

Ministério da Saúde
Fundação Oswaldo Cruz
INSTITUTO OSWALDO CRUZ

Doutorado em Biologia Computacional e Sistemas

PROSPECÇÃO COMPUTACIONAL DE ALVOS PARA O
DESENVOLVIMENTO DE UMA VACINA CONTRA *LEISHMANIA* spp.

LEILANE OLIVEIRA GONÇALVES

Rio de Janeiro
Maio de 2021

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

LEILANE OLIVEIRA GONÇALVES

PROSPECÇÃO COMPUTACIONAL DE ALVOS PARA O
DESENVOLVIMENTO DE UMA VACINA CONTRA *LEISHMANIA* spp.

Tese apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutora em Ciências, com área de concentração em Genômica Funcional, Evolução e Filogenômica.

Orientador (es): Prof. Dr. Jeronimo Conceição Ruiz

Profa. Dra. Daniela de Melo Resende

RIO DE JANEIRO
Maio de 2021

Gonçalves, Leilane Oliveira.

Prospecção computacional de alvos para o desenvolvimento de uma vacina contra *Leishmania* spp. / Leilane Oliveira Gonçalves. - Rio de Janeiro, 2021.

195f f.; il.

Tese (Doutorado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2021.

Orientador: Jeronimo Conceição Ruiz.

Co-orientadora: Daniela de Melo Resende.

Bibliografia: f. 154-168

1. Imunoinformática. 2. Integração de dados. 3. *Leishmania* spp. 4. Redes de regulação gênica. 5. Transcritômica. I. Título.

Elaborado pelo Sistema de Geração Automática de Ficha Catalográfica da Biblioteca de Manguinhos/Icict/Fiocruz com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Igor Falce Dias de Lima - CRB-7/6930.

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTORA: Leilane Oliveira Gonçalves

Prospecção computacional de alvos para o desenvolvimento de uma vacina
contra *Leishmania* spp

ORIENTADOR (ES): Prof. Dr. Jeronimo Conceição Ruiz

Profa. Dra. Daniela de Melo Resende

Aprovada em: 31/05/2021

EXAMINADORES:

Prof. Dr. Fábio Passetti – ICC/PR (Presidente)

Prof. Dra. Angela Kaysel Cruz - USP/SP (Membro)

Prof. Dr. João Carlos Setúbal - USP/SP (Membro)

Prof. Dr. Edward José de Oliveira - PROCC/RJ (Suplente)

Prof. Dr. Ernesto Raúl Caffarena – IRR/MG (Suplente)

Rio de Janeiro, 31 de maio de 2021

DEDICATÓRIA

Dedico esta tese ao meu pai, que mesmo estando longe nunca deixou de se fazer presente.

Epígrafe

Em nossas escolhas encontra-se o nosso destino
"O Labirinto do Fauno"

Agradecimentos

Ao meu orientador, Dr. Jeronimo Ruiz, por ter me orientado em mais essa etapa da minha formação acadêmica. Agradeço pelos ensinamentos, pela confiança e pelas inúmeras oportunidades concedidas.

À Dra. Daniela Resende pela coorientação, pela disponibilidade, pelas palavras de apoio e motivação, pelas valiosas contribuições e companhia nos cafés durante esses quatro anos de doutorado.

À Dra. Maria Gabriela pelo incentivo, disponibilidade e confiança.

À Dra. Carolina Melo pela amizade, pela ajuda nos momentos de necessidade e pelas inúmeras conversas.

À Dra. Marta Polak pela supervisão, disponibilidade e por ter contribuído de maneira ímpar para o meu crescimento pessoal e profissional.

A todos do Grupo *Systems Immunology*, em especial à Gabriela, ao Andrés e à Ana, por terem me recebido de maneira tão acolhedora, por toda a disponibilidade e ensinamentos.

À Universidade de *Southampton*, na pessoa do Dr. David Wilson, pela inestimável ajuda, amizade e suporte nos momentos de necessidade.

Aos doutores Fabio Passetti, Angela Cruz, João Setúbal, Ernesto Caffarena e Edward de Oliveira por aceitarem avaliar este trabalho

À Dra. Silvane Murta por todo apoio, incentivo e disponibilidade.

À Sue, por ter me recebido e acolhido em sua casa. Muito obrigada pela paciência e pelas palavras de incentivo frente aos desafios encontrados. À Maiara, pela companhia e apoio em um país que não era o nosso.

À coordenação de Pós-graduação, em especial à Dra. Ana Carolina e ao Dr. Ernesto Caffarena, por possibilitarem o desenvolvimento deste trabalho em duas unidades da Fiocruz.

À Simone, por todo auxílio em relação às demandas que surgiram em diversos momentos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa disponibilizada.

À todos os amigos do Grupo Informática de Biosistemas e Genômica pelos momentos compartilhados, pelo apoio e pela ajuda indispensável em todos os momentos de necessidade. A amizade de vocês foi essencial durante o

desenvolvimento deste projeto. À Grace pela amizade, incentivo e ajuda durante todo esse tempo de convivência. Ao Jader pela imensa ajuda na resolução dos problemas técnicos ao longo do caminho. Ao Paul, pela ajuda no desenvolvimento dos programas, pela amizade e mensagens motivacionais. Ao Davi, pela amizade, ajuda e pelas diversas discussões científicas. Ao Fred, por toda a amizade, incentivo e ajuda. Ao Alexandre e ao Thiago pela companhia nos cafés, pelos momentos de descontração e por aceitarem todas as fotos do meu *Baby Yoda* de forma tão receptiva.

Ao Paulo e a Ana Paula por todas as conversas e momentos de descontração.

Ao Fernando, Virgínia, Luiz e Luiza pela amizade e suporte.

Às minhas grandes amigas Danielle Cirino, Ana Dian e Elvira que sempre se mantiveram presentes, nas palavras de apoio e por esperarem pacientemente minhas demoradas respostas.

Em especial, agradeço aos meus pais e irmãos pelo incentivo e por todo apoio recebido durante a minha caminhada.

INSTITUTO OSWALDO CRUZ

PROSPECÇÃO COMPUTACIONAL DE ALVOS PARA O DESENVOLVIMENTO DE UMA VACINA CONTRA *LEISHMANIA* sp.

RESUMO

TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Leilane Oliveira Gonçalves

A vacinação é uma das iniciativas de saúde pública mais bem-sucedidas. Apesar do sucesso, vacinas para muitas doenças ainda não foram desenvolvidas e diversas estratégias foram elaboradas para identificar candidatos vacinais específicos e imunogênicos. As Leishmanioses são um conjunto de doenças causadas por parasitos do gênero *Leishmania* para as quais, até o momento, não existem vacinas para uso humano e a quimioterapia contra as formas da doença é extremamente problemática. Dentro desse contexto, a prospecção computacional de vacinas, através da integração de dados biológicos, representa uma abordagem inovadora que, através da integração de múltiplas variáveis biológicas, viabiliza a identificação de perfis sistêmicos de resposta imune e uma compreensão mais detalhada da patogênese. Neste estudo, foram utilizados dados de sequenciamento de RNA de *Leishmania infantum*, *L. amazonensis*, macrófagos humanos infectados com *Leishmania* e de macrófagos humanos não infectados, além do proteoma predito de nove espécies de *Leishmania*, obtidos através do banco *TritrypDB*. A identificação de novos transcritos, permitiu a identificação de aproximadamente 2 mil sequências que até o momento não estavam anotadas no genoma desses parasitos. Além disso, forneceu um conjunto de dados que, somado ao proteoma predito, serviu de base para a realização de predições de imunoinformática. As análises de expressão gênica, possibilitaram a identificação de transcritos expressos de forma constitutiva entre as formas promastigotas e amastigotas do parasito, aumentando a chance de uma possível exposição do epítipo ao sistema imune. A comparação de linhagens sensíveis e resistentes pode identificar epítipos que, por serem mais expressos nas linhagens resistentes, podem ser incluídos nas análises como forma de garantir que a vacina também seja capaz de oferecer

cobertura contra a infecção por essas linhagens. Por sua vez, as redes de regulação gênica forneceram indicativos de como a modulação do sistema imune ocorre e quais são os genes que os parasitos expressam que podem estar envolvidos nessa modulação. Associadas às informações acima, após as predições de imunoinformática, foram selecionadas proteínas que apresentam epítomos para no mínimo 40 alelos humanos e 3 alelos de camundongo, além de serem preditos em proteínas com localização na membrana plasmática ou extracelular. Posteriormente, essas proteínas foram comparadas com o resultado do *OrthoMLC* com o objetivo de excluir proteínas que apresentassem similaridade de sequências com sequências proteicas de humano, camundongo e cachorro. Ao final dessas análises foram selecionadas 1.119 proteínas. Considerando a análise de imunogenicidade realizada com o programa *VaxiJen*, 102 epítomos, foram caracterizados como candidatos antigênicos. Foram priorizadas na seleção dos epítomos, para estudos futuros envolvendo validação experimental, um total de 20 sequências. Esses epítomos foram priorizados por apresentarem predição para alelos mais frequentes na população brasileira, segundo o banco de dados *Allele Frequency Net Database*, maior escore de imunogenicidade e maior porcentagem de conservação de sequências. Das sequências obtidas através da identificação de novos transcritos, duas delas apresentaram epítomos com potencial imunogênico e estão incluídas na análise. Com isso, espera-se que a perspectiva de integração de dados ômicos forneça uma contribuição instrumental para o desenvolvimento de vacinas para patógenos para os quais as abordagens tradicionais falharam até agora.

Palavras-chave: *Leishmania*, imunoinformática, redes de regulação gênica, transcritoma

INSTITUTO OSWALDO CRUZ

COMPUTATIONAL PROSPECTION OF TARGETS FOR THE DEVELOPMENT OF A VACCINE AGAINST *LEISHMANIA* sp.

ABSTRACT

DOCTORAL THESIS IN BIOLOGIA COMPUTACIONAL E SISTEMAS

Leilane Oliveira Gonçalves

Vaccination is one of the most successful public health initiatives. Despite the success, vaccines for many human diseases have not been developed yet and several strategies have been developed to identify specific and immunogenic vaccine candidates. Leishmaniasis is a group of diseases caused by parasites of the genus *Leishmania*. To date, there are no vaccines for human use against leishmaniasis, and chemotherapy against the disease is extremely problematic. Within this context, computational prospecting of vaccines, through the integration of biological data, represents an innovative approach that, through the integration of multiple biological variables, enables the identification of systemic immune response profiles and a more detailed understanding of the pathogenesis. In this study, RNA sequencing data from *Leishmania infantum*, *L. amazonensis*, macrophages infected with *Leishmania* and uninfected macrophages were used, in addition to the predicted proteome of nine species of *Leishmania* obtained through TritypDB. The identification of new transcripts allowed the identification of approximately 2,000 sequences that had not been recorded in the genome of these parasites, and in addition to providing a set of data that added to the predicted proteome served as the basis for making immunoinformatics predictions. The analysis of gene expression enabled the identification of transcripts expressed constitutively between the promastigote and amastigote forms of the parasite. The comparison between sensitive and resistant strains can identify epitopes more expressed in resistant strains, that can be included in the analysis as a way of ensuring that the vaccine is also able to offer protection against infection by these strains. In turn, the gene regulation networks provided indications of how modulation of the immune system occurs and what are the genes that parasites express that may be involved in this modulation. Associated with the information above, after the immunoinformatics

predictions, proteins were selected considering the prediction of epitopes for at least 40 human alleles and 3 mouse alleles, in addition to being predicted in proteins with localization in the plasma or extracellular membrane. Subsequently, these proteins were compared with OrthoMLC results in order to exclude proteins that presented sequence similarity with human, mouse and dog. At the end of these analyzes, 1,119 proteins were selected. Considering the immunogenicity analysis carried out by VaxiJen program, 102 epitopes, of the 124 previously selected, were characterized as antigenic candidates. In the selection of epitopes, for a future study involving experimental validation, a total of 20 sequences were prioritized. These epitopes were prioritized because they present a prediction for more frequent alleles in the Brazilian population, according to the Allele Frequency Net Database, a higher immunogenicity score and a higher percentage of sequence conservation. Of the sequences obtained through the identification of new transcripts, two of them presented epitopes with immunogenic potential and are included in the analysis. With this, it is expected that the perspective of integrating omic data will provide an instrumental contribution to the development of vaccines, for pathogens for which traditional approaches have so far failed.

Key words: *Leishmania*, immunoinformatics, gene regulatory networks, transcriptomics

ÍNDICE

RESUMO.....	ix
ABSTRACT.....	xi
ÍNDICE DE FIGURAS.....	xvi
LISTA DE TABELAS.....	xviii
LISTA DE SIGLAS E ABREVIATURAS.....	xix
1 Introdução.....	22
1.1 Imunoinformática.....	22
1.1.1 Imunoinformática aplicada a doenças negligenciadas.....	25
1.2 Integração de dados.....	26
1.3 Redes de regulação gênica.....	27
1.4 Leishmanioses.....	31
1.4.1 Tratamento das leishmanioses.....	35
1.4.2 Ciclo de vida e papel do sistema imune nas infecções por <i>Leishmania</i> sp.....	37
1.4.3 Vacinas para uso humano contra as leishmanioses.....	39
1.5 Justificativa.....	41
1.6. Hipótese.....	42
2 Objetivos.....	43
2.1 Objetivo geral.....	43
2.2 Objetivos específicos.....	43
3 Metodologia.....	44
3.1 Desenho experimental.....	45
3.1.1 Sequenciamento do transcrito de <i>Leishmania infantum</i> PP75.....	45
3.2 Recuperação de dados biológicos.....	46
3.2.1 Genoma e proteomas de referência de <i>Leishmania</i>	46
3.2.2 Transcritoma de macrófagos humanos infectados por <i>Leishmania</i> sp.....	48
3.2.3 Avaliação inicial do transcrito.....	49
3.3 Identificação de novos transcritos.....	53
3.3.1 Montagem <i>de novo</i> do transcrito.....	53
3.4 Mapeamento das leituras no genoma de referência e análise de expressão diferencial.....	59
3.4.1 Mapeamento das leituras no genoma de referência.....	59
3.4.2 Análise de expressão diferencial.....	59
3.5 Construção das redes de regulação gênica de macrófagos infectados com <i>Leishmania</i>	61

3.6	Predição de epítomos utilizando a abordagem de vacinologia reversa.....	63
3.6.1	Predição de epítomos com afinidade pelo MHC-II.....	63
3.6.2	Predição de epítomos com afinidade pelo MHC-I.....	65
3.6.3	Predição de epítomos de linfócitos B.....	67
3.6.4	Predição de localização subcelular, peptídeo sinal e topologia.....	67
3.7	Identificação de proteínas homólogas.....	70
3.8	Integração de dados em um banco de dados relacional e realização das consultas..	76
3.9	Análise de imunogenicidade e conservação.....	77
4	Resultados e discussão.....	79
4.1	Análise de qualidade dos dados.....	80
4.1.1	Seleção do conjunto de dados para análise através do controle de qualidade das amostras.....	80
4.1.2	Pré processamento dos dados de macrófagos infectados por <i>Leishmania</i> sp.....	86
4.2	Identificação de novos transcritos em <i>Leishmania infantum</i>	88
4.3	Identificação dos perfis de expressão em <i>Leishmania</i> sp.....	97
4.3.1	Mapeamento no genoma de referência e análise de expressão diferencial de <i>Leishmania infantum</i>	97
4.3.2	Mapeamento nos genomas de referência e análise de expressão diferencial dos dados de macrófagos infectados com <i>Leishmania</i>	116
4.4	Construção das redes de regulação gênica de macrófagos infectados.....	127
4.4.1	Identificação de agrupamentos gênicos e vias ativadas pela infecção por <i>Leishmania</i>	129
4.4.2	Anotação funcional dos agrupamentos encontrados e busca na literatura para modelagem das redes de regulação gênica.....	133
4.4.3	Construção das redes de regulação gênica.....	135
4.5	Integração de dados para a seleção das proteínas contendo epítomos imunogênicos.....	143
5.	Conclusão.....	150
6	Perspectivas.....	152
	REFERÊNCIAS.....	153
	Apêndices.....	168
	Apêndice A - Métricas relacionadas à análise de qualidade das amostras sequenciadas.....	168
	Apêndice B - Genes regulados positivamente e enriquecidos para a categoria do <i>Gene Ontology</i> de Processo Biológico.....	171
	Apêndice C - Transcritos regulados negativamente e enriquecidos para a categoria do <i>Gene Ontology</i> de Processo Biológico.....	173

Apêndice D - Transcritos sem enriquecimento funcional para a categoria do <i>Gene Ontology</i> de Processo Biológico.....	174
Apêndice E - Métricas relacionadas ao mapeamento das amostras sequenciadas.....	176
Apêndice F - Termos do <i>Gene Ontology</i> para o conjunto de genes DE em macrófagos.....	180
Apêndice G - Termos do <i>Gene Ontology</i> para o conjunto de transcritos regulados positivamente em <i>L. amazonensis</i>	182
Apêndice H - Apêndice G - Termos do <i>Gene Ontology</i> para o conjunto de transcritos regulados negativamente em <i>L. amazonensis</i>	184
Apêndice I - Média dos valores de leituras mapeadas para cada fator de transcrição em cada um dos <i>time-block</i> no modelo da rede na via de sinalização da IL-17.....	186
Apêndice J - Epítomos preditos.....	188

ÍNDICE DE FIGURAS

Figura 1.1 - Exemplo dos elementos de uma rede de regulação gênica.....	30
Figura 1.2 - Status da endemicidade da LV no mundo no ano de 2019.....	32
Figura 1.3 - Status da endemicidade da LC no mundo no ano de 2019.....	34
Figura 1.4 - Ciclo de vida do parasito.....	38
Figura 3.1 - Visão resumida das metodologias empregadas no trabalho.....	44
Figura 4.1 - Visão resumida das etapas do trabalho.....	80
Figura 4.2 - Resultado do BioAnalyzer para as amostras de <i>L. infantum</i>	82
Figura 4.3 - Distribuição do conteúdo GC das amostras de <i>L. infantum</i>	84
Figura 4.4 - Variação da qualidade das bases sequenciadas antes e depois dos cortes de qualidade para as amostras de <i>L. infantum</i>	85
Figura 4.5 - Variação da qualidade das bases sequenciadas do estudo SRP062278.....	87
Figura 4.6 - Distribuição do conteúdo GC das amostras do estudo SRP062278.....	88
Figura 4.7 - Fluxograma geral das principais etapas empregadas na identificação de novos transcritos.....	89
Figura 4.8 - Fluxograma geral das principais etapas empregadas na análise de expressão diferencial.....	98
Figura 4.9 - Resultado da análise de PCA para os dados de <i>L. infantum</i>	100
Figura 4.10 - Diagrama de Venn representando a distribuição dos transcritos diferencialmente expressos nas categorias do <i>Gene Ontology</i>	104
Figura 4.11 - Distribuição dos transcritos diferencialmente expressos nas categorias do <i>Gene Ontology</i> para o conjunto de dados regulados positivamente.....	106
Figura 4.12 - Distribuição dos transcritos diferencialmente expressos nas categorias do <i>Gene Ontology</i> para o conjunto de dados regulados negativamente.....	113
Figura 4.13 - Fluxograma geral das principais etapas empregadas na análise de expressão diferencial com dados de macrófagos humanos infectados com <i>Leishmania</i>	116
Figura 4.14 - Resultado da análise de PCA para os dados de macrófagos.....	118
Figura 4.15 - Resultado da análise de PCA para os dados de macrófagos após a remoção dos <i>outliers</i>	119
Figura 4.16 - Principais categorias enriquecidas considerando a análise com o <i>Gene Ontology</i> para os dados de macrófagos infectados.....	121

Figura 4.17 - Principais categorias enriquecidas considerando a análise com o Gene Ontology para os transcritos regulados positivamente em <i>L. amazonensis</i>	124
Figura 4.18 - Principais categorias enriquecidas considerando a análise com o Gene Ontology para os transcritos regulados negativamente em <i>L. amazonensis</i>	126
Figura 4.19 - Fluxograma geral das principais etapas empregadas para a construção das redes de regulação gênicas com dados de macrófagos humanos infectados com <i>Leishmania</i>	128
Figura 4.20 - Parâmetros definidos no <i>Graphia Pro</i> para a construção das redes.....	130
Figura 4.21 - Resultado do agrupamento realizado pelo MCL.....	130
Figura 4.22 - Agrupamento anotado como contendo diferenças de expressão entre as amostras infectadas e amostras controle.....	132
Figura 4.23 - Agrupamento anotado como contendo diferenças de expressão relacionadas a macrófagos infectados com <i>L. major</i>	133
Figura 4.24 - Diagrama da rede construída.....	137
Figura 4.25 - Resultado da simulação inicial da rede.....	139
Figura 4.26- Exemplos de alterações inseridas na rede.....	140
Figura 4.27 - Resultado da simulação da rede com as alterações nos fatores de transcrição.	141

LISTA DE TABELAS

Tabela 3.1 - Amostras e concentração de droga utilizada.....	45
Tabela 3.2 - Espécies de <i>Leishmania</i> utilizadas na identificação de novos transcritos e na análise de homologia entre diferentes espécies de <i>Leishmania</i> e seus principais hospedeiros vertebrados.....	47
Tabela 3.3 – Valores de Qualidade e acurácia das leituras.....	51
Tabela 4.1 – Resultado da avaliação de qualidade das leituras.....	83
Tabela 4.2 - Resumo da análise de qualidade do estudo SRP062278.....	86
Tabela 4.3 - Resultado da etapa de montagem de novo do transcrito.....	90
Tabela 4.4 – Organismos utilizados para a construção do banco de dados.....	93
Tabela 4.5 – Resultado da etapa de mapeamento das leituras no genoma de referência.....	99
Tabela 4.6 - Transcritos diferencialmente expressos entre as linhagens sensível e resistente de <i>L. infantum</i> classificados de acordo com a anotação funcional do <i>Gene Ontology</i> (GO).....	103
Tabela 4.7 - Resumo dos resultados do mapeamento das leituras nos genomas de referência.....	117
Tabela 4.8 - Genes diferencialmente expressos em macrófagos.....	120
Tabela 4.9 - Genes diferencialmente expressos em <i>Leishmania</i>	122
Tabela 4.10 - Termos usados para busca no <i>Pubmed</i> e no Periódicos Capes/MEC.....	135
Tabela 4.11 - Epítomos priorizados.....	145
Tabela 4.12 - Resultado do <i>InterproScan</i> para as proteínas anotadas como hipotéticas.....	148

LISTA DE SIGLAS E ABREVIATURAS

BD2K - *Big Data to Knowledge*

BLAST - *Basic Local Alignment Search Tool*

BP - processo biológico

CC - Componente celular

CDD - *Conserved domain database*

cDNA - DNA codificante, do inglês *coding DNA*

DE - diferencialmente expressos

DNT - doenças negligenciadas tropicais

ENA - *European Nucleotide Archive*

FDA - *Food and Drug Administration*

FM - função molecular

FT - fatores de transcrição

GFF - *General Feature Format*

GNR - *Gene Regulation Networks*

GO - *Gene Ontology*

HLA - *Human Leukocyte Antigens - HLA*

HPI - horas pós infecção

IEDB - *Immune Epitope Database and Analysis Resource*

IGV - *Integrative genome viewer*

IL-17 - interleucina 17

IL-2 - interleucina 2

IL-4 - interleucinas 4

LC - leishmaniose cutânea

LC - leishmaniose cutânea

LMC - mucocutânea

LT - leishmaniose tegumentar

LV - leishmaniose visceral

MCL - *Markov Clustering Algorithm*

MHC I - Complexo de Histocompatibilidade de Classe I

MHC II - Complexo de Histocompatibilidade de Classe II

mRNA – RNA mensageiro

MRPA - proteína multirresistente a drogas

NCBI - *National Center for Biotechnology Information*

TNF - *Tumor Necrosis Factor*

NGS - *do inglês, Next Generation Sequencing*

NID - *Neglected Infectious Diseases*

NIH - *National Institutes of Health*

OMS - Organização Mundial da saúde

ORF - *Open Reading Frames*

PCA - *do inglês, Principal Component Analysis*

Pfam - *Protein family database*

RBP - proteínas de ligação ao RNA

RIN - *RNA integrity number*

RNAseq - sequenciamento de RNA em larga escala

RRM - motivo de reconhecimento de RNA

rRNA - RNA ribossomal

SAM - *Sequence Alignment/Map format*

Sb^{III} - antimônio trivalente

Sb^V - antimônio pentavalente

SPN - *Stochastic Petri Nets*

SQL - *do inglês, Structured Query Language*

SRA - *Sequence Read Archive*

TNF - *Tumor Necrosis Factor*

1 Introdução

1.1 Imunoinformática

Atualmente, a vacinação é uma das medidas mais eficazes para o controle das doenças infecciosas. Estima-se que ela seja responsável por salvar cerca de 2-3 milhões de vidas anualmente em todo o mundo (Gregorio and Rappuoli 2012; Backert and Kohlbacher 2015). Atualmente, vacinas profiláticas, contra diferentes doenças humanas, são amplamente utilizadas e o desenvolvimento de vacinas que possam ser utilizadas em estratégias terapêuticas estão em estudo (Backert and Kohlbacher 2015).

Com o advento de novas tecnologias de sequenciamento *high-throughput* ocorreram inovações técnicas, tecnológicas e computacionais que reduziram custos e tempo de análise, mas que em contrapartida conduziram ao aumento da quantidade de dados gerados (Joyce and Palsson 2006).

Diversos bancos de dados foram criados para armazenar e disponibilizar dados biológicos associados a características imunológicas, um deles é o *Immune Epitope Database and Analysis Resource* (IEDB) (Brusic et al. 2000; Tomar and De 2010a, 2014).

Devido ao crescimento vertiginoso desses bancos de dados e das informações neles contidas, o desenvolvimento de novas abordagens computacionais que auxiliem no melhor entendimento das propriedades de todo o sistema imunológico se torna necessário (Tomar and De 2010a; Patronov Atanas and Doytchinova Irimi 2012, 2012).

Nesse contexto, surge a Imunoinformática como uma subárea da Bioinformática dedicada ao estudo e à predição computacional das características necessárias para identificação de proteínas imunogênicas (Bambini and Rappuoli 2009; Resende et al. 2012; Karch and Burkhard 2016).

Apesar de quase todas as vacinas existentes serem desenvolvidas a partir de metodologias tradicionais, com a triagem de alguns candidatos por vez partindo de

características conhecidas dos patógenos, a primeira abordagem de vacinologia reversa foi descrita por Rino Rappuoli para a identificação de vacinas através da utilização de dados genômicos (Rappuoli 2001; Bambini and Rappuoli 2009). A base da vacinologia reversa é a análise do genoma de patógenos com o objetivo de se identificar genes que codificam proteínas que possuam características imunogênicas para ser um bom candidato vacinal (Serruto and Rappuoli 2006). Dessa forma, após a escolha do organismo modelo a próxima etapa é identificar quais genes/proteínas possuem potencial imunogênico. Isso pode ser realizado com a utilização de diversos programas que visam a identificar se essas proteínas são de superfície ou proteínas secretadas, se são fatores de virulência, se possuem epítomos que serão reconhecidos por células T e células B ou até mesmo se são expressas em grande quantidade em condições fisiológicas ou se são superexpressas em condições de estresse (De Groot et al. 2008).

Essas análises de expressão possibilitam a identificação de variações que ocorrem não somente no patógeno, como também em seu hospedeiro (Scarselli et al. 2005a). Essa relação com a expressão já foi reportada em estudos envolvendo câncer, onde são identificados genes com regulação positiva no tecido tumoral e que não são expressos no tecido normal (De Groot et al. 2008; Arya et al. 2020). Couto e colaboradores (2021) também utilizaram dados de *RNAseq* para identificar genes que codificam proteínas com potencial imunogênico no sialotranscritoma de *Rhipicephalus bursa* para o desenvolvimento de uma vacina (Couto et al. 2021).

De maneira geral, a grande disponibilidade de dados ômicos gerados através das novas tecnologias de sequenciamento levou a uma melhor acurácia das ferramentas de imunoinformática, principalmente porque muitas delas são baseadas em métodos de aprendizado de máquina e utilizam esses dados ômicos como forma de treinamento (Kazi et al. 2018). A realização de um *screening*, utilizando abordagens computacionais, em proteomas preditos a partir de genomas completamente sequenciados, com intuito de selecionar um conjunto de dados menor e mais promissor para a realização das etapas experimentais tem se tornado cada vez mais comum (Tomar and De 2010b, 2014; Zhang et al. 2014).

No ano 2000, a vacinologia reversa foi aplicada para o desenvolvimento de uma vacina contra *Neisseria meningitidis*, atualmente já aprovada pelo *Food and Drug Administration* (FDA) (Kelly and Rappuoli 2005; Scarselli et al. 2005b).

Alguns outros estudos têm demonstrado que a utilização de análises computacionais para a identificação de regiões potencialmente imunogênicas tem apresentado resultados promissores, como, por exemplo, nos trabalhos de Nosrati e colaboradores (2016) e Pandey e colaboradores (2018). Estes estudos propuseram uma vacina multiepítopo contra o vírus da Hepatite C e o vírus da imunodeficiência adquirida, respectivamente (Nosrati et al. 2017; Pandey et al. 2018; Vakili et al. 2018).

As vacinas baseadas em epítomos são estratégias que envolvem a utilização de peptídeos relativamente curtos para induzir uma resposta imunológica efetiva. Esses peptídeos contêm epítomos que serão potencialmente reconhecidos por células T e B (Schubert et al. 2015).

Estratégias experimentais para a identificação desses epítomos são laboriosas e caras quando comparadas com a utilização da imunoinformática para a realização de uma triagem inicial (Lohia and Baranwal 2020). Além disso, vacinas que foram desenvolvidas através da utilização de abordagens experimentais convencionais são, em grande parte, constituídas de várias proteínas ou um patógeno inativado e segundo Oli e colaboradores (2020) isso pode constituir uma carga antigênica desnecessária e que aumentaria as chances de indução de algum processo alérgico e/ou autoimune (Oli et al. 2020).

As vacinas desenvolvidas a partir de imunógenos são capazes de desencadear respostas imunes altamente específicas, além de possibilitarem a utilização de diferentes epítomos em sua construção. Elas possuem diversas vantagens quando comparadas com vacinas desenhadas com um único epítopo ou vacinas clássicas, estando elas relacionadas a: I) apresentarem um conjunto de diferentes epítomos obtidos a partir de alvos distintos de determinado organismo; II) podem ser reconhecidos mais facilmente por alelos de antígenos leucocitários humanos (do inglês, *Human Leukocyte Antigens* - HLA); III) podem ser facilmente adjuvados; IV) podem ativar respostas imunológicas mediadas por células T CD4+,

T CD8+ e epítomos de células B; e V) apresentam uma baixa chance de causar efeitos adversos e/ou respostas imunológicas indesejadas (Zhang 2018; Lohia and Baranwal 2020; Oli et al. 2020).

1.1.1 Imunoinformática aplicada a doenças negligenciadas

Uma aplicação muito importante dessas estratégias de imunoinformática está na identificação de potenciais candidatos vacinais para patógenos nos quais as metodologias convencionais de desenvolvimento de vacinas não tiveram sucesso até o momento, como, por exemplo, para os patógenos causadores das doenças classificadas como negligenciadas (Scarselli et al. 2005b; Oli et al. 2020).

Segundo a Organização Mundial da Saúde (OMS), as doenças infecciosas negligenciadas (do inglês, *Neglected Infectious Diseases* - NIDs) são responsáveis por afetar populações que vivem em condições socioeconômicas precárias. Como essas doenças compartilham fatores socioambientais, elas frequentemente coexistem em determinadas áreas geográficas (WHO 2020; CDC 2020; PAHO/WHO, 2020). As NIDs são causadas por diferentes patógenos, sendo, em sua maioria, causadores de doenças crônicas com efeitos de longa duração na saúde, afetando populações com baixa visibilidade. Esse fato leva, por sua vez, à diminuição da produtividade do trabalho e à diminuição de geração de renda (WHO 2020; PAHO/WHO 2020).

Atualmente, a OMS classifica 20 doenças no grupo de doenças negligenciadas tropicais (DNT), a grande maioria delas possuem ocorrência reconhecida no Brasil (Luna et al. 2020). Elas constituem um grupo de doenças parasitárias, como amebíase, Doença de Chagas, esquistossomose e leishmaniose. Apesar de sua relevância para a saúde pública global, os avanços no desenvolvimento de vacinas para essas doenças ainda é muito incipiente (WHO 2020; PAHO/WHO 2020). Segundo Bethony e colaboradores (2011), a ausência de incentivos financeiros faz com que as empresas farmacêuticas não tenham interesse na pesquisa e desenvolvimento para vacinas contra DNT, ficando a cargo dos governos o incentivo e investimento em pesquisas (Bethony et al. 2011).

Até o momento, não existem vacinas verdadeiramente eficazes para a maioria das DNT, embora diversas iniciativas estejam em andamento (Robleda-Castillo et al. 2021). Segundo Terry e colaboradores (2014), a disponibilidade de dados de sequenciamento de genomas de diversos patógenos emergentes e associados a doenças tropicais negligenciadas possibilita o emprego de ferramentas de imunoinformática, facilitando a identificação e testagem de potenciais candidatos vacinais (Terry et al. 2015).

1.2 Integração de dados

Com o avanço das tecnologias de sequenciamento, a disponibilização de conjuntos de dados biológicos em bancos de dados públicos tem se tornado cada vez mais frequente. Podemos citar, como exemplo, o *GenBank* que, em dezembro de 1982, disponibilizava 606 sequências biológicas. Em fevereiro de 2021, o mesmo banco registrava 226.241.476 sequências (National Center for Biotechnology Information 2021).

Em 2013, o Instituto Nacional de Saúde dos EUA (do inglês, *National Institutes of Health* - NIH) deu início ao programa “*Big Data to Knowledge*” (*BD2K*) para promover o desenvolvimento e aperfeiçoamento de novas estratégias para analisar a grande quantidade de dados disponíveis atualmente (Bui and Van Horn 2017; Noor et al. 2019). Uma das propostas da iniciativa *BD2K* está relacionada à integração de dados heterogêneos com o intuito de se encontrar novos relacionamentos através de uma visão abrangente da saúde, que integra características moleculares, genômicas e dados clínicos aos diferentes fenótipos humanos (Ulfenborg 2019).

Essas estratégias permitem investigar mecanismos complexos através de diferentes níveis de informação, como no trabalho de Woo e colaboradores (2017) que integraram dados de genômica, epigenética e transcrito para analisar os diferentes perfis de carcinoma hepatocelular. Isso levou à identificação de três diferentes perfis moleculares que poderiam estar relacionados aos diferentes prognósticos da doença (Woo et al. 2017).

Por sua vez, Zhu e colaboradores (2017) analisaram perfis de genômica, epigenética, transcrito e dados clínicos de aproximadamente 3.000 indivíduos e

14 diferentes tipos de câncer e observaram que características relacionadas à expressão de microRNAs são as que mais variam entre as amostras, sugerindo que a sua utilização deve ser considerada no aperfeiçoamento das análises prognósticas atualmente realizadas (Zhu et al. 2017).

Ulfenborg e colaboradores (2019) sugerem que as principais questões biológicas que podem ser respondidas com base nas aplicações das abordagens integrativas podem ser divididas em três áreas, sendo elas a identificação e classificação de doenças com base em perfis multiômicos; a predição de biomarcadores para diversas aplicações e, por último, a obtenção de informações sobre a biologia de diferentes patologias (Ulfenborg 2019).

No entanto, ainda existe um grande desafio quanto à extração da informação desses conjuntos de dados disponibilizados publicamente, muitas vezes devido à falta de padronização dos formatos de arquivos que são disponibilizados e à falta de ligação entre os diferentes bancos de dados. Além disso, segundo Pinu e colaboradores (2019), os estudos multiômicos requerem o desenvolvimento de um projeto experimental adequado, a coleta de dados quantitativos e metadados e a revisão sobre as ferramentas mais adequadas para integração e interpretação dos dados (Pinu et al. 2019; Graw et al. 2021).

1.3 Redes de regulação gênica

Elucidar as relações entre os genes e os produtos que eles codificam continua sendo um dos desafios centrais da biologia experimental e computacional.

As redes de regulação gênica (do inglês, *Gene Regulation Networks* - GRN) demonstram como os genes podem interagir uns com os outros, através da construção de um grafo direcionado no qual os reguladores da expressão gênica são conectados por nós e arestas. Os reguladores da expressão gênica incluem fatores de transcrição (FT) que podem atuar como ativadores ou repressores (Chandrasekaran et al. 2011; Balaguer et al. 2017; Sinha et al. 2020).

As GRN podem definir estados transcricionais específicos de diferentes tipos de células em diferentes condições fisiológicas, representando uma ferramenta fundamental para estudo e modelagem de processos biológicos complexos,

ajudando na descoberta de novas interações entre entidades biológicas (Delgado and Gómez-Vela 2019; Sinha et al. 2020).

A abordagem de GRN descreve como os reguladores (FT) afetam a expressão de um conjunto de genes alvos (Steele et al. 2009; Ghedira et al. 2011). A identificação de elementos regulatórios pode auxiliar para um melhor entendimento de como os sinais de entrada (feitos pelos FT) promovem, como saída, diferentes resultados, causados pelas mudanças nos níveis de expressão gênica (Chandrasekaran et al. 2011; Peter and Davidson 2017). As GRN são frequentemente representadas como um grafo em que os nós são genes/produtos gênicos (TF e seus genes-alvo) e as arestas correspondem às relações entre nós (Steele et al. 2009; Ghedira et al. 2011). Este método recebe como entrada dados de expressão gênica e fontes de informação prévia (dados de interação gênica obtidos através de buscas na literatura e que tenham sido validados experimentalmente), e produz relações regulatórias entre FTs e seus genes alvo que explicam os níveis de expressão gênica observados (Jackson et al. 2020).

Durante os últimos anos, muitos pesquisadores têm se concentrado na modelagem de GRN através de análises de expressão gênica. Vários estudos relataram a predição computacional de GRN usando dados de transcrito, gerados por tecnologias como microarranjos e *RNAseq* (Muraro et al. 2013; Gutiérrez et al. 2016; Yang et al. 2016; Banf and Rhee 2017; Delgado and Gómez-Vela 2019; Liang et al. 2019). A reconstrução das GRN é necessária para entender como a desregulação da expressão gênica contribui para o desenvolvimento de doenças (Barabási et al. 2011; Yu et al. 2017; Jackson et al. 2020).

A abordagem GRN utilizando dados de sequenciamento de RNA já foi aplicada no diagnóstico de carcinoma hepatocelular e para a identificação de biomarcadores moleculares para monitoramento e tratamento da progressão do câncer e também para a compreensão de como o sistema imunológico poderia lidar com infecções (Singh et al. 2014; Pennisi et al. 2016; Yang et al. 2016; Liang et al. 2019).

De acordo com Singh e colaboradores (2014), a abordagem de GRN tem sido amplamente utilizada para a modelagem do sistema imunológico, a fim de elucidar

as diferenças entre os tipos de células imunológicas e sua diferenciação e ativação em resposta a patógenos ou antígenos não próprios (Singh et al. 2014; Castiglione et al. 2016; Palma et al. 2018). Palma e colaboradores (2018) e Castiglione e colaboradores (2016) usaram a modelagem de GRN para corroborar os estados de polarização dos macrófagos a fim de explicar as mudanças entre os diferentes fenótipos dos macrófagos, sugerindo que macrófagos anti-inflamatórios são resilientes para voltar ao fenótipo pró-inflamatório (Castiglione et al. 2016; Palma et al. 2018). A polarização dos macrófagos é um processo fundamental para a resposta do sistema imunológico às doenças infecciosas. Os dois principais fenótipos de macrófagos são os M1 e M2. O fenótipo M1 corresponde a macrófagos pró-inflamatórios, enquanto o M2 é um subtipo de macrófago anti-inflamatório/regulador (Tomiotto-Pellissier et al. 2018).

As abordagens qualitativas para a construção de GRNs frequentemente se baseiam na utilização de equações (booleanas ou multiníveis). Por sua vez, as redes do tipo *Petri Nets* oferecem outra estrutura complementar para analisar as propriedades dinâmicas de sistemas, de um ponto de vista qualitativo ou quantitativo (Marwan et al., 2011).

As *Petri Nets* são grafos bipartidos ponderados e direcionados que foram aplicados com sucesso à biologia de sistemas, inicialmente, para a reconstrução e a comparação de vias metabólicas e de mecanismos de transdução de sinal na modelagem de processos estocásticos (discretos) e determinísticos (contínuos) (Chaouiya et al., 2007).

1.3.1 Petri Nets

As *Petri Nets* são uma abordagem matemática para descrever sistemas distribuídos, sendo amplamente utilizadas para modelagem de diferentes tipos de sistemas, incluindo os biológicos (Hardy and Robillard 2004; O'Hara et al. 2016).

Essas abordagens compartilham uma série de características, como por exemplo serem geralmente modeladas como redes direcionais bipartidas, ou seja, com uma estrutura recorrente de nó-transição-nó (ou, no caso de diagramas de processo, entidade-processo-entidade), onde os nós são considerados lugares ou

transições conectadas por arestas. As transições representam interações entre entidades (genes) ou a transição entre os diferentes estados de uma entidade (figura 1.1) (Livigni et al. 2016).

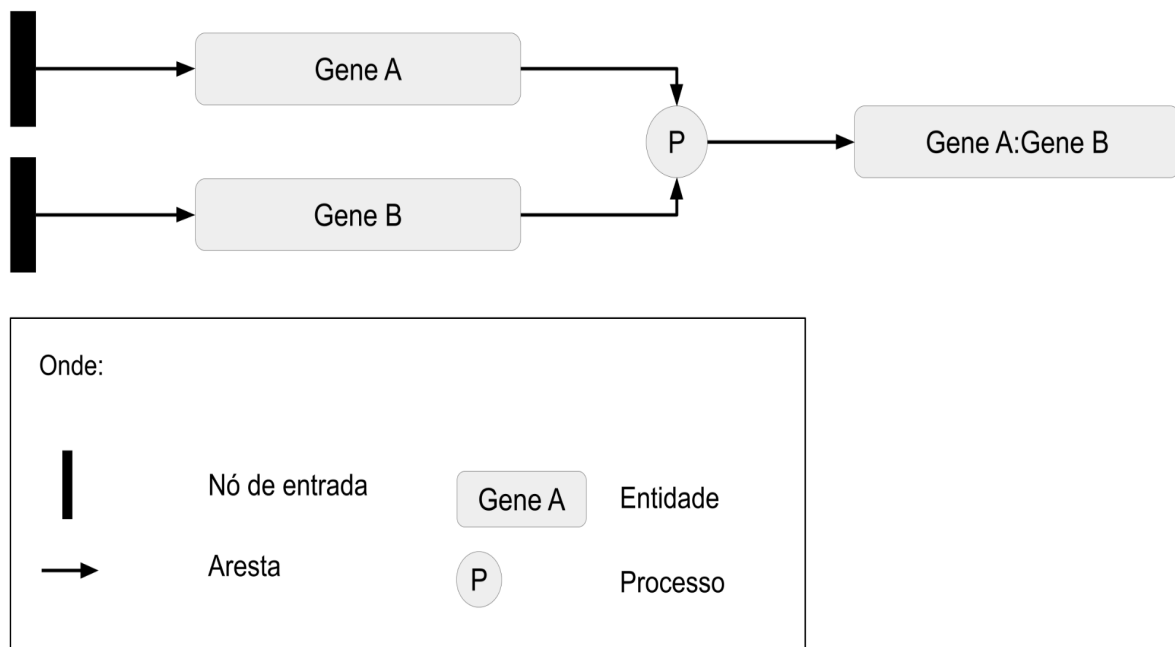


Figura 1.1 - Exemplo dos elementos de uma rede de regulação gênica. O modelo de uma GNR representa uma estrutura recorrente de nó-transição-nó, onde os nós são considerados lugares ou transições conectadas por arestas. As transições representam interações entre entidades (genes) ou a transição entre os diferentes estados de uma entidade.

Fonte: Dados do trabalho

Suas principais vantagens estão relacionadas à facilidade de interpretação gráfica e à representação da dinâmica e da estrutura de um sistema em vários níveis de informação (Miyagi et al. 2002).

Para a sua execução, os modelos criados através da utilização de abordagens de mineração de texto baseadas em semântica precisam ser parametrizados, geralmente utilizando dados de expressão gênica, definindo assim a condição inicial do modelo. Essa parametrização define para onde e como ocorre o fluxo de informação (O’Hara et al. 2016).

Essas abordagens têm sido amplamente utilizadas para a modelagem de vias bioquímicas e metabólicas. Sua primeira utilização foi descrita em 1993 por Reddy e colaboradores, que modelaram diferentes vias metabólicas (Reddy et al. 1993; Hardy and Robillard 2004). Após isso, sua utilização já foi descrita para a modelagem de processos de comunicação celular, para a descrição de mecanismos de ação de drogas contra tuberculose e para modelar as diferentes respostas imunológicas (Janowski et al. 2010; Pennisi et al. 2016; Singh et al. 2020).

1.4 Leishmanioses

As leishmanioses são um conjunto de doenças negligenciadas causadas por mais de 20 espécies de protozoários parasitos do gênero *Leishmania* (Cuervo et al. 2010). Nas Américas, as leishmanioses estão presentes em 18 países sendo a leishmaniose tegumentar (LT) na sua forma cutânea a forma clínica mais comum (Brasil et al.). Segundo estimativas da OMS, cerca de 50-90.000 novos casos, relacionados à leishmaniose visceral (LV) humana, e cerca de 600 mil a um milhão de novos casos de leishmaniose tegumentar, dos tipos cutânea (LC) e mucocutânea (LMC) humana, foram registrados em aproximadamente 100 países endêmicos (WHO 2021a).

Mais de 90% dos casos de LV são reportados em apenas cinco países, sendo eles Brasil, Etiópia, Índia, Sudão do Sul e Sudão (figura 1.2) (WHO 2021).

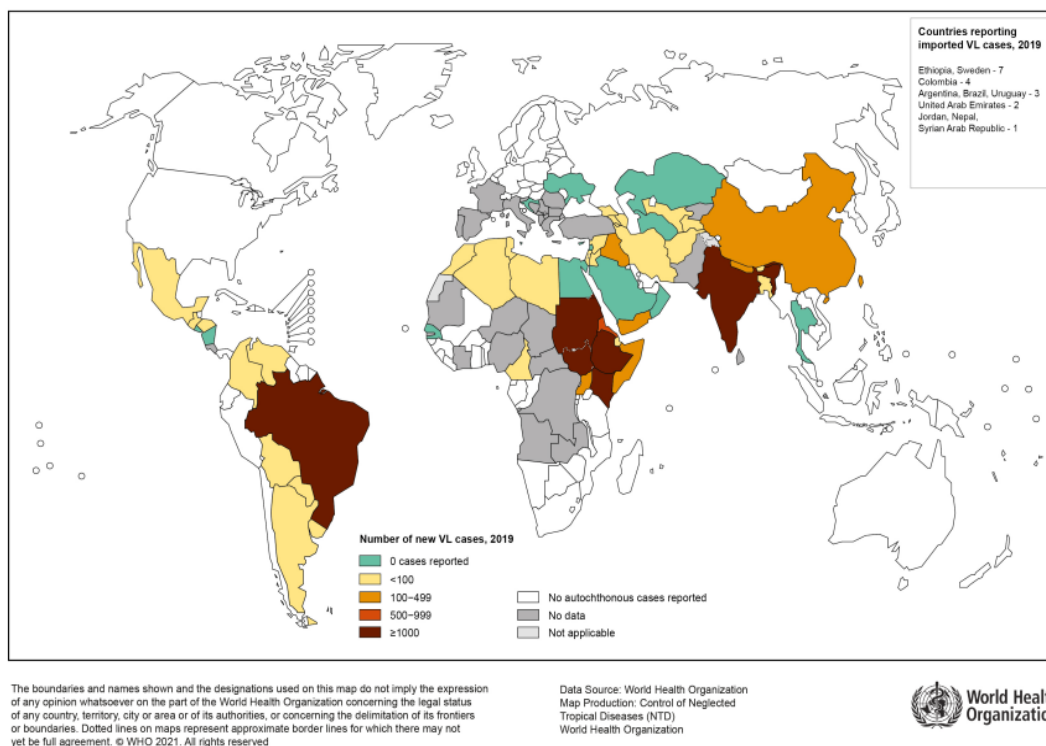


Figura 1.2 - Status da endemidade da LV no mundo no ano de 2019. Países em que foram reportados casos de LV no ano de 2019, onde mais de 90% dos casos de LV são reportados em apenas cinco países, sendo eles, Brasil, Etiópia, Índia, Sudão do Sul e Sudão

Fonte: https://cdn.who.int/media/docs/default-source/ntds/leishmaniasis/leishmaniasis-vl-2019.pdf?sfvrsn=a9c1920e_9

A LV é caracterizada pela ocorrência de febre prolongada, esplenomegalia (aumento do baço), hepatomegalia (aumento do fígado), perda substancial de peso, anemia progressiva, pancitopenia e hipergamaglobulinemia (aumento das imunoglobulinas no sangue) (Sundar 2001). É a forma mais grave da doença, sendo letal caso não diagnosticada e tratada. Considerando os indivíduos sintomáticos, segundo Ready (2014), a LV apresenta uma taxa de mortalidade de 75%–95% (Ready 2014; Ibarra-Meneses et al. 2020).

A LV é comumente causada pelas espécies *Leishmania donovani* e *Leishmania infantum*. As características clínicas da LV causada por espécies diferentes não são uniformes e cada parasito possui um padrão epidemiológico distinto. No subcontinente indiano, a doença é quase exclusivamente causada por *L.*

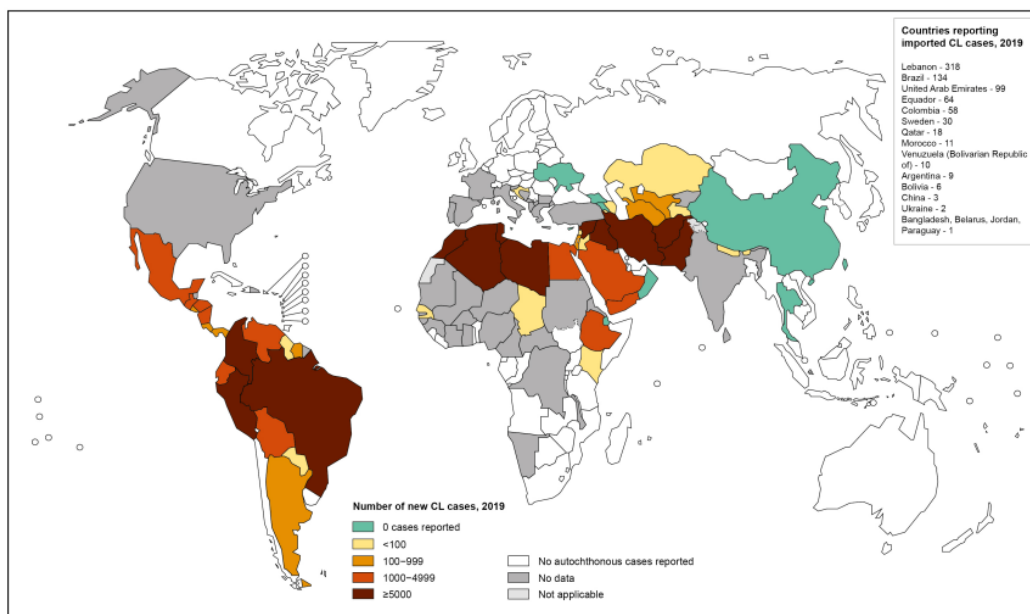
donovani, enquanto nas Américas ocorre a prevalência da espécie *L. infantum*, (Sundar and Rai 2002).

No Brasil, o primeiro relato de LV ocorreu em 1913, em um paciente de Boa Esperança, Mato Grosso. Até o início da década de 1950, haviam sido notificados somente 379 casos de LV. No entanto, nos últimos 40 anos, a transmissão da LV foi descrita em diversos municípios brasileiros e, principalmente, a partir da década de 1980, são relatados inúmeros casos e surtos de LV humana (Maia-Elkhoury et al. 2008).

Segundo dados do Ministério da Saúde, em 2019, no Brasil, foram confirmados aproximadamente 2.500 novos casos de LV humana, sendo que 26,8% dessas notificações foram em crianças menores de cinco anos. A LV apresenta uma taxa de incidência de 1,2 casos a cada 100 mil habitantes, sendo a região Nordeste responsável por 49,1% das notificações (Brasil 2021).

De acordo com Sundar (2002), o aumento da prevalência da infecção pelo vírus da imunodeficiência humana (HIV) e a coinfeção por HIV-*Leishmania* spp na população adulta tem sido relatada com frequência. Segundo a OMS, as altas taxas de coinfeção são reportadas no Brasil, Etiópia e em Bihar, na Índia. No Brasil, a coinfeção *Leishmania*/HIV representou 11,1% do total de casos confirmados em 2019, sendo registrados principalmente nas regiões Norte e Centro-Oeste (Sundar and Rai 2002; Brasil 2021).

A LC é a forma mais comum da doença, sendo causada por diferentes espécies de *Leishmania*, entre elas *Leishmania major*, *Leishmania amazonensis*, *Leishmania braziliensis*, *Leishmania guyanensis* e *Leishmania mexicana* (Martinez and Petersen 2014). Em 2019, a grande maioria dos casos reportados ocorreram em 10 países, sendo eles Afeganistão, Algéria, Brasil, Colômbia, Irã, Bolívia, Iraque, Paquistão, Tunísia e Síria (figura 1.3). Em relação à LMC, cerca de 90% dos casos reportados ocorreram na Bolívia, no Brasil, na Etiópia e no Peru (WHO 2021).



The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement. © WHO 2021. All rights reserved.

Data Source: World Health Organization
Map Production: Control of Neglected Tropical Diseases (NTD)
World Health Organization



Figura 1.3 - Status da endemidade da LC no mundo no ano de 2019. Em 2018, a grande maioria dos casos reportados ocorreu em 10 países, sendo eles Afeganistão, Algéria, Brasil, Colômbia, Irã, Bolívia, Iraque, Paquistão, Tunísia e Síria

Fonte: https://cdn.who.int/media/docs/default-source/ntds/leishmaniasis/leishmaniasis-cl-2019.pdf?sfvrsn=7b1eedab_7

A grande maioria dos casos de LC são caracterizados por apresentarem lesões de difícil cicatrização, deixando cicatrizes para toda a vida e deficiências graves (Vega-López 2003).

No Brasil, segundo dados do Ministério da Saúde, em 2019 foram confirmados aproximadamente 16 mil casos novos de LT. A região com o maior número de notificações foi a região Norte, com 42,8% dos casos reportados. Do total de notificações, 5,3% dos diagnósticos foram relacionados à LMC e em 1% houve registro da coinfeção com HIV. Destes casos, aproximadamente 61% dos pacientes evoluíram para cura clínica, enquanto 1,9% abandonaram o tratamento (Brasil 2021; Brasil et al. 2020).

1.4.1 Tratamento das leishmanioses

Até o momento não existem vacinas contra leishmaniose para uso humano e o tratamento é realizado através da utilização de compostos derivados de antimônio pentavalente (Sb^V), como por exemplo o estibogluconato de sódio e o antimoniato de meglumina (Pentostam® e Glucantime®) (Guerin et al. 2002).

No entanto, o tratamento das leishmanioses com esses fármacos de primeira escolha não é seguro e eficaz, e a falta de incentivo por parte da iniciativa farmacêutica é um fator que agrava essa situação. Entre 1975 e 2004 foram aprovadas cerca de 1.556 novas moléculas para o tratamento de doenças infecciosas. No entanto, apenas 21 dessas moléculas foram desenvolvidas para o tratamento de DNT (Surur et al. 2020).

O Pentostam® e o Glucantime® são fármacos que estão em uso há mais de 70 anos e estão associados a efeitos adversos graves e requerem, por parte do paciente, uma adesão prolongada ao tratamento. Esses quimioterápicos são considerados pró-fármacos, uma vez que são administradas em sua forma inativa (Sb^V) e são ativadas após o processo de biotransformação, no qual o Sb^V é reduzido à forma do antimônio trivalente (Sb^{III}) (Ouellette et al. 2004).

Segundo Anversa (2018), o sucesso do tratamento contra as leishmanioses envolve diversos fatores, como por exemplo resposta imunológica e apresentação clínica da doença; dosagem, duração e conclusão do tratamento, além de características do parasito, como a falta de resistência ao medicamento (Anversa et al. 2018).

Entretanto, a quimioterapia contra as infecções por *Leishmania* apresenta uma série de limitações, como a elevada toxicidade do fármaco e efeitos colaterais severos (Croft et al. 2006; Moore and Lockwood 2010). Além disso, são contraindicados para o tratamento de pacientes que sofram de doenças cardíacas, renais e hepáticas e também para mulheres grávidas, uma vez que esses fármacos são capazes de atravessar a barreira placentária e afetar o tecido nervoso fetal (Anversa et al. 2018).

Outro fator importante que dificulta a eficácia do tratamento nos pacientes infectados está relacionada à emergência de parasitos resistentes aos fármacos de primeira escolha (Iqbal et al. 2016).

Nos últimos anos, o número de insucessos terapêuticos devido ao surgimento de parasitos resistentes aos antimoniais tem se tornado frequente. Isso limita a utilização desses fármacos em diversos países, como na Índia, onde já foram observadas falhas terapêuticas em mais de 60% dos casos de LV (Ready 2014).

Atualmente existem alguns fármacos de segunda escolha para o tratamento das leishmanioses, no entanto elas também apresentam uma série de limitações. No Brasil, segundo dados do Ministério da Saúde, a escolha de cada um dos medicamentos deverá considerar a faixa etária, presença de gravidez e comorbidades (Brasil, 2011).

A anfotericina B é um dos fármacos de segunda escolha com maior potencial leishmanicida, disponível comercialmente, atuando contra as formas promastigotas e amastigotas, tanto *in vitro* quanto *in vivo* (THAKUR; NARAYAN, 2004; Anversa et al. 2018). Atualmente, dois métodos de apresentação para a anfotericina B são disponibilizados pelo Ministério da Saúde para o tratamento dos pacientes infectados, o desoxicolato de anfotericina B e a anfotericina B lipossomal (Brasil, 2020). O desoxicolato de anfotericina B tornou-se comercialmente disponível na forma de uma solução na qual o sal desoxicolato de sódio é utilizado como agente solubilizante, (VYAS; GUPTA, 2006; LABORÍN; VARGAS, 2009). Já a anfotericina B na sua formulação lipossomal é um fármaco menos tóxico e que provoca menos efeitos colaterais, sendo indicado para pacientes que apresentem complicações renais e/ou cardíacas, e também para pacientes como gestantes, pessoas soropositivas e pacientes para os quais os fármacos de primeira escolha falharam. No entanto, esse tratamento apresenta um custo elevado para os sistemas públicos de saúde (Tuon et al. 2008; Anversa et al. 2018).

Por sua vez, a pentamidina é utilizada principalmente em regiões onde ocorre falha no tratamento com os fármacos de primeira escolha (principalmente na Índia, Guiana Francesa e Suriname) e em casos individuais de resistência à primeira terapia de escolha, sendo indicada para o tratamento de leishmaniose tegumentar causada por *L. guyanensis*. No entanto, é contraindicada em casos de gravidez,

diabetes, insuficiência renal, cardíaca e hepática (Croft et al. 2006; Anversa et al. 2018).

A paromomicina apresenta baixo custo e efeitos colaterais pouco frequentes, no entanto as taxas de cura relatadas são frequentemente menores do que as observadas com o uso de antimoniais pentavalentes (Tuon et al. 2008).

A miltefosina é um fármaco que deve possuir um regime de administração de 28 dias, duas vezes ao dia. Apesar dos efeitos gastrointestinais, apresenta altas taxas de cura nos casos de LV. No entanto, a biodisponibilidade oral pode levar ao uso indevido e a efeitos indesejáveis, como toxicidade, descontinuação prematura da terapia, falha terapêutica e desenvolvimento de resistência. Outros aspectos importantes estão relacionados à toxicidade grave, que ocorre em cerca de 3% dos pacientes, e o potencial teratogênico desse fármaco (Sundar and Olliaro 2007; Sunyoto et al. 2018).

Devido a todas as limitações relacionadas à quimioterapia das leishmanioses, pesquisas com o intuito de identificar novas alternativas para o tratamento dessas doenças vêm sendo consideradas como estratégicas e prioritárias pela OMS.

1.4.2 Ciclo de vida e papel do sistema imune nas infecções por *Leishmania* spp.

Durante o seu ciclo de vida, esses parasitos apresentam dois estágios principais de desenvolvimento: a) formas promastigotas, flageladas, que são encontradas no intestino de insetos flebotomíneos, do gênero *Phlebotomus* e *Lutzomyia* e b) formas amastigotas, que residem, principalmente, nos macrófagos dos hospedeiros vertebrados (figura 1.4) (Cuervo et al. 2010).

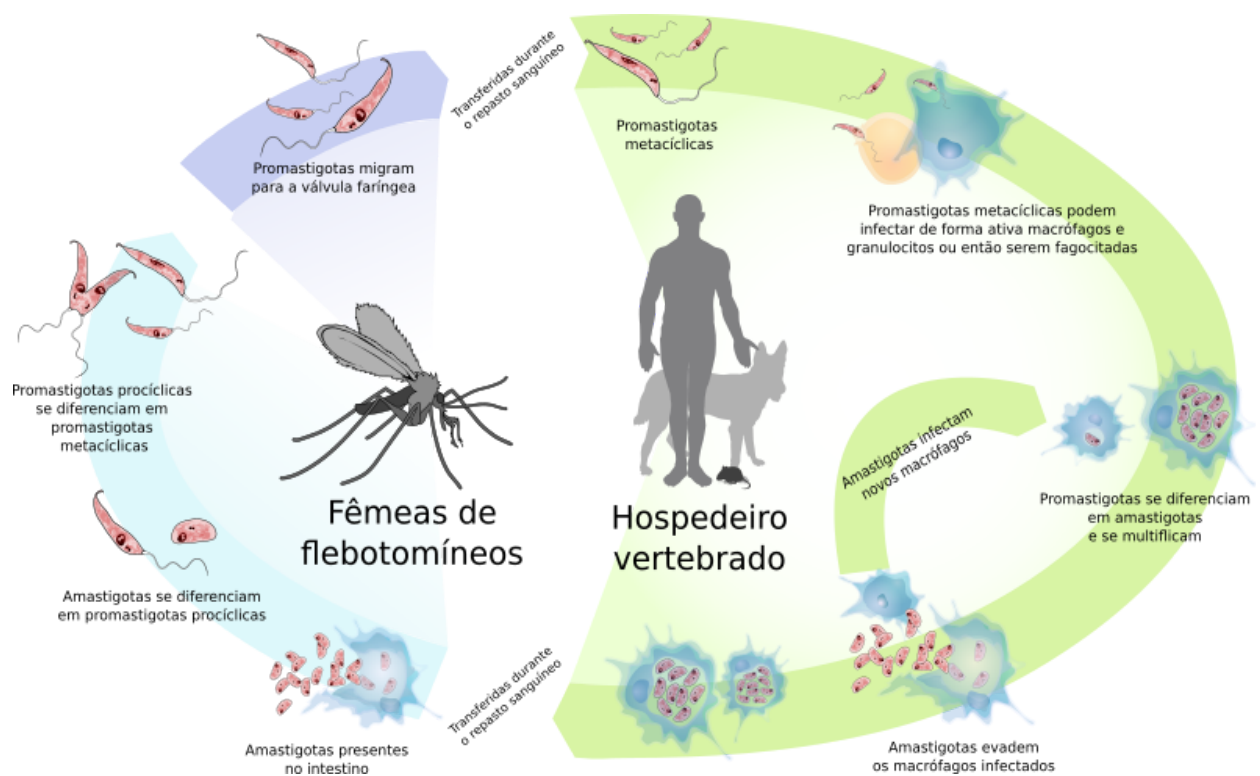


Figura 1.4 - Ciclo de vida do parasito. Durante o seu ciclo de vida, esses parasitos apresentam dois estágios principais de desenvolvimento: a) formas promastigotas, flageladas, que são encontradas no intestino de insetos flebotomíneos, do gênero *Phlebotomus* e *Lutzomyia* e b) formas amastigotas, que residem, principalmente, nos macrófagos dos hospedeiros vertebrados.

Fonte: Adaptado de Wikimedia Commons (https://commons.wikimedia.org/wiki/File:Leishmaniasis_life_cycle_diagram_en.svg)

Ao entrar em contato com o organismo do hospedeiro vertebrado, as formas promastigotas de *Leishmania* são fagocitadas principalmente por macrófagos e neutrófilos. No entanto, pelo fato dos neutrófilos apresentarem uma curta vida útil, os macrófagos se tornam as células hospedeiras definitivas para a proliferação desses parasitos (Liu and Uzonna 2012).

Após o processo de internalização dos parasitos pelos macrófagos, as formas promastigotas se diferenciam em amastigotas, que se multiplicam por fissão binária, podendo, eventualmente, romper o macrófago e infectar outras células (Cuervo et al. 2010; Liu and Uzonna 2012; Afrin et al. 2019).

O processo de fagocitose das formas promastigotas pelas células hospedeiras é mediado por receptores, sendo que um grande número de moléculas

de superfície estão diretamente envolvidas nesse mecanismo como, por exemplo, os receptores do complemento 1 e 3, receptores de fibronectina e de manose (Handman 2001; McConville et al. 2007; Naderer and McConville 2008). Após a fagocitose, os parasitos precisam evitar a ativação e o reconhecimento dos macrófagos infectados pelas células T e, por isso, desenvolveram uma série de mecanismos de escape das atividades inflamatórias e antimicrobianas realizadas por macrófagos e neutrófilos, alterando, por exemplo, várias vias de sinalização da célula hospedeira (Ghosh and Bandyopadhyay 2004; Naderer and McConville 2008; Isnard et al. 2012).

As infecções por *Leishmania* spp. promovem, no organismo hospedeiro, uma série de respostas imunológicas. Dentre elas, pode-se observar o aumento de células do tipo T CD4+, que promovem a resposta do tipo Th1 com a produção de interferon- γ , interleucina (IL) 2 e IL-12, os quais irão estimular os macrófagos a combaterem as células infectadas (Isnard et al. 2012; Mutiso et al. 2013). No entanto, se a resposta imunológica for do tipo Th2, serão produzidas IL-4 e IL-10, que atuam na inibição da ativação dos macrófagos, consequentemente promovendo o agravamento da doença (Isnard et al. 2012; Liu and Uzonna 2012; Mutiso et al. 2013; Tibúrcio et al. 2019). Assim como nos macrófagos, esses parasitos também atuam no processo de sinalização intracelular das células dendríticas, alterando as funções de apresentação de antígenos e consequentemente a habilidade de indução de uma resposta imune eficiente (Liu and Uzonna 2012).

1.4.3 Vacinas para uso humano contra as leishmanioses

Devido aos diversos problemas associados à quimioterapia das leishmanioses, medidas profiláticas, focadas não somente no controle vetorial, têm se tornado cada vez mais necessárias. Tais medidas incluem o desenho de vacinas capazes de induzir tanto a ativação de células T CD4+ quanto T CD8+ para promover uma resposta Th1 robusta, específica contra o parasito (para evitar efeitos adversos em hospedeiros mamíferos), além de exibir um alto grau de homologia entre diferentes espécies de *Leishmania* (Isnard et al. 2012; Mutiso et al. 2013).

Os primeiros estudos envolvendo a utilização de parasitos vivos como forma de imunização contra as leishmanioses tiveram seu início na Europa, em 1910, sendo denominados de “leishmanização”. Inicialmente, o material de lesão

infecciosa ou o inóculo do parasito eram utilizados para inocular humanos não infectados. Esse método foi em grande parte descontinuado devido a uma série de razões, incluindo a dificuldade no controle de qualidade e padronização, além da persistência do parasito (Kedzierski 2010; Oliveira et al. 2013).

Posteriormente, tiveram início os estudos envolvendo vacinas de primeira geração que se baseiam em parasitos mortos ou frações do parasito, no entanto foi reportada uma baixa eficácia em ensaios clínicos (Kedzierski et al. 2006; Kedzierski 2010). No Brasil, os primeiros ensaios utilizando o parasito morto tiveram início na década de 1940 (Mutiso et al. 2013).

Já as vacinas de segunda geração incluem parasitos geneticamente modificados, vacinas de subunidades ou bactérias e vírus recombinantes que expressam antígenos de *Leishmania*. Diversas moléculas já foram testadas, incluindo glicoproteínas de superfície (GP63), proteínas de membrana (*Parasite Surface Antigen 2 - PSA-2*), e proteínas expressas ao longo de todo o ciclo celular do parasito (*Leishmania activated C kinase - LACK*), além da cisteína proteinase A2 e de proteínas de *heat shock* (Kedzierski et al. 2006; Kedzierski 2010).

Por sua vez, vacinas com parasitos vivos atenuados podem ser geradas através de alterações genéticas, eliminando o risco de reversão para o fenótipo virulento. Parasitos atenuados podem fornecer um conjunto completo de antígenos para as células apresentadoras de antígenos (Mutiso et al. 2013; Oliveira et al. 2013).

Um dos principais requisitos para uma vacina ideal contra as leishmanioses está relacionada à sua eficácia contra mais de uma espécie, capaz de proteger indivíduos contra as diferentes formas da doença ou quando ocorre sobreposição geográfica dessas espécies (Kedzierski et al. 2006).

Nos últimos anos, diversos trabalhos têm reportado a utilização de estratégias de imunoinformática para a identificação de alvos vacinais contra *Leishmania* (Vakili et al. 2018; Akya et al. 2019; Singh et al. 2019; Brito et al. 2020; Rabienia et al. 2020; Yadav et al. 2020). No entanto, em todos estes trabalhos ainda existe uma lacuna no que diz respeito à aplicação de abordagens integrativas associadas às estratégias de imunoinformática.

Apesar da complexidade em se desenvolver essas vacinas devido a fatores como virulência e diferenças genéticas entre as espécies de *Leishmania*, além de fatores do hospedeiro, a utilização de abordagens computacionais irá permitir a identificação de genes/proteínas homólogas entre as espécies, levando à identificação de epítomos compartilhados ou espécie-específicos, que podem ser incluídos em uma vacina multiepítomo contra *Leishmania*.

Adicionalmente, a utilização de estratégias computacionais para o estudo das leishmanioses foi identificada como uma abordagem promissora na prospecção de tratamentos alternativos e medidas preventivas para essas doenças. Além disso, é de extrema importância a aplicação e integração de novas estratégias, como a identificação de novos transcritos no genoma desses parasitos, além das modelagem de redes de regulação gênica para identificar esses potenciais candidatos vacinais, uma vez que essas abordagens, ao serem utilizadas em conjunto, possibilitam a busca por alvos em proteomas completos e reduzem, de forma considerável, o tempo e o custo de todo o processo.

1.5 Justificativa

As leishmanioses são um conjunto de doenças causadas por protozoários parasitos do gênero *Leishmania*, sendo o Brasil um país endêmico para todas as formas da doença. Segundo dados da OMS, estima-se que cerca de 700.000 - um milhão de novos casos sejam reportados anualmente no mundo. A vacinação é uma das iniciativas de saúde pública mais bem-sucedidas, promovendo a erradicação e a prevenção de diversas doenças. Pesquisas com o intuito de identificar novas alternativas para o tratamento de doenças negligenciadas vêm sendo consideradas como estratégicas e prioritárias pela OMS. Apesar do sucesso, vacinas para muitas doenças humanas ainda não foram desenvolvidas e estratégias foram elaboradas para identificar candidatos vacinais específicos e imunogênicos na esperança de induzir uma resposta imunológica terapêutica ou profilática. À medida que o conhecimento sobre os alvos de respostas imunológicas cresce, a seleção mais precisa dos alvos vacinais se torna cada vez mais importante. Tais estratégias incluem a prospecção computacional de imunógenos, que tem se apresentado como uma estratégia vantajosa, reduzindo custos, experimentações e tempo de desenvolvimento. Isso é relevante quando pensamos em doenças negligenciadas,

como as leishmanioses. Até o momento não existem vacinas para uso humano, e a quimioterapia contra as formas da doença são problemáticas. Idealmente, uma vacina contra as leishmanioses deve induzir tanto a ativação de células T CD4+ quanto T CD8+, promovendo uma resposta Th1 robusta, específica contra o parasito. Dentro desse contexto, a prospecção computacional de vacinas, através da integração de dados biológicos, representa uma abordagem inovadora que, através da integração de múltiplas variáveis biológicas, viabiliza a identificação de perfis sistêmicos de resposta imune e uma compreensão mais detalhada da patogênese. Além disso, espera-se que a perspectiva de integração de dados ômicos forneça uma contribuição instrumental para o desenvolvimento de vacinas para patógenos para os quais as abordagens tradicionais falharam até agora.

1.6. Hipótese

A prospecção *in silico* de epítomos com potencial imunogênico para o posterior desenvolvimento de vacinas contra as leishmanioses pode ser beneficiada pela utilização de abordagens de integração de dados biológicos de diferentes fontes.

2 Objetivos

2.1 Objetivo geral

Prospectar computacionalmente moléculas indutoras de resposta imune como potenciais candidatos vacinais contra as leishmanioses através da integração de dados biológicos de diferentes fontes.

2.2 Objetivos específicos

- Identificar transcritos não anotados previamente no genoma de *Leishmania infantum* JPCM5 e o perfil de expressão gênica de populações sensíveis e resistentes ao antimônio
- determinar o perfil de expressão gênica de macrófagos humanos não infectados e infectados por *Leishmania major* Friedlin e por *Leishmania amazonensis* PH8 construindo redes de regulação gênica que possam estar associadas à infecção por esses parasitos
- identificar regiões potencialmente imunogênicas em espécies de *Leishmania* spp.

3 Metodologia

A figura 3.1 resume as abordagens metodológicas que foram empregadas no desenvolvimento deste trabalho. Estão elencadas as principais grandes áreas de acordo com o tipo de dado biológico que foi utilizado, a estratégia (objetivos específicos) e a abordagem metodológica empregada para a realização da estratégia delineada para que o objetivo geral seja alcançado.

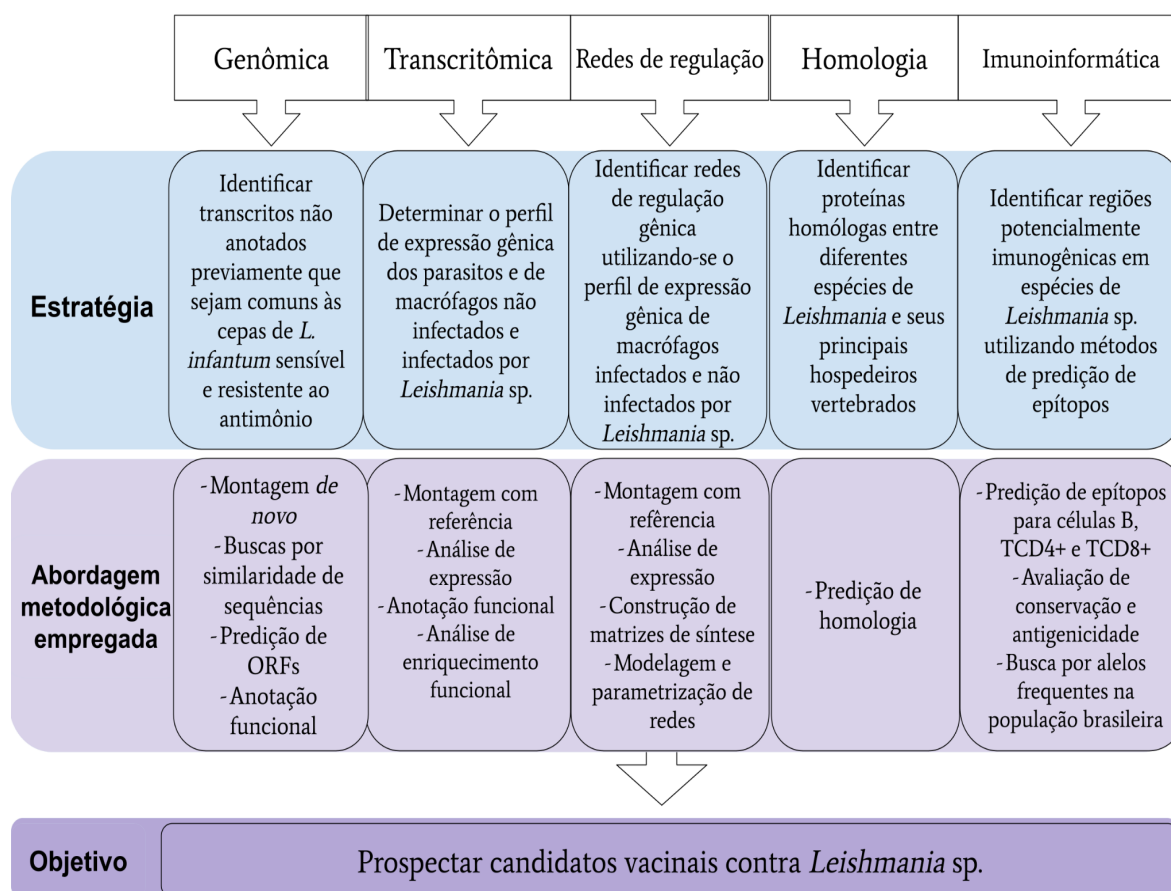


Figura 3.1 - Visão resumida das metodologias empregadas no trabalho. Grandes áreas do conhecimento, organizadas de acordo com o tipo de dado biológico que foi utilizado para o desenvolvimento do trabalho para se alcançar o objetivo final.

Fonte: Dados do trabalho

Todas as análises de bioinformática foram realizadas em sistema operacional Linux a partir dos servidores disponíveis no Grupo Informática de Biosistemas, sendo eles dois servidores SGI, totalizando 112 núcleos (AMD Opteron(TM)) e 512 GB de memória RAM; Storage SGI com 24 núcleos (Intel(R))

Xeon(R)) e 46 TB de armazenamento redundante; dois clusters PowerEdge totalizando 128 núcleos (AMD Opteron™), 1024 GB de memória RAM e 8 TB de armazenamento.

3.1 Desenho experimental

3.1.1 Sequenciamento do transcrito de *Leishmania infantum* PP75

Neste trabalho foram utilizadas formas promastigotas de *L. infantum* (MHOM/BR/74/PP75) na fase logarítmica de crescimento. As amostras resistentes (*LiSbR*) foram obtidas a partir da linhagem sensível de *L. infantum* (*LiWTS*), submetida a pressão seletiva causada pela administração gradual e contínua com Sb^{III} , conforme descrito por Liarte e Murta (2010), apresentando um índice de resistência quatro vezes maior do que o seu par sensível (Liarte and Murta 2010).

A linhagem resistente foi cultivada apenas na presença de Sb^{III} , recebendo a concentração de 0,06 mg/mL, que corresponde ao IC_{50} da amostra (quantidade de droga que inibe 50% do crescimento do parasito, ou seja, significa que na concentração de droga utilizada metade dos parasitos foram resistentes) (tabela 5.1).

Tabela 3.1 - Amostras e concentração de droga utilizada.

Amostras	Concentração de Sb^{III} utilizada	Abreviação utilizada no trabalho
<i>L. infantum</i> sensível	-	LiWTS_0
<i>L. infantum</i> resistente	0,06 mg/ml	LiWTS_0.06

Fonte: Dados do trabalho

As formas promastigotas de *L. infantum* foram lisadas e homogeneizadas na presença de um agente altamente desnaturante de acordo com as instruções do fabricante (*Invitrogen, CA, USA*). O RNA total foi obtido utilizando o *Kit* de Extração de RNA (*RNeasy-QIAGEN, Valencia, CA, USA*), também de acordo com as instruções do fabricante.

Após o processo de extração, o RNA total foi tratado com DNase (*Ambion® TURBO DNA-free™* - Invitrogen) para a remoção do DNA genômico contaminante e avaliado quanto à sua qualidade e integridade no *Bioanalyzer 2100* (*Agilent Bioanalyzer* - Santa Clara, CA, EUA) sendo em seguida submetido à síntese de cDNA.

A criação das bibliotecas foi realizada como descrito no protocolo de construção de bibliotecas, utilizando 10µg de mRNA para cada biblioteca do tipo *single-end* (Illumina, Inc., San Diego, CA, EUA).

Foram obtidas três réplicas biológicas para cada uma das condições (sensível e resistente). O sequenciamento foi realizado pela Plataforma de Sequenciamento do Instituto Pasteur, França, utilizando a tecnologia *HiSeq 2000* (Illumina, Inc., San Diego, CA, EUA).

As etapas experimentais descritas acima foram realizadas em colaboração com o Grupo Genômica Funcional de Parasitos do Instituto René Rachou (IRR).

3.2 Recuperação de dados biológicos

3.2.1 Genoma e proteomas de referência de *Leishmania*

Para a etapa de mapeamento dos dados de *RNAseq*, foi utilizada como referência a sequência genômica de *Leishmania infantum* *JPCM5*, além do arquivo no formato *General Feature Format* (GFF) contendo as anotações gênicas do genoma de referência. Esses dados foram obtidos a partir do banco de dados do *European Nucleotide Archive* (ENA), com o código de acesso ena-STUDY-CBMSO-04-04-2017 (Fuente et al. 2017).

Para as etapas de identificação de novos transcritos no genoma de *L. infantum* e análise de homologia entre diferentes espécies de *Leishmania* e seus principais hospedeiros vertebrados utilizamos os transcritomas preditos de 13 espécies de *Leishmania* (tabela 5.2), obtidos a partir do banco de dados *TritrypDB* (<http://tritrypdb.org/tritrypdb/>) (Aslett et al. 2010), versão 43.

Tabela 3.2 - Espécies de *Leishmania* utilizadas na identificação de novos transcritos e na análise de homologia entre diferentes espécies de *Leishmania* e seus principais hospedeiros vertebrados.

Espécie	Linhagem	Patogênica para humano
<i>Leishmania aethiopica</i>	L147	Sim
<i>Leishmania amazonensis</i>	MHOM/BR/71973/2269	Sim
<i>Leishmania arabica</i>	LEM/1108	Não
<i>Leishmania braziliensis</i>	MHOM/BR/75/M2903	Sim
	MHOM/BR/75/M2904	Sim
	MHOM/BR/75/M2904 _ 2019	Sim
<i>Leishmania donovani</i>	BPK282A1	Sim
	CL-SL	Sim
	LV9	Sim
<i>Leishmania enriettii</i>	LEM3045	Não
<i>Leishmania gerbilli</i>	LEM 452	Não
<i>Leishmania major</i>	Friedlin	Sim
	LV39c5	Sim
	SD75	Sim
<i>Leishmania mexicana</i>	MHOM/GT/2001/1103	Sim
<i>Leishmania panamensis</i>	MHOM/COL/81/13	Sim
	MHOM/94/PSC1	Sim
<i>Leishmania tarentolae</i>	ParrotTarII	Não
<i>Leishmania tropica</i>	L590	Sim
<i>Leishmania turanica</i>	LEM423	Sim

Fonte: TriTrypDB

Já para etapa de predição de epítomos foram utilizados os proteomas preditos de *L. infantum* JPCM5, *L. braziliensis* M2903, *L. braziliensis* M2904_2019 e *L. mexicana* 1103.

3.2.2 Transcritoma de macrófagos humanos infectados por *Leishmania* sp.

Os dados referentes ao sequenciamento do transcritoma de 66 amostras de macrófagos humanos foram baixados do banco de dados *Sequence Read Archive* (SRA) usando o *SRA toolkit* versão 2.10.2 (<https://github.com/ncbi/sra-tools>) com o código de acesso SRP062278 (Fernandes et al. 2016). Os pacotes *prefetch* e *fastq-dump* foram utilizados para obter os arquivos no formato SRA e convertê-los em arquivos no formato *Fastq* (formato de arquivo que contém a sequência da leitura e os valores de qualidade para cada uma das bases sequenciadas). Todas as amostras foram sequenciadas utilizando a plataforma *Illumina HiSeq 1500* a partir da construção de bibliotecas de mRNA do tipo *paired-end* com um tamanho de 100pb (Fernandes et al. 2016).

O estudo SRP062278 compreende um experimento de série temporal com quatro momentos diferentes (4, 24, 48 e 72 horas pós infecção (hpi)) de macrófagos humanos infectados com *Leishmania amazonensis* (IFLA/BR/67/PH8), macrófagos humanos infectados com *Leishmania major* (clone VI, MHOM/IL/80/Friedlin), macrófagos humanos não infectados e macrófagos humanos que fagocitaram *beads* de látex.

Os dados referentes ao genoma de referência de *Homo sapiens* foram obtidos do banco de dados do *Ensembl genome database* (<https://www.ensembl.org/index.html>), referentes à versão GRCh38.p13 do genoma humano, e os dados do genoma de referência de *L. amazonensis* e *L. major* foram obtidos do *TriTrypDB*, versão 43.

3.2.3 Avaliação inicial do transcrito

3.2.3.1 Análise de qualidade

O *PRINSEQ* versão 0.20.4 (Schmieder and Edwards 2011) foi utilizado para a avaliação da qualidade dos dados que foram sequenciados. Para a execução do *Prinseq*, os arquivos de entrada podem estar no formato *Fastq* ou *Sequence Alignment/Map format* (SAM) (Korpelainen et al. 2015)

O *Prinseq* avalia os dados de entrada quanto à qualidade das bases sequenciadas, quantas leituras foram geradas e seus tamanhos, porcentagem do conteúdo GC das amostras e também a presença de bases ambíguas. Abaixo é possível observar a linha de comando utilizada.

```
prinseq-lite -fastq arquivo_entrada.fastq -graph_data arquivo.gd
```

Legenda dos parâmetros:

-fastq: Arquivo contendo as leituras no formato fastq

-graph_data: Arquivo com a extensão *gd* que será gerado contendo o resultado da análise realizada pelo *Prinseq*

O formato de saída padrão do *Prinseq* é um arquivo texto que contém o resultado da análise de qualidade. No entanto, é necessário converter esse arquivo texto para um arquivo no formato *Hyper Text Markup Language* (HTML), para que se obtenha os gráficos relacionados a esse resultado. O comando utilizado para esta conversão está descrito abaixo.

```
prinseq-graphs -i arquivo.gd -html_all -png_all -log arquivo.log
```

Legenda dos parâmetros:

-i: Arquivo com a extensão *gd* que foi gerado na etapa anterior contendo o resultado da análise realizada pelo *Prinseq*

-html_all: Gera um arquivo no formato HTML com todas as informações contidas no

arquivo de entrada

-png_all: Gera os gráficos no formato png

-log: Nome do arquivo que será gerado contendo o log da execução

3.2.3.2 Remoção de sequências adaptadoras e bases de baixa qualidade

Para a remoção das sequências adaptadoras e bases de baixa qualidade o *script Trimmomatic* versão 0.30 (Bolger et al. 2014) foi utilizado. Este programa inclui uma grande variedade de opções para filtragem das leituras sequenciadas. No entanto, uma de suas principais aplicações está relacionada à remoção das sequências de adaptadores e de bases de baixa qualidade (se as bases de baixa qualidade forem identificadas no início ou no final das leituras, esses trechos são removidos. Mas caso seja no meio da sequência, a sequência toda é removida) (Korpelainen et al. 2015).

A remoção de sequências de baixa qualidade é realizada a partir da análise dos valores de qualidade que seguem uma escala chamada *Phred score*, onde, por exemplo, um valor de qualidade *Phred* igual a 20 significa que a probabilidade de erro é de 1 base a cada 100 bases sequenciadas (Illumina 2011, 2014). Na tabela 5.3 podem ser observados alguns valores de qualidade *Phred* e suas respectivas acurácias. Para este trabalho utilizamos um valor de qualidade *Phred* de no mínimo 30 e um tamanho mínimo de leitura de 64 pb.

Tabela 3.3 – Valores de Qualidade e acurácia das leituras.

Valor de qualidade <i>Phred</i>	Acurácia da leitura	Probabilidade de leitura incorreta da base (erro)
10	90%	1 em 10
20	99%	1 em 100
30	99,9%	1 em 1.000
40	99,99%	1 em 10.000
50	99,999%	1 em 100.000

Adaptado de: Illumina, 2011 e Illumina, 2014.

Para a remoção das sequências dos adaptadores utilizados durante o processo de sequenciamento das amostras de *L. infantum*, utilizamos o arquivo *TruSeq3-SE*, que contém os adaptadores padrões utilizados pela Illumina para o sequenciamento do tipo *single-end*.

Essa etapa não foi realizada nos dados referentes ao estudo SRP062278 uma vez que estes dados foram depositados no SRA sem as sequências de adaptadores e com as leituras de baixa qualidade já removidas. A remoção de sequências adaptadoras e de baixa qualidade para as demais amostras foi realizada conforme descrito abaixo.

```
java -jar /storage/bin/Trimmomatic-0.30/trimmomatic-0.30.jar SE -threads 60
-trimlog          arquivo.log          arquivo_saida.fq          arquivo_entrada
ILLUMINACLIP:TruSeq2-SE.fa:2:30:10          LEADING:10          TRAILING:10
SLIDINGWINDOW:4:30 MINLEN:64
```

Legenda dos parâmetros:

SE: Indica que as leituras são do tipo *single end*

-threads: Número de processadores utilizados

-trimlog: Arquivo *log* que será gerado

arquivo_saida.fq: Arquivo no formato *fastq* contendo as leituras de alta qualidade e sem as sequências de adaptadores

arquivo_entrada: Arquivo contendo todas as leituras do sequenciamento no formato *fastq*

-ILLUMINACLIP: Arquivo com as sequências adaptadoras

LEADING: Número de bases removidas no início da sequência

TRAILING: Número de bases removidas no final da sequência

SLIDINGWINDOW: Percorre a sequência em uma janela de 4 bases com um valor mínimo de qualidade de 20

MINLEN: Tamanho mínimo das sequências que serão mantidas no arquivo final

3.2.3.3 Segunda análise de qualidade dos dados

Após a remoção das sequências de adaptadores e/ou de baixa qualidade, foi realizada uma nova análise de qualidade para verificar se a etapa descrita anteriormente foi realizada de forma correta, ou seja, se as sequências de adaptadores e sequências contendo bases de baixa qualidade foram realmente removidas. As linhas de comando estão descritas abaixo.

prinseq-lite -fastq arquivo_entrada -graph_data arquivo.gd

Legenda dos parâmetros:

-fastq: Arquivo contendo as leituras no formato *fastq* (arquivo de saída do Trimmomatic)

-graph_data: Arquivo com a extensão *gd* que será gerado contendo o resultado da análise realizada pelo Prinseq

prinseq-graphs -i arquivo.gd -html_all -png_all -log arquivo_trimmed.log

Legenda dos parâmetros:

-i: Arquivo com a extensão *gd* que foi gerado na etapa anterior contendo o resultado da análise realizada pelo Prinseq

-html_all: Gera um arquivo no formato *HTML* com todas as informações contidas no arquivo de entrada

-png_all: Gera os gráficos no formato *png*

-log: Gera o arquivo com o log da execução

O script *MultiQC* versão 1.8 (Ewels et al. 2016) foi utilizado para gerar uma visualização gráfica que integrasse todos os arquivos *HTML* que foram gerados.

3.3 Identificação de novos transcritos

3.3.1 Montagem *de novo* do transcrito

As montagens *de novo* da amostra resistente (*LiSbR*) e da amostra sensível (*LiWTS*) foram realizadas com a utilização do script *Trinity* versão 2.4.0 (Grabherr et al. 2011), conforme a linha de comando descrita abaixo.

```
Trinity --seqType fq --max_memory 150G --left
arquivo_saída_Trimmomatic.fastq --CPU 60 --min_contig_length 300
--output trinity_assembly --no_normalize_reads
```

Legenda dos parâmetros:

-seqType: Formato de entrada das leituras

-max_memory: Tamanho máximo de memória que o programa pode alocar para utilização

-left: Arquivo de entrada, no formato *fastq*, contendo as leituras

-CPU: Número de processadores que serão utilizados pelo programa

-min_contig_length: Tamanho mínimo do *contig* a ser gerado

-output: Nome do diretório onde os resultados serão salvos

-no_normalize_reads: Não normaliza as leituras

Avaliamos a qualidade das montagens dos transcritomas das linhagens sensível e resistente com o script *TrinityStats.pl*, disponível no pacote *Trinity*. Esse programa utiliza como entrada o arquivo dos transcritos, em formato fasta, gerado na etapa anterior. Abaixo é possível visualizar a linha de comando para execução do script *TrinityStats.pl*.

TrinityStats.pl arquivo_entrada.fasta

Legenda dos parâmetros

-arquivo_entrada.fasta: Arquivo, no formato fasta, contendo as sequências dos transcritos montados

3.4.2 Identificação de transcritos não anotados previamente em *L. infantum* JPCM5

Todos os transcritos montados foram analisados quanto à similaridade de sequência contra transcritos já anotados no genoma de referência (*L. infantum* JPCM5), utilizando o programa *Basic Local Alignment Search Tool (BLAST)* versão 2.2.31+ (Altschul et al. 1990; Camacho et al. 2009), conforme linha de comando descrita abaixo.

makeblastdb -in Linfantum_annotated_transcripts.fasta -db_type nucl

Legenda dos parâmetros:

-in: Nome do arquivo de entrada

-db_type: Informa se as sequências presentes no arquivo de entrada são de proteínas (prot) ou nucleotídeos (nucl)

**blastn -query arquivo-transcritos.fasta -db Linf-annotated_transcripts
-num_alignments 500 -num_descriptions 500 -evaluate 0.00001 -out
arquivo_saida.txt -num_threads 60**

Legenda dos parâmetros:

-query: Arquivo de entrada
-db: Banco de dados
-num_alignments: Número de alinhamentos reportados no arquivo de saída
-num_descriptions: Número de descrições reportadas no arquivo de saída
-evaluate: Valor de "esperança" do alinhamento, indica a chance do mesmo ter acontecido ao acaso
-out: Nome do arquivo de saída
-num_threads: Número de processadores

Todos os transcritos que apresentaram uma identidade menor ou igual a 60% e uma cobertura mínima de 70% em relação ao tamanho total do transcrito foram mantidos para as análises posteriores.

Com as sequências restantes, a identificação de transcritos comuns entre as duas amostras foi realizada com a utilização do programa *CD-HIT-est-2d* (Li and Godzik 2006), que compara sequências de nucleotídeos a partir de dois conjuntos de dados, tendo como critério de corte uma identidade mínima de 95%.

```
cdhit-est-2d -i transcritos-Linf_res.fasta -i2 transcritos_Linf_sens.fasta -o resultado_Linf.fasta -c 0.95 -T 0
```

Legenda dos parâmetros

-i: Nome do arquivo de entrada 1
-i2: Nome do arquivo de entrada 2
-o: Nome do arquivo de saída
-c: Valor mínimo de identidade
-T: Utiliza todos os processadores disponíveis para execução

A partir deste resultado, os transcritos comuns foram agrupados de modo a se obter uma única sequência e receberam um identificador que permitisse a identificação como transcrito compartilhado entre a amostra resistente e a sensível.

Para os transcritos não compartilhados entre elas, as *flags* Res e Sens foram adicionadas aos identificadores das sequências para uma melhor identificação das mesmas.

Após essa etapa, o script *TransDecoder* (Haas et al. 2013), versão 5.5.0, foi utilizado para a identificação de possíveis regiões codificantes. O primeiro script utilizado foi o *TransDecoder.LongOrfs*, que realiza a predição das possíveis regiões codificantes, conforme a linha de comando descrita abaixo.

```
TransDecoder.LongOrfs -t Trinity.fasta -G Protozoan
```

Legenda dos parâmetros:

-t: Arquivo de entrada, no formato fasta, contendo as sequências dos transcritos
-G: Tabela de código genético utilizada

Para identificação das *Open Reading Frames (ORF)* que possam ter algum domínio funcional, foram realizadas buscas por similaridade de sequências utilizando o programa BLAST e os seguintes arquivos de entrada: arquivo, no formato fasta, contendo as ORF preditas anteriormente (arquivo com extensão .pep gerada na etapa descrita acima); os bancos de dados do *UniProt* (The UniProt Consortium 2021), *Conserved Domain Database (CDD)* (Lu et al. 2020) e *Protein family database (Pfam)* (Mistry et al. 2021); e um banco criado com as sequências das espécies de *Leishmania* descritas na tabela 5.2.

Os domínios proteicos foram identificados utilizando o programa *hmmscan*, versão 3.1b2, disponível no pacote HMMER (Finn et al. 2011), conforme a linha de comando descrita abaixo.

```
blastp -query longest_orfs.pep -db banco de dados -max_target_seqs 1  
-outfmt 6 -evaluate 0.000001 -num_threads 60 -out  
leishmania_ORFs_blastp-1hit.tab
```

Legenda dos parâmetros:

-query: Arquivo contendo as sequências de busca
-db: Banco de dados que será utilizado
-max_target_seqs: Número máximo de resultados no arquivo de saída
-outfmt: Formato do arquivo de saída (6 = formato tabular)
-evalue: Valor que indica a chance do alinhamento ocorrer aleatoriamente
-num_threads: Número de processadores utilizados pelo programa
-out: Nome do arquivo de saída

```
hmmsearch --cpu 60 --domtblout pfam.domtblout Pfam-A.hmm
arquivo_entrada.fasta
```

Legenda dos parâmetros:

--cpu: Número de processadores utilizados pelo programa
--domtblout: Salvar o arquivo em formato tabular de análises
Pfam-A.hmm: Caminho para o banco de dados Pfam

Após a execução das etapas anteriores, os resultados gerados foram utilizados pelo *TransDecoder.Predict* para identificar as sequências que continham domínios funcionais, utilizando a linha de comando descrita abaixo.

```
TransDecoder.Predict -t arquivo_entrada.fasta --retain_pfam_hits
pfam.domtblout --retain_blastp_hits blastp.outfmt6
```

Legenda dos parâmetros:

-t: Arquivo no formato fasta contendo as sequências dos transcritos
--retain_pfam_hits: Arquivo da análise do hmmsearch contra o banco de dados pfam no formato tabular
--retain_blastp_hits: Arquivo com os resultados do blastp no formato tabular

As sequências obtidas com a montagem foram submetidas a uma nova busca por similaridade de sequências, utilizando o programa *BLASTn*, tendo como banco

de dados o genoma de referência de *Leishmania infantum* JPCM5 com o intuito de identificar a localização dessas sequências no genoma.

A anotação funcional dessas sequências foi realizada utilizando os resultados das buscas por similaridade de sequências descritas acima e também a anotação nas categorias funcionais do *Gene Ontology*, realizada com o programa *OmicsBox*, (Ashburner et al. 2000; The Gene Ontology Consortium 2015). O processo de curadoria manual foi realizado utilizando o programa *Artemis* versão 13.0.0, para verificar se as ORFS estavam anotadas na fase de leitura correta (Rutherford et al. 2000). Todos estes dados foram integrados aos dados da anotação original. As sequências que não apresentaram similaridade contra alguma sequência já depositada em bancos de dados públicos não foram consideradas para a realização deste trabalho.

3.4 Mapeamento das leituras no genoma de referência e análise de expressão diferencial

3.4.1 Mapeamento das leituras no genoma de referência

Para o mapeamento das leituras obtidas através do sequenciamento de RNA, no genoma de referência, foi utilizado o programa *STAR*, versão 2.5 (Dobin et al. 2013; Dobin and Gingeras 2015). A primeira etapa foi criar um índice a partir do genoma de referência que seria utilizado para guiar a etapa de mapeamento.

Para as amostras de *Leishmania infantum* JPCM5 foi criado um índice a partir do genoma de referência disponível no banco de dados ENA. Já para os dados de macrófagos humanos infectados, outros dois índices foram criados, o primeiro contendo a sequência genômica de *Homo sapiens* e *Leishmania amazonensis* e o segundo contendo a sequência genômica de *Homo sapiens* e *Leishmania major*, obtidos a partir do *Ensembl* e do *TriTrypDB*, respectivamente.

Após a realização desse processo de indexação, a etapa posterior foi o mapeamento das amostras no genoma de referência, conforme linhas de comando descritas abaixo.

```
STAR --runMode genomeGenerate --genomeDir index/ --genomeFastaFiles  
Genoma_da_referencia.fasta --sjdbGTFfeatureExon Arquivo_annot_ref.gff
```

Legenda dos parâmetros:

--runMode: Define o modo de execução do STAR

--genomeDir: Define o diretório em que será gerado o index do genoma a ser usado como referência

--genomeFastaFiles: Define o genoma de referência a ser indexado

--sjdbGTFfeatureExon: Define que atributos do arquivo de anotação utilizar para indexar os genes da referência

```
STAR --runMode alignReads --genomeDir index/ --readFilesIn arquivo_R1.fq
```

Legenda dos parâmetros:

--outFileNamePrefix nome da amostra

--genomeDir: Diretório com a indexação das referências criado anteriormente

--readFilesIn: Arquivos com as leituras R1 e R2 a serem mapeadas

--outFileNamePrefix: Prefixo adicionado em cada arquivo de saída gerado após o mapeamento das amostras

Para a conversão dos arquivos binários gerados no mapeamento, para arquivos no formato SAM, foi utilizado o pacote *SAMtools* versão 1.3.1 (Li et al. 2009) e para a visualização dos dados, o *Integrative Genome Viewer* (IGV) (Robinson et al. 2011).

3.4.2 Análise de expressão diferencial

A contagem do número das leituras mapeadas para a realização das análises de expressão gênica foi realizada com o *HTSeq count* versão 0.5.4p3 (Anders et al. 2014). Vale ressaltar que para os dados de *L. infantum* os novos transcritos foram incluídos no arquivo GFF, que contém as informações da anotação gênica, para os quais serão extraídas as informações referentes ao número de leituras mapeadas para cada transcrito.

```
htseq-count -s no -i gene_id arquivo.sam arquivo_anotação.gff > leituras.txt
```

Legenda dos parâmetros:

-s: Define se a biblioteca foi fita específica ou não

-i: Define qual campo do arquivo gff será utilizado

arquivo.sam: Arquivo sam de entrada

arquivo_anotação.gff: Arquivo de anotação

>: Direcionador de saída

leituras .txt: Nome do arquivo de saída

Para as análises de expressão dos dados de *L. infantum* o pacote utilizado foi o *DEseq2*, versão 1.6.3 (Love et al. 2014), implementado na linguagem de programação *R*. Já para os dados de macrófagos humanos infectados com *Leishmania*, os pacotes utilizados foram *edgeR* (Robinson et al. 2010) e *Sleuth* (Pimentel et al. 2016). Para os dados de *L. infantum*, as amostras resistentes foram comparadas com as amostras sensíveis. Já para os dados de macrófagos humanos, foram comparados:

- Macrófagos não infectados com macrófagos que fagocitaram *beads* de latex, em cada um dos tempos de infecção (4 hpi, 24 hpi, 48 hpi e 72 hpi);
- Macrófagos infectados com *L. amazonensis* com macrófagos não infectados, em cada um dos tempos de infecção (4 hpi, 24 hpi, 48 hpi e 72 hpi);
- Macrófagos infectados com *L. major* com macrófagos não infectados, em cada um dos tempos de infecção (4 hpi, 24 hpi, 48 hpi e 72 hpi);
- Formas promastigotas de *L. amazonensis* com *L. amazonensis* no tempo de infecção de 4 hpi;
- Formas promastigotas de *L. major* com *L. major* no tempo de infecção de 4 hpi;

Para se definir a significância, um *p*-valor ajustado menor que 0,05 e um *Fold Change* maior do que 2 foram considerados.

A anotação e o enriquecimento funcional foram realizados com o *software Blast2GO* (Conesa et al. 2005; Conesa and Götz 2008), e a identificação de vias metabólicas ativadas pelo parasita nos macrófagos foi realizada com o programa *David* (Dennis et al. 2003)

3.5 Construção das redes de regulação gênica de macrófagos humanos infectados com *Leishmania*

A abordagem de redes de regulação gênica (do inglês, *Gene Regulatory Network*) descreve como os fatores de transcrição (FT) afetam a expressão de um conjunto de genes (alvos) (Steele et al. 2009; Ghedira et al. 2011). A identificação de elementos regulatórios poderia auxiliar para um melhor entendimento de como os sinais de entrada (produzidos pelos FTs) promovem, como saída, diferentes resultados, como consequência de mudanças nos níveis de expressão gênica (Peter and Davidson 2017). As GRNs são frequentemente representadas como um grafo em que os nós são genes e/ou seus produtos gênicos (FTs e seus genes-alvo) e as arestas correspondem à relação entre os nós.

A análise das GRN foi realizada utilizando o programa *Graphia Pro* (Freeman et al. 2020), disponível no pacote *BioLayout Express* (<http://biolayout.org/>) onde como entrada fornecemos uma matriz de correlação para todos os genes humanos identificados como diferencialmente expressos nas amostras do estudo SRP062278. Essa matriz foi calculada a partir dos valores de *counts per million* (CPM), obtidos com o pacote *Sleuth*. Como resultado, uma rede foi gerada utilizando um coeficiente de correlação de Pearson de $r \geq 0,85$. Neste contexto, os nós da rede representam os genes na matriz de correlação.

Utilizando o *Graphia*, os dados foram agrupados a partir dos genes que apresentaram perfis de expressão semelhantes usando o algoritmo *MCL* com um valor de *inflation* de 1.7. Esse valor de *inflation* está diretamente associado ao tamanho dos agrupamentos a serem gerados, onde valores de *inflation* maiores tendem a gerar agrupamentos com um menor número de genes.

Os genes contidos nos 30 agrupamentos mais significativos (ou seja, que apresentaram maior variação de expressão entre os grupos de comparação), regulados positivamente em macrófagos humanos infectados com *Leishmania* sp., foram posteriormente anotados funcionalmente usando as ferramentas *ToppGene* (Chen et al. 2009) e *David* (Dennis et al. 2003)

Para a montagem do modelo da rede de regulação gênica, foi realizada uma busca na literatura por artigos que continham as informações de interesse referentes aos genes selecionados, suas interações (como por exemplo ativação e ligação), seus parceiros e possíveis FT.

Essa informação foi categorizada como componentes da rede, recebendo a seguinte nomenclatura:

- *input node*: fatores de transcrição ativados a partir do estímulo recebido e que interagem com os genes encontrados nos agrupamentos selecionados (infecção do parasito);
- *transmission node*: genes que são ativados durante o curso da infecção;
- *output node*: genes que são ativados para uma determinada resposta imunológica;
- *mode of interaction*: forma como esses genes interagem, se é por ativação, ligação, inibição, fosforilação, entre outros.

Para serem incluídas na construção do modelo, essas interações deveriam ser confirmadas, de forma experimental, por pelo menos três artigos científicos diferentes.

O diagrama da rede de regulação gênica foi construído usando *yEd Graph Editor*, (<https://www.yworks.com/products/yed>), seguindo a notação *mEPN*, conforme descrito por Livigni e colaboradores (2018), e analisado no *Graphia Pro*.

As arestas que conectam os nós e as transições determinam a direção do fluxo de informação, representando o progresso do processo biológico ao longo do tempo (denominado pelo programa como *time-blocks*). Neste estudo, cada tempo de

infecção no desenho experimental representa um *time-block* que foi utilizado para a simulação. Dessa forma os *time-blocks* foram definidos como:

- 4 horas pós-infecção (*time-block* 0-25);
- 24 horas pós-infecção (*time-block* 26-50);
- 48 horas pós-infecção (*time-block* 51-75); e
- 72 horas pós-infecção (*time-block* 76-100).

As médias dos valores normalizados em CPM das contagens das leituras mapeadas para cada um dos fatores de transcrição nos diferentes tempos de infecção foram utilizadas como valores de parametrização para cada nó de entrada na rede. As simulações foram executadas no *Graphia Pro* usando como parâmetros 100 *time-blocks*, 500 *runs*, filtro de *Uniform distribution in Stochastic Petri Nets* (SPN) e *Distribution and Consumptive Transition in SPN Transition Type* ativados.

3.6 Predição de epítomos utilizando a abordagem de vacinologia reversa

As predições foram realizadas para epítomos de células T CD8+ com afinidade pelo Complexo de Histocompatibilidade de Classe I (MHC-I), para epítomos de células T CD4+ com afinidade pelo MHC de Classe II (MHC-II) e para epítomos de células B.

3.6.1 Predição de epítomos com afinidade pelo MHC-II

Para a predição de epítomos de MHC-II foi utilizado o script *NetMHCII* (Nielsen et al. 2003; Nielsen and Lund 2009), versão 2.2, conforme linha de comando descrita abaixo.

```
netMHCII -a alelo proteínas.fasta > netmhcii_$alelo.txt
```

Legenda dos parâmetros

-a: Nome do alelo;

proteínas.fasta: Arquivo de entrada

>: Direcionamento do resultado para um arquivo de saída.

netmhc2_\$alelo.txt: Nome do arquivo de saída que foi gerado

A predição foi realizada para um total de 61 alelos diferentes, sendo 7 alelos de camundongo e 54 alelos de humanos. Para selecionar as sequências de epítomos de interesse (identificados com as siglas WB (*week binding*) e SB (*strong binding*) no arquivo contendo a predição) no resultado da análise do *NetMHCII* e realizar a extração dos resultados, foi gerado um arquivo em formato tabular utilizando a linha de comando abaixo.

```
cat *.txt | grep -E "WB|SB" | awk '{print  
$10"\t"$3"\t"$2"\t"$2+15"\t"$1"\t"$6"\t"$11"\t"$4}' > resultado.tab
```

Legenda dos parâmetros:

cat: Comando *bash* para abrir todos os arquivos com extensão .txt

grep: Comando *bash* para realizar buscas por padrões

awk: Comando *bash* para realizar o processamento das informações selecionadas pelo grep

print: Imprime os campos selecionados

>: Direciona os campos selecionados para um arquivo

resultado.tab: Nome do arquivo de saída

5.6.2 Predição de epítomos com afinidade pelo MHC-I

A predição de epítomos de MHC-I foi realizada utilizando os scripts NetCTL versão 1.2b (Larsen et al. 2005) e NetMHC versão 4.0 (Andreatta and Nielsen 2016) conforme linhas de comandos descritas abaixo.

```
netMHC -a $allele arquivo.fasta > netmhc_$allelo.txt
```

Legenda dos parâmetros:

-a: Nome do alelo

arquivo.fasta: Arquivo contendo as sequências de proteínas

>: Direcionamento do resultado para um arquivo de saída

netmhc_\$allelo.txt: Nome do arquivo de saída que foi gerado

```
netCTL -longid -s $allele arquivo.fasta > netctl_$allelo.txt;
```

Legenda dos parâmetros:

-longid: Define o número de campos permitidos para o identificador

-s: Nome do alelo

arquivo.fasta: Arquivo contendo as sequências de proteínas

>: Direcionamento do resultado para um arquivo de saída

netctl_\$allelo.txt: Nome do arquivo de saída que foi gerado

Considerando o NetMHC, a predição foi realizada para um total de 95 alelos diferentes, sendo 6 alelos de camundongo e 89 alelos de humanos de onde foram selecionados os epítomos de interesse (identificados com as siglas WB (week binding) e SB (strong binding) no arquivo contendo o resultado geral da predição).

Por sua vez, considerando o NetCTL, foram realizadas predições para 12 supertipos de alelos. Estes supertipos de alelos são definidos como agrupamentos que compartilham características que permitem a ligação de um conjunto de peptídeos que apresentam resíduos específicos nas posições dois e nove. Os alelos

dentro de um mesmo supertipo são funcionalmente semelhantes entre si e funcionalmente diferentes de um outro supertipo. Por exemplo, as moléculas do supertipo A2 compartilham especificidade para peptídeos com resíduos hidrofóbicos na posição 2 e na posição C terminal, enquanto as moléculas do supertipo A3 reconhecem peptídeos com resíduos alifáticos na posição 2 e resíduos básicos na posição C terminal (Sidney, 2008; Francisco, 2015).

Para selecionar as sequências de epítomos de interesse nos resultados do NetCTL e do NetMHC foram desenvolvidos *scripts* que geram arquivos em formato tabular para cada um dos resultados, de acordo com as seguintes linhas de comando.

Para parsear o resultado do NetMHC

```
cat *.netMHC | grep -E "WB|SB" | awk '{print $11"\t"$3"\t"$1+1"\t"$1+9"\t"$2"\t"$12"\t"$16"\t"$4}' > resultado.tab
```

Legenda dos parâmetros:

cat: Comando *bash* para abrir todos os arquivos com extensão .txt

grep: Comando *bash* para realizar buscas por padrões

awk: Comando *bash* para realizar o processamento das informações selecionadas pelo grep

print: Imprime os campos selecionados

>: Direciona os campos selecionados para um arquivo

resultado.tab: Nome do arquivo de saída

Para parsear o resultado do NetCTL

```
for allele in A1 A24 A26 A2 A3 B27 B39 B44 B58 B62 B7 B8; do grep 'E$' *$allele.txt | awk '{print $3"\t"$5"\t"$1"\t"$1+8"\t"$allele"\t"$15}' > $allele.tmp ; done
```

Legenda dos parâmetros:

for: Comando *bash* para colocar todos os arquivos contendo os resultados das predições dos alelos (A1 A24 A26 A2 A3 B27 B39 B44 B58 B62 B7 B8) em uma variável

grep: Comando *bash* para realizar buscas por padrões, neste caso que

apresente o caractere “E” no final da linha

awk: Comando *bash* para realizar o processamento das informações selecionadas pelo *grep*

print: Imprime os campos selecionados

>: Direciona os campos selecionados para um arquivo

Sallele.tmp: Nome do arquivo de saída

3.6.3 Predição de epítomos de linfócitos B

A predição de epítomos de linfócitos B foi realizada utilizando o script *BepiPred* versão 1.0b ((Jespersen et al. 2017). Abaixo é possível visualizar a linha de comando utilizada para a execução do programa.

```
bepipred -s arquivo.fasta > arquivo.txt
```

Legenda dos parâmetros:

-s: Inclusão das sequências contidas no arquivo de entrada no arquivo de saída;

>: Direcionamento do resultado para um arquivo de saída.

Para selecionar os epítomos de interesse gerados pelo script *BepiPred* foi utilizado um script denominado *bepipred_parser.pl*, desenvolvido em linguagem de programação Perl. Para a sua execução é necessário o arquivo contendo o resultado do *BepiPred* e qual é o tamanho dos epítomos que o programa deverá selecionar. O resultado do programa é um arquivo em formato tabular.

3.6.4 Predição de localização subcelular, peptídeo sinal e topologia

A predição da localização subcelular foi realizada com a utilização do programa WoLF PSORT, versão 0.2 (Horton et al. 2007), conforme linha de comando descrita abaixo.

wolfPredict animal < arquivo.fasta > arquivo.txt

Legenda dos parâmetros:

animal: Define o táxon para qual a predição será realizada

<: Direcionamento do arquivo de entrada no formato fasta;

arquivo.fasta: Nome do arquivo de entrada

>: Direcionamento do resultado para um arquivo de saída.

arquivo.txt: Arquivo de saída

As informações de interesse no resultado completo do *WoLF PSORT* foram selecionadas e armazenadas em um arquivo em formato tabular utilizando a seguinte linha de comando.

grep -v '#' *.txt | awk '{print \$1"\t"\$2"\t"\$3}' | sed 's/,//g' > arquivo.tab

Legenda dos parâmetros:

grep -v: Seleciona as linhas que não apresentarem o padrão contido nas aspas simples (match inverso);

***.txt:** Seleciona qualquer nome de arquivo que tenha a extensão .txt

awk: Comando *bash* para realizar o processamento das informações selecionadas pelo grep

print: Imprime os campos selecionados

sed 's/,//g': Retira todas as vírgulas do resultado obtido pelo comando awk;

>: Direcionamento do resultado para um arquivo de saída

arquivo.tab: Arquivo de saída

Já a predição de peptídeo sinal foi realizada utilizando o programa *Sigcleave* (<https://www.bioinformatics.nl/cgi-bin/emboss/sigcleave>), versão 6.6.0.0, (von Heijne 1986), com a seguinte linha de comando.

sigcleave -minweight 3.5 -send 70 arquivo.fasta arquivo.txt

Legenda dos parâmetros:

-minweight: Valor de peso de *score* mínimo para a predição do local de clivagem (foi utilizado o valor recomendado como padrão);

-send: número de resíduos a ser avaliado na porção N-terminal para a presença de peptídeo sinal.

Para filtrar a partir do resultado do *Sigcleave* somente as informações de interesse, foi desenvolvido um programa na linguagem de programação *Perl*. Esse programa (*sigcleave_parser.pl*) realiza a extração dos resultados e gera um arquivo de saída em formato tabular.

Já a predição da topologia dos domínios transmembrana foi realizada utilizando o script TMHMM (Krogh et al. 2001), versão 2.0c., conforme descrito abaixo

tmhmm arquivo.fasta > arquivo.txt

Legenda dos parâmetros:

arquivo.fasta: Nome do arquivo de entrada

>: Direcionamento do resultado para um arquivo de saída.

arquivo.txt: Nome do arquivo de saída

Posteriormente, para selecionar os resultados de interesse (regiões preditas como *outside*) foi desenvolvido um script na linguagem de programação *awk*. Esse

programa realiza a extração dos resultados e gera um arquivo em formato tabular no final.

```
parser_tmhmm.awk resultado.tmhmm > arquivo.tab
```

Legenda dos parâmetros:

resultado.tmhmm : Nome do arquivo de entrada

>: Direcionamento do resultado para um arquivo de saída.

arquivo.tab: Nome do arquivo de saída

3.7 Identificação de proteínas homólogas

A identificação de proteínas que sejam homólogas entre as espécies de *Leishmania* e seus principais hospedeiros vertebrados foi realizada de modo a se tentar reduzir as chances de reações cruzadas, com potencial de provocar respostas autoimunes.

A análise de homologia foi realizada utilizando o OrthoMCL versão 2.2 (Li 2003). Utilizamos os proteomas preditos de *L. infantum* JPCM5, *L. braziliensis* M2903, *L. braziliensis* M2904 e *L. amazonensis*, além dos proteomas preditos de *Homo sapiens* (humano), *Mus musculus* (camundongo) e *Canis lupus familiaris* (cachorro). Os proteomas preditos de *Homo sapiens*, *Mus musculus* e *Canis lupus familiaris* foram obtidos a partir do NCBI, com os respectivos códigos de acesso, GRCh38.p13, GRCm39, ROS_Cfam_1.0.

O OrthoMCL contém um conjunto de scripts que são responsáveis desde a formatação adequada das sequências a serem fornecidas como entrada até a realização do agrupamento e identificação das sequências homólogas. Dessa forma, o primeiro script a ser executado foi o *orthomclAdjustFasta*, que foi responsável por acrescentar um código de três caracteres, definidos pelo usuário, e que deve ser

único para cada proteoma utilizado. Esse código foi acrescentado aos identificadores das sequências nos arquivos no formato fasta.

orthomclAdjustFasta código_3_caracteres arquivo.fasta 3

Legenda dos parâmetros:

código_3_caracteres: Abreviação para o nome do arquivo de saída que será gerado

arquivo.fasta: Nome do arquivo de entrada

3: Número de campos que serão selecionados

Após essa etapa é necessário separar desses arquivos todas as sequências com um tamanho inferior a 70 aminoácidos ou com uma porcentagem maior do que 5% de *stop codons* em sua sequência. O script *orthomclFilterFasta* foi utilizado para isso, com a seguinte linha de comando.

orthomclFilterFasta fasta-files/ 70 05

Legenda dos parâmetros:

fasta-files: Nomes dos arquivos de entrada, no formato fasta

70: Critério de corte de tamanho mínimo das sequências

05: Critério de corte da porcentagem máxima de stop codons

Esse script gera dois arquivos, um denominado *goodProteins.fasta*, que é um arquivo contendo as sequências que passaram pelos critérios de corte definidos (tamanho mínimo de 70 aminoácidos e no máximo 5% de *stop codons* na sequência), e um outro arquivo denominado *poorProteins.fasta*, que é um arquivo contendo as sequências que não passaram pelos critérios de corte definidos.

A próxima etapa, após essa filtragem, é realizar uma busca por similaridade de sequências, utilizando o BLAST (Altschul et al. 1990), tendo como arquivo referente ao banco de dados o *goodProteins.fasta* e como arquivo de entrada o mesmo arquivo *goodProteins.fasta*. Este tipo de estratégia, em relação a utilização do mesmo conjunto de dados como arquivo de entrada e banco de dados é conhecida como *all-versus-all*, sendo bastante utilizada na busca por potenciais homologias.

Utilizamos como parâmetros para a execução do BLAST, gerando um arquivo no formato tabular, um *e-value* de 10^{-5} e um número máximo de descrições de 1.000, conforme linha de comando mostrada abaixo.

Para gerar os arquivos do banco de dados:

```
formatdb -i goodProteins.fasta -p T
```

Legenda dos parâmetros

-i: Nome do arquivo de entrada

-p: Define que o banco de dados a ser gerado é de contém sequências de proteínas

Para executar o BLAST:

```
blastall -p blastp -d blastdb/goodProteins -i goodProteins.fasta -o  
goodProteins--vs--goodProteins-blastp-e5-m8-b400.tab -e 0.00001 -m 8 -b  
1000 -a 60
```

Legenda dos parâmetros:

blastall: Programa blast

-p: Tipo de blast realizado (Blastp: quando o banco de dados contém sequências de proteínas e o arquivo de entrada também contém sequências de proteínas)

-d: Banco de dados utilizado

-i: Arquivo de entrada

-o: Arquivo de saída

- e:** *e-value* utilizado
- m:** Formato do arquivo de saída (8=tabular)
- b:** Número de alinhamentos
- a:** Número de processadores

Após a etapa acima, para se iniciar o processamento do resultado da busca por similaridade de sequências é necessário utilizar o script *orthomclInstallSchema* para criar o esquema do banco de dados, conforme descrito abaixo.

orthomclInstallSchema -config_file InstallSchema

Legenda dos parâmetros:

orthomclInstallSchema: Script utilizado

-config_file: Parâmetro com o arquivo contendo as informações para criar o banco (Ex.: nome do banco e tabelas a serem criadas)

Para realizar a importação do resultado da busca por similaridade de sequências no banco de dados foi necessário realizar o parseamento do resultado gerado através do script *orthomclBlastParser*. Esse arquivo foi então carregado no banco com a utilização do script *orthomclLoadBlast*, conforme descrito abaixo.

orthomclBlastParser

```
goodProteins--vs--goodProteins-blastp-e5-m8-b400.tab fasta-checked/ >>  
my_orthomcl_dir/similarSequences.txt
```

Legenda dos parâmetros:

orthomclBlastParser: Script utilizado

goodProteins--vs--goodProteins-blastp-e5-m8-b400.tab: Arquivo de entrada

fasta-files/: Diretório onde estão os dados (arquivos fasta - resultado do script "orthomclAdjustFasta")

my_orthomcl_dir/similarSequences.txt: Arquivo de saída

orthomclLoadBlast config_file similarSequences.txt

Legenda dos parâmetros:

orthomclLoadBlast: Script utilizado

similarSequences.txt: Arquivo contendo o resultado gerado a partir do blast

config_file: Arquivo contendo as informações para carregar o arquivo "similarSequences.txt" no banco criado na etapa anterior (Esse arquivo contém a porcentagem mínima de similaridade e o *e-value* utilizado para o blast)

A partir do resultado anteriormente selecionado, utilizamos o script *orthomclPairs* para gerar uma tabela com as informações necessárias para a execução do MCL (para agrupar as seqüências) (Dongen 2008; Satuluri et al. 2010) e o script *orthomclDumpPairsFiles* para exportá-las, conforme descrito abaixo.

orthomclPairs config_file-v2 file.log cleanup=no

Legenda dos parâmetros:

orthomclPairs: Script utilizado

config_file: Arquivo contendo as configurações do banco (mesmo da etapa anterior)

file.log: Nome do arquivo log

cleanup=no: Não remover os arquivos temporários

orthomclDumpPairsFiles config_file

Legenda dos parâmetros:

orthomclDumpPairsFiles: Script utilizado

config_file: Arquivo contendo as configurações do banco (mesmo das etapas anteriores)

As etapas descritas anteriormente geram um arquivo texto denominado *mclInput* que foi utilizado para a execução do algoritmo *MCL*, responsável por realizar o agrupamento das sequências. Essa etapa gera um arquivo, denominado *mclOutput*, contendo os agrupamentos formados. Neste arquivo é adicionado então um identificador sequencial para cada um dos agrupamentos com o *script orthomclMclToGroups*. Após essa etapa, o *script orthomclSingletons* é utilizado para gerar um arquivo com os identificadores das sequências não agrupadas, conforme descrito abaixo.

```
mcl mclInput -te 40 --abc -I 1.5 -o mclOutput
```

Legenda dos parâmetros:

mcl: Programa que realiza o agrupamento

mclInput: Arquivo de entrada (gerado no passo anterior)

-te: Número de processadores utilizados

--abc: Indica que o formato do arquivo de entrada é um arquivo tabular

-I: Valor do parâmetro de "*inflation*"

-o: Nome do arquivo de saída

```
orthomclMclToGroups cluster- 1 < mclOutput > groups.txt
```

Legenda dos parâmetros:

orthomclMclToGroups : Script utilizado

cluster: Prefixo identificador do cluster

1: Identificação sequencial dos clusters

mclOutput: Arquivo de entrada contendo as informações dos grupos (gerado pelo MCL)

groups.txt: Arquivo de saída com as informações dos grupos

```
orthomclSingletons ../goodProteins.fasta groups.txt >> singletons.txt
```

Legenda dos parâmetros:

orthomclSingletons: Script utilizado

../goodProteins.fasta: Arquivo fasta contendo todas as proteínas que foram utilizadas para o blast

groups.txt: Arquivo com o resultado do MCL

singletons.txt: Arquivo de saída gerado

As sequências não agrupadas e as de agrupamentos somente das espécies de *Leishmania* utilizadas, referentes a proteínas consideradas não homólogas com humano, camundongo ou cachorro, foram priorizadas na etapa de seleção de proteínas que apresentaram epítomos potencialmente imunogênicos.

3.8 Integração de dados em um banco de dados relacional e realização das consultas

Para viabilizar as consultas e as análises comparativas de todos os dados obtidos pelas análises de homologia, predição de epítomos e pelas localizações subcelulares entre os organismos estudados, foi criado um banco de dados relacional utilizando o sistema gerenciador de bancos de dados *PostgreSQL*.

Neste banco de dados foram armazenados todos os resultados obtidos das predições de imunogenicidade, expressão e homologia. Em seguida, foram realizadas consultas em linguagem SQL (do inglês, *Structured Query Language*) para iniciar a seleção dos epítomos.

Para a identificação dos potenciais candidatos vacinais, consideramos os seguintes critérios:

a) possuir pelo menos um epítomo para linfócito B predito pelo *BepiPred* com um valor de *score* ≥ 1 e tamanho mínimo de 10 e máximo de 20 resíduos de aminoácidos;

b) localização subcelular definida como *plasma membrane* ou *extracellular* pelo programa *WoLF PSORT*;

- c) apresentar predição de peptídeo sinal pelo programa *Sigcleave*.
- d) possuir epítomos de células T CD4+ com afinidade pelo máximo possível de alelos de MHC-II;
- e) possuir epítomos de células T CD8+ com afinidade pelo máximo possível de alelos de MHC-I;
- f) baixa ou nenhuma similaridade com sequências de proteínas presentes nos hospedeiros vertebrados que foram avaliados na análise de homologia realizada com o OrthoMCL;
- g) ser expressa nos dados dos transcritomas analisados (*L. infantum* e *L. amazonensis*).

As proteínas que atendessem os critérios acima mencionados e que apresentassem epítomos preditos com afinidade por, no mínimo, 40 alelos humanos e 3 alelos de camundongo pelo *NetMHCII*, 40 alelos humanos e 3 alelos de camundongo pelo *NetMHC* e 12 supertipos de alelos pelo programa *NetCTL* foram selecionadas. A inclusão de alelos de camundongo foi considerada pela possibilidade de validação em modelo murino.

Posteriormente, foi realizada uma análise da topologia das proteínas selecionadas utilizando o *TMHMM*. Os epítomos que foram preditos dentro de uma região predita como *outside* pelo *TMHMM*, ou seja, externa à membrana plasmática, foram selecionados.

Adicionalmente, foi realizada uma busca no banco de dados *Allele Frequency Net Database* (Gonzalez-Galarza et al. 2020) com o intuito de se identificar quais seriam os alelos mais frequentes na população brasileira para priorizar a seleção de epítomos.

3.9 Análise de imunogenicidade e conservação

Após a seleção dos epítomos, foi realizada uma análise de imunogenicidade dos mesmos utilizando o programa *VaxiJen* (Doytchinova and Flower 2006, 2007), versão 2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). Para o

parâmetro *Target Organism* foi escolhido *Parasites* e como critério de corte um valor de *score* de 0,5.

A análise de conservação dos epítomos foi realizada utilizando o programa *Conservation Across Antigens* (Bui et al. 2007) disponível na base de dados IEDB (<http://tools.iedb.org/conservancy/>). O programa foi parametrizado para realizar análises para epítomos lineares, com identidade $\geq 80\%$ com o proteoma predito de todas as espécies de *Leishmania* detalhadas na tabela 5.2 . Esse programa é utilizado através de um servidor *web* .

Ao final, foram selecionados 20 epítomos que apresentaram o maior valor de *score* encontrado pelo programa *VaxiJen* e a maior taxa de conservação entre as proteínas de *Leishmania*.

Para as proteínas anotadas como hipotéticas o *InterproScan*, versão 5.32-71.0 (<https://www.ebi.ac.uk/interpro/search/sequence/>) (Quevillon et al. 2005; Mitchell et al. 2015) foi utilizado com o objetivo de se encontrar algum domínio funcional ou outra informação sobre a função biológica dessas sequências.

4 Resultados e discussão

Com o objetivo de detalhar a estratégia analítica implementada no desenvolvimento do presente trabalho, bem como prover uma forma assertiva de correlacionar os objetivos delineados, as estratégias computacionais escolhidas e os resultados alcançados, foi construído o fluxograma detalhado na figura 4.1.

A referida figura (figura 4.1), resume as abordagens metodológicas que foram empregadas no desenvolvimento deste trabalho. Na figura são elencadas as grandes áreas de atuação e desenvolvimento do presente trabalho (vinculadas ao tipo e origem do dado biológico analisado), a estratégia (vinculada aos objetivos específicos do projeto) e a abordagem metodológica empregada para alcance dos resultados desejados. Adicionalmente, são detalhadas as contribuições gerais alcançadas dentro de cada grande área (genômica, transcritômica, redes de regulação, homologia e imunoinformática) e seu vínculo com os objetivos estabelecidos.



Figura 4.1 - Visão resumida das etapas do trabalho. As abordagens metodológicas que foram empregadas no desenvolvimento deste trabalho estão elencadas de acordo com as grandes áreas em que o projeto agrega informações. Assim como as contribuições gerais de cada objetivo específico para que o objetivo principal seja alcançado

Fonte: Dados do trabalho

4.1 Análise de qualidade dos dados

4.1.1 Seleção do conjunto de dados para análise através do controle de qualidade das amostras

A avaliação da integridade e da qualidade do material biológico a ser sequenciado é fundamental para o sucesso do sequenciamento e das análises subsequentes (Imbeaud et al. 2005), sendo as estratégias convencionais, como análises das bandas 5S, 18S e 28S do RNA ribossomal (rRNA) por meio de eletroforese em gel de agarose ainda comumente utilizadas para avaliar a

integridade do RNA extraído (Schroeder et al. 2006). Neste estudo a integridade do RNA extraído foi avaliada pelo equipamento BioAnalyzer (Agilent Technologies®).

O BioAnalyzer é um equipamento que foi desenvolvido para separação e análise de amostras de DNA, RNA e proteínas. Esse equipamento utiliza microfluídica para a separação eletroforética das amostras biológicas de forma automatizada a partir de pequenas quantidades de amostras de RNA. As amostras são aplicadas em *chips* que realizam a separação do rRNA de acordo com o peso molecular para posterior detecção de fluorescência induzida por laser e que se correlaciona com a quantidade de RNA, de determinado peso molecular, presente na amostra. Os resultados são visualizados em gráficos de eletroferogramas em uma escala que varia de 1-10, onde 1 representa uma amostra cujo material (RNA) se encontra completamente degradado e 10 onde o RNA a ser sequenciado se encontra íntegro (Schroeder et al. 2006). Essa escala, conhecida como RNA Integrity Number (RIN), possibilita uma avaliação padronizada da qualidade do RNA, e sua interpretação facilita a comparação de amostras e garante a reprodutibilidade dos experimentos (Jahn, 2008).

Com a análise realizada pelo *BioAnalyzer* foi possível observar que todas as amostras apresentaram um valor de $RIN \geq 6,8$. O resultado do *BioAnalyzer* está representado na figura 4.2.

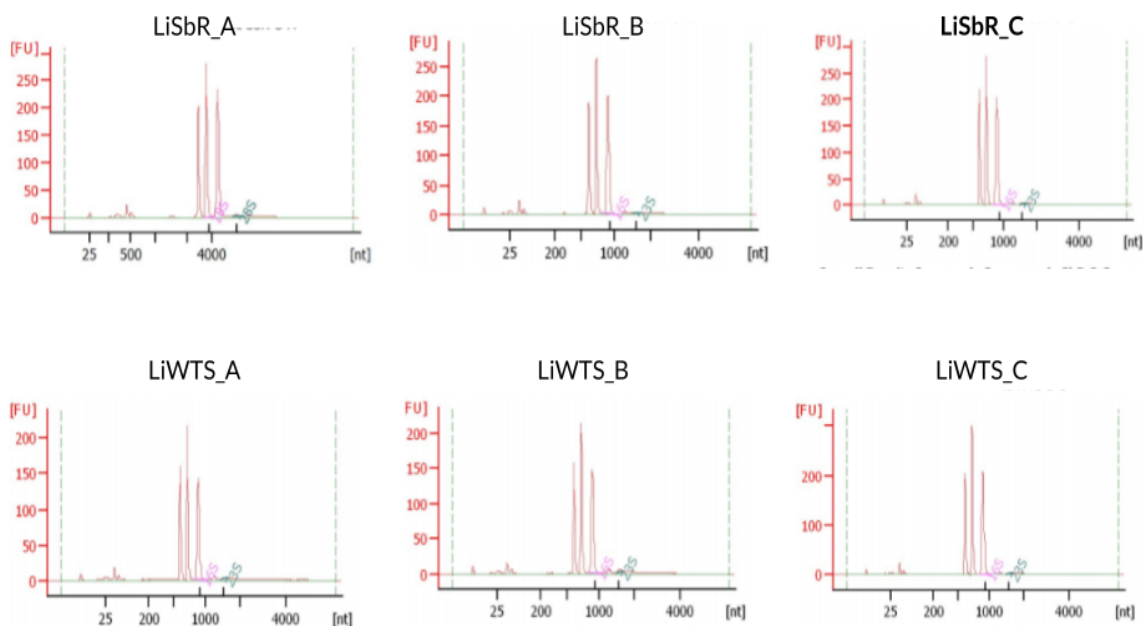


Figura 4.2 - Resultado do BioAnalyzer para as amostras de *L. infantum*. No eixo das abscissas está representado o tempo e no eixo das ordenadas, a intensidade de fluorescência. As iniciais *Li* são referentes à abreviação de *Leishmania infantum*, enquanto as abreviações *SbR* e *WTS* são referentes às amostras resistentes e às amostras sensíveis, respectivamente. As letras A, B e C se referem às réplicas biológicas.

Fonte: Dados do trabalho.

As leituras obtidas pelo sequenciamento foram analisadas quanto à qualidade utilizando-se, como critério de corte, um valor de qualidade *Phred* de no mínimo 30 e tamanho mínimo de sequência de 64 pb.

Na tabela 4.1 podem ser observados os resultados gerais da análise de qualidade que foi realizada, detalhando o total de leituras sequenciadas, a porcentagem de leituras removidas pelo *Trimmomatic* para cada uma das amostras e a cobertura de cada amostra em relação ao transcrito predito de *L. infantum* JPCM5.

A cobertura média do sequenciamento, calculada em relação ao tamanho do transcrito predito (obtido do *TriTrypDB*), foi de 261,38 vezes. A amostra LiSbR_A foi a que apresentou um maior valor de cobertura do sequenciamento (363,39

vezes) enquanto a amostra LiWTS_B apresentou o valor de cobertura mais baixo (174,95 vezes) (Tabela 4.1).

Tabela 4.1 – Resultado da avaliação de qualidade das leituras.

Condição	Amostra	Total de leituras sequenciadas	% de leituras removidas após avaliação de qualidade	Cobertura do sequenciamento
<i>L. infantum</i> selvagem	LiWTS_A	64.323.456	0,75%	290,28
	LiWTS_B	37.998.505	0,79%	174,95
	LiWTS_C	59.509.129	1,69%	237,71
<i>L. infantum</i> resistente	LiSbR_A	102.433.194	1,44%	363,39
	LiSbR_B	57.220.492	0,89%	207,41
	LiSbR_C	81.008.236	0,93%	294,80

Fonte: Dados do trabalho

Adicionalmente, as amostras foram analisadas quanto ao seu conteúdo GC e seus valores de qualidade. A figura 4.3 ilustra a distribuição do conteúdo GC para as amostras sequenciadas, mostrando os perfis de distribuição de conteúdo GC nas amostras sequenciadas.

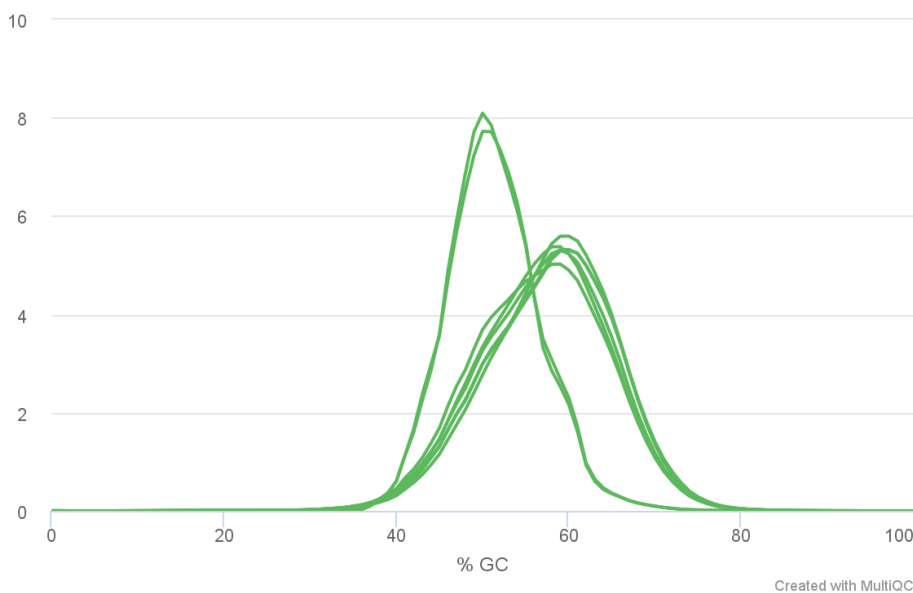


Figura 4.3 - Distribuição do conteúdo GC das amostras de *L. infantum*.

No eixo das abscissas está representada a porcentagem do conteúdo GC enquanto no eixo das ordenadas está representado o número relativo de sequências. A figura ilustra a distribuição do conteúdo GC para as amostras sequenciadas.

Tendo em vista a distribuição normal observada na análise de conteúdo GC, consideramos as amostras como ausentes de qualquer tipo de interferente (possíveis contaminantes, problemas na etapa de construção das bibliotecas ou degradação de RNA). Esses possíveis interferentes seriam observados se o gráfico apresentasse uma distribuição bimodal (onde seriam observados vários picos com diferentes porcentagens de conteúdo GC). Apesar de ser esperado em trabalhos de metagenômica, onde vários organismos são sequenciados de forma conjunta, não é o desejável quando trabalhamos com o sequenciamento de RNA de um único organismo (Schmieder and Edwards 2011).

Em relação à qualidade das bases sequenciadas, de um total de 402.493.012 leituras geradas, foram removidas 1,12% (4.525.995 leituras) que não atingiram os critérios de corte definidos, restando aproximadamente 98% do total original (397.967.017 leituras) (tabela 4.1).

A amostra LiWTS_C foi a que apresentou a maior porcentagem de leituras removidas (1,69%). A figura 4.4 ilustra a variação da qualidade das bases

sequeenciadas para as diferentes amostras antes e depois da realização dos cortes de qualidade.

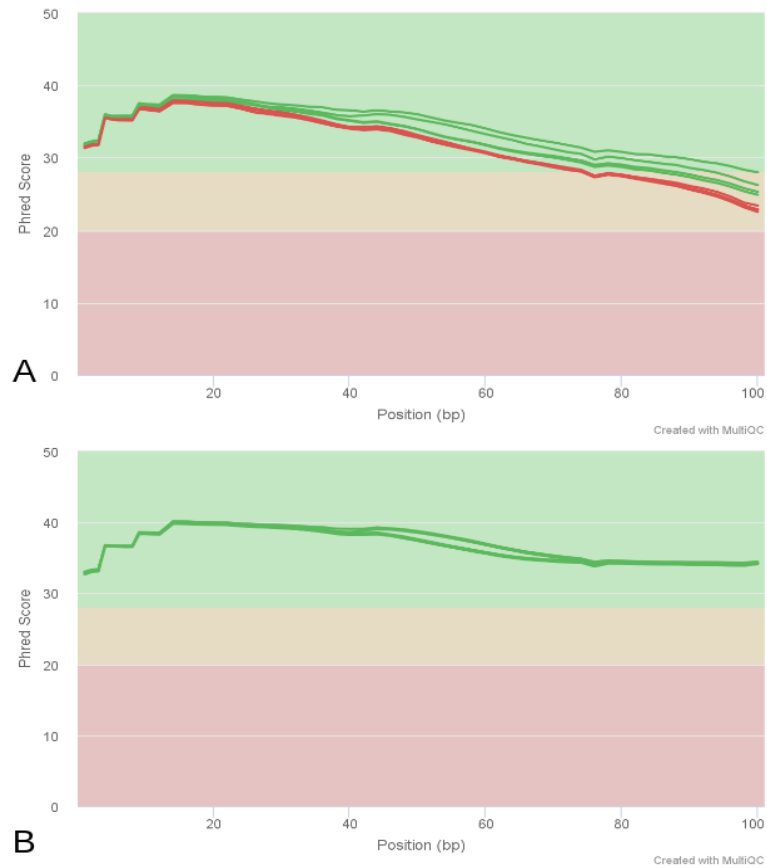


Figura 4.4 - Variação da qualidade das bases sequenciadas antes e depois dos cortes de qualidade para as amostras de *L. infantum*. No eixo das abscissas está representada a posição da base em relação à leitura, enquanto no eixo das ordenadas está representado o valor de qualidade de bases, na escala *Phred*. Em A, está representada a variação da qualidade das bases sequenciadas para as amostras de *Leishmania infantum*. Já em B, está representada a variação da qualidade das bases sequenciadas, após a aplicação dos critérios de corte, para as amostras de *Leishmania infantum*.

Fonte: Dados do trabalho

Todas as leituras geradas pelo sequenciamento foram depositadas no *Sequence Read Archive* (SRA), do *NCBI*, com os seguintes códigos de acesso: SRX2833233, SRX2833324, SRX2833326, SRX2833322, SRX2833327, SRX2833329, SRX2833328 e SRX2833325.

4.1.2 Pré processamento dos dados de macrófagos humanos infectados por *Leishmania* sp.

Os dados referentes às 66 amostras do estudo SRP062278 que foram obtidos através do SRA foram sequenciados utilizando a plataforma *Illumina HiSeq* 1500 a partir da construção de bibliotecas de mRNA do tipo *paired-end*, com fragmentos que apresentavam um tamanho de 100pb.

Na tabela 4.2 podem ser observados os resultados gerais da análise de qualidade que foi realizada, onde são detalhados o total médio de leituras e a média de cobertura do sequenciamento de cada um dos tratamentos em relação ao transcrito. Os dados, detalhados por amostra, estão descritos no Apêndice A.

Tabela 4.2 - Resumo da análise de qualidade do estudo SRP062278.

Tratamento	Média do total de leituras	Média da cobertura do sequenciamento
Macrófagos que fagocitaram <i>beads</i> de latex	43.546.335	151,64
Macrófagos infectados com <i>Leishmania amazonensis</i>	41.152.051	143,31
Macrófagos infectados com <i>Leishmania major</i>	40.934.609	142,55
Formas promastigotas metacíclicas de <i>Leishmania major</i> e <i>L. amazonensis</i>	31.817.938	127,37
Macrófagos não infectados	43.361.292	151,00

Fonte: Dados do trabalho

Em relação à qualidade das bases sequenciadas, foram obtidas um total de 2.791.262.864 de leituras, onde, para todas as amostras, foi possível observar que o *score Phred* de qualidade das bases foi de no mínimo 25, conforme pode ser observado na figura 4.5.

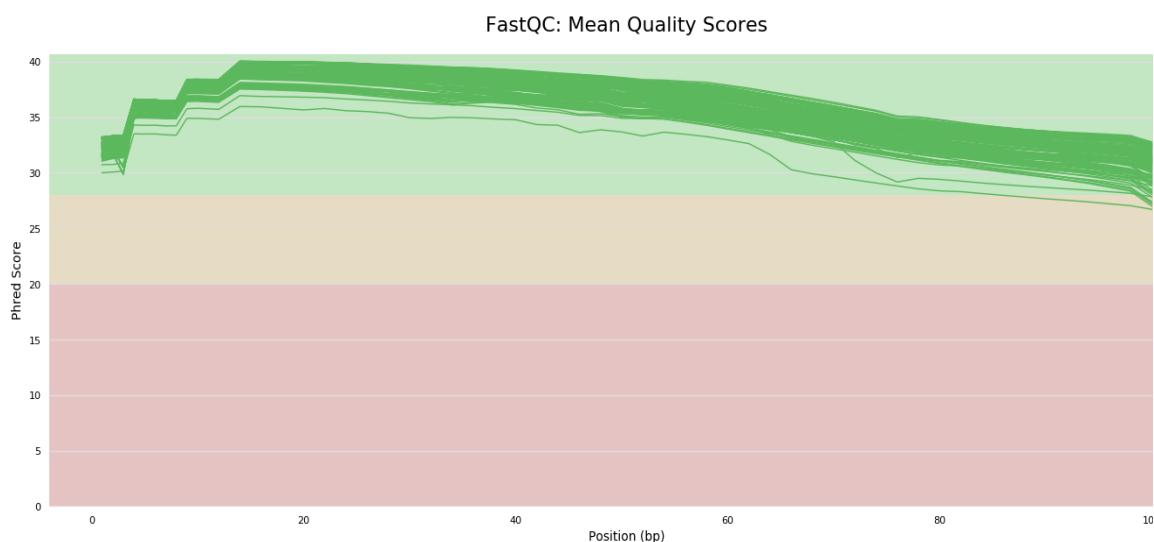


Figura 4.5 - Variação da qualidade das bases sequenciadas do estudo SRP062278. No eixo das abscissas está representada a posição da base em relação à leitura, enquanto no eixo das ordenadas está representado o valor de qualidade de bases, na escala *Phred*. A figura ilustra a variação da qualidade das bases sequenciadas para todas as amostras do estudo SRP062278.

Fonte: Dados do trabalho.

Já em relação à porcentagem de conteúdo GC das amostras (figura 4.6), é possível observar que nem todas as amostras apresentaram uma distribuição com tendência a uma curva normal. Isso pode ser explicado pelo fato de que essas amostras são referentes aos dados de macrófagos infectados com *Leishmania*. Ou seja, estamos visualizando a distribuição do conteúdo GC de dois organismos diferentes.

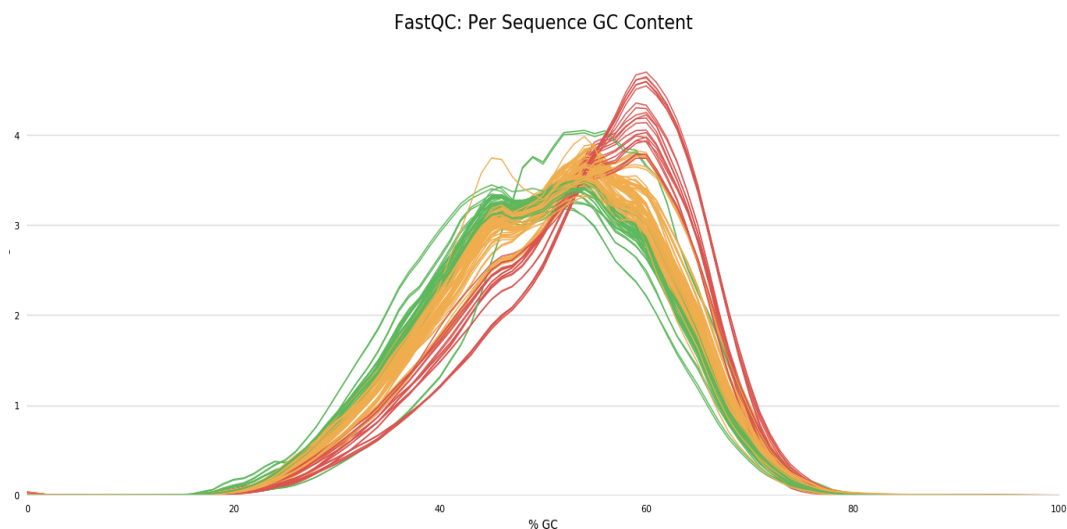


Figura 4.6 - Distribuição do conteúdo GC das amostras do estudo SRP062278. No eixo das abscissas está representada a porcentagem do conteúdo GC enquanto no eixo das ordenadas está representado o número de sequências. A figura ilustra a distribuição do conteúdo GC para todas as amostras do estudo SRP062278. Em vermelho estão representadas as amostras referentes aos macrófagos não infectados e as amostras referentes aos macrófagos que fagocitaram as *beads* de latex, enquanto em verde e em laranja estão representados as amostras referentes aos macrófagos infectados com *L. amazonensis* e macrófagos infectados com *L. major*, respectivamente.

Fonte: Dados do trabalho.

4.2 Identificação de novos transcritos em *Leishmania infantum*

Com o objetivo de detalhar a estratégia analítica implementada no desenvolvimento do presente trabalho, bem como prover uma forma assertiva de correlacionar os objetivos delineados, as estratégias computacionais escolhidas e os resultados alcançados, foi construído o fluxograma detalhando as principais etapas para a identificação de novos transcritos, detalhados na figura 4.7.

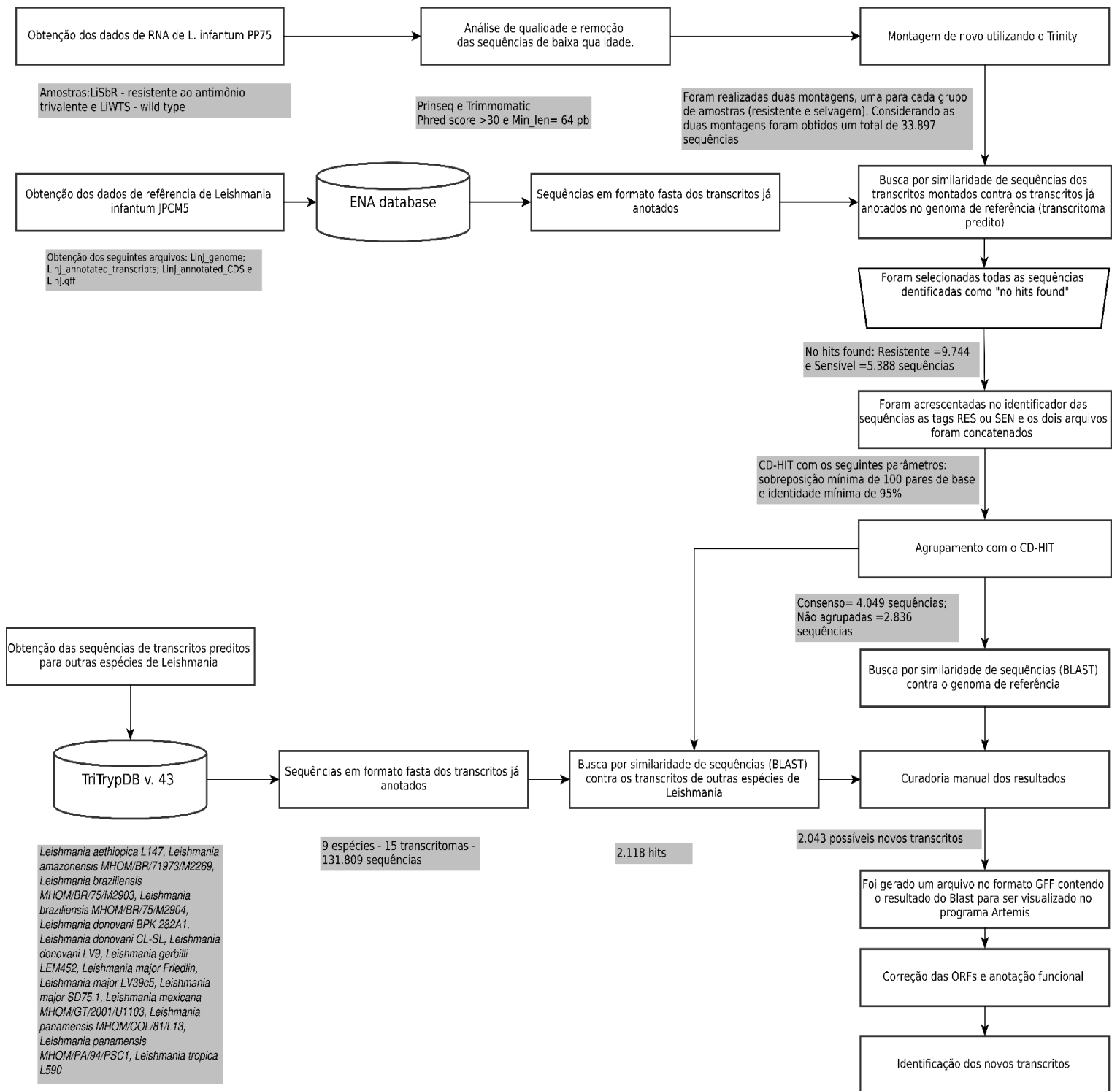


Figura 4.7 - Fluxograma geral das principais etapas empregadas na identificação de novos transcritos. Fonte: Dados do trabalho.

Visando à reconstrução de novos transcritos e considerando as particularidades relacionadas à organização genômica e da biologia de tripanosomatídeos, dentre as quais podemos citar o agrupamento sintênico, a alta conservação gênica (mais de 99% entre as espécies *L. major*, *L. infantum* e *L. braziliensis*), a transcrição policistrônica e a ausência de íntrons, optamos por realizar duas montagens (uma montagem para a amostra sensível e outra montagem para a amostra resistente), utilizando o programa *Trinity*. Utilizamos como corte um tamanho mínimo de sequência de 400 pares de base e um *K-mer* de tamanho 25.

Esta etapa identificou, para a amostra resistente, um total de 19.959 sequências, com tamanho médio de 972 pares de bases e um valor de N50 de 1.536. Já para a amostra sensível, foi possível identificar um total de 13.938 sequências com tamanho médio de 1.339 pares de bases e um valor de N50 de 2.670. Na tabela 4.3 é possível observar o resultado da etapa de montagem.

Tabela 4.3 - Resultado da etapa de montagem *de novo* do transcritoma.

Amostras	Total de transcritos montados	N50	Tamanho médio dos contigs
<i>Leishmania infantum</i> resistente	19.959	1.536	972
<i>Leishmania infantum</i> sensível	13.938	2.670	1.339
Total	33.897	-	-

Fonte: Dados do trabalho

Foi realizada uma busca por similaridade de sequências, utilizando o programa *BLASTn*, dos 33.897 transcritos montados pelo *Trinity* contra o transcritoma predito de *L. infantum* JPCM5, obtido a partir do banco de dados do *TriTrypDB* versão 43.

Observando o número de transcritos identificados por amostra (19.959 para a linhagem resistente e 13.938 para a linhagem sensível) verificamos um número maior de potenciais genes preditos pela abordagem *de novo* quando comparado ao descrito na literatura (~8.000 genes) (Clayton and Shapira 2007; Peacock et al. 2007; Fuente et al. 2017; Clayton 2019).

É interessante ressaltar que após a realização de uma busca no *TriTrypDB* referente ao histórico de atualizações das sequências de *Leishmania sp.* nele depositadas, observamos que ocorreu um aumento em relação ao número de genes preditos em depósitos ou atualizações mais recentes, como os de *L. donovani* CL-SL, *L. major* Friedlin e *L. braziliensis* MHOM/BR/75/M2903, que datam de 2019.

Trabalhos anteriores sugerem que essas espécies possuem cerca de 8.000 genes, no entanto, estudos recentes que utilizaram abordagens comparativas com outras espécies de *Leishmania* e obtiveram um aumento considerável no número de sequências preditas (Leifso et al. 2007; Peacock et al. 2007; Downing et al. 2011; Rogers et al. 2011; Alonso et al. 2016; Fuente et al. 2017; Zhang et al. 2017; Samarasinghe et al. 2018).

Dentro dos critérios de corte adotados, selecionamos as sequências que não apresentaram similaridade contra o banco de dados (resultado = “*no hits found*”). Essas sequências resultaram em um conjunto de 9.744 sequências da amostra resistente e 5.388 sequências da amostra sensível, totalizando 15.132 sequências.

Essas sequências foram concatenadas em um único arquivo e o programa CD-HIT foi utilizado com o intuito de reduzir a redundância desse conjunto de dados. Para tal, uma sobreposição de 100 pb e uma identidade mínima de 95% foram consideradas como critério para que o agrupamento ocorresse.

Um aspecto relevante, a ser destacado nesta etapa, diz respeito justamente ao uso do programa *CD-HIT* para a realização da identificação e do agrupamento das sequências comuns entre a linhagem resistente e a linhagem sensível. As abordagens atualmente descritas na literatura e voltadas à identificação de transcritos novos consideram modelos gênicos eucariotos com potencial ocorrência de eventos de *splicing* alternativo (Roberts et al. 2011).

No entanto, em *Leishmania*, assim como em outros tripanosomatídeos, eventos de *splicing* alternativo são extremamente raros e quase todos os genes não possuem íntrons. Esses parasitos apresentam transcrição policistrônica, gerando RNAs imaturos contendo mais de um gene, como ocorre em organismos procaríotos (Papadopoulou et al. 2004; Haile and Papadopoulou 2007; Gaudenzi et al. 2011; Kramer 2012).

Dessa forma, considerando os agrupamentos gerados após a utilização do programa *CD-HIT*, foram obtidas um total de 4.049 sequências consenso e 2.836 sequências não foram agrupadas. Todas essas sequências foram novamente submetidas a uma busca por similaridade de sequências contra o transcrito predito de *L. infantum* JPCM5 e nenhum *hit* significativo foi encontrado (considerando uma similaridade mínima de 40%).

Diversos estudos comparativos demonstraram que os genomas de diferentes espécies de *Leishmania* são altamente conservados e que o número de genes espécie-específicos é relativamente pequeno (Peacock et al. 2007; Smith et al. 2007; Rogers et al. 2011; Pawar et al. 2014).

O sequenciamento de linhagens de diferentes espécies de *Leishmania* também demonstrou que os genes espécie-específicos se mantêm conservados entre diferentes isolados de um mesmo subgênero de espécies (*Leishmania* *Leishmania* sp. ou *Leishmania Viannia* sp.) ou entre linhagens que foram isoladas em localizações geográficas semelhantes (Peacock et al. 2007; Smith et al. 2007; Rogers et al. 2011; Pawar et al. 2014; Tschoeke et al. 2014).

Segundo Pawar (2014), a alta sintonia apresentada por esses parasitos têm sido de grande valia quando são aplicadas abordagens *in silico*, uma vez que podem contribuir para a anotação de genomas recém sequenciados (Pawar et al. 2014). Nesse contexto, realizamos uma busca por similaridade de sequências com o *BLASTn* contra um banco de dados contendo 15 transcritos preditos de nove espécies de *Leishmania* (tabela 4.4), obtidos do *TriTrypDB* versão 43, com o objetivo de identificar no nosso conjunto de dados os genes já anotados em outras espécies.

Tabela 4.4 – Organismos utilizados para a construção do banco de dados.

Organismo	Total de transcritos preditos	Total de pseudogenes
<i>Leishmania aethiopica</i> L147	8.826	0
<i>Leishmania amazonensis</i> MHOM/BR/71973/M2269	8.127	337
<i>Leishmania braziliensis</i> MHOM/BR/75/M2903	9.379	0
<i>Leishmania braziliensis</i> MHOM/BR/75/M2904	8.493	178
<i>Leishmania donovani</i> BPK 282A1	8.135	54
<i>Leishmania donovani</i> CL-SL	9.757	0
<i>Leishmania donovani</i> LV9	8.387	832
<i>Leishmania gerbilli</i> LEM452	8.700	0
<i>Leishmania major</i> Friedlin	9.495	95
<i>Leishmania major</i> LV39c5	9.101	0
<i>Leishmania major</i> SD75.1	8.931	0
<i>Leishmania mexicana</i> MHOM/GT/2001/U1103	8.677	102
<i>Leishmania panamensis</i> MHOM/COL/81/L13	8.778	0
<i>Leishmania panamensis</i> MHOM/PA/94/PSC1	8.580	185
<i>Leishmania tropica</i> L590	8.938	0

Fonte: TriTrypDB.

Utilizando essa abordagem comparativa com outras espécies de *Leishmania*, foram identificados um total de 2.118 sequências, incluindo 75 pseudogenes, que foram excluídos das análises posteriores. Todos os 2.043 possíveis novos transcritos passaram por um processo de curadoria manual, utilizando o programa *Artemis*, onde foram ajustados as *ORFs*, os *stop codons* (códon de parada) e a metionina inicial. Todas as informações, incluindo a anotação funcional realizada utilizando os transcriptomas preditos das diferentes espécies de *Leishmania* e as informações referentes ao resultado do *BLASTn* dos transcritos preditos contra o genoma de *L. infantum* JPCM5 (*e-value*, similaridade e identidade) foram integradas ao arquivo contendo a anotação original.

Um exemplo do emprego de abordagens comparativas e de integração de dados de genoma e transcriptoma é o trabalho de Alonso e colaboradores (2016) que, a partir do ressequenciamento do genoma de *L. major* Friedlin, identificou que aproximadamente 200.000 leituras sequenciadas não mapeavam no genoma de referência utilizado para a montagem. Essas leituras foram agrupadas em 17 *contigs* e, após uma busca por similaridade de sequências contra dados depositados no *TriTrypDB*, no *NCBI* e no *GeneDB*, foi observado que elas apresentaram alta similaridade contra diferentes espécies de *Leishmania*. Utilizando dados de sequenciamento de RNA (*RNAseq*), foi observado que algumas leituras também mapeavam nesses *contigs* recém montados. Isso possibilitou a identificação de seis novos genes em *L. major* Friedlin (LmjF.15.1475, LmjF.15.0285, LmjF.24.0765, LmjF.14.0860, LmjF.19.0305 e LmjF.27.2035), que não tinham sido previamente anotados devido a problemas na montagem do genoma de referência, e a correção de outros dois genes (LmjF.15.1480 e LmjF.27.2030). Esse trabalho demonstra que a utilização de dados de sequenciamento *high throughput*, como o *RNAseq*, pode contribuir para melhorar as anotações genômicas já disponíveis permitindo, dessa forma, a identificação de aspectos moleculares ainda não conhecidos (Alonso et al. 2016).

Por sua vez, Aoki e colaboradores (2017) identificaram 443 transcritos que também não estavam mapeados em versões anteriores do genoma de *L.*

amazonensis. Essa identificação foi realizada utilizando o programa *Cufflinks*, fornecendo como entrada dados de RNAseq (Aoki et al. 2017).

Em contraponto ao que foi utilizado (leituras não mapeadas) no trabalho de Alonso e colaboradores (2016), consideramos, para a identificação de possíveis transcritos novos, apenas as leituras que mapearam em regiões do genoma para qual não se tinha, previamente, predição estrutural e funcional (Alonso et al. 2016).

Considerando o conjunto de 2.043 possíveis novos transcritos, observamos que aproximadamente 50% deles (1.055 transcritos) foram identificados como sendo comuns à linhagem sensível e à linhagem resistente. Aproximadamente 68% dessas sequências estão anotadas como proteínas hipotéticas, o que é condizente com o reportado na literatura referente à anotação funcional das sequências de tripanosomatídeos (Ivens 2005; Peacock et al. 2007). Em contrapartida, apesar da ausência de função predita, diversos estudos têm demonstrado o alto potencial imunogênico dessas sequências anotadas como hipotéticas, o que as tornam promissores alvos para intervenções terapêuticas (Coelho et al. 2012; Ravooru et al. 2014; Martins et al. 2015; Dias et al. 2018; Ribeiro et al. 2018).

Avaliando o conjunto de sequências com função predita (aproximadamente 32% dos dados) foi possível a identificação de diversas sequências já descritas na literatura como importantes para o desenvolvimento do fenótipo de resistência a antimoniais pentavalentes, como por exemplo os transportadores ABC (*ATP Binding Cassette transporter protein*) e os genes da via da tripanotiona.

Os transportadores do tipo ABC são uma família de proteínas transmembrana que está envolvida no sequestro do tiol conjugado à membrana (Basu et al. 2008). Em *Leishmania*, esses transportadores estão relacionados ao fenótipo de resistência por atuarem como “bombas de efluxo”, retirando compostos derivados de antimônio de dentro do parasito (Arcari et al. 2017). Por sua vez, a tripanotiona está envolvida nas vias de detoxificação desenvolvidas pelo parasito para se proteger das espécies reativas de oxigênio (ROS) que são produzidas tanto pelos macrófagos quanto pelos parasitos (El-Awady et al. 2017; Ilari et al. 2017).

Avaliando os transcritos identificados apenas na linhagem sensível, obtivemos um total de 386 sequências, sendo que 67% são anotadas como

hipotéticas, enquanto na linhagem resistente foram encontrados um total de 602 transcritos, sendo 52% anotados como hipotéticos. As sequências proteicas desses transcritos foram obtidas com a utilização do *software Artemis* e foram incluídas no arquivo contendo o proteoma predito para serem utilizadas nas etapas de predição posteriores.

No que diz respeito às sequências que não apresentaram nenhum *hit*, ou seja, não apresentaram similaridade de sequências contra alguma sequência depositada nos bancos de dados utilizados, elas não foram utilizadas no desenvolvimento deste trabalho, não sendo incluídas nas análises posteriores. Análises adicionais são necessárias para a verificação da real existência desses transcritos no genoma de *L. infantum*.

O aperfeiçoamento das tecnologias de sequenciamento associado a projetos de sequenciamento em larga escala aumentou a disponibilidade de genomas de patógenos e hospedeiros, em bases de dados públicas, tornando as abordagens de bioinformática mais assertivas para a identificação de potenciais candidatos vacinais e para a aplicação em testes de diagnóstico (Wu et al. 2012; Culligan et al. 2014). No entanto, até o presente momento, não foram descritos para *L. infantum* estudos que realizassem a identificação de transcritos não anotados previamente no genoma utilizando dados de expressão gênica. Além disso, os perfis transcricionais aqui relatados demonstram diferenças substanciais entre as amostras resistentes e sensíveis, além de terem possibilitado a identificação de transcritos até então não reportados por provavelmente apresentarem uma baixa expressão.

Nesse contexto, identificar de forma abrangente a expressão de transcritos é importante para uma compreensão mais abrangente de suas funções biológicas. Esses perfis do transcritoma podem fornecer uma ponte entre os estudos do proteoma e do genoma. Uma vez que os dados gerados por este trabalho se baseiam e melhoram os conjuntos de dados gênicos existentes atualmente e as anotações estruturais fornecem uma interpretação mais detalhada da biologia de *L. infantum*, eles têm o potencial de fornecer dados adicionais para a descoberta de drogas e esforços de desenvolvimento de vacinas.

Em resumo, a identificação de novos transcritos permitiu a seleção de aproximadamente mil novos transcritos comuns entre as amostras resistentes e sensíveis. Cerca de 300 transcritos foram identificados somente na amostra sensível, enquanto aproximadamente 600 transcritos foram identificados na amostra resistente. Conforme já esperado, a maior parte dessas sequências foram anotadas como codificantes de proteínas hipotéticas, no entanto também foi observada a ocorrência de proteínas que participam de diferentes vias relacionadas a mecanismos de resistência a drogas, processamento de RNA e manutenção do ciclo celular.

4.3 Identificação dos perfis de expressão em *Leishmania* sp

4.3.1 Mapeamento no genoma de referência e análise de expressão diferencial de *Leishmania infantum*

Com o objetivo de detalhar a estratégia analítica implementada no desenvolvimento do presente trabalho, bem como prover uma forma assertiva de correlacionar os objetivos delineados, as estratégias computacionais escolhidas e os resultados alcançados, foi construído o fluxograma detalhando as principais etapas para a análise de expressão diferencial, detalhados na figura 4.8.

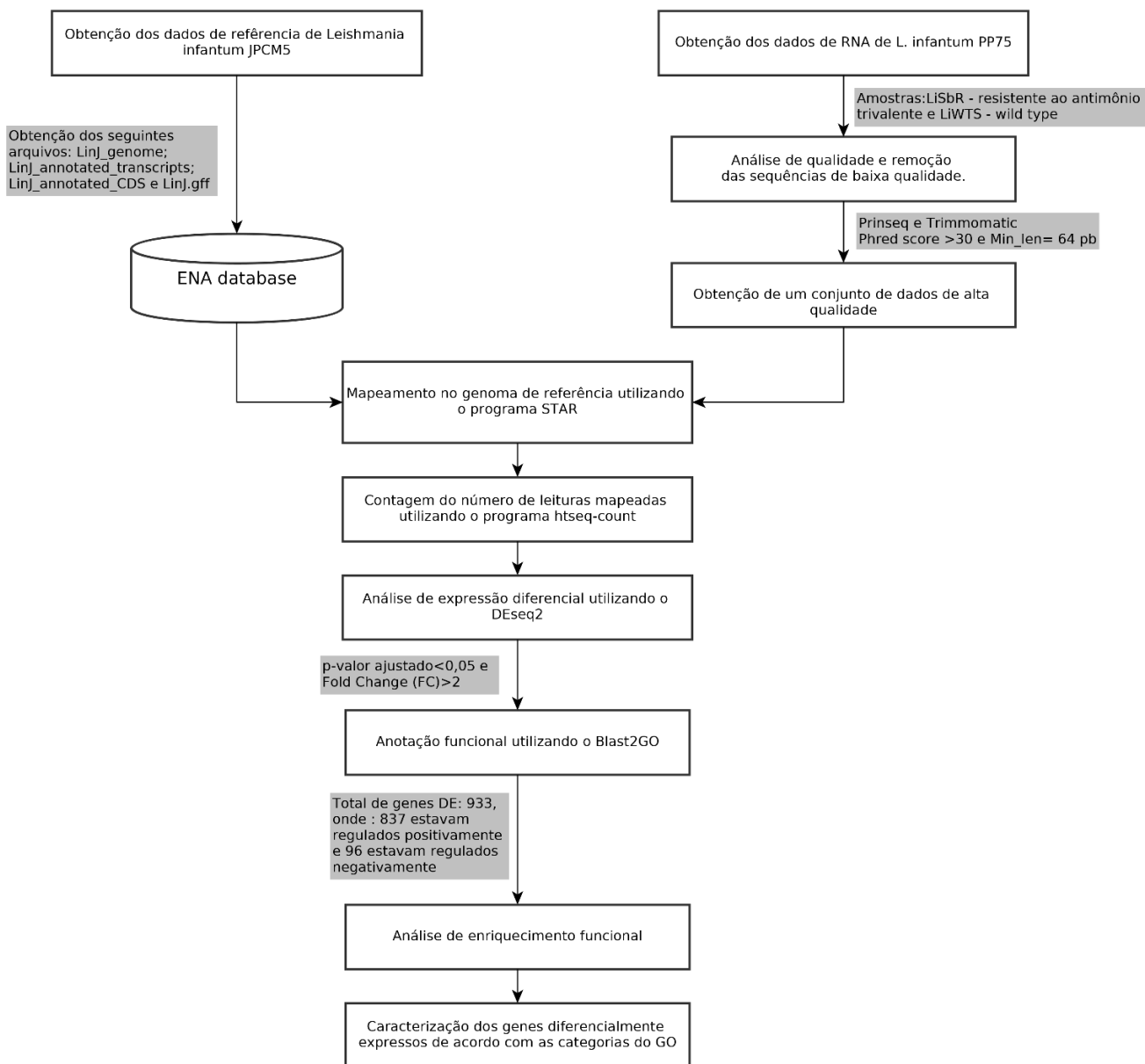


Figura 4.8 - Fluxograma geral das principais etapas empregadas na análise de expressão diferencial. Fonte: Dados do trabalho.

A sequência genômica de *L. infantum* JPCM5 foi obtida através do banco de dados *ENA*. Para otimização do processo de indexação de dados, todas as 36 sequências, referentes aos cromossomos, foram agrupadas em uma única molécula utilizando um *script* desenvolvido na linguagem *Perl*. Após essa etapa, o genoma foi então indexado e utilizado como referência para o mapeamento das leituras, que foi realizado com o programa *STAR*.

Todas as leituras que mapearam em mais de uma posição no genoma foram descartadas. Em média 9,8% das leituras não foram mapeadas no genoma de referência, sendo que a amostra que apresentou a menor porcentagem de mapeamento foi a LiSbR_A, com 12,19% de leituras não mapeadas, como pode ser observado na tabela 4.5.

Tabela 4.5 – Resultado da etapa de mapeamento das leituras no genoma de referência.

Condição	Amostra	Total de leituras	% de leituras mapeadas	% de leituras não mapeadas
<i>Leishmania infantum</i> selvagem	LiWTS_A	64.323.456	93,15%	6,85%
	LiWTS_B	37.998.505	94,02%	5,98%
	LiWTS_C	59.509.129	89,36%	10,64%
<i>Leishmania infantum</i> resistente	LiSbR_A	102.433.194	87,81%	12,19%
	LiSbR_B	57.220.492	90,43%	9,57%
	LiSbR_C	81.008.236	89,47%	10,53%

Fonte: Dados do trabalho.

Após a realização do mapeamento, as leituras foram quantificadas em termos do total de leituras associadas a cada um dos genes preditos, com o *htseq-count*, gerando uma tabela de contagem contendo o identificador dos genes preditos e o total de leituras mapeadas para cada um deles. A tabela de contagem gerada foi então utilizada para as análises posteriores.

A Análise de Componente Principal (do inglês, *Principal Component Analysis* - PCA), foi realizada com a tabela de contagem de leituras mapeadas. O PCA pode ser usado para reduzir as dimensões de um conjunto de dados, permitindo uma melhor descrição desses conjuntos e de sua variância com um número reduzido de

variáveis. As semelhanças entre os conjuntos de dados são correlacionados com as distâncias na projeção do espaço definido pelos componentes principais (PC1 e PC2) (Abdi and Williams 2010; Bro and K. Smilde 2014).

O PCA também pode ser usado para identificar *outliers* em relação aos componentes principais. O componente 1 (PC1) deve transmitir a variação máxima entre os pontos de dados (amostras) e conter erro mínimo. O componente principal 2 (PC2) é a segunda linha que atende ao PC1, perpendicularmente, e explica a segunda maior variação nos dados (Abdi and Williams 2010; Bro and K. Smilde 2014).

Essa análise mostrou que a amostra LiSbR_B apresentava um perfil de expressão diferente das demais amostras da mesma condição, dessa forma ela foi removida de todas as análises posteriores (figura 4.9).

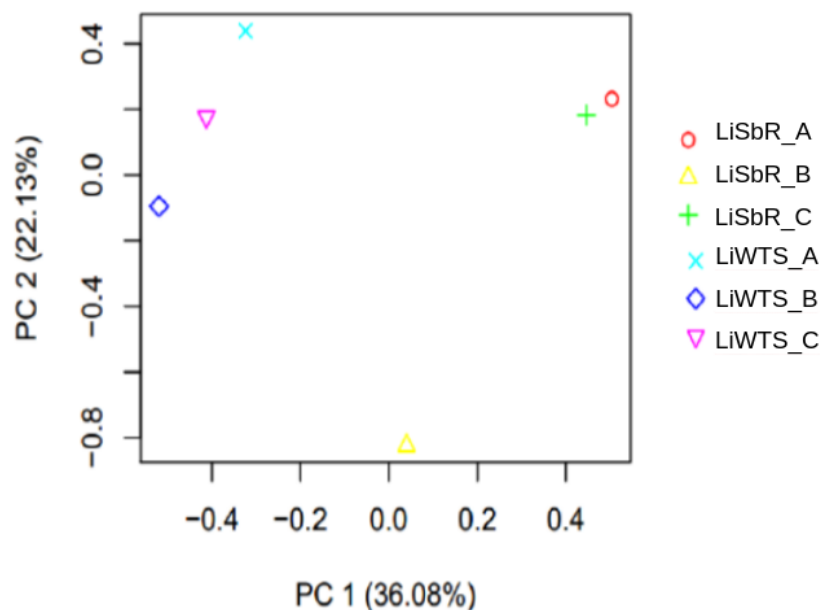


Figura 4.9 - Resultado da análise de PCA para os dados de *L. infantum*. O componente 1 (PC1) representa a variação máxima entre os pontos de dados, enquanto o componente principal 2 (PC2) descreve a segunda maior variação nos dados. Cada ponto no gráfico corresponde a uma das amostras analisadas.

Fonte: Dados do trabalho.

Para a realização da análise estatística de expressão diferencial, utilizamos o pacote *DESeq2* (baseado em uma distribuição binomial negativa) implementado na

linguagem de programação *R* e, para definirmos se um transcrito estava diferencialmente expresso, consideramos como critério de corte um *p*-valor ajustado $< 0,05$ e um \log_2 *Fold Change* (FC) > 2 .

Até o momento não existem vacinas contra leishmaniose para uso humano e o tratamento é realizado através da utilização de compostos derivados de Sb^V , como por exemplo o estibogluconato de sódio e o antimoniato de meglumina (Guerin et al. 2002). Esses quimioterápicos são considerados pró-fármacos, uma vez que são administradas em sua forma inativa (Sb^V) e são ativadas após o processo de biotransformação, no qual o Sb^V é reduzido a forma do antimônio trivalente (Sb^{III}) (Ouellette et al. 2004). Entretanto, a quimioterapia contra as infecções por *Leishmania* apresenta uma série de limitações, como a elevada toxicidade da droga, efeitos colaterais severos e elevado custo do tratamento, no caso da utilização da anfotericina B, utilizada como fármaco de segunda escolha no Brasil (Moore and Lockwood 2010; Croft and Olliaro 2011). Outro fator importante que dificulta a eficácia do tratamento nos pacientes infectados está relacionada à emergência de parasitos resistentes às drogas de primeira escolha (Iqbal et al. 2016). Assim, a compreensão dos mecanismos moleculares de resistência em *Leishmania* spp. é muito importante para identificar potenciais alvos de drogas para prevenir ou reverter tais mecanismos.

Com isso, pesquisas com o intuito de identificar novas alternativas para a terapia dessas doenças vêm sendo consideradas como estratégicas e prioritárias pela OMS.

Atualmente a vacinação é uma das iniciativas de saúde pública mais bem-sucedidas, promovendo a erradicação de doenças como, por exemplo, a varíola. Apesar do sucesso, vacinas para muitas doenças humanas ainda não foram desenvolvidas. Diversas estratégias foram elaboradas para identificar candidatos vacinais específicos na esperança de induzir uma resposta imunológica terapêutica ou profilática. À medida que o conhecimento sobre os alvos de respostas imunológicas cresce, a seleção mais precisa dos componentes vacinais se torna cada vez mais importante.

Diante disso, realizamos a análise de expressão gênica, com o intuito de

priorizar possíveis epítomos que sejam preditos em genes que foram encontrados como sendo expressos pelos parasitos. Dessa forma aumentamos a chance de uma possível exposição do epítomo ao sistema imune e ao compararmos linhagens sensíveis e resistentes podemos encontrar epítomos que por serem mais expressos nas linhagens resistentes, possam ser incluídos nas análises como forma de garantir que uma vacina também possa oferecer cobertura contra a infecção por essas linhagens.

Dessa forma, a comparação das amostras possibilitou a identificação de 933 transcritos diferencialmente expressos (DE) entre a linhagem de *L. infantum* resistente ao antimonial trivalente e a linhagem sensível ao mesmo tratamento. Dos 933 transcritos, 837 estão regulados positivamente (*up-regulated*) e 96 estão regulados negativamente (*down-regulated*).

Dos 933 transcritos DE, 504 (504/933, 54,01%) apresentaram anotação funcional e 429 (429/933, 45,99%) foram atribuídos como proteínas hipotéticas (sem função predita) (tabela 6.6). Um total de 837 (837/933, 89,71%) transcritos foram regulados positivamente e 96 (96/933, 10,29%) foram regulados negativamente em *L. infantum* resistente ao Sb^{III} (tabela 4.6).

Tabela 4.6 - Transcritos diferencialmente expressos entre as linhagens sensível e resistente de *L. infantum* classificados de acordo com a anotação funcional do *Gene Ontology* (GO).

	Categoria do GO: Processo Biológico		Categoria do GO: Componente celular ou Função molecular	Sem categoria do GO	Total
	Enriquecido	Não enriquecido			
Genes com anotação funcional					
Regulados positivamente	111 (11,90%)	13 (1,39%)	289 (31,94%)	18 (1,93%)	431 (46,19%)
Regulados negativamente	37 (3,96%)	0	6 (0,64%)	30 (3,21%)	73 (7,82%)
Proteínas hipotéticas (sem função predita)					
Regulados positivamente	32 (3,43%)	13 (1,39%)	142 (15,22%)	219 (23,47%)	406 (43,51%)
Regulados negativamente	1 (0,11%)	0	0	22 (2,36%)	23 (2,79%)
Total de transcritos diferencialmente expressos	181	26	437	289	933

Fonte: Dados do trabalho.

A análise de enriquecimento funcional dos 933 transcritos obtidos neste estudo foi realizada no *software* Blast2GO. De 933 transcritos, 644 (644/933, 69,02%) foram associadas às categorias de processo biológico (BP), função molecular (MF) ou componente celular (CC) e 289 transcritos (298/933, 30,98%) não apresentaram nenhum termo GO associado (Tabela 4.6). Um total de 207 (207/644, 32,14%) transcritos DE apresentaram termos do GO anotados na categoria de processo biológico, sendo 181 transcritos (181/207, 87,44%) enriquecidos funcionalmente e 26 transcritos (26/207, 12,56%) não enriquecidos (Tabela 4.6)

A distribuição dos 644 transcritos diferencialmente expressos, mapeados nas três ontologias do GO, está representada no diagrama de Venn (figura 4.10).

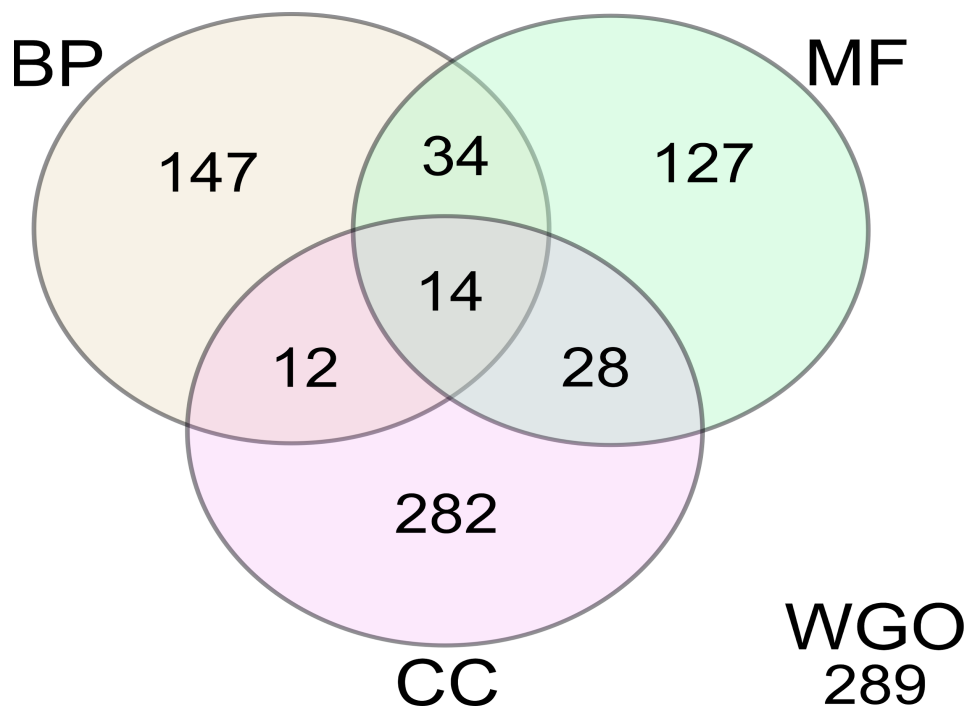


Figura 4.10 - Diagrama de Venn representando a distribuição dos transcritos diferencialmente expressos nas categorias do *Gene Ontology*. Foram encontrados um total de 14 transcritos DE, que foram anotados nas três categorias do GO, além de 289 transcritos que não apresentaram nenhuma anotação. Os demais transcritos foram anotados em duas ou apenas uma categoria.

Fonte: Dados do trabalho.

Muitos transcritos estão associados a mais de uma ontologia GO. A maioria dos transcritos correspondem às categorias de componente celular (282/644, 47,79%), seguidos por processo biológico (147/644, 22,83%) e função molecular (127/644, 19,72%). Quatorze transcritos apresentam mapeamento em todas as três categorias (Figura 4.10). Um total de 289 transcritos não foram mapeados em nenhuma das categorias do GO.

A análise de enriquecimento funcional (teste de *Fisher*) foi realizada usando como conjunto de teste a lista de transcritos DE regulados positivamente e a lista de transcritos DE regulados negativamente e como "conjunto de referência" (*background*) o proteoma predito de *L. infantum* JPCM5. Um FDR inferior a 0,05 foi estabelecido como o limite para definir a significância do enriquecimento funcional.

Dos 431 (431/834, 51,19%) transcritos enriquecidos com regulação positiva na linhagem de *L. infantum* resistente ao *Sb^{III}*, 124 (124/431, 28,77%) apresentaram

termos GO associados a processo biológico (111 enriquecidos e 13 sem enriquecimento), 289 (289/431, 67,05%) transcritos não apresentaram termos GO associados a processo biológico e 18 (18/431, 4,18%) transcritos não apresentaram termos GO associados.

4.3.1.1 Transcritos regulados positivamente em *L. infantum*

Alguns dos transcritos diferencialmente expressos identificados neste estudo corroboram análises proteômicas anteriores em linhagens de *L. infantum* resistentes ou suscetíveis ao antimônio (Matrangolo et al. 2013).

Além disso, diversos transcritos identificados neste estudo também foram investigados anteriormente pelo Grupo de Genômica funcional de Parasitos do IRR (FIOCRUZ), como a gama-glutamilcisteína sintetase (Fonseca et al. 2017), fator de alongação 2 (Moreira and Murta 2016) e manosiltransferase (Ribeiro et al. 2019), confirmando sua associação com o fenótipo de resistência em *Leishmania* spp.

A análise de enriquecimento funcional dos 837 transcritos regulados positivamente em *L. infantum* resistente ao *Sb^{III}* mostrou que os termos super-representados estavam relacionados a *quorum sensing* (GO: 0009372, GO: 0052097 e GO: 0052106), interação parasita-hospedeiro (GO: 0044764, GO : 0044114, GO: 0044115 e GO: 0044145), modificações de proteínas (GO: 1903320 e GO: 0032446), modificações pós-translacionais, fosforilação de proteínas (GO: 0006468) e ubiquitinação de proteínas (GO: 0016567), movimento baseado em microtúbulos (GO: 0007018) e regulação da distribuição de lipídios da membrana (GO: 0097035) (figura 4.11, Apêndice B).

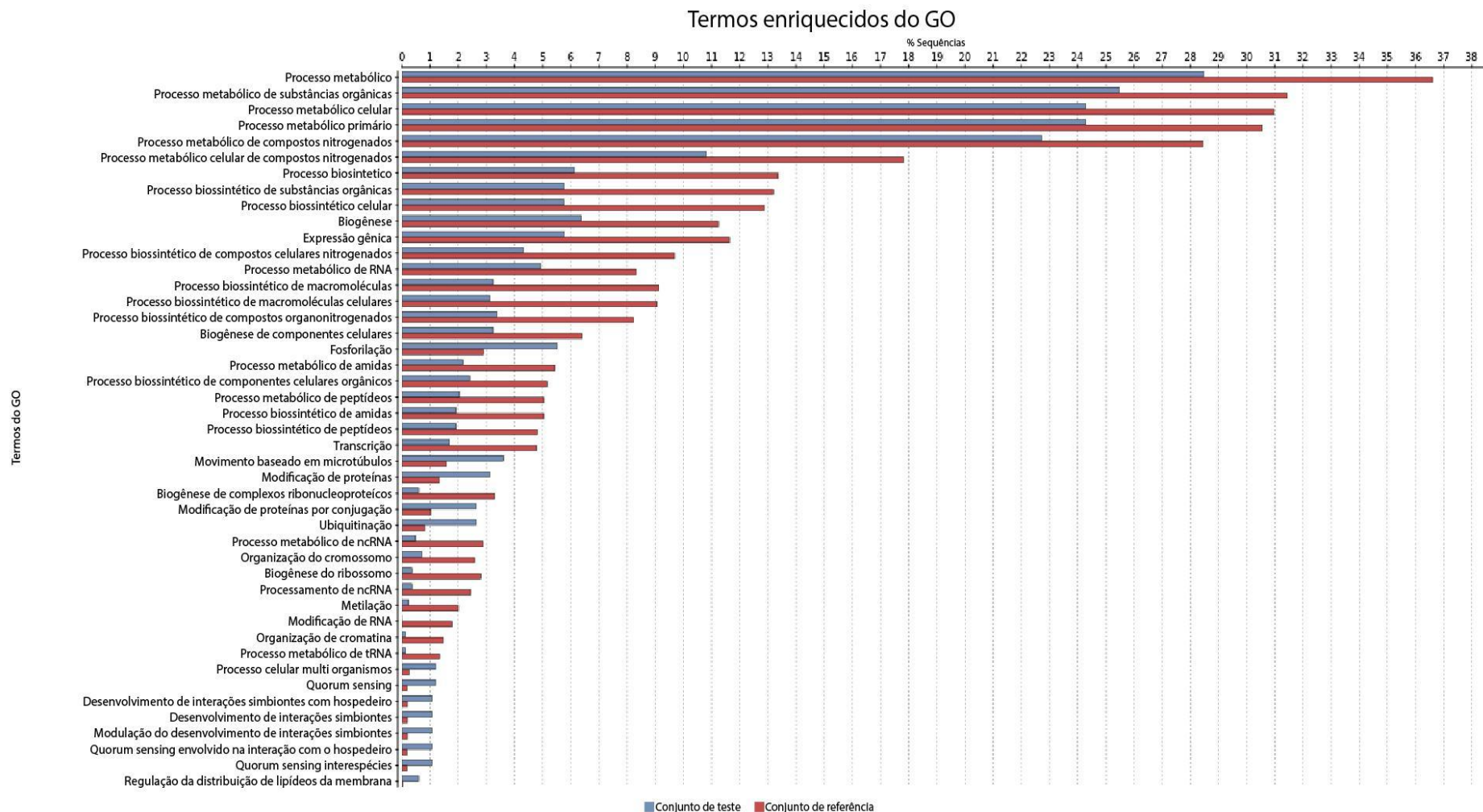


Figura 4.11 - Distribuição dos transcritos diferencialmente expressos nas categorias do Gene Ontology para o conjunto de dados regulados positivamente. A análise de enriquecimento funcional foi realizada usando como "conjunto de teste" a lista de transcritos DE regulados positivamente (representado em azul) e como conjunto de referência (background) o proteoma predito de *L. infantum* JPCM5 (representado em vermelho). Um FDR inferior a 0,05 foi estabelecido como o limite para definir a significância do enriquecimento funcional.

Fonte: Dados do trabalho.

Uma categoria que apresentou transcritos DE foi a de fosforilação de proteínas. A fosforilação de proteínas é uma das modificações pós-traducionais mais importantes que regulam vários processos de sinalização. A análise de enriquecimento GO mostrou que 37 transcritos pertencentes à categoria de fosforilação de proteína estão expressos de 2,04 a 8,93 vezes quando comparados com seu par sensível (Apêndice B).

As proteínas quinases são reguladores importantes de muitos processos celulares, como o controle da transcrição, progressão do ciclo celular, diferenciação e resposta ao estresse (Naula et al. 2005; Ardito et al. 2017). Eles representam promissores alvos de drogas, uma vez que alguns deles são essenciais para a viabilidade de parasitos e suas sequências possuem diferenças significativas com sequências homólogas de mamíferos (Naula et al. 2005).

Uma análise proteômica e fosfoproteômica comparativa de linhagens resistentes e suscetíveis a *Sb^{III}* de *L. braziliensis* identificou vários candidatos para redes bioquímicas ou de sinalização associadas ao fenótipo de resistência ao antimônio neste parasito (Matrangolo et al. 2013; Moreira et al. 2015). Na linhagem de *L. infantum* antimônio-resistente, também observamos que diferentes quinases e fosfatases são expressas diferencialmente. RAB GTPases, desempenham um papel chave na regulação das vias exocíticas e endocíticas em células eucarióticas (Matrangolo et al. 2013; Chamakh-Ayari et al. 2017). Foi demonstrado que RAB GTPases de *L. major* são altamente imunogênicas em indivíduos imunes à leishmaniose cutânea e visceral. A superexpressão de RAB6 de *L. donovani* mostrou um fenótipo resistente, permitindo que os parasitos tratados com trans-dibenzalacetona aumentassem os níveis de tiol e aumentassem a atividade da bomba MRP (Chauhan et al. 2019).

O fator de alongação 2 (EF2), é um fator relevante para a produção de proteínas, podendo ser regulado por meio da fosforilação inibitória na treonina 56 pela EF2 quinase (Ryazanov and Davydova 1989). Os transcritos desta enzima foram regulados positivamente em 3,23 vezes na linhagem resistente. Estudos mostraram que o clone de *L. braziliensis* com superexpressão de EF2 era mais resistente ao inibidor EF2K do que a linhagem sensível. Surpreendentemente, esse inibidor aumentou o efeito leishmanicida do *Sb^{III}*, sugerindo que esta associação

pode ser uma estratégia valiosa para a quimioterapia da leishmaniose (Matrangolo et al. 2013; Moreira et al. 2015).

Outros transcritos associados à desfosforilação, como a proteína fosfatase e a proteína fosfatase 2C, foram identificados com uma expressão de 17,52 a 2,28 vezes, respectivamente. A análise proteômica usando essas mesmas linhagens de *L. infantum* mostrou que ambas as enzimas também eram mais abundantes na linhagem resistente ao *Sb^{III}* (Matrangolo et al. 2013).

A categoria de ubiquitinação de proteínas também foi uma categoria cujos transcritos foram expressos diferencialmente nos parasitos resistentes. A ubiquitinação é um processo crucial em todos os organismos eucarióticos. Está envolvida em várias funções essenciais, como degradação de proteínas desnaturadas, reparo de DNA, endocitose, regulação dos níveis de proteínas, transcrição e apoptose (Wang et al. 2017; Marín 2018).

Vinte transcritos que são regulados positivamente de 2,03 a 9,14 vezes na linhagem resistente foram atribuídos a esta categoria, esses transcritos são anotados como: ubiquitina, ubiquitina-transferase (domínio HECT e domínio SPRY), *cullin protein* (envolvida na ubiquitinação por meio da participação em complexos de ubiquitina ligase) e *zinc fingers*. Semelhante aos nossos resultados, o isolado de *L. tropica* resistente ao antimônio também mostrou superexpressão de ubiquitina (Kazemi-Rad et al. 2013a). Esses dados sugerem que níveis aumentados de ubiquitinação de proteínas podem contribuir para a degradação de proteínas oxidadas, protegendo o parasito contra o estresse oxidativo causado pelo antimônio.

As proteínas do tipo *zinc finger*, serina palmitoiltransferase e RNA helicase dependente de ATP, anotadas respectivamente nas categorias "processo celular", "processo biossintético" e "montagem do complexo de ribonucleoproteína" também tiveram seus transcritos diferencialmente expressos. As proteínas do tipo *zinc finger* são proteínas de ligação de RNA envolvidas em diversos processos biológicos por meio da ligação de ácidos nucleicos ou participando de processos de transcrição ou tradução por meio da mediação de interações proteína-proteína e associação de membrana (Laity et al. 2001).

Proteínas com domínios do tipo *zinc finger* foram recentemente propostos

como potenciais alvos para Sb^{III} , devido à capacidade do Sb^{III} de competir com o zinco (Zn^{II}). Estudos anteriores sugeriram que a interação do Sb^{III} com proteínas com domínios do tipo *zinc fingers* podem modular a ação farmacológica de drogas antimoniais (Gourbal et al. 2004; Demicheli et al. 2008; Frézard et al. 2012). Em nosso estudo, quatro transcritos que codificam diferentes proteínas *zinc finger* (tipo C3HC4 - RING *finger* e FYVE) foram regulados positivamente de 2,15 a 3,77 vezes na linhagem resistente. O domínio *FYVE* é um pequeno módulo de ligação de zinco que reconhece o fosfatidilinositol 3-fosfato e a maioria dessas proteínas estão implicadas no tráfego de membrana e transdução de sinalização (Kutateladze 2006).

A serina palmitoiltransferase catalisa a primeira etapa limitante na via biossintética *de novo* de esfingolipídios (Denny et al. 2004). Essa enzima foi regulada positivamente 2,99 vezes na linhagem resistente. A análise metabolômica revelou diferenças nos conteúdos de fosfolipídios e esfingolipídios entre isolados de *L. donovani* susceptíveis e resistentes ao antimônio (t'Kindt et al. 2010). De acordo com Zhang e Beverley (2010), essas duas classes de lipídios são abundantes e críticas para a virulência e a viabilidade em *Leishmania* (Zhang and Beverley 2010).

Os transcritos que codificam helicases RNA dependentes foram encontrados regulados de 2,38 a 3,34 vezes. Esses transcritos desempenham uma função essencial no metabolismo do RNA, incluindo a degradação, tradução, regulação e edição do RNA (Møller et al. 1996; Padmanabhan et al. 2016). Foi demonstrado que um membro das RNA helicases "*DDX3 DEAD-Box*" de *Leishmania* desempenha um papel central na prevenção de danos mediados por espécies reativas de oxigênio e na manutenção do controle de qualidade da proteína mitocondrial (Padmanabhan et al. 2016).

Quatro transcritos relacionados à ATPase do tipo P foram regulados positivamente na linha LiSbR. Esses transcritos foram mapeados nas categorias de "regulação da distribuição de lipídios da membrana e translocação de fosfolipídios". ATPases são proteínas de membrana que realizam transporte ativo de íons através de membranas, sendo encontradas em bactérias e todas as células eucarióticas, incluindo *Leishmania* (Ouameur et al. 2008). Fernandez-Prada et al. (2016) demonstraram que diferentes mutações pontuais em um transportador de ATPase do tipo P em *L. infantum* estão implicadas com resistência cruzada à miltefosina e à

anfotericina B (Fernandez-Prada et al. 2016).

As categorias de "processo metabólico celular" e "processo metabólico primário"; "processo biossintético de macromoléculas celulares"; "processo metabólico de compostos nitrogenados celulares" também apresentaram transcritos DE em parasitos resistentes ao Sb^{III}, como HSPs e proteínas DNAJ. As HSPs têm funções importantes no dobramento, secreção, montagem, localização intracelular, regulação e degradação de outras proteínas (Young et al. 2004). Em geral, a resposta ao choque térmico é um mecanismo homeostático que protege as células dos efeitos deletérios do estresse ambiental, como calor e exposição a drogas (Folgueira and Requena 2007).

Relatos da literatura mostram que a superexpressão de HSPs em isolados resistentes ao antimônio de *L. donovani* (Vergnes et al. 2007; Biyani et al. 2011; Kumar et al. 2012), *L. braziliensis* e *L. infantum* (Brotherton et al. 2014; Requena et al. 2015). Um transcrito que codifica uma proteína de choque térmico de 100 KDa foi regulado positivamente em 2,86 vezes na linhagem resistente.

As proteínas DNAJ, também conhecidas como HSP40s, são parceiros cruciais para as chaperonas HSP70, e grande parte da diversidade funcional das HSP70s é impulsionada por esta classe diversa de cofatores (Requena et al. 2015). Neste trabalho, o transcrito DNAJ foi regulado positivamente em 2,13 vezes. Essa proteína desempenha um papel relevante no processo de diferenciação do estágio de promastigota para amastigota em *L. infantum*, uma vez que sofre um aumento dramático na fosforilação. Curiosamente, a expressão de HSP40 também foi encontrada aumentada em *L. donovani* resistente à artemisinina (Higgins 1992; Verma et al. 2019).

Em nosso estudo, encontramos transcritos pertencentes à classe de transportadores do tipo ABC regulados positivamente. Esses transcritos foram agrupados na categoria de "regulação da distribuição de lipídios na membrana; translocação de fosfolipídios". Os transportadores ABC compreendem uma superfamília de proteínas de membrana envolvidas no transporte dependente de ATP de uma variedade de moléculas através das membranas biológicas, incluindo aminoácidos, açúcares, peptídeos, lipídeos, íons e drogas quimioterápicas (Higgins

1992).

Esses termos têm sido associados à resistência aos medicamentos em várias doenças. Em *Leishmania*, a primeira proteína ABC identificada foi MRPA (proteína multirresistente a drogas - PgpA), que é membro da subfamília ABCC, capaz de conferir resistência a antimoniais sequestrando conjugados tiol-metal em uma vesícula intracelular (Ouellette et al. 1998; Moreira et al. 2013). Estudos anteriores mostraram uma associação da amplificação cromossômica do gene MRPA com o fenótipo de resistência a drogas em *Leishmania* spp resistente a Sb^{III} (Moreira et al. 2013). Da mesma forma, foi demonstrado que o nocaute em *L. infantum* para o gene MRPA torna o parasito mais sensível ao Sb (Douanne et al. 2020).

Uma vez que a regulação da expressão gênica em tripanossomatídeos ocorre amplamente em níveis pós-transcricionais, os principais pontos de controle na expressão gênica são a degradação e a tradução do mRNA (Clayton and Shapira 2007). As proteínas de ligação ao RNA (RBP) desempenham papéis essenciais na regulação do processamento, transporte, localização, tradução e degradação do RNA. RBPs contêm vários motivos estruturais, como motivo de reconhecimento de RNA (RRM), domínio de ligação de dsRNA, *zinc fingers* e outros (Lunde et al. 2007).

As proteínas ribossomais desempenham um papel importante na tradução e também regulam o crescimento celular e a apoptose. Em nosso estudo, o L23a, um componente da subunidade 60S, foi encontrado 2,07 vezes regulado positivamente. Uma análise proteômica mostrou que essa proteína também foi superexpressa em isolados de *L. donovani* resistentes ao antimônio (Kumar et al. 2010). Curiosamente, a linhagem de *L. donovani* com superexpressão de L23a foi mais resistente ao Sb^V, miltefosina e paromomicina (Das et al. 2013).

4.3.1.2 Transcritos regulados negativamente em *L. infantum*

Em contrapartida, a análise de enriquecimento de GO de 96 transcritos regulados negativamente em *L. infantum* resistente a Sb^{III} mostrou super-representação de todos os termos GO ligados ao processamento de rRNA (GO: 0006364), montagem de nucleossomo (GO: 0034622, GO: 0022607, GO: 0065003, GO : 0022618, GO: 0070925, GO: 0042255, GO: 0000028, GO: 0006334,

GO: 0031497, GO: 0006333, GO: 0065004 e GO: 0034063) e manutenção da fidelidade translacional (GO: 1990145) (figura 4.12 e Apêndice C).

De acordo com a análise de enriquecimento, um grupo de termos relacionados ao ribossomo estava sobrerrepresentado no conjunto de dados DE. Os transcritos que codificam proteínas ribossomais, tais como: proteínas ribossomais 40S e 60S, proteínas nucleolares e nucleares além de subunidades ribossomais, como S3a, S4, S15, S16, S17, S17, S18, S19, S21, S23 S33, L5, L7a, L11, L13, L18a, L21, L22, L31, L35 e L37, foram encontrados de 2,0 a 3,12 vezes regulados negativamente (Apêndice C).

Os transcritos que codificam as histonas H2A, H2B, H3 e H4 foram encontradas 2,0 a 2,56 vezes regulados negativamente, e foram incluídos na categoria de montagem de nucleossomos (Apêndice C).

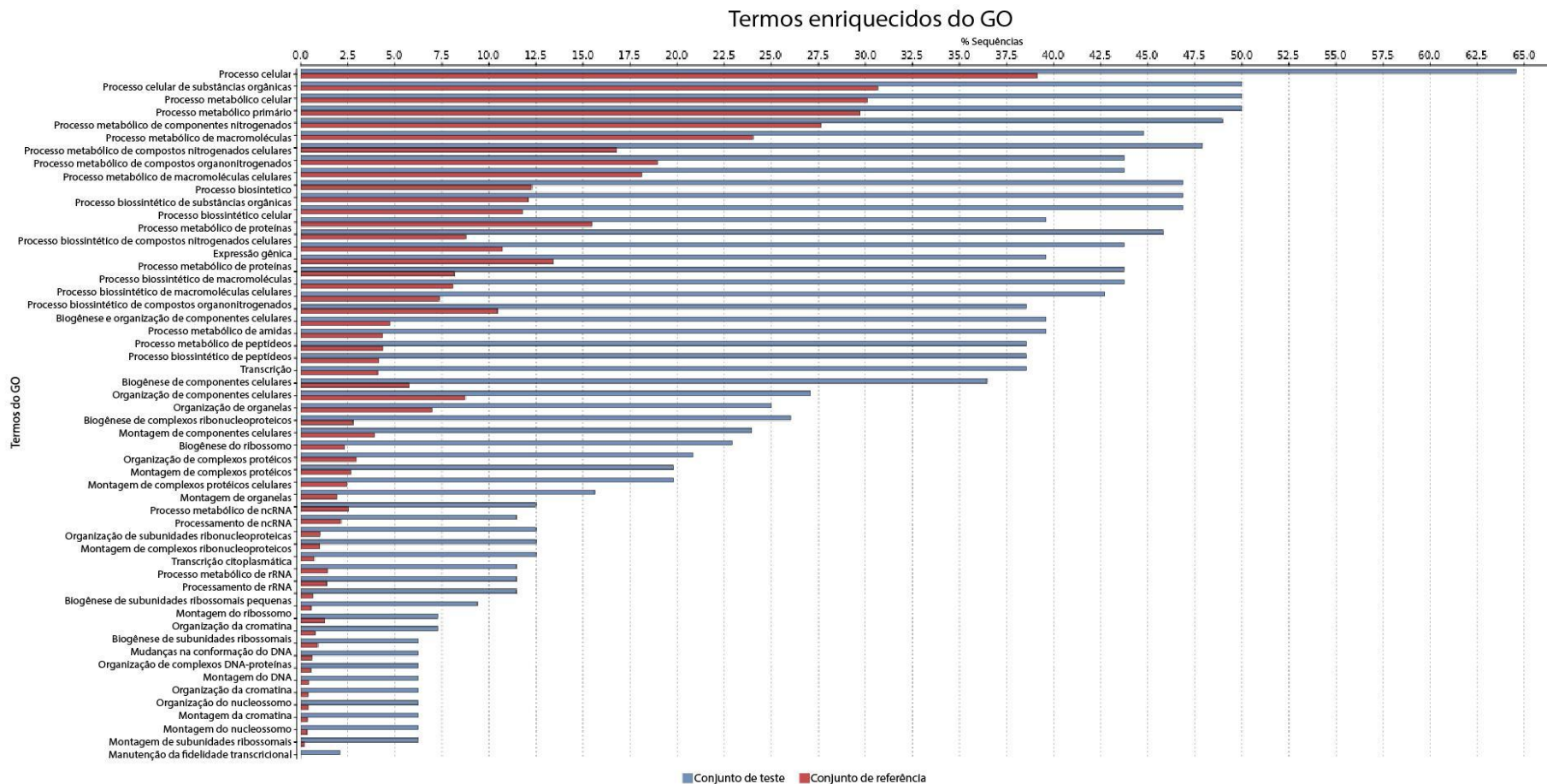


Figura 4.12 - Distribuição dos transcritos diferencialmente expressos nas categorias do *Gene Ontology* para o conjunto de dados regulados negativamente. A análise de enriquecimento funcional foi realizada usando como "conjunto de teste" a lista de transcritos DE regulados negativamente (representado em azul) e como "conjunto de referência" (*background*) o proteoma predito de *L. infantum* JPCM5 (representado em vermelho). Um FDR inferior a 0,05 foi estabelecido como o limite para definir a significância do enriquecimento funcional.

Fonte: Dados do trabalho.

4.3.1.3 Transcritos sem enriquecimento funcional para a categoria de processo biológico

Alguns transcritos diferencialmente expressos não foram relacionados a nenhuma categoria na análise de enriquecimento GO, incluindo, por exemplo: transcrito ribossomal 60S, citocromo b5 e citocromo P450 redutase, gama-glutamilcisteína sintetase, mannosyltransferase, e dois transcritos que codificam a proteína classificada como amastina (Apêndice D).

A gama-glutamilcisteína sintetase (γ -GCS) é a primeira enzima da via da glutatona que produz a γ -glutamilcisteína, um precursor direto da glutatona (Mukherjee et al. 2009). Os resultados mostraram que um transcrito que codifica essa enzima foi encontrado 2,6 vezes regulado positivamente na linhagem resistente. Foi demonstrado que o γ -GCS é essencial para a viabilidade de *L. infantum*, pois confere proteção contra o estresse oxidativo e contra o Sb^V (Mukherjee et al. 2009). O aumento dos níveis de mRNA de GSH1 também foi relatado em algumas amostras de *L. tarentolae* com resistência induzida *in vitro* ao antimônio (Guimond et al. 2003) e alguns isolados de *L. donovani* resistentes a Sb^V (Mukherjee et al. 2007). A superexpressão de γ -GCS está associada ao fenótipo de resistência ao antimônio em *L. guyanensis* (Fonseca et al. 2017).

O citocromo b5 e o P450, que estão envolvidos na atividade de oxidorreductase, foram 6,37 e 4,33 vezes regulados positivamente, respectivamente. O citocromo b5 é uma flavohemoproteína associada a reações oxidativas, como o catabolismo de xenobióticos e compostos do metabolismo endógeno (Schenkman and Jansson 2003). Mukherjee et al. (2012) observaram que a deficiência em *L. major* no domínio da oxidorreductase do citocromo b5 apresentou uma diminuição da síntese de linoleato, seguida de aumento do estresse oxidativo e morte celular por apoptose (Mukherjee et al. 2012).

O glicosilfosfatidilinositol é uma molécula de superfície importante para as interações parasito-hospedeiro. A manosiltransferase (GPI-14) é uma enzima essencial para adicionar a manose ao grupo glicosilfosfatidil. Nossos dados mostraram que um transcrito que codifica essa enzima é regulado positivamente 2,54 vezes. Um estudo superexpressou o gene GPI-14 em *L. braziliensis* e observou

o envolvimento da enzima GPI-14 no fenótipo de resistência a Sb^{III} em *L. braziliensis* (Ribeiro et al. 2019).

4.3.1.4 Transcritos anotados como *Hypothetical proteins*

Um total de 429 transcritos diferencialmente expressos foram identificados como codificantes de proteínas hipotéticas (sem função predita). Destes, 406 transcritos foram regulados positivamente e 23 foram regulados negativamente.

Dos 429 transcritos DE, 46 apresentaram ontologia GO no processo biológico (32 funcionalmente enriquecidos e 13 sem enriquecimento), 142 transcritos não apresentaram ontologia GO na categoria de processo biológico e 241 transcritos não tiveram nenhum termo GO associado.

De acordo com a análise de enriquecimento de GO, alguns termos relacionados a processos biológicos estavam sub ou super-representados nos transcritos DE. Semelhante aos transcritos DE com função predita, os principais termos enriquecidos foram movimento baseado em microtúbulos, fosforilação de proteínas, ubiquitinação de proteínas, *quorum sensing* envolvido na interação com o hospedeiro e formação de complexos ribonucleoproteicos.

Em resumo, foram encontrados 933 transcritos DE, sendo 837 regulados positivamente e 96 regulados negativamente em *L. infantum* resistente ao Sb^{III}. Foi identificado um conjunto de dados robusto que evidencia mecanismos de resistência a fármacos através da ativação de vias bioquímicas que têm o potencial de proteger o parasito do stress oxidativo causado pelo antimônio. Essas análises também possibilitaram a identificação de transcritos expressos em formas promastigotas que podem ser utilizados como potenciais candidatos vacinais contra *Leishmania*, através da construção de uma vacina multiepítipo que também seja eficaz contra a infecção por parasitos resistentes, uma vez que, no Brasil, estima-se que cerca de 25% dos insucessos terapêuticos estejam relacionados à resistência.

4.3.2 Mapeamento nos genomas de referência e análise de expressão diferencial dos dados de macrófagos infectados com *Leishmania*

Com o objetivo de detalhar a estratégia analítica implementada no desenvolvimento do presente trabalho, bem como prover uma forma assertiva de correlacionar os objetivos delineados, as estratégias computacionais escolhidas e os resultados alcançados, foi construído o fluxograma detalhando as principais etapas para a identificação de genes diferencialmente expressos em dados de macrófagos humanos infectados com *Leishmania*, detalhados na figura 4.13.

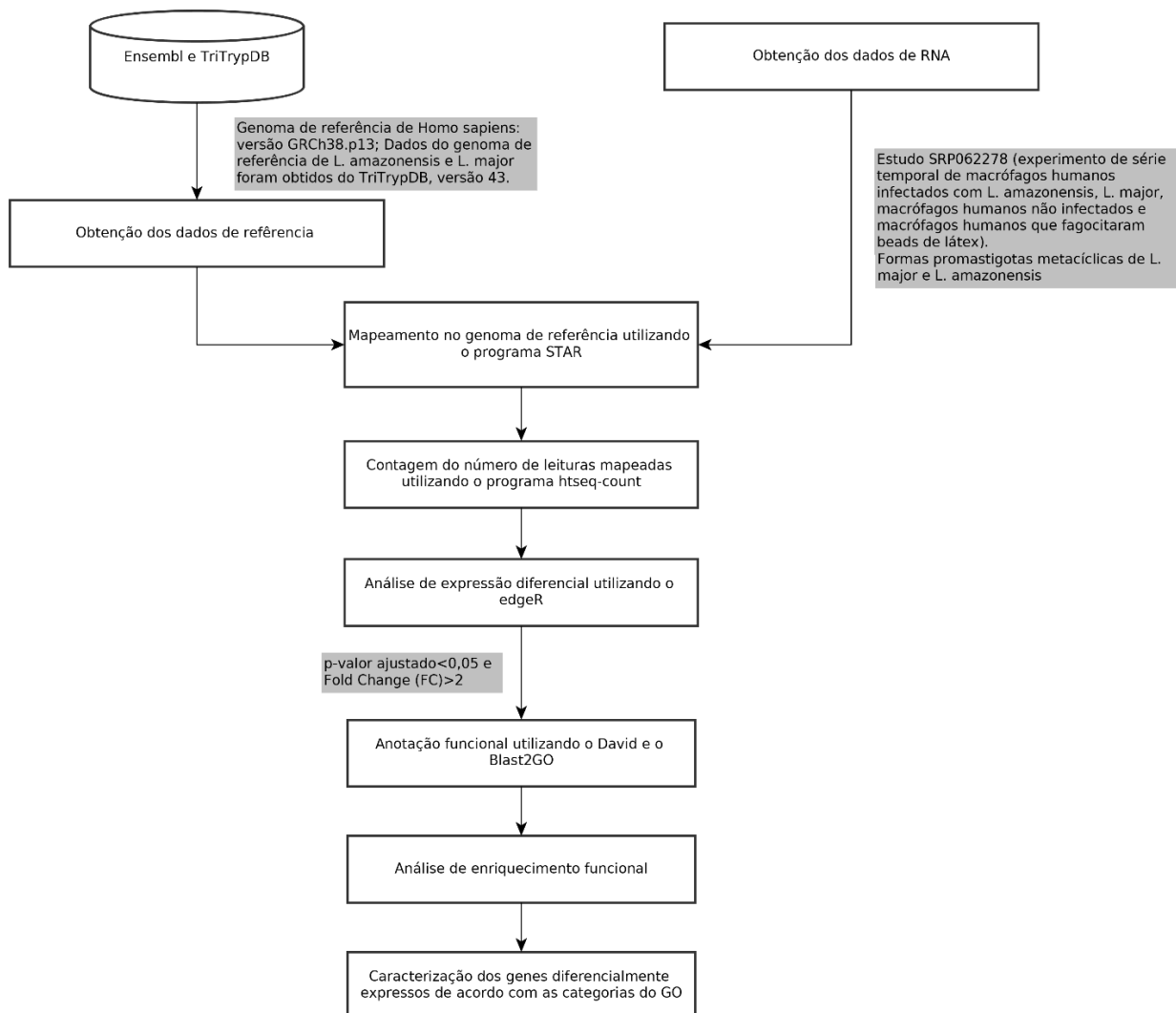


Figura 4.13 - Fluxograma geral das principais etapas empregadas na análise de expressão diferencial com dados de macrófagos humanos infectados com *Leishmania*. Fonte: Dados do trabalho.

Os dados de macrófagos infectados por *Leishmania* foram mapeados contra os genomas de referência (*Homo sapiens* e *L. amazonensis* ou *L. major*), utilizando o programa STAR. A taxa de mapeamento médio entre as amostras foi de 85%. Na tabela 4.7 é possível observar a porcentagem de mapeamento e total de leituras mapeadas para cada um dos tratamentos utilizados. A tabela contendo os resultados para cada uma das amostras está no Apêndice E.

Tabela 4.7 - Resumo dos resultados do mapeamento das leituras nos genomas de referência.

Tratamento	Média do total de leituras mapeadas	% de leituras mapeadas
Macrófagos que fagocitaram <i>beads</i> de látex	38.635.922,55	88,72
Macrófagos infectados com <i>Leishmania amazonensis</i>	33.350.101,92	81,04
Macrófagos infectados com <i>Leishmania major</i>	34.481.747,26	84,24
Formas promastigotas metacíclicas de <i>Leishmania major</i> e <i>L. amazonensis</i>	27.045.247,72	86,07
Macrófagos não infectados	38.749.292,96	89,36

Fonte: Dados do trabalho

Após o mapeamento, o *htseq-count* foi utilizado para gerar as tabelas de contagens que seriam utilizadas para a análise de expressão diferencial. A análise destes dados com o pacote *Sleuth* mostrou que seria necessário remover alguns *outliers*, conforme demonstrado na análise de PCA (figura 4.14).

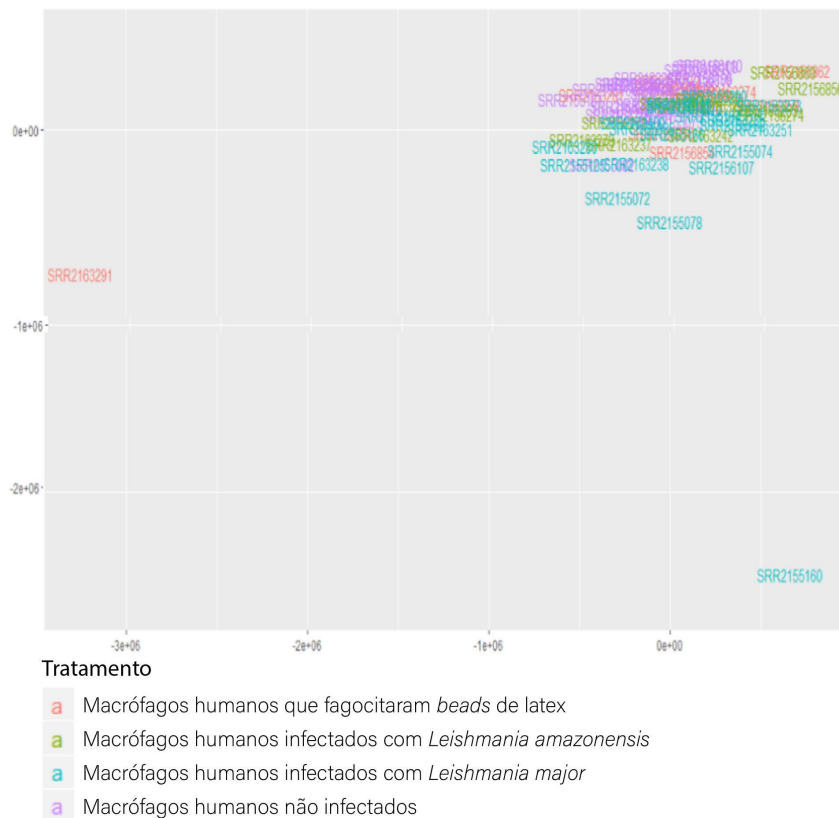


Figura 4.14 - Resultado da análise de PCA para os dados de macrófagos. O componente 1 (PC1) representa a variação máxima entre os pontos de dados (amostras), enquanto o componente principal 2 descreve a segunda maior variação nos dados. Em rosa estão representadas as amostras de macrófagos que fagocitaram *beads* de látex, em verde as amostras de macrófagos infectados com *L. amazonensis*, em azul as amostras de macrófagos infectados com *L. major* e em roxo as amostras de macrófagos não infectados. Cada ponto no gráfico representa uma amostra sequenciada.

Fonte: Dados do trabalho

As amostras SRR2163291 (da condição de macrófago que fagocitaram *beads* de látex) e SRR2155160 (da condição de macrófagos infectados com *L. major* no tempo de 24h pós-infecção) foram identificadas como *outliers* e removidas das análises. Uma nova análise de PCA foi realizada e está demonstrada na figura 4.15.

tempo de 24 horas pós-infecção e nenhum gene DE para os tempos de 48 horas e 72 horas pós-infecção. Por se tratarem de genes cuja expressão foi potencialmente induzida pelo processo de fagocitose e não pela infecção do parasito, eles foram removidos, caso encontrados, da lista de genes diferencialmente expressos provenientes da comparação das amostras de macrófagos não infectados com macrófagos infectados.

Foram identificados como genes (humanos) diferencialmente expressos em macrófagos infectados por *L. major* quando comparados com macrófagos não infectados, um total de 3.082, 1.389, 1.193 e 642 genes para os tempos de infecção de 4 hpi, 24 hpi, 48 hpi e 72 hpi, respectivamente

Já para o conjunto de dados que compreende os macrófagos infectados com *L. amazonensis*, um total de 2.452, 1.332, 301 e 327 genes (humanos) foram identificados como diferencialmente expressos para cada tempo de infecção (4 hpi, 24 hpi, 48 hpi e 72 hpi, respectivamente). Os resultados estão descritos na tabela 4.8.

Tabela 4.8 - Genes diferencialmente expressos em macrófagos.

Tempo de infecção	Macrófagos infectados com <i>L. major</i>		Macrófagos infectados com <i>L. amazonensis</i>	
	Regulados positivamente	Regulados negativamente	Regulados positivamente	Regulados negativamente
4 horas pós-infecção	1.604	1.478	1.407	1.045
24 horas pós-infecção	700	698	652	680
48 horas pós-infecção	689	504	84	217
72 horas pós-infecção	361	281	192	135

Fonte: Dados do trabalho

Os genes mais expressos em macrófagos infectados foram anotados funcionalmente quanto às categorias funcionais do *Gene Ontology*. Para a categoria de processo biológico, as principais ontologias identificadas foram de resposta

celular ao estímulo causado pelo dano ao DNA (GO:0006974), reparo de *DNA* (GO:0006281), processo metabólico de *DNA* (GO:0006259), reparo (GO:0006260), e modificações em lisinas (GO:0018205). Considerando a categoria de função molecular, as principais ontologias identificadas estavam relacionadas à inserção ou deleção de bases de DNA (GO:0032135), ligação ao DNA (GO:0003684), atividade de ATPase DNA dependente (GO:0008094). Já para a categoria de componente celular foram identificadas ontologias relacionadas a cromatina (GO:0000785) e centróssomo (GO:0005813).

Após a análise de enriquecimento funcional, os termos do *Gene Ontology* super-representados no conjunto de dados foram agrupados com base na semântica utilizando o programa *Revigo* (figura 4.16, Apêndice F). Essa etapa foi realizada devido à grande quantidade de termos encontrados. Dessa forma, foi observado que a maioria das funções encontradas estavam relacionadas a alterações no ciclo celular, resposta ao estresse, mecanismos de reparo de DNA, processos de modificações em proteínas e diversificação de receptores imunológicos.

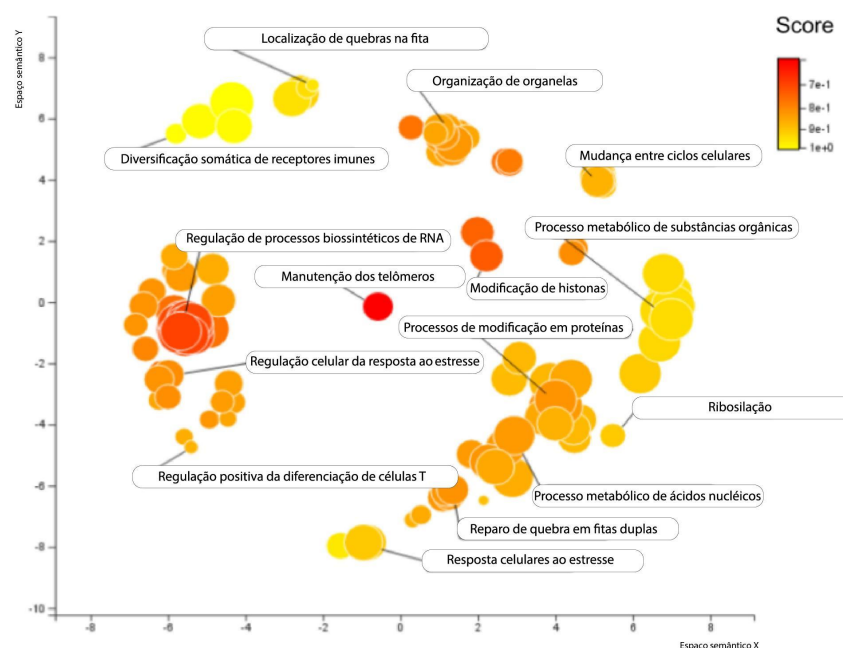


Figura 4.16 - Principais categorias enriquecidas considerando a análise com o *Gene Ontology* para os dados de macrófagos infectados. Termos do *Gene Ontology* super-representados no conjunto de dados foram agrupados com base na semântica utilizando o programa *Revigo*. As cores representam a variação dos valores de score.

Fonte: Dados do trabalho

4.3.2.2 Genes diferencialmente expressos entre as espécies de *Leishmania*

Considerando as duas espécies de *Leishmania* utilizadas para a infecção em macrófagos, além da utilização de amostras onde foram sequenciados parasitos na forma promastigota metacíclica (forma infectiva do parasito), foram realizadas a análise de expressão gênica de cada uma dessas espécies nos diferentes tempos de infecção. Dessa forma foram comparados:

- dados de sequenciamento das formas promastigotas de *L. major* e *L. amazonensis* com os dados de infecção por *L. major* e *L. amazonensis* no tempo de 4 hpi;
- tempos de 4 hpi com 24 hpi;
- tempos de 24 hpi com 48 hpi; e
- tempos de 48 hpi com 72 hpi.

A análise de expressão diferencial com o edgeR mostrou que o maior número de transcritos DE foram encontrados quando foram comparadas amostras de formas promastigotas de *L. major* ou *L. amazonensis* com amostras no tempo de 4hpi (tabela 4.9).

Tabela 4.9 - Genes diferencialmente expressos em *Leishmania*.

Tempo de infecção	<i>L. major</i>		<i>L. amazonensis</i>	
	Regulados positivamente	Regulados negativamente	Regulados positivamente	Regulados negativamente
Promastigota metacíclica - 4hpi	2.631	1.792	2.169	1.327
4hpi - 24hpi	516	327	776	401
24hpi - 48hpi	102	23	45	17
48hpi - 72hpi	12	0	5	0

Fonte: Dados do trabalho

Apesar das informações inicialmente geradas considerarem os dados de *L. major*, devido à falta de incidência dessa espécie no Brasil (Akhoundi et al. 2016; Anversa et al. 2018), a discussão do conjunto de dados diferencialmente expressos estará focada em *L. amazonensis*.

A comparação das amostras possibilitou a identificação de 4.740 transcritos DE em *L. amazonensis*. Dos 4.740 transcritos, 2.169 estão regulados positivamente (*up-regulated*) e 1.327 estão regulados negativamente (*down-regulated*) em *L. amazonensis* no tempo de 4 hpi quando comparado com a forma promastigota.

Dos 4.740 transcritos DE, 2.502 (52,78%) foram anotados como proteínas hipotéticas (sem função predita) e 2.238 (47,32%) apresentaram anotação funcional.

A análise de enriquecimento funcional (teste de Fisher) foi realizada usando como "conjunto de teste" a lista de transcritos regulados positivamente e a lista de transcritos regulados negativamente diferencialmente expressos e como "conjunto de referência" (*background*) o proteoma predito *L. amazonensis*. Um FDR inferior a 0,05 foi estabelecido como o limite para definir a significância do enriquecimento funcional. Somente no tempo de 4 hpi foram identificadas funções enriquecidas.

4.3.2.2.1 Genes regulados positivamente em *L. amazonensis*

Considerando a análise de enriquecimento funcional, aproximadamente 30% do conjunto de dados regulado positivamente em *L. amazonensis*, no tempo de infecção de 4hpi, foi identificado como contendo funções enriquecidas.

Essas funções estão relacionadas a mecanismos de resposta a estímulos, transdução de sinais e processos celulares como por exemplo transporte, processamento de rRNA e glicosilação. Os termos mais representativos, segundo as categorias do *Gene Ontology* foram transporte de íons (GO:0006811), processo metabólico de ácidos nucleicos (GO:0090304), transporte (GO:0006810), localização (GO:0051179), glicosilação de proteínas (GO:0006486), regulação da sinalização (GO:0023051), regulação da resposta a estímulos (GO:0048583), transporte intracelular de proteínas (GO:0006886) (figura 4.17, Apêndice G).

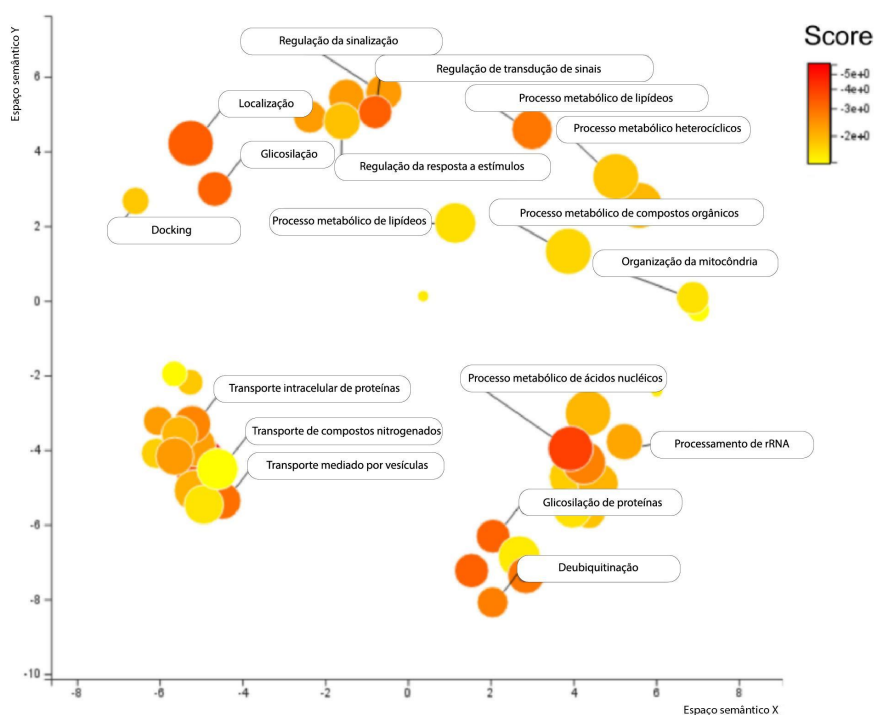


Figura 4.17 - Principais categorias enriquecidas considerando a análise com o Gene Ontology para os transcritos regulados positivamente em *L. amazonensis*. Termos do Gene Ontology super-representados no conjunto de dados foram agrupados com base na semântica utilizando o programa Revigo. As cores representam a variação dos valores de score.

Fonte: Dados do trabalho

Em relação aos transcritos DE, foi observado que as funções previstas desses transcritos foram muito semelhantes aos transcritos encontrados na comparação realizada com os dados de *L. infantum*. Em comum, foram observados transcritos relacionados a transportadores do tipo ABC, amastinas, proteínas kinases e com domínios do tipo *zinc fingers*.

Diferente da comparação realizada em *L. infantum*, foram encontrados como diferencialmente expressos, no tempo de 4 hpi, os transcritos anotados funcionalmente como *glutathione synthetase* e *tryparedoxin peroxidase* com *Fold change* variando de 1,90-3,02 e 2,75-3,27, respectivamente.

Em *Leishmania*, a triparredoxina peroxidase participa na proteção do parasito contra o estresse oxidativo, atuando como um agente antioxidante (Moreira et al. 2013). Corroborando nossos achados, já foi reportado que formas amastigotas de *L. amazonensis* apresentaram uma alta expressão de triparedoxina peroxidase e que isso pode estar associado ao aumento da resistência do parasito às defesas antioxidantes dos macrófagos (Henard et al. 2014). Henard e colaboradores demonstraram que parasitos que superexpressam essa enzima eram mais resistentes às citocinas ativadas por macrófagos quando comparado com as formas promastigotas. Em modelo murino, essa superexpressão contribuiu para o aumento do crescimento parasitário e desenvolvimento de lesões (Henard et al. 2014).

4.3.2.2.1 Genes regulados negativamente em *L. amazonensis*

Considerando a análise de enriquecimento funcional, aproximadamente 30% do conjunto de dados regulado negativamente em *L. amazonensis*, no tempo de infecção de 4hpi, foram identificados como contendo funções enriquecidas.

Essas funções estão relacionadas a processos de sinalização e comunicação celular, metabolismo de ácidos graxos e carboidratos, transdução de sinais, fosforilação de proteínas e processos baseados em microtúbulos (figura 4.18, Apêndice H).

De forma similar ao abordado no item anterior, houve um consenso muito grande em relação às funções preditas desses transcritos. Foi observada a baixa expressão de transcritos que codificam proteínas ribossomais, *RNA binding motifs*, algumas proteínas de transporte e quinases, além de um grande número de transcritos que codificam proteínas anotadas como hipotéticas.

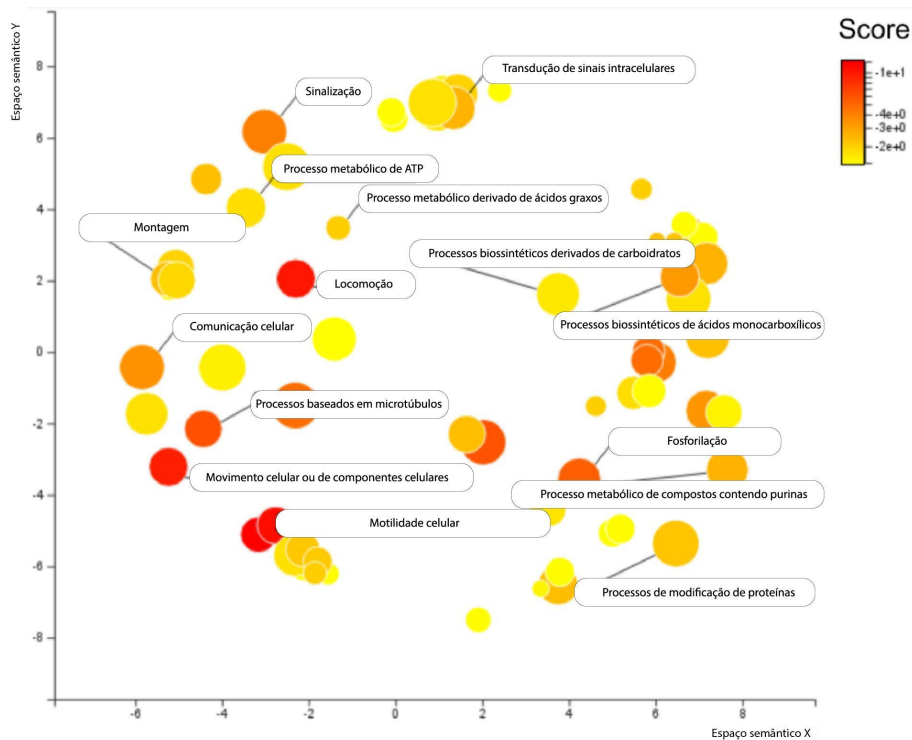


Figura 4.18 - Principais categorias enriquecidas considerando a análise com o Gene Ontology para os transcritos regulados negativamente em *L. amazonensis*. Termos do Gene Ontology super-representados no conjunto de dados foram agrupados com base na semântica utilizando o programa Revigo. As cores representam a variação dos valores de score.

Fonte: Dados do trabalho

Em resumo, as análises de expressão diferencial possibilitaram identificar cerca de 3.500 transcritos DE nos parasitos no tempo de 4 hpi quando comparado com promastigotas metacíclicas. Após o tempo inicial de infecção, o número de transcritos DE reduziu de 3 a 20 vezes nos demais tempos pós infecção. Foi observado um aumento da expressão de processos de glicosilação e deubiquitinação, além de uma redução da expressão de transcritos associados à locomoção e a processos baseados em microtúbulos. A identificação de genes expressos de forma constitutiva entre as diferentes formas é importante, no que diz respeito à identificação de candidatos vacinais, uma vez que fornece evidências quanto à possível expressão das proteínas que eles codificam.

4.4 Construção das redes de regulação gênica de macrófagos infectados

Com o objetivo de detalhar a estratégia analítica implementada no desenvolvimento do presente trabalho, bem como prover uma forma assertiva de correlacionar os objetivos delineados, as estratégias computacionais escolhidas e os resultados alcançados, foi construído um fluxograma detalhando as principais etapas para a construção das redes de regulação gênicas (Figura 4.19).

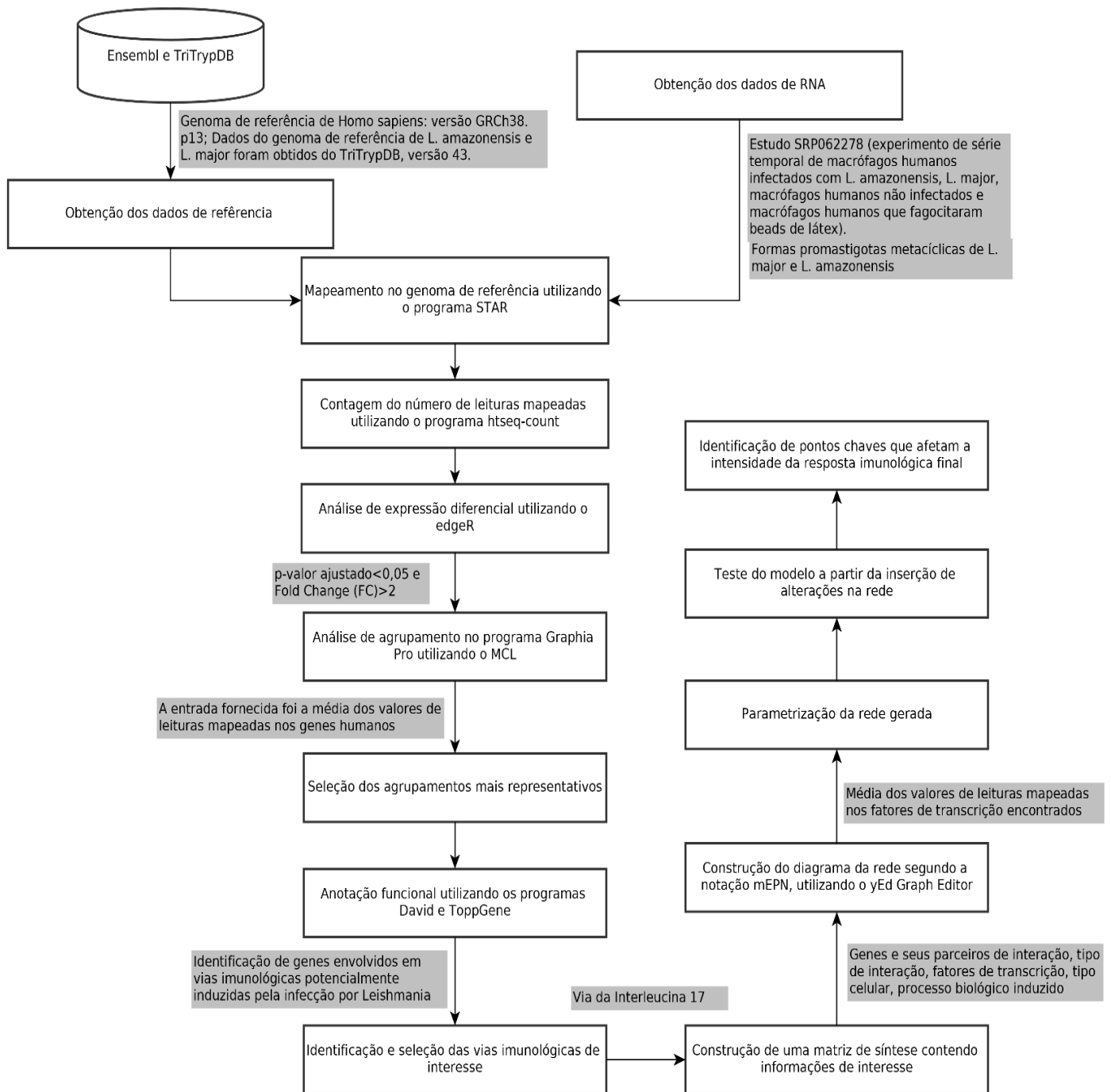


Figura 4.19 - Fluxograma geral das principais etapas empregadas para a construção das redes de regulação gênica com dados de macrófagos humanos infectados com Leishmania. Fonte: Dados do trabalho.

4.4.1 Identificação de agrupamentos gênicos e vias ativadas pela infecção por *Leishmania*

Para a identificação de perfis de expressão relacionados aos macrófagos infectados por *Leishmania* e posterior identificação de vias ativadas por esses parasitos nos macrófagos, uma tabela contendo os valores normalizados (obtidos através da análise com o *Sleuth*) para os genes diferencialmente expressos foi utilizada como entrada para o programa Graphia Pro, disponível no *BioLayout Express*.

Além das informações referentes aos valores normalizados resultantes da contagem das leituras mapeadas para os genes humanos, foram adicionados como campo de descrição o identificador do gene, o identificador do transcrito, o nome do gene e a função predita (todas essas informações foram obtidas a partir do banco de dados do *Ensembl*). Como cabeçalho da tabela, foram adicionadas as informações referentes aos nomes das amostras, códigos dos experimentos, tempo de infecção (4hpi, 24hpi, 48hpi e 72hpi), condição (macrófago infectado e macrófago não infectado), tipo de condição (infectado por *L. infantum*, infectado por *L. major*, fagocitose com *beads* e não infectado).

Um grafo de rede não direcional foi gerado com um valor de coeficiente de correlação de Pearson de $r \geq 0,85$. Como método de normalização, a opção TMM foi a escolhida. A figura 4.20A mostra a variação do tamanho da rede que será gerada de acordo com o coeficiente de correlação estabelecido, quanto maior for esse coeficiente menor será o número de elementos no grafo a ser gerado, ou seja, o grupo conterá somente elementos altamente correlacionados entre si. A variância interna, no grupo, é baixa e a entre grupos é elevada. Já na figura 4.20B, é possível observar a distribuição do grau dos nós na rede.

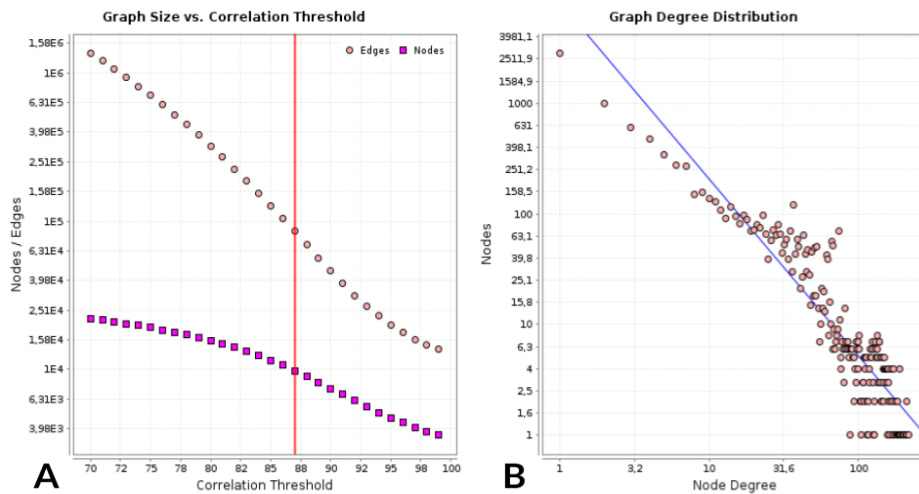


Figura 4.20 - Parâmetros definidos no Graphia Pro para a construção das redes. A figura A, mostra a variação do tamanho da rede que será gerada de acordo com o coeficiente de correlação estabelecido, quanto maior for esse coeficiente menor será o número de elementos no grafo a ser gerado. Já na figura B, é possível observar a distribuição do grau dos nós na rede.

Fonte: Dados do trabalho

Uma matriz de correlação gênica foi calculada e agrupada utilizando o *Markov Clustering Algorithm* (MCL) com um valor de *inflation* de 1,7 (que define como será a tamanho dos agrupamentos gerados). O resultado desse agrupamento pode ser observado na figura 4.21, onde pode ser observada a existência de agrupamentos maiores, com um grande número de genes que compartilham perfis de expressão semelhantes (Figura 4.21A) e alguns agrupamentos menores com um número reduzido de genes (Figura 4.21B).

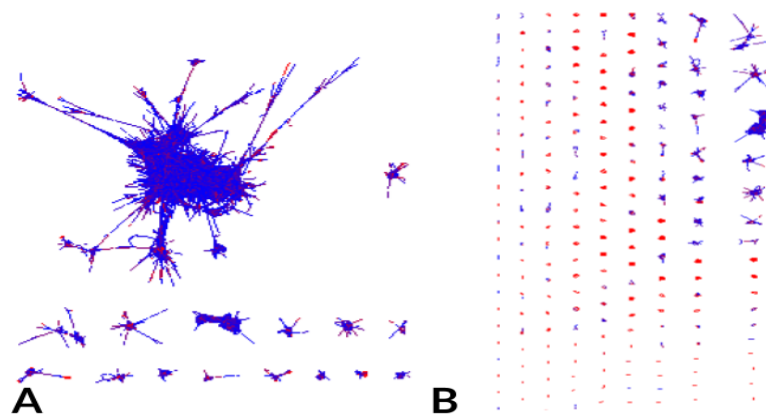


Figura 4.21 - Resultado do agrupamento realizado pelo MCL. A existência de agrupamentos maiores pode ser observada, com um grande número de genes que

compartilham perfis de expressão semelhantes (Figura A) e alguns agrupamentos menores com um número reduzido de genes (Figura B).

Fonte: Dados do trabalho

Utilizando a função “*Class Viewer*” disponível no Graphia Pro, os agrupamentos resultantes do MCL foram visualizados e analisados quanto à sua similaridade em relação à expressão dos genes em cada uma das redes.

Na figura 4.22, é possível observar que os dados estão agrupados segundo o esquema de cores que representam os termos descritivos das amostras que foram fornecidos como entrada para o programa. Dessa forma, no eixo das abscissas estão representados: na primeira linha, os tempos de infecção (amarelo para o tempo de 4 hpi, rosa escuro para o tempo de 24 hpi, azul para o tempo de 48 hpi e rosa claro para o tempo de 72 hpi); na segunda linha a condição, sendo verde claro para as amostras não infectadas e verde escuro para as amostras infectadas, entre as amostras infectadas; na terceira linha foi separado por espécie que causou a infecção, sendo azul escuro para as amostras de macrófagos que foram infectados com *L. amazonensis* e lilás para as amostras de macrófagos que foram infectados com *L. major*. Na última linha, os dados foram agrupados pelo tipo de condição, sendo macrófagos que fagocitaram *beads* de látex representados pela cor rosa, macrófagos não infectados representados pela cor cinza, macrófagos infectados por *L. major* representados na cor amarela e macrófagos infectados por *L. amazonensis* na cor rosa.

É possível observar que, no tempo de infecção de 4h, as amostras de macrófagos infectados apresentaram uma expressão elevada de um determinado conjunto de genes quando comparado com as amostras controle. Esses genes foram selecionados e utilizados para as etapas posteriores, de anotação funcional, identificação de possíveis vias alteradas pela infecção causada pelo parasito e construção de uma rede de regulação gênica para uma dessas vias.

A partir de todos os agrupamentos gerados, foram selecionados um total de 30 em que os genes apresentaram um perfil de expressão mais significativo, ou seja, que apresentaram maior variação de expressão entre os grupos de comparação.

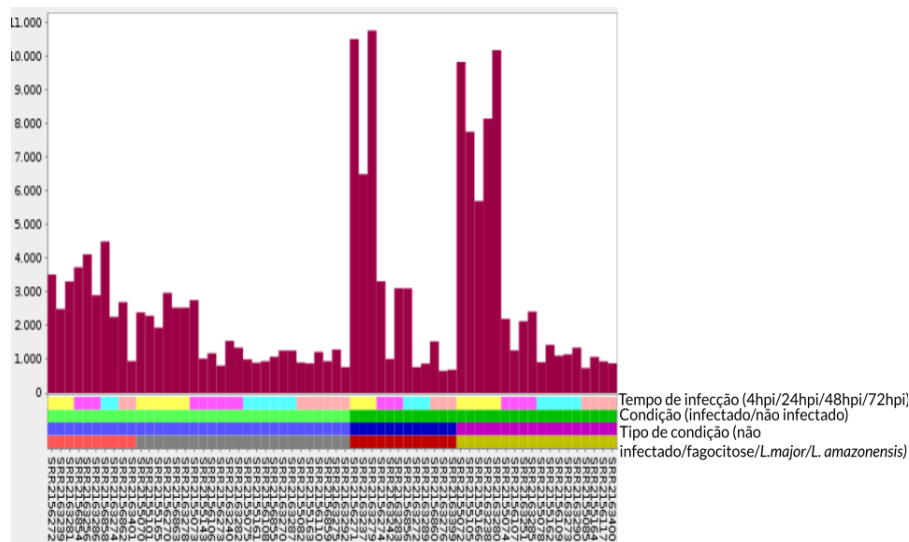


Figura 4.22 - Agrupamento anotado como contendo diferenças de expressão entre as amostras infectadas e amostras controle.

No eixo das abscissas estão representados: na primeira linha, os tempos de infecção (amarelo para o tempo de 4 hpi, rosa escuro para o tempo de 24 hpi, azul para o tempo de 48 hpi e rosa claro para o tempo de 72 hpi); na segunda linha a condição, sendo verde claro para as amostras não infectadas e verde escuro para as amostras infectadas, entre as amostras infectadas; na terceira linha foi separado por espécie que causou a infecção, sendo azul escuro para as amostras de macrófagos que foram infectados com *L. amazonensis* e lilás para as amostras de macrófagos que foram infectados com *L. major*. Na última linha, os dados foram agrupados pelo tipo de condição, sendo macrófagos que fagocitaram beads de látex representados pela cor rosa, macrófagos não infectados representados pela cor cinza, macrófagos infectados por *L. major* representados na cor amarela e macrófagos infectados por *L. amazonensis* representados na cor rosa.

Fonte: Dados do trabalho.

Foram encontrados um total de 26 agrupamentos gênicos com perfis de expressão diferentes entre os macrófagos infectados pelas diferentes espécies de *Leishmania*, sendo 17 deles correspondentes aos macrófagos infectados por *L. major* e nove referentes a macrófagos infectados por *L. amazonensis*. A figura 4.23, exemplifica um desses resultados, onde é possível observar que esse conjunto de genes foi expresso somente nas amostras referentes a macrófagos infectados por *L. major*.

A nossa hipótese é de que esses agrupamentos possam estar relacionados aos diferentes desfechos clínicos causados por essas duas espécies.

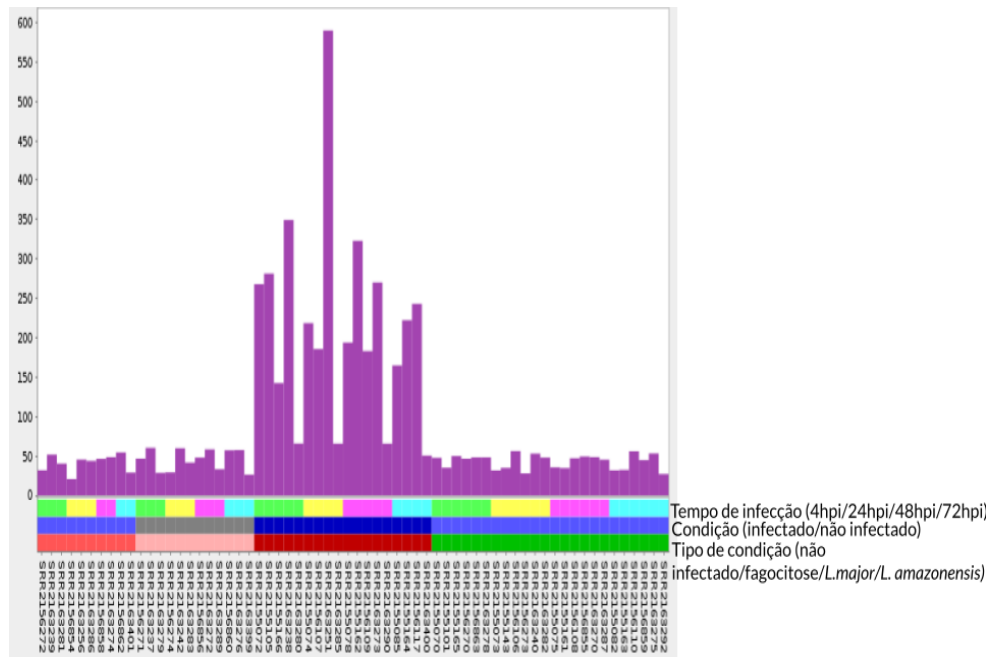


Figura 4.23 - Agrupamento anotado como contendo diferenças de expressão relacionadas a macrófagos infectados com *L. major*. No eixo das abscissas estão representados: na primeira linha, os tempos de infecção (verde para o tempo de 4 hpi, amarelo para o tempo de 24 hpi, rosa para o tempo de 48 hpi e azul para o tempo de 72 hpi); na terceira linha e última linha, os dados foram agrupados pelo tipo de condição, sendo macrófagos que fagocitaram *beads* de látex representados pela cor rosa escuro, macrófagos não infectados pela cor rosa claro, macrófagos infectados por *L. major* representados na cor vermelha e macrófagos infectados por *L. amazonensis* representados na cor verde.

Fonte: Dados do trabalho.

4.4.2 Anotação funcional dos agrupamentos encontrados e busca na literatura para modelagem das redes de regulação gênica

Os 30 agrupamentos encontrados com expressão gênica semelhante entre os dados de macrófagos infectados pelas diferentes espécies de *Leishmania* foram anotados funcionalmente utilizando o *David* e o *ToppGene*.

Essa anotação funcional possibilitou a identificação de genes envolvidos em vias imunológicas potencialmente induzidas pela infecção por *Leishmania*, como a via de sinalização de TNF (*Tumor Necrosis Factor*), via de sinalização de *NF-kappa B*, via de interação de receptores do tipo *Cytokine-cytokine*, via de sinalização de MAPK, via de sinalização de receptores do tipo *Toll*, via de sinalização de Interleucina 10 (IL-10) e via de sinalização de IL-17.

Dentre essas vias encontradas, foi realizada uma busca na literatura com o intuito de se identificar as vias imunológicas mais relevantes frente à infecção por *Leishmania*.

As infecções por *Leishmania* spp. podem promover, no organismo hospedeiro, diferentes tipos de respostas imunológicas. Dentre eles, é possível observar o aumento de células T CD4 + que pode ocasionar uma resposta Th1 com a produção de interferon- γ , IL-2 e IL-12, que irão estimular macrófagos a combaterem células infectadas (Isnard et al. 2012; Mutiso et al. 2013). Porém, se a resposta imune for do tipo Th2, poderá ocorrer a produção de IL-4 e IL-10, que atuam na inibição da ativação de macrófagos, consequentemente promovendo o agravamento da doença (Isnard et al. 2012; Liu and Uzonna 2012; Mutiso et al. 2013; Tibúrcio et al. 2019).

A via de sinalização da IL-17 está envolvida em muitas doenças infecciosas humanas e funções biológicas conflitantes da IL-17 na patologia das leishmanioses foram relatados na literatura (Bhattacharya and Ali 2013). Não está claro se as citocinas ativadas por essa via participam como um mecanismo de defesa ou no agravamento da patologia dessas doenças (Pitta et al. 2009; Hu et al. 2010; Gonzalez-Lombana et al. 2013; Gonçalves-de-Albuquerque et al. 2017; Banerjee et al. 2018).

As células Th17 são moduladores cruciais da imunidade adaptativa contra os parasitos do gênero *Leishmania*, atuando principalmente no recrutamento de neutrófilos (Gonzalez-Lombana et al. 2013).

Dessa forma, a via de sinalização de IL-17 foi a escolhida devido ao fato de estudos anteriores demonstrarem que esses elevados níveis de IL-17 podem ter um importante papel pró-inflamatório no desenvolvimento da doença e sugerem que essa via pode ser um alvo promissor para o desenvolvimento de estratégias imunoterápicas (Vargas-Inchaustegui et al. 2008; Boaventura et al. 2010; Castilho et al. 2010; Katara et al. 2012, 2013; Gonzalez-Lombana et al. 2013; Banerjee et al. 2018) .

Considerando a via de IL-17, foi realizada uma busca na literatura utilizando os termos descritos na tabela 4.10, nas páginas do *Pubmed* e Periódicos

Capes/MEC, entre os meses de fevereiro e março de 2020. Um total de 327 artigos foram encontrados a partir da busca realizada. As informações desses artigos foram extraídas e uma matriz de síntese foi construída com base nessas interações. Todas as informações foram posteriormente analisadas e categorizadas de modo a serem utilizadas como entrada para a construção das redes de regulação gênica.

Tabela 4.10 - Termos usados para busca no Pubmed e no Periódicos Capes/MEC.

Termos	Total de artigos encontrados
<i>IL-17 OR Interleukin 17 signaling AND transcriptional factors</i>	42
<i>IL-17 OR Interleukin 17 AND expression AND macrophage stimulation</i>	102
<i>IL-17 OR Interleukin 17 AND macrophages AND expression</i>	75
<i>IL-17 OR Interleukin 17 AND macrophages AND regulation</i>	89
<i>IL-17 OR Interleukin 17 AND Leishmania OR trypanosomatids AND regulation AND macrophages</i>	19
Total de artigos	327

4.4.3 Construção das redes de regulação gênica

Esses 327 artigos foram lidos, e, quando disponíveis, as informações referentes aos genes de interesse obtidos através da análise de expressão diferencial foram selecionadas. Essas informações foram categorizadas em :

- Parceiro 1 (gene que atua regulando de alguma forma um segundo gene);
- Parceiro 2 (gene que sofre algum tipo de indução/inibição pelo parceiro 1);

- Fatores de transcrição que interagem com os parceiros 1 e/ou 2;
- Tipo celular em que essa interação foi observada;
- Tipo de interação;
- Processo biológico induzido;
- Desfecho da interação entre os parceiros 1 e 2 (ativação de outros genes, indução de citocinas ou quimiocinas).

Essas informações foram categorizadas como componentes da rede, recebendo a seguinte nomenclatura:

- *input node*: fatores de transcrição ativados a partir do estímulo recebido e que interagem com os genes encontrados nos agrupamentos selecionados (infecção do parasito);
- *transmission node*: genes que são ativados durante o curso da infecção;
- *output node*: genes que são ativados para uma determinada resposta imunológica;
- *mode of interaction*: forma como esses genes interagem, se é por ativação, ligação, inibição, fosforilação, entre outros.

O diagrama da rede (figura 4.24) foi construído utilizando o yEd Graph Editor segundo a notação mEPN, descrita por Livigni e colaboradores (2018), utilizando as informações obtidas através da busca realizada na literatura (Livigni et al. 2016)

Os componentes (genes) na GNR da IL-17 são representados por retângulos conectados por setas que indicam as interações moleculares encontradas durante a etapa de revisão da literatura. Os nós de início do sistema (fatores de transcrição) estão representados por retângulos coloridos em preto. O sinal de saída da GRN é representado no retângulo no lado direito do diagrama.

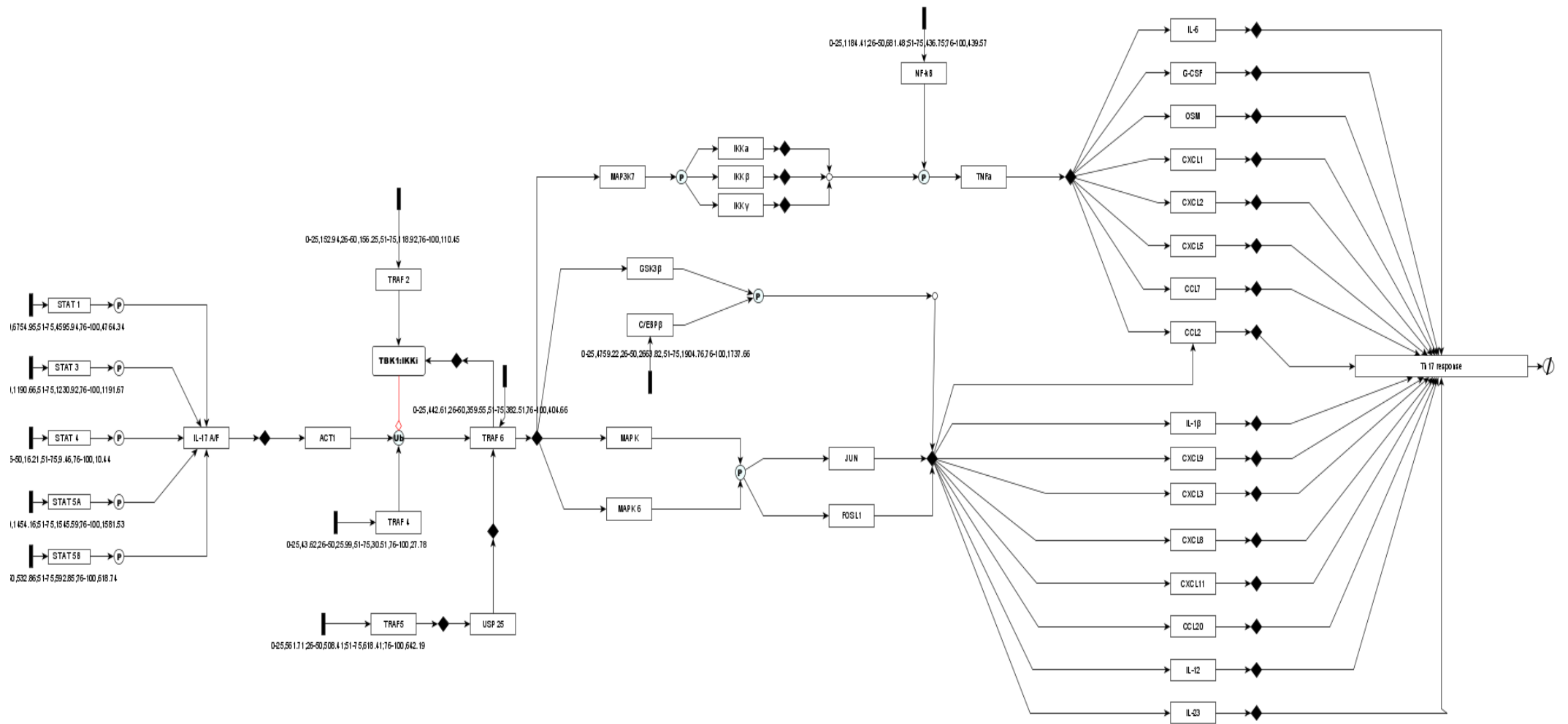


Figura 4.24 - Diagrama da rede construída. O diagrama da rede foi construído utilizando o yEd Graph Editor segundo a notação mEPN, descrita por Livigni e colaboradores (2018), utilizando as informações obtidas através da busca realizada na literatura. Os componentes (genes) na GNR da IL-17 são representados por retângulos conectados por setas que indicam as interações moleculares encontradas durante a etapa de revisão da literatura. Os nós de início do sistema (fatores de transcrição) estão representados por retângulos coloridos em preto. O sinal de saída da GRN é representado no retângulo no lado direito do diagrama. **Fonte:** Dados do trabalho

As arestas que conectam os nós e as transições determinam a direção do fluxo de informação, representando o progresso do processo biológico ao longo do tempo (denominado pelo programa como *time-blocks*). Cada tempo de infecção no desenho experimental representa um *time-block* que foi utilizado para a simulação. Dessa forma os *time-blocks* foram definidos como:

- 4 horas pós-infecção (*time-block* 0-25);
- 24 horas pós-infecção (*time-block* 26-50);
- 48 horas pós-infecção (*time-block* 51-75); e
- 72 horas pós-infecção (*time-block* 76-100).

A média das contagens das leituras mapeadas para cada um dos fatores de transcrição nos diferentes tempos de infecção foram utilizados como valores de parametrização para cada nó de entrada na rede (Apêndice I). As simulações foram executadas no Graphia Pro usando como parâmetros 100 *time-blocks*, 500 *runs*, filtro de *Uniform distribution in Stochastic Petri Nets* (SPN) e *Distribution and Consumptive Transition in SPN Transition Type* ativados.

O resultado dessa simulação é um gráfico (figura 4.25) demonstrando a intensidade do sinal de saída da GRN, representado no retângulo no lado direito do diagrama da rede (figura 4.24). Essa intensidade de sinal é gerada a partir dos valores de entrada, correspondentes aos fatores de transcrição. É possível observar que, em condições normais, como pode ser observado na figura 4.25A (macrófagos não infectados), essa intensidade do sinal é cerca de 2-3 vezes menor do que quando comparado com as amostras de macrófagos infectados por *L. major* e *L. amazonensis* (figura 4.25B e 4.25C respectivamente).

Para ser gerado, o fluxo de informação inicial começa na ativação dos fatores de transcrição (*input node*). Esses fatores irão interagir com uma série de genes (*transmission node*) através de processos (*mode of interaction*) como, por exemplo, fosforilação, ubiquitinação e oxidação. Esses processos serão responsáveis por intensificar ou reduzir a intensidade do sinal gerado como resposta.

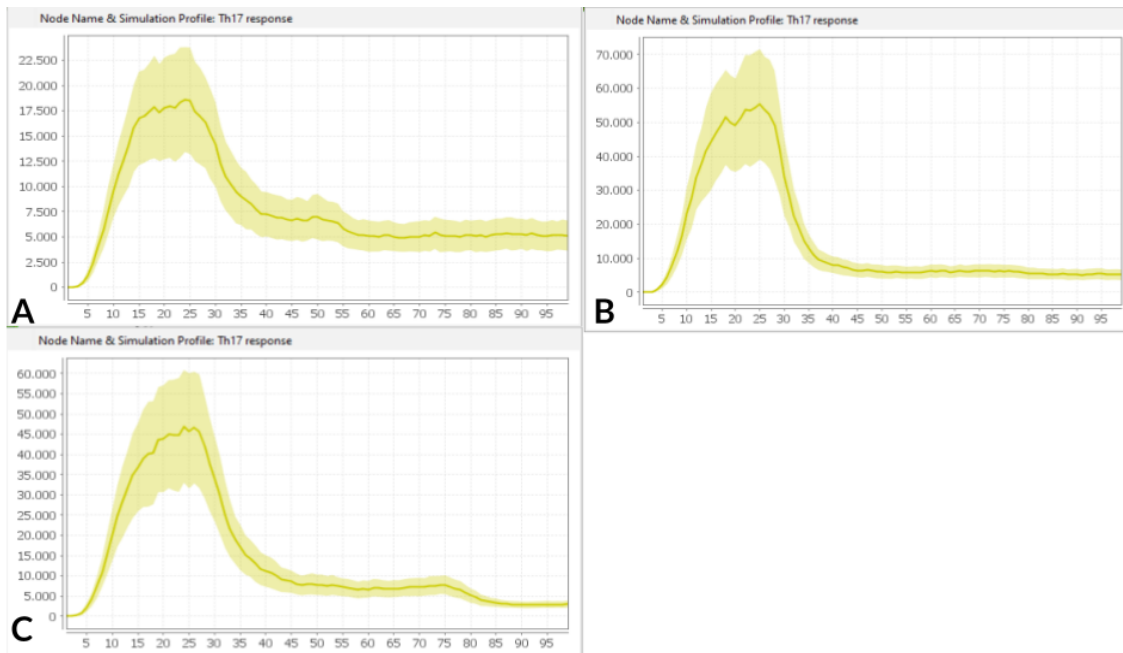


Figura 4.25 - Resultado da simulação inicial da rede. Em condições normais, como pode ser observado em A (macrófagos não infectados), a intensidade do sinal de resposta final é cerca de 2-3 vezes menor do que quando comparado com as amostras de macrófagos infectados por *L. major* e *L. amazonensis* (representadas em B e em C, respectivamente). Para ser gerado, o fluxo de informação inicial começa na ativação dos fatores de transcrição (*input node*). Esses fatores irão interagir com uma série de genes (*transmission node*) através de processos (*mode of interaction*) como, por exemplo, fosforilação, ubiquitinação e oxidação. Esses processos serão responsáveis por intensificar ou reduzir a intensidade do sinal gerado como resposta.

Fonte: Dados do trabalho

Após a parametrização da rede, pontos de inibição nas arestas entre determinadas interações foram inseridos para se observar como a intensidade do sinal se alterava a partir dessas modificações no modelo (Figura 4.26).

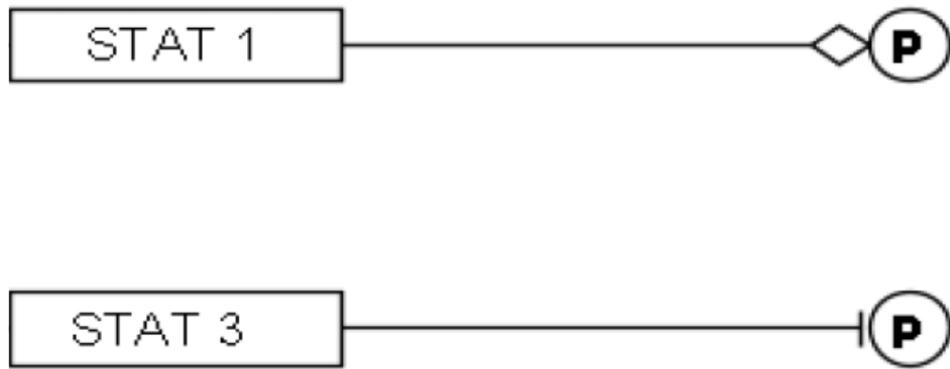


Figura 4.26 - Exemplos de alterações inseridas na rede. Exemplos da representação de alterações inibitórias e/ou bloqueadoras que foram inseridas na rede modelada

Fonte: Dados do trabalho

A Figura 4.26 exemplifica duas modificações realizadas, onde, para o fator de transcrição STAT1, foi inserida uma alteração na aresta, o que reduz a intensidade em que a fosforilação será realizada no gene alvo para a sua ativação, enquanto no fator STAT3 foi inserida uma aresta que inibe completamente o sinal gerado por esse FT, ou seja, ele não irá fosforilar o gene alvo.

Além dessas duas modificações nos fatores de transcrição STAT1 e STAT3, modificações inibitórias parciais também foram inseridas nos fatores de transcrição *CEBP β* (*CEBPB* - *CCAAT/enhancer-binding protein beta*) e *NF- κ β* (*nuclear factor kappa-light-chain-enhancer of activated B cells*). Essas modificações foram as que mais causaram diferenças na intensidade do sinal gerado como resposta (Figura 4.27), ainda assim não sendo suficientes para um bloqueio completo da via nos macrófagos infectados (Figura 4.27B e 4.27C) quando comparado com os macrófagos não infectados (Figura 4.27A).

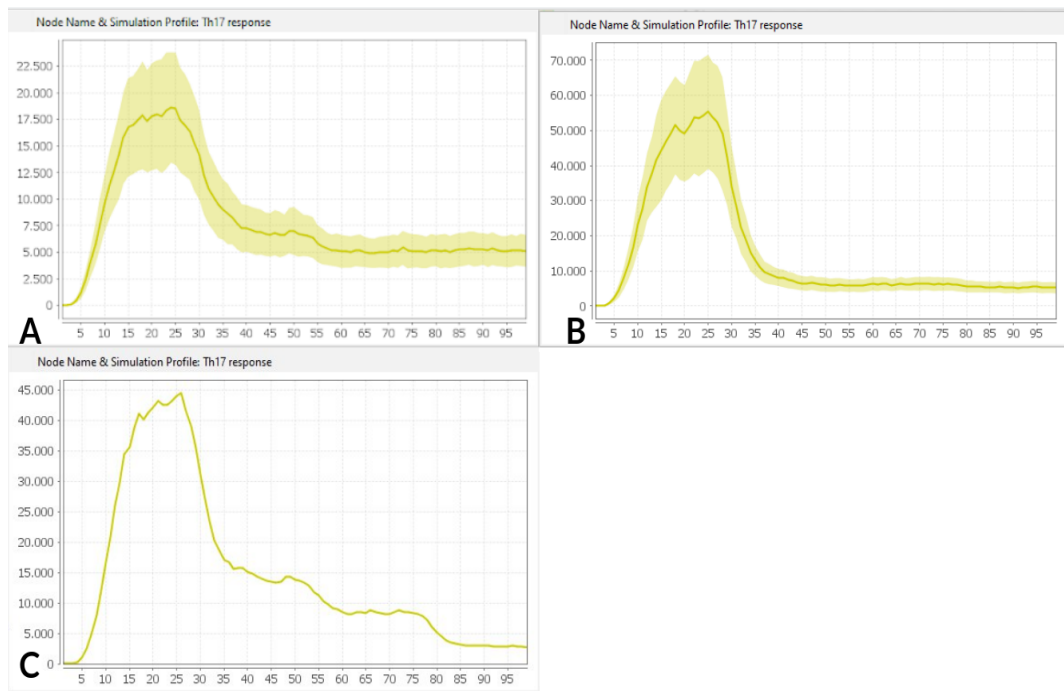


Figura 4.27 - Resultado da simulação da rede com as alterações nos fatores de transcrição. Em condições normais, como pode ser observado em A (macrófagos não infectados), a intensidade do sinal de resposta final é menor do que quando comparado com as amostras de macrófagos infectados por *L. major* e *L. amazonensis* representadas em B e em C, respectivamente.

Fonte: Dados do trabalho

Hipotetizamos que a dificuldade em se bloquear completamente essa via pode estar relacionada à ausência de algum gene alvo na reconstrução do modelo ou devido à interação dos genes da via da IL-17 com outras vias de sinalização.

No entanto, conforme demonstrado, a redução da intensidade do sinal na resposta Th17 via inibição de pontos na rede pode ser um alvo para o tratamento das leishmanioses. Por essas alterações ocorrerem principalmente no início da infecção pelos parasitos, os genes que esses parasitos expressam no momento da infecção podem ser considerados importantes candidatos para o desenvolvimento de uma vacina contra *Leishmania* spp..

Nenhum outro trabalho, considerando macrófagos infectados por *Leishmania*, foi reportado utilizando essa estratégia de GRN. Até o presente momento apenas um estudo foi reportado empregando essa estratégia de GRN em *Leishmania major*, no entanto, diferente deste trabalho, o foco foi a reconstrução do metabolismo de

lipídios desses parasitos, com o objetivo de identificar possíveis alvos para fármacos (Patel et al. 2008).

Em humanos essa abordagem já foi empregada com sucesso na reconstrução da GRN responsável pelo desenvolvimento cardíaco, onde foi observado que mutações em componentes dessa rede estavam relacionados ao desenvolvimento de doenças cardíacas congênitas e que a identificação desses alvos possibilitou o desenvolvimento de testes genéticos para a identificação precoce dessas alterações (Olson 2006; Munshi Nikhil V. 2012).

Já no estudo de Lopes-Ramos (2020), a construção da GRN possibilitou a identificação de diferenças na expressão gênica entre homens e mulheres, em 29 tecidos humanos, que podem guiar o desenvolvimento de estratégias terapêuticas “sexo específicas”. Essas GRN foram reconstruídas a partir de 8.279 perfis de expressão disponíveis no projeto *Genotype-Tissue Expression* (GTEx). Foi observado que apesar dos FT expressos serem os mesmos, diversos deles apresentaram padrões de expressão diferentes, incluindo alvos diferentes. Outro ponto muito interessante reportado por esse trabalho é que, considerando o tecido cerebral, os genes associados a doenças neurodegenerativas, como, por exemplo, o Parkinson e o Alzheimer, são direcionados por diferentes conjuntos de FT em cada sexo (Lopes-Ramos et al. 2020).

Em resumo, a via de TNF, via de sinalização de *NF-kappa B*, via de interação de receptores do tipo *Cytokine-cytokine*, via de sinalização de MAPK, via de sinalização de receptores do tipo Toll, via de sinalização de IL-10 e de IL-17 foram as principais vias imunológicas identificadas como alteradas em macrófagos infectados com *Leishmania*. A identificação dessas vias irá permitir uma melhor caracterização de como ocorre a modulação do sistema imune por parte do parasito, além da identificação de pontos nessas redes que, quando bloqueados, podem reduzir a resposta imunológica gerada.

4.5 Integração de dados para a seleção das proteínas contendo epítomos imunogênicos

A predição de epítomos foi realizada conforme descrito na seção 5.7 Foram preditos um total de 124.900.639 epítomos para MHC de classe I e 158.951.774 epítomos para MHC de classe II, considerando todas as espécies utilizadas neste trabalho.

Considerando as quatro espécies de *Leishmania* para as quais foram realizadas as predições, inicialmente foram identificadas 1.727 proteínas que apresentavam epítomos de LB com tamanho mínimo de 10 e máximo de 20 resíduos de aminoácidos. Considerando a avaliação de estudos que indicam que quanto maior o número de alelos de MHC de classe I e de classe II com epítomos preditos em uma proteína, maior será a sua imunogenicidade, foram selecionadas proteínas que apresentaram epítomos para no mínimo 40 alelos humanos, 3 alelos de camundongo e 12 supertipos de alelos (Brito et al. 2017). Os alelos de camundongos foram considerados para possíveis validações experimentais utilizando modelos murinos. Por fim, foram selecionadas proteínas com localização subcelular de membrana plasmática ou extracelular.

A localização subcelular é uma característica essencial para a caracterização funcional de uma proteína (Cai et al. 2010), sendo assim, essa etapa é importante na identificação de epítomos imunogênicos, uma vez que é necessária a interação direta desses trechos com células protetoras do sistema imune, como linfócitos T e linfócitos B, para ativação do sistema imunológico.

Posteriormente, essas proteínas foram comparadas com o resultado do *OrthoMCL* com o objetivo de excluir proteínas que apresentassem similaridade com sequências proteicas de humano, camundongo e cachorro. Ao final dessas análises foram selecionadas 1.119 proteínas.

Para as espécies *L. infantum* e *L. amazonensis*, selecionamos apenas as proteínas cujos transcritos apresentavam expressão (não necessariamente sendo diferencialmente expressos em uma das condições) nos dados de RNAseq. Esse critério foi utilizado pois dessa forma temos uma evidência de que possivelmente a proteína que continha o epítomo predito também seria expressa, o que foi

comprovado, em alguns casos, pela literatura, em trabalhos que utilizaram estratégias de proteômica.

Em seguida foi realizado um refinamento no processo de seleção de epítomos, avaliando-se a topologia completa da proteína, ou seja, a avaliação de todas as hélices transmembranas e suas orientações (interno/externo) em relação à membrana. Dessa forma, foram considerados os epítomos que foram preditos dentro de uma região predita como *outside* pelo *TMHMM*, ou seja, externa à membrana plasmática. Ao final desse processo, foram selecionadas 113 proteínas e 124 epítomos distintos.

Considerando a análise de imunogenicidade realizada pelo programa *VaxiJen*, 102 epítomos, dos 124 selecionados anteriormente, foram caracterizados como candidatos antigênicos, sendo, portanto, candidatos promissores ao desenvolvimento de um vacina contra as leishmanioses (Apêndice H).

Foram priorizados na seleção descrita anteriormente, para estudos futuros envolvendo validação experimental, um total de 20 epítomos (Tabela 4.11). Esses epítomos foram priorizados por apresentarem predição para alelos mais frequentes na população brasileira, segundo o banco de dados *Allele Frequency Net Database*, maior escore de imunogenicidade segundo o *VaxiJen* e maior porcentagem de conservação de sequência segundo análise realizada através do *Epitope Conservancy Analysis*, disponível no banco de dados do IEDB. Das sequências obtidas através da identificação de novos transcritos, duas delas apresentaram epítomos com potencial imunogênico e estão incluídas na análise.

Tabela 4.11 - Epítomos priorizados.

Identificador da proteína	Função predita	Sequência do epítopo	Score do VaxiJen
LAMA_000598900	<i>hypothetical protein, conserved</i>	AHLRRYLAATAIASL*	0.6002
Trinity_LINF_09.8745	<i>hypothetical protein - conserved</i>	AHLRRYLAATAIASL*	0.6002
LmxM.29.1260	<i>hypothetical protein, conserved</i>	AHLRRYLAATAIASL*	0.6002
LAMA_000017700	<i>hypothetical protein, conserved</i>	FVFPASFTL*	0.6046
LINF_360060000	<i>32 kDa ER-associated protein - putative</i>	FVFPASFTL*	0.6046
LmxM.36.5050	<i>hypothetical protein, conserved</i>	FVFPASFTL*	0.6046
LAMA_000175700	<i>Amino acid permease, putative</i>	FHRCMYLSMATMSLI*	0.6166
LmxM.11.0520	<i>amino acid permease/transporter, putative</i>	FHRCMYLSMATMSLI*	0.6166
LbrM.14.0490	<i>hypothetical protein, unknown function</i>	SSSFALSMAAFVLHA*	0.8436
LbrM.14.0500	<i>hypothetical protein, unknown function</i>	SSSFALSMAAFVLHA*	0.8436
LAMA_000281600	<i>hypothetical protein, conserved</i>	YVMELFFTL*	0.9154
LbrM.18.1750	<i>hypothetical protein, conserved</i>	YVMELFFTL*	0.9154
LmxM.18.0840	<i>hypothetical protein, conserved</i>	YVMELFFTL*	0.9154
LAMA_000808000	<i>hypothetical protein, conserved</i>	FTYAAFTAL*	1.0168
LmxM.34.4140	<i>hypothetical protein, unknown function</i>	FTYAAFTAL*	1.0168
LAMA_000362500	<i>Nuclease-related domain containing protein, putative</i>	FVAAYWHLL	1.1801
LmxM.36.0650	<i>GPI transamidase component Tta2, putative</i>	ELALWYVRFIPAVLA*	1.2915
LINF_360012100	<i>GPI transamidase component Tta2 - putative</i>	ALWYVRFIPAVLAKP*	1.3118
LAMA_000045000	<i>hypothetical protein, conserved</i>	YMFVASYYM*	1.3945
LmxM.03.0490	<i>hypothetical protein, conserved</i>	YMFVASYYM*	1.3945
Trinity_LINF_29.2679	<i>hypothetical protein - conserved</i>	YVMKASVTL	1.4765
LINF_320045500	<i>hypothetical protein - conserved</i>	YMFLFVYVL	1.5994
LINF_230014100	<i>hypothetical protein - conserved</i>	YVYKIWLEL	1.6186

LAMA_000443100	<i>hypothetical protein, conserved</i>	YVYKIWVEL	1.629
LAMA_000366500	<i>hypothetical protein, conserved</i>	YLWYMFMPI*	2.1203
LmxM.36.5970	<i>Flagellum attachment zone protein 5</i>	YLWYMFMPI*	2.1203

*Epítópos comuns a mais de uma espécie de *Leishmania*.

Fonte: Dados do trabalho.

Foi observada uma grande similaridade entre as sequências de diversos epítópos preditos, sendo eles preditos em duas ou mais proteínas homólogas entre as espécies de *Leishmania*. Esses epítópos são de grande interesse, uma vez que podem compor uma vacina que tem o potencial de oferecer proteção contra a infecção por mais de uma espécie de *Leishmania*.

Em relação à anotação funcional dessas proteínas, foi observada a ocorrência de funções relacionadas a nucleases, transaminases, permeases, *flagellum attachment*, *emp24/gp25L/p24/GOLD* e *ER-associated*. As sequências referentes aos transcritos novos identificados neste trabalho foram preditas como proteínas hipotéticas, ou seja, sem função conhecida.

Em *Leishmania*, as nucleases desempenham um papel vital no fornecimento de purinas que serão utilizadas para o crescimento e desenvolvimento dos parasitos, uma vez que eles são auxotróficos para purinas, o que significa que esses parasitos são incapazes de produzir purinas *de novo* (Farajnia et al. 2011; Guimarães-Costa et al. 2014). Segundo Guimarães-Costa (2014), essa enzima foi associada pela primeira vez com a nutrição do parasito devido à sua atividade poder gerar nucleotídeos e fosfato a partir de ácidos nucleicos. Uma vez que não há uma enzima homóloga com propriedades semelhantes em mamíferos, esta enzima apresenta grande potencial para ser utilizada como candidato vacinal (Guimarães-Costa et al. 2014).

Por sua vez, a família proteica com domínios *emp24/gp25L/p24/GOLD* está envolvida no transporte e exportação de proteínas, muitas delas consideradas fatores de virulência (Grace et al. 2021). Essas proteínas foram encontradas como sendo expressas tanto em formas promastigotas quanto amastigotas e o domínio GOLD que elas apresentam pode estar associado com a mediação de diversas

interações proteína-proteína, incluindo interações com o hospedeiro (El-Sayed et al. 2005; Fernandes et al. 2016; Grace et al. 2021).

As permeases são proteínas de membrana envolvidas no transporte de aminoácidos para a célula. Alguns estudos sugerem que elas podem estar envolvidas nos processos de detoxificação de drogas derivadas de compostos antimoniais (Akerman et al. 2004; Kazemi-Rad et al. 2013a, 2013b).

As proteínas anotadas como hipotéticas, ou seja, sem função predita, representaram 75,67% (28/37 proteínas) do conjunto de dados priorizado para uma validação experimental futura.

Estudos anteriores já demonstraram que apesar da ausência de função predita, proteínas anotadas como hipotéticas desempenham papéis importantes no organismo, como, por exemplo, no desenvolvimento de resistência ao arsênio e em funções celulares, incluindo adaptação ao organismo hospedeiro e habilidade de evasão do sistema imune (Mazandu and Mulder 2012; Costa et al. 2018).

Apesar da ausência de função predita, essas proteínas representam um conjunto de dados de extrema importância e que deve ser analisado mais profundamente. Em *Leishmania*, já foi reportada a capacidade dessas proteínas induzirem resposta imune. Um exemplo é a proteína hipotética LiHyP, que foi reportada como antigênica em uma trabalho de imunoproteômica em LV canina (Ribeiro et al. 2018).

Já a proteína LiHyS foi avaliada quanto à sua antigenicidade, imunogenicidade e eficácia protetora contra LV. Em relação à antigenicidade, foram observados elevados valores de sensibilidade e especificidade para a proteína recombinante (rLiHyS) no diagnóstico da LV. Ao avaliar a imunogenicidade da LiHyS, os ensaios proliferativos com células mononucleares do sangue periférico de indivíduos saudáveis ou pacientes com LV mostraram alto índice proliferativo em ambas as amostras. Posteriormente, em camundongos BALB/c que foram desafiados com formas promastigotas de *L. infantum*, a vacina composta por rLiHyS induziu a produção de *interferon-γ* e IL-12 e foi observado que ocorreu uma redução significativa da carga parasitária no fígado, baço, medula óssea e linfonodos (Dias et al. 2017, 2018; Ribeiro et al. 2018).

Já foi reportado em diversos trabalhos que, em *Leishmania*, a porcentagem de genes que codificam proteínas hipotéticas ultrapassa a faixa dos 60% de todo o genoma (Choi and El-Sayed 2012; Pawar et al. 2014; Ravooru et al. 2014; Bhardwaj et al. 2016). Portanto, diversas abordagens computacionais têm sido empregadas para analisar essas sequências.

Essas 28 sequências anotadas como hipotéticas foram analisadas com a utilização do *InterproScan* com o objetivo de se encontrar alguma anotação biológica para elas. Das 28 sequências inicialmente utilizadas foi possível encontrar alguma correlação em outros bancos de dados para oito delas, conforme descrito na tabela 4.12.

Tabela 4.12 - Resultado do *InterproScan* para as proteínas anotadas como hipotéticas.

Identificador	Banco de dados	Função predita
LbrM.14.0490	PANTHER	<i>Amastin surface glycoprotein</i>
LbrM.14.0490	Pfam	<i>Amastin surface glycoprotein</i>
LbrM.14.0500	Pfam	<i>Amastin surface glycoprotein</i>
LINF_320045500	Pfam	<i>Cornichon protein, putative</i>
LINF_320045500	SMART	<i>Cornichon protein, putative</i>
LINF_070014400	ProSiteProfiles	<i>Prokaryotic membrane lipoprotein lipid attachment site profile</i>
LINF_230014100	PANTHER	<i>Rab11-family interacting protein class I</i>
LAMA_000443100	PANTHER	<i>Rab11-family interacting protein class I</i>

Fonte: Dados do trabalho

As amastinas são uma família multigênica de glicoproteínas de superfície, que em *Leishmania* apresentam mais de 45 cópias. É uma proteína transmembrana, apresentando quatro domínios membranares que podem estar associados ao transporte de íons ou em mecanismos de sinalização celular (Kumar et al. 2010, 2012). Diversos estudos têm reportado o papel imunogênico dessas sequências em modelo murino (Stober et al. 2006; De Luca and Macedo 2016; Zhang et al. 2020). Além de serem consideradas biomarcadores para LV e fatores de virulência em Tripanossomatídeos (Rafati et al. 2006; Salotra et al. 2006).

Quanto à proteína anotada como “*Cornichon protein*”, não foi encontrada na literatura nenhuma informação sobre a função que ela desempenha em *Leishmania*, ou sobre seu potencial imunogênico. No entanto, em um estudo de proteômica, do tipo *label-free*, essa proteína também foi encontrada como sendo expressa em *Leishmania*. Em outros organismos, essa proteína atua no transporte e secreção de proteínas (Castro et al. 2007).

Por sua vez, os domínios *Prokaryotic membrane lipoprotein lipid attachment site profile* foram reportados em bactérias como envolvidos na entrega de moléculas efetoras às células eucarióticas alvo, causando uma série de mudanças na fisiologia da célula hospedeira durante a infecção (Pj 2001).

Já as proteínas com domínios *Rab11-family interacting protein class I* são consideradas um marcador de divisão flagelar (Ambit et al. 2011). Essas GTPases são fatores regulatórios importantes para a formação de vesículas, motilidade e fusão (Stenmark and Olkkonen 2001; Patel et al. 2008).

5. Conclusão

A utilização de dados de sequenciamento de RNA de *L. infantum* permitiu não somente a identificação de genes diferencialmente expressos, como também a identificação de aproximadamente 2.000 transcritos não anotados no genoma desses parasitos. Esses transcritos foram mapeados em diferentes vias relacionadas a mecanismos de resistência a drogas, processamento de RNA e manutenção do ciclo celular. A identificação desses transcritos para a utilização em estratégias de imunoinformática é extremamente importante por ampliar o conjunto de dados utilizado para as predições, não se limitando a sequências já conhecidas. Das sequências obtidas através da identificação de novos transcritos, duas delas apresentaram epítomos com potencial imunogênico e estão incluídas na análise. Dois novos transcritos foram identificados contendo epítomos com potencial imunogênico.

Por sua vez, através da análise de expressão diferencial foi possível identificar um conjunto de dados robusto que evidencia diversos mecanismos de resistência a drogas através da ativação de vias bioquímicas que têm o potencial de proteger o parasito do *stress* oxidativo causado pelo antimônio. Essas análises também possibilitaram a identificação de transcritos expressos em formas promastigotas que podem ser utilizados como potenciais candidatos vacinais contra *Leishmania* através da construção de uma vacina multiepítomo que também seja eficaz contra a infecção por parasitos resistentes, uma vez que, no Brasil estima-se que cerca de 25% dos insucessos terapêuticos estejam relacionados à resistência. Além disso, a identificação de genes expressos de forma constitutiva entre as formas promastigotas e amastigotas pode contribuir para um maior sucesso em intervenções imunológicas, uma vez que não se limita a epítomos que sejam expressos em apenas uma forma.

Conforme já esperado, a maior parte dessas sequências identificadas foram anotadas como codificadoras de proteínas hipotéticas. No entanto, apesar da ausência de função predita, essas proteínas representam um conjunto de dados de extrema importância e que deve ser analisado mais profundamente.

Além disso, a análise de redes de regulação gênica possibilitou a identificação das vias de TNF, via de sinalização de *NF-kappa B*, via de interação de

receptores do tipo *Cytokine-cytokine*, via de sinalização de MAPK, via de sinalização de receptores do tipo *Toll*, via de sinalização de IL-10 e de IL-17 como alteradas em macrófagos infectados com *Leishmania*. A identificação e modelagem dessas vias irá permitir uma melhor caracterização de como ocorre a modulação do sistema imune por parte do parasito, além da identificação de pontos nessas redes que, quando bloqueados, podem reduzir a resposta imunológica gerada frente à infecção pelo parasito.

A integração desses dados possibilitou a identificação de proteínas nunca antes identificadas como potenciais alvos para o desenvolvimento de vacinas. Apesar de, até o momento, não ser possível atribuir função biológica para um grande número delas, esse grupo apresenta características potencialmente imunogênicas que devem ser consideradas.

Foram priorizados, para estudos futuros envolvendo validação experimental, um total de 20 epítomos. Esses epítomos foram priorizados por apresentarem predição para alelos mais frequentes na população brasileira, por serem altamente imunogênicos, não serem homólogos com proteínas de humano e camundongo, além de conservados entre as diferentes espécies de *Leishmania*. A grande similaridade entre os epítomos encontrados é um fator de extrema relevância para o desenvolvimento de uma vacina multiepítomos, que seja capaz de oferecer proteção contra mais de uma espécie de *Leishmania*, o que é especialmente relevante em regiões com sobreposição geográfica de espécies.

6 Perspectivas

Apesar dos epítomos identificados neste trabalho terem passado por vários processos de análise imunoinformática que visam a garantir uma seleção mais assertiva de possíveis candidatos vacinais, quando comparado às metodologias convencionais, uma avaliação de sua eficácia *in vitro* e em modelo experimental é extremamente necessária.

Além disso, a modelagem de outras redes de regulação gênica que interagem com a sinalização da via da IL-17 poderá agregar conhecimento sobre as alterações que esses parasitos causam nos macrófagos infectados e possibilitar a identificação de pontos para a intervenção terapêutica.

REFERÊNCIAS

- Abdi H, Williams LJ. Principal component analysis. *WIREs Computational Statistics*. 2010;2(4):433–59.
- Afrin F, Khan I, Hemeg HA. Leishmania-Host Interactions—An Epigenetic Paradigm. *Front Immunol* [Internet]. 2019 [cited 2020 Nov 17];10. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00492/full>
- Akerman M, Shaked-Mishan P, Mazareb S, Volpin H, Zilberstein D. Novel motifs in amino acid permease genes from *Leishmania*. *Biochem Biophys Res Commun*. 2004 Dec 3;325(1):353–66.
- Akhoundi M, Kuhls K, Cannet A, Votýpka J, Marty P, Delaunay P, et al. A Historical Overview of the Classification, Evolution, and Dispersion of *Leishmania* Parasites and Sandflies. Bañuls A-L, editor. *PLOS Neglected Tropical Diseases*. 2016 Mar 3;10(3):e0004349.
- Akya A, Farasat A, Ghadiri K, Rostamian M. Identification of HLA-I restricted epitopes in six vaccine candidates of *Leishmania tropica* using immunoinformatics and molecular dynamics simulation approaches. *Infection, Genetics and Evolution*. 2019 Nov 1;75:103953.
- Alonso G, Rastrojo A, López-Pérez S, Requena JM, Aguado B. Resequencing and assembly of seven complex loci to improve the *Leishmania major* (Friedlin strain) reference genome. *Parasites & Vectors* [Internet]. 2016 Dec [cited 2018 Feb 16];9(1). Available from: <http://www.parasitesandvectors.com/content/9/1/74>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990 Oct 5;215(3):403–10.
- Ambit A, Woods KL, Cull B, Coombs GH, Mottram JC. Morphological events during the cell cycle of *Leishmania major*. *Eukaryot Cell*. 2011 Nov;10(11):1429–38.
- Anders S, Pyl PT, Huber W. HTSeq – A Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2014 Sep 25;btu638.
- Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. 2016 Feb 15;32(4):511–7.
- Anversa L, Tiburcio MGS, Richini-Pereira VB, Ramirez LE, Anversa L, Tiburcio MGS, et al. Human leishmaniasis in Brazil: A general review. *Revista da Associação Médica Brasileira*. 2018 Mar;64(3):281–9.
- Aoki JI, Muxel SM, Zampieri RA, Laranjeira-Silva MF, Müller KE, Nerland AH, et al. RNA-seq transcriptional profiling of *Leishmania amazonensis* reveals an arginase-dependent gene expression regulation. *PLOS Neglected Tropical Diseases*. 2017 Oct 27;11(10):e0006026.
- Arcari T, Manzano JI, Gamarro F. ABCI3 Is a New Mitochondrial ABC Transporter from *Leishmania major* Involved in Susceptibility to Antimonials and Infectivity. *Antimicrobial Agents and Chemotherapy*. 2017 Dec 1;61(12):e01115-17.
- Ardito F, Giuliani M, Perrone D, Troiano G, Muzio LL. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med*. 2017 Aug;40(2):271–80.
- Arya A, Arora SK, Arya A, Arora SK. Epitope Based Vaccine Designing- A mini review. *Journal of Vaccines and Immunology*. 2020 Nov 24;6(1):038–41.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000 May;25(1):25–9.
- Aslett M, Aurrecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucl Acids Res*. 2010 Jan 1;38(suppl 1):D457–62.
- Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome Medicine*. 2015 Nov 20;7(1):119.
- Balaguer MA de L, Fisher AP, Clark NM, Fernandez-Espinosa MG, Möller BK, Weijers D, et al. Predicting gene regulatory networks by combining spatial and temporal gene

- expression data in Arabidopsis root stem cells. PNAS. 2017 Sep 5;114(36):E7632–40.
- Bambini S, Rappuoli R. The use of genomics in microbial vaccine development. *Drug Discov Today*. 2009 Mar;14(5–6):252–60.
- Banerjee A, Bhattacharya P, Dagur PK, Karmakar S, Ismail N, Joshi AB, et al. Live Attenuated *Leishmania donovani* Centrin Gene–Deleted Parasites Induce IL-23–Dependent IL-17–Protective Immune Response against Visceral Leishmaniasis in a Murine Model. *The Journal of Immunology*. 2018 Jan 1;200(1):163–76.
- Banf M, Rhee SY. Computational inference of gene regulatory networks: Approaches, limitations and opportunities. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. 2017 Jan 1;1860(1):41–52.
- Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*. 2011 Jan;12(1):56–68.
- Basu JM, Mookerjee A, Banerjee R, Saha M, Singh S, Naskar K, et al. Inhibition of ABC Transporters Abolishes Antimony Resistance in *Leishmania* Infection. *Antimicrobial Agents and Chemotherapy*. 2008 Mar 1;52(3):1080–93.
- Bethony JM, Cole RN, Guo X, Kamhawi S, Lightowers MW, Loukas A, et al. Vaccines to combat the neglected tropical diseases. *Immunol Rev*. 2011 Jan;239(1):237–70.
- Bhardwaj R, Kumar R, Singh SK, Selvaraj C, Dubey VK. Understanding the importance of conservative hypothetical protein LdBPK_070020 in *Leishmania donovani* and its role in subsistence of the parasite. *Arch Biochem Biophys*. 2016 Apr 15;596:10–21.
- Bhattacharya P, Ali N. Involvement and interactions of different immune cells and their cytokines in human visceral leishmaniasis. *Rev Soc Bras Med Trop*. 2013 Apr;46(2):128–34.
- Biyani N, Singh AK, Mandal S, Chawla B, Madhubala R. Differential expression of proteins in antimony-susceptible and -resistant isolates of *Leishmania donovani*. *Molecular and Biochemical Parasitology*. 2011 Oct;179(2):91–9.
- Boaventura VS, Santos CS, Cardoso CR, de Andrade J, Dos Santos WLC, Clarêncio J, et al. Human mucosal leishmaniasis: neutrophils infiltrate areas of tissue damage that express high levels of Th17-related cytokines. *Eur J Immunol*. 2010 Oct;40(10):2830–6.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114–20.
- Brasil, Ministério da Saúde, Secretária de Vigilância em Saúde. Manual de vigilância da leishmaniose tegumentar [Internet]. [cited 2018 Jun 29]. Available from: <https://files.zotero.net/19941522731/Manual%20de%20vigil%C3%A2ncia%20da%20leishmaniose%20tegumentar%20%5Br.pdf>
- Brasil S de V em S. Boletim Epidemiológico - Doenças tropicais negligenciadas.
- Brito RCF, Guimarães FG, Velloso JPL, Corrêa-Oliveira R, Ruiz JC, Reis AB, et al. Immunoinformatics Features Linked to *Leishmania* Vaccine Development: Data Integration of Experimental and In Silico Studies. *International Journal of Molecular Sciences*. 2017 Feb;18(2):371.
- Brito RCFD, Ruiz JC, Cardoso JM de O, Ostolin TLVDP, Reis LES, Mathias FAS, et al. Chimeric Vaccines Designed by Immunoinformatics-Activated Polyfunctional and Memory T Cells That Trigger Protection against Experimental Visceral Leishmaniasis. *Vaccines*. 2020 Jun;8(2):252.
- Bro R, K. Smilde A. Principal component analysis. *Analytical Methods*. 2014;6(9):2812–31.
- Brotherton M-C, Bourassa S, Légaré D, Poirier GG, Droit A, Ouellette M. Quantitative proteomic analysis of amphotericin B resistance in *Leishmania infantum*. *Int J Parasitol Drugs Drug Resist*. 2014 May 16;4(2):126–32.
- Brusic V, Zeleznikow J, Petrovsky N. Molecular immunology databases and data repositories. *Journal of Immunological Methods*. 2000 Apr 21;238(1):17–28.
- Bui AAT, Van Horn JD. Envisioning the future of 'big data' biomedicine. *Journal of Biomedical Informatics*. 2017 May 1;69:115–7.
- Bui H-H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy

- analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2007 Sep 26;8(1):361.
- Cai Y-D, Lu L, Chen L, He J-F. Predicting subcellular location of proteins using integrated-algorithm method. *Mol Divers*. 2010 Aug 1;14(3):551–8.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009 Dec 15;10:421.
- Castiglione F, Tieri P, Palma A, Jarrah AS. Statistical ensemble of gene regulatory networks of macrophage differentiation. *BMC Bioinformatics*. 2016 Dec 22;17(19):506.
- Castilho TM, Goldsmith-Pestana K, Lozano C, Valderrama L, Saravia NG, McMahon-Pratt D. Murine model of chronic *L. (Viannia) panamensis* infection: Role of IL-13 in disease. *Eur J Immunol*. 2010 Oct;40(10):2816–29.
- Castro CP, Piscopo D, Nakagawa T, Derynck R. Cornichon regulates transport and secretion of TGF α -related proteins in metazoan cells. *J Cell Sci*. 2007 Jul 15;120(14):2454–66.
- CDC. CDC - Neglected Tropical Diseases - Diseases [Internet]. 2020 [cited 2021 Apr 10]. Available from: <https://www.cdc.gov/globalhealth/ntd/diseases/index.html>
- Chamakh-Ayari R, Chenik M, Chakroun AS, Bahi-Jaber N, Aoun K, Meddeb-Garnaoui A. *Leishmania major* large RAB GTPase is highly immunogenic in individuals immune to cutaneous and visceral leishmaniasis. *Parasites & Vectors* [Internet]. 2017 Dec [cited 2019 Feb 18];10(1). Available from: <http://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-017-2127-3>
- Chandrasekaran S, Ament SA, Eddy JA, Rodriguez-Zas SL, Schatz BR, Price ND, et al. Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc Natl Acad Sci USA*. 2011 Nov 1;108(44):18020–5.
- Chauhan IS, Rao GS, Singh N. Enhancing the copy number of *Ldrab6* gene in *Leishmania donovani* parasites mediates drug resistance through drug-thiol conjugate dependent multidrug resistance protein A (MRPA). *Acta Tropica*. 2019 Nov 1;199:105158.
- Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009 Jul 1;37(Web Server issue):W305–11.
- Choi J, El-Sayed NM. Functional genomics of trypanosomatids. *Parasite Immunology*. 2012;34(2–3):72–9.
- Clayton C. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biology*. 2019;9(6):190072.
- Clayton C, Shapira M. Post-transcriptional regulation of gene expression in trypanosomes and leishmaniasis. *Mol Biochem Parasitol*. 2007 Dec;156(2):93–101.
- Coelho VTS, Oliveira JS, Valadares DG, Chávez-Fumagalli MA, Duarte MC, Lage PS, et al. Identification of Proteins in Promastigote and Amastigote-like *Leishmania* Using an Immunoproteomic Approach. *PLOS Neglected Tropical Diseases*. 2012 Jan 17;6(1):e1430.
- Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*. 2008;2008:619832.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005 Sep 15;21(18):3674–6.
- Costa WLO da, Araújo CL de A, Dias LM, Pereira LC de S, Alves JTC, Araújo FA, et al. Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. *PLOS ONE*. 2018 Jun 25;13(6):e0198965.
- Couto J, Seixas G, Stutzer C, Olivier NA, Maritz-Olivier C, Antunes S, et al. Probing the *Rhipicephalus bursa* Sialomes in Potential Anti-Tick Vaccine Candidates: A Reverse Vaccinology Approach. *Biomedicine*. 2021 Apr;9(4):363.
- Croft SL, Olliaro P. Leishmaniasis chemotherapy--challenges and opportunities. *Clin Microbiol Infect*. 2011 Oct;17(10):1478–83.
- Croft SL, Sundar S, Fairlamb AH. Drug Resistance in Leishmaniasis. *Clin Microbiol Rev*. 2006 Jan 1;19(1):111–26.

- Cuervo P, Domont GB, De Jesus JB. Proteomics of trypanosomatids of human medical importance. *Journal of Proteomics*. 2010 Mar 10;73(5):845–67.
- Culligan EP, Sleator RD, Marchesi JR, Hill C. Metagenomics and novel gene discovery. *Virulence*. 2014 Apr 1;5(3):399–412.
- Das S, Shah P, Baharia RK, Tandon R, Khare P, Sundar S, et al. Over-Expression of 60s Ribosomal L23a Is Associated with Cellular Proliferation in SAG Resistant Clinical Isolates of *Leishmania donovani*. Bates PA, editor. *PLoS Neglected Tropical Diseases*. 2013 Dec 5;7(12):e2527.
- De Groot AS, Moise L, McMurry JA, Martin W. Epitope-Based Immunome-Derived Vaccines: A Strategy for Improved Design and Safety. *Clinical Applications of Immunomics*. 2008 Jul 31;2:39–69.
- De Luca PM, Macedo ABB. Cutaneous Leishmaniasis Vaccination: A Matter of Quality. *Front Immunol* [Internet]. 2016 [cited 2021 Apr 8];7. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2016.00151/full>
- Delgado FM, Gómez-Vela F. Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. *Artificial Intelligence in Medicine*. 2019 Apr 1;95:133–45.
- Demicheli C, Frézard F, Mangrum JB, Farrell NP. Interaction of trivalent antimony with a CCHC zinc finger domain: potential relevance to the mechanism of action of antimonial drugs. *Chem Commun (Camb)*. 2008 Oct 21;(39):4828–30.
- Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*. 2003 Aug 14;4(9):R60.
- Denny PW, Goulding D, Ferguson MAJ, Smith DF. Sphingolipid-free *Leishmania* are defective in membrane trafficking, differentiation and infectivity. *Molecular Microbiology*. 2004 Apr 1;52(2):313–27.
- Dias DS, Martins VT, Ribeiro PAF, Ramos FF, Lage DP, Tavares GSV, et al. Antigenicity, immunogenicity and protective efficacy of a conserved *Leishmania* hypothetical protein against visceral leishmaniasis. *Parasitology*. 2018 May;145(6):740–51.
- Dias DS, Ribeiro PAF, Martins VT, Lage DP, Portela ÁSB, Costa LE, et al. Recombinant small glutamine-rich tetratricopeptide repeat-containing protein of *Leishmania infantum*: Potential vaccine and diagnostic application against visceral leishmaniasis. *Mol Immunol*. 2017;91:272–81.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15–21.
- Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Current Protocols in Bioinformatics*. 2015;51(1):11.14.1-11.14.19.
- Dongen S. Graph Clustering Via a Discrete Uncoupling Process. *SIAM J Matrix Anal Appl*. 2008;
- Douanne N, Wagner V, Roy G, Leprohon P, Ouellette M, Fernandez-Prada C. MRPA-independent mechanisms of antimony resistance in *Leishmania infantum*. *Int J Parasitol Drugs Drug Resist*. 2020 Apr 2;13:28–37.
- Downing T, Imamura H, Decuypere S, Clark TG, Coombs GH, Cotton JA, et al. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res*. 2011 Jan 12;21(12):2143–56.
- Doytchinova D, Flower D. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine*. 2006 Sep 28;25(5):856–66.
- Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007 Jan 5;8(1):4.
- El-Awady R, Saleh E, Hashim A, Soliman N, Dallah A, Elrasheed A, et al. The Role of Eukaryotic and Prokaryotic ABC Transporter Family in Failure of Chemotherapy. *Front Pharmacol* [Internet]. 2017 [cited 2019 Aug 27];7. Available from: <https://www.frontiersin.org/articles/10.3389/fphar.2016.00535/full>

- El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al. Comparative Genomics of Trypanosomatid Parasitic Protozoa. *Science*. 2005 Jul 15;309(5733):404–9.
- Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016 Oct 1;32(19):3047–8.
- Farajnia S, Rahbarnia L, Maleki zanjani B, Alimohammadian MH, Abdoli Oskoe S, Beh-pajoo A, et al. Molecular Cloning and Characterization of P4 Nuclease from *Leishmania infantum*. *Enzyme Res* [Internet]. 2011 Jun 28 [cited 2021 Apr 8];2011. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3132502/>
- Fernandes MC, Dillon LAL, Belew AT, Bravo HC, Mosser DM, El-Sayed NM. Dual Transcriptome Profiling of *Leishmania*-Infected Human Macrophages Reveals Distinct Reprogramming Signatures. *mBio* [Internet]. 2016 Jul 6 [cited 2020 Jan 16];7(3). Available from: <https://mbio.asm.org/content/7/3/e00027-16>
- Fernandez-Prada C, Vincent IM, Brotherton M-C, Roberts M, Roy G, Rivas L, et al. Different Mutations in a P-type ATPase Transporter in *Leishmania* Parasites are Associated with Cross-resistance to Two Leading Drugs by Distinct Mechanisms. *PLoS Negl Trop Dis*. 2016 Dec;10(12):e0005171.
- Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*. 2011 Jul 1;39(suppl_2):W29–37.
- Folgueira C, Requena JM. A postgenomic view of the heat shock proteins in kinetoplastids. *FEMS Microbiol Rev*. 2007 Jul;31(4):359–77.
- Fonseca MS, Comini MA, Resende BV, Santi AMM, Zoboli AP, Moreira DS, et al. Ornithine decarboxylase or gamma-glutamylcysteine synthetase overexpression protects *Leishmania* (Vianna) *guyanensis* against antimony. *Experimental Parasitology*. 2017 Apr;175:36–43.
- Freeman TC, Horsewell S, Patir A, Harling-Lee J, Regan T, Shih BB, et al. Graphia: A platform for the graph-based visualisation and analysis of complex data. *bioRxiv*. 2020 Sep 3;2020.09.02.279349.
- Frézard F, Silva H, Pimenta AM de C, Farrell N, Demicheli C. Greater binding affinity of trivalent antimony to a CCCH zinc finger domain compared to a CCHC domain of kinetoplastid proteins. *Metallomics*. 2012 May;4(5):433–40.
- Fuente SG la, Peiró-Pastor R, Rastrojo A, Moreno J, Carrasco-Ramiro F, Requena JM, et al. Resequencing of the *Leishmania infantum* (strain JPCM5) genome and de novo assembly into 36 contigs. *Scientific Reports*. 2017 Dec 22;7(1):18050.
- Gaudenzi JGD, Noé G, Campo VA, Frasch AC, Cassola A. Gene expression regulation in trypanosomatids. *Essays In Biochemistry*. 2011 Oct 24;51:31–46.
- Ghedira K, Hornischer K, Konovalova T, Jenhani A-Z, Benkahla A, Kel A. Identification of key mechanisms controlling gene expression in *Leishmania* infected macrophages using genome-wide promoter analysis. *Infection, Genetics and Evolution*. 2011 Jun 1;11(4):769–77.
- Ghosh M, Bandyopadhyay S. Interaction of *Leishmania* parasites with dendritic cells and its functional consequences. *Immunobiology*. 2004 Aug 23;209(1):173–7.
- Gonçalves-de-Albuquerque S da C, Pessoa-e-Silva R, Trajano-Silva LAM, de Goes TC, de Moraes RCS, da C. Oliveira CN, et al. The Equivocal Role of Th17 Cells and Neutrophils on Immunopathogenesis of Leishmaniasis. *Front Immunol* [Internet]. 2017 [cited 2020 Oct 23];8. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2017.01437/full>
- Gonzalez-Galarza FF, McCabe A, Santos EJM dos, Jones J, Takeshita L, Ortega-Rivera ND, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Research*. 2020 Jan 8;48(D1):D783–8.
- Gonzalez-Lombana C, Gimblet C, Bacellar O, Oliveira WW, Passos S, Carvalho LP, et al. IL-17 Mediates Immunopathology in the Absence of IL-10 Following *Leishmania* major Infection. *PLOS Pathogens*. 2013 Mar 21;9(3):e1003243.
- Gourbal B, Sonuc N, Bhattacharjee H, Legare D, Sundar S, Ouellette M, et al. Drug uptake

- and modulation of drug resistance in *Leishmania* by an aquaglyceroporin. *J Biol Chem*. 2004 Jul 23;279(30):31010–7.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011 May 15;29(7):644–52.
- Grace CA, Forrester S, Silva VC, Aare A, Kilford H, Chew YP, et al. Multiple targets of balancing selection in *Leishmania donovani* complex parasites. *bioRxiv*. 2021 Mar 2;2021.03.02.433528.
- Graw S, Chappell K, L. Washam C, Gies A, Bird J, S. Robeson M, et al. Multi-omics data integration considerations and study design for biological systems and disease. *Molecular Omics* [Internet]. 2021 [cited 2021 Apr 13]; Available from: <https://pubs.rsc.org/en/content/articlelanding/2021/mo/d0mo00041h>
- Gregorio ED, Rappuoli R. Vaccines for the future: learning from human immunology. *Microbial Biotechnology*. 2012;5(2):149–55.
- Guerin PJ, Olliaro P, Sundar S, Boelaert M, Croft SL, Desjeux P, et al. Visceral leishmaniasis: current status of control, diagnosis, and treatment, and a proposed research and development agenda. *The Lancet Infectious Diseases*. 2002 Aug;2(8):494–501.
- Guimarães-Costa AB, DeSouza-Vieira TS, Paletta-Silva R, Freitas-Mesquita AL, Meyer-Fernandes JR, Saraiva EM. 3'-Nucleotidase/Nuclease Activity Allows *Leishmania* Parasites To Escape Killing by Neutrophil Extracellular Traps. *Infect Immun*. 2014 Apr;82(4):1732–40.
- Guimond C, Trudel N, Brochu C, Marquis N, El Fadili A, Peytavi R, et al. Modulation of gene expression in *Leishmania* drug resistant mutants as determined by targeted DNA microarrays. *Nucleic Acids Res*. 2003 Oct 15;31(20):5886–96.
- Gutiérrez V, B. Seabra A, M. Reguera R, Khandare J, Calderón M. New approaches from nanomedicine for treating leishmaniasis. *Chemical Society Reviews*. 2016;45(1):152–68.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc* [Internet]. 2013 Aug [cited 2021 Apr 15];8(8). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3875132/>
- Haile S, Papadopoulou B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Current Opinion in Microbiology*. 2007 Dec;10(6):569–77.
- Handman E. *Leishmania* virulence: it's a knock out! *Trends in Parasitology*. 2001 Feb 1;17(2):60.
- Hardy S, Robillard PN. Modeling and simulation of molecular biology systems using petri nets: modeling goals of various approaches. *J Bioinform Comput Biol*. 2004 Dec 1;02(04):619–37.
- von Heijne G. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res*. 1986 Jun 11;14(11):4683–90.
- Henard CA, Carlsen ED, Hay C, Kima PE, Soong L. *Leishmania amazonensis* Amastigotes Highly Express a Tryparedoxin Peroxidase Isoform That Increases Parasite Resistance to Macrophage Antimicrobial Defenses and Fosters Parasite Virulence. *PLOS Neglected Tropical Diseases*. 2014 Jul 17;8(7):e3000.
- Higgins CF. ABC Transporters: From Microorganisms to Man. *Annual Review of Cell Biology*. 1992;8(1):67–113.
- Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*. 2007 Jul;35(Web Server issue):W585–7.
- Hu Y, Ota N, Peng I, Refino CJ, Danilenko DM, Caplazi P, et al. IL-17RC is required for IL-17A- and IL-17F-dependent signaling and the pathogenesis of experimental autoimmune encephalomyelitis. *J Immunol*. 2010 Apr 15;184(8):4307–16.
- Ibarra-Meneses AV, Moreno J, Carrillo E. New Strategies and Biomarkers for the Control of Visceral Leishmaniasis. *Trends in Parasitology*. 2020 Jan 1;36(1):29–38.
- Ilari A, Fiorillo A, Genovese I, Colotti G. Polyamine-trypanothione pathway: an update. *Future*

- Med Chem. 2017;9(1):61–77.
- Illumina I. Quality Scores for Next-Generation Sequencing. 2011 Oct 31 [cited 2016 Jan 22]; Available from: http://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf
- Illumina I. Understanding Illumina Quality Scores. 2014 Apr 23 [cited 2016 Jan 22]; Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_understanding_quality_scores.pdf
- Iqbal H, Ishfaq M, Wahab A, Abbas MN, Ahmad I, Rehman A, et al. Therapeutic modalities to combat leishmaniasis, a review. *Asian Pacific Journal of Tropical Disease*. 2016 Jan;6(1):1–5.
- Isnard A, Shio MT, Olivier M. Impact of *Leishmania* metalloprotease GP63 on macrophage signaling. *Front Cell Infect Microbiol* [Internet]. 2012 [cited 2019 Apr 3];2. Available from: <https://www.frontiersin.org/articles/10.3389/fcimb.2012.00072/full>
- Ivens AC. The Genome of the Kinetoplastid Parasite, *Leishmania major*. *Science*. 2005 Jul 15;309(5733):436–42.
- Jackson CA, Castro DM, Saldi G-A, Bonneau R, Gresham D. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. Barkai N, Weigel D, editors. *eLife*. 2020 Jan 27;9:e51254.
- Janowski S, Kormeier B, Töpel T, Hippe K, Hofestädt R, Willassen N, et al. Modeling of Cell-to-Cell Communication Processes with Petri Nets Using the Example of Quorum Sensing. *In Silico Biology*. 2010 Jan 1;10(1,2):27–48.
- Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 2017 Jul 3;45(Web Server issue):W24–9.
- Joyce AR, Palsson BØ. The model organism as a system: integrating “omics” data sets. *Nature Reviews Molecular Cell Biology*. 2006 Mar;7(3):198–210.
- Karch CP, Burkhard P. Vaccine technologies: From whole organisms to rationally designed protein assemblies. *Biochemical Pharmacology*. 2016 Nov 15;120:1–14.
- Katara GK, Ansari NA, Singh A, Ramesh V, Salotra P. Evidence for Involvement of Th17 Type Responses in Post Kala Azar Dermal Leishmaniasis (PKDL). *PLOS Neglected Tropical Diseases*. 2012 Jun 19;6(6):e1703.
- Katara GK, Raj A, Kumar R, Avishek K, Kaushal H, Ansari NA, et al. Analysis of localized immune responses reveals presence of Th17 and Treg cells in cutaneous leishmaniasis due to *Leishmania tropica*. *BMC Immunology*. 2013 Nov 22;14(1):52.
- Kazemi-Rad E, Mohebbi M, Khadem-Erfan MB, Hajjarian H, Hadighi R, Khamesipour A, et al. Overexpression of Ubiquitin and Amino Acid Permease Genes in Association with Antimony Resistance in *Leishmania tropica* Field Isolates. *Korean J Parasitol*. 2013a Aug;51(4):413–9.
- Kazemi-Rad E, Mohebbi M, Khadem-Erfan MB, Saffari M, Raoofian R, Hajjarian H, et al. Identification of antimony resistance markers in *Leishmania tropica* field isolates through a cDNA-AFLP approach. *Experimental Parasitology*. 2013b Oct;135(2):344–9.
- Kazi A, Chuah C, Majeed ABA, Leow CH, Lim BH, Leow CY. Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. *Pathog Glob Health*. 2018 May;112(3):123–31.
- Kedzierski L. Leishmaniasis Vaccine: Where are We Today? *J Glob Infect Dis*. 2010 May;2(2):177–85.
- Kedzierski L, Zhu Y, Handman E. *Leishmania* vaccines: progress and problems. *Parasitology*. 2006;133 Suppl:S87–112.
- Kelly DF, Rappuoli R. Reverse vaccinology and vaccines for serogroup B *Neisseria meningitidis*. *Adv Exp Med Biol*. 2005;568:217–23.
- t'Kindt R, Scheltema RA, Jankevics A, Brunker K, Rijal S, Dujardin J-C, et al. Metabolomics to Unveil and Understand Phenotypic Diversity between Pathogen Populations. *PLOS Neglected Tropical Diseases*. 2010 Nov 30;4(11):e904.

- Korpelainen E, Tuimala J, Somervuo P, Huss M, Wong G. RNA-seq Data Analysis A Practical Approach. CRC Press; 2015. (CHAPMAN & HALL/CRC Mathematical and Computational Biology Series).
- Kramer S. Developmental regulation of gene expression in the absence of transcriptional control: The case of kinetoplastids. *Molecular and Biochemical Parasitology*. 2012 Feb;181(2):61–72.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001 Jan 19;305(3):567–80.
- Kumar A, Sisodia B, Misra P, Sundar S, Shasany AK, Dube A. Proteome mapping of overexpressed membrane-enriched and cytosolic proteins in sodium antimony gluconate (SAG) resistant clinical isolate of *Leishmania donovani*: Overexpressed proteins in SAG resistant *L. donovani*. *British Journal of Clinical Pharmacology*. 2010 Oct;70(4):609–17.
- Kumar D, Singh R, Bhandari V, Kulshrestha A, Negi NS, Salotra P. Biomarkers of antimony resistance: need for expression analysis of multiple genes to distinguish resistance phenotype in clinical isolates of *Leishmania donovani*. *Parasitol Res*. 2012 Jul 1;111(1):223–30.
- Kutateladze T. Phosphatidylinositol 3-phosphate recognition and membrane docking by the FYVE domain. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. 2006 Aug;1761(8):868–77.
- Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol*. 2001 Feb;11(1):39–46.
- Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, et al. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol*. 2005 Aug 1;35(8):2295–303.
- Leifso K, Cohen-Freue G, Dogra N, Murray A, McMaster WR. Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: The *Leishmania* genome is constitutively expressed. *Molecular and Biochemical Parasitology*. 2007 Mar 1;152(1):35–46.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078–9.
- Li L. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*. 2003 Sep 1;13(9):2178–89.
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006 Jul 1;22(13):1658–9.
- Liang X, Young WC, Hung L-H, Raftery AE, Yeung KY. Integration of Multiple Data Sources for Gene Network Inference Using Genetic Perturbation Data. *J Comput Biol*. 2019 Oct 1;26(10):1113–29.
- Liarte DB, Murta SMF. Selection and phenotype characterization of potassium antimony tartrate-resistant populations of four New World *Leishmania* species. *Parasitol Res*. 2010 Apr 7;107(1):205–12.
- Liu D, Uzonna JE. The early interaction of *Leishmania* with macrophages and dendritic cells and its influence on the host immune response. *Front Cell Infect Microbiol* [Internet]. 2012 [cited 2019 Apr 3];2. Available from: <https://www.frontiersin.org/articles/10.3389/fcimb.2012.00083/full>
- Livigni A, O'Hara L, Polak ME, Angus T, Smith LB, Freeman TC. Petri Net-Based Graphical and Computational Modelling of Biological Systems. *bioRxiv*. 2016 Jun 22;047043.
- Lohia N, Baranwal M. An Immunoinformatics Approach in Design of Synthetic Peptide Vaccine Against Influenza Virus. In: Tomar N, editor. *Immunoinformatics* [Internet]. New York, NY: Springer US; 2020 [cited 2021 Apr 9]. p. 229–43. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-0716-0389-5_11
- Lopes-Ramos CM, Chen C-Y, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, et al. Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues.

- Cell Reports. 2020 Jun 23;31(12):107795.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014 Dec 5;15(12):550.
- Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020 Jan 8;48(D1):D265–8.
- Luna EJ de A, Campos SR de SL da C, Luna EJ de A, Campos SR de SL da C. Vaccine development against neglected tropical diseases. *Cadernos de Saúde Pública* [Internet]. 2020 [cited 2021 Apr 10];36. Available from: http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0102-311X2020001402005&lng=en&nrm=iso&tlng=en
- Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*. 2007 Jun;8(6):479–90.
- Maia-Elkhoury ANS, Alves WA, Sousa-Gomes ML de, Sena JM de, Luna EA. Visceral leishmaniasis in Brazil: trends and challenges. *Cadernos de Saúde Pública*. 2008 Dec;24(12):2941–7.
- Marín I. Origin and evolution of fungal HECT ubiquitin ligases. *Sci Rep*. 2018 Apr 23;8(1):1–13.
- Martinez PA, Petersen CA. Chronic infection by *Leishmania amazonensis* mediated through MAPK ERK mechanisms. *Immunol Res*. 2014 Aug;59(0):153–65.
- Martins VT, Duarte MC, Chávez-Fumagalli MA, Menezes-Souza D, Coelho CSP, de Magalhães-Soares DF, et al. A *Leishmania*-specific hypothetical protein expressed in both promastigote and amastigote stages of *Leishmania infantum* employed for the serodiagnosis of, and as a vaccine candidate against, visceral leishmaniasis. *Parasites & Vectors*. 2015 Jul 11;8(1):363.
- Matrangolo FSV, Liarte DB, Andrade LC, de Melo MF, Andrade JM, Ferreira RF, et al. Comparative proteomic analysis of antimony-resistant and -susceptible *Leishmania braziliensis* and *Leishmania infantum chagasi* lines. *Mol Biochem Parasitol*. 2013 Aug;190(2):63–75.
- Mazandu GK, Mulder NJ. Function Prediction and Analysis of *Mycobacterium tuberculosis* Hypothetical Proteins. *International Journal of Molecular Sciences*. 2012 Jun;13(6):7283–302.
- McConville MJ, de Souza D, Saunders E, Likic VA, Naderer T. Living in a phagolysosome; metabolism of *Leishmania amastigotes*. *Trends Parasitol*. 2007 Aug;23(8):368–75.
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D412–9.
- Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. *Nucl Acids Res*. 2015 Jan 28;43(D1):D213–21.
- Miyagi PE, Villani E, Gustin GDB, Maruyama N, Santos Filho DJ. Petri Net approach for modelling system integration in intelligent buildings. *Journal of the Brazilian Society of Mechanical Sciences*. 2002 Nov;24(4):341–50.
- Møller JV, Juul B, le Maire M. Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochim Biophys Acta*. 1996 May 6;1286(1):1–51.
- Moore EM, Lockwood DN. Treatment of Visceral Leishmaniasis. *J Glob Infect Dis*. 2010;2(2):151–8.
- Moreira D de S, Pescher P, Laurent C, Lenormand P, Späth GF, Murta SMF. Phosphoproteomic analysis of wild-type and antimony-resistant *Leishmania braziliensis* lines by 2D-DIGE technology. *Proteomics*. 2015 Sep 1;15(17):2999–3019.
- Moreira DS, Monte Neto RL, Andrade JM, Santi AMM, Reis PG, Frézard F, et al. Molecular characterization of the MRPA transporter and antimony uptake in four New World *Leishmania* spp. susceptible and resistant to antimony. *Int J Parasitol Drugs Drug Resist*. 2013 Sep 5;3:143–53.
- Moreira DS, Murta SMF. Involvement of nucleoside diphosphate kinase b and elongation factor 2 in *Leishmania braziliensis* antimony resistance phenotype. *Parasites &*

- Vectors. 2016 Dec 13;9(1):641.
- Mukherjee A, Padmanabhan PK, Singh S, Roy G, Girard I, Chatterjee M, et al. Role of ABC transporter MRPA, γ -glutamylcysteine synthetase and ornithine decarboxylase in natural antimony-resistant isolates of *Leishmania donovani*. *J Antimicrob Chemother*. 2007 Feb 1;59(2):204–11.
- Mukherjee A, Roy G, Guimond C, Ouellette M. The γ -glutamylcysteine synthetase gene of *Leishmania* is essential and involved in response to oxidants. *Molecular Microbiology*. 2009 Nov 1;74(4):914–27.
- Mukherjee S, Santara SS, Das S, Bose M, Roy J, Adak S. NAD(P)H Cytochrome b5 Oxidoreductase Deficiency in *Leishmania major* Results in Impaired Linoleate Synthesis Followed by Increased Oxidative Stress and Cell Death. *J Biol Chem*. 2012 Dec 10;287(42):34992–5003.
- Munshi Nikhil V. Gene Regulatory Networks in Cardiac Conduction System Development. *Circulation Research*. 2012 May 25;110(11):1525–37.
- Muraro D, Voß U, Wilson M, Bennett M, Byrne H, De Smet I, et al. Inference of the genetic network regulating lateral root initiation in *Arabidopsis thaliana*. *IEEE/ACM Trans Comput Biol Bioinform*. 2013 Feb;10(1):50–60.
- Mutiso JM, Macharia JC, Kiio MN, Ichagichu JM, Rikoi H, Gicheru MM. Development of *Leishmania* vaccines: predicting the future from past and present experience. *J Biomed Res*. 2013 Mar;27(2):85–102.
- Naderer T, McConville MJ. The *Leishmania*-macrophage interaction: a metabolic perspective. *Cell Microbiol*. 2008 Feb;10(2):301–8.
- National Center for Biotechnology Information. GenBank and WGS Statistics [Internet]. [cited 2019 Aug 31]. Available from: <https://www.ncbi.nlm.nih.gov/genbank/statistics/>
- Naula C, Parsons M, Mottram JC. Protein kinases as drug targets in trypanosomes and *Leishmania*. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2005 Dec;1754(1–2):151–9.
- Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics*. 2009 Sep 18;10:296.
- Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science*. 2003 May 1;12(5):1007–17.
- Noor E, Cherkaoui S, Sauer U. Biological insights through omics data integration. *Current Opinion in Systems Biology*. 2019 Jun 1;15:39–47.
- Nosrati M, Mohabatkar H, Behbahani M. A Novel Multi-Epitope Vaccine For Cross Protection Against Hepatitis C Virus (HCV): An Immunoinformatics Approach. *Research in Molecular Medicine*. 2017 Feb 10;5(1):17–26.
- O'Hara L, Livigni A, Theo T, Boyer B, Angus T, Wright D, et al. Modelling the Structure and Dynamics of Biological Pathways. *PLOS Biology*. 2016 Aug 10;14(8):e1002530.
- Oli AN, Obialor WO, Ifeanyichukwu MO, Odimegwu DC, Okoyeh JN, Emechebe GO, et al. Immunoinformatics and Vaccine Development: An Overview. *ITT*. 2020 Feb 26;9:13–30.
- Oliveira KL, Pacheco D, Dourado NM. Vaccines Vs Leishmaniasis: A Review. 2013;11.
- Olson EN. Gene Regulatory Networks in the Evolution and Development of the Heart. *Science*. 2006 Sep 29;313(5795):1922–7.
- Ouameur A, Girard I, Legare D, Ouellette M. Functional analysis and complex gene rearrangements of the folate/biopterin transporter (FBT) gene family in the protozoan parasite *Leishmania*. *Molecular and Biochemical Parasitology*. 2008 Dec;162(2):155–64.
- Ouellette M, Drummelsmith J, Papadopoulou B. Leishmaniasis: drugs in the clinic, resistance and new developments. *Drug Resistance Updates*. 2004 Aug;7(4–5):257–66.
- Ouellette M, Légaré D, Haimour A, Grondin K, Roy G, Brochu C, et al. ABC transporters in *Leishmania* and their role in drug resistance. *Drug Resistance Updates*. 1998 Mar;1(1):43–8.
- Padmanabhan PK, Zghidi-Abouzid O, Samant M, Dumas C, Aguiar BG, Estaquier J, et al.

- DDX3 DEAD-box RNA helicase plays a central role in mitochondrial protein quality control in *Leishmania*. *Cell Death & Disease*. 2016 Oct;7(10):e2406–e2406.
- PAHO/WHO. General Information on Neglected Infectious Diseases Integrated Approach [Internet]. Pan American Health Organization / World Health Organization. [cited 2021 Apr 10]. Available from: https://www.paho.org/hq/index.php?option=com_content&view=article&id=5753:2011-informacion-general-enfoque-integrado-enfermedades-infecciosas-desatendidas&Itemid=4141&lang=en
- Palma A, Jarrah AS, Tieri P, Cesareni G, Castiglione F. Gene Regulatory Network Modeling of Macrophage Differentiation Corroborates the Continuum Hypothesis of Polarization States. *Front Physiol* [Internet]. 2018 [cited 2020 Aug 10];9. Available from: <https://www.frontiersin.org/articles/10.3389/fphys.2018.01659/full>
- Pandey RK, Ojha R, Aathmanathan VS, Krishnan M, Prajapati VK. Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. *Vaccine*. 2018 19;36(17):2262–72.
- Papadopoulou B, Huang X-F, Boucher N, McNicoll F. Stage-Specific Regulation of Gene Expression in *Leishmania*. *ASM*. 2004;69:282–8.
- Patel N, Singh SB, Basu SK, Mukhopadhyay A. *Leishmania* requires Rab7-mediated degradation of endocytosed hemoglobin for their growth. *PNAS*. 2008 Mar 11;105(10):3980–5.
- Patronov Atanas, Doytchinova Irini. T-cell epitope vaccine design by immunoinformatics. *Open Biology*. 2012;3(1):120139.
- Pawar H, Kulkarni A, Dixit T, Chaphekar D, Patole MS. A bioinformatics approach to reanalyze the genome annotation of kinetoplastid protozoan parasite *Leishmania donovani*. *Genomics*. 2014 Dec;104(6):554–61.
- Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nature Genetics*. 2007 Jul;39(7):839–47.
- Pennisi M, Cavalieri S, Motta S, Pappalardo F. A methodological approach for using high-level Petri Nets to model the immune system response. *BMC Bioinformatics*. 2016 Dec 22;17(19):498.
- Peter IS, Davidson EH. Assessing regulatory information in developmental gene regulatory networks. *PNAS*. 2017 Jun 6;114(23):5862–9.
- Pimentel H, L BN, Puente S, Melsted P, Pachter L. Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv*. 2016 Jun 10;058164.
- Pinu FR, Beale DJ, Paten AM, Kouremenos K, Swarup S, Schirra HJ, et al. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites* [Internet]. 2019 Apr 18 [cited 2021 Apr 13];9(4). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6523452/>
- Pitta MGR, Romano A, Cabantous S, Henri S, Hammad A, Kouriba B, et al. IL-17 and IL-22 are associated with protection against human kala azar caused by *Leishmania donovani*. *J Clin Invest*. 2009 Aug;119(8):2379–87.
- Pj C. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol Microbiol*. 2001 Apr 1;40(2):294–305.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. *Nucl Acids Res*. 2005 Jan 7;33(suppl 2):W116–20.
- Rabienia M, Roudbari Z, Ghanbariasad A, Abdollahi A, Mohammadi E, Mortazavidehkordi N, et al. Exploring membrane proteins of *Leishmania major* to design a new multi-epitope vaccine using immunoinformatics approach. *European Journal of Pharmaceutical Sciences*. 2020 Sep 1;152:105423.
- Rafati S, Hassani N, Taslimi Y, Movassagh H, Rochette A, Papadopoulou B. Amastin Peptide-Binding Antibodies as Biomarkers of Active Human Visceral Leishmaniasis. *Clin Vaccine Immunol*. 2006 Jan 10;13(10):1104–10.
- Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine*. 2001 Mar 21;19(17–19):2688–91.

- Ravooru N, Ganji S, Sathyanarayanan N, Nagendra HG. Insilico analysis of hypothetical proteins unveils putative metabolic pathways and essential genes in *Leishmania donovani*. *Front Genet* [Internet]. 2014 [cited 2017 Jan 29];5. Available from: <http://journal.frontiersin.org/article/10.3389/fgene.2014.00291/abstract>
- Ready PD. Epidemiology of visceral leishmaniasis. *Clin Epidemiol*. 2014 May 3;6:147–54.
- Reddy VN, Mavrovouniotis ML, Liebman MN. Petri Net Representations in Metabolic Pathways. 1993;9.
- Requena JM, Montalvo AM, Fraga J. Molecular Chaperones of *Leishmania* : Central Players in Many Stress-Related and -Unrelated Physiological Processes. *BioMed Research International*. 2015;2015:1–21.
- Resende DM, Rezende AM, Oliveira NJ, Batista IC, Corrêa-Oliveira R, Reis AB, et al. An assessment on epitope prediction methods for protozoa genomes. *BMC Bioinformatics*. 2012;13:309.
- Ribeiro CV, Rocha BFB, Moreira D de S, Peruhype-Magalhães V, Murta SMF. Mannosyltransferase (GPI-14) overexpression protects promastigote and amastigote forms of *Leishmania braziliensis* against trivalent antimony. *Parasites & Vectors*. 2019 Jan 25;12(1):60.
- Ribeiro PAF, Dias DS, Novais MVM, Lage DP, Tavares GSV, Mendonça DVC, et al. A *Leishmania* hypothetical protein-containing liposome-based formulation is highly immunogenic and induces protection against visceral leishmaniasis. *Cytokine*. 2018;111:131–9.
- Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*. 2011 Jun 21;btr355.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotech*. 2011 Jan;29(1):24–6.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010 Jan 1;26(1):139–40.
- Robleda-Castillo R, Ros-Lucas A, Martinez-Peinado N, Alonso-Padilla J. An Overview of Current Uses and Future Opportunities for Computer-Assisted Design of Vaccines for Neglected Tropical Diseases. *Adv Appl Bioinform Chem*. 2021 Feb 15;14:25–47.
- Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Research*. 2011 Dec 1;21(12):2129–42.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, et al. Artemis: sequence visualization and annotation. *Bioinformatics*. 2000;16(10):944–5.
- Ryazanov AG, Davydova EK. Mechanism of elongation factor 2 (EF-2) inactivation upon phosphorylation. Phosphorylated EF-2 is unable to catalyze translocation. *FEBS Lett*. 1989 Jul 17;251(1–2):187–90.
- Salotra P, Duncan RC, Singh R, Subba Raju BV, Sreenivas G, Nakhasi HL. Upregulation of surface proteins in *Leishmania donovani* isolated from patients of post kala-azar dermal leishmaniasis. *Microbes Infect*. 2006 Mar;8(3):637–44.
- Samarasinghe SR, Samaranyake N, Kariyawasam UL, Siriwardana YD, Imamura H, Karunaweera ND. Genomic insights into virulence mechanisms of *Leishmania donovani*: evidence from an atypical strain. *BMC Genomics* [Internet]. 2018 Nov 28 [cited 2019 Aug 27];19. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6262978/>
- Sampaio RB, Fonseca BPF, Bahulkar A, Szymanski BK. Network analysis to support public health: evolution of collaboration among leishmaniasis researchers. *Scientometrics*. 2017 Jun 1;111(3):2001–21.
- Satuluri V, Parthasarathy S, Ucar D. Markov clustering of protein interaction networks with improved balance and scalability. In: *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology* [Internet]. New York, NY, USA: Association for Computing Machinery; 2010 [cited 2021 Apr 15]. p. 247–56. (BCB '10). Available from: <https://doi.org/10.1145/1854776.1854812>

- Scarselli M, Giuliani MM, Adu-Bobie J, Pizza M, Rappuoli R. The impact of genomics on vaccine design. *Trends in Biotechnology*. 2005a Feb 1;23(2):84–91.
- Scarselli M, Giuliani MM, Adu-Bobie J, Pizza M, Rappuoli R. The impact of genomics on vaccine design. *Trends Biotechnol*. 2005b Feb;23(2):84–91.
- Schenkman JB, Jansson I. The many roles of cytochrome b5. *Pharmacol Ther*. 2003 Feb;97(2):139–52.
- Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011 Mar 15;27(6):863–4.
- Schubert B, Brachvogel H-P, Jürges C, Kohlbacher O. EpiToolKit—a web-based workbench for vaccine design. *Bioinformatics*. 2015 Jul 1;31(13):2211–3.
- Serruto D, Rappuoli R. Post-genomic vaccine development. *FEBS Lett*. 2006 May 22;580(12):2985–92.
- Singh G, Pritam M, Banerjee M, Singh AK, Singh SP. Genome based screening of epitope ensemble vaccine candidates against dreadful visceral leishmaniasis using immunoinformatics approach. *Microbial Pathogenesis*. 2019 Nov 1;136:103704.
- Singh GP, Jha M, Singh M, Naina. Modeling the mechanism pathways of first line drug in Tuberculosis using Petri nets. *Int J Syst Assur Eng Manag*. 2020 Jul;11(S2):313–24.
- Singh H, Khan AA, Dinner AR. Gene regulatory networks in the immune system. *Trends in Immunology*. 2014 May 1;35(5):211–8.
- Sinha S, Jones BM, Traniello IM, Bukhari SA, Halfon MS, Hofmann HA, et al. Behavior-related gene regulatory networks: A new level of organization in the brain. *PNAS [Internet]*. 2020 Jul 13 [cited 2020 Aug 6]; Available from: <https://www.pnas.org/content/early/2020/07/10/1921625117>
- Smith DF, Peacock CS, Cruz AK. Comparative genomics: From genotype to disease phenotype in the leishmaniasis. *International Journal for Parasitology*. 2007 Sep 1;37(11):1173–86.
- Steele E, Tucker A, 't Hoen P a. C, Schuemie MJ. Literature-based priors for gene regulatory networks. *Bioinformatics*. 2009 Jul 15;25(14):1768–74.
- Stenmark H, Olkkonen VM. The Rab GTPase family. *Genome Biol*. 2001;2(5):REVIEWS3007.
- Stober CB, Lange UG, Roberts MTM, Gilmartin B, Francis R, Almeida R, et al. From genome to vaccines for leishmaniasis: Screening 100 novel vaccine candidates against murine *Leishmania major* infection. *Vaccine*. 2006 Mar 24;24(14):2602–16.
- Sundar S. Drug resistance in Indian visceral leishmaniasis. *Tropical Medicine & International Health*. 2001 Nov 1;6(11):849–54.
- Sundar S, Olliaro PL. Miltefosine in the treatment of leishmaniasis: Clinical evidence for informed clinical risk management. *Ther Clin Risk Manag*. 2007 Oct;3(5):733–40.
- Sundar S, Rai M. Laboratory Diagnosis of Visceral Leishmaniasis. *Clin Diagn Lab Immunol*. 2002 Sep 1;9(5):951–8.
- Sunyoto T, Potet J, Boelaert M. Why miltefosine—a life-saving drug for leishmaniasis—is unavailable to people who need it the most. *BMJ Glob Health [Internet]*. 2018 May 3 [cited 2021 Apr 11];3(3). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5935166/>
- Surur AS, Fekadu A, Makonnen E, Hailu A. Challenges and Opportunities for Drug Discovery in Developing Countries: The Example of Cutaneous Leishmaniasis. *ACS Med Chem Lett*. 2020 Nov 12;11(11):2058–62.
- Terry FE, Moise L, Martin RF, Torres M, Pilotte N, Williams SA, et al. Time for T? Immunoinformatics addresses vaccine design for neglected tropical and emerging infectious diseases. *Expert Rev Vaccines*. 2015 Jan 2;14(1):21–35.
- The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucl Acids Res*. 2015 Jan 28;43(D1):D1049–56.
- The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D480–9.
- Tibúrcio R, Nunes S, Nunes I, Rosa Ampuero M, Silva IB, Lima R, et al. Molecular Aspects of Dendritic Cell Activation in Leishmaniasis: An Immunobiological View. *Front Immunol*

- [Internet]. 2019 [cited 2019 Apr 3];10. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00227/full>
- Tomar N, De RK. Immunoinformatics: an integrated scenario. *Immunology*. 2010a Oct 1;131(2):153–68.
- Tomar N, De RK. Immunoinformatics: an integrated scenario. *Immunology*. 2010b Oct;131(2):153–68.
- Tomar N, De RK. Immunoinformatics: a brief review. *Methods Mol Biol*. 2014;1184:23–55.
- Tomiotto-Pellissier F, Bortoleti BT da S, Assolini JP, Gonçalves MD, Carloto ACM, Miranda-Sapla MM, et al. Macrophage Polarization in Leishmaniasis: Broadening Horizons. *Front Immunol* [Internet]. 2018 [cited 2020 Aug 6];9. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.02529/full>
- Tschoeke DA, Nunes GL, Jardim R, Lima J, Dumaresq AS, Gomes MR, et al. The Comparative Genomics and Phylogenomics of *Leishmania amazonensis* Parasite. *Evol Bioinform Online*. 2014 Sep 23;10:131–53.
- Tuon FF, Amato VS, Graf ME, Siqueira AM, Nicodemo AC, Neto VA. Treatment of New World cutaneous leishmaniasis – a systematic review with a meta-analysis. *International Journal of Dermatology*. 2008;47(2):109–24.
- Ulfenborg B. Vertical and horizontal integration of multi-omics data with miodin. *BMC Bioinformatics*. 2019 Dec 10;20(1):649.
- Vakili B, Eslami M, Hatam GR, Zare B, Erfani N, Nezafat N, et al. Immunoinformatics-aided design of a potential multi-epitope peptide vaccine against *Leishmania infantum*. *International Journal of Biological Macromolecules*. 2018 Dec 1;120:1127–39.
- Vargas-Inchaustegui DA, Xin L, Soong L. *Leishmania braziliensis* Infection Induces Dendritic Cell Activation, ISG15 Transcription, and the Generation of Protective Immune Responses. *J Immunol*. 2008 Jun 1;180(11):7537–45.
- Vega-López F. Diagnosis of cutaneous leishmaniasis. *Current Opinion in Infectious Diseases*. 2003 Apr;16(2):97–101.
- Veras PST, Bezerra de Menezes JP. Using Proteomics to Understand How *Leishmania* Parasites Survive inside the Host and Establish Infection. *International Journal of Molecular Sciences*. 2016 Aug;17(8):1270.
- Vergnes B, Gourbal B, Girard I, Sundar S, Drummelsmith J, Ouellette M. A Proteomics Screen Implicates HSP83 and a Small Kinetoplastid Calpain-related Protein in Drug Resistance in *Leishmania donovani* Clinical Field Isolates by Modulating Drug-induced Programmed Cell Death. *Mol Cell Proteomics*. 2007 Jan 1;6(1):88–101.
- Verma A, Ghosh S, Salotra P, Singh R. Artemisinin-resistant *Leishmania* parasite modulates host cell defense mechanism and exhibits altered expression of unfolded protein response genes. *Parasitol Res*. 2019 Sep;118(9):2705–13.
- Wang Z, Zhu W-G, Xu X. Ubiquitin-like modifications in the DNA damage response. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2017 Oct 1;803–805:56–75.
- WHO WHO. Leishmaniasis [Internet]. WHO. 2015a [cited 2016 Jan 28]. Available from: <http://www.who.int/leishmaniasis/en/>
- WHO WHO. Leishmaniasis [Internet]. WHO. 2015b [cited 2016 Jan 28]. Available from: <http://www.who.int/leishmaniasis/en/>
- WHO WHO. Leishmaniasis [Internet]. WHO. 2015c [cited 2016 Jan 28]. Available from: <http://www.who.int/leishmaniasis/en/>
- Woo HG, Choi J-H, Yoon S, Jee BA, Cho EJ, Lee J-H, et al. Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer. *Nature Communications*. 2017 Oct 10;8(1):839.
- Wu P-Y, Phan JH, Wang MD. The Effect of Human Genome Annotation Complexity on RNA-Seq Gene Expression Quantification. *IEEE Int Conf Bioinform Biomed Workshops*. 2012 Oct;2012:712–7.
- Yadav S, Prakash J, Shukla H, Das KC, Tripathi T, Dubey VK. Design of a multi-epitope subunit vaccine for immune-protection against *Leishmania* parasite. *Pathogens and Global Health*. 2020 Nov 16;114(8):471–81.

- Yang F, Ouma WZ, Li W, Doseff AI, Grotewold E. Establishing the Architecture of Plant Gene Regulatory Networks. *Methods Enzymol.* 2016;576:251–304.
- Young JC, Agashe VR, Siegers K, Hartl FU. Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol.* 2004 Oct;5(10):781–91.
- Yu D, Lim J, Wang X, Liang F, Xiao G. Enhanced construction of gene regulatory networks using hub gene information. *BMC Bioinformatics.* 2017 Mar 23;18(1):186.
- Zhang GL, Sun J, Chitkushev L, Brusic V. Big Data Analytics in Immunology: A Knowledge-Based Approach. *Biomed Res Int [Internet].* 2014 [cited 2019 Aug 30];2014. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4090507/>
- Zhang J, He J, Li J, Zhou Q, Chen H, Zheng Z, et al. The immunogenicity and protective immunity of multi-epitopes DNA prime-protein boost vaccines encoding Amastin-Kmp-11, Kmp11-Gp63 and Amastin-Gp63 against visceral leishmaniasis. *PLOS ONE.* 2020 Mar 16;15(3):e0230381.
- Zhang K, Beverley SM. Phospholipid and sphingolipid metabolism in Leishmania. *Molecular and Biochemical Parasitology.* 2010 Apr 1;170(2):55–64.
- Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cellular & Molecular Immunology.* 2018 Feb;15(2):182–4.
- Zhang W-W, Lypaczewski P, Matlashewski G. Optimized CRISPR-Cas9 Genome Editing for Leishmania and Its Use To Target a Multigene Family, Induce Chromosomal Translocation, and Study DNA Break Repair Mechanisms. *mSphere [Internet].* 2017 Jan 18 [cited 2019 Aug 27];2(1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5244264/>
- Zhu B, Song N, Shen R, Arora A, Machiela MJ, Song L, et al. Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers. *Scientific Reports.* 2017 Dec 5;7(1):16954.

Apêndices

Apêndice A - Métricas relacionadas à análise de qualidade das amostras sequenciadas.

Identificador da corrida no SRA	Identificador da amostra	Tratamento/tempo de infecção	Total de leituras	Cobertura
SRR2156854	HPGL0441	<i>Macrófagos que fagocitaram beads de latex, 24 hpi</i>	27543191	95,91
SRR2163256	HPGL0464	<i>Macrófagos que fagocitaram beads de latex, 24 hpi</i>	49212698	171,38
SRR2163286	HPGL0502	<i>Macrófagos que fagocitaram beads de latex, 24 hpi</i>	49261429	171,54
SRR2156272	HPGL0438	<i>Macrófagos que fagocitaram beads de latex, 4 hpi</i>	32435553	112,95
SRR2163239	HPGL0460	<i>Macrófagos que fagocitaram beads de latex, 4 hpi</i>	49261383	171,54
SRR2163281	HPGL0498	<i>Macrófagos que fagocitaram beads de latex, 4 hpi</i>	48013008	167,20
SRR2156858	HPGL0444	<i>Macrófagos que fagocitaram beads de latex, 48 hpi</i>	47775253	166,37
SRR2163274	HPGL0468	<i>Macrófagos que fagocitaram beads de latex, 48 hpi</i>	48695862	169,58
SRR2163291	HPGL0506	<i>Macrófagos que fagocitaram beads de latex, 48 hpi</i>	53716605	187,06
SRR2156862	HPGL0452	<i>Macrófagos que fagocitaram beads de latex, 72 hpi</i>	43300491	150,79
SRR2163401	HPGL0510	<i>Macrófagos que fagocitaram beads de latex, 72 hpi</i>	29794211	103,75
SRR2156274	HPGL0440	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 24 hpi</i>	27427610	95,51
SRR2163242	HPGL0462	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 24 hpi</i>	44837168	156,14
SRR2163283	HPGL0500	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 24 hpi</i>	45280151	157,68
SRR2156271	HPGL0437	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 4 hpi</i>	49018481	170,70
SRR2163237	HPGL0458	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 4 hpi</i>	49928664	173,87
SRR2163279	HPGL0496	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 4 hpi</i>	43965989	153,10

SRR2156856	HPGL0443	Macrófagos infectados com <i>Leishmania amazonensis</i> (PH8 strain), 48 hpi	42539912	148,14
SRR2163272	HPGL0466	Macrófagos infectados com <i>Leishmania amazonensis</i> (PH8 strain), 48 hpi	49364066	171,90
SRR2163289	HPGL0504	Macrófagos infectados com <i>Leishmania amazonensis</i> (PH8 strain), 48 hpi	32097879	111,78
SRR2156860	HPGL0446	Macrófagos infectados com <i>Leishmania amazonensis</i> (PH8 strain), 72 hpi	39776226	138,51
SRR2163276	HPGL0470	Macrófagos infectados com <i>Leishmania amazonensis</i> (PH8 strain), 72 hpi	44208239	153,95
SRR2163399	HPGL0508	Macrófagos infectados com <i>Leishmania amazonensis</i> (PH8 strain), 72 hpi	25380222	88,38
SRR2155074	HPGL0368	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 24 hpi	36216045	126,12
SRR2155160	HPGL0378	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 24 hpi	38286499	133,33
SRR2156107	HPGL0401	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 24 hpi	42658371	148,55
SRR2163251	HPGL0463	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 24 hpi	44582958	155,25
SRR2163285	HPGL0501	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 24 hpi	40598925	141,38
SRR2155072	HPGL0366	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi	43011845	149,78
SRR2155105	HPGL0376	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi	41952510	146,09
SRR2155166	HPGL0399	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi	45233522	157,52
SRR2163238	HPGL0459	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi	43124122	150,17
SRR2163280	HPGL0497	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi	43348262	150,95
SRR2155078	HPGL0370	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 48 hpi	39275069	136,77
SRR2155162	HPGL0380	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 48 hpi	37570407	130,83

		<i>hpi</i>		
		<i>Macrófagos infectados com Leishmania major (Friedlin strain), 48 hpi</i>		
SRR2156109	HPGL0403		49012593	170,68
		<i>Macrófagos infectados com Leishmania major (Friedlin strain), 48 hpi</i>		
SRR2163273	HPGL0467		41250707	143,65
		<i>Macrófagos infectados com Leishmania major (Friedlin strain), 48 hpi</i>		
SRR2163290	HPGL0505		46518427	161,99
		<i>Macrófagos infectados com Leishmania major (Friedlin strain), 72 hpi</i>		
SRR2155085	HPGL0372		33976311	118,32
		<i>Macrófagos infectados com Leishmania major (Friedlin strain), 72 hpi</i>		
SRR2155164	HPGL0382		35690452	124,29
		<i>Macrófagos infectados com Leishmania major (Friedlin strain), 72 hpi</i>		
SRR2156117	HPGL0405		44500276	154,96
		<i>Macrófagos infectados com Leishmania major (Friedlin strain), 72 hpi</i>		
SRR2163400	HPGL0509		30950266	107,78
SRR2155073	HPGL0367	<i>Macrófagos não infectados, 24 hpi</i>	42757070	148,89
SRR2155143	HPGL0377	<i>Macrófagos não infectados, 24 hpi</i>	38326882	133,47
SRR2156106	HPGL0400	<i>Macrófagos não infectados, 24 hpi</i>	49186866	171,29
SRR2156273	HPGL0439	<i>Macrófagos não infectados, 24 hpi</i>	25646097	89,31
SRR2163240	HPGL0461	<i>Macrófagos não infectados, 24 hpi</i>	51880109	180,66
SRR2163282	HPGL0499	<i>Macrófagos não infectados, 24 hpi</i>	48593257	169,22
SRR2155070	HPGL0365	<i>Macrófagos não infectados, 4 hpi</i>	45694346	159,12
SRR2155101	HPGL0375	<i>Macrófagos não infectados, 4 hpi</i>	39235553	136,63
SRR2155165	HPGL0398	<i>Macrófagos não infectados, 4 hpi</i>	50097600	174,46
SRR2156270	HPGL0436	<i>Macrófagos não infectados, 4 hpi</i>	43839665	152,66
SRR2156863	HPGL0457	<i>Macrófagos não infectados, 4 hpi</i>	53886954	187,65
SRR2163278	HPGL0495	<i>Macrófagos não infectados, 4 hpi</i>	50563242	176,08
SRR2155075	HPGL0369	<i>Macrófagos não infectados, 48 hpi</i>	38915463	135,52
SRR2155161	HPGL0379	<i>Macrófagos não infectados, 48 hpi</i>	38141815	132,82
SRR2156108	HPGL0402	<i>Macrófagos não infectados, 48 hpi</i>	43319195	150,85
SRR2156855	HPGL0442	<i>Macrófagos não infectados, 48 hpi</i>	39727322	138,34
SRR2163270	HPGL0465	<i>Macrófagos não infectados, 48 hpi</i>	49694900	173,05
SRR2163287	HPGL0503	<i>Macrófagos não infectados, 48 hpi</i>	47397331	165,05
SRR2155082	HPGL0371	<i>Macrófagos não infectados, 72 hpi</i>	39992869	139,27
SRR2155163	HPGL0381	<i>Macrófagos não infectados, 72 hpi</i>	36802757	128,16
SRR2156110	HPGL0404	<i>Macrófagos não infectados, 72 hpi</i>	52188462	181,74
SRR2156859	HPGL0445	<i>Macrófagos não infectados, 72 hpi</i>	35745917	124,48
SRR2163275	HPGL0469	<i>Macrófagos não infectados, 72 hpi</i>	51249010	178,47
SRR2163292	HPGL0507	<i>Macrófagos não infectados, 72 hpi</i>	27788324	96,77

Apêndice B - Genes regulados positivamente e enriquecidos para a categoria do *Gene Ontology* de Processo Biológico.

Identificador	Descrição (Função predita)	Fold Change	p-valor
<i>Protein phosphorylation</i>			
LINF_320013700	<i>CYC2-like cyclin - putative*</i>	5,56	3,34E-42
LINF_340027100	<i>Dual specificity protein phosphatase - putative</i>	8,93	4,34E-107
LINF_360054700	<i>Elongation factor-2 kinase-like protein</i>	3,23	2,10E-16
LINF_350046100	<i>Kinetoplastid kinetochore protein 3 - putative</i>	2,02	9,69E-13
LINF_250012200	<i>Myosin heavy chain kinase c-like protein</i>	3,37	3,77E-20
LINF_300023300	<i>Phosphatidylinositol kinase - putative*</i>	3,97	4,10E-42
LINF_360034100	<i>Protein kinase - putative*</i>	3,88	5,44E-42
LINF_010011100	<i>Rab-GTPase-TBC domain containing protein</i>	2,13	1,17E-16
LINF_340047100	<i>Target of rapamycin kinase 3*</i>	2,74	2,33E-21
<i>Microtubule-based movement</i>			
LINF_070010300	<i>Dynein heavy chain*</i>	5,90	6,51E-54
LINF_130011900	<i>Kinesin - putative*</i>	5,87	1,11E-70
LINF_330038800	<i>Present in the outer mitochondrial membrane</i>	4,06	1,74E-37
LINF_260024800	<i>Microtubule-associated protein - putative</i>	3,73	1,67E-59
LINF_270017700	<i>Intraflagellar transport protein 88 - putative</i>	2,16	2,53E-19
LINF_040010500	<i>Tetratricopeptide repeat</i>	2,14	2,06E-19
LINF_210017400	<i>WD40 repeat-containing protein</i>	2,06	4,35E-21
<i>Biosynthetic process</i>			
LINF_350008300	<i>Serine palmitoyltransferase - putative</i>	2,99	2,23E-35
LINF_170011100	<i>Phenazine biosynthesis-like protein</i>	2,15	3,76E-08
<i>Cellular process</i>			
LINF_360020600	<i>N-terminal region of Chorein</i>	4,95	3,27E-59
LINF_190015000	<i>FYVE zinc finger containing protein</i>	3,66	2,95E-27
LINF_270009800	<i>Glycosomal transporter (GAT3) - putative</i>	3,39	1,48E-39
LINF_240020050	<i>Multi drug resistance protein-like</i>	3,27	1,11E-33
LINF_350025400	<i>Ankyrin repeat protein - putative</i>	2,67	1,11E-21
<i>Cellular component organization</i>			
LINF_130020500	<i>Phosphoprotein phosphatase - putative</i>	2,28	1,71E-11
<i>Protein ubiquitination</i>			
LINF_350059500	<i>SPRY domain; HECT-domain (ubiquitin-transferase)*</i>	9,14	4,18E-81
LINF_360053000	<i>Ubiquitin protein ligase - putative*</i>	7,30	3,59E-111
LINF_220017000	<i>Zinc finger - C3HC4 type (RING finger)</i>	3,77	2,20E-40
LINF_110017600	<i>Protein transport protein SEC31 - putative</i>	3,57	6,80E-54

LINF_160017100	<i>WW domain containing protein</i>	2,63	5,06E-15
LINF_160018000	<i>Cullin family</i>	2,23	1,06E-15
Ribonucleoprotein complex assembly			
LINF_170016300	<i>ATP-dependent RNA helicase - putative</i>	2,38	1,49E-26
Stress granule assembly			
LINF_180019600	<i>Pumilio protein 2 — putative</i>	5,59	6,98E-64
Cellular metabolic process			
LINF_270018400	<i>RING-H2 zinc finger</i>	3,33	9,20E-24
LINF_290018600	<i>Heat shock protein 100 kDa</i>	2,86	5,67E-32
Primary metabolic process; Cellular macromolecule biosynthetic process; Cellular nitrogen compound metabolic process			
LINF_300022800	<i>DNAJ domain protein - putative</i>	2,13	9,30E-14
Primary metabolic process; Macromolecule metabolic process; Nitrogen compound metabolic process			
LINF_120008300	<i>Myotubularin-related protein - putative</i>	4,00	4,38E-52
Multi-organism cellular process			
LINF_310038500	<i>Acetyl-CoA carboxylase - putative</i>	2,74	5,28E-33
Regulation of membrane lipid distribution; Phospholipid translocation			
LINF_340032300	<i>Phospholipid-transporting ATPase-like protein*</i>	3,69	2,87E-39
Regulation of membrane lipid distribution			
LINF_260032600	<i>ATP-binding cassette protein subfamily B - member 2 - putative</i>	2,21	7,28E-13
Quorum sensing involved in interaction with host; Multi-organism cellular process; Modulation of symbiont involved in interaction with host			
LINF_330022900	<i>RNA recognition motif (RRM/RBD or RNP domain)*</i>	14,4	1,07E-153
LINF_330023000	<i>RNA-binding protein - putative</i>	4,01	1,25E-12

* Genes com mais de uma cópia.

Apêndice C - Transcritos regulados negativamente e enriquecidos para a categoria do Gene Ontology de Processo Biológico.

Identificador	Descrição (Função predita)	Fold Change	p-valor
Nucleosome assembly			
LINF_100016800	<i>histone H3 — putative</i>	2.54	5.91E-28
LINF_170019500	<i>histone H2B*</i>	2.34	4.29E-09
LINF_310040900	<i>histone H4</i>	2.15	6.51E-18
Ribonucleoprotein complex subunit organization; Ribosome biogenesis			
LINF_110013300	<i>40S ribosomal protein S21 - putative*</i>	2.77	2.71E-31
LINF_070010600	<i>60S ribosomal protein L7a - putative*</i>	2.41	3.03E-17
LINF_010009200	<i>ribosomal protein S7 - putative</i>	2.36	9.19E-23
LINF_360008600	<i>nuclear protein family a - putative</i>	2.11	4.18E-08
Translation			
LINF_290032300	<i>60S ribosomal protein L13 - putative*</i>	2.19	2.28E-16
LINF_130017200	<i>40S ribosomal protein S4 - putative</i>	2.02	8.88E-14
Ribosome biogenesis			
LINF_060009400	<i>Ribosomal protein L19e - putative*</i>	3.14	1.54E-35
LINF_350023700	<i>60S ribosomal protein L5 - putative</i>	3.00	9.31E-33
LINF_360015100	<i>40S ribosomal protein S18 - putative</i>	2.35	9.75E-17
Ribonucleoprotein complex subunit organization			
LINF_210018200	<i>40S ribosomal protein S23 - putative*</i>	2.62	1.59E-24
LINF_030007300	<i>ribosomal protein L38 - putative</i>	2.24	1.49E-22
Ribonucleoprotein complex subunit organization; rRNA processing; Ribosome biogenesis			
LINF_260021300	<i>40S ribosomal protein S33 - putative</i>	2.66	1.95E-25
LINF_340050700	<i>nucleolar protein family a - putative</i>	2.10	2.15E-09
Cytoplasmic translation			
LINF_330028300	<i>60S ribosomal protein L37*</i>	2.27	7.94E-20
LINF_350009300	<i>40S ribosomal protein S3A - putative</i>	2.00	3.20E-12

* Genes com mais de uma cópia

Apêndice D - Transcritos sem enriquecimento funcional para a categoria do Gene Ontology de Processo Biológico.

Identificador	Descrição (Função predita)	Fold Change	p-valor
Regulados positivamente			
LINF_190016700	<i>Ankyrin repeat - putative</i>	4.37	1.47E-42
LINF_310029900	<i>Armadillo-like helical domain</i>	2.15	4.19E-21
LINF_060012700	<i>C2 domain/Ankyrin repeats</i>	2.57	2.01E-17
LINF_270021600	<i>Cyclic nucleotide-binding domain containing protein</i>	2.18	1.00E-13
LINF_260029300	<i>Fibronectin type III domain containing protein</i>	6.23	4.99E-79
LINF_230012700	<i>Galactose oxidase</i>	5.69	1.34E-78
LINF_250017200	<i>Galactose oxidase</i>	5.96	3.96E-64
LINF_300016400	<i>IQ calmodulin-binding motif containing protein</i>	3.28	5.47E-27
LINF_150013800	<i>Kelch motif/Galactose oxidase</i>	2.01	5.93E-13
LINF_050010400	<i>kinetoplast-associated protein-like protein</i>	2.15	1.28E-18
LINF_200012200	<i>Midasin - putative</i>	13.72	2.03E-215
LINF_090016200	<i>N-terminal region of Chorein</i>	4.52	1.74E-56
LINF_330026100	<i>OTT 1508-like deaminase - putative</i>	2.69	6.35E-08
LINF_090012200	<i>PH domain containing protein - putative Pfam:PF00169</i>	2.31	7.47E-15
LINF_210019600	<i>phosphatidylinositol 3-kinase - putative</i>	2.8	1.42E-15
LINF_360052600	<i>predicted C2 domain protein</i>	2.28	1.09E-17
LINF_200019800	<i>protein kinase - putative</i>	2.68	4.01E-12
LINF_210023500	<i>Raptor N-terminal CASPase like domain containing protein - putative Pfam:PF14538</i>	2.4	2.71E-13
Regulação negativa			
LINF_330036400	<i>Complex 1 protein (LYR family) - putative Pfam:PF05347</i>	2.47	1.15E-14
LINF_310018400	<i>Cytochrome b5-like Heme/Steroid binding domain containing protein - putative Pfam:PF00173</i>	2.15	1.95E-10
LINF_220021500	<i>Dephospho-CoA kinase - putative</i>	2.06	1.71E-08
LINF_270022300	<i>DNA-directed RNA polymerase II-like protein</i>	2.08	6.66E-08
LINF_360008300	<i>Dolicholphosphate-mannose synthase - putative</i>	2.05	8.03E-07
LINF_050005600	<i>Dynein light chain - putative</i>	2.11	1.27E-09
LINF_280028000	<i>Dynein light chain LC6 - flagellar outer arm - putative</i>	2.06	2.86E-08
LINF_340014000	<i>Elongation factor 1-beta</i>	2.01	1.22E-06
LINF_250012900	<i>Eukaryotic initiation factor 5a - putative</i>	3.23	2.07E-27
LINF_010010200	<i>Fatty acyl CoA synthetase 2 - putative</i>	2.15	4.19E-06
LINF_210016700	<i>Histone H2A - putative</i>	2.56	6.01E-23
LINF_300024900	<i>Intraflagellar transport (IFT) protein - putative</i>	2	4.91E-08
LINF_350027300	<i>Kinetoplastid membrane protein-11</i>	2.27	1.26E-03
LINF_040012700	<i>Nascent polypeptide associated complex subunit-like protein - copy</i>	2.36	4.00E-18
LINF_040012800	<i>Nascent polypeptide associated complex subunit-like protein - copy 2</i>	2.06	2.46E-10

LINF_230026500	<i>Nucleoside 2-deoxyribosyltransferase - putative Pfam:PF05014</i>	2.04	2.81E-15
LINF_320036800	<i>Nucleoside diphosphate kinase b</i>	2.5	1.33E-22
LINF_360067000	<i>Phosphoprotein lepp12</i>	2.07	8.78E-08
LINF_250014000	<i>Prefoldin subunit - putative Pfam:PF01920</i>	2.38	3.98E-21
LINF_230012800	<i>Putative tRNA binding domain containing protein - putative Pfam:PF01588</i>	2.04	6.81E-12
LINF_320034100	<i>Ribosomal protein L27 - putative</i>	2.41	9.80E-28
LINF_360042600	<i>Ribosomal protein L29 - putative</i>	2.09	1.19E-14
LINF_230018650	<i>Small hydrophilic endoplasmic reticulum-associated protein (sherp)</i>	2.79	1.18E-02
LINF_260021900	<i>Sphingolipid delta 4 desaturase - putative</i>	2.16	9.53E-05
LINF_300032800	<i>Translation machinery associated TMA7 - putative Pfam:PF09072</i>	2.2	5.21E-11
LINF_240020600	<i>Translationally controlled tumor protein (TCTP) - putative</i>	2.2	1.70E-02
LINF_310033900	<i>Ubiquinol-cytochrome-c reductase-like protein</i>	2.29	2.64E-18
LINF_310028300	<i>Ubiquitin-fusion protein</i>	2.5	1.94E-26
LINF_310026800	<i>Ubiquitin-fusion protein</i>	2.56	4.15E-28
LINF_350049500	<i>Zinc-binding domain containing protein - putative Pfam:PF13695</i>	2.15	1.33E-11

Apêndice E - Métricas relacionadas ao mapeamento das amostras sequenciadas.

Identificador da corrida no SRA	Tratamento/tempo de infecção	Total de leituras mapeadas	Total de leituras não mapeadas	% de leituras mapeadas
SRR2156274	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 24 hpi</i>	23770461	3657149	86.7%
SRR2163242	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 24 hpi</i>	31736385	13100783	70.8%
SRR2163283	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 24 hpi</i>	39298923	5981228	86.8%
SRR2156271	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 4 hpi</i>	41225111	7793370	84.1%
SRR2163237	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 4 hpi</i>	37884293	12044371	75.9%
SRR2163279	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 4 hpi</i>	37321001	6644988	84.9%
SRR2156856	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 48 hpi</i>	36879540	5660372	86.7%
SRR2163272	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 48 hpi</i>	36259760	13104306	73.5%
SRR2163289	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 48 hpi</i>	28316723	3781156	88.2%
SRR2156860	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 72 hpi</i>	33446009	6330217	84.1%
SRR2163276	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 72 hpi</i>	32095100	12113139	72.6%
SRR2163399	<i>Macrófagos infectados com Leishmania amazonensis (PH8 strain), 72 hpi</i>	21967917	3412305	86.6%
SRR2155074	<i>Macrófagos infectados com Leishmania major (Friedlin strain), 24 hpi</i>	30232226	5983819	83.5%
SRR2155160	<i>Macrófagos infectados com Leishmania major (Friedlin strain), 24 hpi</i>	28323212	9963287	74.0%
SRR2156107	<i>Macrófagos infectados com Leishmania major (Friedlin strain), 24 hpi</i>	36294164	6364207	85.1%
SRR2163251	<i>Macrófagos infectados com</i>	34692898	9890060	77.8%

	<i>Leishmania major</i> (Friedlin strain), 24 hpi			
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 24 hpi			90.3%
SRR2163285		36658050	3940875	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi			81.8%
SRR2155072		35182002	7829843	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi			81.1%
SRR2155105		34040224	7912286	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi			87.5%
SRR2155166		39589448	5644074	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi			80.5%
SRR2163238		34700121	8424001	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 4 hpi			89.1%
SRR2163280		38626653	4721609	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 48 hpi			83.9%
SRR2155078		32971244	6303825	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 48 hpi			82.5%
SRR2155162		30988110	6582297	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 48 hpi			87.1%
SRR2156109		42707089	6305504	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 48 hpi			82.7%
SRR2163273		34134129	7116578	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 48 hpi			89.1%
SRR2163290		41468342	5050085	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 72 hpi			86.3%
SRR2155085		29322869	4653442	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 72 hpi			84.9%
SRR2155164		30283453	5406999	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 72 hpi			83.3%
SRR2156117		37063059	7437217	
	Macrófagos infectados com <i>Leishmania major</i> (Friedlin strain), 72 hpi			90.1%
SRR2163400		27875905	3074361	
	Macrófagos não infectados, 24 hpi			92.1%
SRR2155073		39373512	3383558	
	Macrófagos não infectados, 24 hpi			90.6%
SRR2155143		34711939	3614943	
	Macrófagos não infectados, 24 hpi			88.8%
SRR2156106		43675028	5511838	

SRR2156273	<i>Macrófagos não infectados, 24 hpi</i>	22590407	3055690	88.1%
SRR2163240	<i>Macrófagos não infectados, 24 hpi</i>	45999398	5880711	88.7%
SRR2163282	<i>Macrófagos não infectados, 24 hpi</i>	42873353	5719904	88.2%
SRR2155070	<i>Macrófagos não infectados, 4 hpi</i>	40523682	5170664	88.7%
SRR2155101	<i>Macrófagos não infectados, 4 hpi</i>	35400828	3834725	90.2%
SRR2155165	<i>Macrófagos não infectados, 4 hpi</i>	45573814	4523786	91.0%
SRR2156270	<i>Macrófagos não infectados, 4 hpi</i>	40188424	3651241	91.7%
SRR2156863	<i>Macrófagos não infectados, 4 hpi</i>	48631101	5255853	90.2%
SRR2163278	<i>Macrófagos não infectados, 4 hpi</i>	45333865	5229377	89.7%
SRR2155075	<i>Macrófagos não infectados, 48 hpi</i>	35258619	3656844	90.6%
SRR2155161	<i>Macrófagos não infectados, 48 hpi</i>	35210289	2931526	92.3%
SRR2156108	<i>Macrófagos não infectados, 48 hpi</i>	38154149	5165046	88.1%
SRR2156855	<i>Macrófagos não infectados, 48 hpi</i>	35221378	4505944	88.7%
SRR2163270	<i>Macrófagos não infectados, 48 hpi</i>	43806613	5888287	88.2%
SRR2163287	<i>Macrófagos não infectados, 48 hpi</i>	41156682	6240649	86.8%
SRR2155082	<i>Macrófagos não infectados, 72 hpi</i>	34862515	5130354	87.2%
SRR2155163	<i>Macrófagos não infectados, 72 hpi</i>	33442513	3360244	90.9%
SRR2156110	<i>Macrófagos não infectados, 72 hpi</i>	45988506	6199956	88.1%
SRR2156859	<i>Macrófagos não infectados, 72 hpi</i>	31338984	4406933	87.7%
SRR2163275	<i>Macrófagos não infectados, 72 hpi</i>	45757150	5491860	89.3%
SRR2163292	<i>Macrófagos não infectados, 72 hpi</i>	24910282	2878042	89.6%
SRR2156854	<i>Macrófagos que fagocitaram beads de latex, 24 hpi</i>	24506163	3037028	89.0%
SRR2163256	<i>Macrófagos que fagocitaram beads de latex, 24 hpi</i>	44241787	4970911	89.9%
SRR2163286	<i>Macrófagos que fagocitaram beads de latex, 24 hpi</i>	43919361	5342068	89.2%
SRR2156272	<i>Macrófagos que fagocitaram beads de latex, 4 hpi</i>	29119421	3316132	89.8%

SRR2163239	<i>Macrófagos que fagocitaram beads de latex, 4 hpi</i>	44056349	5205034	89.4%
SRR2163281	<i>Macrófagos que fagocitaram beads de latex, 4 hpi</i>	43923366	4089642	91.5%
SRR2156858	<i>Macrófagos que fagocitaram beads de latex, 48 hpi</i>	42614313	5160940	89.2%
SRR2163274	<i>Macrófagos que fagocitaram beads de latex, 48 hpi</i>	43891946	4803916	90.1%
SRR2163291	<i>Macrófagos que fagocitaram beads de latex, 48 hpi</i>	43292336	10424269	80.6%
SRR2156862	<i>Macrófagos que fagocitaram beads de latex, 72 hpi</i>	38359537	4940954	88.6%
SRR2163401	<i>Macrófagos que fagocitaram beads de latex, 72 hpi</i>	27070569	2723642	90.9%

Apêndice F - Termos do *Gene Ontology* para o conjunto de genes DE em macrófagos.

Termo	Descrição	Frequência
GO:0002200	<i>somatic diversification of immune receptors</i>	0.01%
GO:0008152	<i>metabolic process</i>	63.54%
GO:0009411	<i>response to UV</i>	0.05%
GO:0045588	<i>positive regulation of gamma-delta T cell differentiation</i>	0.00%
GO:0051301	<i>cell division</i>	1.49%
GO:1990166	<i>protein localization to site of double-strand break</i>	0.00%
GO:0006471	<i>protein ADP-ribosylation</i>	0.02%
GO:0044770	<i>cell cycle phase transition</i>	0.06%
GO:0007049	<i>cell cycle</i>	1.77%
GO:0006996	<i>organelle organization</i>	4.21%
GO:0006807	<i>nitrogen compound metabolic process</i>	46.32%
GO:0006260	<i>DNA replication</i>	1.56%
GO:0051345	<i>positive regulation of hydrolase activity</i>	0.40%
GO:0046483	<i>heterocycle metabolic process</i>	22.18%
GO:0030592	<i>DNA ADP-ribosylation</i>	0.00%
GO:1904262	<i>negative regulation of TORC1 signaling</i>	0.01%
GO:0032465	<i>regulation of cytokinesis</i>	0.04%
GO:0000018	<i>regulation of DNA recombination</i>	0.05%
GO:1902115	<i>regulation of organelle assembly</i>	0.07%
GO:0034641	<i>cellular nitrogen compound metabolic process</i>	26.81%
GO:0043170	<i>macromolecule metabolic process</i>	35.08%
GO:0051726	<i>regulation of cell cycle</i>	0.60%
GO:0090305	<i>nucleic acid phosphodiester bond hydrolysis</i>	2.71%
GO:0044238	<i>primary metabolic process</i>	50.80%
GO:0006479	<i>protein methylation</i>	0.39%
GO:0010506	<i>regulation of autophagy</i>	0.07%
GO:0043412	<i>macromolecule modification</i>	16.14%
GO:0071704	<i>organic substance metabolic process</i>	56.06%
GO:0044237	<i>cellular metabolic process</i>	56.28%
GO:0008213	<i>protein alkylation</i>	0.39%
GO:0018205	<i>peptidyl-lysine modification</i>	0.43%
GO:0006974	<i>cellular response to DNA damage stimulus</i>	2.58%
GO:0070647	<i>protein modification by small protein conjugation or removal</i>	1.14%
GO:0006357	<i>regulation of transcription by RNA polymerase II</i>	1.69%
GO:0019222	<i>regulation of metabolic process</i>	13.19%
GO:0006261	<i>DNA-dependent DNA replication</i>	0.57%
GO:0018193	<i>peptidyl-amino acid modification</i>	2.77%
GO:0006310	<i>DNA recombination</i>	1.76%

GO:0006139	<i>nucleobase-containing compound metabolic process</i>	19.25%
GO:0044260	<i>cellular macromolecule metabolic process</i>	30.23%
GO:0071539	<i>protein localization to centrosome</i>	0.01%
GO:0044267	<i>cellular protein metabolic process</i>	19.87%
GO:0032200	<i>telomere organization</i>	0.13%
GO:0033683	<i>nucleotide-excision repair, DNA incision</i>	0.01%
GO:0015031	<i>protein transport</i>	2.84%
GO:0006259	<i>DNA metabolic process</i>	5.49%
GO:0030031	<i>cell projection assembly</i>	0.33%
GO:0080135	<i>regulation of cellular response to stress</i>	0.18%
GO:0000723	<i>telomere maintenance</i>	0.13%
GO:0006296	<i>nucleotide-excision repair, DNA incision, 5'-to lesion</i>	0.00%
GO:0007098	<i>centrosome cycle</i>	0.03%
GO:0000278	<i>mitotic cell cycle</i>	0.40%
GO:0000724	<i>double-strand break repair via homologous recombination</i>	0.12%
GO:0090304	<i>nucleic acid metabolic process</i>	13.98%
GO:0006355	<i>regulation of transcription, DNA-templated</i>	9.96%
GO:0032392	<i>DNA geometric change</i>	0.70%
GO:0000725	<i>recombinational repair</i>	0.16%
GO:0006298	<i>mismatch repair</i>	0.17%
GO:1902017	<i>regulation of cilium assembly</i>	0.02%
GO:0006302	<i>double-strand break repair</i>	0.21%
GO:0051276	<i>chromosome organization</i>	2.08%
GO:0006270	<i>DNA replication initiation</i>	0.13%

Apêndice G - Termos do Gene Ontology para o conjunto de transcritos regulados positivamente em *L. amazonensis*.

Termo	Descrição	Frequência
GO:0006264	<i>mitochondrial DNA replication</i>	0.00%
GO:0032042	<i>mitochondrial DNA metabolic process</i>	0.01%
GO:0000002	<i>mitochondrial genome maintenance</i>	0.04%
GO:0006490	<i>oligosaccharide-lipid intermediate biosynthetic process</i>	0.06%
GO:0140029	<i>exocytic process</i>	0.07%
GO:0048278	<i>vesicle docking</i>	0.07%
GO:0022406	<i>membrane docking</i>	0.10%
GO:0006821	<i>chloride transport</i>	0.17%
GO:0006816	<i>calcium ion transport</i>	0.21%
GO:0016579	<i>protein deubiquitination</i>	0.24%
GO:0001522	<i>pseudouridine synthesis</i>	0.36%
GO:0007005	<i>mitochondrion organization</i>	0.36%
GO:0006873	<i>cellular ion homeostasis</i>	0.37%
GO:1902531	<i>regulation of intracellular signal transduction</i>	0.45%
GO:0006486	<i>protein glycosylation</i>	0.51%
GO:0009100	<i>glycoprotein metabolic process</i>	0.56%
GO:0070085	<i>glycosylation</i>	0.66%
GO:0015698	<i>inorganic anion transport</i>	0.90%
GO:0010646	<i>regulation of cell communication</i>	0.94%
GO:0023051	<i>regulation of signaling</i>	0.94%
GO:0006364	<i>rRNA processing</i>	1.09%
GO:0070647	<i>protein modification by small protein conjugation or removal</i>	1.14%
GO:0048583	<i>regulation of response to stimulus</i>	1.15%
GO:0016192	<i>vesicle-mediated transport</i>	1.34%
GO:0006886	<i>intracellular protein transport</i>	1.37%
GO:0030001	<i>metal ion transport</i>	1.49%
GO:0009451	<i>RNA modification</i>	1.96%
GO:0051641	<i>cellular localization</i>	2.35%
GO:0044255	<i>cellular lipid metabolic process</i>	3.28%
GO:0006812	<i>cation transport</i>	3.57%
GO:0033036	<i>macromolecule localization</i>	3.64%
GO:0006396	<i>RNA processing</i>	3.77%
GO:0006629	<i>lipid metabolic process</i>	4.27%
GO:0071705	<i>nitrogen compound transport</i>	4.66%
GO:0006259	<i>DNA metabolic process</i>	5.49%
GO:0006508	<i>proteolysis</i>	5.70%
GO:0016070	<i>RNA metabolic process</i>	8.11%
GO:0006811	<i>ion transport</i>	10.22%
GO:0090304	<i>nucleic acid metabolic process</i>	13.98%

GO:0006810	<i>transport</i>	18.12%
GO:0051179	<i>localization</i>	18.99%
GO:0006139	<i>nucleobase-containing compound metabolic process</i>	19.25%
GO:0046483	<i>heterocycle metabolic process</i>	22.18%
GO:0006725	<i>cellular aromatic compound metabolic process</i>	22.22%
GO:1901360	<i>organic cyclic compound metabolic process</i>	22.95%

Apêndice H - Apêndice G - Termos do Gene Ontology para o conjunto de transcritos regulados negativamente em *L. amazonensis*.

Termo	Descrição	Frequência
GO:0001539	<i>cilium or flagellum-dependent cell motility</i>	0.40%
GO:0023052	<i>signaling</i>	7.02%
GO:0040011	<i>locomotion</i>	1.01%
GO:0065007	<i>biological regulation</i>	22.83%
GO:0030031	<i>cell projection assembly</i>	0.33%
GO:0006928	<i>movement of cell or subcellular component</i>	1.00%
GO:0007017	<i>microtubule-based process</i>	0.70%
GO:0006468	<i>protein phosphorylation</i>	4.25%
GO:0035556	<i>intracellular signal transduction</i>	3.86%
GO:0007154	<i>cell communication</i>	7.56%
GO:0046034	<i>ATP metabolic process</i>	1.35%
GO:1901568	<i>fatty acid derivative metabolic process</i>	0.03%
GO:0017144	<i>drug metabolic process</i>	0.13%
GO:0072330	<i>monocarboxylic acid biosynthetic process</i>	1.07%
GO:0044281	<i>small molecule metabolic process</i>	16.28%
GO:0006091	<i>generation of precursor metabolites and energy</i>	3.61%
GO:0016052	<i>carbohydrate catabolic process</i>	1.38%
GO:0006793	<i>phosphorus metabolic process</i>	14.80%
GO:0005975	<i>carbohydrate metabolic process</i>	6.09%
GO:0006452	<i>translational frameshifting</i>	0.00%
GO:0042593	<i>glucose homeostasis</i>	0.06%
GO:0072521	<i>purine-containing compound metabolic process</i>	2.77%
GO:0043243	<i>positive regulation of protein-containing complex disassembly</i>	0.02%
GO:0046950	<i>cellular ketone body metabolic process</i>	0.01%
GO:1901264	<i>carbohydrate derivative transport</i>	0.26%
GO:0051674	<i>localization of cell</i>	0.62%
GO:1901137	<i>carbohydrate derivative biosynthetic process</i>	4.82%
GO:0017013	<i>protein flavinylation</i>	0.03%
GO:0010608	<i>posttranscriptional regulation of gene expression</i>	1.51%
GO:0009190	<i>cyclic nucleotide biosynthetic process</i>	0.14%
GO:0042737	<i>drug catabolic process</i>	0.01%
GO:0055086	<i>nucleobase-containing small molecule metabolic process</i>	5.31%
GO:1901642	<i>nucleoside transmembrane transport</i>	0.07%
GO:0046952	<i>ketone body catabolic process</i>	0.01%
GO:0072522	<i>purine-containing compound biosynthetic process</i>	1.82%
GO:0006850	<i>mitochondrial pyruvate transmembrane transport</i>	0.02%
GO:1904659	<i>glucose transmembrane transport</i>	0.02%
GO:0030030	<i>cell projection organization</i>	0.67%
GO:0072525	<i>pyridine-containing compound biosynthetic process</i>	0.47%

GO:0009187	<i>cyclic nucleotide metabolic process</i>	0.16%
GO:0006751	<i>glutathione catabolic process</i>	0.06%
GO:0019673	<i>GDP-mannose metabolic process</i>	0.07%
GO:0050789	<i>regulation of biological process</i>	21.13%
GO:0007099	<i>centriole replication</i>	0.01%
GO:0042866	<i>pyruvate biosynthetic process</i>	0.01%
GO:0042759	<i>long-chain fatty acid biosynthetic process</i>	0.01%
GO:0043171	<i>peptide catabolic process</i>	0.08%
GO:0044273	<i>sulfur compound catabolic process</i>	0.07%
GO:0030497	<i>fatty acid elongation</i>	0.02%
GO:0009166	<i>nucleotide catabolic process</i>	0.21%
GO:0098656	<i>anion transmembrane transport</i>	2.55%
GO:0032787	<i>monocarboxylic acid metabolic process</i>	2.84%
GO:0044283	<i>small molecule biosynthetic process</i>	7.03%
GO:0033559	<i>unsaturated fatty acid metabolic process</i>	0.04%
GO:0046939	<i>nucleotide phosphorylation</i>	0.63%
GO:0006636	<i>unsaturated fatty acid biosynthetic process</i>	0.03%
GO:0046434	<i>organophosphate catabolic process</i>	0.52%
GO:0001676	<i>long-chain fatty acid metabolic process</i>	0.06%
GO:0016310	<i>phosphorylation</i>	8.29%
GO:0006083	<i>acetate metabolic process</i>	0.08%
GO:0010468	<i>regulation of gene expression</i>	11.92%
GO:0019222	<i>regulation of metabolic process</i>	13.19%
GO:0006464	<i>cellular protein modification process</i>	13.93%
GO:0070925	<i>organelle assembly</i>	0.63%
GO:0009165	<i>nucleotide biosynthetic process</i>	3.08%
GO:0033500	<i>carbohydrate homeostasis</i>	0.06%
GO:0034219	<i>carbohydrate transmembrane transport</i>	0.30%
GO:0071616	<i>acyl-CoA biosynthetic process</i>	0.21%

Apêndice I - Média dos valores de leituras mapeadas para cada fator de transcrição em cada um dos *time-block* no modelo da rede na via de sinalização da IL-17.

Fator de transcrição	Macrófagos não infectados	Macrófagos infectados por <i>L. major</i>	Macrófagos infectados por <i>L. amazonensis</i>
STAT1	0-25,5877.61; 26-50,6754.95; 51-75,4595.94; 76-100,4764.34	0-25,5372.16; 26-50,5297.05; 51-75,7787.30; 76-100,6128.42	0-25,4952.21; 26-50,9402.04; 51-75,7038.73; 76-100,2346.24
STAT4	0-25,103.36; 26-50,16.21; 51-75,9.46; 76-100,10.44	0-25,316.41; 26-50,22.79; 51-75,9.43; 76-100,7.51	0-25,219.78; 26-50,56.48; 51-75,28.13; 76-100,9.38
STAT5A	0-25,1377.02; 26-50,1454.16; 51-75,1545.59; 76-100,1581.53	0-25,1127.25; 26-50,990.51; 51-75,1210.43; 76-100,1422.85	0-25,887.90; 26-50,1214.90; 51-75,1210.77; 76-100,403.59
STAT5B	0-25,524.35; 26-50,532.86; 51-75,592.85; 76-100,618.74	0-25,407.32; 26-50,319.05; 51-75,464.02; 76-100,502.93	0-25,429.26; 26-50,405.40; 51-75,472.34; 76-100,157.45
TRAF4	0-25,43.62; 26-50,25.99; 51-75,30.51; 76-100,27.78	0-25,54.67; 26-50,28.18; 51-75,63.88; 76-100,48.00	0-25,55.00; 26-50,48.88; 51-75,55.27; 76-100,18.42
TRAF6	0-25,442.61; 26-50,359.55; 51-75,382.51; 76-100,404.66	0-25,654.83; 26-50,268.30; 51-75,338.58; 76-100,414.34	0-25,652.17; 26-50,374.58; 51-75,368.61; 76-100,122.87
NFKB1	0-25,1184.41; 26-50,681.48; 51-75,436.75; 76-100,439.57	0-25,5017.29; 26-50,694.29; 51-75,580.69; 76-100,460.40	0-25,4894.69; 26-50,1270.40; 51-75,695.09; 76-100,231.70
TRAF3	0-25,1635.59; 26-50,1600.42; 51-75,1291.95; 76-100,1372.07	0-25,2203.29; 26-50,1746.89; 51-75,1465.59; 76-100,1247.54	0-25,2435.34; 26-50,1915.48; 51-75,1686.70; 76-100,1487.19
TRAF5	0-25,561.71; 26-50,508.41; 51-75,618.41; 76-100,642.19	0-25,1832.44; 26-50,401.60; 51-75,594.85; 76-100,634.43	0-25,1460.42; 26-50,436.62; 51-75, 606.95; 76-100,631.53
NFKBIZ	0-25,280.17; 26-50,421.33; 51-75,208.93; 76-100,179.65	0-25,661.13; 26-50,408.78; 51-75,241.42; 76-100,173.16	0-25,880.30; 26-50,677.47; 51-75, 381.54; 76-100,231.80
CEBPD	0-25,422.55;	0-25,184.11;	0-25,163.55;

	26-50,157.70; 51-75,166.76; 76-100,145.51	26-50,51.71; 51-75,73.86; 76-100,94.86	26-50,47.14; 51-75, 51.91; 76-100,41.46
CEBPB	0-25,4759.21; 26-50,2663.82; 51-75,1904.76; 76-100,1737.65	0-25,10652.21; 26-50,5337.31; 51-75,4190.46; 76-100,3264.60	0-25,11066.34; 26-50,5726.38; 51-75,4000.16; 76-100,3344.22
AP-1	0-25,2401.65; 26-50,1359.71; 51-75,1260.27; 76-100,1272.11	0-25,8360.52; 26-50,2622.15; 51-75,1768.60; 76-100,2044.06	0-25,9936.31; 26-50,3633.69; 51-75,2268.36; 76-100,2052.19

Apêndice J - Epítomos preditos

ID	Função predita	Epítopo	Score do VaxiGen	% de Conservação
LAMA_000017700	<i>hypothetical protein, conserved</i>	FVFPASFTL*#	0.6046	0.01% (1/16169)
LAMA_000045000	<i>hypothetical protein, conserved</i>	YMVFASYYM*#	1.3945	0.01% (1/16169)
LAMA_000133200	<i>hypothetical protein, conserved</i>	YTFATTPPV*#	0.2191	0.01% (1/16169)
LAMA_000175700	<i>Amino acid permease, putative</i>	FHRCMYLSMATMSLI*#	0.6166	0.01% (2/16169)
LAMA_000229400	<i>hypothetical protein, conserved</i>	FLALAFLSAVRASAI	0.8823	0.01% (1/16169)
LAMA_000281600	<i>hypothetical protein, conserved</i>	YVMELFFTL*#	0.9154	0.01% (2/16169)
LAMA_000353900	<i>hypothetical protein, conserved</i>	YMMTAMPQM*#	0.4116	0.01% (1/16169)
LAMA_000362500	<i>Nuclease-related domain containing protein, putative</i>	FVAAYWHLL	1.1801	0.01% (1/16169)
LAMA_000366500	<i>hypothetical protein, conserved</i>	YLWYMFMPI*#	2.1203	0.01% (1/16169)
LAMA_000443100	<i>hypothetical protein, conserved</i>	YVYKIWVEL*#	1.629	0.01% (2/16169)
LAMA_000535100	<i>hypothetical protein, conserved</i>	FTMQSSFLL*#	0.3133	0.01% (2/16169)
LAMA_000598900	<i>hypothetical protein, conserved</i>	AHLRRYLAATAIASL*#	0.6002	0.01% (2/16169)

LAMA_000808000	<i>hypothetical protein, conserved</i>	FTYAAFTAL*#	1.0168	0.01% (1/16169)
LbrM.03.0720	<i>hypothetical protein (fragment)</i>	LQRLRIFLSAFAAGV	0.7802	0.01% (1/16169)
LbrM.08.1030	<i>amastin-like protein</i>	PGRLMLFRLAQALAV	0.1338	0.01% (1/16169)
LbrM.10.0630	<i>hypothetical protein, conserved</i>	YVMKAIVTL	0.6959	0.01% (2/16169)
LbrM.10.0630		FTFYFTTAV	0.2791	0.01% (1/16169)
LbrM.14.0490	<i>hypothetical protein, unknown function</i>	SSSFALSMAAFVLHA*#	0.8436	0.01% (2/16169)
LbrM.14.0500	<i>hypothetical protein, unknown function</i>	SSSFALSMAAFVLHA*#	0.8436	0.01% (2/16169)
LbrM.16.0140	<i>hypothetical protein</i>	QRWRIYTSTFARVAA	0.8957	0.00% (0/16169)
LbrM.16.0520	<i>hypothetical protein, unknown function</i>	TLLLSFAMASAVAV	0.7989	0.01% (1/16169)
LbrM.18.1730	<i>P-type H⁺-ATPase, putative</i>	MLSFLTYRISATLQL	0.594	0.02% (3/16169)
LbrM.18.1750	<i>hypothetical protein, conserved</i>	YVMELFFTL*#	0.9154	0.01% (2/16169)
LbrM.23.1510	<i>hypothetical protein, conserved</i>	EPLFYVIAAIRPLAA	0.5708	0.01% (2/16169)
LbrM.24.0880	<i>emp24/gp25L/p24 family/GOLD, putative</i>	QYIYETFASMRMALV	0.5954	0.01% (2/16169)
LbrM.24.1470	<i>hypothetical predicted transmembrane protein</i>	FVAPGGYSL	0.3406	0.01% (2/16169)

LbrM.26.2520	<i>Paraquat-inducible protein A, putative</i>	FMYATLLSI	0.9	0.01% (2/16169)
LbrM.34.4140	<i>hypothetical protein, unknown function</i>	FAYAAFTAL	0.8312	0.01% (1/16169)
LBRM2903_00000960 0	<i>P-type H⁺-ATPase, putative</i>	LSFLTYRISATLQLV	1.0183	0.02% (3/16169)
LBRM2903_02001110 0	<i>hypothetical protein, conserved</i>	FIYYYIFPV*#	1.7615	0.01% (1/16169)
LBRM2903_03001300 0	<i>hypothetical protein, conserved</i>	YLMAVAWAL	0.7802	0.01% (1/16169)
LBRM2903_14001010 0	<i>Amastin surface glycoprotein, putative</i>	LQRLRIFLSAFAAGV	0.8436	0.01% (1/16169)
LBRM2903_15001310 0	<i>hypothetical protein, conserved</i>	SSSFALSMAAFVLHA	0.0362	0.01% (2/16169)
LBRM2903_16001120 0	<i>Amastin surface glycoprotein, putative</i>	FLAFYMSSTLRRAS	0.7989	0.01% (2/16169)
LBRM2903_16002150 0	<i>hypothetical protein, conserved</i>	AHAFLSALRAAKLAR	0.6214	0.01% (2/16169)
LBRM2903_18000960 0	<i>Amastin surface glycoprotein, putative</i>	TLLLSFAMASAVAV	0.3511	0.01% (1/16169)
LBRM2903_18001470 0	<i>hypothetical protein, conserved</i>	EVTMLRYLQAAALYL	0.9428	0.01% (2/16169)
LBRM2903_20001850 0	<i>amastin-like surface protein, putative</i>	NRFRLAQAFALISIF	0.9154	0.04% (6/16169)
LBRM2903_20001850 0	<i>amastin-like surface protein, putative</i>	YVMELFFTL	0.0882	0.01% (2/16169)
LBRM2903_20001850 0	<i>amastin-like surface protein, putative</i>	GHLMLRYAAEAFAL	0.1344	0.01% (2/16169)

LBRM2903_20003540	<i>amastin-like surface protein, putative</i>	PRRLRLFRAAGVLAI	0.5708	0.02% (3/16169)
LBRM2903_23002570	<i>hypothetical protein, conserved</i>	EPLFYVIAAIRPLAA	0.5954	0.01% (2/16169)
LBRM2903_24001480	<i>emp24/gp25L/p24 family/GOLD, putative</i>	QYIYETFASMRMALV	0.3406	0.01% (2/16169)
LBRM2903_25001560	<i>hypothetical protein, conserved</i>	YQYSSLAAL	0.9	0.01% (1/16169)
LBRM2903_26003130	<i>Paraquat-inducible protein A, putative</i>	FMYATLLSI	1.0009	0.01% (2/16169)
LBRM2903_31000790		YQFTKTFSL	0.8312	0.01% (2/16169)
LBRM2903_31000790	<i>MFS/sugar transport protein, putative</i>	FMFTLCFTF	0.3129	0.01% (2/16169)
LBRM2903_35003890	<i>hypothetical protein, conserved</i>	FVFKSLYTL	0.4018	0.01% (1/16169)
LINF_040009300	<i>hypothetical protein - conserved</i>	YTYAWTSAF	0.1501	0.00% (0/16169)
LINF_050012400	<i>phosphatase-like protein</i>	PASFRFVRSRLARIVR	0.0136	0.00% (0/16169)
LINF_070014400	<i>hypothetical protein - conserved</i>	YVMGRHPL*#	0.2101	0.00% (0/16169)
LINF_080009500	<i>hypothetical protein - conserved</i>	TRLVASLRASPVALR	0.2191	0.01% (1/16169)
Sen_LINF_29.2679	<i>hypothetical protein - conserved</i>	YVMKASVTL*#	1.4765	0.01% (2/16169)

		YVFKHNYTL	0.3	0.01% (1/16169)
		FTNFTFWAL	0.0367	0.01% (1/16169)
LINF_110009700	<i>hypothetical protein</i>	YVAAAFLLPI	0.7898	0.00% (0/16169)
LINF_150014500	<i>hypothetical protein - conserved</i>	AHVFLSALRAAKLAR	0.7565	0.01% (2/16169)
LINF_210022000	<i>hypothetical protein - conserved</i>	LYFSSFVLSRANAAA	0.7758	0.01% (1/16169)
LINF_210024200	<i>hypothetical protein - conserved</i>	RYLIALAAAHVVVSA	0.4684	0.01% (2/16169)
LINF_230014100	<i>hypothetical protein - conserved</i>	YVYKIWLEL*#	1.6186	0.01% (1/16169)
LINF_240013900	<i>emp24/gp25L/p24 family/GOLD - putative</i>	YIYETFASM*#	0.4314	0.01% (2/16169)
LINF_270012200	<i>hypothetical protein - conserved</i>	FTMQSSFLL*#	0.3133	0.01% (2/16169)
LINF_280006200	<i>hypothetical protein - conserved</i>	VVMFLLSSLPMVIIE	0.6286	0.01% (1/16169)
LINF_300013600	<i>hypothetical protein - conserved</i>	FSVLWFSPAPSILFT	0.866	0.01% (1/16169)
		HLRRYLAATAIASLW	0.653	0.01% (2/16169)
Res_LINF_09.8745	<i>hypothetical protein - conserved</i>	AHLRRYLAATAIASL*#	0.6002	0.01% (2/16169)
LINF_300030400	<i>hypothetical protein - conserved</i>	FFYVNFLASTAAYWV	0.9262	0.01% (2/16169)

LINF_310021200	<i>hypothetical protein - conserved</i>	TLKVMRAFAAAAMPSL	0.0911	0.01% (1/16169)
LINF_310021400	<i>hypothetical protein - conserved</i>	NTLKVMRAFAAAAMPS	0.2213	0.01% (1/16169)
LINF_310021500	<i>hypothetical protein - conserved</i>	TLKVMRAFAAAAMPSL	0.0911	0.01% (1/16169)
LINF_320010700	<i>Sulfite exporter TauE/SafE - putative</i>	MATFFIAFSSALHLL	0.2466	0.01% (2/16169)
LINF_320010700		GMATFFIAFSSALHL	0.0783	0.01% (2/16169)
LINF_320024500	<i>Cation efflux family - putative</i>	ISTSFLIASMQALVL	0.6544	0.01% (2/16169)
LINF_320045500	<i>hypothetical protein - conserved</i>	YMFLFVYVL*#	1.5994	0.01% (2/16169)
LINF_330018800	<i>hypothetical protein - conserved</i>	FTYHAPTAL	1.1266	0.00% (0/16169)
LINF_360012100	<i>GPI transamidase component Tta2 - putative</i>	ALWYVRFIPAVLAKP*#	1.3118	0.01% (1/16169)
LINF_360031900	<i>hypothetical protein - conserved</i>	YIFENFHPM	0.4826	0.00% (0/16169)
LINF_360048100	<i>hypothetical protein - conserved</i>	IFLRSFLAAMGQPTL	0.8206	0.01% (1/16169)
LINF_360060000	<i>32 kDa ER-associated protein - putative</i>	FVFPASFTL*#	0.6046	0.01% (1/16169)
LmxM.03.0490	<i>hypothetical protein, conserved</i>	YMFVASYYM*#	1.3945	0.01% (1/16169)
LmxM.07.0830	<i>hypothetical protein, conserved</i>	YVMSGRHPL*#	0.2101	0.00% (0/16169)

LmxM.08.0440	<i>hypothetical protein, conserved</i>	YTFATTPPV*#	0.2191	0.01% (1/16169)
LmxM.08.0760	<i>amastin-like protein, putative</i>	RDNFRLAQAFIISI	0.1143	0.02% (3/16169)
LmxM.11.0520	<i>amino acid permease/transporter, putative</i>	FHRCLMYLSMATMSLI*#	0.6166	0.01% (2/16169)
LmxM.11.0660	<i>Protein Associated with Differentiation, putative</i>	NSSFLFVALQSAPVS	0.5228	0.01% (1/16169)
LmxM.15.0350	<i>hypothetical protein, conserved</i>	FLALAFLSAVRVSAI	0.7387	0.01% (1/16169)
LmxM.18.0840	<i>hypothetical protein, conserved</i>	YVMELFFTL*#	0.9154	0.01% (2/16169)
LmxM.24.0870	<i>emp24/gp25L/p24 family/GOLD, putative</i>	YIYETFASM*#	0.4314	0.01% (2/16169)
LmxM.25.0790	<i>hypothetical protein, unknown function</i>	FVFRMVSPI	0.5857	0.01% (1/16169)
LmxM.27.0620	<i>hypothetical protein, conserved</i>	FTMQSSFLL*#	0.3133	0.01% (2/16169)
LmxM.29.1260	<i>hypothetical protein, conserved</i>	AHLRRYLAATAIASL*#	0.6002	0.01% (2/16169)
LmxM.31.0540	<i>sulfite exporter TauE/SafE, putative</i>	ATFFIAFSSALQLLI	0.4511	0.01% (1/16169)
LmxM.33.1560a	<i>hypothetical protein</i>	VQHRFRAAQAFIIS	0.3243	0.01% (1/16169)
LmxM.34.4140	<i>hypothetical protein, unknown function</i>	FTYAAFTAL*#	1.0168	0.01% (1/16169)
LmxM.34.4350	<i>PGAP1-like protein, putative</i>	AFIFASVWASVVAMR	0.8975	0.01% (1/16169)

	<i>GPI transamidase component Tta2,</i>			0.01%
LmxM.36.0650	<i>putative</i>	ELALWYVRFIPAVLA*#	1.2915	(1/16169)
	<i>hypothetical protein, conserved</i>			0.01%
LmxM.36.2450		YIFETFHPI	0.0626	(1/16169)
	<i>hypothetical protein, conserved</i>			0.01%
LmxM.36.3920		IVFLRSFLAAMGQPT	0.8515	(1/16169)
	<i>hypothetical protein, conserved</i>			0.01%
LmxM.36.4680		YMMTAMPQM*#	0.4116	(1/16169)
	<i>hypothetical protein, conserved</i>			0.01%
LmxM.36.5050		FVFPASFTL*#	0.6046	(1/16169)
	<i>Flagellum attachment zone protein 5</i>			0.01%
LmxM.36.5970		YLWYMFMPI*#	2.1203	(1/16169)
