO; WILSON, 2010).

O paradigma empírico de desenvolvimento de vacinas foi muito eficiente na criação de vacinas contra várias doenças, mas ineficiente na criação de vacinas contra diversas outras (LEVITZ; GOLENBOCK, 2012; NAKAYA; PULENDRAN, 2015; POLAND et al., 2013). A criação de novas vacinas esbarra em múltiplos fatores, sendo um destes a falta de conhecimento dos fatores determinantes de respostas imunológicas bem-sucedidas (PULENDRAN; AHMED, 2011). O estudo desses mecanismos pode não apenas auxiliar na compreensão do sistema imune, mas também informar o desenvolvimento de novas vacinas (BHATT et al., 2015; HAGAN et al., 2015).

1.2 BREVE INTRODUÇÃO À IMUNOLOGIA DE VACINAS

As respostas imunológicas induzidas por vacinas são mediadas pela atividade concertada de múltiplas células, moléculas e processos do sistema imune inato e adaptativo. De maneira geral: após a inoculação, os antígenos e adjuvantes presentes nas vacinas são detectados por receptores de células do sistema imune inato, que respondem produzindo citocinas e quimiocinas pró-inflamatórias. Uma vez instaurado um microambiente inflamatório, suportado pelo recrutamento adicional de células do sistema imune inato, monócitos diferenciam-se em macrófagos e células dendríticas são ativadas. Essas células dendríticas migram para linfonodos por meio de vasos linfáticos (SIEGRIST, 2013).

No linfonodo, células T ativadas por células dendríticas interagem com células B que entraram em contato com antígenos livres. Essas células B são ativadas e podem então se diferenciar em plasmócitos de vida curta, ou iniciar uma reação de centro germinativo (PULENDRAN; AHMED, 2006). Durante a reação de centro germinativo, células B sofrem expansão clonal, hipermutação somática, seleção de

clones com maior afinidade a antígenos e mudança de classe de imunoglobulinas. Como resultado desse processo, que envolve a atuação de citocinas, células T foliculares e células dendríticas foliculares, células B maduras deixam os centros germinativos como plasmócitos de vida longa e células B de memória (MCHEYZER-WILLIAMS; MCHEYZER-WILLIAMS, 2005). Por fim, a vacinação bem-sucedida é capaz de induzir proteção de longo prazo através de diversos mecanismos efetores. Esses incluem a atividade de linfócitos T $CD4^+$ e $CD8^+$, e principalmente de anticorpos produzidos por células B (SIEGRIST, 2013).

Vacinas diminuem taxas de incidência de doenças em populações vacinadas, mas não induzem proteção em todas as pessoas vacinadas (ORENSTEIN et al., 1985). A incapacidade de um individuo em gerar uma resposta imune efetiva pode ser atribuída a múltiplos fatores. Por exemplo, polimorfismos em genes codificadores de receptores celulares (*SLAM* e *CD46*), citocinas (*IL4*, *IL12B*, *IL10*) e genes do antígeno leucocitário humano (*HLA*) estão associados a variações nas respostas induzidas por vacinas (POLAND et al., 2007). Esses e muitos outros genes participam dos diversos processos mencionados nos parágrafos anteriores. Outros fatores não genéticos que interferem na resposta vacinal incluem a idade, doenças crônicas, microbiota, dieta, sexo, estresse e etnia (HAGAN et al., 2015).

1.3 BIOLOGIA DE SISTEMAS E CIÊNCIAS ÔMICAS

Organismos são sistemas complexos compostos por muitas partes interdependentes cuja homeostase é subordinada a múltiplos processos estocásticos (MENG; SOMANI; DHAR, 2004). O controle desses processos é exercido por várias alças de retroalimentação, interruptores moleculares, sensores moleculares, subsistemas redundantes, subsistemas de amplificação de sinais, acoplamento de diferentes subsistemas, etc. Em outras palavras, os comportamentos dos sistemas biológicos são determinados por múltiplas relações de causa e efeito, organizadas em complexas redes biológicas (SAUER; HEINEMANN; ZAMBONI, 2007). Achados da pesquisa biológica originaram-se, em sua maioria, pelo estudo isolado das partes e processos desses sistemas. Essa abordagem de pesquisa é conhecida como reducionismo metodológico. Apesar de poderosa, a estratégia de construção incremental de esquemas conectando os componentes dos sistemas biológicos não permite o entendimento completo das

propriedades que emergem destes (KITANO, 2002).

A necessidade de se obter um entendimento compreensivo dos sistemas biológicos em estado de homeostase e perturbação fomentou o desenvolvimento de um paradigma científico denominado Biologia de Sistemas. A Biologia de Sistemas é o campo científico que estuda como emergem as propriedades de sistemas biológicos a partir da interação de múltiplos componentes. Essa abordagem de pesquisa foi também catalisada pelos crescentes avanços tecnológicos em nossa capacidade de identificar/quantificar diversas classes de moléculas biológicas (AHN et al., 2006). O alargamento de escopo e resolução em nossa capacidade de interrogação de fenótipos levou à inauguração de novos campos de pesquisa, cujos nomes são compostos com o sufixo “-ômica”. A identificação e/ou quantificação compreensiva dos genes, metabólitos, proteínas e transcritos de um organismo, por exemplo, referem-se à genômica, metabolômica, proteômica e transcritômica, respectivamente (VAILATI-RIBONI; PALOMBO; LOOR, 2017).

1.4 VACINOLOGIA DE SISTEMAS

A resposta imunológica pode ser entendida como o resultado das interações (arestas) de múltiplos componentes biológicos (nós) (POLAND et al., 2013). A maior parte das descobertas relacionadas ao sistema imune adveio do exame diligente das partes e processos que o compõem. Apesar de poderosa, essa abordagem é limitada, já que não considera o conjunto dos componentes funcionais do sistema imune em um contexto amplo de perturbação, como observado em infecções, doenças autoimunes e após a vacinação (DAVIS; TATO; FURMAN, 2017). Ademais, nem sempre os achados obtidos em modelos animais são traduzíveis para a interpretação da imunologia humana (PULENDRAN; AHMED, 2011). Nesse contexto, a vacinologia de sistemas surgiu como uma abordagem de análise holística que traz a promessa de aumentar nosso entendimento dos mecanismos moleculares responsáveis pela eficácia das vacinas. Essa estratégia consiste na aplicação de técnicas da biologia de sistemas para a criação de redes de interação e modelos computacionais que explicam as perturbações induzidas por vacinas em seres humanos (NAKAYA; PULENDRAN, 2015; POLAND et al., 2013).

Diferentes estudos de vacinologia de sistemas abordaram variados aspectos da resposta vacinal. Exemplos incluem a descoberta de assinaturas gênicas que

predizem a resposta de anticorpos e atividade de células T $CD8^+$ após vacinação contra a febre amarela (QUEREC et al., 2009) e gripe (NAKAYA et al., 2011). Abordagens sistêmicas também permitiram o estudo de fatores influenciadores da resposta vacinal, tais como o sexo (FURMAN et al., 2014), a idade (THAKAR et al., 2015) e infecções prévias (FURMAN et al., 2015). Análises do transcritoma e metaboloma revelaram que a atividade do ciclo de Krebs e da resposta de estresse do retículo endoplasmático predizem a suscetibilidade a efeitos adversos causados pela infecção com YF-17D (CHAN et al., 2019). Análises multi-ômicas permitiram o estabelecimento de conexões entre redes metabólicas e transcricionais que atuam em resposta à vacinação contra herpes (LI et al., 2017a). Outros exemplos interessantes incluem comparações entre a resposta induzida pela vacina inativada contra a gripe e pela vacina conjugada antipneumocócica (OBERMOSER et al., 2013), e comparações entre as respostas induzidas por cinco diferentes vacinas (LI et al., 2014). Esses dois trabalhos revelaram que as alterações transcricionais induzidas por diferentes vacinas em sangue periférico diferem qualitativa e quantitativamente.

A vacinologia de sistemas também permite a geração de novas hipóteses relacionadas aos mecanismos moleculares responsáveis pela resposta vacinal. A descoberta do papel regulatório do gene *CAMK4* na produção de anticorpos só foi possível após a observação da correlação negativa entre a expressão deste gene com mudanças em títulos de anticorpos induzidas pela vacina inativada contra gripe (NAKAYA et al., 2011). Outra descoberta notável foi motivada pela observação de que mudanças na expressão de *TLR5* nos dias 3-7 após a vacinação contra gripe correlacionavam-se com a magnitude da resposta humoral no dia 28. A proteína TLR5 é um receptor da flagelina bacteriana, e não tinha sido antes associada à resposta imune contra vírus (PULENDRAN, 2014). Esse achado motivou o estudo do papel da microbiota na resposta imune induzida por vacinas em camundongos: a imunização de camundongos *Tlr5*$^{-/-}$ com uma vacina inativada contra gripe indicou um papel previamente não apreciado da microbiota intestinal na modulação da resposta humoral contra vírus (OH et al., 2014). Mais recentemente, o estudo dos efeitos da microbiota intestinal humana na resposta vacinal foi feito por meio da administração de antibióticos, seguida da administração de uma vacina inativada contra gripe (HAGAN et al., 2019).

1.5 RNAS NÃO-CODIFICADORES LONGOS E O SISTEMA IMUNE

A proteção induzida por vacinas depende da ativação coordenada de processos relacionados ao sistema imune inato e adaptativo. As atividades das células de ambos os braços do sistema imune, por sua vez, dependem da integridade de subsistemas contendo receptores de membrana, cascatas de ativação, fatores de transcrição, citocinas, entre outros. Uma classe de moléculas com atividade regulatória, denominada RNAs Não-Codificadores Longos (lncRNAs), vem também sendo associada a múltiplos processos do sistema imune (ATIANAND; CAFFREY; FITZGERALD, 2017; ATIANAND; FITZGERALD, 2014; CHEN; SATPATHY; CHANG, 2017).

Os lncRNAs são caracterizados, de maneira geral, como transcritos com mais de 200 nucleotídeos que não codificam proteínas (MATTICK; RINN, 2015). Assim como os RNAs mensageiros (mRNAs), são comumente poliadenilados, submetidos a 5' *capping* e *splicing.* Estas características não são, porém, universais, já que alguns lncRNAs não são poliadenilados ou submetidos a 5' *capping* (AYUPE et al., 2015). Análises do projeto *Encyclopedia of DNA Elements* (ENCODE), revelaram que mais de 80% do genoma humano é transcrito em ao menos uma das 15 linhagens celulares humanas estudadas (DJEBALI et al., 2012). Uma grande fração desses RNAs é composta por transcritos anotados como lncRNAs. O interesse no estudo de lncRNAs é justificado pela descoberta de funções biológicas por eles exercidas. Processos que envolvem lncRNAs incluem o remodelamento da cromatina, o *imprinting* genômico e o processamento de mRNAs (KUNG; COLOGNORI; LEE, 2013). Os lncRNAs podem regular a expressão de genes próximos (*cis*) ou de genes localizados em regiões distantes (*trans*). O *XIST* é o exemplo clássico do primeiro caso. Esse transcrito é expresso no centro de inativação do cromossomo X, e intermedeia a inativação deste cromossomo inteiro (BROWN et al., 1991). Um exemplo do segundo caso é o *HOTAIR*, um lncRNA localizado no cromossomo 12 capaz de afetar a expressão do *locus HOXD* no cromossomo 2 (RINN et al., 2007). Uma estimativa recente aponta para a existência de ao menos 27,919 lncRNAs, dos quais 19,176 apresentam evidencias de relevância funcional (HON et al., 2017). Considerando que apenas um pequeno número desses transcritos tiveram suas funções estudadas (YAN et al., 2017), a existência de potencial funcional a ser ainda investigado no amplo panorama

transcricional dos lncRNAs é indisputável.

Diversas características dos lncRNAs são empregadas em suas categorizações. De acordo com suas posições genômicas em relação a outros genes, os lncRNAs podem ser classificados em intergênicos, antisenso ou intrônicos. De acordo com suas posições genômicas em relação a regiões regulatórias, os lncRNAs podem ser classificados como *enhancer associated* e *promoter associated* (Figura 1). De acordo com suas sublocalizações celulares, os lncRNAs são caracterizados como nucleares e citoplasmáticos. Os lncRNAs podem ser categorizados de diversas outras maneiras, não obstante, com o incremento no conhecimento associado a esses transcritos, novas categorias de lncRNAs podem ser ainda propostas (ST LAURENT; WAHLESTEDT; KAPRANOV, 2015).

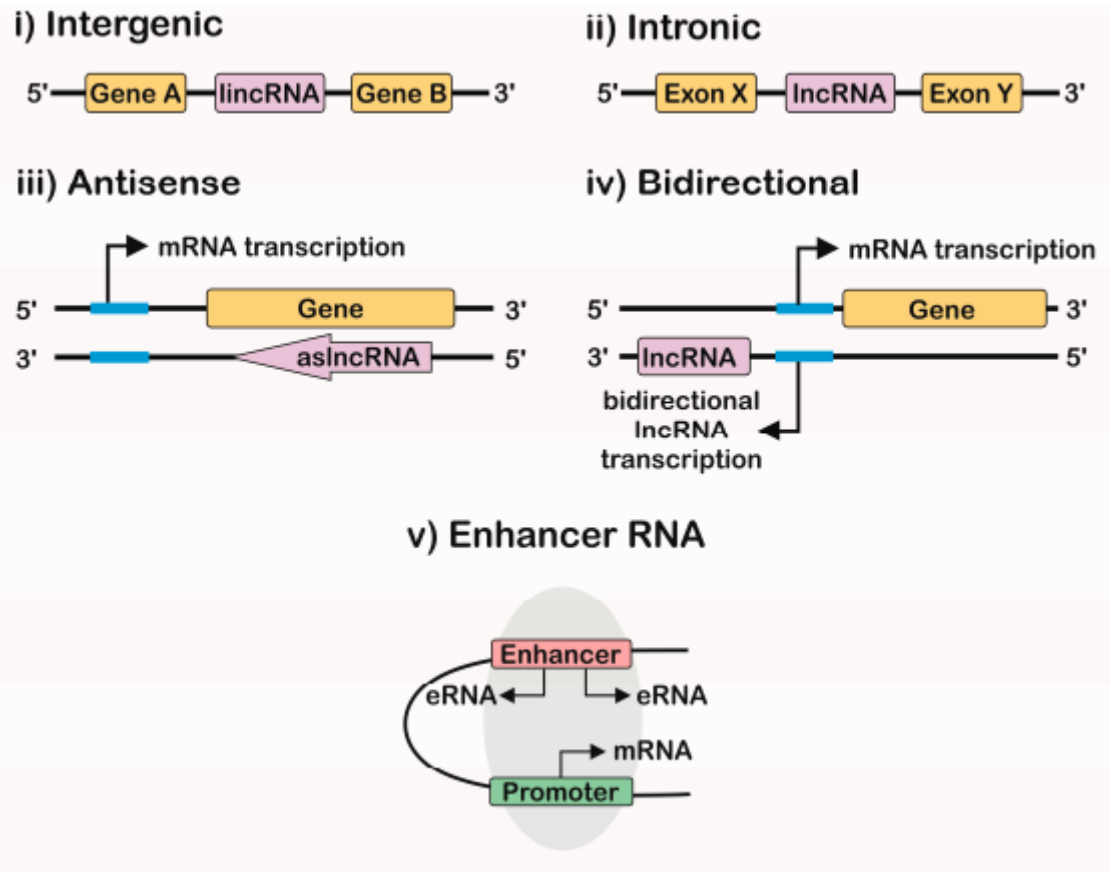

Figura 1 - Exemplos de classes de lncRNAs (Adaptado de (FERNANDES et al., 2019))

Diversos lncRNAs foram associados a processos do sistema imune. O lnc-DC, por exemplo, foi implicado na regulação da diferenciação e atividade de células dendríticas por meio da inibição da desfosforilação de STAT3 (WANG et al., 2014).

Outros lncRNAs foram associados à produção de citocinas (IMAMURA et al., 2014), ativação de células $CD8^+$ na tuberculose (WANG et al., 2015), modulação da expressão de *Foxp3* em células T regulatórias (ZEMMOUR et al., 2017) e diferenciação de células T_h2 (ZHANG et al., 2013).

Além dos lncRNAs associados funcionalmente ao sistema imune, a expressão de muitos outros foi detectada em linhagens de células linfoides e mieloides. Como exemplo, podemos citar a identificação de 4516 lncRNAs expressos em diversos estágios de desenvolvimento e ativação de células B de camundongos (BRAZÃO et al., 2016). Um trabalho mais recente, feito com sequenciamento de RNAs (RNA-*seq*), indicou que 30% do genoma humano é transcrito em ao menos algum dos estágios de diferenciação de linfócitos B (AGIRRE et al., 2019). Caracterizações amplas da expressão de lncRNAs também foram feitas para os estágios de diferenciação da linhagem eritroide de camundongos (ALVAREZ-DOMINGUEZ et al., 2014) e para a linhagem linfoide humana (RANZANI et al., 2015).

Apesar da importância dos lncRNAs na regulação do sistema imunológico, não temos conhecimento de relatos associando estes transcritos à resposta imunológica induzida por vacinas. Se o estudo da resposta transcricional induzida por vacinas for limitada ao exame de genes codificadores de proteína, então deixamos de considerar a importante camada de regulação do sistema imune composta por lncRNAs.

1.6 REANOTAÇÃO DE PLATAFORMAS DE *MICROARRAYS*

Um *microarray* é, em linhas gerais, um substrato contendo numerosos pontos de oligonucleotídeos (sondas*)* aderidos e dispostos em uma posição conhecida (GOVINDARAJAN et al., 2012). O fenômeno em que se baseia a tecnologia de *microarrays* é a hibridização de sequências complementares de nucleotídeos, o que implica que as sondas sejam desenhadas para ter como alvos transcritos previamente conhecidos. As sondas representadas em plataformas comerciais de *microarray* foram escolhidas com base em bancos de dados de *Expressed Sequence Tags* (ESTs) (LIU; BEBU; LI, 2010). Desta forma, os alvos destas sondas não necessariamente consistem em genes codificadores de proteínas, mas as ESTs representadas nesses repositórios (ENGELHARDT; STADLER, 2012).

O aumento do conhecimento associado aos transcritos representados em

sondas de plataformas comerciais de *microarrays* torna obsoletas as anotações fornecidas pelos fabricantes. O remapeamento das sondas, levando em consideração esse novo conhecimento, foi proposto como uma solução capaz de aumentar a precisão da anotação (SANDBERG; LARSSON, 2007). Outro benefício desse procedimento é a possibilidade de análise focalizada no estudo de lncRNAs (JOHNSON, 2012; MICHELHAUGH et al., 2011; ZHANG et al., 2012).

Diversos trabalhos de vacinologia de sistemas estudaram alterações transcricionais de genes codificadores de proteína utilizando plataformas comerciais de *microarrays*. Concomitantemente, essas plataformas também mediram o nível de expressão de milhares de lncRNAs (JOHNSON, 2012; MICHELHAUGH et al., 2011; ZHANG et al., 2012). A plataforma *Affymetrix Human Genome U133 Plus 2.0 Array,* muito utilizada em estudos de vacinologia de sistemas, possui conjuntos de sondas (*probe sets*) capazes de medir a expressão de pelo menos 1.970 diferentes lncRNAs (ZHANG et al., 2012). Outra plataforma amplamente utilizada, a *Affymetrix Human Exon 1.0 ST array,* possui 10.207 *probe sets* com essa capacidade (DU et al., 2013). Trabalhos pioneiros estudaram a expressão de lncRNAs em *arrays* desenhados especialmente para este fim (NAKAYA et al., 2007; REIS et al., 2004). Não obstante, a alta representação de lncRNAs em plataformas comerciais aponta para a subutilização de uma rica fonte de informação escondida em conjuntos de dados já consolidados (ENGELHARDT; STADLER, 2012), o que nos motivou a desenvolver um *pipeline* de reanotação de sondas.

A utilização de plataformas comerciais de *microarray* com foco na busca por lncRNAs foi empregada em diversos estudos. Exemplos incluem a identificação de perfis de coexpressão de lncRNAs em estágios de desenvolvimento de células B humanas (PETRI et al., 2015) e sepse (PELLEGRINA et al., 2017). Outros estudos fizeram reanálises de dados públicos provenientes de quatro tipos de câncer (DU et al., 2013) e de glioma (ZHANG et al., 2012). O reaproveitamento de amostras de *microarray* pode também servir para a confirmação de hipóteses formuladas *a priori*. Análises de correlação permitiram o aumento do corpo de evidências comprobatórias da atividade exercida pelo lncRNA *ANRASSF1* no controle epigenético da expressão do gene RASSF1A, por exemplo (BECKEDORFF et al., 2013).

É importante mencionar que a utilização de *microarrays* na quantificação da expressão gênica vem sendo suplantada por tecnologias de sequenciamento de

RNA (RNA-*seq*). Estas apresentam diversas vantagens técnicas em relação às tecnologias de *microarrays*, tais como a possibilidade de identificação/quantificação de novos transcritos, isoformas e variantes genéticas, além de uma faixa dinâmica de quantificação de transcritos mais larga. Ademais, *microarrays* introduzem vieses técnicos relacionados ao desenho dos oligonucletídeos tais como a hibridização cruzada, hibridização não especifica e heterogeneidade no sinal de sondas representando os mesmos transcritos (ZHAO et al., 2014). Esses vieses impactam a quantificação de transcritos pouco abundantes, cujos sinais se confundem com o ruído de fundo, e de transcritos muito abundantes, que podem apresentar sinais saturados (HSIAO et al., 2002). Apesar dessas questões, a massiva quantidade de dados de *microarray* disponíveis publicamente enseja a sua reutilização na formulação de novas questões científicas.

1.7 METANÁLISES COM DADOS DE EXPRESSÃO GÊNICA

A ampla adoção da tecnologia dos *microarrays* pela comunidade científica levou ao acúmulo de milhões de amostras em repositórios públicos, tais como o *ArrayExpress* e o *Gene Expression Omnibus* (GEO). A organização dessas amostras em bancos de dados permite não apenas a reprodução das análises conduzidas por seus autores, mas também a reutilização para outros fins não previstos originalmente. A condução de metanálises consiste em uma possibilidade de reutilização desses dados.

Metanálises consistem em sumarizações de resultados de múltiplos estudos com o propósito geral de permitir a estimativa de estatísticas mais precisas e generalizáveis pelo aumento do número de amostras. A possibilidade de formulação de questões não abordadas nos estudos em que os dados foram originalmente utilizados constitui uma vantagem adicional trazida por metanálises (HIGGINS; GREEN, 2011). No entanto, múltiplos fatores devem ser considerados antes da realização de uma metanálise. Por exemplo, os estudos a serem integrados devem ser suficientemente similares em aspectos relevantes para a questão científica proposta (HIGGINS; GREEN, 2011).

Metanálises com dados de *microarray* são mais comumente focadas na identificação de Genes Diferencialmente Expressos (DEGs, acrônimo em inglês de *Differentially Expressed Genes*) (TSENG; GHOSH; FEINGOLD, 2012). Contudo,

outros propósitos podem ser contemplados. Por exemplo, metanálises com dados de *microarray* podem também ter foco na classificação de fenótipos, estudo da ativação de vias (*pathway analysis)* e construção de redes consenso de coexpressão (TSENG; GHOSH; FEINGOLD, 2012).

Ao conduzir metanálises com *microarrays,* é preciso decidir se um novo conjunto de dados com todas as amostras será criado, ou se resultados de análises conduzidas nos dados de diferentes estudos serão sumarizados. No primeiro caso, métodos de normalização cruzada devem ser aplicados a fim de mitigar os vieses técnicos inerentes à realização de experimentos em diferentes laboratórios (WALSH et al., 2015). Apesar de existirem vantagens na análise de um *dataset* com maior número de amostras, a sumarização dos tamanhos de efeitos dos estudos facilita o controle da heterogeneidade interlaboratorial (RUNG; BRAZMA, 2013).

Estudos a serem integrados podem ter desfechos medidos em razões de possibilidades, coeficientes de correlação ou diferença de médias entre grupos, por exemplo. Estas medidas, denominadas tamanhos de efeito, podem ser utilizadas na determinação de uma estatística que resume os múltiplos estudos. A estimativa dos resumos de tamanhos de efeito pode ser feita de duas maneiras, a depender das premissas adotadas. Partindo-se da premissa de que os resultados das comparações conduzidas em diferentes coortes estimam um tamanho de efeito comum à população, sendo os desvios causados unicamente por erros de amostragem, a estimativa de um resumo pode ser feita com um modelo de metanálise de efeitos fixos. Contudo, essa premissa nem sempre pode ser adotada com segurança. Se considerarmos que as medidas tomadas em diferentes coortes consistem em estimadores de parâmetros específicos para as diferentes subpopulações estudadas, um modelo de metanálise de efeitos aleatórios deve ser aplicado. Desconsiderar a existência de heterogeneidade nos tamanhos de efeitos de diferentes coortes pode gerar estimativas de resumo excessivamente precisas. Em outras palavras, os intervalos de confiança podem ser mais estreitos do que aqueles suportados pelos dados (RILEY; HIGGINS; DEEKS, 2011).

1.8 REDES BIOLÓGICAS E REDES DE COEXPRESSÃO

Funções biológicas discretas raramente podem ser atribuídas a moléculas individuais. Em realidade, a função de componentes celulares é mais comumente

exercida pela interação modular com outros componentes celulares (HARTWELL et al., 1999). O conjunto destas interações, ou interactoma, pode conter relações de regulação, modificação pós-traducional, participação em uma mesma via bioquímica e interação física, por exemplo. A quantidade de moléculas biológicas que servem como nós no interactoma, incluindo moléculas de RNA com atividade funcional, proteínas e metabólitos excede facilmente o número de 100.000 (BARABÁSI; GULBAHCE; LOSCALZO, 2011). A aplicação de métodos de análise de redes é capaz de auxiliar na compreensão dos fenômenos biológicos controlados por redes biológicas (BARABÁSI; OLTVAI, 2004).

Conceitos da teoria de redes comumente empregados na interpretação de redes biológicas incluem a modularidade, a distribuição dos graus e a centralidade de nós (BARABÁSI; GULBAHCE; LOSCALZO, 2011). Modularidade refere-se à organização de redes em comunidades de nós altamente conectados entre si. Esse conceito é particularmente importante para a biologia, pois funções celulares são conduzidas de maneira modular (BARABÁSI; OLTVAI, 2004; HARTWELL et al., 1999). A distribuição dos graus de uma rede refere-se à caracterização global de seus padrões de conexão. O conceito de “grau” está relacionado à caracterização da conectividade de nós individuais. O grau de um nó consiste em seu número total de conexões. A análise de redes de diferentes tipos revelou que as distribuições de graus destas redes comumente seguem uma lei de potência. Isso significa que poucos nós são muito conectados, enquanto a maioria dos outros nós são pouco conectados. Redes que seguem esse padrão foram denominadas redes livres de escala (BARABÁSI; ALBERT, 1999). Finalmente, nós altamente conectados, comumente referenciados como *hubs,* têm maior importância para a estabilidade de sistemas biológicos do que outros nós menos conectados (JEONG et al., 2001).

A massiva quantidade de dados gerados com *microarrays* fomentou a aplicação de estratégias de interpretação que iam além da mera identificação de transcritos diferencialmente regulados entre duas condições. Em um exemplo clássico, a clusterização de dados de expressão gênica revelou que genes que exercem funções similares são coexpressos (EISEN et al., 1998). Posteriormente, um *framework* de construção de redes de coexpressão baseado no ajuste a um modelo livre de escala, proposto por Langfelder e Horvath, foi amplamente adotado pela comunidade científica (LANGFELDER; HORVATH, 2008). A premissa motivadora para a construção de redes de coexpressão é a de que genes que

compartilham padrões similares de expressão são controlados por um mesmo programa regulatório, são relacionados funcionalmente, ou são membros de uma mesma via ou complexo proteico (WEIRAUCH, 2011). Não obstante, outros fatores devem ser considerados na interpretação do fenômeno da coexpressão, tal como a covariância na expressão de genes marcadores de tipos celulares em amostras de RNA provenientes de misturas de células (*bulk* RNA) (GAITERI et al., 2014).

1.9 JUSTIFICATIVA

Neste trabalho fizemos metanálises com dados de expressão gênica oriundos de diferentes coortes de pessoas vacinadas. Nossa hipótese inicial era de que lncRNAs estavam de alguma forma associados à resposta imunológica induzida pela vacinação. Todas as análises foram conduzidas separadamente em cada conjunto de dados e posteriormente integradas. Uma vez que as coortes foram imunizadas em diferentes anos, com formulações contendo diferentes cepas, e diferentes plataformas de *microarray* foram utilizadas, partimos do pressuposto de que as medidas de expressão apresentavam heterogeneidade com origem de difícil identificação. Por esse motivo, adotamos um modelo de metanálises de efeitos aleatórios. Adicionalmente, construímos redes consenso de coexpressão e identificamos comunidades nestas redes. Nossos resultados trazem evidências de que lncRNAs estão associados a múltiplas frentes da resposta imunológica induzida por vacinas e apontam para a necessidade de estudos adicionais confirmando a associação desses transcritos à regulação dos processos da imunologia humana.

2 OBJETIVOS

2.1 OBJETIVO PRINCIPAL

Estudar o papel de lncRNAs na resposta imune humana induzida pela vacinação por meio da reanálise de dados de expressão gênica de indivíduos imunizados com vacinas inativadas contra gripe (*Inactivated Influenza Vaccines,* ou IIVs), a vacina atenuada febre amarela (YF-17D), e uma vacina atenuada contra gripe (*Live Attenuated Influenza Vaccine,* ou LAIV).

2.2 OBJETIVOS ESPECÍFICOS

a) Identificar lncRNAs diferencialmente expressos após a vacinação.
b) Determinar correlações entre lncRNAs com mudanças em títulos de anticorpos induzidas pela vacinação.
c) Determinar as correlações entre a expressão de lncRNAs e mRNAs localizados na mesma região genômica em busca de evidências de regulação em *cis*.
d) Construir redes de coexpressão dos mRNAs e lncRNAs em busca de evidências de regulação em *trans*.
e) Aplicar métodos de integração dos resultados.
f) Propor anotações funcionais de lncRNAs
g) Avaliar diferenças de expressão de lncRNAs após imunização de crianças com LAIV em dados de RNA-*seq*.

3 MATERIAIS E MÉTODOS

Fizemos reanálises de bancos de dados públicos de *microarrays* com enfoque na detecção de lncRNAs. Conduzimos metanálises de expressão diferencial, correlação com títulos de anticorpos, correlação entre lncRNA e mRNAs próximos e coexpressão. Tais procedimentos foram feitos a fim de identificar lncRNAs relacionados a diferentes aspectos da resposta imune induzida por vacinas. As diferentes etapas foram sumarizadas na Figura 2.

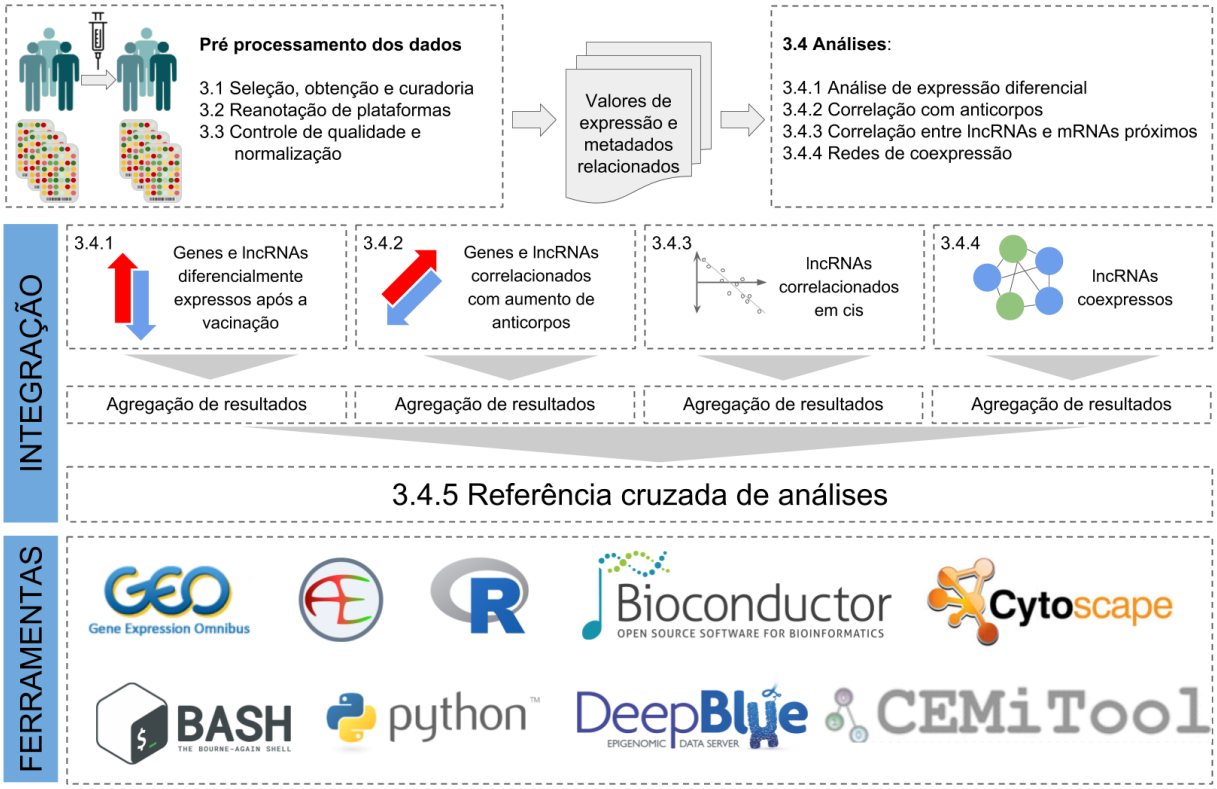

Figura 2 - Representação esquemática da metodologia. Números correspondem às seções do texto.

3.1 SELEÇÃO, OBTENÇÃO E CURADORIA DOS DADOS

3.1.1 Dados públicos de expressão gênica

Buscas com os termos “influenza vaccination” e ‘”yellow fever vaccination” foram realizadas nos repositórios *Gene Expression Omnibus* (GEO) e ArrayExpress. Os resultados das buscas foram pré-filtrados para a seleção de estudos conduzidos em plataformas comerciais de *microarray* com pelo menos 10 amostras. Os dados de expressão gênica foram, exceto quando explicitamente especificado, coletados do GEO. Um *script* que baixa as matrizes de expressão e as armazenam em pastas pré-determinadas foi utilizado para esse fim. Os metadados associados a cada estudo foram baixados por um *script* que acessa a página do GEO e indexa as informações de cada estudo em um arquivo de texto. Dados adicionais foram coletados manualmente no repositório ArrayExpress e em arquivos suplementares de artigos.

Uma etapa fundamental para a correta realização das análises é a padronização dos metadados. Nesta etapa aplicamos os critérios de

inclusão/exclusão de amostras e fizemos a curadoria dos metadados. Selecionamos apenas os estudos em que amostras foram coletadas no momento da vacinação e em pelo menos mais uma ocasião posterior. Além disso, removemos estudos em que os códigos identificadores dos indivíduos vacinados, o tipo de vacina aplicada e os tempos experimentais não estavam disponíveis. Selecionamos apenas as amostras provenientes de sangue total ou *Peripheral Blood Mononuclear Cells* (PBMC), coletadas de indivíduos com idade entre 18 e 49 anos, por meio de punção venosa. A terminologia utilizada no corpo e cabeçalho dos arquivos de anotação foi padronizada, a fim de tornar possível a execução automática das análises.

Medidas de títulos de anticorpos, feitas por meio de ensaios de inibição da hemaglutinação (HAI) ou neutralização (MN), estavam disponíveis para alguns estudos. Nesses casos, esses dados foram coletados, transformados e concatenados aos arquivos de anotação. Os títulos de anticorpos correspondem ao número recíproco da medida de diluição. As razões das medidas de títulos de anticorpos (logFCs), foram calculadas de acordo com a seguinte fórmula.

$$logFC = log_2(z_x) - log_2(z_0)$$

Onde z_x corresponde ao título de anticorpo do indivíduo z no tempo x após a vacinação e z_0 corresponde ao título de anticorpo do indivíduo z antes da vacinação. Nos casos em que foram dosados anticorpos contra diferentes cepas, selecionamos aquele que apresentou o maior logFC. Com isto, as dosagens de anticorpos deixaram de serem registros em tempos específicos, para tornarem-se medidas proporcionais às alterações de concentração de anticorpos induzidas pela vacinação em cada indivíduo. Todas as etapas de padronização e transformação de dados foram feitas em *scripts* escritos em R (R CORE TEAM, 2015) para cada estudo.

3.1.2 Dados de expressão gênica de células do sistema imune

Dados normalizados de expressão gênica (FPKM) de diversos tipos celulares foram coletados por meio do uso da *Application Programming Interface* (API) DeepBlue (disponível em http://deepblue.mpi-inf.mpg.de/). Essa API possibilita a requisição padronizada dos dados de expressão gênica e marcas epigenéticas oriundas de amostras coletadas em consórcios que incluem o *Encyclopedia of DNA*

Elements (ENCODE), o *NIH Roadmap Epigenomics Mapping Consortium* (ROADMAP), o *BLUEPRINT Epigenome Project* e o *German Epigenome Programme* (DEEP) (ALBRECHT et al., 2016). Requisições foram enviadas para a API por meio de um *script* escrito na linguagem de programação Python (https://www.python.org/). A seleção das amostras foi feita com base nas ontologias *Uber anatomy ontology* (Uberon), *Cell Ontology* (CL) e *Experimental Factor Ontology* (EFO) (MALLADI et al., 2015). Requisições à API foram feitas com os termos filhos de 'B cell', 'T cell', 'monocyte', 'blood', 'dendritic cell', 'macrophage', 'eosinophil', 'basophil', 'neutrophil', 'natural killer cell' e 'hematopoietic stem cell'.

3.1.3 RNA-*seq* de coorte imunizada com LAIV

Dados inéditos de expressão gênica, obtidos e processados por colaboradores, foram utilizados neste trabalho. Crianças com idades entre 2 a 4 anos (n = 52) foram imunizadas com a vacina atenuada contra gripe (LAIV, Serum Institute of India Pvt Ltd). Amostras de sangue foram coletadas em tubos PAXgene nos dias 0 e 2 após a vacinação (clinicaltrials.gov, identificador NCT02972957). Etapas de preparação de bibliotecas e sequenciamento de RNA (Illumina Hiseq X, *paired-end* e *reads* com 100 pares de bases) foram conduzidas no Beijing Genome Institute. Controle de qualidade das *reads* foi realizado com FastQC. Alinhamento ao genoma humano (GRCh38), foi feito com DART (versão 1.3.0) (LIN; HSU, 2018). *Reads* duplicadas com distância inferior a 2500 pixels na mesma *tile* foram removidas com Picard MarkDuplicates. Contagem de *reads* foi feita com a função featureCounts do pacote Rsubread (LIAO; SMYTH; SHI, 2014). Dados brutos e processados foram disponibilizados no repositório GEO sob o código de acesso GSE1282224.

3.2 REANOTAÇÃO DAS PLATAFORMAS DE *MICROARRAYS*

Um *pipeline* de reanotação foi desenvolvido a fim de atribuir novas anotações às sondas presentes em plataformas de *microarray.* Sequências das sondas foram alinhadas com a montagem HG38 do genoma humano usando BLAT (KENT, 2002) com limiares de *scores* de alinhamento > 90% e *scores* de identidade > 90%. As sondas com múltiplos alinhamentos no genoma foram descartadas, e as sondas

restantes foram atribuídas a genes anotados na versão v24 do Gencode usando BEDTools (QUINLAN; HALL, 2010). Sondas atribuídas a mais de um gene foram também descartadas. A visualização das regiões genômicas das sondas reanotadas foi realizada com a ferramenta *Genome Browser* (KENT et al., 2002).

3.3 CONTROLE DE QUALIDADE E NORMALIZAÇÃO DE DADOS DE EXPRESSÃO

O controle de qualidade das amostras foi realizado com o pacote ArrayQualityMetrics (KAUFFMANN; GENTLEMAN; HUBER, 2009). Esse pacote realiza testes estatísticos com intensidades de expressão e identifica amostras *outliers.* As amostras que reprovaram em dois dos três testes foram consideradas aberrantes, removidas e registradas em uma lista. Esse processo foi repetido iterativamente com as amostras restantes até que não ocorressem reprovações. As amostras listadas foram removidas da análise.

Em seguida, as intensidades de expressão foram normalizadas e submetidas a uma transformação logarítmica. Dados coletados em plataformas da *Affymetrix* foram normalizados com o algoritmo *Robust Multi-Array Average* (RMA) (IRIZARRY et al., 2003). Dados oriundos de outras plataformas foram normalizados com a função *normalizeBetweenArrays* do pacote limma (RITCHIE et al., 2015). Finalmente, as sondas desenhadas para o mesmo gene foram sumarizadas por meio da seleção daquela com maior média de expressão entre as amostras.

3.4 ANÁLISES

Um *pipeline* de análises foi escrito com o pacote Snakemake (KOSTER; RAHMANN, 2012). Esse *pipeline* consiste em uma série de regras de execução dos *scripts* que realizam as seguintes tarefas: 1) abrem e processam arquivos de expressão e seus metadados associados; 2) detectam genes diferencialmente expressos; 3) determinam a correlação entre pares lncRNAs-mRNAs em *cis*; 4) determinam a correlação entre a expressão de lncRNAs com a dosagem de anticorpos (quando disponíveis) e finalmente; 5) detectam redes de coexpressão. Esse procedimento foi realizado para garantir não apenas a reprodutibilidade, mas também a escalabilidade das análises, o que significa que a adição de um novo conjunto de dados não demanda a reformulação dos *scripts*.

Os resultados das análises foram armazenados em tabelas. Essas foram integradas posteriormente por meio de *scripts* escritos em R. As visualizações dos resultados foram feitas com os pacotes ggplot2 e ComplexHeatmap (GU; EILS; SCHLESNER, 2016). Visualizações de redes foram construídas com o *software* Gephi (BASTIAN; HEYMANN; JACOMY, [s.d.]).

3.4.1 Detecção de genes diferencialmente expressos

Genes diferencialmente expressos após a vacinação foram detectados através da aplicação de um teste t moderado fornecido no pacote limma (RITCHIE et al., 2015). Os valores de intensidade de expressão de cada gene após a vacinação (e.g. dias 1, 3, 7) foram comparados aos valores do dia 0 (*baseline)* em uma análise pareada pelo indivíduo vacinado. Esse procedimento foi conduzido separadamente por estudo, sendo os resultados armazenados em tabelas. Essas continham as diferenças das médias de expressão entre grupos experimentais, os p-valores associados às comparações e os intervalos de confiança de 95%.

3.4.2 Detecção de genes correlacionados a títulos de anticorpos

As diferenças dos logaritmos dos *fold changes* (logFC) de cada gene para cada indivíduo foram calculados separadamente em cada estudo conforme com a seguinte fórmula:

$$logFC = log2(yz_x) - log2(yz_0)$$

Onde "yz_x" corresponde à intensidade de expressão do gene y no indivíduo z no tempo x e "yz_0" corresponde à intensidade de expressão do gene y no indivíduo z no tempo 0. Os valores resultantes foram correlacionados aos logFCs dos títulos de anticorpos de cada indivíduo.

3.4.3 Detecção de pares genes-lncRNAs correlacionados em *cis*

Pares de lncRNAs e mRNAs em um mesmo cromossomo separados por no máximo 200.000 pares de bases foram armazenados em uma tabela de

correspondência. Um *script* que compara a posição genômica de lncRNAs com as posições de mRNAs foi escrito em R (R CORE TEAM, 2015) com o pacote GenomicRanges (LAWRENCE et al., 2013) obtido do projeto Bioconductor (GENTLEMAN et al., 2004).

O cálculo da correlação de Pearson entre os logFCs de expressão dos lncRNAs e de seus respectivos mRNAs próximos foi realizado. Nesta etapa foram utilizadas a tabela de correspondência lncRNA-mRNA e os dados de logFCs calculados para cada estudo conforme descrito no item 3.4.2. Os resultados foram armazenados em um arquivo para cada estudo e integrados em uma metanálise de correlação.

3.4.4 Detecção de módulos de coexpressão

A detecção dos módulos de coexpressão foi realizada com o pacote CEMiTool, publicado por nosso grupo de pesquisa (RUSSO et al., 2018). Esse pacote consiste em uma aplicação *user-friendly* que estende as funcionalidades do algoritmo de detecção de módulos de coexpressão WGCNA (ZHANG; HORVATH, 2005). As tabelas pré-processadas e metadados de cada estudo foram submetidas a funções do CEMiTool utilizando parâmetros padrão. O resultado obtido consistiu em uma tabela por estudo contendo a correspondência entre cada um dos genes selecionados e o módulo a ele atribuído.

A integração dos módulos de coexpressão em redes consenso foi implementada no pacote CEMiTool de acordo com o seguinte algoritmo:

➔ Para cada resultado (correspondência genes-módulos em um estudo):
    Para cada módulo de coexpressão do estudo:
        Ligar todas as combinações gene-gene (arestas) dentro do módulo.
        Calcular coeficiente de correlação da expressão dos pares de genes.
        Armazenar as arestas e coeficientes em uma lista de arestas.
    Concatenar listas de arestas de todos os módulos do estudo em uma tabela.
➔ Unir listas de arestas de todos os estudos em uma tabela.
➔ Selecionar as arestas com base na representação em estudos e/ou correlação segundo os seguintes critérios:
    1) Estavam representadas em pelo menos duas coortes;
    2) Tinham uma média de coeficiente de correlação maior que 0.6 caso estivesse representada em duas coortes;
    3) Tinham uma média de coeficiente de correlação maior que 0.55 caso estivesse representada em três coortes;

4) Tinham uma média de coeficiente de correlação maior que 0.5 caso estivesse representada em quatro coortes;
5) Tinham uma média de coeficiente de correlação maior que 0.45 caso estivesse representada em cinco coortes (apenas para IIVs);
6) Estavam representadas em seis ou mais coortes (IIVs) ou cinco ou mais coortes (para YF-17D).

Esse algoritmo foi aplicado duas vezes, uma para as coortes vacinadas com IIVs e outra para as coortes imunizadas com YF-17D. Como resultado, obtivemos redes compreendendo todos os pares de genes consistentemente coexpressos na maioria dos estudos. Contudo, o procedimento implica a perda da informação de modularidade originalmente presente nos resultados de cada estudo. A fim de detectar módulos consenso a partir da rede consenso, implementamos no CEMiTool uma função de detecção de comunidades, definidas como agrupamentos de nós em uma rede, com base no pacote iGraph (CSARDI; NEPUSZ, 2006). Neste trabalho utilizamos o algoritmo de particionamento de redes baseado em um modelo *spin-glass* (REICHARDT; BORNHOLDT, 2006), e selecionamos comunidades com um mínimo de 15 genes.

3.4.5 Estratégias de Metanálise

Os tamanhos de efeitos obtidos das análises de expressão diferencial, correlação entre pares de genes vizinhos e correlação com títulos de anticorpos foram integrados por meio da aplicação de um modelo de efeitos aleatórios implementado no pacote metafor (VIECHTBAUER, 2010). Os tamanhos de efeitos das análises de expressão diferencial consistiam na diferença de médias de expressão. No caso das análises de correlação, os tamanhos de efeitos correspondiam aos coeficientes de correlação de Pearson. Tamanhos de efeito e suas variâncias foram submetidas à função rma do pacote metafor para a estimativa dos resumos dos tamanhos de efeito. Estudos foram ponderados pelo inverso da variância. Metanálises baseadas na contagem das vezes em que genes passavam por um limiar de p-valor pré-determinado (*vote counting)* foram também realizadas.

3.4.6 Enriquecimento funcional

A determinação da relevância biológica das comunidades consenso foi feita

de duas maneiras. Primeiramente, a anotação funcional foi feita por meio da comparação de listas de genes com *Blood Transcriptional Modules* (BTMs) (LI et al., 2014). Para isso, utilizamos o teste exato de Fisher, implementado no pacote clusterProfiler (YU et al., 2012). Mudanças congruentes na atividade destas comunidades foram quantificadas em cada coorte por meio da aplicação do teste de enriquecimento fornecido no pacote fgsea (SERGUSHICHEV, 2016). Comunidades com p-valor ajustado < 0.05 foram selecionadas, e seus *Normalized Enrichment Scores* foram representados em *heatmaps.*

4 RESULTADOS

4.1 DADOS DE EXPRESSÃO E REANOTAÇÃO DE PLATAFORMAS

De acordo com os critérios apresentados na seção 3.1.1, coletamos dados públicos de expressão oriundos de 17 coortes vacinadas. Dessas, doze continham amostras de indivíduos imunizadas com IIVs e cinco continham amostras de indivíduos imunizadas com YF-17D. Diversas coortes continham também dados de dosagem de anticorpos, tomados antes e depois da vacinação (Tabela 1 e Figura 3).

Tabela 1 - Bancos de dados analisados. Abreviações: ID – Identificador interno; AB – Segundo dia de dosagem de anticorpos. Estudos com código de acesso GSE13486 e GSE13699 foram desagregados por conterem dados de diferentes coortes.

Referência	ID	Código de Acesso	AB	Vacina	Platforma
(QUEREC et al., 2009)	YF-17D_01	GSE13486	-	YF-17D	GPL570
(QUEREC et al., 2009)	YF-17D_02	GSE13486	-	YF-17D	GPL570
(GAUCHER et al., 2008)	YF-17D_03	GSE13699	-	YF-17D	GPL6104
(GAUCHER et al., 2008)	YF-17D_04	GSE13699	-	YF-17D	GPL6104
(HOU et al., 2017)	YF-17D_05	GSE82152	-	YF-17D	GPL15207
(TSANG et al., 2014)	IV_09_4	GSE47353	70	IIV	GPL6244
(BUCASAS et al., 2011)	IV_08_2	GSE48018	28	IIV	GPL6947
(BUCASAS et al., 2011)	IV_09_3	GSE48023	28	IIV	GPL10558
(OBERMOSER et al., 2013)	IV_09_2	GSE48762	-	IIV	GPL6947
(THAKAR et al., 2015)	IV_10_2	GSE59635	-	IIV	GPL10558
(THAKAR et al., 2015)	IV_11_2	GSE59654	-	IIV	GPL10558
(SOBOLEV et al., 2016)	IV_09_5	E-MTAB-2313	63	IIV	GPL4133
(NAKAYA et al., 2015)	IV_10_1	GSE74817/GSE29619	28	IIV	GPL570
(NAKAYA et al., 2015)	IV_11_1	GSE74817/GSE29619	28	IIV	GPL570
(NAKAYA et al., 2015)	IV_07_1	GSE74817/GSE29619	28	IIV	GPL570
(NAKAYA et al., 2015)	IV_08_1	GSE74817/GSE29619	28	IIV	GPL570
(NAKAYA et al., 2015)	IV_09_1	GSE74817/GSE29619	28	IIV	GPL570

Diferentes estudos foram desenhados para responder diferentes questões científicas. Os bancos de dados reanalisados neste trabalho foram originalmente utilizados na determinação de preditores de resposta de anticorpos (TSANG et al., 2014), comparação entre diferentes vacinas (OBERMOSER et al., 2013),

comparação entre a resposta vacinal de jovens e idosos (THAKAR et al., 2015), estudo da vacina inativada contra gripe suína (H1N1) contendo o adjuvante AS03 (SOBOLEV et al., 2016) e compreensão da resposta da vacina YF-17D contra febre-amarela (GAUCHER et al., 2008; HOU et al., 2017; QUEREC et al., 2009). Outros conjuntos de dados não puderam ser utilizados devido a particularidades em seus desenhos experimentais. No *dataset* GSE58970, por exemplo, só estavam disponíveis amostras coletadas antes da vacinação. Em GSE81692, por sua vez, as amostras de sangue foram coletadas unicamente por punção digital. No segundo caso, as diferenças transcricionais observadas entre amostras de sangue venoso e capilar podem afetar a comparabilidade de resultados (STEIN et al., 2016).

Amostras que não passaram nos critérios do controle de qualidade ou que estavam em desacordo com nossa proposta foram removidas. Nossa análise integrativa foi conduzida com um menor número de amostras também devido aos diferentes regimes de coleta. Nos estudos de vacinação contra febre amarela, todos os bancos de dados continham amostras referentes aos dias 3 e 7. Nos estudos de vacinação contra gripe, os estudos apresentaram uma maior variação no número total de amostras e nos dias de coleta, sendo mais frequentes as amostras coletadas nos dias 1, 3 e 7. A figura 3 indica o número final de amostras analisadas em cada coorte. Dados de dosagem de anticorpos estavam disponíveis para diversas coortes.

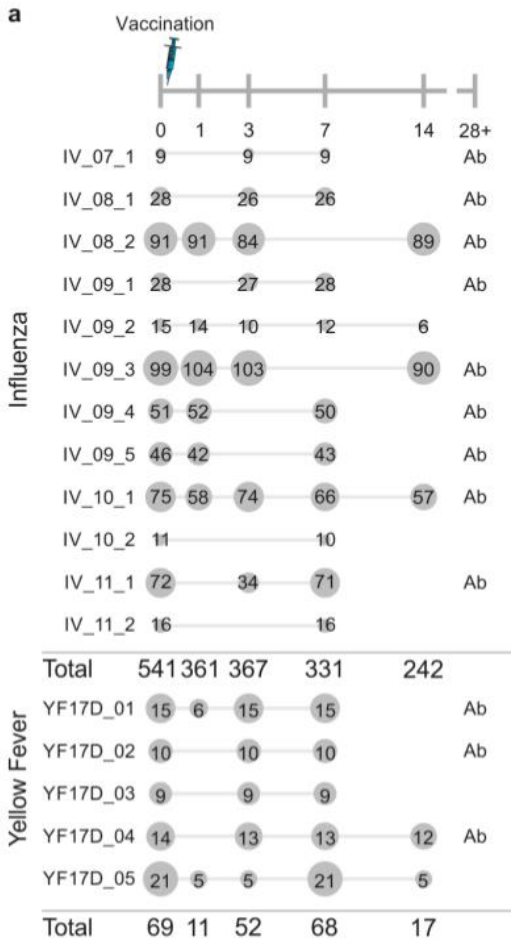

Figura 3 - Número de amostras analisadas por estudo. Tamanho e número dos círculos são proporcionais à quantidade de amostras disponíveis para o dia de coleta correspondente. Datasets com informação de dosagem de anticorpos estão identificados pela palavra “Ab”.

As plataformas empregadas nos bancos de dados foram reanotadas. Todas as plataformas continham sondas anotadas como lncRNAs. As classes mais frequentes de lncRNAs foram, nos dois casos, de lncRNAs antisenso e intergênicos (lincRNAs). Porém, lncRNAs classificados como intrônicos e *sense overlapping* foram também detectados (Figura 4). Além disso, a análise de sondas com anotação oficial diferente de nossa reanotação revelou que sondas originalmente anotadas como tendo alvo em genes codificadores representavam, na verdade, lncRNAs antisenso, intrônico ou intergênicos (Figura 5). A análise da intersecção das sondas reanotadas como lncRNAs nas diferentes plataformas, revelou que 204 e 91 lncRNAs estavam representados concomitantemente em todas as plataformas utilizadas nas coortes imunizadas com IIVs (Figura 6A) e YF-17D, respectivamente (Figura 6B).

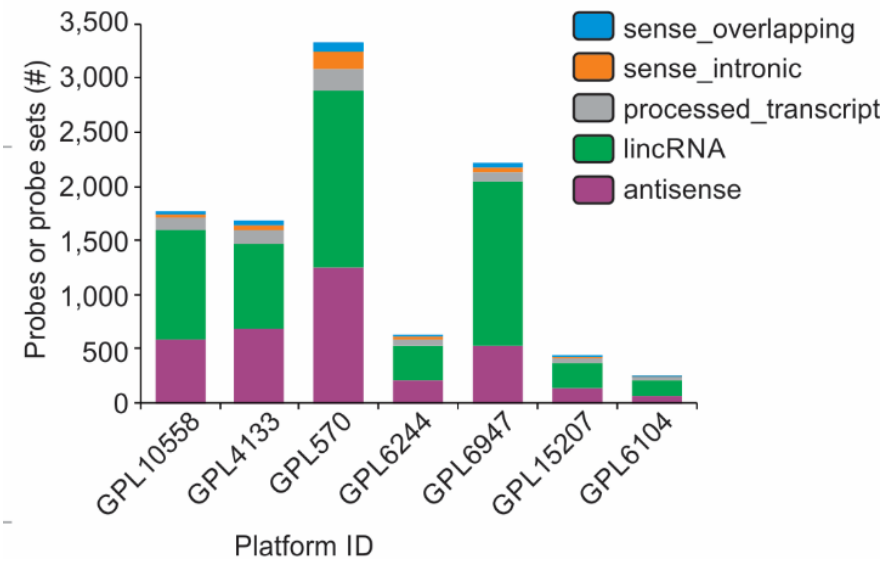

Figura 4 - Classes de lncRNAs em plataformas reanotadas. Classes foram definidas pelo Gencode v24.

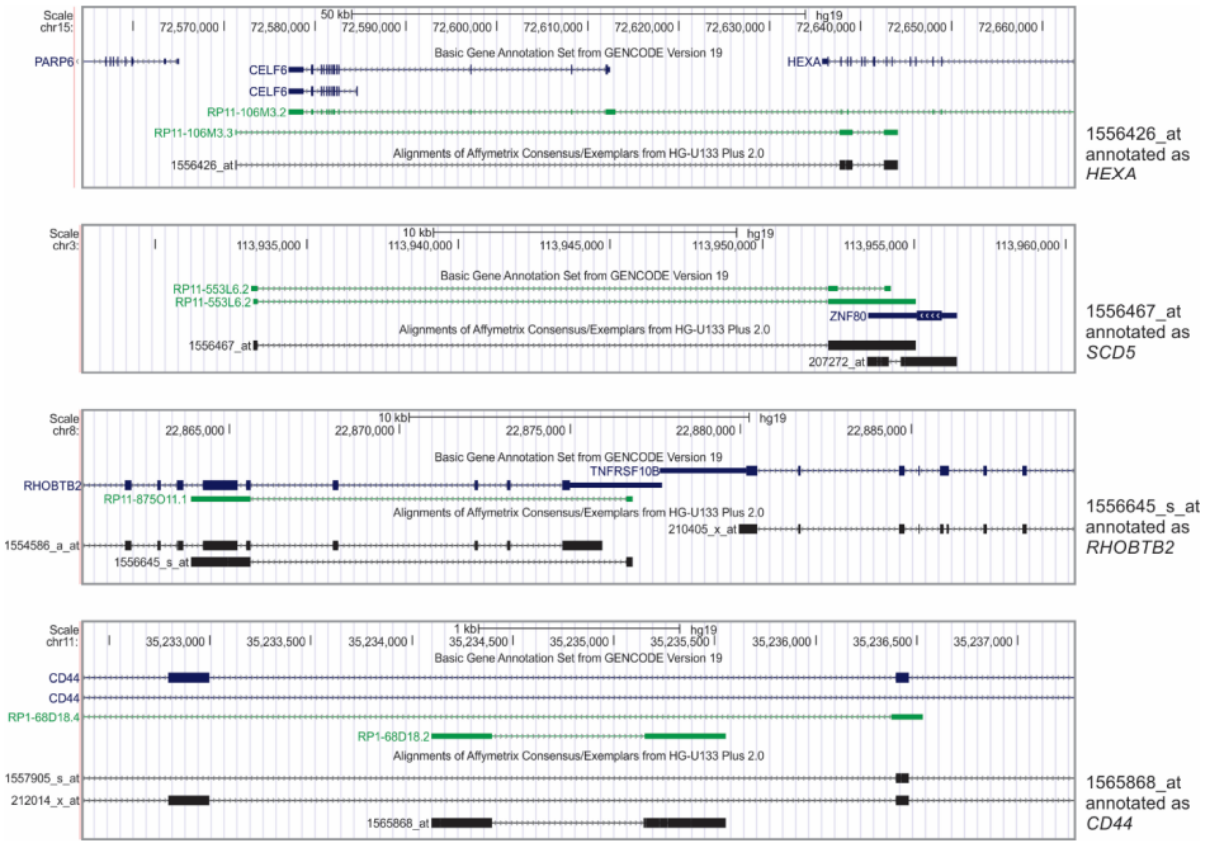

Figura 5 – Exemplos de sondas mal anotadas na plataforma GPL570. Cada seção contém o resultado da busca pela sonda no *Genome Browser.* Caixas representam genes ou sondas.

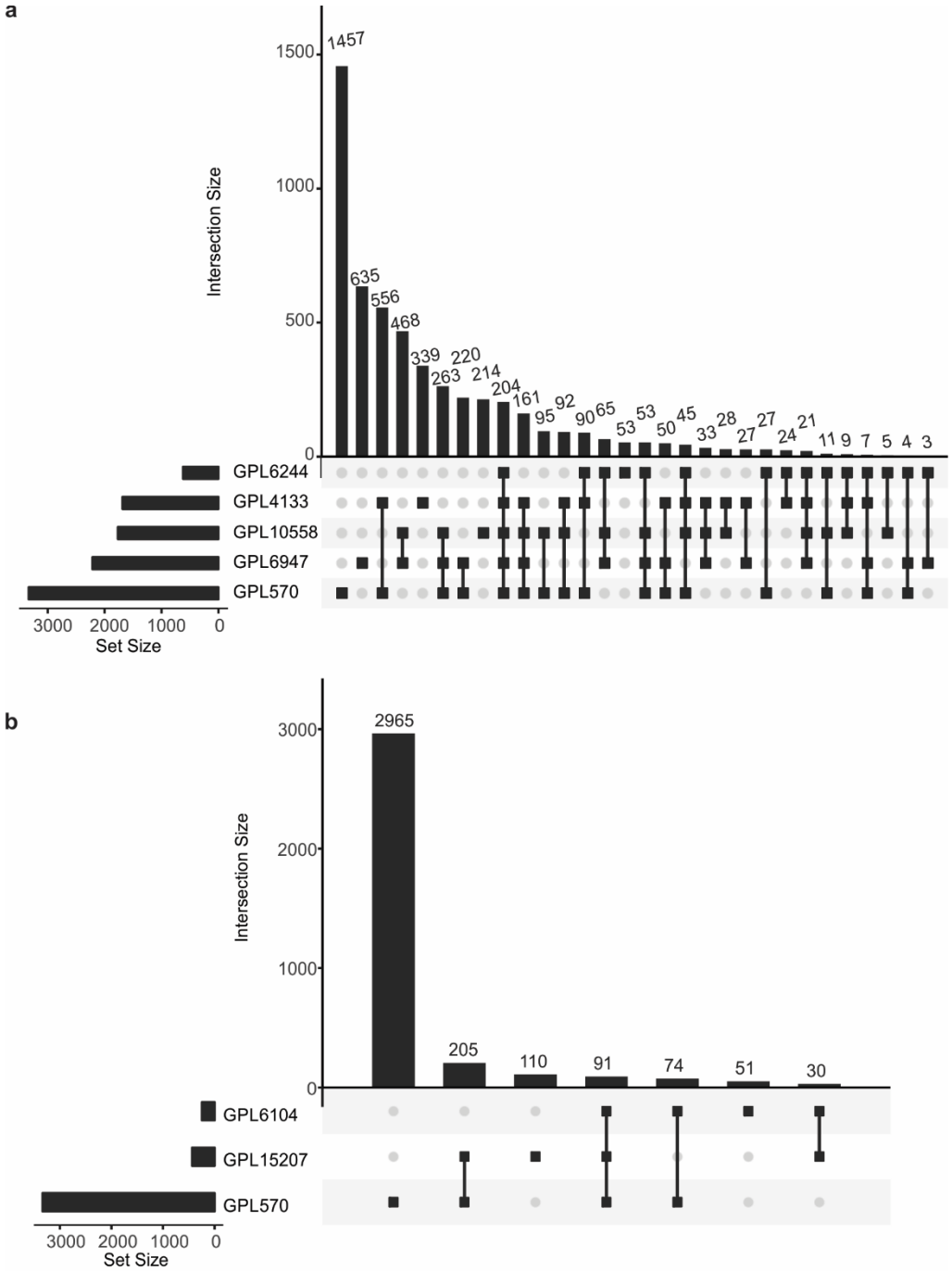

Figura 6 – Intersecção de lncRNAs representados nas plataformas. Barras verticais representam o tamanho da intersecção de lncRNAs representados nas plataformas utilizadas nos estudos de IIVs (a) ou YF-17D (b). Barras horizontais representam o número total de sondas com alvo em lncRNAs em cada plataforma.

4.2 EXPRESSÃO DIFERENCIAL DE GENES E LNCRNAS APÓS A VACINAÇÃO

Genes Diferencialmente Expressos (DEGs) foram detectados após a vacinação. A estratégia de metanálise baseada em contagem de votos (*vote counting)* indicou que intersecções entre DEGs detectados em comparações similares de coortes imunizadas com IIVs (Figura 7A) e YF-17D (Figura 7B) em múltiplos estudos foram consideravelmente baixas. Não obstante, genes consistentemente diferencialmente expressos detectados em nossa análise coincidiram com os resultados originalmente publicados pelos autores dos bancos de dados. Por exemplo, os genes *CD38*, *TNFRSF17* e *MZB1* estavam presentes em assinaturas gênicas induzidas pela vacina contra a gripe no dia 7 (NAKAYA et al., 2011; TSANG et al., 2014). Em nossa análise integrativa eles apresentaram uma consistente expressão diferencial no dia 7 após IIVs (Figura 7A). Esses resultados servem para confirmar nossa capacidade de reproduzir resultados prévios. Os genes consistentemente diferencialmente expressos nas coortes de YF-17D foram também previamente reportados na literatura como associados à resposta transcricional induzida por YF-17D. Por exemplo, genes relacionados à resposta de interferon foram detectados como diferencialmente expressos em todas as coortes no dia 7 após a vacinação (dados não apresentados). Essa ocorrência foi também previamente descrita pelos autores originais desses conjuntos de dados (GAUCHER et al., 2008; HOU et al., 2017).

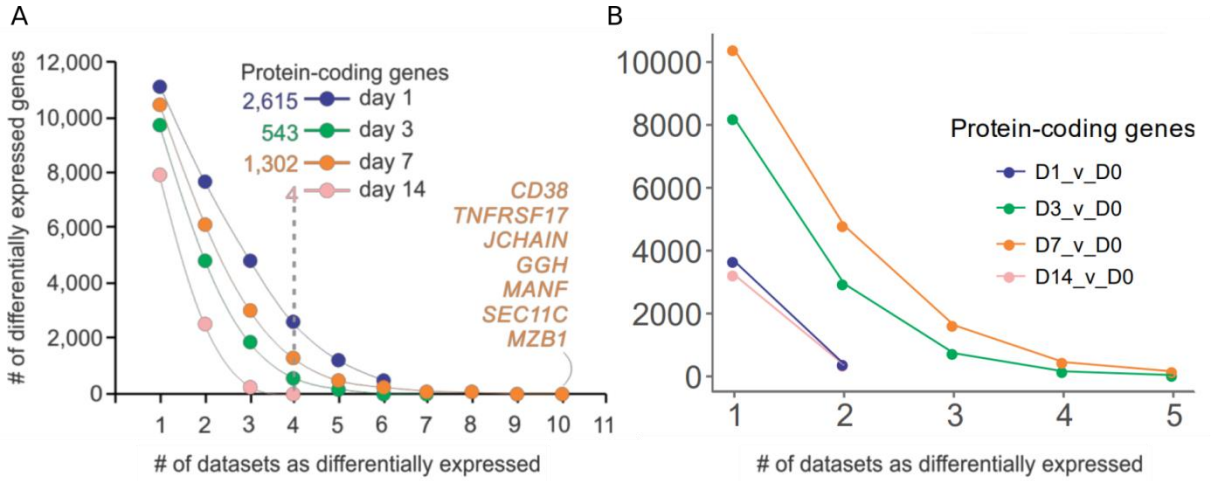

Figura 7 - Intersecção entre genes diferencialmente expressos em coortes imunizadas com IIVs (A) e YF-17D (B). Eixo x representa a quantidade mínima de coortes em que genes foram considerados diferencialmente expressos sob um limiar de p-valor nominal de 0.05. No eixo y está representada a soma cumulativa de genes diferencialmente expressos.

A mesma abordagem de metanálise de contagem de votos revelou diversos lncRNAs diferencialmente expressos em comparações similares de diferentes estudos de IIVs e YF-17D (Figura 8). Ainda assim, uma abordagem de metanálise baseada unicamente na contagem de vezes em que um gene passa por um critério arbitrário de significância não é capaz de extrair toda a informação disponível nos múltiplos conjuntos de dados. Medidas de intensidade e direção do tamanho de efeito, por exemplo, são completamente desconsideradas.

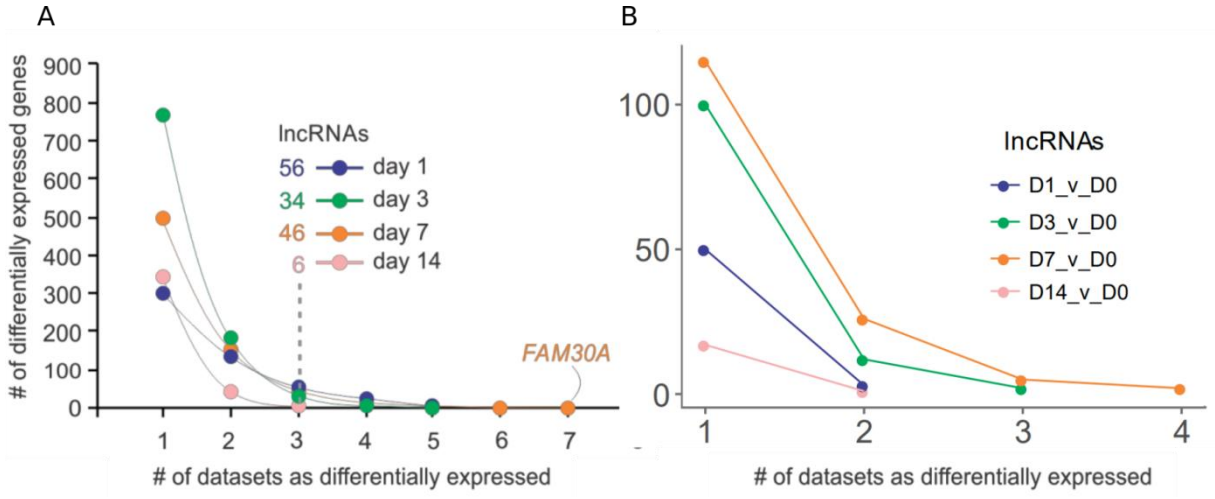

Figura 8 - Intersecção entre lncRNAs diferencialmente expressos em coortes imunizadas com IIVs (A) e YF-17D (B). Eixo x representa a quantidade mínima de coortes em que lncRNAs foram considerados diferencialmente expressos sob um limiar de p-valor nominal de 0.05. No eixo y está representada a soma cumulativa de lncRNAs diferencialmente expressos.

A fim de calcular uma estatística resumindo os tamanhos de efeito, utilizamos um método de metanálise de efeitos aleatórios fornecido no pacote metafor. Com essa abordagem, pudemos obter o resumo dos tamanhos de efeito relacionados à diferença de expressão de um gene em uma mesma comparação em múltiplos estudos, o intervalo de confiança de 95% e um p-valor da metanálise. Resultados das metanálises possibilitaram a identificação de centenas de lncRNAs diferencialmente regulados nos estudos de pacientes imunizados com IIVs (Figura 9) e YF-17D (Figura 10).

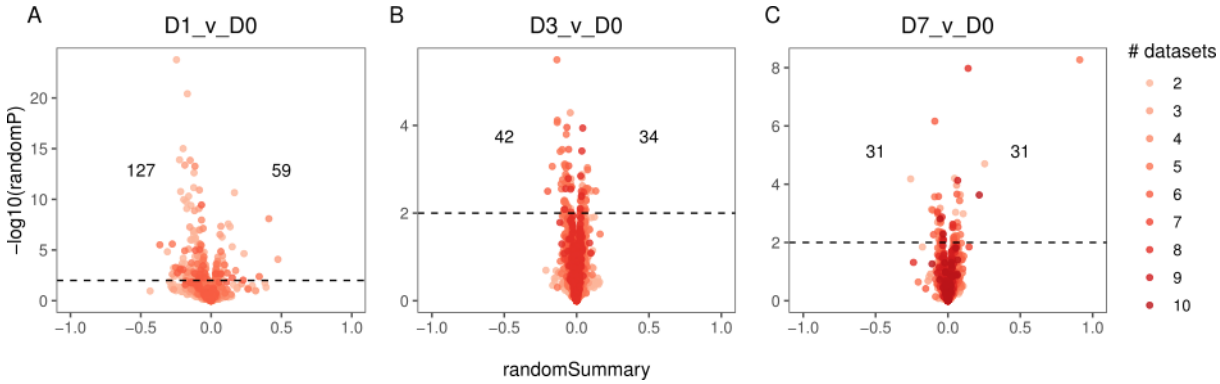

Figura 9 – *Volcano plots* das metanálises de expressão diferencial de lncRNAs nas coortes imunizadas com IIVs. Resumo dos tamanhos de efeito (eixo x) e logaritmos negativos dos p-valores das metanálises (eixo y). Linhas tracejadas representam limiares de p-valor nominal de 0,01. Número de lncRNAs regulados negativamente e positivamente (sob o limiar) estão representados nas áreas correspondentes dos gráficos. Cores dos pontos representam o número de coortes em que o lncRNA foi medido.

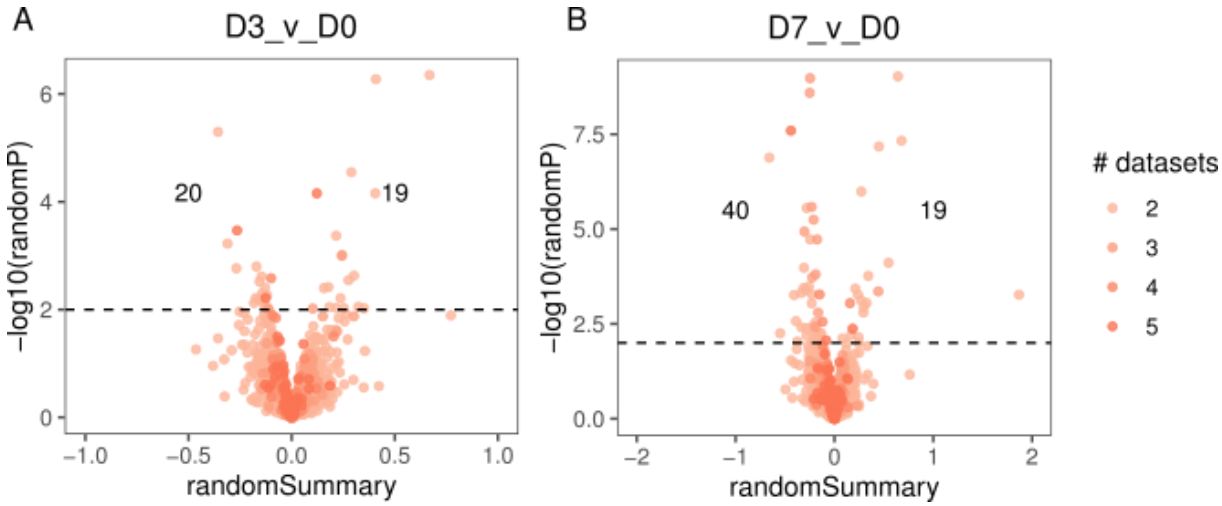

Figura 10 - *Volcano plots* das metanálises de expressão diferencial de lncRNAs nas coortes imunizadas com YF-17D. Resumo dos tamanhos de efeito (eixo x) e logaritmos negativos dos p-valores das metanálises (eixo y). Linhas tracejadas representam limiares de p-valor nominal de 0,01. Número de lncRNAs regulados negativamente e positivamente (sob o limiar) estão representados nas áreas correspondentes dos gráficos. Cores dos pontos representam o número de coortes em que o lncRNA foi medido.

Diversos lncRNAs apresentaram diferenças de expressão após 1, 3, 7 ou 14 dias após IIVs ou YF-17D (Figura 11). Nas coortes imunizadas com IIVs, *MIAT* (resumo = -0,365 e p-valor nominal = $3{,}05 \times 10^{-6}$) e *MCM3AP-AS1* (resumo = -0,112 e p-valor nominal = $1{,}06 \times 10^{-5}$) estavam regulados negativamente no dia 1 após a vacinação. *LINC01133* estava regulado positivamente no dia 3 (resumo = 0,075 e p-valor nominal = 0,000917) e *DANCR* no dia 7 após a vacinação (resumo = 0,139 e p-valor nominal = $1{,}05 \times 10^{-8}$). O gene codificador *TNFRSF17* estava, como esperado,

regulado positivamente no dia 7 em todas as coortes (resumo = 1,33 e p-valor nominal = $1,787 \times 10^{-11}$) (Figura 12).

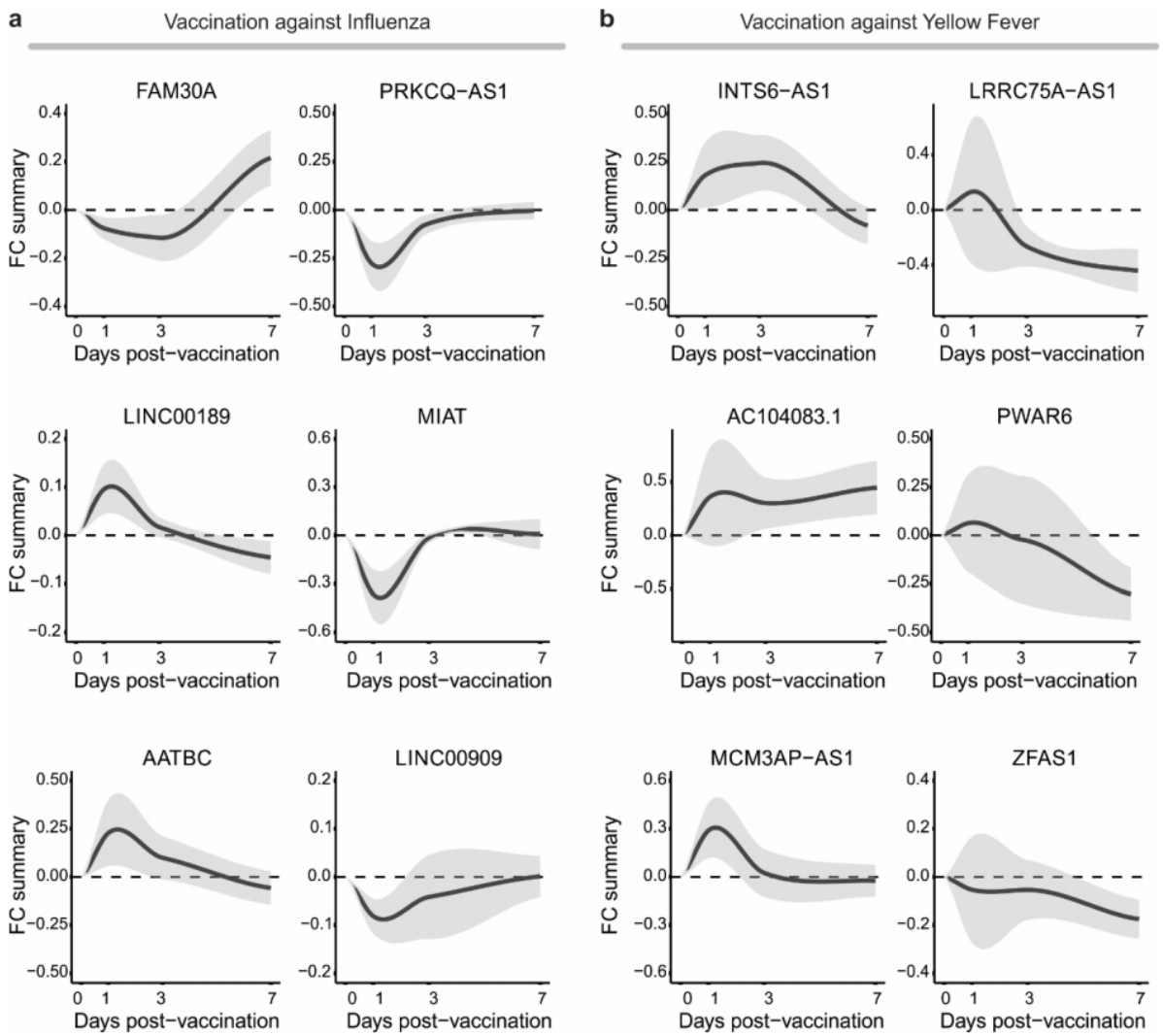

Figura 11 – Alterações na expressão de lncRNAs após a vacinação. (a) Exemplos de lncRNAs com expressão alterada após IIVs. (b) Exemplos de lncRNAs com expressão alterada após YF-17D. Curvas pretas e regiões sombreadas representam o resumo do logFC em relação ao *baseline* e seus intervalos de confiança de 95%, respectivamente.

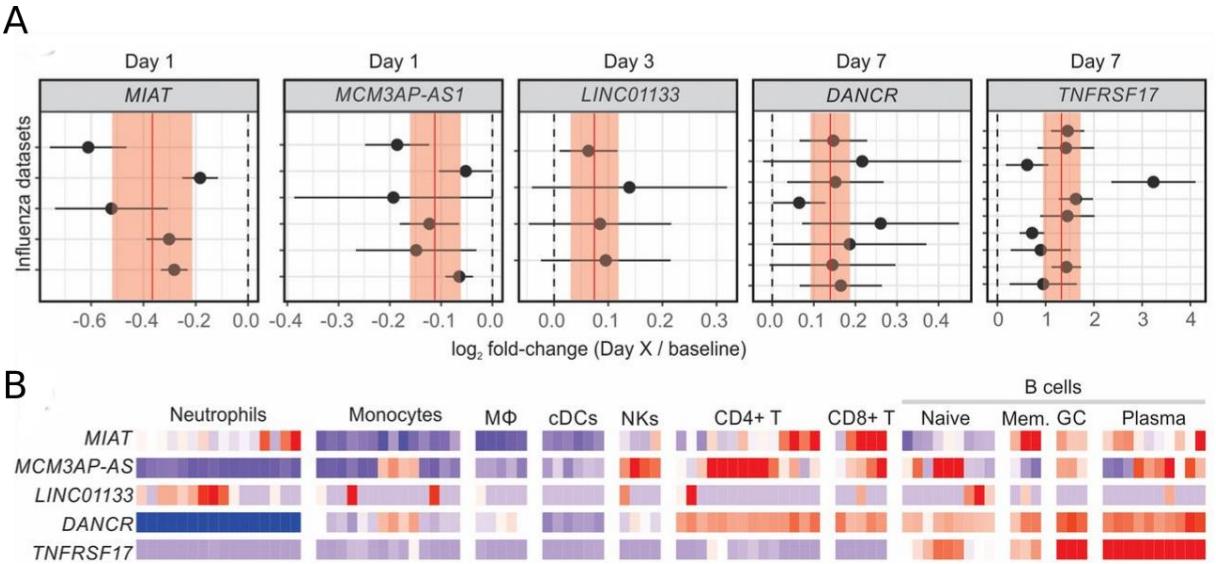

Figura 12 – Metanálise de lncRNAs representativos com expressão diferencial reiterada em coortes imunizadas com IIVs. (A) LogFCs e seus respectivos intervalos de confiança de 95% (eixo x) de cada coorte disponível (eixo y). Linhas verticais vermelhas e regiões sombreadas representam o resumo dos tamanhos de efeito, bem como os seus correspondentes intervalos de confiança de 95%, respectivamente. O resultado da metanálise com o gene codificador TNFRSF17 foi também representado. (B) *Heatmap* de dados de expressão gênica de células imunes humanas coletados do *Blueprint*. Intensidades representam *z-scores* dos $\log_2$(FPKM) de cada amostra.

Nas coortes imunizadas com YF-17D, lncRNAs diferencialmente expressos incluíram *AATBC* (resumo = 0,157 e p-valor nominal = 0,0008938 no dia 7), *MAPKAPK5-AS1* (resumo = -0,151 e p-valor nominal = 0,0005284 no dia 7) e *LRRC75A-AS1* (resumo = -0,265 e p-valor nominal = 0,000339 no dia 3; resumo = -0,441 e p-valor nominal = $2{,}524 \times 10^{-8}$ no dia 7). O aumento na expressão de genes estimulados por interferon é um marcador da resposta transcricional à vacinação com YF-17D no dia 7 (GAUCHER et al., 2008; QUEREC et al., 2009). Em nossos resultados, detectamos aumentos acentuados na expressão de *IFI27* em todas coortes imunizadas com YF-17D no dia 7 (resumo = 3,778 e p-valor nominal = $6{,}393 \times 10^{-45}$) (Figura 13).

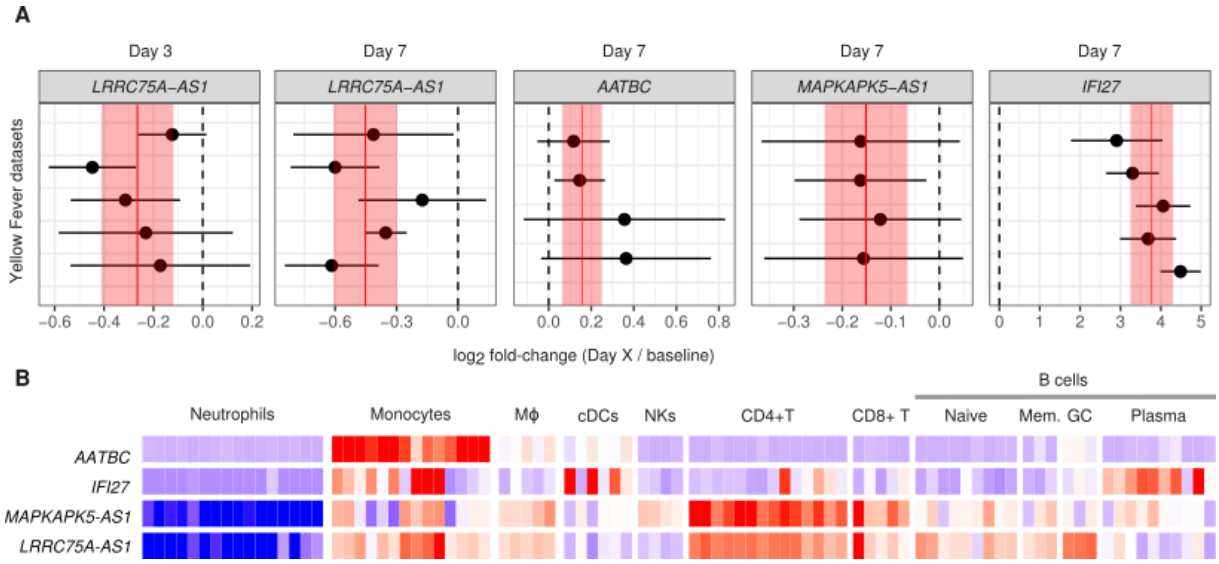

Figura 13 – Metanálise de lncRNAs representativos com expressão diferencial reiterada em coortes imunizadas com YF-17D. (A) LogFCs e seus respectivos intervalos de confiança de 95% (eixo x) de cada coorte disponível (eixo y). Linhas verticais vermelhas e regiões sombreadas representam respectivamente o resumo dos tamanhos de efeito, bem como os seus correspondentes intervalos de confiança de 95%. O resultado da metanálise com o gene codificador IFI27 foi também representado. (B) *Heatmap* de dados de expressão gênica de células imunes humanas coletados do *Blueprint*. Intensidades representam *z-scores* dos $\log_2$(FPKM) de cada amostra.

4.3 CORRELAÇÃO DA EXPRESSÃO DE LNCRNAS E TÍTULOS DE ANTICORPOS

Análises de expressão diferencial de indivíduos vacinados não consideram a capacidade da vacina de induzir proteção. Títulos de anticorpos são usados como correlatos de proteção para diversas vacinas. Mais ainda, a maioria das vacinas licenciadas induz proteção por intermédio da atividade exercida por anticorpos (PLOTKIN, 2008). Nesse contexto, estudos de vacinologia de sistemas identificaram genes potencialmente associados à produção de anticorpos por meio de análises de correlação entre a expressão gênica e aumentos nos títulos de anticorpos (NAKAYA et al., 2015; QUEREC et al., 2009; TSANG et al., 2014). Com o mesmo procedimento, pudemos associar mudanças na expressão de lncRNAs com mudanças na produção de anticorpos após imunização com IIVs. Em seguida aplicamos um modelo de metanálise de efeitos aleatórios a fim de determinar um resumo das correlações entre as diversas coortes. Diversos lncRNAs estavam correlacionados negativamente e positivamente com aumentos nos títulos de anticorpos em coortes imunizadas com IIVs (Figura 14).

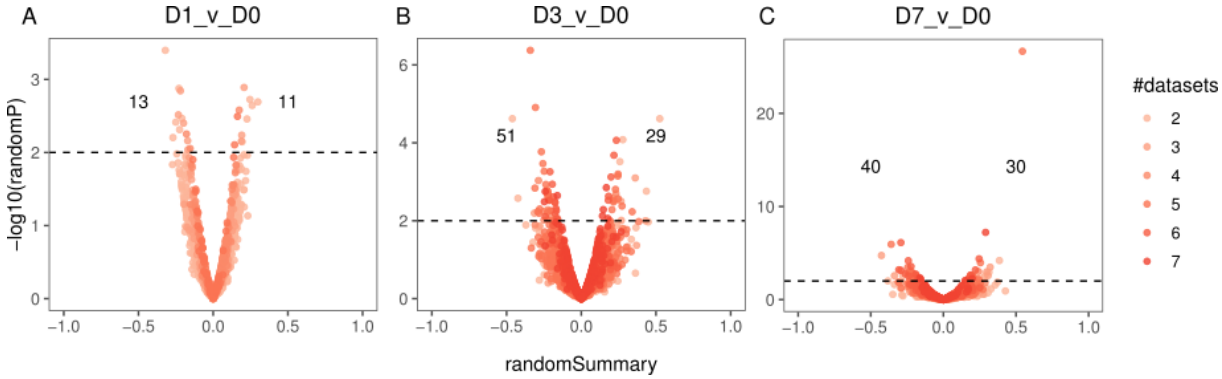

Figura 14 - *Volcano plots* das metanálises de correlação de lncRNAs com títulos de anticorpos nas coortes imunizadas com IIVs. Resumo dos tamanhos de efeito (eixo x) e logaritmos negativos dos p-valores das metanálises (eixo y). Linhas tracejadas representam limiares de p-valor nominal de 0,01. Número de lncRNAs correlacionados negativamente e positivamente (sob o limiar) estão representados nas áreas correspondentes dos gráficos. Cores dos pontos representam o número de coortes em que o lncRNA foi medido.

A análise desses lncRNAs pode apontar para suas funções em potencial. O transcrito *GAS6-AS1*, por exemplo, está correlacionado negativamente com a produção de anticorpos no dia 7 (resumo = -0,241 e p-valor = 0,00006541). Interessantemente, o gene *GAS6*, codificador complementar a esse lncRNA, está correlacionado positivamente, ainda que marginalmente, com a produção de anticorpos no mesmo dia (resumo = 0,18313 e p-valor = 0,013). O lncRNA *CDC42-IT1*, transcrito intrônico ao gene *CDC42*, está correlacionado negativamente com a produção de anticorpos no dia 7. Outros transcritos negativamente correlacionados com a produção de anticorpos em diferentes dias incluíram *TRIM52-AS1* e *C1orf132* (Figura 15).

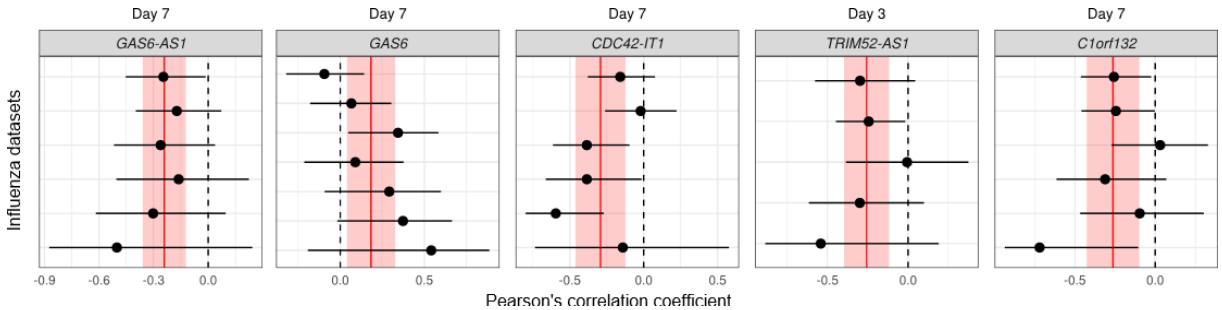

Figura 15 - *Forest plots* de lncRNAs correlacionados com títulos de anticorpos em coortes imunizadas com IIVs. Coeficientes de correlação de Pearson e seus respectivos intervalos de confiança de 95% (eixo x) de cada coorte disponível (eixo y). Linhas verticais vermelhas e regiões sombreadas representam o resumo dos tamanhos de efeito da metanálise, bem como os seus correspondentes intervalos de confiança de 95%, respectivamente.

4.4 CORRELAÇÃO ENTRE LNCRNAS E MRNAS NA MESMA REGIÃO GENÔMICA

Diversos lncRNAs tem ação limitada à região genômica em que estão localizados. A fim de buscar possíveis alvos de regulação de lncRNAs em *cis*, fizemos uma metanálise das correlações entre a expressão de lncRNAs e mRNAs localizados na mesma região genômica. Diversos pares lncRNAs-mRNAs estavam correlacionados negativamente e positivamente nas coortes imunizadas com IIVs (Figura 16).

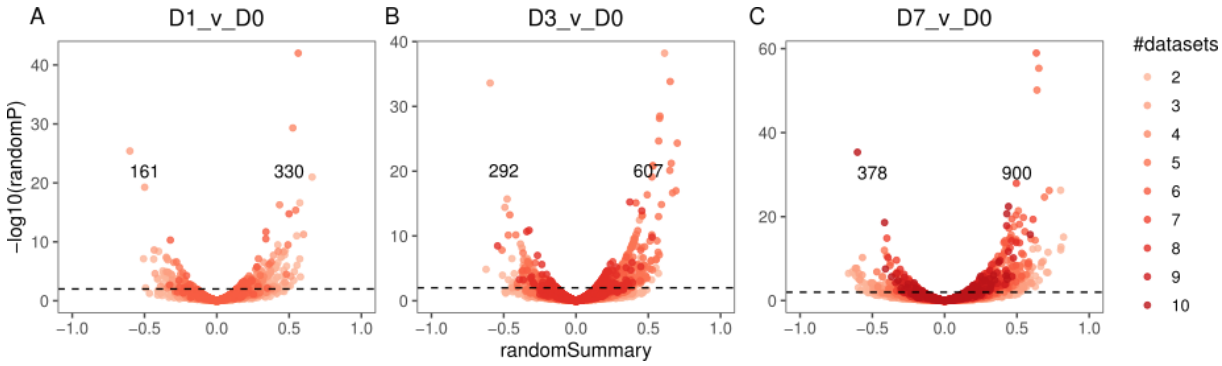

Figura 16 - *Volcano plots* das metanálises de correlação de lncRNAs com mRNAs próximos nas coortes imunizadas com IIVs. Resumo dos tamanhos de efeito (eixo x) e logaritmos negativos dos p-valores das metanálises (eixo y). Linhas tracejadas representam limiares de p-valor nominal de 0,01. Número de pares correlacionados negativamente e positivamente (sob o limiar) estão representados nas áreas correspondentes dos gráficos. Cores dos pontos representam o número de coortes em que o par foi medido.

4.5 DETECÇÃO DE REDES CONSENSO DE COEXPRESSÃO

Os resultados de análises de expressão diferencial consistem em listas de genes diferencialmente regulados entre dois grupos experimentais. Esse tipo de análise pode gerar resultados com baixa intersecção em comparações similares de diferentes estudos e com alta sensibilidade a ruído (CHAUSSABEL et al., 2008; IOANNIDIS, 2005; MICHIELS; KOSCIELNY; HILL, 2005). Apesar de permitirem a detecção de mudanças de expressão consistentes entre diversas coortes, a condução de metanálises de expressão diferencial não permite a detecção de padrões de expressão coordenados entre múltiplos transcritos. Fizemos a identificação de grupos de genes coexpressos com o CEMiTool, um pacote de detecção de redes de coexpressão publicado por nosso grupo de pesquisa (RUSSO et al., 2018). Posteriormente, construímos redes consenso de coexpressão por meio

da comparação das arestas detectadas nos trabalhos conduzidos com as mesmas vacinas. Esse procedimento tem o propósito de aumentar a relevância biológica das redes de coexpressão (LEE et al., 2004).

Duas redes consenso de coexpressão foram construídas: uma relativa às coortes imunizadas com IIVs (Figura 17A) e outra relativa às coortes imunizadas com YF-17D (Figura 18A). A primeira rede continha 2.574 genes, 64.922 interações e foi particionada em 16 comunidades. A atribuição de função às comunidades consenso e, por conseguinte, aos lncRNAs nelas presentes, foi feita por meio de análises de enriquecimento funcional. Para isto, comparamos os genes de cada uma das comunidades com *Blood Transcriptional Modules* (BTMs) (LI et al., 2014). As comunidades da rede construída com dados de coortes vacinadas com IIVs estavam associadas a grupos de genes característicos de células B e plasmócitos (CM7), células T (CM5), monócitos (CM1), células NK (CM9) e resposta antiviral mediada por interferon (IFN) (CM6) (Figura 17B).

A fim de avaliar se as comunidades consenso apresentavam alterações síncronas de expressão após a vacinação, fizemos análises de enriquecimento de vias. Nossa abordagem foi capaz de recapitular resultados previamente obtidos em estudos de vacinologia de sistemas. Nos dias 1 e 3 após IIV, por exemplo, detectamos um consistente aumento de expressão de uma comunidade associada à resposta mediada por IFN e ativação de células dendríticas (CM6). A progressiva diminuição de atividade desta comunidade nos dias 3 e 7 indica que a sinalização mediada por IFN é limitada ao início da resposta induzida por IIVs, como previamente relatado (ATHALE et al., 2017; NAKAYA et al., 2015). Comunidades associadas a células T e células NK apresentaram diminuição de atividade em todas as coortes no dia 1. No dia 7, por outro lado, foi possível detectar aumentos na expressão da comunidade CM7 (associada a plasmócitos) em todas as coortes (Figura 17C). Esse resultado condiz com o aumento na frequência de células produtoras de anticorpos, observado aproximadamente 7 dias após imunização com IIVs (SOBOLEV et al., 2016; TSANG et al., 2014).

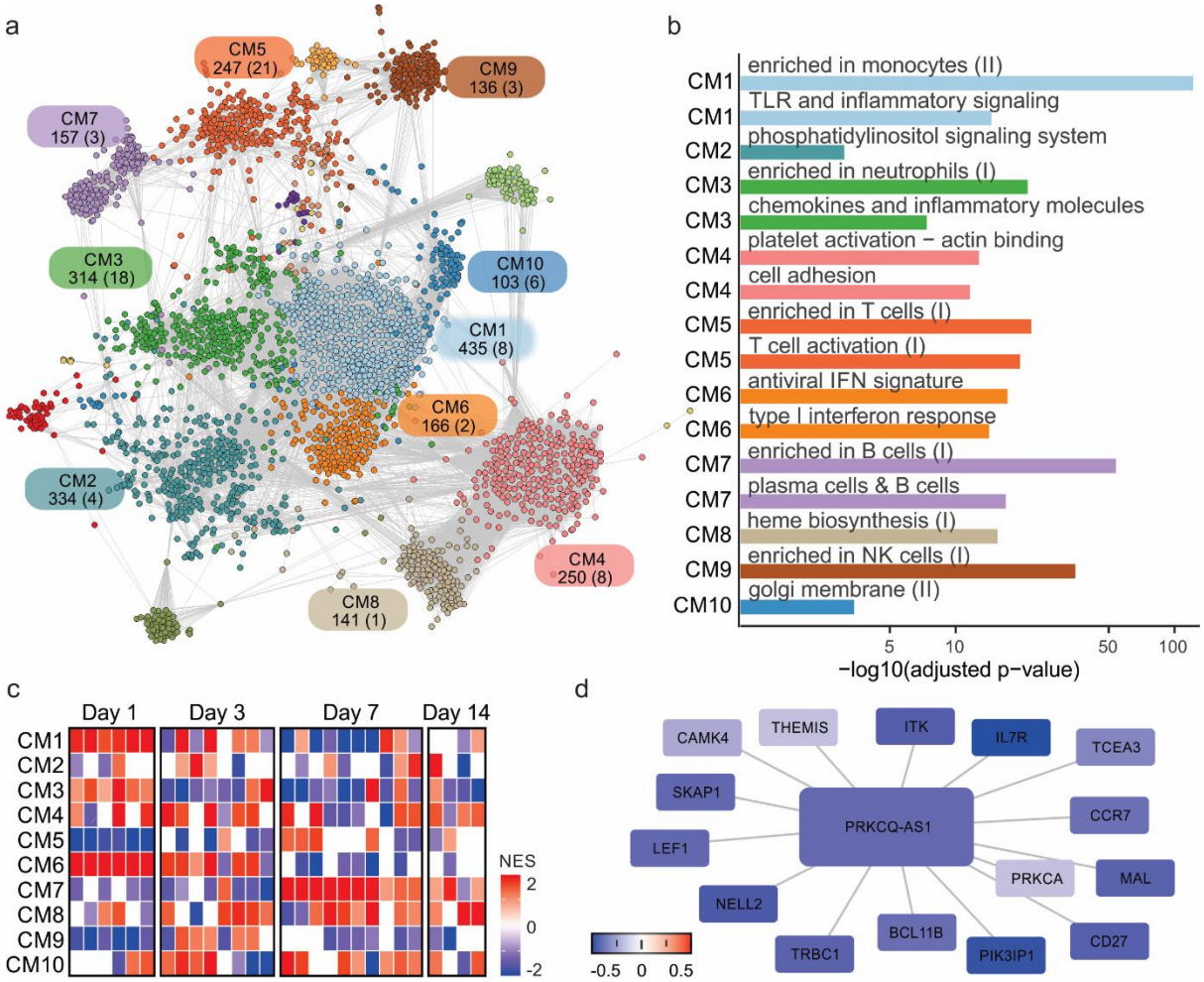

Figura 17 - Visão geral da rede consenso de coexpressão de IIVs. (a) Cores indicam a que comunidade os nós foram atribuídos. Retângulos coloridos com cantos arredondados contém o nome da comunidade, o número total de genes e o número de lncRNAs (em parênteses). (b) Resultados do enriquecimento funcional das comunidades usando BTMs. Gráfico de barras representam o logaritmo negativo na base 10 dos p-valores ajustados do teste exato de Fisher. (c) Resultados de enriquecimento de vias (GSEA) de coortes vacinadas com IIVs. Comunidades consenso foram usadas como vias (linhas) e média dos *fold-changes* dos genes de cada coorte como *ranks* (colunas). Intensidade de cores representa os *Normalized Enrichment Score* (NES) (d) Conexões representativas de PRKCQ-AS1 na CM5. As cores nos nós representam o resumo do logFC no dia 1 em relação ao *baseline.*

Múltiplos lncRNAs estavam presentes nas comunidades consenso de coexpressão de coortes imunizadas com IIVs (Quadro 1). A análise dos genes conectados a esses lncRNAs pode ser usada na inferência de suas funções em potencial. Por exemplo, o lncRNA *PRKCQ-AS1* é um membro da comunidade CM5, relacionada a células T na rede consenso de IIVs. A análise do enriquecimento de vias revelou que CM5 tem expressão diminuída no dia 1 após IIVs em múltiplas coortes (Figura 17C). Uma sub-rede de genes conectados a *PRKCQ-AS1* também

tem expressão regulada negativamente no dia 1 (Figura 17D). Diversos membros dessa sub-rede são genes relacionados a células T, como *CCR7*, *CD27*, *ITK*, *LEF1*, *THEMIS* e *SKAP1*. Mais ainda, o gene codificador *PRKCQ* está relacionado à atividade de células T (ZHANG; KONG; ALTMAN, 2013). Esses achados sugerem que o lncRNA *PRKCQ-AS1* pode estar envolvido com a função das células T.

Quadro 1 – lncRNAs em comunidades da rede consenso de coexpressão de coortes vacinadas com IIVs.

Comunidade	lncRNAs
CM1	*AATBC, AC104809.4, DAPK1-IT1, LINC00877, LINC01127, LINC01503, MIR223, RP3-525N10.2*
CM2	*MAPKAPK5-AS1*
CM3	*ALMS1-IT1, CTA-29F11.1, CTD-2033D15.2, CTD-2286N8.2, IL10RB-AS1, LINC00936, MIR222HG, MIR22HG, MIR24-2, RP11-127B20.3, RP11-194N12.2, RP11-373D23.2, RP11-386J22.3, RP11-489E7.4, RP11-670E13.6, RP11-769O8.3, SNHG15, SNHG17*
CM4	*AP001189.4, AP003068.23, C15orf54, RP11-588K22.2, RP11-874J12.4, RP11-879F14.2, RP11-90C4.1, WDR11-AS1*
CM5	*BX322557.10, CHRM3-AS2, CKMT2-AS1, CTD-2555O16.4, FOXP1-IT1, LEF1-AS1, LINC01550, MIR568, PRKCQ-AS1, PWAR6, RP11-138A9.1, RP11-158G18.1, RP11-159N11.4, RP11-161M6.2, RP11-283I3.6, RP11-327P2.7, RP11-395I6.3, RP11-526I2.5, RP11-549J18.1, RP11-664D1.1, SLC25A25-AS1*
CM6	*AC079630.4, RP11-476D10.1*
CM7	*FAM30A, GUSBP11, LINC00926*
CM8	*AC012146.7*
CM9	*A2M-AS1, MIAT, SNHG9*
CM10	*CARD8-AS1, HCP5, ILF3-AS1, LINC00324, RP1-39G22.7, RP11-448G15.3, RP11-53O19.3, RP11-747H7.3, RP11-97C16.1*
CM11	*LRRC75A-AS1, SNHG6*
CM12	*C9orf62, LINC01618, RP11-262I2.2*
CM13	*ITPK1-AS1, LINC00957, LINC01205, RP11-463J7.2, ST8SIA6-AS1*
CM14	*CTB-89H12.4, CTD-2196E14.6, FTX, RP11-104N10.2, SNHG5, SNHG7*
CM15	*LINC01089*

A rede construída com dados de coortes imunizadas com YF-17D continha 1.240 genes, 22.883 interações e foi particionada em 17 comunidades. Diversos lncRNAs eram membros de comunidades associadas a quimiocinas, diferenciação de células T e ativação de plaquetas (Figura 18B). A análise de enriquecimento de vias, feito nas comunidades, revelou que após IIVs, a resposta mediada por IFN restringia-se aos dias 1 e 3 (Figura 17C). No caso das coortes vacinadas com YF-17D, uma ativação da comunidade relacionada à resposta de IFN era ainda observada no dia 7 (Figura 18C).

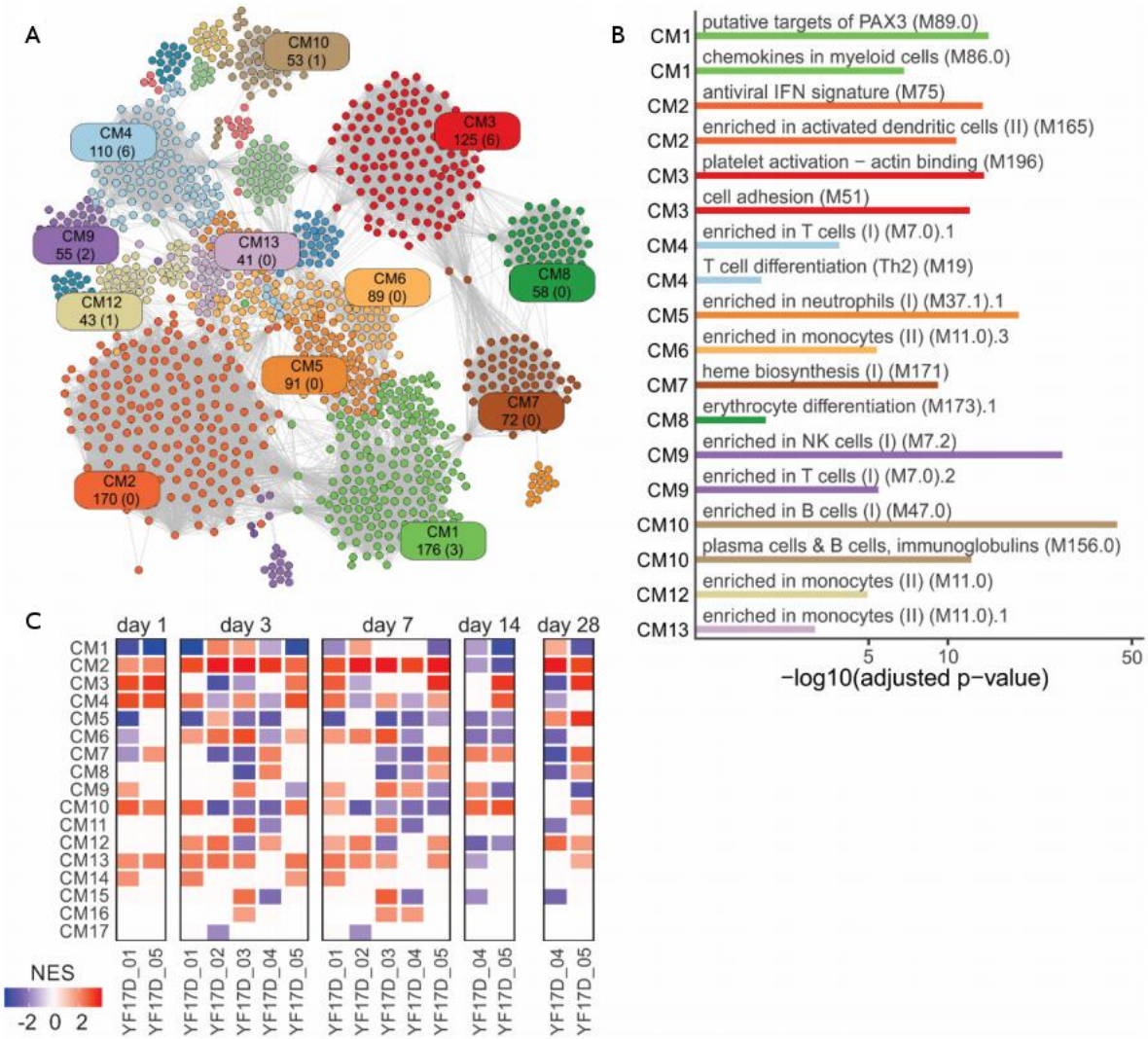

Figura 18 - Visão geral da rede consenso de coexpressão de YF-17D. (a) Cores indicam a que comunidade os nós foram atribuídos. Retângulos coloridos com cantos arredondados contém o nome da comunidade, o número total de genes e o número de lncRNAs (em parênteses). (b) Resultados do enriquecimento funcional das comunidades usando BTMs. Gráfico de barras representam o logaritmo negativo na base 10 dos p-valores ajustados do teste exato de Fisher. (c) Resultados de enriquecimento de vias (GSEA) de coortes vacinadas com YF-17D. Comunidades consenso foram usadas como vias (linhas) e média dos *fold-changes* dos genes de cada coorte como *ranks* (colunas). Intensidade de cores representa os *Normalized Enrichment Score* (NES)

4.6 EXPRESSÃO DE LNCRNAS EM CÉLULAS IMUNES

Mudanças na abundância de transcritos em misturas de diferentes células podem ser causadas tanto por regulação da expressão quanto por alterações nas frequências relativas destas populações celulares (CHIKINA; ZASLAVSKY; SEALFON, 2015; TSANG et al., 2014). Essa ocorrência é particularmente relevante

para o estudo de lncRNAs, já que a expressão destes transcritos tende a ter maior especificidade para tecidos do que os genes codificadores (CABILI et al., 2011; DERRIEN et al., 2012). A fim de identificar em quais linhagens celulares do sistema imune os lncRNAs discutidos neste trabalho tem maior probabilidade de serem expressos, utilizamos medidas de expressão gênica oriundas de linhagens imunes (ADAMS et al., 2012). A premissa motivadora para a condução desse procedimento é de que a identificação dos tipos celulares em que lncRNAs são expressos pode auxiliar na interpretação funcional desses. Por exemplo, lncRNAs relacionados à manutenção de pluripotência estão expressos especificamente em células-tronco (HON et al., 2017).

A maior parte dos lncRNAs membros da comunidade CM5 das coortes vacinadas com IIVs apresentaram alta expressão em células T $CD4^+$ em relação a outras linhagens imunes. O lncRNA *DANCR* estava expressos em células de origem linfoide (Figura 12B). O lncRNA *AATBC*, associado previamente com proliferação celular e apoptose (ZHAO et al., 2015), estava expresso especificamente em monócitos (Figura 13B). A expressão do gene codificador *GAS6* é elevada em plasmócitos e células dendríticas convencionais. Por outro lado, o lncRNA *GAS6-AS1* apresenta expressão maior em células B *Naive*, células B de memória e células B de centro germinativo do que em plasmócitos (Figura 20).

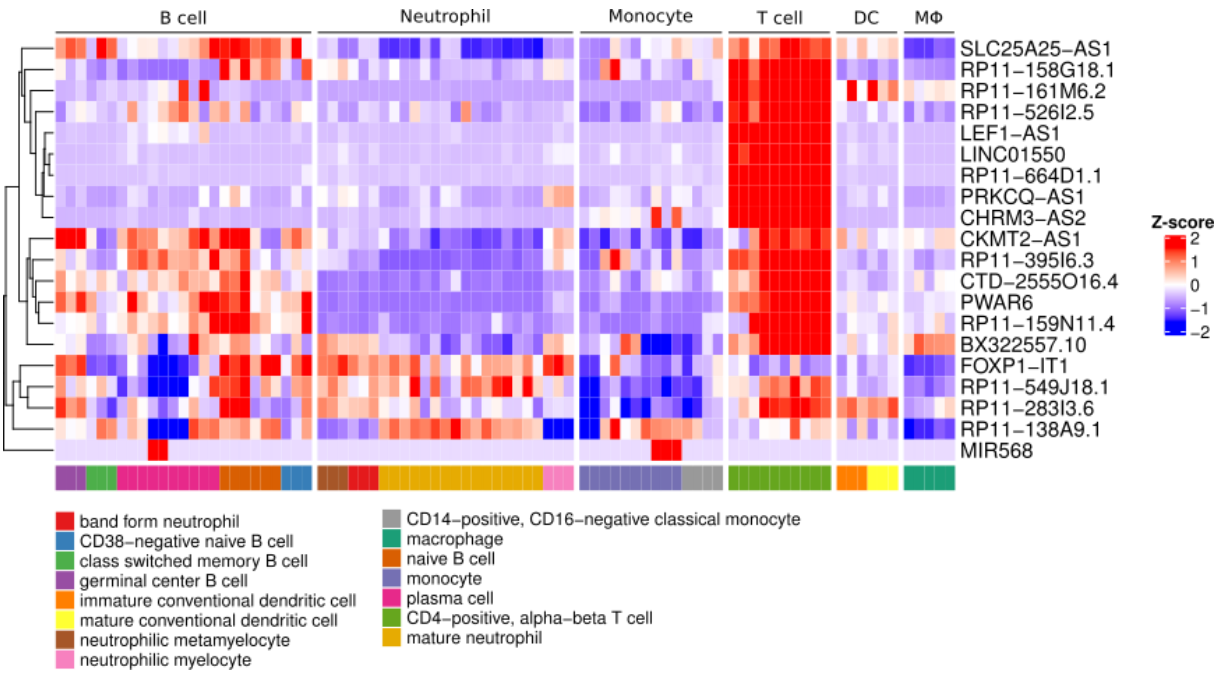

Figura 19 - Expressão de lncRNAs da comunidade CM5 das coortes vacinadas com IIVs em linhagens de células imunes. Colunas representam diferentes grupos de linhagens celulares.

Barra inferior representa origens específicas das células. O agrupamento dos lncRNAs, representado no dendrograma, foi feito por meio de um algoritmo de *clustering* hiearárquico.

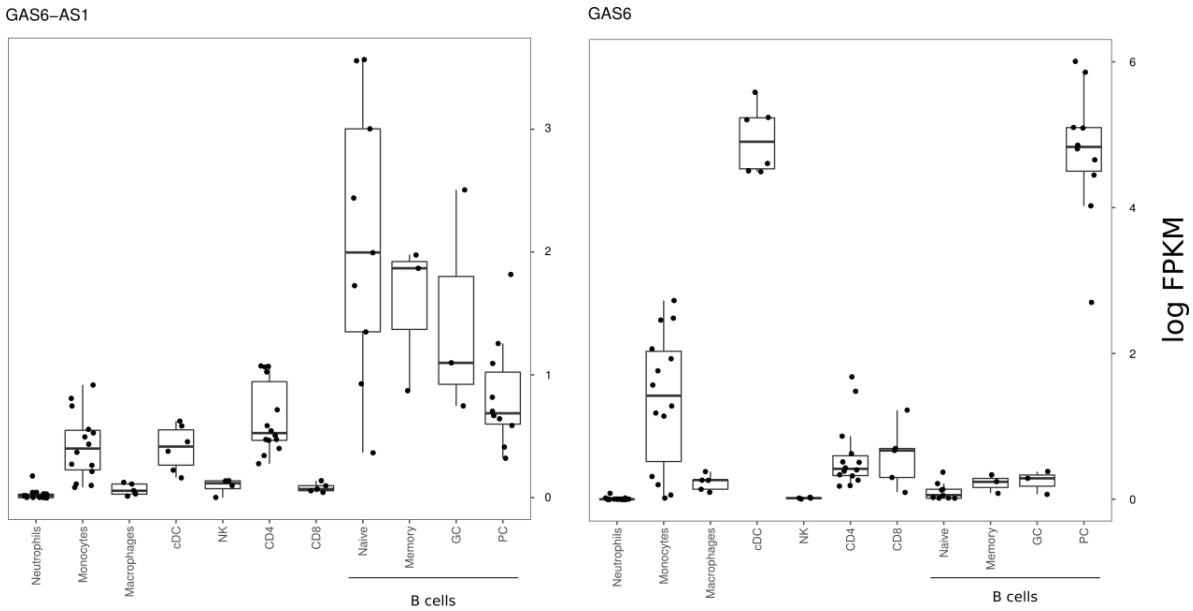

Figura 20 – Expressão de *GAS6-AS1* e *GAS6* em linhagens imunes.

4.7 COMPARAÇÃO DOS RESULTADOS E *FAM30A*

As análises conduzidas neste trabalho permitiram a detecção de diversos lncRNAs consistentemente relacionados a diferentes aspectos da resposta imunológica induzida por duas vacinas. A fim de identificar lncRNAs que apresentam mais de uma linha de evidência apoiando seu papel na resposta vacinal, fizemos uma comparação dos resultados das metanálises de expressão diferencial, correlação com anticorpos, correlação de pares de lncRNAs e mRNAs e coexpressão conduzidas nos estudos de IIVs.

O resultado mais proeminente identificado por meio dessa comparação foi o *FAM30A*, um transcrito expresso em orientação antisenso ao *locus* de Cadeias Pesadas de Imunoglobulina (*IGH*). Esse lncRNA apresenta evidências que indicam uma associação com a resposta imunológica induzida pela vacinação contra gripe. Nomeadamente, o *FAM30A* estava diferencialmente expresso na maioria dos estudos no dia 7 (Figura 8A), correlacionado positivamente com mudanças nos títulos de anticorpos (resumo = 0,291 e p-valor nominal = $5{,}96 \times 10^{-8}$) (Figura 21A), expresso preferencialmente em células B (Figura 21B) e correlacionado positivamente com sondas de transcritos do *locus IGH* (Figura 21C).

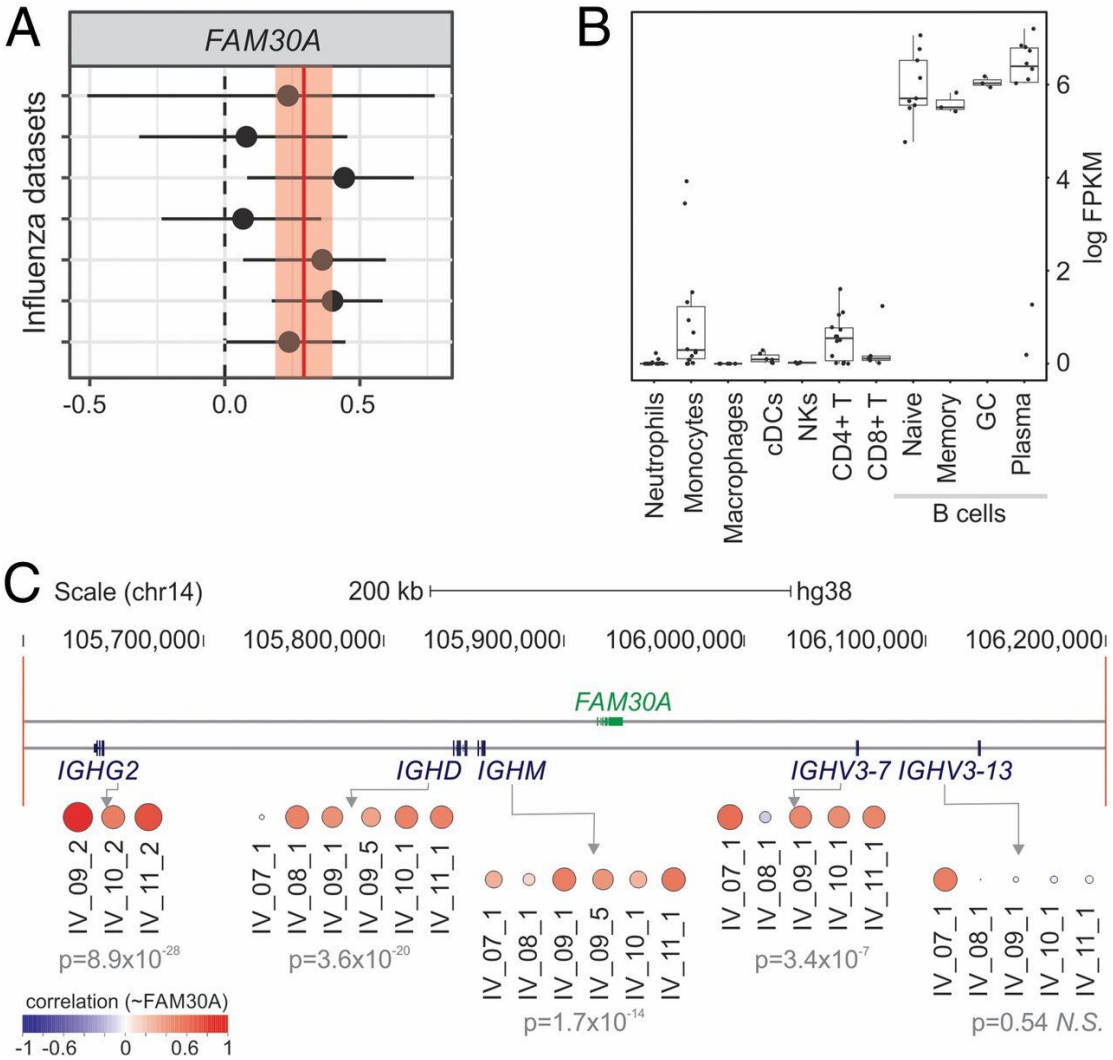

Figura 21 – Expressão de *FAM30A* correlaciona-se com a produção de anticorpos e com segmentos gênicos no *locus* IGH. (A) Correlações entre *FAM30A* e títulos de anticorpos, bem como seus intervalos de confiança de 95% (eixo x) estão representados para cada coorte (eixo y). Linha vermelha vertical e região sombreada representam o resumo das correlações e seu intervalo de confiança de 95%, respectivamente. (B) Expressão de *FAM30A* é maior em linhagens de células B (ADAMS et al., 2012). (C) Correlação entre logFCs de *FAM30A* e segmentos gênicos no *locus* IGH no dia 7 após IIVs. Os círculos representam coeficientes de correlação de Pearson em coortes imunizadas com IIVs. P-valores da metanálise de correlação estão abaixo dos identificadores das coortes.

4.8 MUDANÇAS DE EXPRESSÃO DE LNCRNAS APÓS LAIV

Nossas metanálises foram feitas com base em dados públicos de *microarray.* Não obstante, a identificação de novos transcritos pode ser mais bem conduzida com o uso da tecnologia de sequenciamento de RNAs (RNA-*seq*). A fim identificar lncRNAs relevantes para vacinação, fizemos análises de expressão diferencial em um conjunto de dados inédito de RNA-*seq* de crianças vacinadas. Foram detectados 55 lncRNAs diferencialmente expressos em sangue total no dia 2 sob um limiar de p-

valor ajustado de 0,05 (Figura 22). Diversos lncRNAs presentes na metanálise estavam também diferencialmente expressos nessa coorte, incluindo *FAM225A*, *LRRC75A-AS1*, *MAPKAPK5-AS1* e *DANCR*.

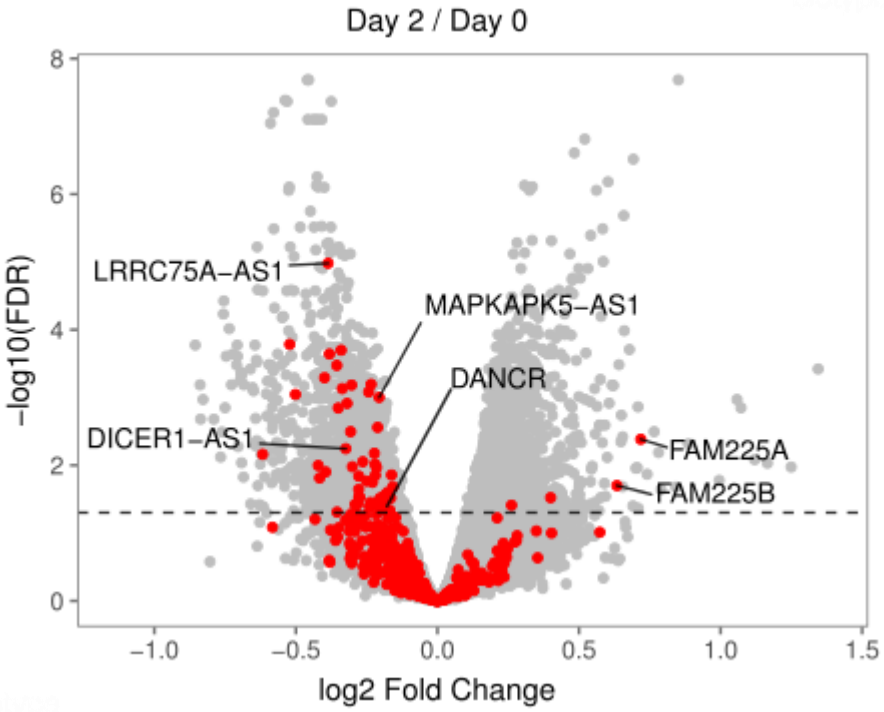

Figura 22 – *Volcano plot* dos resultados de RNA-*seq* de coorte de crianças vacinadas. Comparação entre dia 2 após a vacinação e dia 0. Cada ponto representa um gene. Pontos vermelhos representam transcritos anotados como lncRNAs.

5 DISCUSSÃO

Múltiplos processos orquestrados estão envolvidos na resposta imune à vacinação e infecção. Nossas análises revelaram que lncRNAs estão associados a alguns desses processos. Enquanto YF-17D contém um vírus atenuado, que precisa replicar antes de ativar o sistema imune de indivíduos previamente não expostos, IIVs contém proteínas do vírus influenza que potencializam a imunidade preexistente. No dia 7 após a vacinação, a assinatura transcricional periférica de YF-17D é relacionada à resposta antiviral mediada por IFN, enquanto no mesmo dia após IIVs a assinatura está associada a células B secretoras de anticorpos. Tais diferenças podem explicar a observação de diferentes grupos de lncRNAs associados com cada vacina. Não obstante, identificamos lncRNAs com expressão induzida por mais de uma vacina, o que pode apontar para a ativação de vias

relacionadas à geração da imunidade induzida por vacinas em geral.

LncRNAs cujas funções foram descritas em outros contextos, como o *MIAT*, *LINC01133* e *DANCR* foram identificados como diferencialmente expressos simultaneamente em múltiplas coortes. O *MIAT*, também conhecido como Gomafu, foi descrito em 2006 como um transcrito oriundo de uma região genômica contendo SNPs que conferem maior suscetibilidade ao infarto do miocárdio (ISHII et al., 2006). Relatos mais recentes mostram que o *MIAT* pode estar envolvido com apoptose, proliferação, oncogênese e migração celular por diferentes mecanismos (SUN et al., 2018). Sua consistente diminuição de expressão no dia 1 após IIVs em múltiplas coortes pode apontar para um papel na resposta a vacinas. Aumentos na expressão de *LINC01133* foram detectados em múltiplos cânceres (WU et al., 2017; ZENG; QIU; FENG, 2017; ZHANG; ZHU; CHEN, 2015). Em nossas análises, a expressão de LINC01133 estava elevada no dia 3 após IIVs, o que sugere que esse transcrito pode também atuar na regulação da proliferação de células imunes após a vacinação. Outros lncRNAs podem também estar associados à resposta imune adaptativa. Por exemplo, aumentos na expressão do transcrito *DANCR* foram detectados no dia 7 em jovens com alta resposta de anticorpos após IIVs (THAKAR et al., 2015). Outros estudos indicaram que *DANCR* influencia a atividade de Enhancer of Zeste Homolog 2 (*EZH2*) em múltiplos cânceres (JIA et al., 2016; LI et al., 2017b; MAO et al., 2017). Modificações epigenéticas mediadas por Ezh2 estão associadas à diferenciação de células B e produção de anticorpos em camundongos (GUO et al., 2018). Aumentos na expressão de *DANCR* após o dia 7 sugerem que este transcrito está relacionado à diferenciação de células secretoras de anticorpos, cujas frequências em sangue periférico aumentam no dia 7 após IIVs (SOBOLEV et al., 2016; TSANG et al., 2014).

Outros lncRNAs, como *MCM3AP-AS1*, *MAPKAPK5-AS1* e *LRRC75A-AS1* ainda não foram estudados a fundo. Não obstante, suas posições genômicas podem apontar para suas funções em potencial. Por exemplo, o gene *MCM3AP* codifica a proteína GANP, que está relacionada à proliferação de células B em centros germinativos (KUWAHARA et al., 2000). A diminuição de *MCM3AP-AS1* no dia 1 pode indicar que este transcrito atua na atividade de células B após imunização com IIVs. *MAPKAPK5* codifica uma serina/treonina quinase que ativa a proteína *Heat Shock Protein 27*. Esta, por sua vez, foi associada à replicação viral (RAJAIYA et al., 2012). A diminuição da expressão de *MAPKAPK5-AS1* no dia 7 após YF-17D e no

dia 2 após LAIV sugere um papel em potencial para este transcrito na replicação de vacinas virais atenuadas. Ainda nessa linha, a diminuição da expressão de *LRRC75A-AS1* nos dias 1 e 7 após YF-17D, dia 2 após LAIV, e dia 1 após IIVs sugere um papel para este transcrito na resposta de múltiplas vacinas.

Outro exemplo interessante é o *CDC42-IT1*, um transcrito totalmente intrônico (TIN) ao gene *CDC42*, este associado à regulação do ciclo celular (OLSON; ASHWORTH; HALL, 1995). Um trabalho seminal já mostrou que lncRNAs intrônicos a genes relacionados com a regulação do ciclo celular apresentam expressão elevada (ENGELHARDT; STADLER, 2012; NAKAYA et al., 2007). Outros exemplos notáveis podem gerar *insights* relacionados à imunologia de tumores e imunologia geral. O transcrito *GAS6-AS1*, por exemplo, é capaz de regular a expressão do gene codificador *GAS6* em câncer gástrico por meio de formação de um duplex mRNA-lncRNA (ZHANG et al., 2019). Nossas análises indicaram que a expressão de *GAS6-AS1* correlaciona-se negativamente com a produção de anticorpos, enquanto a expressão de seu gene codificador cognato apresenta correlação positiva. Além disso, a expressão de *GAS6-AS1* está aumentada em várias classes de células B em relação a outras linhagens imunes (Figura 20). Esses resultados podem indicar, por exemplo, um possível papel do transcrito *GAS6-AS1* na atividade de linfócitos B infiltrantes de tumores (NELSON, 2010). É importante ainda mencionar que a proteína GAS6 é agonista do receptor AXL. Múltiplas vias relacionadas à regulação da atividade de células dendríticas e células NK, bem como ao desenvolvimento, crescimento e disseminação de tumores, estão subordinadas à ativação de AXL (GAY; BALAJI; BYERS, 2017). A sinalização mediada por GAS6 foi também associada à sobrevivência de plasmócitos em nichos da medula óssea (TARTE et al., 2003). Nossos resultados sugerem, portanto, que a relação regulatória entre *GAS6-AS1* e *GAS6*, descrita previamente no estudo de câncer, pode também ser relevante para o entendimento da resposta imune humoral induzida por IIVs.

Nossa análise integrativa permitiu o levantamento de evidências que apontam para uma função em potencial do transcrito *FAM30A*. O *FAM30A* foi inicialmente identificado na linhagem celular KG-1, no baço e em sangue periférico. (NAGASE et al., 1995). Dentre as análises aqui conduzidas, a determinação da correlação entre pares de lncRNAs e mRNAs próximos mostrou-se imprescindível para a sua identificação. Outros trabalhos já propuseram que a correlação entre a expressão de lncRNAs e mRNAs de um mesmo *locus* pode ser um indicador de relações

regulatórias (DERRIEN et al., 2012; NAKAYA et al., 2007; PETRI et al., 2015). No caso específico do *FAM30A* essa ocorrência é ainda mais relevante. Isso porque lncRNAs inseridos em *loci* sujeitos a recombinação foram propostos como responsáveis por mediar as interações na cromatina necessárias para a recombinação V(D)J em camundongos (ABARRATEGUI; KRANGEL, 2007; VERMA-GAUR et al., 2012). É importante também mencionar que além de ter sido detectado em um módulo relacionado à resposta de células B em nossas análises, o *FAM30A* foi também detectado em um módulo gênico correlacionado com a resposta humoral induzida por IIVs em outro estudo (OBERMOSER et al., 2013). Em suma, nossos resultados indicam que o *FAM30A* está provavelmente relacionado à biologia de linfócitos B e com a resposta de anticorpos induzida por IIVs.

A aplicação anterior do WGCNA revelou que o transcritoma de amostras homogeneizadas de cérebro organiza-se em módulos representativos da composição celular subjacente (OLDHAM et al., 2008). Da mesma forma, nossos resultados indicaram que as comunidades consenso de coexpressão capturaram concomitantemente os sinais de coexpressão relacionados à ativação de programas transcricionais (e.g. ativação de resposta de interferon) e à covariação de transcritos marcadores de populações celulares (e.g. células T, monócitos, células B). A anotação funcional dessas comunidades, baseada na aplicação da heurística de *Guilty By Association,* permitiu a proposição de funções em potencial para lncRNAs nelas presentes (WOLFE; KOHANE; BUTTE, 2005). É importante também mencionar que o uso de BTMs na anotação das comunidades justifica-se por sua maior relevância na análise do transcritoma do sangue em relação a vias canônicas (LI et al., 2014). A análise da expressão de linhagens celulares imunes reiterou a especificidade da expressão de diversos lncRNAs presentes em comunidades, e portanto suas possíveis relevâncias funcionais na atividade destas linhagens.

Os lncRNAs aqui discutidos correspondem a apenas uma fração da informação gerada em nossas análises. Os conjuntos de dados e metadados de coortes vacinadas com IIVs e YF-17D podem ser facilmente obtidos por qualquer pessoa interessada. Não obstante, esses dados não estavam padronizados, normalizados ou ao menos organizados em torno de uma questão científica comum. Por essa razão, e a fim de facilitar o acesso da comunidade acadêmica a resultados relacionados à resposta vacinal, construímos um banco de dados que armazena os resultados de nossas metanálises (disponível em http://vaccinedb.sysbio.tools). Ao

selecionar uma vacina (IIVs ou YF-17D), uma das três análises (correlação com anticorpos, correlação de lncRNAs com genes vizinhos ou expressão diferencial) e um gene ou par gene-lncRNA, usuários podem visualizar os resultados da requisição.

Nossa abordagem mostrou-se adequada para a confirmação de que diversos lncRNAs têm expressão alterada por vacinas. Não obstante, algumas escolhas analíticas podem aperfeiçoar a detecção de lncRNAs relevantes para a resposta vacinal. As diferenças entre vacinas, exposição prévia dos indivíduos vacinados, sexo e anos de aplicação compõem grupos de fatores que, se controlados, podem gerar resultados mais precisos para a questão levantada neste trabalho, em especial para vacinas contra a gripe (NAKAYA et al., 2015). Além disso, as alterações transcricionais induzidas pela vacinação ocorrem de maneira diferente em indivíduos que apresentam diferentes níveis de resposta à vacina contra gripe (BUCASAS et al., 2011). A detecção de lncRNAs diferencialmente expressos com maior relevância para a resposta vacinal pode ser aperfeiçoada com a estratificação desses indivíduos.

As limitações deste estudo devem também ser mencionadas. Uma vez que as sondas utilizadas em plataformas de *microarray* podem ser desenhadas com base em diferentes regiões do mesmo gene, heterogeneidade nas medidas de expressão gênica entre diferentes coortes pode ser causada por uma mistura de fatores técnicos (como vieses de transcriptase reversa) e biológicos (como uso diferencial de éxons). É importante também ressaltar que correlações entre pares de lncRNAs e mRNAs próximos não necessariamente implicam a existência de relações de causalidade (GOYAL et al., 2017). Mais ainda, a ocorrência de coexpressão local pode também ser explicada por fenômenos relacionados à colocalização genômica, tais como a abertura da cromatina ou o compartilhamento de um mesmo elemento regulatório (MICHALAK, 2008). Finalmente, as funções atribuídas a transcritos anotados como lncRNAs podem ser, em realidade, exercidas por peptídeos por eles codificados (HUANG et al., 2017). Nesse contexto, propomos que os achados descritos neste trabalho apontam para a existência de relevância funcional de transcritos anotados como lncRNAs em um contexto de vacinação, mas não são evidências suficientes para alegações categóricas de funcionalidade. Essas devem ser feitas por meio da condução de experimentos funcionais cuidadosamente desenhados (KOPP; MENDELL, 2018).

6 CONCLUSÕES

Nossa análise integrativa apontou para a existência de associações entre a atividade de lncRNAs e a resposta imunológica induzida pela vacinação. A utilização de múltiplas abordagens analíticas permitiu a detecção de lncRNAs relacionados a diferentes aspectos dessa resposta. A ocorrência de alterações transcricionais similares induzidas por três diferentes vacinas evidenciou a existência de lncRNAs com potencial função compartilhada. Os resultados obtidos são coerentes com a literatura científica e apontam para a existência de múltiplos papéis ainda não apreciados de lncRNAs na resposta imune induzidas por vacinas.

Os lncRNAs aqui identificados parecem participar de processos que incluem a regulação do ciclo celular, o silenciamento epigenético, a resposta ao estresse e a recombinação V(D)J. Outros lncRNAs estavam coexpressos com genes associados a diferentes linhagens celulares e processos biológicos. Uma integração entre resultados de coexpressão, expressão diferencial e correlação com dosagem de anticorpos permitiu a detecção de modificações modulares em padrões de expressão induzidas pela vacinação com IIVs e YF-17D. A determinação dos níveis de expressão de lncRNAs em diferentes linhagens celulares reitera os resultados de nossas análises e aponta para a existência de papéis diversificados de lncRNAs na vacinação. A análise de dados inéditos de RNA-seq, obtidos de uma coorte de crianças vacinadas com LAIV, fornece mais evidências suportando nossa alegação de relevância funcional de lncRNAs na imunidade induzida por vacinas.

REFERÊNCIAS

ABARRATEGUI, I.; KRANGEL, M. S. Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination. **The EMBO journal**, v. 26, n. 20, p. 4380–90, 17 out. 2007.

ADAMS, D. et al. {BLUEPRINT} to decode the epigenetic signature written in blood. **Nature Biotechnology**, v. 30, n. 3, p. 224–226, mar. 2012.

AGIRRE, X. et al. Long non-coding RNAs discriminate the stages and gene regulatory states of human humoral immune response. **Nature Communications**, v.

10, n. 1, p. 821, 18 dez. 2019.

AHN, A. C. et al. The limits of reductionism in medicine: could systems biology offer an alternative? **PLoS medicine**, v. 3, n. 6, p. e208, maio 2006.

ALBRECHT, F. et al. {DeepBlue} epigenomic data server: programmatic data retrieval and analysis of epigenome region sets. **Nucleic Acids Research**, v. 44, n. W1, p. W581-6, 8 jul. 2016.

ALVAREZ-DOMINGUEZ, J. R. et al. Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. **Blood**, v. 123, n. 4, p. 570–81, 23 jan. 2014.

ATHALE, S. et al. Influenza vaccines differentially regulate the interferon response in human dendritic cell subsets. **Science Translational Medicine**, v. 9, n. 382, p. eaaf9194, 22 mar. 2017.

ATIANAND, M. K.; CAFFREY, D. R.; FITZGERALD, K. A. Immunobiology of Long Noncoding RNAs. **Annual Review of Immunology**, v. 35, n. 1, p. 177–198, 26 abr. 2017.

ATIANAND, M. K.; FITZGERALD, K. A. Long non-coding RNAs and control of gene expression in the immune system. **Trends in Molecular Medicine**, v. 20, n. 11, p. 623–631, nov. 2014.

AYUPE, A. C. et al. Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. **RNA biology**, v. 12, n. 8, p. 877–92, 2015.

BARABÁSI, A.-L.; ALBERT, R. Emergence of Scaling in Random Networks. **Science**, v. 286, n. 5439, p. 509–512, 15 out. 1999.

BARABÁSI, A.-L.; GULBAHCE, N.; LOSCALZO, J. Network medicine: a network-based approach to human disease. **Nature Reviews Genetics**, v. 12, n. 1, p. 56–68, 17 jan. 2011.

BARABÁSI, A.-L.; OLTVAI, Z. N. Network biology: understanding the cell’s functional organization. **Nature reviews. Genetics**, v. 5, n. 2, p. 101–113, 2004.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. **Gephi : An Open Source Software for Exploring and Manipulating Networks Visualization and Exploration of Large Graphs**. [s.l: s.n.]. Disponível em: <www.aaai.org>. Acesso em: 13 jun. 2019.

BECKEDORFF, F. C. et al. The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. **PLoS genetics**, v. 9, n. 8, p. e1003705, 2013.

BHATT, K. et al. Quest for correlates of protection against tuberculosis. **Clinical and vaccine immunology : CVI**, v. 22, n. 3, p. 258–66, mar. 2015.

BRAZÃO, T. F. et al. Long noncoding RNAs in B-cell development and activation. **Blood**, v. 128, n. 7, p. e10–e19, 18 ago. 2016.

BROWN, C. J. et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. **Nature**, v. 349, n. 6304, p. 38–44, 3 jan. 1991.

BUCASAS, K. L. et al. Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans. **The Journal of infectious diseases**, v. 203, n. 7, p. 921–9, 1 abr. 2011.

CABILI, M. N. et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. **Genes & Development**, v. 25, n. 18, p. 1915–1927, 15 set. 2011.

CHAN, K. R. et al. Metabolic perturbations and cellular stress underpin susceptibility to symptomatic live-attenuated yellow fever infection. **Nature Medicine**, v. 25, n. 8, p. 1218–1224, 15 ago. 2019.

CHAUSSABEL, D. et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. **Immunity**, v. 29, n. 1, p. 150–164, 18 jul. 2008.

CHEN, Y. G.; SATPATHY, A. T.; CHANG, H. Y. Gene regulation in the immune system by long noncoding RNAs. **Nature Immunology**, v. 18, n. 9, p. 962–972, 22 ago. 2017.

CHIKINA, M.; ZASLAVSKY, E.; SEALFON, S. C. CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. **Bioinformatics**, v. 31, n. 10, p. 1584–1591, 15 maio 2015.

CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. **InterJournal**, v. Complex Sy, p. 1695, 2006.

D'ARGENIO, D. A.; WILSON, C. B. A Decade of Vaccines: Integrating Immunology and Vaccinology for Rational Vaccine Design. **Immunity**, v. 33, n. 4, p. 437–440, 29 out. 2010.

DAVIS, M. M.; TATO, C. M.; FURMAN, D. Systems immunology: just getting started. **Nature Immunology**, v. 18, n. 7, p. 725–732, 20 jun. 2017.

DE GREGORIO, E.; RAPPUOLI, R. From empiricism to rational design: a personal perspective of the evolution of vaccine development. **Nature Reviews Immunology**, v. 14, n. 7, p. 505–514, 13 jul. 2014.

DERRIEN, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. **Genome research**, v. 22, n. 9, p. 1775–89, set. 2012.

DJEBALI, S. et al. Landscape of transcription in human cells. **Nature**, v. 489, n. 7414, p. 101–108, 5 set. 2012.

DU, Z. et al. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. **Nature structural & molecular biology**, v. 20, n. 7, p. 908–13, jul. 2013.

EISEN, M. B. et al. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 25, p. 14863–8, 8 dez. 1998.

ENGELHARDT, J.; STADLER, P. F. Hidden treasures in unspliced EST data. **Theory in Biosciences**, v. 131, n. 1, p. 49–57, 8 maio 2012.

FERNANDES, J. C. R. et al. Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. **Non-coding RNA**, v. 5, n. 1, 17 fev. 2019.

FURMAN, D. et al. Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination. **Proceedings of the National Academy of Sciences**, v. 111, n. 2, p. 869–874, 14 jan. 2014.

FURMAN, D. et al. Cytomegalovirus infection enhances the immune response to influenza. **Science translational medicine**, v. 7, n. 281, p. 281ra43, 1 abr. 2015.

GAITERI, C. et al. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. **Genes, Brain and Behavior**, v. 13, n. 1, p. 13–24, 1 jan. 2014.

GAUCHER, D. et al. Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. **The Journal of experimental medicine**, v. 205, n. 13, p. 3119–31, 22 dez. 2008.

GAY, C. M.; BALAJI, K.; BYERS, L. A. Giving AXL the axe: targeting AXL in human malignancy. **British Journal of Cancer**, v. 116, n. 4, p. 415–423, 10 fev. 2017.

GENTLEMAN, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biology**, v. 5, n. 10, p. R80, jan. 2004.

GOVINDARAJAN, R. et al. Microarray and its applications. **Journal of pharmacy & bioallied sciences**, v. 4, n. Suppl 2, p. S310-2, ago. 2012.

GOYAL, A. et al. A cautionary tale of sense-antisense gene pairs: independent regulation despite inverse correlation of expression. **Nucleic Acids Research**, v. 45, n. 21, p. 12496–12508, 1 dez. 2017.

GU, Z.; EILS, R.; SCHLESNER, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847–2849, 15 set. 2016.

GUO, M. et al. {EZH2} Represses the B Cell Transcriptional Program and Regulates Antibody-Secreting Cell Metabolism and Antibody Production. **Journal of Immunology**, v. 200, n. 3, p. 1039–1052, fev. 2018.

HAGAN, T. et al. Systems vaccinology: Enabling rational vaccine design with systems biological approaches. **Vaccine**, v. 33, n. 40, p. 5294–301, 6 abr. 2015.

HAGAN, T. et al. Antibiotics-Driven Gut Microbiome Perturbation Alters Immunity to Vaccines in Humans. **Cell**, v. 178, n. 6, p. 1313- 1328.e13, 5 set. 2019.

HARTWELL, L. H. et al. From molecular to modular cell biology. **Nature**, v. 402, n. S6761, p. C47–C52, 2 dez. 1999.

HIGGINS, J.; GREEN, S. **Cochrane Handbook for Systematic Reviews of InterventionsCochrane Handbook for Systematic Reviews of Interventions Version 5.1.0**, 2011.

HON, C.-C. et al. An atlas of human long non-coding RNAs with accurate 5′ ends. **Nature**, v. 543, n. 7644, p. 199–204, 1 mar. 2017.

HOU, J. et al. A Systems Vaccinology Approach Reveals Temporal Transcriptomic Changes of Immune Responses to the Yellow Fever 17D Vaccine. **Journal of immunology (Baltimore, Md. : 1950)**, v. 199, n. 4, p. 1476–1489, 15 ago. 2017.

HSIAO, L.-L. et al. Correcting for Signal Saturation Errors in the Analysis of Microarray Data. **BioTechniques**, v. 32, n. 2, p. 330–336, fev. 2002.

HUANG, J.-Z. et al. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. **Molecular cell**, v. 68, n. 1, p. 171- 184.e6, 5 out. 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**, v. 53, n. 3, p. 393–406, 6 fev. 2014.

IOANNIDIS, J. P. Microarrays and molecular research: noise discovery? **The Lancet**, v. 365, n. 9458, p. 454–455, fev. 2005.

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, 1 abr. 2003.

ISHII, N. et al. Identification of a novel non-coding {RNA}, {MIAT}, that confers risk of myocardial infarction. **Journal of Human Genetics**, v. 51, n. 12, p. 1087–1099, out. 2006.

JEONG, H. et al. Lethality and centrality in protein networks. **Nature**, v. 411, n. 6833, p. 41–42, maio 2001.

JIA, J. et al. Long noncoding {RNA} {DANCR} promotes invasion of prostate cancer through epigenetically silencing expression of {TIMP2}/3. **Oncotarget**, v. 7, n. 25, p. 37868–37881, jun. 2016.

JOHNSON, R. Long non-coding RNAs in Huntington's disease neurodegeneration. **Neurobiology of disease**, v. 46, n. 2, p. 245–54, maio 2012.

KAUFFMANN, A.; GENTLEMAN, R.; HUBER, W. arrayQualityMetrics - A bioconductor package for quality assessment of microarray data. **Bioinformatics**, v. 25, n. 3, p. 415–416, 2009.

KENT, W. J. BLAT---The BLAST-Like Alignment Tool. **Genome Research**, v. 12, n. 4, p. 656–664, 20 mar. 2002.

KENT, W. J. et al. The human genome browser at UCSC. **Genome research**, v. 12, n. 6, p. 996–1006, 1 jun. 2002.

KITANO, H. Computational systems biology. **Nature**, v. 420, n. 6912, p. 206–210, 14 nov. 2002.

KOFF, W. C.; GUST, I. D.; PLOTKIN, S. A. Toward a Human Vaccines Project. **Nature Immunology**, v. 15, n. 7, p. 589–592, 18 jun. 2014.

KOPP, F.; MENDELL, J. T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. **Cell**, v. 172, n. 3, p. 393–407, 25 jan. 2018.

KOSTER, J.; RAHMANN, S. Snakemake--a scalable bioinformatics workflow engine. **Bioinformatics**, v. 28, n. 19, p. 2520–2522, 1 out. 2012.

KUNG, J. T. Y.; COLOGNORI, D.; LEE, J. T. Long Noncoding RNAs: Past, Present, and Future. **Genetics**, v. 193, n. 3, p. 651–669, 1 mar. 2013.

KUWAHARA, K. et al. A novel nuclear phosphoprotein, {GANP}, is up-regulated in centrocytes of the germinal center and associated with {MCM3}, a protein essential for {DNA} replication. **Blood**, v. 95, n. 7, p. 2321–2328, abr. 2000.

LANGFELDER, P.; HORVATH, S. WGCNA: an R package for weighted correlation network analysis. **BMC bioinformatics**, v. 9, p. 559, dez. 2008.

LAWRENCE, M. et al. Software for computing and annotating genomic ranges. **{PLoS} Computational Biology**, v. 9, n. 8, p. e1003118, 8 ago. 2013.

LEE, H. K. et al. Coexpression analysis of human genes across many microarray data sets. **Genome research**, v. 14, n. 6, p. 1085–94, jun. 2004.

LEVITZ, S. M.; GOLENBOCK, D. T. Beyond empiricism: informing vaccine development through innate immunity research. **Cell**, v. 148, n. 6, p. 1284–92, 16 mar. 2012.

LI, S. et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. **Nature Immunology**, v. 15, n. 2, p. 195–204, fev. 2014.

LI, S. et al. Metabolic Phenotypes of Response to Vaccination in Humans. **Cell**, v. 169, n. 5, p. 862- 877.e17, 18 maio 2017a.

LI, Z. et al. The degradation of {EZH2} mediated by {lncRNA} {ANCR} attenuated the invasion and metastasis of breast cancer. **Cell Death and Differentiation**, v. 24, n. 1, p. 59–71, 2017b.

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, 1 abr. 2014.

LIN, H.-N.; HSU, W.-L. DART: a fast and accurate RNA-seq mapper with a partitioning strategy. **Bioinformatics**, v. 34, n. 2, p. 190–197, 15 jan. 2018.

LIU, H.; BEBU, I.; LI, X. Microarray probes and probe sets. **Frontiers in bioscience (Elite edition)**, v. 2, p. 325–38, 1 jan. 2010.

MALLADI, V. S. et al. Ontology application and use at the ENCODE DCC. **Database : the journal of biological databases and curation**, v. 2015, 2015.

MAO, Z. et al. LncRNA DANCR promotes migration and invasion through suppression of lncRNA-LET in gastric cancer cells. **Bioscience reports**, v. 37, n. 6, 22 dez. 2017.

MATTICK, J. S.; RINN, J. L. Discovery and annotation of long noncoding RNAs. **Nature Structural & Molecular Biology**, v. 22, n. 1, p. 5–7, 1 jan. 2015.

MCHEYZER-WILLIAMS, L. J.; MCHEYZER-WILLIAMS, M. G. ANTIGEN-SPECIFIC MEMORY B CELL DEVELOPMENT. **Annual Review of Immunology**, v. 23, n. 1, p. 487–513, abr. 2005.

MENG, T. C.; SOMANI, S.; DHAR, P. Modeling and simulation of biological systems with stochasticity. **In silico biology**, v. 4, n. 3, p. 293–309, 2004.

MICHALAK, P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. **Genomics**, v. 91, n. 3, p. 243–248, 1 mar. 2008.

MICHELHAUGH, S. K. et al. Mining Affymetrix microarray data for long non-coding {RNAs}: altered expression in the nucleus accumbens of heroin abusers. **Journal of Neurochemistry**, v. 116, n. 3, p. 459–466, fev. 2011.

MICHIELS, S.; KOSCIELNY, S.; HILL, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. **The Lancet**, v. 365, n. 9458, p. 488–492, fev. 2005.

NAGASE, T. et al. Prediction of the coding sequences of unidentified human genes. {IV}. The coding sequences of 40 new genes ({KIAA0121}-{KIAA0160}) deduced by analysis of {cDNA} clones from human cell line {KG}-1. **{DNA} Research**, v. 2, n. 4, p. 167- 174,199, ago. 1995.

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.

NAKAYA, H. I. et al. Systems biology of vaccination for seasonal influenza in

humans. **Nature Immunology**, v. 12, n. 8, p. 786–795, 10 jul. 2011.

NAKAYA, H. I. et al. Systems Analysis of Immunity to Influenza Vaccination across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures. **Immunity**, v. 43, n. 6, p. 1186–1198, 15 dez. 2015.

NAKAYA, H. I.; PULENDRAN, B. Vaccinology in the era of high-throughput biology. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 370, n. 1671, 19 jun. 2015.

NELSON, B. H. {CD20}+ B cells: the other tumor-infiltrating lymphocytes. **Journal of Immunology**, v. 185, n. 9, p. 4977–4982, 1 nov. 2010.

OBERMOSER, G. et al. Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. **Immunity**, v. 38, n. 4, p. 831–844, 18 abr. 2013.

OH, J. Z. et al. TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. **Immunity**, v. 41, n. 3, p. 478–492, 18 set. 2014.

OLDHAM, M. C. et al. Functional organization of the transcriptome in human brain. **Nature neuroscience**, v. 11, n. 11, p. 1271–82, nov. 2008.

OLSON, M. F.; ASHWORTH, A.; HALL, A. An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. **Science (New York, N.Y.)**, v. 269, n. 5228, p. 1270–2, 1 set. 1995.

ORENSTEIN, W. A. et al. Field evaluation of vaccine efficacy. **Bulletin of the World Health Organization**, v. 63, n. 6, p. 1055–68, 1985.

PELLEGRINA, D. et al. Insights into the Function of Long Noncoding RNAs in Sepsis Revealed by Gene Co-Expression Network Analysis. **Non-Coding RNA**, v. 3, n. 4, p. 5, 26 jan. 2017.

PETRI, A. et al. Long Noncoding RNA Expression during Human B-Cell Development. **PLOS ONE**, v. 10, n. 9, p. e0138236, 22 set. 2015.

PLOTKIN, S. A. Vaccines: Correlates of Vaccine-Induced Immunity. **Clinical Infectious Diseases**, v. 47, n. 3, p. 401–409, 1 ago. 2008.

PLOTKIN, S.; PLOTKIN, S. A Short History of Vaccination. In: [s.l: s.n.]. p. 1-15.e8.

POLAND, G. A. et al. Heterogeneity in Vaccine Immune Response: The Role of Immunogenetics and the Emerging Field of Vaccinomics. **Clinical Pharmacology & Therapeutics**, v. 82, n. 6, p. 653–664, 31 dez. 2007.

POLAND, G. A. et al. Vaccinomics, adversomics, and the immune response network theory: individualized vaccinology in the 21st century. **Seminars in Immunology**, v. 25, n. 2, p. 89–103, abr. 2013.

PULENDRAN, B. Systems vaccinology: Probing humanity's diverse immune systems with vaccines. **Proceedings of the National Academy of Sciences**, v. 111, n. 34, p. 12300–12306, 18 ago. 2014.

PULENDRAN, B.; AHMED, R. Translating innate immunity into immunological memory: implications for vaccine development. **Cell**, v. 124, n. 4, p. 849–63, 24 fev. 2006.

PULENDRAN, B.; AHMED, R. Immunological mechanisms of vaccination. **Nature Immunology**, v. 12, n. 6, p. 509–517, jun. 2011.

QUEREC, T. D. et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. **Nature immunology**, v. 10, n. 1, p. 116–25, 23 jan. 2009.

QUINLAN, A. R.; HALL, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. **Bioinformatics**, v. 26, n. 6, p. 841–842, 15 mar. 2010.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**Vienna, Austria, 2015. Disponível em: <https://www.r-project.org/>

RAJAIYA, J. et al. Heat Shock Protein 27 Mediated Signaling in Viral Infection. **Biochemistry**, v. 51, n. 28, p. 5695–5702, 17 jul. 2012.

RANZANI, V. et al. The long intergenic noncoding RNA landscape of human

lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. **Nature immunology**, v. 16, n. 3, p. 318–325, mar. 2015.

RAPPUOLI, R. et al. Vaccines, new opportunities for a new society. **Proceedings of the National Academy of Sciences**, v. 111, n. 34, p. 12288–12293, 18 ago. 2014a.

RAPPUOLI, R. et al. Vaccines, new opportunities for a new society. **Proceedings of the National Academy of Sciences**, v. 111, n. 34, p. 12288–12293, 18 ago. 2014b.

REICHARDT, J.; BORNHOLDT, S. Statistical mechanics of community detection. **Physical Review E**, v. 74, n. 1, jul. 2006.

REIS, E. M. et al. Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. **Oncogene**, v. 23, n. 39, p. 6684–6692, 28 ago. 2004.

RIEDEL, S. Edward Jenner and the history of smallpox and vaccination. **Proceedings (Baylor University. Medical Center)**, v. 18, n. 1, p. 21, 2005.

RILEY, R. D.; HIGGINS, J. P. T.; DEEKS, J. J. Interpretation of random effects meta-analyses. **BMJ (Clinical research ed.)**, v. 342, p. d549, 10 fev. 2011.

RINN, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. **Cell**, v. 129, n. 7, p. 1311–23, 2007.

RITCHIE, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. **Nucleic Acids Research**, v. 43, n. 7, p. e47, 20 jan. 2015.

RUNG, J.; BRAZMA, A. Reuse of public genome-wide gene expression data. **Nature Reviews Genetics**, v. 14, n. 2, p. 89–99, 27 fev. 2013.

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, 20 dez. 2018.

SANDBERG, R.; LARSSON, O. Improved precision and accuracy for microarrays using updated probe set definitions. **{BMC} Bioinformatics**, v. 8, p. 48, fev. 2007.

SAUER, U.; HEINEMANN, M.; ZAMBONI, N. GENETICS: Getting Closer to the Whole Picture. **Science**, v. 316, n. 5824, p. 550–551, 27 abr. 2007.

SERGUSHICHEV, A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. **BioRxiv**, p. 060012, 20 jun. 2016.

SIEGRIST, C.-A. 2 - Vaccine immunology. In: PLOTKIN, S. A.; ORENSTEIN, W. A.; OFFIT, P. A. (Eds.). . **Vaccines (Sixth Edition)**. Sixth Edit ed. London: W.B. Saunders, 2013. p. 14–32.

SOBOLEV, O. et al. Adjuvanted influenza-H1N1 vaccination reveals lymphoid signatures of age-dependent early responses and of clinical adverse events. **Nature immunology**, v. 17, n. 2, p. 204–13, fev. 2016.

ST LAURENT, G.; WAHLESTEDT, C.; KAPRANOV, P. The Landscape of long noncoding RNA classification. **Trends in genetics : TIG**, v. 31, n. 5, p. 239–51, maio 2015.

STEIN, D. F. et al. Gene expression profiles are different in venous and capillary blood: Implications for vaccine studies. **Vaccine**, v. 34, n. 44, p. 5306–5313, out. 2016.

SUN, C. et al. Long non-coding {RNA} {MIAT} in development and disease: a new player in an old game. **Journal of Biomedical Science**, v. 25, n. 1, p. 23, mar. 2018.

TARTE, K. et al. Gene expression profiling of plasma cells and plasmablasts: toward a better understanding of the late stages of B-cell differentiation. **Blood**, v. 102, n. 2, p. 592–600, 15 jul. 2003.

THAKAR, J. et al. Aging-dependent alterations in gene expression and a mitochondrial signature of responsiveness to human influenza vaccination. **Aging**, v. 7, n. 1, p. 38–52, jan. 2015.

TOGNOTTI, E. The eradication of smallpox, a success story for modern medicine and public health: What lessons for the future? **The Journal of Infection in Developing Countries**, v. 4, n. 05, p. 264–266, 15 maio 2010.

TSANG, J. S. et al. Global analyses of human immune variation reveal baseline

predictors of postvaccination responses. **Cell**, v. 157, n. 2, p. 499–513, 10 abr. 2014.

TSENG, G. C.; GHOSH, D.; FEINGOLD, E. Comprehensive literature review and statistical considerations for microarray meta-analysis. **Nucleic acids research**, v. 40, n. 9, p. 3785–99, maio 2012.

VAILATI-RIBONI, M.; PALOMBO, V.; LOOR, J. J. What Are Omics Sciences? In: **Periparturient Diseases of Dairy Cows**. Cham: Springer International Publishing, 2017. p. 1–7.

VERMA-GAUR, J. et al. Noncoding transcription within the Igh distal V(H) region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 42, p. 17004–9, 16 out. 2012.

VIECHTBAUER, W. Conducting Meta-Analyses in R with the metafor Package. **Journal of statistical software**, v. 36, n. 3, 2010.

WALSH, C. et al. Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. **Microarrays**, v. 4, n. 3, p. 389–406, 21 ago. 2015.

WANG, P. et al. The STAT3-Binding Long Noncoding RNA lnc-DC Controls Human Dendritic Cell Differentiation. **Science**, v. 344, n. 6181, p. 310–313, 18 abr. 2014.

WANG, Y. et al. Long noncoding RNA derived from CD244 signaling epigenetically controls CD8+ T-cell immune responses in tuberculosis infection. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 29, p. E3883-92, 6 jul. 2015.

WEIRAUCH, M. T. Gene Coexpression Networks for the Analysis of DNA Microarray Data. In: **Applied Statistics for Network Biology**. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2011. p. 215–250.

WOLFE, C. J.; KOHANE, I. S.; BUTTE, A. J. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. **BMC bioinformatics**, v. 6, p. 227, 14 set. 2005.

WU, X.-S. et al. {LncRNA}-{PAGBC} acts as a {microRNA} sponge and promotes gallbladder tumorigenesis. **{EMBO} Reports**, v. 18, n. 10, p. 1837–1853, set. 2017.

YAN, P. et al. Cis - and trans -acting lncRNAs in pluripotency and reprogramming. **Current Opinion in Genetics & Development**, v. 46, p. 170–178, out. 2017.

YU, G. et al. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. **OMICS: A Journal of Integrative Biology**, v. 16, n. 5, p. 284–287, maio 2012.

ZEMMOUR, D. et al. *Flicr* , a long noncoding RNA, modulates Foxp3 expression and autoimmunity. **Proceedings of the National Academy of Sciences**, v. 114, n. 17, p. E3472–E3480, 25 abr. 2017.

ZENG, H.-F.; QIU, H.-Y.; FENG, F.-B. Long Noncoding {RNA} {LINC01133} Sponges {miR}-422a to Aggravate the Tumorigenesis of Human Osteosarcoma. **Oncology Research**, mar. 2017.

ZHANG, B.; HORVATH, S. A general framework for weighted gene co-expression network analysis. **Statistical applications in genetics and molecular biology**, v. 4, p. Article17, 2005.

ZHANG, E. Y.; KONG, K.-F.; ALTMAN, A. The yin and yang of protein kinase C-theta (PKCθ): a novel drug target for selective immunosuppression. **Advances in pharmacology (San Diego, Calif.)**, v. 66, p. 267–312, 2013.

ZHANG, H. et al. Profiling of human CD4+ T-cell subsets identifies the TH2-specific noncoding RNA GATA3-AS1. **Journal of Allergy and Clinical Immunology**, v. 132, n. 4, p. 1005–1008, out. 2013.

ZHANG, J.; ZHU, N.; CHEN, X. A novel long noncoding {RNA} {LINC01133} is upregulated in lung squamous cell cancer and predicts survival. **Tumour Biology**, v. 36, n. 10, p. 7465–7471, set. 2015.

ZHANG, P. et al. Long non-coding antisense RNA GAS6-AS1 supports gastric cancer progression via increasing GAS6 expression. **Gene**, v. 696, p. 1–9, 15 maio 2019.

ZHANG, X. et al. Long non-coding RNA expression profiles predict clinical phenotypes in glioma. **Neurobiology of disease**, v. 48, n. 1, p. 1–8, out. 2012.

ZHAO, F. et al. Knockdown of a novel lincRNA AATBC suppresses proliferation and induces apoptosis in bladder cancer. **Oncotarget**, v. 6, n. 2, p. 1064–78, 20 jan. 2015.

ZHAO, S. et al. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. **PLoS ONE**, v. 9, n. 1, p. e78644, 16 jan. 2014.

ANEXOS

Janus - Sistema Administrativo da Pós-Graduação

Universidade de São Paulo
Faculdade de Ciências Farmacêuticas
Documento sem validade oficial
FICHA DO ALUNO

9142 - 4481634/1 - Diógenes Saulo de Lima

Email:	diogenes.lima@usp.br
Data de Nascimento:	08/06/1990
Cédula de Identidade:	RG - 46.148.284-8 - SP
Local de Nascimento:	Estado de São Paulo
Nacionalidade:	Brasileira
Graduação:	Farmacêutico-Bioquímico - Faculdade de Ciências Farmacêuticas - Universidade de São Paulo - São Paulo - Brasil - 2015

Curso:	Doutorado Direto
Programa:	Farmácia (Fisiopatologia e Toxicologia)
Área:	Fisiopatologia
Data de Matrícula:	12/01/2016
Início da Contagem de Prazo:	12/01/2016
Data Limite para o Depósito:	13/01/2020
Orientador:	Prof(a). Dr(a). Helder Takashi Imoto Nakaya - 20/02/2017 até o presente. Email: hnakaya@gmail.com
Proficiência em Línguas:	Inglês, Aprovado em 12/01/2016
Data de Aprovação no Exame de Qualificação:	Aprovado em 14/03/2018
Data do Depósito do Trabalho:	
Título do Trabalho:	
Data Máxima para Aprovação da Banca:	
Data de Aprovação da Banca:	
Data Máxima para Defesa:	
Data da Defesa:	
Resultado da Defesa:	
Histórico de Ocorrências:	Primeira Matrícula em 12/01/2016

Aluno matriculado no Regimento da Pós-Graduação USP (Resolução nº 6542 em vigor de 20/04/2013 até 28/03/2018).
Última ocorrência: Matrícula de Acompanhamento em 15/07/2019
Impresso em: 02/01/2020 19:42:20

Janus - Sistema Administrativo da Pós-Graduação

Universidade de São Paulo
Faculdade de Ciências Farmacêuticas
Documento sem validade oficial
FICHA DO ALUNO

9142 - 4481634/1 - Diógenes Saulo de Lima

Sigla	Nome da Disciplina	Início	Término	Carga Horária	Cred.	Freq.	Conc.	Exc.	Situação
FBC5792-3/3	Tópicos em Análises Clínicas III	08/03/2016	20/06/2016	15	1	87	A	N	Concluída
BMI5862-9/2	Seminário Didático-Cientifico em Imunologia I (Instituto de Ciências Biomédicas - Universidade de São Paulo)	10/03/2016	30/06/2016	60	4	100	A	N	Concluída
FBC5748-4/2	Trabalhos Científicos: da Elaboração à Publicação	05/04/2016	17/05/2016	60	4	90	A	N	Concluída
FBC5766-4/3	Tópicos em Análises Clínicas IV	02/08/2016	14/11/2016	15	1	100	A	N	Concluída
BMI5863-8/2	Seminário Didático-Cientifico em Imunologia II (Instituto de Ciências Biomédicas - Universidade de São Paulo)	18/08/2016	30/11/2016	60	4	100	A	N	Concluída
PCS5031-1/2	Introdução à Ciência dos Dados (Escola Politécnica - Universidade de São Paulo)	12/09/2016	04/12/2016	120	8	100	A	N	Concluída
FBC5707-6/1	Biologia de Sistemas para Ciências da Vida	26/10/2016	08/12/2016	60	4	100	A	N	Concluída
FBC5710-7/1	Vacinas e Adjuvantes	13/11/2017	17/12/2017	60	0	-	-	N	Matrícula cancelada
FBC5710-8/1	Vacinas e Adjuvantes	19/11/2018	09/12/2018	90	6	80	A	N	Concluída

	Créditos mínimos exigidos		Créditos obtidos
	Para exame de qualificação	Para depósito de tese	
Disciplinas:	0	25	32
Estágios:			
Total:	0	25	32

Créditos Atribuídos à Tese: 167

Conceito a partir de 02/01/1997:

A - Excelente, com direito a crédito; B - Bom, com direito a crédito; C - Regular, com direito a crédito; R - Reprovado; T - Transferência.

Um(1) crédito equivale a 15 horas de atividade programada.

Última ocorrência: Matrícula de Acompanhamento em 15/07/2019
Impresso em: 02/01/2020 19:42:20

Diógenes Saulo de Lima

Endereço para acessar este CV: http://lattes.cnpq.br/3974838381500683
ID Lattes: **3974838381500683**
Última atualização do currículo em 02/01/2020

Doutorado em andamento no Laboratório de Biologia de Sistemas Computacional (CSBL) da Faculdade de Ciências Farmacêuticas da Universidade de São Paulo (FCF-USP). Graduado em Farmácia-Bioquímica na Universidade de São Paulo (2015). Iniciação Científica concluída em Biologia de Sistemas (2015) e Iniciação Científica concluída em Toxicologia (2012). Possui experiência em bioinformática. **(Texto informado pelo autor)**

Identificação

Nome	Diógenes Saulo de Lima
Nome em citações bibliográficas	LIMA, D. S.;DE LIMA, DIÓGENES SAULO;LIMA, DIÓGENES SAULO;LIMA, DIOGENES SAULO DE;LIMA, DIÓGENES S.;DE LIMA, DIÓGENES S.
Lattes iD	http://lattes.cnpq.br/3974838381500683

Endereço

Formação acadêmica/titulação

2016	Doutorado em andamento em Farmácia (Fisiopatologia e Toxicologia) (Conceito CAPES 7). Universidade de São Paulo, USP, Brasil. Título: RNAs Não Codificadores Longos Associados à imunidade Induzida por Vacinação, Orientador: Helder Takashi Imoto Nakaya. Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil. Palavras-chave: lncRNAs; Vacinologia de Sistemas; Bioinformática. Grande área: Ciências Biológicas Grande Área: Ciências Biológicas / Área: Imunologia.
2010 - 2015	Graduação em Farmácia e Bioquímica. Universidade de São Paulo, USP, Brasil. Título: IDENTIFICAÇÃO DE MÓDULOS GÊNICOS DE COEXPRESSÃO EM PROCESSOS INFLAMATÓRIOS. Orientador: Helder Takashi Imoto Nakaya.

Formação Complementar

2018 - 2018	São Paulo School of Advanced Science on Vaccines. (Carga horária: 90h). Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, Brasil.
2017 - 2017	BRAZIL ADVANCED COURSE ON VACCINES. (Carga horária: 40h). International Union of Immunological Societies, IUIS, Austria.
2015 - 2015	Curso de Inverno em Imunologia. (Carga horária: 80h). Universidade de São Paulo, USP, Brasil.
2015 - 2015	Curso de Verão em Bioinformática. (Carga horária: 38h). Associação Brasileira de Bioinformática e Biologia Computacional, AB3C, Brasil.
2014 - 2014	Bioestatística: Análise de Dados Biológicos. (Carga horária: 12h). Faculdade de Ciências Farmacêuticas - USP, FCF-USP, Brasil.
2014 - 2014	Estágio Avançado de Bioquímica/Biologia Molecular. (Carga horária: 180h). Instituto de Química - USP, IQ-USP, Brasil.
2013 - 2013	Uso de modelos animais em transt. psiquiátricos. (Carga horária: 4h). Faculdade de Ciências Farmacêuticas - USP, FCF-USP, Brasil.
2013 - 2013	Farmacoeconomia. (Carga horária: 8h). Faculdade de Ciências Farmacêuticas - USP, FCF-USP, Brasil.

Atuação Profissional

Faculdade de Ciências Farmacêuticas - USP, FCF-USP, Brasil.

Vínculo institucional
2017 - 2017 — Vínculo: Bolsista, Enquadramento Funcional: Monitor, Carga horária: 6

Vínculo institucional
2015 - 2015 — Vínculo: Livre, Enquadramento Funcional: Iniciação Científica, Carga horária: 30
Outras informações — Atuação em projeto de iniciação científica com título "Desenvolvimento de módulos gênicos de inflamação com utilização de bancos de dados de acesso livre" na área de Biologia de Sistemas sob orientação do Professor Doutor Helder Takashi Imoto Nakaya.

Vínculo institucional
2015 - 2015 — Vínculo: Livre, Enquadramento Funcional: Monitor, Carga horária: 4
Outras informações — Monitoria na disciplina de Imunodiagnóstico - Curso de Farmácia-Bioquímica

Vínculo institucional
2011 - 2012 — Vínculo: Bolsista, Enquadramento Funcional: Iniciação Científica, Carga horária: 20
Outras informações — Desenvolvimento de projeto de iniciação científica com título: 'Determinação de barbitúricos em amostra postmortem de fígado humano por cromatografia gasosa acoplada à espectrometria de massas' sob orientação da Dra. Carolina Dizioli Rodrigues de Oliveira. Auxílio Financeiro: Bolsa Institucional RUSP

Áreas de atuação

1. Grande área: Ciências Biológicas / Área: Imunologia / Subárea: Bioinformática.
2. Grande área: Ciências da Saúde / Área: Farmácia / Subárea: Avaliação e analises toxicológicas.
3. Grande área: Ciências Biológicas / Área: Imunologia.

Idiomas

Português — Compreende Bem, Fala Bem, Lê Bem, Escreve Bem.
Inglês — Compreende Bem, Fala Bem, Lê Bem, Escreve Bem.
Espanhol — Compreende Razoavelmente, Fala Pouco, Lê Razoavelmente, Escreve Pouco.

Produções

Produção bibliográfica

Artigos completos publicados em periódicos

Ordenar por

Ordem Cronológica ▼

1. GONZALEZ-DIAS, PATRÍCIA ; LEE, EVA K. ; SORGI, SARA ; **DE LIMA, DIÓGENES S.** ; URBANSKI, ALYSSON H. ; SILVEIRA, EDUARDO LV ; NAKAYA, HELDER I. . Methods for predicting vaccine immunogenicity and reactogenicity. Human Vaccines & Immunotherapeutics JCR, v. 1, p. 1-8, 2019.

2. **DE LIMA, DIÓGENES S.**; CARDOZO, LUCAS E. ; MARACAJA-COUTINHO, VINICIUS ; SUHRBIER, ANDREAS ; MANE, KARIM ; JEFFRIES, DAVID ; SILVEIRA, EDUARDO L. V. ; AMARAL, PAULO P. ; RAPPUOLI, RINO ; DE SILVA, THUSHAN I. ; NAKAYA, HELDER I. . Long noncoding RNAs are involved in multiple immunological pathways in response to vaccination. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA (ONLINE) JCR, v. NA, p. 201822046-NA, 2019.

3. RUSSO, PEDRO S. T. ; FERREIRA, GUSTAVO R. ; CARDOZO, LUCAS E. ; BÜRGER, MATHEUS C. ; ARIAS-CARRASCO, RAUL ; MARUYAMA, SANDRA R. ; HIRATA, THIAGO D. C. ; **LIMA, DIÓGENES S.** ; PASSOS, FERNANDO M. ; FUKUTANI, KIYOSHI F. ; LEVER, MELISSA ; SILVA, JOÃO S. ; MARACAJA-COUTINHO, VINICIUS ; NAKAYA, HELDER I. . CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. BMC BIOINFORMATICS JCR, v. 19, p. 19-56, 2018.

4. OLIVEIRA, ÉRICA APARECIDA DE ; **LIMA, DIOGENES SAULO DE** ; CARDOZO, LUCAS ESTEVES ; SOUZA, GARCIA FERREIRA DE ; DE SOUZA, NAYANE ; ALVES-FERNANDES, DEBORA KRISTINA ; FAIÃO-FLORES, FERNANDA ; QUINCOCES, JOSÉ AGUSTÍN PABLO ; BARROS, SILVIA BERLANGA DE MORAES ; NAKAYA, HELDER I. ; MONTEIRO, GISELE ; MARIA-ENGLER, SILVYA STUCHI . Toxicogenomic and bioinformatics platforms to identify key molecular

mechanisms of a curcumin-analogue DM-1 toxicity in melanoma cells. PHARMACOLOGICAL RESEARCH JCR, v. 125, p. 178-187, 2017.

5. MENCK, RAFAEL ALMEIDA ; OLIVEIRA, CAROLINA DIZIOLI RODRIGUES ; **LIMA, DIÓGENES SAULO** ; GOES, LIGIA ENDO ; LEYTON, VILMA ; PASQUALUCCI, CARLOS AUGUSTO ; MUÑOZ, DANIEL ROMERO ; YONAMINE, MAURICIO . Hollow fiber-liquid phase microextraction of barbiturates in liver samples. Forensic Toxicology (Tokyo. Print) JCR, v. 31, p. 31-36, 2013.

Citações: WEB OF SCIENCE 10 | SCOPUS 9

6. MENCK, RAFAEL ALMEIDA ; **DE LIMA, DIÓGENES SAULO** ; SEULIN, SASKIA CAROLINA ; LEYTON, VILMA ; PASQUALUCCI, CARLOS AUGUSTO ; MUÑOZ, DANIEL ROMERO ; OSSELTON, MICHAEL DAVID ; YONAMINE, MAURICIO . Hollow-fiber liquid-phase microextraction and gas chromatography-mass spectrometry of barbiturates in whole blood samples. Journal of Separation Science (Print) JCR, v. 35, p. 3361-3368, 2012.

Citações: WEB OF SCIENCE 3 | SCOPUS 6

Apresentações de Trabalho

1. **DE LIMA, DIÓGENES SAULO**. Biologia de Sistemas Aplicada a Doenças Infecciosas. 2017. (Apresentação de Trabalho/Conferência ou palestra).
2. **LIMA, D. S.**; ALMEIDA, R. M. ; YONAMINE, M. ; OLIVEIRA, C. D. R. . Determinação de barbitúricos em amostras postmortem de fígado humano por cromatografia em fase gasosa acoplada à espectrometria de massas.. 2012. (Apresentação de Trabalho/Simpósio).
3. ALMEIDA, R. M. ; OLIVEIRA, C. D. R. ; **LIMA, D. S.** ; SEULIN, SASKIA CAROLINA ; LEYTON, V. . Análise de barbitúricos em tecidos postmortem com finalidade forense: Método alternativo utilizando óleo essencial. 2012. (Apresentação de Trabalho/Congresso).

Demais tipos de produção técnica

1. NAKAYA, H. ; **LIMA, DIÓGENES SAULO** . Monitor no Curso de Bioinformática para o CRID. 2017. (Curso de curta duração ministrado/Outra).