



**UNIVERSIDADE FEDERAL DA BAHIA**  
**INSTITUTO DE CIÊNCIAS DA SAÚDE**  
**PROGRAMA DE PÓS-GRADUAÇÃO MULTICÊNTRICO EM BIOQUÍMICA E**  
**BIOLOGIA MOLECULAR**

**THIAGO MENDONÇA DOS SANTOS**

**IDENTIFICAÇÃO DE BIOMARCADORES TRANSCRICIONAIS**  
**ASSOCIADOS ÀS PRINCIPAIS MANIFESTAÇÕES CLÍNICAS DA**  
**INFECÇÃO PELO HTLV-1**

Salvador

2020

**THIAGO MENDONÇA DOS SANTOS**

**IDENTIFICAÇÃO DE BIOMARCADORES TRANSCRICIONAIS  
ASSOCIADOS ÀS PRINCIPAIS MANIFESTAÇÕES CLÍNICAS DA  
INFECÇÃO PELO HTLV-1**

Dissertação apresentada ao Programa de Pós-Graduação Multicêntrico em Bioquímica e Biologia Molecular, Instituto de Ciências da Saúde, Universidade Federal da Bahia, como requisito para obtenção do grau de Mestre em Bioquímica e Biologia Molecular.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Aline Cristina Andrade Mota  
Miranda Mascarenhas

Salvador

2020

Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas (SIBI/UFBA),  
com os dados fornecidos pelo(a) autor(a).

**S237** Santos, Thiago  
IDENTIFICAÇÃO DE BIOMARCADORES TRANSCRICIONAIS  
ASSOCIADOS ÀS PRINCIPAIS MANIFESTAÇÕES CLÍNICAS DA  
INFECÇÃO PELO HTLV-1 / Thiago Santos. -- Salvador,  
2020.

88 f. : il

Orientadora: Aline Cristina Andrade Mota Miranda  
Mascarenhas.

Dissertação (Mestrado - Bioquímica e Biologia  
Molecular) -- Universidade Federal da Bahia, ICS,  
2020.

1. HTLV-1. 2. Expressão gênica. 3. Metanálise. 4.  
ATLL. 5. HAM/TSP. I. Andrade Mota Miranda  
Mascarenhas, Aline Cristina. II. Título.

**CDU 575.112**



Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq  
Universidade Federal da Bahia - UFBA  
Instituto de Ciências da Saúde - ICS

**PMBqBM**

Programa Multicêntrico de  
Pós-Graduação em Bioquímica e Biologia Molecular



## TERMO DE APROVAÇÃO DO BANCA DE DEFESA DE DISSERTAÇÃO

THIAGO MENDONÇA DOS SANTOS

### POTENCIAL BIOMARCADOR DE PROGRESSÃO PARA MANIFESTAÇÃO DE DOENÇA EM INDIVÍDUOS INFECTADOS PELO HTLV-1.

Salvador, Bahia, 27 de fevereiro de 2020.

#### COMISSÃO EXAMINADORA:

*Aline Cristina A. M. Miranda Mascarenhas*

Profa. Dra. Aline Cristina Andrade Mota Miranda Mascarenhas (Orientadora/UFBA)

(Presidente)

*Antonio Ricardo Khouri Cunha*

Prof. Dr. Antonio Ricardo Khouri Cunha

Instituto de Pesquisa Gonçalo Moniz, Laboratório de Enfermidades Infecciosas  
Transmitidas por Vetores (LEITV)

*Glória Regina Franco*

Profa. Dra. Glória Regina Franco (UFMG)

*Dedico esse trabalho aos pacientes infectados pelo HTLV-1,  
frequentemente esquecidos pelos poderes  
públicos e até mesmo pelas pesquisas.*

## AGRADECIMENTOS

A minha orientadora, Professora Aline, pela extraordinária orientação e valiosas sugestões, sempre tão didática e atenciosa. Sou extremamente grato pela oportunidade que tive de tê-la como orientadora, com certeza foi a experiência mais engrandecedora que tive na jornada da carreira acadêmica.

A Mauro, pelo apoio e incentivo nos momentos de desânimo. Sempre me mostrando que sou capaz, obrigado!

Aos bioinformatas com quem cruzei durante a pesquisa, em momentos de necessidade, especialmente Dr. Pablo Ivan Ramos pela gentil disposição no auxílio técnico. Felizmente muitos pesquisadores entendem que a Ciência não deve ser contida em seus laboratórios, mas sim compartilhada.

À minha família pelo apoio, sem eles nada disso seria possível.

À Dra. Luciane Amorim Santos pelo auxílio na elaboração do projeto.

Ao meu grupo de pesquisa, Jéssica Laís Almeida dos Santos, Jéssica Duarte Sousa, Cinthia Cristina Oliveira S. Costa e Beatriz Rolim Lobo pela oportunidade de estar sempre me aprimorando e agregando conhecimento em nossas sessões científicas.

Aos meus colegas de Mestrado Jéssica Teles e Victor, por estarem sempre solícitos quando em necessidade. A Jéssica, por ser meu modelo em alguns pontos (“como Jéssica faria?”). E a Victor, pelas reuniões que sempre clareavam dúvidas e me forneciam mais ideias.

Ao Programa de Pós-Graduação Multicêntrico em Bioquímica e Biologia Molecular, pelo constante interesse em pról da formação de seus alunos.

A histórica Universidade Federal da Bahia, que, diante de tantos ataques, ainda é a casa mãe de milhares de alunos e pesquisadores.

A CAPES, pelo financiamento pessoal durante todo o período do mestrado.

## RESUMO

O Vírus Linfotrófico de Células T do adulto 1 (HTLV-1) é um retrovírus humano sexualmente transmissível que apresenta tropismo preferencial por células T CD4<sup>+</sup>. Embora a maioria dos indivíduos infectados por esse vírus permaneça assintomática (ASS), as principais manifestações clínicas, como a Mielopatia Associada ao HTLV/Paraparesia Espástica Tropical (HAM/TSP) e a Leucemia/Linfoma de Células T do Adulto (ATLL), são de difícil prognóstico. A HAM/TSP é uma manifestação inflamatória do sistema nervoso central que pode culminar com a perda parcial dos movimentos dos membros inferiores ao passo que a ATLL é um tipo de linfoma não-Hodgkin que geralmente leva à morte. Este trabalho tem como objetivo propor um perfil transcricional diferencial, sugestivo das patologias associadas ao HTLV-1, além de investigar as implicações causadas pela desregulação na expressão gênica. Uma metanálise de dez dados de microarranjos disponíveis publicamente foi realizada para identificar genes diferencialmente expressos (GDEs). Foram realizadas análises de vias KEGG enriquecidas. Redes de interação proteína-proteína (IPP), extração de módulos e seleção de genes hubs foram implementados com STRING, MCODE e CytoHubba. Foram identificados ao total 907 GDEs para ATLL, 368 para HAM/TSP e 150 para ASS. A análise KEGG identificou "Vias do câncer" e "Endocitose" como vias enriquecidas para ATLL no conjunto de GDEs superexpresso e subexpresso respectivamente. Não foram obtidas vias enriquecidas significativas para o conjunto total de GDEs de HAM/TSP e ASS. A partir das redes IPP geradas pelo STRING, foram extraídos três módulos para cada uma das manifestações clínicas. Combinando os resultados do MCODE e CytoHubba, cinco genes hub foram identificados para cada condição. Somente ATLL apresentou micro-RNAs diferencialmente expressos, o hsa-mir-21, diretamente envolvido no desenvolvimento da doença e o hsa-mir-6840, ainda pouco conhecido. Este estudo gerou um banco de dados de marcadores genéticos candidatos e vias enriquecidas desreguladas para os status clínicos associados à infecção pelo HTLV-1, o que pode facilitar a definição e a compreensão de biomarcadores terapêuticos prognósticos.

Palavras-chave: HTLV-1, Expressão gênica, Metanálise, ATLL, HAM/TSP

## ABSTRACT

Adult T-cell Lymphotropic Virus 1 (HTLV-1) is a sexually transmitted human retrovirus that exhibits preferential CD4<sup>+</sup> T-cell tropism. Although most individuals infected with this virus remain asymptomatic (ASS), the main clinical manifestations such as HTLV-Associated Myelopathy/Tropical Spastic Paraparesis (HAM/TSP) and Adult T-Cell Leukemia/Lymphoma (ATLL) are troublesome conditions. HAM/TSP is an inflammatory manifestation of the central nervous system that may culminate in partial loss of lower limbs movements. ATLL is a type of non-Hodgkin's lymphoma that usually leads to death. This work aims to propose differential transcriptional profile, suggestive of the HTLV-1-associated pathologies, and to investigate the implications of deregulation on gene expression. A meta-analysis of ten publicly available microarray datasets was performed to identify differentially expressed genes (DEGs). Enriched KEGG pathways analysis was performed. Protein-protein interaction (PPI) networks, module extraction and gene hubs selection were implemented with STRING, MCODE and CytoHubba. A total of 907, 368 and 150 DEGs were identified for ATLL, HAM/TSP, and ASS respectively. The KEGG analysis identified "Cancer Pathways" and "Endocytosis" as ATLL-enriched pathways in the set of upregulated and downregulated DEGs respectively. No enriched pathways were identified for the full set of HAM/TSP and ASS DEGs. From the PPI networks generated by STRING, three modules were extracted for each clinical manifestation. Combining the results from MCODE and CytoHubba, five hub genes selected for each condition. Only ATLL presented differentially expressed micro-RNAs: hsa-mir-21, directly involved in the development of the disease and hsa-mir-6840, which is still poorly known. This study generated a database of candidate genetic markers and enriched pathways for the clinical status associated with HTLV-1 infection, facilitating the definition and understanding of prognostic therapeutic biomarkers.

Keywords: HTLV-1, Gene Expression, Meta-analysis, ATLL, HAM/TSP



## LISTA DE ILUSTRAÇÕES

|   |    |
|---|----|
| <b>Figura 1</b> - Representação da estrutura do HTLV-1.....   | 13 |
| <b>Figura 2</b> - Esquema do genoma do HTLV-1.....  | 14 |
| <b>Figura 3</b> - Epidemiologia do HTLV-1 ao redor do mundo .....   | 15 |
| <b>Figura 4</b> - Fluxograma da metodologia utilizada.....  | 29 |
| <b>Quadro 1</b> - Relação dos Ensaios de expressão gênica de pacientes infectados pelo HTLV-1, selecionados no GEO, para inclusão nas metanálises.....  | 33 |
| <b>Figura 5</b> - Quantidade de genes diferencialmente expressos (GDE) nos três status clínicos associados à infecção pelo HTLV-1.....  | 35 |
| <b>Figura 6</b> - Sobreposição dos perfis de genes desregulados nos três status clínicos associados à infecção pelo HTLV-1. (a) Conjunto total de GDEs, (b) GDEs superexpressos, (c) GDEs subexpressos..... | 37 |
| <b>Figura 7</b> - Rede de Interação Proteína-Proteína de GDEs em indivíduos com ATLL.....   | 40 |
| <b>Figura 8</b> - Módulos extraídos da rede de Interação Proteína-Proteína de ATLL.....   | 41 |
| <b>Figura 9</b> - Rede dos top cinco genes hubs de ATLL.....  | 43 |
| <b>Figura 10</b> - Rede de Interação Proteína-Proteína de GDEs em HAM/TSP.....  | 43 |
| <b>Figura 11</b> - Módulos da rede Interação Proteína-Proteína de HAM/TSP.....  | 44 |
| <b>Figura 12</b> - Rede dos top cinco genes hubs em indivíduos HAM/TSP.....   | 45 |
| <b>Figura 13</b> - Rede de Interação Proteína-Proteína de GDEs em ASS.....  | 46 |
| <b>Figura 14</b> - Módulos da rede de Interação Proteína-Proteína de ASS.....   | 46 |
| <b>Figura 15</b> - Rede dos top cinco genes hubs de indivíduos ASS.....   | 48 |
| <b>Quadro 2</b> - Caracterização das moléculas de miRNAs diferencialmente expressos em indivíduos com ATLL.....   | 49 |
| <b>Quadro 3</b> - GDEs considerados possíveis genes alvo de MDEs em ATLL.....   | 49 |

## LISTA DE TABELAS

|  |    |
|--|----|
| <b>Tabela 1</b> - Composição dos estudos selecionados para as metanálises quanto aos diferentes status clínicos associados à infecção pelo HTLV-1..... | 34 |
| <b>Tabela 2</b> - Destaque para os cinco genes mais superexpressos e subexpressos nos três status clínicos associados à infecção pelo HTLV-1.....      | 36 |
| <b>Tabela 3</b> - Cálculos de Descoberta e Perda nas metanálises dos status clínicos (HAM/TSP, ATLL e ASS) associados à infecção pelo HTLV-1.....      | 38 |
| <b>Tabela 4</b> - Vias enriquecidas para os GDEs em ATLL.....  | 39 |
| <b>Tabela 5</b> - Vias KEGG enriquecidas nos módulos da rede IPP de ATLL.....  | 42 |
| <b>Tabela 6</b> - Vias KEGG enriquecidas nos módulos da rede IPP de HAM/TSP.....   | 44 |
| <b>Tabela 7</b> - Vias KEGG enriquecidas nos módulos da rede IPP de ASS.....   | 47 |
| <b>Tabela 8</b> - Sondas referentes a miRNAs presentes nas plataformas dos estudos utilizados na metanálise.....                                       | 48 |

## LISTA DE ABREVIATURAS E SIGLAS

|         |   |
|---------|---|
| ANOVA   | Análise de variância  |
| ATLL    | <i>Adult T-cell Leukemia/Lymphoma</i> – Leucemia/Linfoma de Células T do Adulto   |
| AMPc    | <i>Cyclic Adenosine Monophosphate</i> – Adenosina Monofosfato Cíclico   |
| AZT     | Azidotimidina ou zidovulina   |
| CCR     | <i>Chemokine Receptors</i> – Receptores de quimiocinas  |
| cDNA    | DNA complementar  |
| DE      | Diferencialmente Expresso   |
| DNA     | <i>Deoxyribonucleic acid</i> - Ácido desoxirribonucleico  |
| EI      | Estudos Individuais   |
| FC      | <i>Fold Change</i>  |
| FDR     | <i>False discovery rate</i> – Taxa de falsa descoberta  |
| GDE     | Genes diferencialmente expressos  |
| GEO     | <i>Gene Expression Omnibus</i>  |
| HAM/TSP | <i>HTLV-1 Associated Myelopathy/Tropical Spastic Paraparesis</i> – Mielopatia associada ao HTLV/ Paraparesia Espástica Tropical |
| HBZ     | <i>HTLV-1 Basic Leucine Zipper</i> – Zíper de Leucina Básico do HTLV-1  |
| HIV     | <i>Human Immunodeficiency Virus</i> – Vírus da Imunodeficiência Humana  |
| HLA     | <i>Human Leukocyte Antigen</i> – Antígenos leucocitários humanos  |
| HSPG    | Proteoglicanos Heparan Sulfato  |
| HSV-2   | Herpes Simplex tipo 2   |
| HTLV-1  | <i>Human T-cell Lymphotropic Virus Type 1</i> – Vírus linfotrópico de células T humanas do tipo 1                               |
| IDD     | <i>Integration driven discovery</i> – Descoberta dirigida pela integração   |
| IDR     | <i>Integration driven rate</i> – Taxa de descoberta pela integração   |
| IFN-g   | Interferon gama   |
| IL-1    | Interleucina - 1  |
| IL-6    | Interleucina - 6  |
| IPP     | Interação Proteína-Proteína   |
| IRR     | <i>Integration Resolution Rate</i> – Taxa de resolução pela integração  |

## LISTA DE ABREVIATURAS E SIGLAS

|                |  |
|----------------|--|
| JAK-STAT       | <i>Janus Kinases-Signal Transducer and Activator of Transcription proteins</i><br>Transdutor do Quinase-Sinal de Janus e Ativador da Transcrição |
| KEGG           | <i>Kyoto Encyclopedia of Genes and Genomes</i> –<br>Enciclopédia de genes e genomas de Quioto  |
| LDH            | Lactato desidrogenase  |
| LMA            | Leucemia Mielóide Aguda  |
| lncRNA         | <i>Long non-coding RNA</i> – RNA longo não codificante   |
| LTR            | <i>Long Terminal Repeat</i> – Repetição Terminal Longa   |
| MA             | Metanálise   |
| MCC            | <i>Maximal Clique Centrality</i> – Centralidade Máxima de Clique   |
| MCODE          | <i>Molecular Complex Detection</i> – Detecção de complexos moleculares   |
| MDE            | Micro-RNA Diferencialmente Expresso  |
| MIAME          | <i>Minimum Information About a Microarray Experiment</i> –<br>Informação mínima sobre um experimento de microarray                               |
| miRNA          | Micro-RNA  |
| NCBI           | <i>National Center for Biotechnology Information</i> –<br>Centro nacional para informação de Biotecnologia                                       |
| NF- $\kappa$ B | <i>Factor nuclear kappa B</i> – Fator nuclear Kappa B  |
| NGS            | <i>Next-generation sequencing</i> – Sequenciamento de nova geração   |
| NIH            | <i>National Human Research Institute</i> – Instituto Nacional de Pesquisa Humana   |
| ORF            | <i>Open Reading Frame</i> – Fase de leitura aberta   |
| PBMC           | <i>Peripheral Blood Mononuclear Cells</i> –<br>Células mononucleares do sangue periférico  |
| PCA            | <i>Principal Component Analysis</i> –<br>Análise dos componentes principais  |
| PKC            | <i>Protein Kinase C</i> - Proteína quinase C   |
| PTB            | <i>Polipirimidin Binding Protein</i> - Proteína de Ligação a Polipirimidina  |
| RISC           | <i>RNA-induced silencing complex</i><br>Complexo de Silenciamento Induzido por RNA   |

## LISTA DE ABREVIATURAS E SIGLAS

|        |   |
|--------|---|
| RNA    | <i>Ribonucleic acid</i> – Ácido ribonucleico  |
| SCF    | SKP1-culina-F-box   |
| SD     | <i>Standard Deviation</i> – Desvio Padrão   |
| STRING | <i>Search Tool for Retrieval of Interacting Genes/Proteins</i> –<br>Ferramenta de busca para recuperação de genes / proteínas<br>em interação |
| SU     | Superfície Celular  |
| TCR    | <i>T-Cell Receptor</i> – Receptor de Células T  |
| TM     | Transmembrana   |
| TNF-a  | <i>Tumor Necrosis Factor-alpha</i> - Fator de necrose tumoral alfa  |
| UTR    | <i>Untranslated Region</i> – Região não traduzida   |

## SUMÁRIO

|  |    |
|--|----|
| 1.INTRODUÇÃO .....   | 14 |
| 1.1 O HTLV-1 .....   | 14 |
| 1.1.1 Aspectos estruturais e genômicos.....  | 15 |
| 1.1.2 Epidemiologia.....   | 18 |
| 1.2 MANIFESTAÇÕES CLÍNICAS ASSOCIADAS AO HTLV-1 .....                                    | 19 |
| 1.2.1 ATLL.....  | 19 |
| 1.2.2 HAM/TSP .....  | 21 |
| 1.3 MANIFESTAÇÃO DE DOENÇA E BIOMARCADORES.....  | 22 |
| 1.4 MICRO-RNA .....  | 25 |
| 1.5 JUSTIFICATIVA.....   | 26 |
| 2.OBJETIVOS .....  | 28 |
| 2.1 OBJETIVO GERAL .....   | 28 |
| 3.METODOLOGIA.....   | 29 |
| 3.1 BUSCA DE DADOS DE EXPRESSÃO GÊNICA: .....  | 29 |
| 3.2 METANÁLISES: .....   | 30 |
| 3.3 ANÁLISE DE VIAS ENRIQUECIDAS:.....   | 31 |
| 3.4 ANÁLISE DE REDE DE INTERAÇÃO PROTEÍNA-PROTEÍNA, SELEÇÃO DE MÓDULOS E GENES HUB:..... | 31 |
| 3.5 ANÁLISE DE MICRO-RNAS: .....   | 32 |
| 4.RESULTADOS .....   | 33 |
| 4.1 BUSCA POR ENSAIOS DE EXPRESSÃO NO GEO.....   | 33 |
| 4.2 METANÁLISES .....  | 34 |
| 4.3 ANÁLISE DE VIAS ENRIQUECIDAS.....  | 38 |
| 4.4. INTERAÇÃO PROTEÍNA-PROTEÍNA, SELEÇÃO DE MÓDULOS E GENES HUBS.....                   | 39 |
| 4.4.1 ATLL.....  | 40 |

|  |    |
|--|----|
| 4.4.2 HAM/TSP .....  | 43 |
| 4.5.3 ASS .....  | 45 |
| 4.6 ANÁLISE DE MICRO-RNAS .....  | 48 |
| 5.DISSCUSSÃO .....   | 50 |
| 5.1 O ARTIFÍCIO DA METANÁLISE .....  | 51 |
| 5.2 A DESREGULAÇÃO DIFERENCIAL NOS STATUS CLÍNICOS ASSOCIADOS AO HTLV-1..... | 52 |
| 5.4 VIAS ENRIQUECIDAS NO CONJUNTO TOTAL DE GDEs.....                         | 54 |
| 5.5 REDES IPP REVELAM MÓDULOS ENRIQUECIDOS PARA VIAS BIOQUÍMICAS .....       | 56 |
| 5.5.1 ATLL.....  | 57 |
| 5.2.2 HAM/TSP .....  | 59 |
| 5.2.3 ASS .....  | 61 |
| 5.3 MICRO-RNAS DESREGULADOS FAVORECEM A ATLL.....                            | 62 |
| 6.CONCLUSÃO.....   | 64 |
| REFERÊNCIAS BIBLIOGRÁFICAS .....   | 66 |
| APÊNDICE A – PCAs DOS ESTUDOS INDIVIDUAIS .....                              | 84 |
| APÊNDICE B – GDES COMUNS ENTRE OS STATUS CLÍNICOS.....                       | 86 |
| APÊNDICE C – GENES HUBS E SEUS SCORES .....                                  | 87 |

## 1. INTRODUÇÃO

### 1.1 O HTLV-1

O Vírus Linfotrófico de Células T Humanas do tipo 1 (HTLV-1), isolado pela primeira vez no Japão em 1979, a partir de amostras de um paciente com linfoma cutâneo de células T, foi o primeiro retrovírus humano descrito (POIESZ et al, 1980). Pertence ao gênero *Deltaretrovirus* e à família *Retroviridae*, a qual também pertence o Vírus da Imunodeficiência Humana (HIV, do inglês: *Human Immunodeficiency Virus*), e ambos os vírus, acredita-se, tenham emergido em populações humanas após transmissão zoonótica de população símia (VANDAMME & SALEMI, 1998; KEELE et al., 2006). Posteriormente três novos vírus foram identificados: HTLV-2, HTLV-3 e HTLV-4. Entretanto, não foram correlacionados a patologias humanas (PAIVA & CASSEB, 2014).

Apesar do HTLV-1 conseguir infectar células como monócitos, células dendríticas e linfócitos T CD8<sup>+</sup>, possui tropismo preferencial por células T CD4<sup>+</sup> (ECHEVARRIA-LIMA et al, 2018). O processo de ligação e entrada do HTLV-1 na célula alvo envolve alguns receptores celulares identificados até então, o Transportador de Glicose 1 (GLUT1), Neuropilina-1 (NRP-1) e Proteoglicanos Heparan Sulfato (HSPG), que interagem com proteínas do envelope viral (GHEZ et al., 2010).

Diferentemente da maioria dos outros vírus, o HTLV-1 é transmitido principalmente célula-a-célula por um fenômeno chamado sinapse viral que compreende o contato de junções estreitas e condúites celulares de uma célula infectada para uma nova célula hospedeira, limitando assim a exposição do vírus a mecanismos de defesa celulares (SATTENTAU, 2008). A transmissão por vírus livre é ineficiente e, portanto, não é encontrada a presença de RNA viral no plasma sanguíneo de indivíduos infectados (DEMONTIS, et al. 2015; GROSS & THOMAKRESS, 2016). O provírus do HTLV-1 permanece integrado ao DNA do hospedeiro ao longo de sua vida e sua multiplicação depende da divisão mitótica da célula infectada (ASQUITH & BANGHAM, 2007). O aspecto da clonalidade, a qual acontece pela proliferação de células T infectadas, é outra característica da infecção pelo HTLV-1 que sustenta a persistência desse vírus no paciente (BANGHAM, 2018).

A transmissão desse vírus pode ocorrer por via sexual, através do contato sexual desprotegido, por via vertical, através da placenta de uma mãe infectada ou amamentação para o bebê e por via parenteral, por transfusão sanguínea/compartilhamento de seringas (LAIRMORE et al., 2012). A transmissão

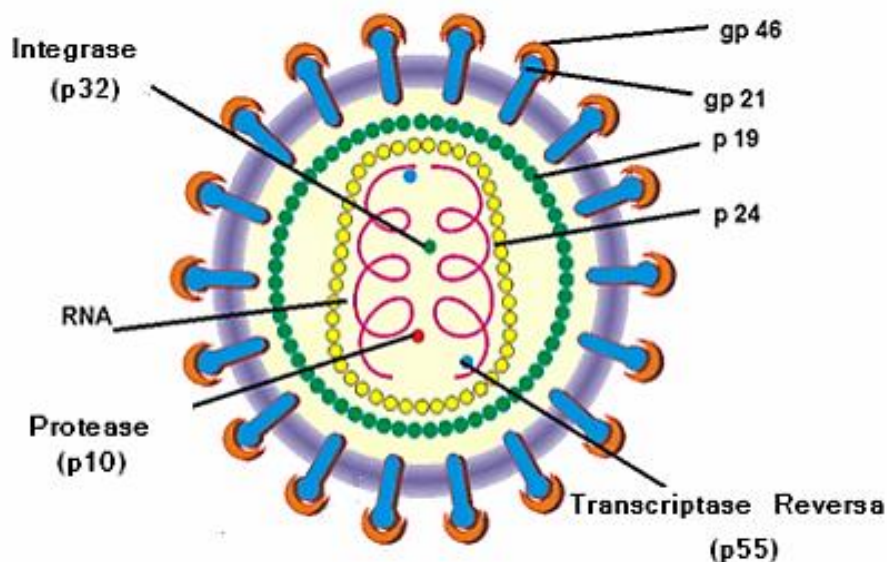


sexual é mais eficiente de homens para mulheres do que de mulheres para homens e pode ser agravada por doenças sexualmente transmissíveis ulcerativas, que causam rupturas na pele e mucosas, como a sífilis e a herpes simplex tipo 2 (HSV-2) (PAIVA & CASSEB, 2014). A forma mais eficiente de transmissão é via exposição intravenosa ao sangue infectado, o que ocorria em grande frequência no passado quando o sangue doado não era testado para HTLV-1. A eficiência da via de transmissão através da amamentação, por sua vez, é por volta de 20% (GONÇALVES et al., 2010). Além disso, a rota da infecção já foi relacionada com a manifestação patológica do HTLV-1. A HAM/TSP (Mielopatia Associada ao HTLV/Paraparesia Espástica Tropical), por exemplo, já foi associada com a via de transmissão sanguínea enquanto que a ATLL (Leucemia/Linfoma de células T do Adulto) com a via de amamentação (FUJINO & NAGATA, 2000; OSAME et al., 1990).

#### 1.1.1 Aspectos estruturais e genômicos

O HTLV-1 é um vírus constituído de um envelope, matriz proteica e nucleocapsídeo, apresentando uma morfologia esférica a pleomórfica medindo de 80 a 100nm de diâmetro (**Fig. 1**). A proteína de superfície (gp46, SU) e proteína transmembrana (gp21, TM), originárias do gene *env*, compõem o envelope viral, exibindo junto à membrana do envelope a proteína da matriz (p19). O nucleocapsídeo, formado pela proteína p24, abriga em seu interior o genoma viral composto por duas fitas simples de RNA de polaridade positiva, às quais estão associadas a proteínas básicas, chamadas de proteínas do nucleocapsídeo além da Protease (p10), responsável pela clivagem dos produtos dos genes *Gag* e *Pol*, e das proteínas Integrase (p32) e Transcriptase Reversa (p55), indispensáveis no processo de integração do DNA viral no genoma do hospedeiro (GREEN & CHEN, 2001).

**Figura 1** - Representação da estrutura do HTLV-1.



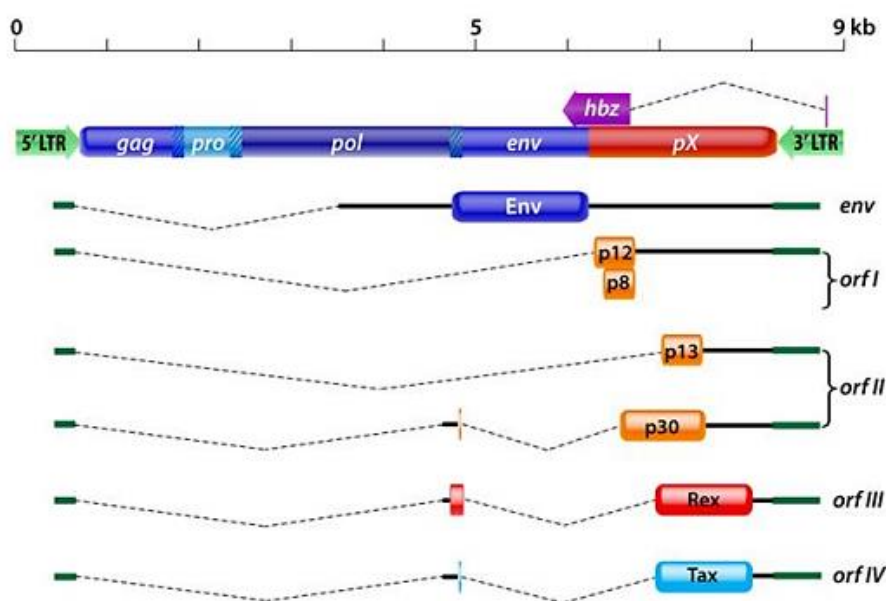
gp21 (Glicoproteína transmembrana), gp46 (Glicoproteína de superfície), p19 (Proteína da matriz), p24 (Proteína do nucleocapsídeo), RNA (Ácido ribonucléico), p55 (Transcriptase Reversa), p32 (Integrase) e p10 (Protease). Fonte: [www.HTLV.com.br](http://www.HTLV.com.br).

O DNA proviral do HTLV-1, de tamanho 9032pb, transcreve proteínas estruturais e regulatórias necessárias para o ciclo replicativo a partir de genes comumente encontrados em retrovírus *gag*, *pol* e *env* além do gene *pro*, a região *pX*, que contém quatro *Open Reading Frames* (ORF) parcialmente sobrepostas. Tais genes estão flanqueadas por duas regiões não codificantes e promotoras conhecidas como regiões 5' e 3' LTR (*Long Terminal Repeat*) (SEIKI et al, 1983; MATSUOKA & JEANG, 2007). A partir do DNA proviral são transcritos três mRNAs: uma fita de RNA completa, que não sofre *splicing*, é utilizada para a tradução dos produtos dos genes *gag*, *pol* e *pro*; a segunda fita é utilizada para a tradução das proteínas do gene *env*; a terceira fita codifica proteínas da região *pX*, dentre elas, Tax e HBZ (KROON & PROIETTI, 2015).

As quatro ORFs na região *pX*, (**Fig. 2**), são sobrepostas e produzem proteínas regulatórias e acessórias através de *splicing* alternativo. A ORF-I produz a proteína p12 que é clivada para gerar a p8, as duas são proteínas acessórias que provavelmente participam da regulação gênica e portanto, contribuem para a

persistência da infectividade da partícula viral. A ORF-II produz as proteínas p13 e p30: p13 participa de diversas funções dependendo do local intracelular onde se encontra, dentre elas, está a de regulador pós-traducional de Tax (SILIC-BENUSSI et al., 2010). p30, por sua vez, exerce junto com a proteína Rex uma regulação pós-transcricional de Tax. A ORF-III e a ORF-IV resultam na expressão de Rex e Tax, respectivamente e, um mRNA *antisense* transcrito a partir da região 3' LTR origina a proteína HBZ (*HTLV-1 Basic Zipper*) (KANNIAN & GREEN, 2010; EDWARDS et al., 2011). A proteína Tax é um potente transativador da expressão gênica do vírus além de ser um conhecido regulador da expressão gênica da célula infectada, incluindo genes envolvidos na proliferação celular, ciclo celular, apoptose e resposta a danos ao DNA. Rex, por sua vez, é um regulador pós-transcricional auxiliador da exportação de mRNA viral a partir do núcleo, exercendo um efeito contrário ao da proteína p30, que reprime a exportação do mRNA para o citoplasma. HBZ reprime a expressão viral ao regular negativamente a transativação proporcionada por Tax. Portanto, em conjunto, essas proteínas contribuem para a infectividade viral, manutenção das cargas virais, ativação das células do hospedeiro e regulação da transcrição gênica (MICHAEL et al., 2004).

**Figura 2 - Esquema do genoma do HTLV-1.**



Os genes estruturais *gag*, *pol* e *env* estão flanqueados pelas regiões 5' e 3' LTR (Setas verdes). O transcrito de *env* e das quatro ORFs, originárias da região *pX*, estão representados abaixo do genoma. O transcrito *antisense* de HBZ está representado pela seta lilás a partir da região 3' LTR. Tax, REX, p12, p8, p13 e p30 (Proteínas regulatórias codificadas por *pX*). Fonte: EDWARDS et al., 2011.

### 1.1.2 Epidemiologia

A distribuição da infecção pelo HTLV-1 revela algumas regiões endêmicas em alguns locais do mundo como o sudoeste do Japão, África subsaariana, América do Sul, Caribe, alguns lugares no Oriente Médio e Oceania (**Fig. 3**). As áreas que exibem mais de 5% de soropositividade para HTLV-1 são consideradas de alta prevalência enquanto que entre 5 e 1% são consideradas de média prevalência e menos de 1% de baixa prevalência (MINISTÉRIO DA SAÚDE, 2013). Estima-se que existam de 15 a 20 milhões de pessoas infectadas, pelo HTLV-1, ao redor do mundo, podendo esse número ser muito maior devido ao processo da estimativa de prevalência ser realizada através da sorologia de doadores de sangue, o que pode tornar esses números subestimados. Dados epidemiológicos da Argentina, Brasil, Colômbia e Peru são, em grande maioria, limitados a doadores de sangue (SANTOS et al., 2015). Além do mais, outro obstáculo é a inexistência de estudos sobre essa infecção em áreas densamente povoadas como a China, Índia e oeste da África (GONÇALVES et al, 2010; GESSAIN & CASSAR, 2012; VIANA et al., 2014).

**Figura 3 - Epidemiologia do HTLV-1 ao redor do mundo**



Fonte: GESSAIN & CASSAR, 2012.

O Brasil pode ter o maior número absoluto de infectados pelo HTLV-1 no mundo devido ao tamanho de sua população e estima-se que de 800.000 a 2,5 milhões de indivíduos estejam infectados (CATALAN-SOARES et al., 2005; GESSAIN & CASSAR, 2012). A região Sul brasileira apresenta a menor prevalência para o HTLV-1 e as regiões Norte e Nordeste as maiores, provavelmente devido ao maior percentual de descendentes de africanos escravizados durante a época do Brasil colônia e império, posto que o continente africano é o provável local de origem deste vírus (SANTOS et al., 2015). A região amazônica é a maior área endêmica brasileira e, Salvador, no estado da Bahia, é a cidade com a maior prevalência da infecção na população geral, com estimativa de 1,8% da população infectada por esse vírus, sendo mais elevada em mulheres (2,0%) em comparação com os indivíduos do sexo masculino (1,2%) (DOURADO et al., 2003; PAIVA & CASSEB, 2015).

## 1.2 MANIFESTAÇÕES CLÍNICAS ASSOCIADAS AO HTLV-1

Apesar de nem sempre refletir a realidade clínica, os escassos dados oficiais conduzidos até o momento, sugerem que a grande maioria dos indivíduos infectados pelo HTLV-1 permanece assintomática e ainda não se entende os motivos que levam os indivíduos a manifestar doença. Tem sido sugerido que o desfecho da infecção pode estar relacionado a fatores do hospedeiro e/ou virais, a exemplo da via de transmissão do vírus, carga proviral, tipo e magnitude da resposta imune do hospedeiro e por componentes genéticos envolvendo os alelos de HLA (COSTA et al, 2009). O HTLV-1 é associado com algumas pequenas desordens clínicas como uveíte (inflamação da úvea), dermatite, artropatia, disfunção erétil, desordens urinárias e neuropatia periférica (WATANABE, 1997; CASKEY et al., 2007) entretanto, as principais manifestações clínicas provocadas por esse vírus são a Mielopatia Associada ao HTLV/Paraparesia Espástica Tropical (HAM/TSP) e a Leucemia/Linfoma de células T do Adulto (ATLL) e, mesmo estas patologias não apresentando cura ou algum tratamento específico, a infecção pelo HTLV-1 é considerada negligenciada (COOK et al, 2013).

### 1.2.1 ATLL

A ATLL foi primeiramente descrita no Japão (UCHIYAMA, 1977) e posteriormente tornou-se a primeira neoplasia maligna humana associado a um vírus (POIESZ et al., 1980). É um linfoma não-Hodgkin, agressivo, que geralmente leva à

morte (TAGAYA & GALLO, 2017). Clinicamente é classificada em quatro subtipos (indolente, agudo, crônico e linfomatoso) de acordo com a contagem de linfócitos anormais circulantes, severidade dos sintomas, nível de lactato desidrogenase (LDH), hipercalcemia, envolvimento dos órgãos e outros pequenos distúrbios (SHIMOYAMA, 1991). Na sua forma aguda, os pacientes apresentam linfadenopatia, febre, lesões na pele, leucocitose e hepatoesplenomegalia devido ao aspecto infiltrativo da doença (LAIRMORE et al., 2012). A ATLL é mais frequente em indivíduos infectados pelo HTLV-1 na infância, permanecendo latente por volta de quatro a cinco décadas, sendo mais associado à via de transmissão vertical, principalmente a via da amamentação (GONÇALVES et al., 2010). Geralmente desenvolve-se em 1 a 5% dos pacientes infectados pelo HTLV-1, podendo sua verdadeira prevalência ser subestimada devido a uma alta taxa de mortalidade nos subtipos agudo ou linfomatoso e a não diferenciação diagnóstica com outras doenças malignas envolvendo leucócitos (LOUREIRO & LOPES, 2015).

Alguns aspectos da ATLL como o longo período de latência e a baixa incidência dessa doença leva à crer que o HTLV-1 somente não é suficiente para o desenvolvimento da ATLL e por isso, outras alterações genômicas, bem como fatores do hospedeiro são fundamentais para a biogênese da doença (LOUREIRO & LOPES, 2015). Os fatores específicos do hospedeiro incluem o seu genótipo, idade a qual ocorreu a infecção e presença de algumas coinfeções, especialmente *Strongyloides stercoralis* (BANGHAM et al., 2015). Indivíduos acometidos por essa manifestação apresentam descontrole na proliferação de células T, causado majoritariamente pela capacidade transativadora da proteína viral Tax, do acúmulo de mutações gênicas no hospedeiro, que acabam proporcionando o escape celular da apoptose, da regulação da transcrição viral e por fim, da modulação de fatores do hospedeiro pela proteína HBZ, única proteína constitutivamente expressa em todos os casos de ATLL (KATAOKA et al., 2015; ZHAO, 2016). Apesar de, no geral, pessoas do sexo feminino apresentarem prevalência maior da infecção pelo HTLV-1 (possivelmente por serem mais suscetíveis à transmissão desse vírus por via sexual) o sexo masculino apresenta uma probabilidade 40% maior de desenvolver ATLL em relação às mulheres (GRANT, 2002). Além disso, existe também um grande número relatado de ATLL em grupos familiares, indicando uma predisposição genética para a manifestação desse tipo de leucemia (YASUNAGA, 2007).

É uma doença extremamente difícil de tratar mesmo após 40 anos de estudos,

de modo que pacientes diagnosticados com ATLL, em média, terão menos de um ano de vida restante (TAGAYA & GALLO, 2017). As formas aguda e linfomatosa apresentam o pior prognóstico, com sobrevivência média de seis a 10 meses, mesmo com auxílio da quimioterapia multi-alvo (YAMADA et al., 2001). Todavia, alguns tratamentos não específicos estão disponíveis para ATLL, dependendo do subtipo da doença, como a quimioterapia, uso da zidovudina (AZT), interferon alfa (com ou sem trióxido de arsênico) e transplante alogênico de medula (TANOSAKI et al., 2008; BANGHAM et al., 2018). Recentemente, Mogamulizumab, um anticorpo anti-CCR4 aprovado no Japão foi desenvolvido especialmente para ATLL e vem apresentando bons resultados (MAKITA & TOBINAI, 2017).

### 1.2.2 HAM/TSP

A HAM/TSP é uma manifestação inflamatória do sistema nervoso central caracterizada principalmente por distúrbios urinários, como a bexiga neurogênica, e a paralisia dos membros inferiores, porém é eventualmente acompanhada de disfunção intestinal, perda de sensibilidade, parestesia (formigamento) e distesia (alteração na sensibilidade do tato) (GRASSI et al., 2011). Dentre os infectados pelo HTLV-1, de 1 a 4%, dependendo da etnia, desenvolvem HAM/TSP (BANGHAM, 2018) e por volta de 50% dos acometidos perderão os movimentos das pernas após 10 anos de progressão da doença (GESSAIN & MAHIEUX, 2012).

A HAM/TSP foi inicialmente detectada em áreas tropicais como na Martinica, Colômbia, Jamaica e associada com o HTLV-1 em 1985 (GESSAIN et al., 1985). Logo após, veio a confirmação com Rodgers-Johnson e cols. (1985), também em 1985, através da detecção de anticorpos contra o HTLV-1 no fluido espinhal, sendo naquele momento, denominada Paraparesia Espástica Tropical (TSP). No ano seguinte, no Japão, Osame e cols. (1986) identificaram uma doença de características semelhantes denominando-a Mielopatia Associada ao HTLV-1 (HAM).

Três importantes características estão associadas com o desenvolvimento de HAM/TSP: aumento da ativação da célula T, proliferação linfocitária descontrolada e produção de citocinas pró-inflamatórias (MENEZES et al., 2014). Indivíduos que desenvolvem HAM/TSP apresentam elevados níveis de carga proviral e exacerbada resposta imune, como pode-se observar pela presença de linfócitos T citotóxicos HTLV-1 específicos, o que, por sua vez, pode ser o fator de imunovigilância contra o desenvolvimento de tumor em células transformadas pelo HTLV-1, já que o mesmo

não é encontrado em pacientes ATLL (HANABUCHI et al, 2000).

A HAM/TSP desenvolve-se quando há infiltração de linfócitos T CD4<sup>+</sup> e T CD8<sup>+</sup> na medula espinhal levando a produção de citocinas pró-inflamatórias como IL-1, IL-6, TNF- $\alpha$  e IFN- $\gamma$  causando uma inflamação crônica (OLIÈRE et al., 2011). Modelos de infiltração já propostos foram resumidos por Silva & Araújo (2006), entretanto, ainda é incerto como ou o que desencadeia essa infiltração. O que se compreende é que há uma ativação de células T CD4<sup>+</sup> mediada por Tax, e que essas células ativadas invadem o sistema nervoso central (BANGHAM, 2017). Danos no trato piramidal e perda de mielina e axônios na medula espinhal inferior também são observados, porém ainda não se tem conhecimento acerca dos fatores que desencadeiam esses danos, se são efeitos diretos mediados pelo HTLV-1 ou efeitos exclusivos do sistema imune (GESSAIN & MAHIEUX, 2012).

A HAM/TSP é mais prevalente em indivíduos de idades mais avançadas, a partir dos 40-50 anos, no entanto, já foi reportada em crianças no Brasil e no Peru (KENDALL et al., 2009). É frequentemente acompanhada por alta carga proviral em células mononucleares do sangue periférico (PBMC), que aumenta ao longo dos anos. Em efeito disso, a análise da carga proviral foi sugerida como marcador de progressão e da atividade da HAM/TSP (OLINDO et al., 2005).

Devido à incerteza a respeito dos mecanismos fisiopatológicos dessa doença, uma terapêutica adequada e direcionada para HAM/TSP ainda não foi desenvolvida. Enquanto ainda não há cura, a terapia utilizada atualmente baseia-se, além do tratamento da sintomatologia (problemas urinários, dor e espasticidade), no uso de corticoides, vitamina C, interferon alfa e antirretrovirais, entretanto, aquém do sucesso desejado (GESSAIN & MAHIEUX, 2012).

### 1.3 MANIFESTAÇÃO DE DOENÇA E BIOMARCADORES

Biomarcadores são sinais biológicos que podem ser observados e medidos em pacientes acometidos por alguma patologia independentemente de manifestação de sintomas. Esses sinais são frequentemente utilizados como ferramentas de predição do status clínico, avaliação da progressão de doença, averiguação da eficácia e personalização de tratamentos e aprimoramento de estudos. Os biomarcadores podem variar de pressão sanguínea e medição de pulso a sinais moleculares, como a alteração da expressão de um gene (STRIMBU & TAVEL, 2010).

Alguns biomarcadores do hospedeiro já foram associados com manifestações



clínicas do HTLV-1. Antígenos leucocitários humanos (HLA) classe 1 (HLA-A, HLA-B, HLA-C) são glicoproteínas presentes na superfície de células nucleadas cuja principal função é a apresentação de antígenos a linfócitos T CD8<sup>+</sup>. No Japão, foi sugerida uma associação entre a presença do HLA-A\*02 ou HLA-Cw\*08, com uma menor carga proviral e uma prevalência de HAM/TSP menor (JEFFERY et al, 1999; JEFFERY et al, 2000). Por outro lado, os alelos HLA-DRB1\*0101 e HLA-B\*5401 apresentam associação com um aumento na susceptibilidade para HAM/TSP (JEFFERY et al, 1999; JEFFERY et al, 2000). Além desses, os receptores de quimiocinas CXCL10, CXCL9 e a Neopterina, presente no fluido cerebrospinal, foram associados com o desenvolvimento de HAM/TSP (YAMAMOTO-TAGUCHI et al., 2013) assim como a subexpressão de FOXP3 induzida pelo HBZ em células T CD4<sup>+</sup> (YAMAUCHI et al., 2019). Para ATLL, estudos sugerem a subexpressão das proteínas CAN2 (*Calpain-2*) e SPTA2 ( *$\alpha$ -II spectrin*) como úteis para classificação dos pacientes com ATLL (ISHIHARA et al., 2013), e alterações nos genes *GBP2* (*Interferon-induced guanylate-binding protein 2*) e *CD40LG* (*CD40 Ligand*) foram sugeridas como biomarcadores (FUKUTANI et al., 2019).

Quanto ao genoma viral do HTLV-1 e sua expressão gênica, entende-se que as proteínas Tax e HBZ são os principais responsáveis pelo desenvolvimento da ATLL. Tax ativa constitutivamente a via do NF- $\kappa$ B, fator que está envolvido na multiplicação das células T promovendo a proliferação desenfreada das células infectadas. A partir de certo ponto há a perda da expressão de Tax, mas de alguma forma as proteínas estimuladas por ela continuam sendo desreguladas (KATAOKA et al., 2015). O surgimento e acúmulo de mutações no genoma das células leucêmicas fornecem um escape da apoptose e de *checkpoints* do ciclo celular regulados pelo sistema imune (RAUCH & RATNER, 2011; KATAOKA et al., 2015). HBZ, por sua vez, regula Tax/Rex, induz a latência viral e estimula as células doentes a se proliferar (GIAM & SEMMES, 2016). Em HAM/TSP, uma das hipóteses para o desenvolvimento da doença sugere que há um mimetismo entre uma proteína das células neuronais, a Ribonucleoproteína Nuclear Heterogênea (hnRNP-A1) e a proteína Tax do HTLV-1, desencadeando um processo inflamatório autoimune, lesionando os neurônios (LEVIN et al., 2002). Outra hipótese é a produção exagerada de citocinas pró-inflamatórias por linfócitos T CD4<sup>+</sup> infectados e os linfócitos T CD8<sup>+</sup> que lesionariam as células neuronais (OSAME, 2002; NAKAMURA et al., 2009). Além desses marcadores virais, há também a utilização da carga proviral sérica como marcador

para progressão de HAM/TSP onde foi constatado que a mesma pode estar relacionada com o processo inflamatório no sistema nervoso central (NAGAI et al 1998).

A definição de um biomarcador em uma condição clínica, bem como a associação de padrões e redes gênicas com manifestações clínicas de determinado patógeno só é possível com a presença de alterações a nível transcricional. Muitas doenças derivam da interação entre a expressão gênica e o ambiente ou o patógeno. O estudo da expressão gênica diferencial é essencial para entender as diferenças biológicas entre os estados sadios e doente (EMILSSON, 2008). Estudos de expressão gênica através de chips de microarranjo são uma das alternativas para tentar entender as mudanças no transcriptoma causadas por um agente patogênico em um tecido ou célula. Apesar de existirem outras alternativas como o RNA-seq, que constitui o sequenciamento de nova geração (NGS, do inglês: *Next-generation sequencing*) do cDNA de todo o RNA intracelular, a tecnologia de microarranjos de DNA, desenvolvida anteriormente, possibilitou o avanço nos estudos em diversas áreas da pesquisa como testes de drogas (YUEN et al., 2018), entendimento de processos gênicos e patológicos (MIKLOS & MALESZKA, 2004), construção de redes de genes (ROBERTS et al., 2000) e identificação de novas regiões de *splicing* (LEE et al., 2017).

Os microarranjos são coleções de sondas de DNA correspondentes a transcritos gênicos que podem se ligar a moléculas alvo produzindo dados quantitativos (expressão gênica) ou qualitativos (técnicas de diagnóstico). No geral, em sua estrutura, as sequências de DNA estão imobilizadas e organizadas em milhões de “*spots*”, os quais representam um gene específico, e se hibridizarão à apenas uma sequência alvo na amostra, emitindo uma fluorescência que será detectada por um *scanner* (MILLER & TANG, 2009). A partir dos resultados de microarranjo são identificados os Genes Diferencialmente Expressos (GDE), sendo aqueles genes que apresentam expressão diferenciada no grupo teste, como em estados patológicos, em comparação com níveis normais de expressão. Em pesquisas farmacêuticas e clínicas, os GDEs podem ser úteis para sugestão de biomarcadores, alvos terapêuticos e assinaturas gênicas para diagnóstico (RODRIGUEZ-ESTEBAN & JIANG, 2017).

Bancos de dados públicos de estudos de microarranjo estão disponíveis na internet, dentre eles o *Gene Expression Omnibus* (GEO). O GEO é um repositório

internacional com suporte do *National Center for Biotechnology Information* (NCBI) que arquiva e disponibiliza dados de microarranjos e RNA-seq depositados por pesquisadores do mundo todo. Dados biológicos disponibilizados no GEO seguem o padrão MIAME (*Minimum Information About a Microarray Experiment*) para padronização e manutenção da qualidade da informação permitindo assim seu uso e replicação (CLOUGH & BARRETT, 2016). Criado em 2000, atualmente esse repositório contém cerca de 114.515 estudos e 19.824 plataformas diferentes.

Diversas plataformas comerciais de microarranjo estão disponíveis para identificação e quantificação de genes ou micro-RNAs diferencialmente expressos. Dentre as plataformas de microarranjos há uma predominância de microarranjos da Affymetrix, Illumina e Agilent diferenciando-se na sua manufatura, repertório e análise dos genes e na composição dos *spots* no chip. Eles podem conter oligonucleotídeos curtos (25-30 bases), longos (50-80 bases) ou cDNA (variável em tamanho) responsáveis por identificar de 40 a 60 mil genes por chip (BARNES et al., 2005).

#### 1.4 MICRO-RNA

Micro-RNAs (miRNAs) são uma classe de moléculas de RNA não codificantes, de aproximadamente 22 nucleotídeos, transcritos a partir de um gene, geralmente apresentando papel regulatório na pós-transcrição gênica ao interagir com a região 3' não traduzida (3' UTR) das moléculas de mRNAs alvo com o objetivo de induzir sua degradação e repressão traducional (RUPAIMOOLE & SLACK, 2017). A interação dos miRNAs e seus genes alvos é dependente da localização celular, nível da expressão de miRNAs e número de cópias dos genes alvo. Podem ser secretadas extracelularmente ou transportadas por vesículas (O'BRIEN et al., 2018). Nas últimas décadas, o entendimento da importância do papel dos miRNAs na regulação se traduz no aumento dos estudos a respeito desse tema, no aprimoramento da anotação gênica e identificação dos genes alvo.

Atualmente sabe-se que os miRNAs estão envolvidos em diversos processos biológicos normais como diferenciação celular (RAN et al., 2017), regulação do sistema imune (MEHTA & BALTIMORE, 2016) e do metabolismo (ZHU & LEUNG, 2018). Por outro lado, a desregulação dos miRNAs está presente em processos patológicos como câncer (ALLAYA et al., 2015), doenças neurodegenerativas (QUINLAN et al., 2017) e doenças infecciosas (ELEDGE & YERUVA, 2018). No câncer, moléculas de miRNAs podem agir como oncogenes por influenciar alguns

mecanismos, como a metástase (ALLAYA et al., 2015). Além do mais, podem agir como supressores ou promotores de tumor, sendo, portanto, importantes biomarcadores (SABARIMURUGAN et al., 2019).

A atividade de regulação exercida por miRNAs sobre um gene é baseada no silenciamento gênico através da atuação de um complexo de proteínas, como a Argonauta, associado ao miRNA processado, conhecido como *RNA-Induced Silencing Complex* (RISC). O miRNA presente no RISC serve como uma molécula guia que direcionará esse complexo silenciador ao mRNA alvo por complementaridade parcial das bases e o sequestrará, induzindo a repressão da tradução ou compartimentalização do mesmo (BARTEL, 2004). Microarranjos são rotineiramente utilizados para estudos de desregulação de miRNAs em estados patológicos versus saudáveis.

### 1.5 JUSTIFICATIVA

A tecnologia do microarranjo permite a identificação de genes diferencialmente expressos (GDEs) porém, os resultados podem variar nos diferentes estudos de acordo com as diferentes abordagens metodológicas utilizadas. Sabe-se que algumas pesquisas realizadas com microarranjos podem não ser reproduzíveis, podendo ter como causas uma análise executada de forma errônea, métodos de validação inadequados, controle insuficiente de falsos positivos e/ou uma divulgação insuficiente da metodologia (NTZANI & IOANNIDIS, 2003; DUPUY & SIMON, 2007). Outro grande problema é o baixo número amostral, geralmente algumas poucas dezenas de pacientes são utilizados para estudos com microarranjos de milhares de genes, o que diminui o poder estatístico da análise. Além disso, a generalização de regiões geográficas ou históricas a partir de um microarranjo com população específica também deve ser evitado (FERGUSON, 2004).

A metanálise de dados de microarranjo é uma abordagem utilizada para superar esses problemas e gerar uma estimativa mais precisa da expressão gênica diferencial por aumentar o poder estatístico da análise através da combinação de estudos individuais, porém relacionados, e conseqüentemente a obtenção de um maior número de amostras, numa mesma avaliação. Devido ao fato de a infecção pelo HTLV-1 ser considerada rara e, conseqüentemente, a obtenção de amostras ser difícil, essa técnica contribuiria de forma significativa para a elaboração desses perfis de expressão gênica diferencial associados às manifestações clínicas de maneira

mais verossímil. Até o momento, nenhuma metanálise integrando ferramentas de bioinformática foi realizada no intuito de revelar aspectos moleculares desregulados em status clínicos associados à infecção por HTLV-1. O que destaca a singularidade do estudo descrito nesse exemplar, bem como a contribuição do mesmo para futuras repercussões no que diz respeito ao manejo clínico desses indivíduos.

A importância da indicação e caracterização correta acerca dos perfis de expressão gênica indica ainda a possibilidade de compreender mecanismos moleculares de manifestação patológica e, a partir deles, criar um banco de dados de possíveis redes ou genes, candidatos essenciais ao desenvolvimento da doença, que podem ser alvo de pesquisas clínicas. Essa importância científica se torna ainda maior levando em consideração que o HTLV-1 é o responsável por duas manifestações clínicas importantes e negligenciadas, além de estar associado com estados assintomáticos não compreendidos. Portanto, estudos para a delimitação do perfil de expressão gênica diferencial, nesses indivíduos, tornam-se ainda mais necessários.

## 2. OBJETIVOS

### 2.1 OBJETIVO GERAL

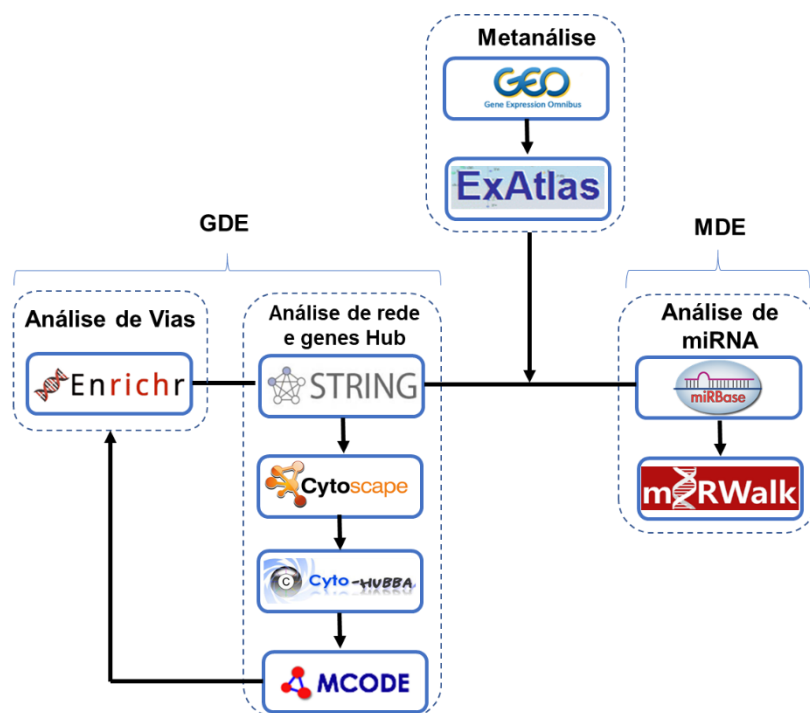
- Avaliar o perfil transcricional das principais manifestações clínicas associadas ao HTLV-1 com uso de ferramentas de bioinformática e banco de dados públicos.

### 2.2 OBJETIVOS ESPECÍFICOS

- Identificar os genes e micro-RNAs diferencialmente expressos (GDEs e MDEs), e determinar quais deles apresentam uma maior ou menor expressão em indivíduos infectados pelo HTLV-1, levando em conta o status clínico (HAM/TSP, ATLL ou ASS) em comparação com indivíduos não infectados;
- Identificar as vias bioquímicas mais enriquecidas, que envolvem os GDEs;
- Avaliar a formação de redes de interação entre os GDEs obtidos nas diferentes condições clínicas com identificação de módulos e genes hubs.

### 3. METODOLOGIA

**Figura 4** - Fluxograma da metodologia utilizada.



#### 3.1 BUSCA DE DADOS DE EXPRESSÃO GÊNICA:

A busca de dados normalizados de expressão gênica em ensaios envolvendo a infecção pelo HTLV-1 foi realizada no repositório do GEO (<<https://www.ncbi.nlm.nih.gov/gds>>) (**Fig. 4**) utilizando os termos HTLV-1, ATLL e HAM/TSP, sob os filtros “*Homo sapiens*” e “*Microarray*” no período de 09/2018 a 10/2018. Os ensaios foram filtrados de acordo com os seguintes critérios de inclusão: 1) perfil de expressão gênica derivado de microarranjo a partir células humanas de pacientes infectados pelo HTLV-1; 2) deveria apresentar ao menos um grupo controle e um grupo teste no estudo, com o grupo controle composto de indivíduos não infectados e o grupo teste composto por pacientes infectados somente por HTLV-1, sem outra co-infecção; 3) sem tratamento aplicado às células.

### 3.2 METANÁLISES:

Os estudos selecionados foram submetidos à ferramenta *online* ExAtlas (Sharov et al., 2015, <<https://lgsun.grc.nia.nih.gov/exatlas/>>) (**Fig. 4**) para realização da metanálise “*cross-platform*”, ou seja, uma metanálise que combina estudos de microarranjo de plataformas diferentes. Após a submissão, uma análise de variância (ANOVA) foi realizada pela ferramenta para remoção de *outliers*, mesmo já tendo sido utilizado dados normalizados do GEO. Uma análise prévia da qualidade dos estudos individuais foi realizada de acordo com parâmetros sugeridos pela ferramenta: remoção de amostras com desvio padrão (SD) > 3 dentro cada estudo e análise do agrupamento de amostras em seus respectivos perfis clínicos através de gráficos de análise de componentes principais (PCA, do inglês: *Principal Component Analysis*).

Em seguida, três metanálises foram realizadas: uma comparando os estudos agrupados de ATLL com controle, outra comparando HAM/TSP com controle e por fim, uma terceira que comparou os dados de indivíduos ASS com controle. As metanálises foram realizadas também no ExAtlas utilizando o método do tamanho de efeito combinado (efeito aleatório) associado ao recurso de ajuste do valor de  $p$  proposto por Benjamini & Hochberg (1995) com os parâmetros *False Discovery Rate* (FDR)  $\leq 0,05$  e *Fold Change* (FC)  $\geq 1,5$  e  $\leq -1,5$  para evitar falsos positivos. Os resultados obtidos são transformados em valores de  $\log_{10}$  para facilitar o entendimento. Também foram obtidos os GDEs de cada estudo individual nos mesmos parâmetros das metanálises com o objetivo de calcular o valor de IDD (*Integration Driven-Discovery*), Perda, IDR (*Integration Driven Rate*) e IRR (*Integration Resolution Rate*) a fim de estimar a eficiência da metanálise (CHOI et al., 2003; STEVENS et al., 2005).

$$\text{IDD} = \text{n}^\circ \text{ genes [DE na MA e não-DE nos EI]}$$

$$\text{Perda} = \text{n}^\circ \text{ genes [DE em pelo menos um EI e não-DE na MA]}$$

$$\text{IDR} = \frac{\text{n}^\circ \text{ genes [DE na MA e não-DE nos EI]}}{\text{n}^\circ \text{ genes [DE in MA]}}$$

$$\text{IRR} = \frac{\text{n}^\circ \text{ genes [DE em pelo menos um EI e não-DE na MA]}}{\text{n}^\circ \text{ genes [DE em pelo menos um EI]}}$$

Onde: DE = Diferencialmente Expresso

MA = Metanálise

EI = Estudo Individual



As três listas de GDEs obtidas nas metanálises foram submetidas ao Venny 2.1 (OLIVEROS, 2007-2015, <<http://bioinfogp.cnb.csic.es/tools/venny/>>) para geração de diagramas de *Venn* a fim de sobrepor e comparar os perfis de expressão gênica diferencial dos três status clínicos.

### 3.3 ANÁLISE DE VIAS ENRIQUECIDAS:

As listas de GDEs dos três grupos em estudo, divididas em genes superexpressos e subexpressos, foram submetidas separadamente ao Enrichr (CHEN et al., 2013; KULESHOV et al., 2016, <<https://amp.pharm.mssm.edu/Enrichr/>>) (**Fig. 4**), ferramenta online de análise de listas gênicas para obtenção das vias enriquecidas nos diferentes grupos através do teste exato de Fisher. A análise utilizou como base o banco de vias da KEGG 2019 (*Kyoto Encyclopedia of Genes and Genomes*) (FDR  $\leq$  0,05).

### 3.4 ANÁLISE DE REDE DE INTERAÇÃO PROTEÍNA-PROTEÍNA, SELEÇÃO DE MÓDULOS E GENES HUB:

As listas de GDEs geradas nas três metanálises foram submetidas ao STRING v11 (*Search Tool for Retrieval of Interacting Genes/Proteins*) (SNEL et al., 2000, <<https://string-db.org/>>) (**Fig. 4**), ferramenta que fornece dados experimentais e preditos de interação física de proteínas para obtenção das redes de Interação Proteína-Proteína (IPP). Proteínas preditas foram adicionadas na rede de HAM/TSP e ASS objetivando alcançar o valor de  $p \leq 0,05$ . Em seguida, as redes IPP do STRING foram importadas para o Cytoscape v3.7.1 (SHANNON et al., 2003) (**Fig. 4**), uma ferramenta que permite a visualização e manipulação das redes gênicas. A partir das redes IPP geradas, módulos densos (sub-redes) de proteínas interconectadas foram identificados através do MCODE (*Molecular Complex Detection*) (BADER et al., 2003) (**Fig. 4**), *plugin* do Cytoscape, utilizando os parâmetros padrão (*Degree Cutoff: 2, K-Core: 2, Node Score Cutoff: 0.2, Profundidade máxima da semente: 100*). Foram selecionados os módulos com valor de *score*  $\geq 10$  para ATLL e  $\geq 5$  para HAM/TSP e ASS. As proteínas presentes nos módulos foram testadas para as vias enriquecidas como descrito na seção 3.3. Genes hub (nós altamente interconectados) foram extraídos das redes IPP através do *plugin* CytoHubba (CHIN et al., 2014) (**Fig. 4**) no Cytoscape utilizando o algoritmo de análise topológica MCC (*Maximal Clique Centrality*), indicado pela

ferramenta por possuir uma melhor performance para células eucarióticas. Os primeiros 5 genes hub, com base no valor do *score* (densidade\*#nós), foram selecionados para cada grupo.

### 3.5 ANÁLISE DE MICRO-RNAS:

Os MDEs obtidos nas metanálises dos três grupos foram classificados de acordo com a qualidade da anotação depositada no banco de dados miRBase (GRIFFITHS-JONES et al., 2006, <<http://mirbase.org>>) (**Fig. 4**). Os MDEs de alta confiabilidade na anotação foram submetidos ao miRWalk 3.0 (STICHT et al., 2018, <<http://mirwalk.umm.uni-heidelberg.de>>) para a identificação dos genes alvo. Somente os genes alvo experimentalmente validados, sinalizados pelo miRTarBase (Banco de dados curados de miRNA), foram mantidos.

## 4. RESULTADOS

### 4.1 BUSCA POR ENSAIOS DE EXPRESSÃO NO GEO

A busca por ensaios de expressão gênica, em indivíduos infectados pelo HTLV-1 com diferentes manifestações clínicas, no banco de dados do GEO retornou 48 resultados, de acordo com os critérios de inclusão/exclusão estabelecidos na seção Metodologia: foram eliminados da seleção os estudos de RNA-seq, estudos com indivíduos sob algum tipo de tratamento, estudos de exons e estudos sem indivíduos controle (não infectados pelo HTLV-1). Ao final, foram mantidos 10 estudos de diferentes plataformas de microarranjo e que haviam sido realizados em grupos celulares semelhantes (**Quadro 1**).

**Quadro 1** - Relação dos Ensaios de expressão gênica de pacientes infectados pelo HTLV-1, selecionados no GEO, para inclusão nas metanálises.

| Estudo                           | Série    | Célula           | Plataforma  | GPL      |
|----------------------------------|----------|------------------|---|----------|
| <b>CHOI et al., 2007</b>         | GSE1466  | CD4 <sup>+</sup> | Affymetrix Human Genome U133A Array                       | GPL9696  |
| <b>PISE-MASISON et al., 2009</b> | GSE14317 | CD4 <sup>+</sup> | Affymetrix Human Genome U133A 2.0 Array                   | GPL571   |
| <b>OLIERE et al., 2010</b>       | GSE19080 | CD4 <sup>+</sup> | Homo sapiens 3K Immunoarray2 HTLV-1 JH                    | GPL9686  |
| <b>KRESS et al., 2010</b>        | GSE17718 | CD4 <sup>+</sup> | Affymetrix Human Genome U133 Plus 2.0 Array               | GPL570   |
| <b>TATTERMUSCH et al., 2012</b>  | GSE29332 | PBMC             | Illumina HumanWG-6 v3.0 expression beadchip               | GPL6884  |
| <b>TATTERMUSCH et al., 2012</b>  | GSE29312 | PBMC             | Illumina HumanHT-12 V3.0 expression beadchip              | GPL6947  |
| <b>YAMAGISHI et al., 2012</b>    | GSE33615 | CD4 <sup>+</sup> | Agilent-014850 Whole Human Genome Microarray 4x44K G4112F | GPL4133  |
| <b>PINTO et al., 2013</b>        | GSE38537 | CD4 <sup>+</sup> | Agilent-014850 Whole Human Genome Microarray 4x44K G4112F | GPL6480  |
| <b>NAKAHATA et al., 2014</b>     | GSE43017 | CD4 <sup>+</sup> | Affymetrix Human Genome U133 Plus 2.0 Array               | GPL570   |
| <b>KOBAYASHI et al., 2014</b>    | GSE55851 | CD4 <sup>+</sup> | Agilent-026652 Whole Human Genome Microarray 4x44K v2     | GPL10332 |

GEO – *Gene Expression Omnibus*. GPL – Código da plataforma; PBMC – Célula mononuclear do sangue periférico.

Os estudos selecionados para a metanálise apresentam diferentes composições quanto à sintomatologia da infecção pelo HTLV-1 (**Tabela 1**). No total, foram obtidas amostras de expressão gênica de 36 indivíduos portadores de HAM/TSP, 175 amostras de portadores de ATLL, 58 amostras de indivíduos assintomáticos e 70 amostras controle.

**Tabela 1** - Composição dos estudos selecionados para as metanálises quanto aos diferentes status clínicos associados à infecção pelo HTLV-1.

| <b>Série</b>    | <b>HAM/TSP</b> | <b>ATLL</b> | <b>ASS</b> | <b>Controle</b> |
|-----------------|----------------|-------------|------------|-----------------|
| <b>GSE1466</b>  | 0              | 70          | 0          | 3               |
| <b>GSE14317</b> | 0              | 18          | 0          | 7               |
| <b>GSE19080</b> | 12             | 7           | 11         | 8               |
| <b>GSE17718</b> | 0              | 2           | 0          | 2               |
| <b>GSE29332</b> | 10             | 0           | 17         | 8               |
| <b>GSE29312</b> | 10             | 0           | 20         | 9               |
| <b>GSE33615</b> | 0              | 52          | 0          | 21              |
| <b>GSE38537</b> | 4              | 0           | 4          | 4               |
| <b>GSE43017</b> | 0              | 7           | 0          | 5               |
| <b>GSE55851</b> | 0              | 12          | 6          | 3               |
| <b>TOTAL</b>    | <b>36</b>      | <b>175</b>  | <b>58</b>  | <b>70</b>       |

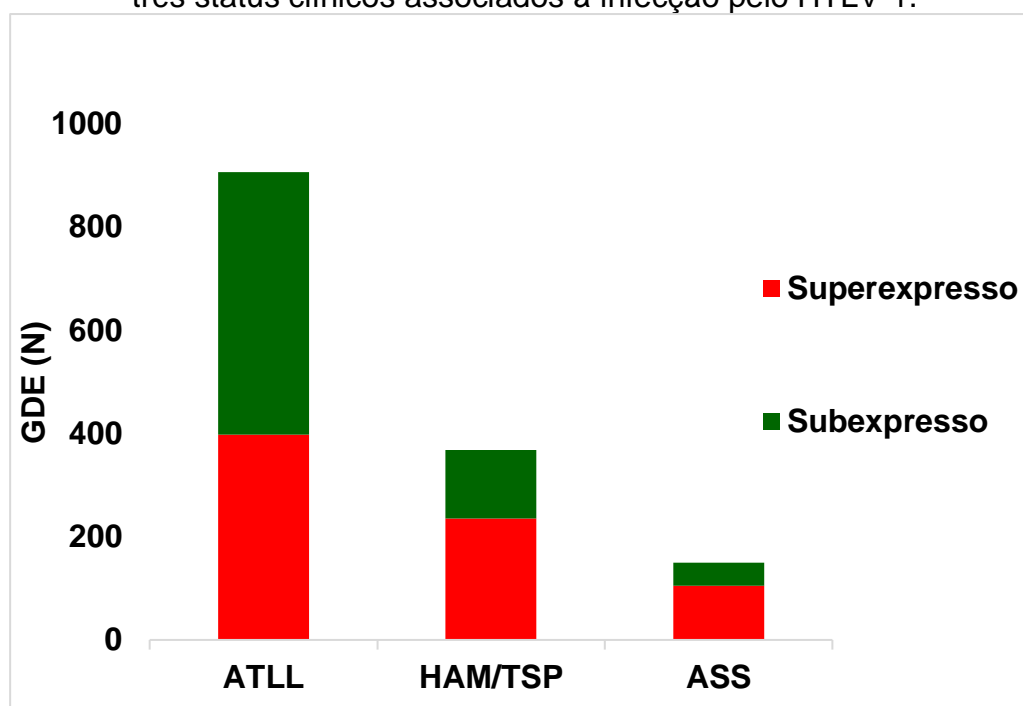
ATLL – Leucemia/linfoma de células T do adulto. ASS – Assintomáticos. HAM/TSP - Mielopatia Associada ao HTLV-1/Paraparesia Espástica Tropical.

#### 4.2 METANÁLISES

Durante a análise de qualidade das planilhas de expressão gênica dos estudos, foi observado o agrupamento das amostras em seus respectivos status clínicos pelos gráficos de análise de componentes principais (PCA) (**Apêndice A**). Além disso, foram removidas duas amostras pertencentes ao ensaio com identificação GSE1466 (GSM24820 e GSM24830), cinco do estudo GSE33615 (GSM831223, GSM831224, GSM831244, GSM831248 e GSM831272) e uma amostra do ensaio identificado como GSE43017 (GSM1055154) por apresentarem  $SD > 0,3$ .

Seguinte à análise de qualidade, as metanálises de ATLL, HAM/TSP e assintomáticos foram realizadas seguindo a metodologia sugerida por Choi e cols. (2003) com o objetivo de encontrar genes diferencialmente expressos (GDE) em comparação com os indivíduos controle não infectados. Foram encontrados, com os parâmetros aplicados ( $FDR \leq 0,05$  e  $-1,5 \leq FC \leq 1,5$ ), 907 GDEs para ATLL, 368 GDEs para HAM/TSP e 150 GDEs para ASS (**Fig. 5**).

**Figura 5** - Quantidade de genes diferencialmente expressos (GDE) nos três status clínicos associados à infecção pelo HTLV-1.



ATLL – Leucemia/linfoma de células T do adulto. ASS – Assintomáticos. FC – *Fold Change*. GDE – Genes Diferencialmente Expressos. HAM/TSP - Mielopatia Associada ao HTLV-1/Paraparesia Espástica Tropical.

Após a identificação da existência de GDEs entre os diferentes grupos, foi possível destacar os 5 genes (Superexpressos e subexpressos) mais desregulados em cada status clínico, como ilustrado na **Tabela 2**.

**Tabela 2** - Destaque para os cinco genes mais superexpressos e subexpressos nos três status clínicos associados à infecção pelo HTLV-1.

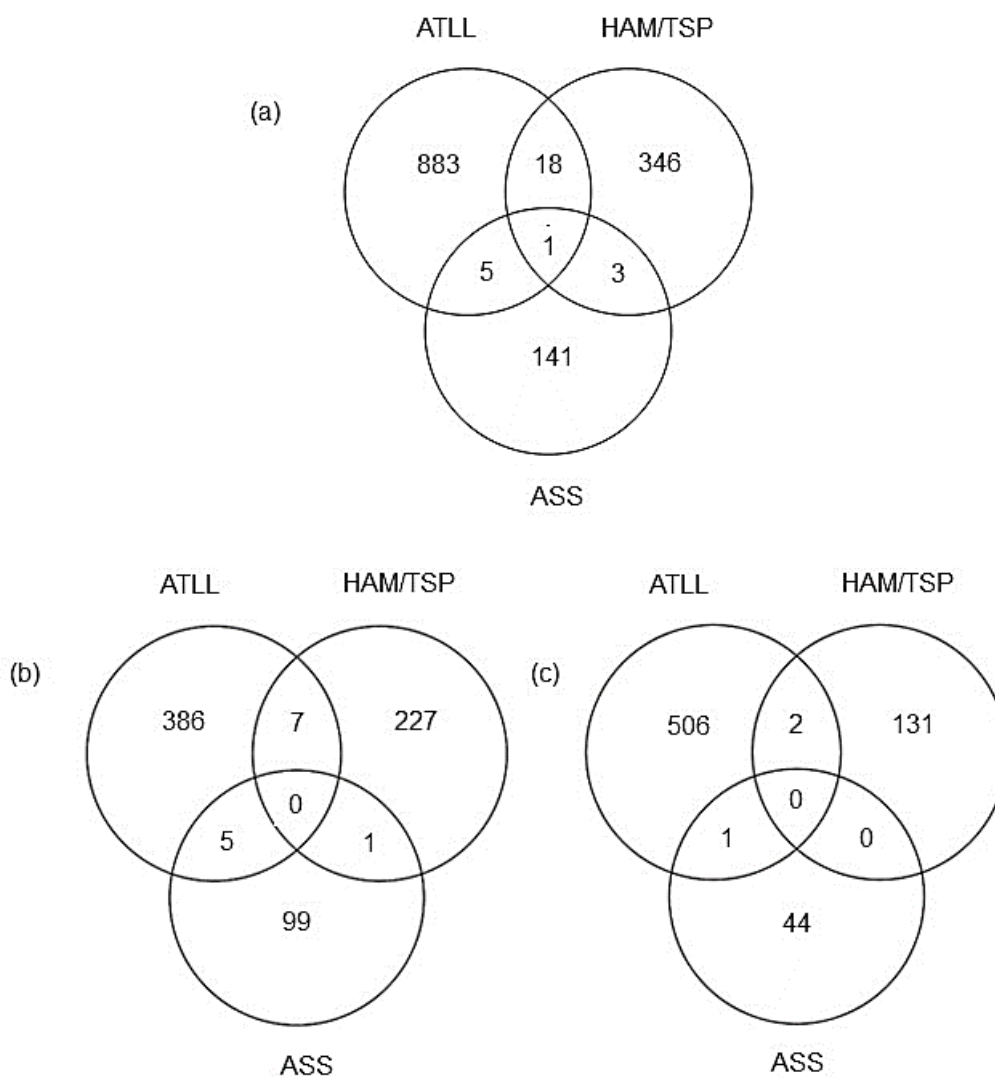
| Expressão     | ATLL           |                      | HAM/TSP       |                      | ASS              |                      |
|---------------|----------------|----------------------|---------------|----------------------|------------------|----------------------|
|               | GDEs           | log <sub>10</sub> FC | GDEs          | log <sub>10</sub> FC | GDEs             | log <sub>10</sub> FC |
| Superexpresso | <i>CADM1</i>   | 1,6701               | <i>BRINP3</i> | 1,1244               | <i>CFAP43</i>    | 0,8468               |
|               | <i>TWIST1</i>  | 1,4943               | <i>PRLHR</i>  | 1,0596               | <i>DNAH12</i>    | 0,8314               |
|               | <i>EVC</i>     | 1,2805               | <i>CAVIN4</i> | 0,9225               | <i>ANKRD30A</i>  | 0,8027               |
|               | <i>RGS13</i>   | 1,1569               | <i>RAVER2</i> | 0,9193               | <i>HORMAD2</i>   | 0,7611               |
|               | <i>MYOF</i>    | 1,1547               | <i>GFM1</i>   | 0,8462               | <i>SYT14P1</i>   | 0,6744               |
| Subexpresso   | <i>SCML1</i>   | -1,2867              | <i>ETFA</i>   | -1.318               | <i>LINC00554</i> | -0,9072              |
|               | <i>LAPTM4B</i> | -1,2263              | <i>STEAP1</i> | -0,9855              | <i>NONOP1</i>    | -0,8085              |
|               | <i>HOOK1</i>   | -1,0896              | <i>LRFN2</i>  | -0,9311              | <i>CT45A2</i>    | -0,6466              |
|               | <i>FAM153A</i> | -1,073               | <i>SORT1</i>  | -0,9204              | <i>OTOL1</i>     | -0,494               |
|               | <i>MAN1C1</i>  | -1,0615              | <i>ASTL</i>   | -0,8643              | <i>LOC644335</i> | -0,4863              |

ATLL – Leucemia/linfoma de células T do adulto. ASS – Assintomáticos. FC – *Fold Change*.

GDE – Genes Diferencialmente Expressos. HAM/TSP - Mielopatia Associada ao HTLV-1/Paraparesia Espástica Tropical.

O resultado da sobreposição dos perfis de genes desregulados nas diferentes manifestações clínicas mostra que o maior compartilhamento desses genes ocorre entre ATLL e HAM/TSP (18) seguido por ATLL e ASS (5) e ASS e HAM/TSP (3) (**Fig. 6 (a)**). Apenas um gene é compartilhado por todas as manifestações, o *ANK1* (Ankyrin 1), gene que codifica a proteína anquirina-R da família das Anquirinas. Ela aparece superexpressa em ATLL ( $\log_{10}FC = 0,1773, 9,03^{-06}$ ) e ASS ( $\log_{10}FC = 0,2662, 7,65^{-05}$ ) mas aparece subexpressa em HAM/TSP ( $\log_{10}FC = -0,5886, FDR = 4,03^{-04}$ ). A sobreposição dos GDEs separados em superexpressos (**Fig. 6 (b)**) e subexpressos (**Fig. 6 (b)**) também está ilustrado.

**Figura 6** - Sobreposição dos perfis de genes desregulados nos três status clínicos associados à infecção pelo HTLV-1. (a) Conjunto total de GDEs, (b) GDEs superexpressos, e (c) GDEs subexpressos.



ATLL – Leucemia/linfoma de células T do adulto. ASS – Assintomáticos. HAM/TSP - Mielopatia Associada ao HTLV-1/Paraparesia Espástica Tropical.

A partir dos resultados das metanálises, foram calculados os parâmetros de Descoberta (IDD, IDR) e Perda (Perda, IRR), de modo que os cálculos realizados para a metanálise com indivíduos HAM/TSP revelaram uma maior taxa de descoberta (IDR), enquanto a menor taxa obtida foi para os indivíduos com ATLL.

**Tabela 3** - Cálculos de Descoberta e Perda nas metanálises dos status clínicos (HAM/TSP, ATLL e ASS) associados à infecção pelo HTLV-1.

|                | <b>GDEs</b> | <b>IDD</b> | <b>Perda</b> | <b>IDR (%)</b> | <b>IRR (%)</b> |
|----------------|-------------|------------|--------------|----------------|----------------|
| <b>ATLL</b>    | 907         | 63         | 10283        | 6,95           | 163,22         |
| <b>HAM/TSP</b> | 368         | 260        | 5609         | 70,65          | 21,57          |
| <b>ASS</b>     | 150         | 89         | 4650         | 59,33          | 52,24          |

ATLL – Leucemia/linfoma de células T do adulto. ASS – Assintomáticos. GDE – Genes Diferencialmente Expressos. HAM/TSP - Mielopatia Associada ao HTLV-1/Paraparesia Espástica Tropical. IDD – *Integration Driven Discovery*. IDR – *Integration Driven Rate*. IRR – *Integration Resolution Rate*.

#### 4.3 ANÁLISE DE VIAS ENRIQUECIDAS

A avaliação para identificação das vias enriquecidas para o conjunto total de GDEs, divididos em superexpressos e subexpressos, das três condições em estudo, indicou que apenas o conjunto de GDEs de indivíduos com ATLL apresentou vias enriquecidas para o banco de dados da KEGG (**Tabela 4**) de acordo com o parâmetro adotado ( $FDR \leq 0,05$ ). Os demais conjuntos de dados dos indivíduos com HAM/TSP e ASS não apresentaram resultados.



**Tabela 4 - Vias enriquecidas para os GDEs em ATLL.**

| ATLL                |                           |              |   |
|---------------------|---------------------------|--------------|---|
| Perfil de Expressão | Via enriquecida           | FDR          | Genes   |
| Superexpressos      | Vias em Câncer (map09020) | $2,05^{-02}$ | <i>CEBPA, MGST3, PTGER3, PIK3R3, MGST2, ADCY1, NRAS, CTNNA1, PMAPI1, ELOB, JAK2, RALGDS, NQO1, JAG1, NCOA3, HGF, WNT5A, AR, CCNE1, IL2RA, FGF18, CTNNB1, CYCS, BAX, PLCB1</i> |
| Subexpressos        | Endocitose (ko04144)      | $1,76^{-02}$ | <i>PRKCI, SMURF2, ARRB1, VPS37B, TGFBR1, IGF1R, TGFBR2, SPART, CYTH3, CYTH4, GRK5, BIN1, PSD4, CHMP3, CCR5, LDLRAP1, CHIMP6, CHIMP7</i>                                       |

ATLL – Leucemia/linfoma de células T do adulto. FDR – *False Discovery Rate*. GDE – Genes Diferencialmente Expressos.

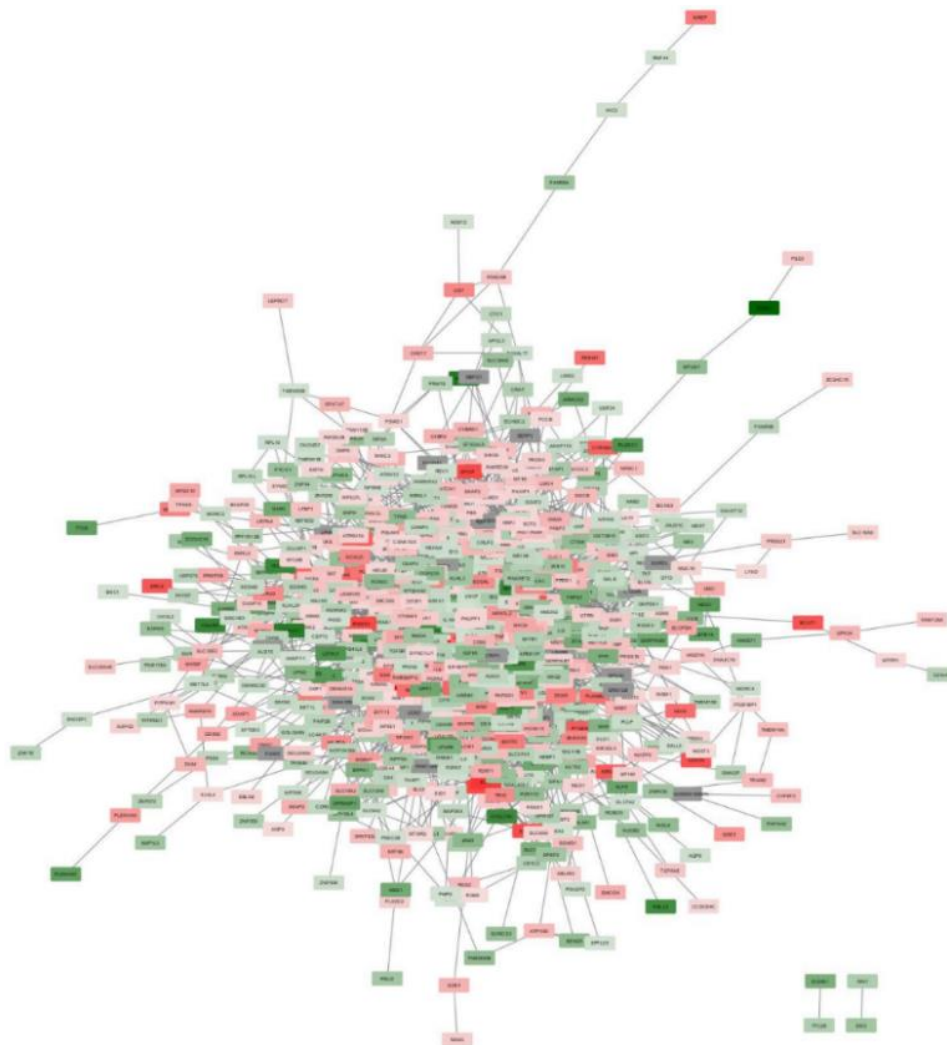
#### 4.4. INTERAÇÃO PROTEÍNA-PROTEÍNA, SELEÇÃO DE MÓDULOS E GENES HUBS

Passada a identificação do perfil de expressão dos genes em cada grupo, bem como as funções moleculares e biológicas dos GDEs identificados, seja de forma exclusiva por cada grupo ou em comum, seguiu-se para a confecção das redes de interação Proteína-Proteína (IPP), para identificação de relação física entre as proteínas codificadas pelos GDEs em cada grupo de dados.

#### 4.4.1 ATLL

A lista dos GDEs superexpressos e subexpressos em indivíduos com ATLL gerou uma rede IPP com 868 nós e 3,147 arestas ( $p = 1,0^{-16}$ ) (**Fig. 7**).

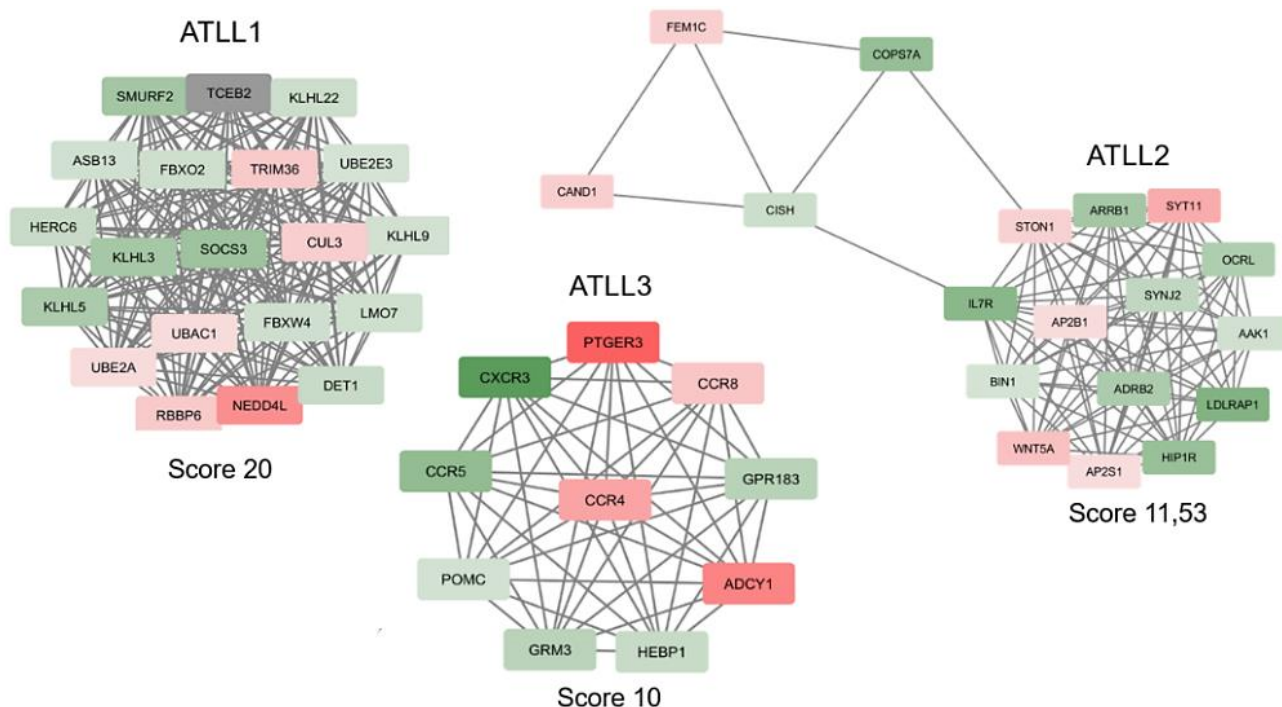
**Figura 7** - Rede de Interação Proteína-Proteína de GDEs em indivíduos com ATLL.



Construída no *Cytoscape* v3.7.1 com base no banco de dados do STRING v11. Cores vermelhas representam superexpressão, cores verdes representam subexpressão e cor cinza representa genes preditos. ATLL – Leucemia/linfoma de células T do adulto. GDEs – Genes Diferencialmente Expressos.

A partir da rede IPP obtida anteriormente, o *plugin* MCODE, também do *Cytoscape*, extraiu 27 módulos, porém, por motivo de simplificação e importância, foram selecionados os módulos com  $score \geq 10$ . Quanto mais numerosas as interações entre o grupo de proteínas do módulo, maior o  $score$  (**Fig. 8**).

**Figura 8** – Módulos extraídos da rede de Interação Proteína-Proteína de ATLL.



Extraídos com MCODE, no *Cytoscape* v3.7.1. Cores vermelhas representam superexpressão, cores verdes representam subexpressão e cor cinza representa genes preditos.

Quando analisadas as composições dos módulos individualmente, foi possível identificar que os três módulos selecionados da rede IPP de indivíduos com ATLL apresentaram diferentes vias enriquecidas (**Tabela 5**), relacionadas a funções biológicas distintas.

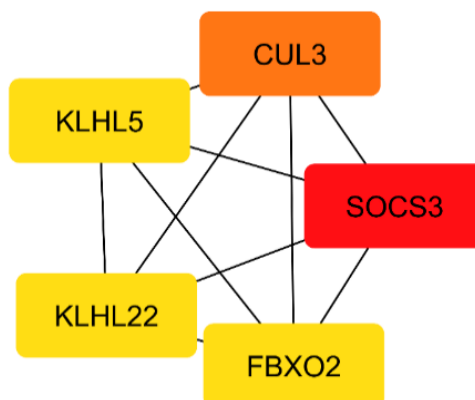
**Tabela 5** - Vias KEGG enriquecidas nos módulos da rede IPP de ATLL.

| <b>Módulo</b> | <b>Vias enriquecidas</b>  | <b>FDR</b>          | <b>Genes</b>  |
|---------------|---|---------------------|---|
| <b>ATLL1</b>  | Proteólise mediada pela ubiquitina (hsa04120)                       | 1,23 <sup>-12</sup> | <i>KLHL9;DET1;SOCS3;SMURF2;FBXO2;UBE2E3;CUL3;NEDD4L;UBE2A</i> |
| <b>ATLL2</b>  | Endocitose (ko04144)  | 6,01 <sup>-04</sup> | <i>BIN1; AP2S1; ARRB1; AP2B1; LDLRAP1</i>                     |
| <b>ATLL3</b>  | Via de sinalização de quimiocinas (hsa04062)                        | 5,48 <sup>-06</sup> | <i>CXCR3;CCR8;ADCY1;CCR5;CCR4</i>                             |
|               | Interação entre receptores citocinas-citocinas (hsa04060)           | 1,38 <sup>-03</sup> | <i>CXCR3;CCR8;CCR5;CCR4</i>                                   |
|               | Infecção pelo citomegalovírus humano (hsa05163)                     | 8,16 <sup>-03</sup> | <i>PTGER3;ADCY1;CCR5</i>                                      |
|               | Via de sinalização AMPc (hsa04024)                                  | 8,22 <sup>-03</sup> | <i>POMC;PTGER3;ADCY1</i>                                      |
|               | Carcinogênese viral (hsa05203)                                      | 8,77 <sup>-03</sup> | <i>CCR8;CCR5;CCR4</i>   |
|               | Infecção pelo herpesvírus associado ao sarcoma de Kaposi (map05167) | 9,29 <sup>-03</sup> | <i>CCR8;CCR5;CCR4</i>   |
|               | Regulação da lipólise em adipócitos (hsa04923)                      | 1,45 <sup>-02</sup> | <i>PTGER3;ADCY1</i>   |
|               | Síntese e secreção de cortisol (ko04927)                            | 1,77 <sup>-02</sup> | <i>POMC;ADCY1</i>   |
|               | Interação ligante-receptor neuroativo (hsa04080)                    | 1,80 <sup>-02</sup> | <i>POMC;GRM3;PTGER3</i>                                       |
|               | Melanogênese (hsa04916)   | 3,10 <sup>-02</sup> | <i>POMC;ADCY1</i>   |
|               | Síntese e secreção de aldosterona (hsa04925)                        | 3,21 <sup>-02</sup> | <i>POMC;ADCY1</i>   |
|               | Sinapse glutamatérgica (hsa04724)                                   | 3,61 <sup>-02</sup> | <i>GRM3;ADCY1</i>   |
|               | Via de sinalização de estrogênio (hsa04915)                         | 4,79 <sup>-02</sup> | <i>POMC;ADCY1</i>   |

ATLL - Leucemia/linfoma de células T do Adulto. FDR – *False Discovery Rate*. IPP – Interação Proteína-Proteína. KEGG – *Kyoto Encyclopedia of Genes and Genomes*.

A partir da rede IPP construída inicialmente para os dados de GDEs em ATLL, também foi possível identificar os cinco genes hub com maior score:

**Figura 9** - Rede dos top cinco genes hubs de ATLL.

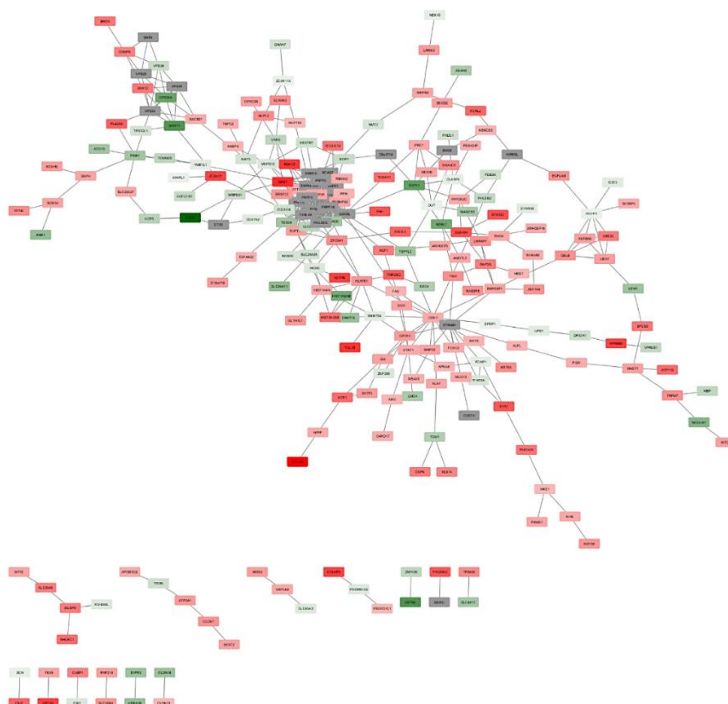


Extraídos com CytoHubba, no Cytoscape v3.7.1. Quanto mais escura a cor, maior o score. ATLL - Leucemia/linfoma de células T do Adulto.

#### 4.4.2 HAM/TSP

A lista dos GDEs superexpressos e subexpressos em indivíduos com HAM/TSP gerou uma rede IPP com 359 nós e 478 arestas ( $p = 3,93^{-06}$ ) (**Fig. 10**).

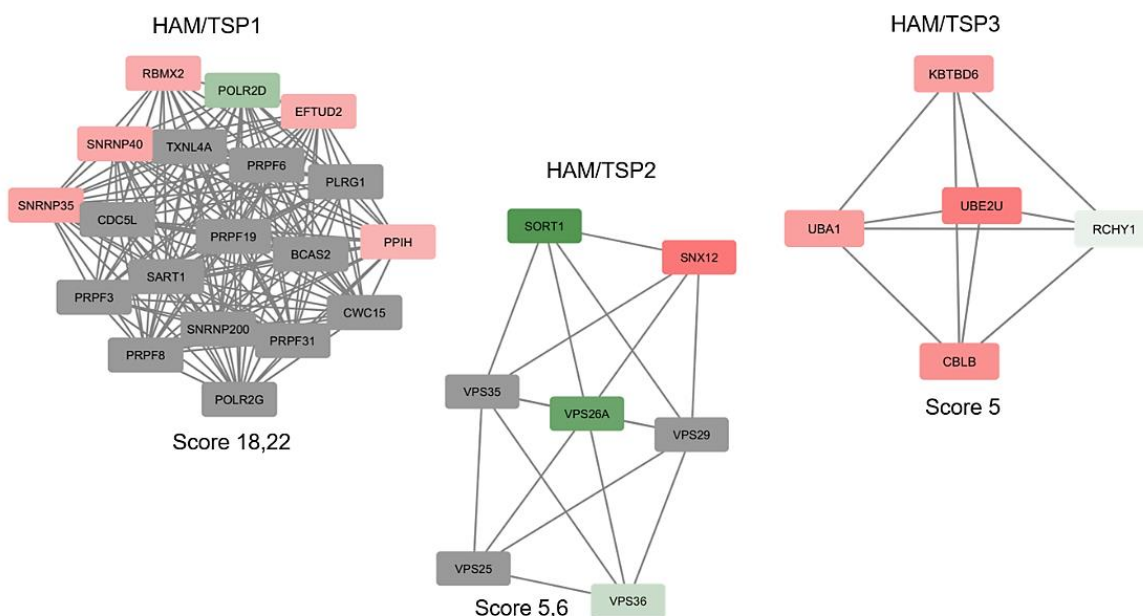
**Figura 10** - Rede de Interação Proteína-Proteína de GDEs em HAM/TSP.



Construída no Cytoscape v3.7.1 com base no banco de dados do STRING v11. Cores vermelhas representam superexpressão, cores verdes representam subexpressão e cor cinza representa genes preditos. HAM/TSP - Mielopatia Associada ao HTLV-1/Paraparesia Espástica Tropical.

Nesse caso, o *plugin* MCODE extraiu, a partir da IPP construída neste grupo, um total de 7 módulos, porém, por motivo de simplificação e grau de importância, foram selecionados os módulos com  $score \geq 5$  (**Fig. 11**).

**Figura 11** - Módulos da rede Interação Proteína-Proteína de HAM/TSP.



Extraídos com MCODE, no *Cytoscape* v3.7.1. Cores vermelhas representam superexpressão, cores verdes representam subexpressão e cor cinza representa genes preditos.

Em análise dos três módulos identificados anteriormente e selecionados da rede IPP de HAM/TSP, também foi possível notar o perfil de ocorrência de diferentes vias enriquecidas (**Tabela 6**).

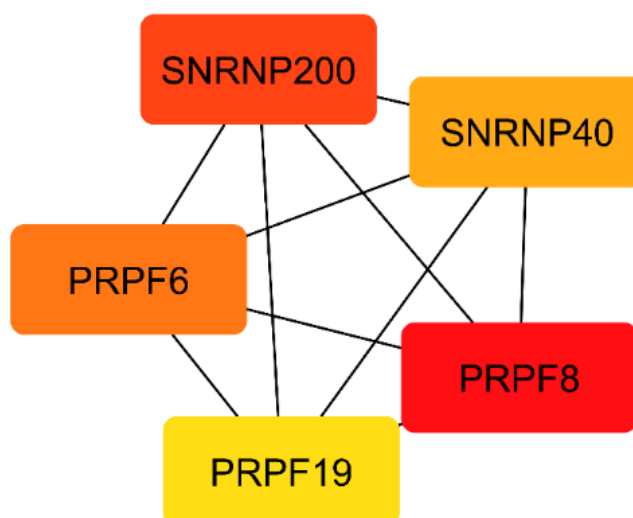
**Tabela 6** - Vias KEGG enriquecidas nos módulos da rede IPP de HAM/TSP.

| Módulo          | Vias enriquecidas                             | FDR                 | Genes  |
|-----------------|---|---------------------|--|
| <b>HAM/TSP1</b> | Spliceossomo (hsa03040)                       | 1,28 <sup>-24</sup> | <i>CDC5L;CWC15;PLRG1;PRPF19;PRPF8; EFTUD2; SART1;SNRNP40;PRPF6; PRPF3;IPPH;PRPF31; SNRNP200;TXNL4A;BCAS2</i> |
| <b>HAM/TSP2</b> | Endocitose (ko04144)                          | 6,62 <sup>-06</sup> | <i>VPS29;SNX12;VPS26A; VPS35;VPS36;VPS25</i>   |
| <b>HAM/TSP3</b> | Proteólise mediada pela ubiquitina (hsa04120) | 3,20 <sup>-06</sup> | <i>UBE2U;CBLB;UBA1;RCHY1</i>   |

FDR – *False Discovery Rate*. HAM/TSP - Mielopatia Associada ao HTLV-1/Paraparesia Espástica Tropical. IPP – Interação Proteína-Proteína. KEGG – *Kyoto Encyclopedia of Genes and Genomes*.

A partir da rede de Interação Proteína-Proteína total do grupo de indivíduos com HAM/TSP, a análise de genes hubs revelou genes pertencentes a duas famílias gênicas distintas: PRPF (*Pre-mRNA Processing Factor*) e SNRN (*Small Nuclear RiboNucleoproteins*), sendo os cinco GDEs com maior score na análise (**Fig. 12**).

**Figura 12** - Rede dos top cinco genes hubs em indivíduos HAM/TSP.

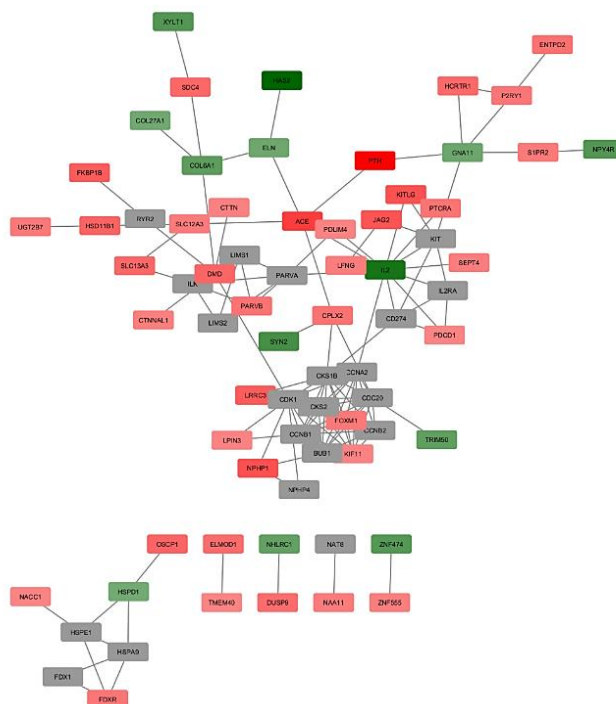


Extraídos com CytoHubba, no *Cytoscape* v3.7.1. Quanto mais escura a cor, maior o score. HAM/TSP - Mielopatia Associada ao HTLV-1/Paraparesia Espástica Tropical.

#### 4.5.3 ASS

A lista dos GDEs superexpressos e subexpressos em indivíduos ASS gerou uma rede IPP ( $p = 0,03$ ) com 146 nós e 128 arestas (**Fig. 13**).

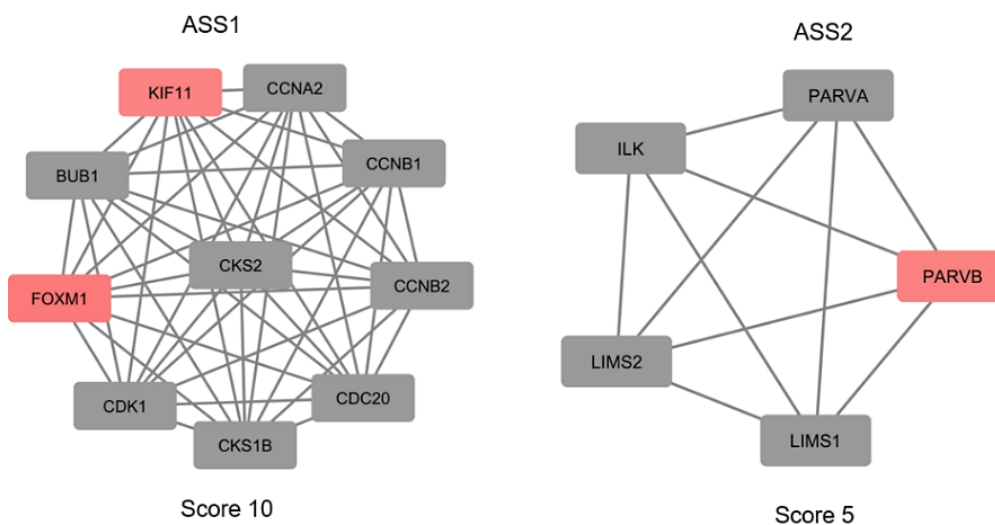
**Figura 13** - Rede de Interação Proteína-Proteína de GDEs em ASS.



Construída no *Cytoscape* v3.7.1 com base no banco de dados do STRING v11. Cores vermelhas representam superexpressão, cores verdes representam subexpressão e cor cinza representa genes preditos.

A partir da rede IPP mostrada anteriormente, o *plugin* MCODE extraiu 4 módulos, a partir dos quais, aqueles que apresentaram  $\text{score} \geq 5$  (**Fig. 14**), foram selecionados, por serem mais relevantes, e por ser uma forma de filtrar os dados apresentados.

**Figura 14** - Módulos da rede de Interação Proteína-Proteína de ASS.



Extraídos com MCODE, no *Cytoscape* v3.7.1. Cor vermelha representa superexpressão e cor cinza representa genes preditos. ASS – Assintomáticos.



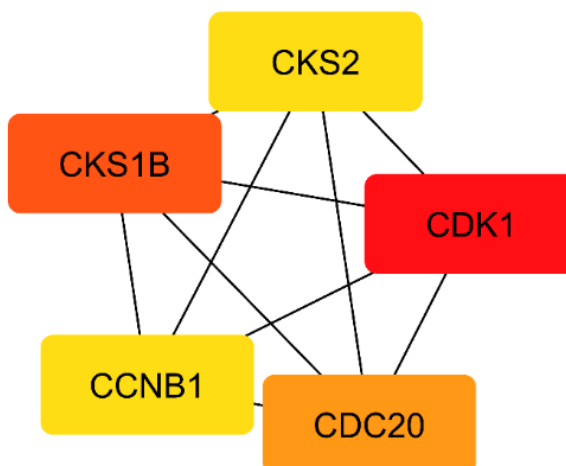
Dentre os três módulos selecionados, com *score*  $\geq 5$ , da rede IPP de indivíduos ASS, foi possível notar diferentes perfis no que diz respeito à formação e função de vias enriquecidas (**Tabela 7**).

**Tabela 7** - Vias KEGG enriquecidas nos módulos da rede IPP de indivíduos ASS.

| <b>Módulo</b> | <b>Vias enriquecidas</b>                                  | <b>FDR</b>   | <b>Genes</b>                                  |
|---------------|---|--------------|---|
| <b>ASS1</b>   | Ciclo celular (hsa04110)                                  | $3,19^{-06}$ | <i>CDC20; CCNA2; CCNB2; CCNB1; CDK1; BUB1</i> |
|               | Maturação de ovócitos mediada por progesterona (hsa04914) | $1,02^{-04}$ | <i>CCNA2; CCNB2; CCNB1; CDK1; BUB1</i>        |
|               | Meiose do oócito (hsa04114)                               | $2,22^{-04}$ | <i>CDC20; CCNB2; CCNB1; CDK1; BUB1</i>        |
|               | Senescência celular (hsa04218)                            | $5,78^{-04}$ | <i>CCNA2; CCNB2; CCNB1; CDK1; FOXM1</i>       |
|               | Via de sinalização p53 (hsa04115)                         | $3,25^{-04}$ | <i>CCNB2; CCNB1; CDK1</i>                     |
|               | Infecção pelo HTLV-1 (hsa05166)                           | $5,65^{-03}$ | <i>CCNB2; CCNB1; CDK1</i>                     |
|               | Carcinogênese viral (hsa05203)                            | $5,85^{-03}$ | <i>CDC20; CCNA2; CDK1</i>                     |
|               | Infecção pelo HIV-1 (hsa05170)                            | $5,87^{-03}$ | <i>CCNB2; CCNB1; CDK1</i>                     |
|               | Câncer de pulmão de pequenas células (hsa05222)           | $3,22^{-02}$ | <i>CKS2; CKS1B</i>                            |
| <b>ASS2</b>   | Adesão focal (hsa04510)                                   | $2,95^{-03}$ | <i>ILK; PARVA; PARVB</i>                      |

ASS – Assintomáticos. FDR – *False Discovery Rate*. IPP – Interação Proteína-Proteína. KEGG – *Kyoto Encyclopedia of Genes and Genomes*.

No que diz respeito, à identificação de genes hubs, também foram identificados cinco genes hubs para os dados de expressão dos indivíduos ASS (**Fig. 15**).

**Figura 15** - Rede dos top cinco genes hubs de indivíduos ASS.

Extraídos com CytoHubba, no *Cytoscape* v3.7.1. Quanto mais escura a cor, maior o score. Quanto mais escura a cor, maior o score. ASS – Assintomáticos.

#### 4.6 ANÁLISE DE MICRO-RNAS

Para análise de MDEs, primeiramente fez-se uma busca de sondas de miRNAs nas plataformas de microarranjo utilizadas pelos estudos individuais que haviam sido selecionados para as três metanálises descritas anteriormente (**Quadro 1**) com o objetivo de ter conhecimento acerca da cobertura desse tipo de análise quanto à quantidade de miRNAs analisados. A **Tabela 8** apresenta a quantidade de sondas referentes às moléculas de miRNAs encontradas nas referidas plataformas, apenas a plataforma GPL9686, utilizada pelo estudo GSE19080 não continha sondas desse tipo. Ao total, eliminando as sondas repetidas entre as plataformas, foi possível identificar 1,847 sondas de miRNAs para todas as manifestações clínicas.

**Tabela 8** - Sondas referentes às moléculas de miRNAs presentes nas plataformas dos ensaios de expressão utilizados na metanálise.

| Série   | GPL  | miRNAs |
|---|--|--------|
| GSE29312, GSE33615, GSE17718, GSE14317, GSE43017, GSE55851, GSE29332, GSE1466 | GPL6947, GPL4133, GPL570, GPL571, GPL10332, GPL6884, GPL9696 | 1,847  |
| GSE38537  | GPL6480  | 3      |
| GSE19080  | GPL9686  | 0      |

GPL – Código da plataforma; miRNAs – Micro-RNAs

No entanto, apenas os indivíduos com ATLL apresentaram MDEs, dentre os GDEs identificados nas metanálises anteriores. Em indivíduos HAM/TSP e ASS não foi possível identificar miRNAs diferencialmente expressos de acordo com os parâmetros adotados. Por outro lado, cinco miRNAs foram considerados DEs em ATLL (**Quadro 2**). De acordo com o mirBase, o maior banco de dados públicos de miRNAs, apenas dois desses MDEs possuem uma alta confiabilidade da anotação, sendo eles o hsa-mir-21 e o hsa-mir-6840.

**Quadro 2** - Caracterização das moléculas de miRNAs diferencialmente expressos em indivíduos com ATLL.

| Micro-RNAs   | Confiança da anotação | Mudança       |
|--------------|-----------------------|---------------|
| hsa-mir-3658 | Sem dados suficientes | superexpresso |
| hsa-mir-21   | Alta                  | superexpresso |
| hsa-mir-631  | Sem dados suficientes | subexpresso   |
| hsa-mir-600  | Sem dados suficientes | subexpresso   |
| hsa-mir-6840 | Alta                  | subexpresso   |

ATLL - Leucemia/linfoma de células T do Adulto. miRNAs – Micro-RNAs

Uma vez mantidos o hsa-mir-21 e o hsa-mir-6840 para as análises seguintes, uma busca dos genes alvo desses miRNAs na ferramenta *online* miRWalk 3.0 (STICHT et al., 2018) foi realizada. Essa busca revelou existirem 219 genes alvo experimentalmente validados para o hsa-mir-21. Desses 219 genes alvo, 12 foram considerados subexpressos em indivíduos com ATLL (**Quadro 3**). O hsa-mir-6840, por sua vez, possui 413 genes alvo validados, desses, 11 foram considerados superexpressos em ATLL (**Quadro 3**).

**Quadro 3** - GDEs considerados possíveis genes alvo de MDEs em ATLL.

| MDE          | Genes alvo diferencialmente expressos em ATLL   |
|--------------|---|
| hsa-mir-21   | <i>PLEKHA1, SLC16A10, ABCB1, DOCK10, RAPGEF6, SATB1, SNRK, AGO2, RECK, NIN, SPTLC3, STAG2</i> |
| hsa-mir-6840 | <i>KLF8, CCR4, TTC22, GDE1, GM2A, SPATS2L, SLC9A7, TBC1D19, CYCS, CHP1, GNS</i>               |

ATLL - Leucemia/linfoma de células T do Adulto. MDE – Micro-RNA Diferencialmente Expresso

## 5. DISCUSSÃO

Uma metanálise contemplando 10 ensaios de expressão gênica, utilizando microarranjo como metodologia, em pacientes infectados pelo HTLV-1, cujos dados foram disponibilizados publicamente no repositório do GEO, foi realizada para identificar genes desregulados no hospedeiro. Tais Genes Diferencialmente Expressos (GDEs), formam perfis de expressão gênica diferenciais em três status clínicos da infecção pelo HTLV-1: a ATLL (leucemia/linfoma), a síndrome conhecida como HAM/TSP (Quadro inflamatório do sistema nervoso central), e o estado assintomático (Que não reflete a completa ausência de sintomas, e sim ausência das outras duas condições clínicas). Este último estado parece representar a maioria dos casos na infecção por esse retrovírus, pelos escassos estudos que investigam esta ocorrência.

Os principais resultados descritos nos estudos selecionados para a metanálise evidenciam alterações na expressão gênica de diferentes formas e com diferentes consequências. O estudo de Choi e cols. (2007) com células CD4<sup>+</sup> de pacientes ATLL sugeriu que a via de sinalização que envolve o HGF (*Hepatocyte Growth Factor*) e seu receptor, por estarem superexpressas, poderiam ser utilizadas como alvo terapêutico. Pise-Masison e cols. (2009) revelaram a importância da Survivina, uma molécula anti-apoptótica, na patogênese da ATLL. No artigo de Olié e colaboradores (2011) é destacada a importância de *SOCS1* (*Suppressor of cytokine signaling 1*) para a evasão do HTLV-1 ao interferon do tipo 1 em pacientes com HAM/TSP e pacientes assintomáticos. Uma elevação na expressão de AMPc regulada pela superexpressão do gene da fosfodiesterase 3 foi descrita em linfócitos T transformados pelo HTLV-1, por Kress e cols. (2010). O trabalho de Tattermusch e cols. (2012) revelou perturbações na via do p53 na infecção pelo HTLV-1 assim como uma assinatura de expressão gênica regulada por interferon em pacientes HAM/TSP. O trabalho de Yamagishi e cols. (2012) fez uma relação entre a perda do MIR-31 e a ativação da via NF-kB dependente de NIK (do inglês: *NF-kB-inducing kinase*) em ATLL. Pinto e cols. (2014), ao investigar a expressão gênica global em células CD4<sup>+</sup> de pacientes HAM/TSP e assintomáticos, evidenciaram a expressão diferencial de genes relacionados à atividade antiviral: migração celular e lise celular. Nakahata e cols. (2014) sugeriram que a subexpressão de *NDRG2* (*NMYC downstream-regulated gene 2*) em ATLL resulta em um aumento na fosforilação de PTEN (*Phosphatase and tensin homolog*), uma proteína supressora tumoral. Por fim, Kobayashi e cols. (2014)

relacionaram a superexpressão de *CADM1* com a expansão clonal de células infectadas pelo HTLV-1 em pacientes com ATLL.

### 5.1 O ARTIFÍCIO DA METANÁLISE

Com a finalidade de abarcar boa parte dos objetivos e propostas de investigação científicos, foram desenvolvidos ao longo dos anos diferentes metodologias para realização de metanálises. Existem três tipos de comparação de dados para metanálise de expressão gênica: através valor de  $p$ , tamanho de efeito (*effect size*) e *combined ranks*. A técnica do valor de  $p$ , que é mais utilizada quando deseja-se obter resultados mais conservativos, envolve submétodos como *Fisher*, *Z-value*, *Stouffer*, método do valor mínimo de  $p$ , o método do valor máximo de  $p$  e valor de  $p$  ordenado por *rth*. *Combined ranks* e seus submétodos (RankProd, RankSum, produto dos *ranks* e soma dos *ranks*) levam em conta a posição de determinado gene em um *rank*. Se sua posição se repete, ela é mantida (CHANG et al., 2013; SIANGPHOE et al., 2019).

A comparação por combinação de tamanho de efeito foi utilizada neste trabalho por ser a mais recomendada para comparação de estudos entre duas classes (controle e teste) (RAMASAMY et al., 2007). O tamanho de efeito enfatiza o tamanho da diferença ao invés da significância estatística. A combinação desse efeito não pode ser realizada através de uma simples média devido às diferenças de precisão entre os estudos, portanto (CHOI et al., 2003). Devido a esse fato, o tamanho de efeito pode ser dividido em efeitos fixos e efeitos aleatórios. O primeiro assume que existe um tamanho de efeito verdadeiro entre todos os estudos e então deve ser aplicado em uma comparação de microarranjos que passaram por processos, metodologias e variáveis com um nível de similaridade alto. O segundo é aconselhável para o caso oposto, onde há uma variedade de procedimentos entre os microarranjos e variáveis não comparáveis (TSENG et al., 2012). Proposto por Choi e cols. (2003), para efeito aleatório, o primeiro passo é calcular o tamanho de efeito e a variância associada com o tamanho de efeito para cada gene em todos os estudos.

Ao final de uma metanálise, são obtidos os GDEs e, a partir deles, o *Integration Driven Discovery* (IDD) poderá ser determinado. O IDD revela os genes que foram considerados não significantes nos estudos individuais e significantes na metanálise. Isso ocorre devido à junção dos “pequenos, porém consistentes” tamanhos de efeito, tratados como falso negativos nos estudos individuais. Esse poder estatístico é efeito

do aumento do tamanho da amostra. Além disso, o erro diminui e, conseqüentemente, a quantidade de falsos positivos também (CHOI et al., 2003). Conjuntamente ao IDD, a metanálise gera o valor da perda, ou seja, a quantidade de genes considerados diferencialmente expressos nos estudos individuais, mas não foram considerados na metanálise.

A metanálise de dados de expressão gênica pode gerar novos GDEs em comparação com aqueles gerados pelos estudos individuais (EI) e pode também não considerar alguns genes ditos diferencialmente expressos nos EI. A **Tabela 3**, por exemplo, apontou que o grupo de indivíduos com HAM/TSP teve a maior quantidade de GDEs novos (260) em relação aos EI, portanto, apresentou a maior taxa de descoberta (IDR) (70,65%). Nos dados dos indivíduos com ATLL, apesar de ter a menor IDR, teve a maior taxa de revisão (IRR) (163,22%) com mais de 10 mil genes considerados falsos positivos nos EI. Uma menor taxa de descoberta para ATLL pode ser atribuída a uma maior amostragem em seus EI, refletindo em uma maior cobertura. Por outro lado, os EI conduzidos em indivíduos com HAM/TSP se baseiam em um menor número de amostras e conseqüentemente o poder da metanálise é maior ao conseguir combinar todos esses estudos.

## 5.2 A DESREGULAÇÃO DIFERENCIAL NOS STATUS CLÍNICOS ASSOCIADOS AO HTLV-1

O perfil de expressão gênica diferencial revelou uma maior desregulação no perfil clínico ATLL seguida pelo perfil clínico HAM/TSP e estado assintomático, evidenciando uma maior alteração na expressão gênica em indivíduos com ATLL em comparação com indivíduos com HAM/TSP. O resultado obtido em indivíduos ASS indica alterações na expressão gênica mesmo sem maiores manifestações fenotípicas. Os três perfis de expressão apresentaram genes compartilhados entre os diferentes status clínicos. O único GDE compartilhado por todas as condições clínicas foi o gene da Anquirina-1 (**Fig. 6 (a)**). Essa família de proteínas intracelulares é responsável por organizar, transportar e ancorar complexos proteicos da membrana à filamentos de actina, geralmente envolvendo canais iônicos, transportadores e moléculas de adesão molecular (CUNHA & MOHLER, 2006). Estudos indicam que a superexpressão de *ANK1* está relacionada com o crescimento de células cancerígenas do pâncreas (OMURA et al., 2016). Os GDEs compartilhados pelo grupo de indivíduos com ATLL e pelo grupo de indivíduos com HAM/TSP participam de

funções e processos diversos, envolvendo em sua maioria fatores de transcrição (*LMO4*, *BATF3*, *HNF4G*, *RAI1*, *HIC2*, *ZNF407*) evidenciando o caráter comum de alteração na regulação da expressão gênica nas duas condições. Além, também, de genes envolvidos na via endocítica (*CHMP6*, *SFT2D3*, *KIF3B*) e outros. Dentre os compartilhados por indivíduos com ATLL e indivíduos ASS aparecem os modeladores do citoesqueleto *POF1B* e *PARVB*. Entre em indivíduos ASS e indivíduos com HAM/TSP são compartilhados três genes diferencialmente expressos com funções distintas, o *NHLRC1* da via de ubiquitinação, o *DUSP9*, que produz uma fosfatase, e *ZNF474*, um fator de transcrição. A lista completa desses genes pode ser consultada no **Apêndice B**.

Separando os GDEs em superexpressos (**Fig. 6 (b)**) e subexpressos (**Fig. 6 (c)**) e analisando o compartilhamento de genes, percebe-se que dentre os grupos de GDEs superexpressos o compartilhamento é maior, principalmente entre ATLL e HAM/TSP, provavelmente devido a uma maior quantidade de GDEs superexpressos quando comparada à quantidade de GDEs subexpressos. Nenhum GDE é compartilhado por todos os status clínicos nessas condições já que *ANK1* mantém a superexpressão em ATLL e ASS e subexpressão em HAM/TSP.

Dentre os genes mais superexpressos em indivíduos com ATLL está o gene *CADM1* que codifica a molécula de adesão celular 1 (**Tabela 2**). Em um dos trabalhos inclusos na metanálise, esse gene já havia sido descrito como superexpresso e relacionado com a expansão clonal de células infectadas pelo HTLV-1 em pacientes com ATLL (KOBAYASHI et al., 2014). Esse gene produz uma proteína de adesão que está superexpressa na superfície da célula infectada, já sendo sugerido como biomarcador para infecção pelo HTLV-1, especialmente em indivíduos com ATLL (MANIVANNAN et al., 2016). *CADM1* inativa reguladores negativos da via do NF- $\kappa$ B, consequentemente mantendo a ativação persistente dessa via em células infectadas pelo HTLV-1 (PUJARI et al., 2015). Entre os genes subexpressos, para este mesmo grupo clínico, está o *MAN1C1* ( $\alpha$ -manosidase IC), que codifica uma  $\alpha$ -manosidase que cataliza a hidrólise de resíduos de  $\alpha$ -manoses. Estudos indicam que esses glicídios auxiliam a adesão e rolamento de monócitos em células epiteliais (CHACKO et al., 2011). Em ATLL uma subexpressão de *MAN1C1* pode estar relacionada com a manutenção desses glicanos na superfície celular das células infectadas para promoção da migração celular.

Em indivíduos com HAM/TSP, destaca-se a superexpressão de *RAVER2*

(Ribonucleoproteína Ligadora de PTB 2) (**Tabela 2**), gene que codifica um regulador negativo da Proteína de Ligação a Polipirimidina (PTB) e provavelmente está relacionado com a regulação do splicing alternativo (BARTOLETTI-STELLA et al., 2015). A superexpressão desse gene foi associada a uma alteração no processamento de mRNA em Leucodistrofia, uma doença desminealizante e, apesar da expressão de *RAVER2* ser ubíqua, é altamente expressa no sistema nervoso central (BARTOLETTI-STELLA et al., 2015). Por outro lado, *SORT1*, gene da Sortilina 1, receptor responsável pelo tráfico de vesículas mediando a distribuição de proteínas nas vias endocíticas, está subexpresso neste grupo clínico. Essa proteína regula vias celulares distintas em vários tecidos e tipos celulares, incluindo a regulação da viabilidade e função neuronal (NYKJAER & WILLNOW, 2012).

Em indivíduos ASS, por sua vez, um dos cinco genes mais superexpressos é o *HORMAD2* (contendo domínio HORMA 2) (**Tabela 2**), que regula o ciclo celular. A superexpressão de *HORMAD2* foi associada a um prolongamento da fase G0/G1 e ao encurtamento da fase S para impedir a mitose celular, bem como promover a apoptose das células cancerígenas da tireoide (LIN et al., 2018). Interessantemente, um dos genes mais subexpressos no grupo de indivíduos ASS é um lncRNA (*Long non-coding RNA*), RNAs longos, não codificantes, que regulam a expressão gênica ativando ou reprimindo os genes (DENIZ & ERMAN, 2017). Uma função conhecida de LINC00554 é de regular positivamente a expressão de *MARCKS* (Substrato C-quinase Rico em Alanina Miristoilada) e *MARCKSL1* (*MARCKS-like1*), genes que codificam proteínas que translocam da membrana para o citosol, modulando a dinâmica da actina citoesquelética e o tráfego vesicular (ELAMRI et al., 2018).

#### 5.4 VIAS ENRIQUECIDAS NO CONJUNTO TOTAL DE GDEs

Outro processo comumente empregado é o estudo de vias enriquecidas. Esse estudo envolve a aplicação de métodos estatísticos para identificar vias que são significativamente representadas em uma lista de genes de interesse, eliminando o que seria esperado por mera chance. A lista de genes é testada com todas as vias de determinado banco de dados através de diversos métodos estatísticos gerando, ao final, vias significantes sugeridas para a lista de genes (CHEN et al., 2013). A análise de vias enriquecidas facilita o estudo molecular por reduzir grandes números de genes a vias normais ou patológicas facilitando a interpretação dos resultados (REIMAND et al., 2019).



O resultado mostrou vias enriquecidas somente para os dados de expressão de indivíduos com ATLL. Alguns motivos podem levar a esse resultado negativo para indivíduos com HAM/TSP e assintomáticos, como a falta de análise para genes sem anotação, vias onde a atividade é controlada por poucos genes ou não controlada por expressão gênica (como regulação pós-transcricional) (REIMAND et al., 2019). “Vias em câncer” foi a via encontrada enriquecida para ATLL (**Tabela 4**), o que era esperado visto que ATLL é um tipo de câncer, alterando assim a expressão de vários genes relacionados à essa doença. A maioria do número de genes nessa via está relacionada a evasão da apoptose, proliferação celular e angiogênese continuada, processos que são fundamentais para permitir que as células cancerosas cresçam e dividam de forma não controlada, invadindo tecidos normais, órgãos e consequentemente se espalhando pelo corpo (MAREEL & LEROY, 2003).

O conjunto de genes subexpressos indicou “Endocitose” como via enriquecida (**Tabela 4**) o que sugere que essa via está em menor atividade em células de indivíduos com ATLL. A endocitose é um processo que objetiva a obtenção de moléculas presentes no lado externo ou na membrana plasmática para internalização, reciclagem ou degradação. Envolve múltiplas vias e a formação de vesículas membranosas internas que agem como transportadoras de cargas como receptores, transportadores e moléculas de adesão (CONNER & SCHMID, 2003). A endocitose pode ser mediada por clatrina, caveolina ou independentemente dessas duas proteínas estruturais que auxiliam na formação da fossa membranosas. Após o estímulo, acontece a invaginação para o interior celular tornando-se uma vesícula intracelular que funde-se à endossomos iniciais onde acontece o direcionamento das proteínas. Então, após transformações químicas e estruturais, maturam-se nos endossomos tardios que eventualmente se fundem ao lisossomo (SCHMID, 2017).

Da mesma maneira em que a endocitose é um mecanismo conhecido por diminuir a ativação de vias por internalizar receptores, sabe-se também que ela é necessária para iniciar algumas outras vias (SCHMID, 2017). Além disso, muitos estudos mostram como a endocitose pode estar desregulada em células cancerígenas (POLO et al., 2004). A atividade anormal da endocitose de proteínas de membrana contribui para transformação das células normais em aberrantes (MOESSON et al., 2008). Esses resultados mostram, pela primeira vez, uma desregulação na via endocítica em indivíduos com ATLL, sugerindo que há uma importância na diminuição da atividade dessa via para o desenvolvimento da

doença. Sendo portanto, pela análise de vias enriquecidas, os primeiros biomarcadores a serem sugeridos para a ATLL: superexpressão de genes relacionados à apoptose, proliferação celular e angiogênese e subexpressão de genes envolvidos com a endocitose, sobretudo porque estes genes/funções também foram sinalizados pela análise das redes de interação proteína-proteína.

## 5.5 REDES IPP REVELAM MÓDULOS ENRIQUECIDOS PARA VIAS BIOQUÍMICAS

Proteínas podem interagir fisicamente com outras proteínas. Uma rede de interação proteína-proteína (IPP) permite ilustrar a relação física que ocorre entre elas. Assim, esta é uma maneira de analisar processos biológicos a nível molecular. Essas redes são dinâmicas e reagem a diferentes estímulos e alterações do ambiente. Até pequenas mudanças nessas redes podem ocasionar fenótipos patológicos (BARABÁSI et al., 2011). A análise das redes IPP possui um importante papel na investigação de processos biológicos complexos, além disso, pode ser útil para descoberta de novas funções proteicas e no desenvolvimento de novas terapias para doenças a partir da análise dos genes hub (proteínas com alta conectividade em uma rede) e da descoberta de módulos a partir das redes gênicas que contribuem para o desenvolvimento da doença (EMILSSON et al., 2008). A respeito das funções biológicas das proteínas, suas interações determinam mecanismos celulares e moleculares e, baseado nessas interações, direcionam-se terapias multi-alvo para agir não só em um gene mas em uma rede de determinada doença multi-gênica complexa (SAFARI-ALIGHIARLOO et al., 2014). Devido à multifuncionalidade e importância das proteínas entende-se a importância dessa análise.

Os GDEs dos três status clínicos, em análise, apontaram a formação de redes IPP significantes. A rede formada com os dados de expressão de indivíduos com ATLL mostrou-se altamente densa e interconectada, enquanto que para as redes de indivíduos com HAM/TSP e ASS foi necessário adicionar proteínas com interações preditas pelo STRING, ou seja, fora dos conjuntos de genes diferencialmente expressos definidos na metanálise, com o objetivo de atingir significância estatística. Os módulos extraídos das três redes IPP foram analisados quanto a presença de vias enriquecidas.

### 5.5.1 ATLL

A ubiquitinação de proteínas, via encontrada como altamente enriquecida no módulo **ATLL1 (Fig. 8)**, é um processo que ocorre em várias etapas e é responsável pela reciclagem de proteínas celulares através do encaminhamento dessas proteínas para o proteassoma. Brevemente, envolve o complexo E1-ativador, que ativa a ubiquitina, E2-conjugador, que recebe a ubiquitina ativada do E1 e transfere para as E3-ligases que ligam a ubiquitina ativada à proteína a ser degradada no proteassoma (YAMAO, 1998). Dentre os GDEs de indivíduos com ATLL nessa via temos *UBE2E3*, participantes da etapa E2 e *KLHL9*, *DET1*, *SOCS3*, *SMURF2*, *FBXO2*, *UBE2E3*, *CUL3*, *NEDD4L*, *UBE2A* da etapa E3 (**Tabela 5**). Assim sendo, a proteólise mediada por ubiquitina influencia vários processos celulares como localização proteica, controle do ciclo celular, regulação da transcrição e controle do receptor por endocitose (SHI & GROSSMAN, 2010; YAMAO, 1998). Além disso, essa via também está envolvida em outros processos como a ativação da resposta inflamatória via NF- $\kappa$ B, a qual está altamente ativada em indivíduos com ATLL devido à indução por Tax e acúmulo de mutações genéticas, como já é descrito na literatura (CHEN & CHEN, 2013; KATAOKA et al., 2015). Células cancerígenas exploram a maquinaria de ubiquitinação para estabilizar a sinalização oncogênica aberrante para sobrevivência e metástase, conseqüentemente, alguns estudos já abriram caminho para a utilização de fármacos, como Bortezomibe e Carfilzomib, que miram componentes da maquinaria de ubiquitinação em cânceres (GALLO et al., 2017).

Os módulos **ATLL2** e **ATLL3**, apesar de apresentarem vias enriquecidas com valor de FDR abaixo de 0,05, não são tão significantes quanto a via enriquecida encontrada em **ATLL1**. Muitas proteínas são compartilhadas entre as vias obtidas para esses dois módulos (**Tabela 5**). No módulo **ATLL2**, enriquecido para “Endocitose”, todos os GDEs estão envolvidos na endocitose mediada por clatrina mas apenas os genes da família AP2 (*Adaptor Protein Complex 2*), envolvidos no recrutamento de clatrina para a formação da fossa que objetiva a internalização da carga a ser endocitada (SCHMID, 2017), estão superexpressos (**Fig. 8**). Portanto, pode haver uma diminuição da atividade dos receptores de endocitose na membrana plasmática celular em indivíduos com ATLL ao passo que a formação do fosso endocítico em si poderá estar mais ativo. Dentre as 13 vias fracamente enriquecidas no módulo **ATLL3**, percebe-se a indicação de vias relacionadas à infecção (“Infecção pelo citomegalovírus humano”, “Infecção pelo herpesvírus associado ao sarcoma de

Kaposi”), sinalização (“Via de sinalização de quimiocinas”, “Via de sinalização via AMPc”, “Via de sinalização de estrogênio”), síntese (“Síntese e secreção de cortisol”, “Síntese e secreção de aldosterona”, “Melanogênese”) e outras (**Tabela 5**). A via de sinalização de quimiocinas envolve genes da família CCR (*Chemokine Receptors*), receptores de quimiocinas que, como efeito final, podem estimular a migração de células e a quimiotaxia, principalmente em leucócitos (HUGHES & NIBBS, 2018). O gene *CCR4* participante da via das quimiocinas, e encontrado superexpresso neste grupo clínico ( $\log_{10}FC = 0,5566$ ,  $FDR = 1,21^{-08}$ ) foi recentemente encontrado com mutações de *frameshift* em pacientes com ATLL e associadas à um pior prognóstico da doença (NAKAGAWA et al., 2014; YOSHIDA et al., 2016). Devido à importância dessa proteína na patologia da ATLL, foi desenvolvido o Mogamulizumab, um anticorpo anti-CCR4 para o tratamento da mesma (MAKITA & TOBINAI, 2017).

Genes hub são aqueles que apresentam alta conectividade em uma rede gênica. Uma alta conectividade representa a importância desses genes para a manutenção da rede. A proteína produzida pelo gene *SOCS3* (*Suppressor of cytokine signaling-3*) foi a que obteve o maior score para essa análise em indivíduos com ATLL (**Fig. 9** e **Apêndice C**). Essa proteína é conhecida por regular negativamente a sinalização de citocinas e está subexpressa em indivíduos com ATLL ( $\log_{10}FC = -0,4627$   $FDR = 1,85^{-04}$ ). A principal via regulada por *SOCS3* é a via da JAK-STAT (*Janus kinase–signal transducer and activator of transcription*) utilizada por muitas citocinas, fatores de crescimento e hormônios resultando em inúmeros processos como divisão celular, apoptose e tumorigênese (ROTTENBERG & CAROW, 2014). *SOCS3* já foi associado ao desenvolvimento de câncer (ALVAREZ et al, 2006) e, em ATLL, foi encontrada como via alterada devido à recorrentes mutações e variações no número de cópias. Além disso, JAK-STAT faz parte da via do TCR-NF-KB, considerada uma via direcionadora à ATLL (KATAOKA, 2015). Outro membro dessa família de proteínas, *SOCS1* (*Suppressor of cytokine signaling 1*), foi citada como tendo papel importante na evasão do HTLV-1 ao interferon do tipo 1 em pacientes com HAM/TSP e pacientes assintomáticos (OLIÈRE et al., 2011).

*CUL3*, outro gene hub do grupo ATLL (**Fig. 9**), é o gene que codifica a proteína Culina 3, uma ligase que participa da via de degradação por Ubiquitina através do proteassoma 26S (PINTARD et al., 2004). A desregulação de *CUL3* foi observada em alguns cânceres humanos, significativamente expresso em câncer invasivo, e podendo ser utilizado como marcador de progressão de câncer de mama

(HAAGENSON et al, 2012). Sua crescente associação com diversas desordens humanas transforma a Culina 3 em um importante alvo terapêutico (ANDÉRICA-ROMERO et al, 2013).

Os outros três genes hubs de ATLL (**Fig. 9**) pontuaram *score* semelhante e também agem na via de ubiquitinação. A família gênica KLHL (*Kelch-like*), a qual estão inseridos *KLHL5* e *KLHL22* comporta um grupo de genes codificantes de proteínas que se ligam à Culina 3 na fase E3-ligase da via de ubiquitinação (DHANOA et al., 2013). A proteína KLHL22, subexpressa em ATLL ( $\log_{10}FC = -0,2175$ , FDR = 0,0393) forma o complexo KLHL-CUL3 que direciona a proteína PKL1 para degradação, a qual é considerada tumorigênica (METZGER et al., 2013). O fármaco Volasertib foi desenvolvido objetivando a inibição de PLK1 em Leucemia Mielóide Aguda (LMA) (GJERTSEN & SCHÖFFSKI, 2015). Em relação a *KLHL5*, uma superexpressão foi associada com uma maior resistência de células cancerígenas a compostos anticâncer (SCHLEIFER et al., 2018), e em ATLL essa proteína está subexpressa ( $\log_{10}FC = -0,3926$ , FDR = 0,027). O FBXO2 (*F-box only protein 2*), por sua vez, ( $\log_{10}FC = -0,2144$ , FDR = 0,0156) é uma proteína da família F-box que constitui uma das subunidades do complexo proteico SCFs (SKP1-culina-F-box) que age como ligase E3 na via de ubiquitinação e já foi sugerido como marcador de metástase em câncer gástrico (SUN et al., 2018). O conjunto de genes hubs resultantes revelam a importância da via JAK-STAT e, principalmente da via de ubiquitinação na patologia da ATLL, podendo ser no futuro alvo de estudos para desenvolvimento de terapias.

### 5.2.2 HAM/TSP

O módulo **HAM/TSP1** é formado por proteínas originárias dos GDEs obtidos nas metanálises, mas também é formado de proteínas com interações preditas pelo STRING (**Fig. 11**). Apesar de não serem GDE, essas proteínas preditas estarão presentes no processo celular normal. O *splicing* é um processo onde o pré-mRNA é processado para remoção de *introns* e consequente formação do mRNA maduro. A maquinaria do spliceossomo é formada por complexos multiprotéicos associados à segmentos de RNA (CVITKOVIC & JURICA, 2012). Esse processo parece estar desregulado em **HAM/TSP1** (**Tabela 6**). Uma desregulação na maquinaria do spliceossomo pode causar defeitos de *splicing* ou alterações que podem ser associadas com doenças humanas (XU et al, 2015). Neste estudo foi mostrado que, provavelmente, a via do spliceossomo pode ter um papel ainda não descoberto na

HAM/TSP.

O módulo **HAM/TSP2** apresenta “Endocitose” como via enriquecida (**Tabela 6**) devido a atividade de proteínas da família gênica VPS (Proteínas direcionadoras de Vacúolo) e ao gene *SNX12*. *VPS29*, *VPS26A*, *VPS35* e a proteína *SNX12* participam da formação dos endossomos iniciais na via endocítica e o *VPS36* e *VPS25* da formação do endossomo tardio, diferentemente do que visto no módulo **ATLL2** onde os GDEs agem diretamente na via de endocitose mediada por clatrina.

A via “Proteólise Mediada por Ubiquitina” desregulada no módulo **HAM/TSP3** apresenta GDEs diferentes dos presentes em **ATLL1**. Em **ATLL1** há uma maioria participante da via da E3 (**Tabela 6**), já em **HAM/TSP3**, além de E3 (*CBLB*, *RCHY1*) há também E1 (*UBA1*) e E2 (*UBE2U*). Desses, apenas *RCHY1* está subexpresso ( $\log_{10}FC = -0,1869$ ,  $FDR = 0,0388$ ). Esse gene produz uma proteína que direciona p53, um conhecido agente supressor tumoral, para degradação no proteassoma (LENG et al, 2003). *UBA1* é um gene supressor de tumor (LEE et al., 2008), *UBE2U* está relacionado a reparo de DNA (GUO et al., 2017) e ambas estão superexpressas (*UBA1*:  $\log_{10}FC = 0,299$ ,  $FDR = 0,0242$ / *UBE2U*:  $\log_{10}FC = 0,4636$ ,  $FDR = 0,005$ ), o que sugere que esse módulo em **HAM/TSP** está proporcionando um efeito oposto à malignidade ocorrente em **ATLL**, apresentando um efeito protetor ao DNA.

Para **HAM/TSP**, dos cinco genes hubs selecionados apenas *SNRN40* foi encontrado diferencialmente expresso na metanálise (**Fig. 12**), os outros quatro genes foram preditos pelo STRING e interagem com os GDEs de **HAM/TSP**. No geral, duas famílias gênicas que participam de etapas da maquinaria do spliceossomo se destacaram nessa análise: PRPF e SNRN. PRPF8, PRPF6 e PRPF19 compreendem proteínas que auxiliam na formação do spliceossomo e SNRNP200, SNRNP40 formam complexos proteína-RNA, também participantes do spliceossomo (Disponível em: [https://www.genecards.org/cgi-bin/carddisp.pl?gene= PRPF8, PRPF6, PRPF19, SNRNP200, SNRNP40](https://www.genecards.org/cgi-bin/carddisp.pl?gene=PRPF8,PRPF6,PRPF19,SNRNP200,SNRNP40)).

Portanto, sinaliza-se para a **HAM/TSP** os genes relacionados a vias de *splicing* como possíveis candidatos a biomarcadores nesta patologia. Genes relacionados à mecanismos endocíticos voltam a ser sinalizados como possíveis marcadores de processos patológicos associados à infecção pelo HTLV-1. Porém, como já discutido anteriormente, é necessário aumentar o poder estatístico da análise com um maior número de estudos/amostras ou uma melhor descrição dos GDEs obtidos para não haver a necessidade inclusão de proteínas fora dos GDE preditos e portanto,

aumentar a precisão dos resultados.

### 5.2.3 ASS

Os módulos da rede IPP de assintomáticos selecionados apresentaram em sua composição uma grande proporção de genes preditos pelo STRING (**Fig. 14**). As vias enriquecidas encontradas para **ASS1** compartilham alguns genes em comum (**Tabela 7**) e valores de FDR semelhantes, o que pode sugerir que qualquer uma dessas vias pode estar desregulada. Alguns estudos indicam que a senescência celular, encontrada enriquecida no módulo **ASS1**, é um mecanismo de defesa antitumoral e antiviral, principalmente contra oncovírus (REDDEL, 2009; BAZ-MARTÍNEZ et al., 2016), o que pode justificar uma ativação dessa via em pacientes assintomáticos para HTLV-1. A via de sinalização p53 é uma parte da via de infecção pelo HTLV-1. Tax, por exemplo, ativa p53, o qual reduz os danos causados por esse vírus no genoma da célula. Como discutido anteriormente, o estudo de Tattermusch e cols. (2012) revelou perturbações na via do p53 na infecção pelo HTLV-1.

A função de “Adesão Focal”, enriquecida no módulo **ASS2** (**Tabela 7**), é uma via responsável pela criação e manutenção de adesões célula-matriz extracelular e possui importante papel em processos biológicos como proliferação celular, motilidade celular, diferenciação celular e frequentemente, alterações na adesão focal levam ao desenvolvimento de patologias humanas (WU, 2007). Huang e cols. (2015), ao estudar expressão gênica de câncer de pulmão, sugeriram que *PARVA* (*parvin-alpha*) ativa *ILK* promovendo metástase nesse tipo de câncer. *PARVB* (*parvin-beta*), por sua vez, quando superexpresso, aumenta a capacidade de migração celular em carcinoma escamoso de língua (ESLAMI et al., 2015). Todas essas alterações estruturais causadas por esse módulo nas células em pacientes assintomáticos podem indicar um possível direcionamento à malignidade.

Os genes hubs selecionados para ASS (**Fig. 15**) não foram encontrados diferencialmente expressos em indivíduos assintomáticos, porém, são genes que interagem com os GDEs de ASS formando a rede de interação. Todos eles estão associados ao amplo processo do ciclo celular. Três deles (*CDC20*, *CCNB1* e *CDK1*) estão relacionados à divisão celular enquanto que os outros dois genes (*CKS2*, *CKS1B*) produzem proteínas cinases, que quando superexpressas burlam o sistema de checkpoint do ciclo celular, permitindo assim a evasão do controle de oncogenes, fornecendo um provável link à oncogênese (MU et al., 2017).

De forma semelhante para os dados de indivíduos HAM/TSP a predição de genes pelo STRING limita o poder da sugestão de biomarcadores para este status clínico, no entanto, como forma de indicar um caminho para análises futuras, destaca-se genes relacionados à infecção pelo HTLV-1 e genes associados à adesão focal como possíveis biomarcadores.

### 5.3 MICRO-RNAS DESREGULADOS FAVORECEM A ATLL

Moléculas de Micro-RNAs são reguladores da expressão gênica e agem diminuindo a expressão de um gene ao direcionar o seu respectivo mRNA para degradação. Apenas indivíduos com ATLL apresentaram miRNAs diferencialmente expressos (**Quadro 2**), apesar da cobertura de 1,847 sondas diferentes de miRNAs nos microarranjos utilizados na realização dos estudos individuais (**Tabela 8**). O hsa-mir-21, transcrito do gene *MIR21*, está superexpresso ( $\log_{10}FC = 0,4012$ ,  $FDR = 0$ ) indicando uma participação desse micro-RNA no processo patológico da ATLL. Na literatura médica, o hsa-mir-21 é um miRNA conhecido por estar associado à alguns tipos de câncer, processos inflamatórios e doenças cardíacas (FENG & TSAO, 2016; SHEEDY, 2015; FU et al., 2011). Em células infectadas pelo HTLV-1, o hsa-mir-21 já foi reportado como superexpresso propiciando uma atividade anti-apoptótica, afetando a proliferação celular e promovendo metástase (SAMPEY et al., 2012; RUGGERO et al., 2010). Devido a essas características, o hsa-mir-21 é considerado um “oncomir” e foi proposto como biomarcador transcricional de malignidade para circulação sanguínea, líquido cefalorraquidiano, escarro e fezes (KUMARSWAMY et al., 2011). Em modelos murinos, inibição de hsa-mir-21 resultou em benefícios significativos, como redução de tumor (HATLEY et al., 2010; MEDINA et al., 2010). Pelo conjunto de informações obtidas a respeito da funcionalidade crítica desse miRNA em processos cancerígenos e inflamatórios, o hsa-mir-21 tornou-se um crescente alvo atraente para pesquisas genéticas e farmacológicas em várias condições de doença.

Por outro lado, *MIR6840*, que codifica o hsa-mir-6840, considerado significativamente subexpresso em ATLL ( $\log_{10}FC = -0,2193$ ,  $FDR = 4,74^{-08}$ ), já foi encontrado subexpresso em amostras de diversos cânceres como o gástrico, de bexiga e no carcinoma escamoso de boca (LUO et al., 2015; CHEN, et al., 2017; KAWAKUBO-YASUKOCHI et al., 2018) porém, sua função ainda não está completamente esclarecida.



A relação observada entre os genes alvo de hsa-mir-21 e de hsa-mir-6840 com os GDEs em ATLL sugere uma possível regulação desses genes por esses miRNAs, alterando sua expressão (**Quadro 3**). Dentre os GDEs subexpressos em ATLL e ao mesmo tempo alvo de hsa-mir-21 há o gene *RECK* (do inglês: *Reversion-inducing-cysteine-rich protein with kazal motifs*), que codifica uma proteína com conhecidas implicações na biologia do câncer, sendo um inibidor de tumor (ALEXIUS-LINDGREN et al., 2014). Dentre os genes regulados por hsa-mir-6840 está o *CCR4*, que está superexpresso em ATLL e é considerado essencial para o desenvolvimento da patologia, como discutido anteriormente (MAKITA & TOBINAI, 2017). Entende-se a necessidade da realização de estudos experimentais adicionais para confirmação dessa possível relação regulatória entre os genes alvo de hsa-mir-21 e de hsa-mir-6840 em comum com os GDEs em ATLL.

Micro-RNAs circulantes são frequentemente propostos na área de diagnósticos como biomarcadores para doenças e para diferenciação entre tipos de câncer pelo fato de seus níveis de expressão refletirem processos patológicos (DE PLANELL-SAGUER & RODICIO, 2011). Nesse sentido, a utilização de hsa-mir-21 e hsa-mir-6840 como biomarcadores para ATLL pode ser viável.

## 6. CONCLUSÃO

Este trabalho foi capaz de identificar diferentes grupos de genes desregulados em três status clínicos associados à infecção pelo HTLV-1 e, a partir desses genes, construir um perfil de expressão gênica diferencial. Mostrando também que houve uma maior desregulação gênica em indivíduos com ATLL, seguido por indivíduos com HAM/TSP e indivíduos assintomáticos.

Análises de vias enriquecidas apontaram vias que são alteradas pela infecção em indivíduos com ATLL como “Vias em Câncer”, que incluem genes responsáveis pela evasão da apoptose, proliferação celular e angiogênese continuada, para os genes superexpressos e, pela primeira vez, apontando uma desregulação na “Endocitose”, para os genes subexpressos, o que pode estar contribuindo para o desenvolvimento da manifestação clínica.

A análise de vias enriquecidas revelou a desregulação das vias de “Proteólise mediada por Ubiquitina”, “Endocitose” e mais 13 vias, dentre elas a “Via de sinalização de quimiocinas” em indivíduos com ATLL. Para o grupo com HAM/TSP as vias do “Spliceossomo”, “Endocitose” e “Proteólise mediada por Ubiquitina” estão desreguladas enquanto que “Adesão focal” juntamente com “Ciclo celular” e mais nove vias estão desregulados em PBMCs de pacientes assintomáticos. Sendo tais genes os mais indicados em futuras investigações e definições de biomarcadores dessas manifestações clínicas.

Além disso, foi possível a sugestão de genes hub e micro-RNAs para os diferentes status clínicos. *CUL3* e genes da família KLHL, fazendo parte da via de ubiquitinação de proteínas, podem ser considerados genes hubs em ATLL juntamente com *SOCS3*, gene integrante da via JAK-STAT. Para HAM/TSP foram selecionados genes participantes da via do spliceossomo, sugerindo uma importância desconhecida dessa via para o desenvolvimento de HAM/TSP.

Já para ASS, os genes hubs selecionados foram genes fora do conjunto de genes diferencialmente expressos, não servindo como biomarcadores mas exibindo importância terapêutica. São genes que, sabidamente, interagem com os GDEs de ASS formando a rede de Interação: *CDC20*, *CCNB1* e *CDK1* assim como dois genes da família *CKS*, refletindo a importância de processos do ciclo celular em indivíduos assintomáticos. Todos esses genes hubs, selecionados através de análises integradas computacionais podem, posteriormente, serem alvo de pesquisas terapêuticas por denotar importância na estruturação da rede de genes desregulados.

Nenhum micro-RNA foi considerado diferencialmente expresso, com os parâmetros adotados, no grupo de indivíduos com HAM/TSP ou ASS. No entanto, dois micro-RNAs de alta confiabilidade foram considerados altamente desregulados no grupo de indivíduos com ATLL, o hsa-mir-21, que parece estar diretamente envolvido no desenvolvimento da doença e o hsa-mir-6840, que ainda apresenta funções desconhecidas, sendo adequados para serem utilizados como biomarcadores para ATLL.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALEXIUS-LINDGREN, Michelle et al. The RECK gene and biological malignancy—its significance in angiogenesis and inhibition of matrix metalloproteinases. *Anticancer research*, v. 34, n. 8, p. 3867-3873, 2014.

ALLAYA, Nesrine et al. Over-expression of miR-10b in NPC patients: correlation with LMP1 and Twist1. *Tumor Biology*, v. 36, n. 5, p. 3807-3814, 2015.

ALVAREZ, James V. et al. Signal transducer and activator of transcription 3 is required for the oncogenic effects of non–small-cell lung cancer–associated mutations of the epidermal growth factor receptor. *Cancer research*, v. 66, n. 6, p. 3162-3168, 2006.

ANDÉRICA-ROMERO, Ana Cristina et al. Cullin 3 as a novel target in diverse pathologies. *Redox biology*, v. 1, n. 1, p. 366-372, 2013.

ARAUJO, Abelardo QC; SILVA, Marcus Tullius T. The HTLV-1 neurological complex. *The Lancet Neurology*, v. 5, n. 12, p. 1068-1076, 2006.

ASQUITH, Becca; BANGHAM, Charles RM. Quantifying HTLV-I dynamics. *Immunology and cell biology*, v. 85, n. 4, p. 280-286, 2007.

BANGHAM, Charles RM. Human T cell leukemia virus type 1: persistence and pathogenesis. *Annual review of immunology*, v. 36, p. 43-71, 2018.

BARABÁSI, Albert-László; GULBAHCE, Natali; LOSCALZO, Joseph. Network medicine: a network-based approach to human disease. *Nature reviews genetics*, v. 12, n. 1, p. 56, 2011.

BARNES, Michael et al. Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. *Nucleic acids research*, v. 33, n. 18, p. 5914-5923, 2005.

BARTEL, David P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*,

v. 116, n. 2, p. 281-297, 2004.

BARTOLETTI-STELLA, Anna et al. Messenger RNA processing is altered in autosomal dominant leukodystrophy. **Human molecular genetics**, v. 24, n. 10, p. 2746-2756, 2015.

BAZ-MARTÍNEZ, Maite et al. Cell senescence is an antiviral defense mechanism. **Scientific reports**, v. 6, p. 37007, 2016.

CASKEY, Marina F. et al. Clinical manifestations associated with HTLV type I infection: a cross-sectional study. **AIDS research and human retroviruses**, v. 23, n. 3, p. 365-371, 2007.

CATALAN-SOARES, Bernadette; CARNEIRO-PROIETTI, Anna Bárbara de F.; PROIETTI, Fernando Augusto. Heterogeneous geographic distribution of human T-cell lymphotropic viruses I and II (HTLV-I/II): serological screening prevalence rates in blood donors from large urban areas in Brazil. **Cadernos de Saúde Pública**, v. 21, p. 926-931, 2005.

CHACKO, Balu K. et al. Endothelial surface N-glycans mediate monocyte adhesion and are targets for anti-inflammatory effects of peroxisome proliferator-activated receptor  $\gamma$  ligands. **Journal of Biological Chemistry**, v. 286, n. 44, p. 38738-38747, 2011.

CHANG, Lun-Ching et al. Meta-analysis methods for combining multiple expression profiles: comparisons, statistical characterization and an application guideline. **BMC bioinformatics**, v. 14, n. 1, p. 368, 2013.

CHEN, Edward Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. **BMC bioinformatics**, v. 14, n. 1, p. 128, 2013.

CHEN, Jueqi; CHEN, Zhijian J. Regulation of NF- $\kappa$ B by ubiquitination. **Current opinion in immunology**, v. 25, n. 1, p. 4-12, 2013.

CHEN, Liang et al. Identification and bioinformatics analysis of miRNAs associated

with human muscle invasive bladder cancer. **Molecular medicine reports**, v. 16, n. 6, p. 8709-8720, 2017.

CHOI, Jung Kyoon et al. Combining multiple microarray studies and modeling interstudy variation. **Bioinformatics**, v. 19, n. suppl\_1, p. i84-i90, 2003.

CHOI, Y. L. et al. A genomic analysis of adult T-cell leukemia. **Oncogene**, v. 26, n. 8, p. 1245, 2007.

CLOUGH, Emily; BARRETT, Tanya. The gene expression omnibus database. In: **Statistical Genomics**. Humana Press, New York, NY, 2016. p. 93-110.

COOK, Lucy B. et al. HTLV-1: persistence and pathogenesis. **Virology**, v. 435, n. 1, p. 131-140, 2013.

COSTA, José Marcos Pereira; SEGURADO, Aluisio Cotrim. Molecular evidence of human T-cell lymphotropic virus types 1 and 2 (HTLV-1 and HTLV-2) infections in HTLV seroindeterminate individuals from São Paulo, Brazil. **Journal of Clinical Virology**, v. 44, n. 3, p. 185-189, 2009.

CUNHA, Shane R.; MOHLER, Peter J. Cardiac ankyrins: Essential components for development and maintenance of excitable membrane domains in heart. **Cardiovascular research**, v. 71, n. 1, p. 22-29, 2006.

CVITKOVIC, Ivan; JURICA, Melissa S. Spliceosome database: a tool for tracking components of the spliceosome. **Nucleic acids research**, v. 41, n. D1, p. D132-D141, 2012.

DE PLANELL-SAGUER, Mariàngels; RODICIO, María Celina. Analytical aspects of microRNA in diagnostics: a review. **Analytica chimica acta**, v. 699, n. 2, p. 134-152, 2011.

DENIZ, Emre; ERMAN, Batu. Long noncoding RNA (lincRNA), a new paradigm in gene expression control. **Functional & integrative genomics**, v. 17, n. 2-3, p. 135-143,

2017.

DHANOVA, Bajinder S. et al. Update on the Kelch-like (KLHL) gene family. **Human genomics**, v. 7, n. 1, p. 13, 2013.

DOURADO, Inês et al. HTLV-I in the general population of Salvador, Brazil: a city with African ethnic and sociodemographic characteristics. **JAIDS Journal of Acquired Immune Deficiency Syndromes**, v. 34, n. 5, p. 527-531, 2003.

DUPUY, Alain; SIMON, Richard M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. **Journal of the National Cancer Institute**, v. 99, n. 2, p. 147-157, 2007.

ECHEVARRIA-LIMA, Juliana et al. Protein profile of blood monocytes is altered in HTLV-1 infected patients: Implications for HAM/TSP disease. **Scientific reports**, v. 8, 2018.

EDWARDS, Dustin et al. Orf-I and orf-II-encoded proteins in HTLV-1 infection and persistence. **Viruses**, v. 3, n. 6, p. 861-885, 2011.

EL AMRI, Mohamed; FITZGERALD, Una; SCHLOSSER, Gerhard. MARCKS and MARCKS-like proteins in development and regeneration. **Journal of biomedical science**, v. 25, n. 1, p. 43, 2018.

ELEDGE, Michael R.; YERUVA, Laxmi. Host and pathogen interface: microRNAs are modulators of disease outcome. **Microbes and infection**, v. 20, n. 7-8, p. 410-415, 2018.

EMILSSON, Valur et al. Genetics of gene expression and its effect on disease. **Nature**, v. 452, n. 7186, p. 423, 2008.

ESLAMI, A. et al. PARVB overexpression increases cell migration capability and defines high risk for endophytic growth and metastasis in tongue squamous cell carcinoma. **British journal of cancer**, v. 112, n. 2, p. 338, 2015.

FENG, Yin-Hsun; TSAO, Chao-Jung. Emerging role of microRNA-21 in cancer. **Biomedical reports**, v. 5, n. 4, p. 395-402, 2016.

FERGUSON, Linda. External validity, generalizability, and knowledge utilization. **Journal of Nursing Scholarship**, v. 36, n. 1, p. 16-22, 2004.

FU, Xiaonan et al. Prognostic role of microRNA-21 in various carcinomas: a systematic review and meta-analysis. **European journal of clinical investigation**, v. 41, n. 11, p. 1245-1253, 2011.

FUJINO, Toshinori; NAGATA, Yukihiro. HTLV-I transmission from mother to child. **Journal of reproductive immunology**, v. 47, n. 2, p. 197-206, 2000.

FUKUTANI, Eduardo Rocha et al. Meta-analysis of HTLV-1-infected patients identifies CD40LG and GBP2 as markers of ATLL and HAM/TSP clinical status: two genes beat as one. **Frontiers in genetics**, v. 10, 2019.

GALLO, L. H.; KO, J.; DONOGHUE, D. J. The importance of regulatory ubiquitination in cancer and metastasis. **Cell Cycle**, v. 16, n. 7, p. 634-648, 2017.

GESSAIN, A. et al. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. **The Lancet**, v. 326, n. 8452, p. 407-410, 1985.

GESSAIN, A.; MAHIEUX, R. Tropical spastic paraparesis and HTLV-1 associated myelopathy: clinical, epidemiological, virological and therapeutic aspects. **Revue neurologique**, v. 168, n. 3, p. 257-269, 2012.

GESSAIN, Antoine; CASSAR, Olivier. Epidemiological aspects and world distribution of HTLV-1 infection. **Frontiers in microbiology**, v. 3, p. 388, 2012.

GHEZ, David et al. Current concepts regarding the HTLV-1 receptor complex. **Retrovirology**, v. 7, n. 1, p. 99, 2010.



GIAM, Chou-Zen; SEMMES, Oliver. HTLV-1 infection and adult T-cell leukemia/lymphoma—a tale of two proteins: Tax and HBZ. **Viruses**, v. 8, n. 6, p. 161, 2016.

GJERTSEN, Bjorn T.; SCHÖFFSKI, Patrick. Discovery and development of the Polo-like kinase inhibitor volasertib in cancer therapy. **Leukemia**, v. 29, n. 1, p. 11, 2015.

GONÇALVES, Denise Utsch et al. Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. **Clinical microbiology reviews**, v. 23, n. 3, p. 577-589, 2010.

GRANT, Christian et al. Human T cell leukemia virus type I and neurologic disease: events in bone marrow, peripheral blood, and central nervous system during normal immune surveillance and neuroinflammation. **Journal of cellular physiology**, v. 190, n. 2, p. 133-159, 2002.

GRASSI, Maria Fernanda Rios et al. Human T cell lymphotropic virus type 1 (HTLV-1) proviral load of HTLV-associated myelopathy/tropical spastic paraparesis (HAM/TSP) patients according to new diagnostic criteria of HAM/TSP. **Journal of medical virology**, v. 83, n. 7, p. 1269-1274, 2011.

GREEN, P.L. & CHEN, I.S.Y. Human T-cell leukemia virus types 1 and 2. Em: KNIPE, D.M.; HOWLEY, P.; GRIFFIN, D.E. et al., ed. **Fields Virology**. Philadelphia, Lippincott Williams & Wilkins, v. 2, p. 1941-1969, 2001.

GRIFFITHS-JONES, Sam et al. miRBase: microRNA sequences, targets and gene nomenclature. **Nucleic acids research**, v. 34, n. suppl\_1, p. D140-D144, 2006.

GROSS, Christine; THOMA-KRESS, Andrea K. Molecular mechanisms of HTLV-1 cell-to-cell transmission. **Viruses**, v. 8, n. 3, p. 74, 2016.

GUO, Yingying et al. An E2-guided E3 screen identifies the RNF17-UBE2U pair as regulator of the radiosensitivity, immunodeficiency, dysmorphic features, and learning difficulties (RIDDLE) syndrome protein RNF168. **Journal of Biological Chemistry**, v.

292, n. 3, p. 967-978, 2017.

HAAGENSON, Kelly K. et al. Cullin-3 protein expression levels correlate with breast cancer progression. **Cancer biology & therapy**, v. 13, n. 11, p. 1042-1046, 2012.

HATLEY, Mark E. et al. Modulation of K-Ras-dependent lung tumorigenesis by MicroRNA-21. **Cancer cell**, v. 18, n. 3, p. 282-293, 2010.

HUANG, Ay-Huey et al. PARVA promotes metastasis by modulating ILK signalling pathway in lung adenocarcinoma. **PLoS One**, v. 10, n. 3, p. e0118530, 2015.

HUGHES, Catherine E.; NIBBS, Robert JB. A guide to chemokines and their receptors. **The FEBS journal**, v. 285, n. 16, p. 2944-2971, 2018.

ISHIHARA, Makoto et al. Preapoptotic protease calpain-2 is frequently suppressed in adult T-cell leukemia. **Blood**, v. 121, n. 21, p. 4340-4347, 2013.

JEFFERY, Katie JM et al. HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy. **Proceedings of the National Academy of Sciences**, v. 96, n. 7, p. 3848-3853, 1999.

JEFFERY, Katie JM et al. The influence of HLA class I alleles and heterozygosity on the outcome of human T cell lymphotropic virus type I infection. **The Journal of Immunology**, v. 165, n. 12, p. 7278-7284, 2000.

KANNIAN, Priya; GREEN, Patrick L. Human T lymphotropic virus type 1 (HTLV-1): molecular biology and oncogenesis. **Viruses**, v. 2, n. 9, p. 2037-2077, 2010.

KATAOKA, Keisuke et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. **Nature genetics**, v. 47, n. 11, p. 1304, 2015.

KAWAKUBO-YASUKOCHI, Tomoyo et al. miR-200c-3p spreads invasive capacity in human oral squamous cell carcinoma microenvironment. **Molecular carcinogenesis**, v. 57, n. 2, p. 295-302, 2018.

KEELE, Brandon F. et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. **Science**, v. 313, n. 5786, p. 523-526, 2006.

KENDALL, Emily A. et al. Early neurologic abnormalities associated with human T-cell lymphotropic virus type 1 infection in a cohort of Peruvian children. **The Journal of pediatrics**, v. 155, n. 5, p. 700-706, 2009.

KOBAYASHI, Seiichiro et al. CADM1 Expression and Stepwise Downregulation of CD7 Are Closely Associated with Clonal Expansion of HTLV-I–Infected Cells in Adult T-cell Leukemia/Lymphoma. **Clinical cancer research**, v. 20, n. 11, p. 2851-2861, 2014.

KRESS, Andrea K. et al. Elevated cyclic AMP levels in T lymphocytes transformed by human T-cell lymphotropic virus type 1. **Journal of virology**, v. 84, n. 17, p. 8732-8742, 2010.

KROON, E. G; PROIETTI, A. B. de F. HTLV-1/2 – O vírus, sua multiplicação e estrutura genômica **Cadernos Hemominas: HTLV**. 6ª ed. v. XVI, p. 15-27, 2015.

KULESHOV, Maxim V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. **Nucleic acids research**, v. 44, n. W1, p. W90-W97, 2016.

KUMARSWAMY, Regalla; VOLKMANN, Ingo; THUM, Thomas. Regulation and function of miRNA-21 in health and disease. **RNA biology**, v. 8, n. 5, p. 706-713, 2011.

LAIRMORE, Michael D.; HAINES, Robyn; ANUPAM, Rajaneesh. Mechanisms of human T-lymphotropic virus type 1 transmission and disease. **Current opinion in virology**, v. 2, n. 4, p. 474-481, 2012.

LEE, Tom V. et al. The E1 ubiquitin-activating enzyme Uba1 in Drosophila controls apoptosis autonomously and tissue growth non-autonomously. **Development**, v. 135, n. 1, p. 43-52, 2008.

LEE, Ahn R. et al. Alternative RNA splicing of the MEAF6 gene facilitates neuroendocrine prostate cancer progression. **Oncotarget**, v. 8, n. 17, p. 27966, 2017.

LENG, Roger P. et al. Pirh2, a p53-induced ubiquitin-protein ligase, promotes p53 degradation. **Cell**, v. 112, n. 6, p. 779-791, 2003.

LEVIN, Michael C. et al. Autoimmunity due to molecular mimicry as a cause of neurological disease. **Nature medicine**, v. 8, n. 5, p. 509, 2002.

LIN, Qiuyu et al. HORMAD 2 methylation-mediated epigenetic regulation of gene expression in thyroid cancer. **Journal of cellular and molecular medicine**, v. 22, n. 10, p. 4640-4652, 2018.

LOUREIRO, Paula; LOPES, Maria S. S. N. Leucemia/Linfoma de Células T do Adulto (ATL). **Cadernos Hemominas: HTLV**. 6ª ed. v. XVI, p. 181-230, 2015.

LUO, Yushuang et al. Bioinformatics identification of potentially involved microRNAs in Tibetan with gastric cancer based on microRNA profiling. **Cancer cell international**, v. 15, n. 1, p. 115, 2015.

MAKITA, Shinichi; TOBINAI, Kensei. Mogamulizumab for the treatment of T-cell lymphoma. **Expert opinion on biological therapy**, v. 17, n. 9, p. 1145-1153, 2017.

MANIVANNAN, Kiruthika et al. CADM1/TSLC1 identifies HTLV-1-infected cells and determines their susceptibility to CTL-mediated lysis. **PLoS pathogens**, v. 12, n. 4, p. e1005560, 2016.

MAREEL, Marc; LEROY, Ancy. Clinical, cellular, and molecular aspects of cancer invasion. **Physiological reviews**, v. 83, n. 2, p. 337-376, 2003.

MATSUOKA, Masao; JEANG, Kuan-Teh. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. **Nature Reviews Cancer**, v. 7, n. 4, p. 270, 2007. MEHTA, Arnav; BALTIMORE, David. MicroRNAs as regulatory elements in immune system logic. **Nature Reviews Immunology**, v. 16, n. 5, p. 279, 2016.

MEDINA, Pedro P.; NOLDE, Mona; SLACK, Frank J. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. **Nature**, v. 467, n. 7311, p. 86, 2010.

METZGER, Thibaud; KLEISS, Charlotte; SUMARA, Izabela. CUL3 and protein kinases: insights from PLK1/KLHL22 interaction. **Cell Cycle**, v. 12, n. 14, p. 2291-2296, 2013.

MICHAEL, Bindhu et al. Human T lymphotropic virus type-1 p30 II alters cellular gene expression to selectively enhance signaling pathways that activate T lymphocytes. **Retrovirology**, v. 1, n. 1, p. 39, 2004.

MIKLOS, George L. Gabor; MALESZKA, Ryszard. Microarray reality checks in the context of a complex disease. **Nature biotechnology**, v. 22, n. 5, p. 615, 2004.

MILLER, Melissa B.; TANG, Yi-Wei. Basic concepts of microarrays and potential applications in clinical microbiology. **Clinical microbiology reviews**, v. 22, n. 4, p. 611-633, 2009.

MINISTÉRIO DA SAÚDE, SECRETARIA DE VIGILÂNCIA EM SAÚDE. Guia de manejo Clínico da Infecção pelo HTLV. 2013.

MOSESSON, Yaron; MILLS, Gordon B.; YARDEN, Yosef. Derailed endocytosis: an emerging feature of cancer. **Nature Reviews Cancer**, v. 8, n. 11, p. 835, 2008.

MU, Ruiling et al. CKS proteins promote checkpoint recovery by stimulating phosphorylation of treslin. **Molecular and cellular biology**, v. 37, n. 20, p. e00344-17, 2017.

NAGAI, Masahiro et al. Analysis of HTLV-I proviral load in 202 HAM/TSP patients and 243 asymptomatic HTLV-I carriers: high proviral load strongly predisposes to HAM/TSP. **Journal of neurovirology**, v. 4, n. 6, p. 586-593, 1998.

NAGAI, Masahiro; OSAME, Mitsuhiro. Human T-cell lymphotropic virus type I and neurological diseases. **Journal of neurovirology**, v. 9, n. 2, p. 228-235, 2003.

NAKAGAWA, Masao et al. Gain-of-function CCR4 mutations in adult T cell leukemia/lymphoma. **Journal of Experimental Medicine**, v. 211, n. 13, p. 2497-2505, 2014.

NAKAHATA, Shingo et al. Loss of NDRG2 expression activates PI3K-AKT signalling via PTEN phosphorylation in ATLL and other cancers. **Nature communications**, v. 5, p. 3393, 2014.

NAKAMURA, Tatsufumi. Immunopathogenesis of HTLV-I-associated myelopathy/tropical spastic paraparesis. **Gazeta Médica da Bahia**, v. 79, n. 1, p. 11-17, 2009.

NTZANI, Evangelia E.; IOANNIDIS, John PA. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. **The Lancet**, v. 362, n. 9394, p. 1439-1444, 2003.

NYKJAER, Anders; WILLNOW, Thomas E. Sortilin: a receptor to regulate neuronal viability and function. **Trends in neurosciences**, v. 35, n. 4, p. 261-270, 2012.

O'BRIEN, Jacob et al. Overview of microRNA biogenesis, mechanisms of actions, and circulation. **Frontiers in endocrinology**, v. 9, p. 402, 2018.

OMURA, Noriyuki et al. Overexpression of ankyrin1 promotes pancreatic cancer cell growth. **Oncotarget**, v. 7, n 23, p. 34977, 2016.

OLIERE, Stephanie et al. HTLV-1 evades type I interferon antiviral signaling by inducing the suppressor of cytokine signaling 1 (SOCS1). **PLoS pathogens**, v. 6, n. 11, p. e1001177, 2010.

OLIERE, Stéphanie et al. Modulation of innate immune responses during human T-cell leukemia virus (HTLV-1) pathogenesis. **Cytokine & growth factor reviews**, v. 22,

n. 4, p. 197-210, 2011.

OLINDO, Stéphane et al. HTLV-1 proviral load in peripheral blood mononuclear cells quantified in 100 HAM/TSP patients: a marker of disease progression. **Journal of the neurological sciences**, v. 237, n. 1-2, p. 53-59, 2005.

OLIVEROS, J.C. Venny. An interactive tool for comparing lists with Venn's diagrams. (2007-2015) <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.

OSAME, Mitsuhiro. HTLV-I associated myelopathy, a new clinical entity. **Lancet**, v. 1, p. 1031-1032, 1986.

OSAME, Mitsuhiro et al. Nationwide survey of HTLV-I-associated myelopathy in Japan: Association with blood transfusion. **Annals of neurology**, v. 28, n. 1, p. 50-56, 1990.

OSAME, Mitsuhiro. Pathological mechanisms of human T-cell lymphotropic virus type I-associated myelopathy (HAM/TSP). **Journal of neurovirology**, v. 8, n. 5, p. 359-364, 2002.

PAIVA, Arthur; CASSEB, Jorge. Sexual transmission of human T-cell lymphotropic virus type 1. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 47, n. 3, p. 265-274, 2014.

PAIVA, Arthur; CASSEB, Jorge. Origin and prevalence of human T-lymphotropic virus type 1 (HTLV-1) and type 2 (HTLV-2) among indigenous populations in the Americas. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 57, n. 1, p. 01-14, 2015.

PINTARD, Lionel; WILLEMS, Andrew; PETER, Matthias. Cullin-based ubiquitin ligases: Cul3–BTB complexes join the family. **The EMBO journal**, v. 23, n. 8, p. 1681-1687, 2004.

PINTO, Mariana Tomazini et al. Genes Related to Antiviral Activity, Cell Migration, and Lysis Are Differentially Expressed in CD4+ T Cells in Human T Cell Leukemia Virus

Type 1-Associated Myelopathy/Tropical Spastic Paraparesis Patients. **AIDS research and human retroviruses**, v. 30, n. 6, p. 610-622, 2014.

PISE-MASISON, Cynthia A. et al. Gene expression profiling of ATL patients: compilation of disease-related genes and evidence for TCF4 involvement in BIRC5 gene expression and cell viability. **Blood**, v. 113, n. 17, p. 4016-4026, 2009.

POIESZ, Bernard J. et al. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. **Proceedings of the National Academy of Sciences**, v. 77, n. 12, p. 7415-7419, 1980.

POLO, Simona; PECE, Salvatore; DI FIORE, Pier Paolo. Endocytosis and cancer. **Current opinion in cell biology**, v. 16, n. 2, p. 156-161, 2004.

PUJARI, Rajeshree et al. Human T-cell leukemia virus type 1 (HTLV-1) tax requires CADM1/TSLC1 for inactivation of the NF- $\kappa$ B inhibitor A20 and constitutive NF- $\kappa$ B signaling. **PLoS pathogens**, v. 11, n. 3, p. e1004721, 2015.

QUINLAN, Sean et al. MicroRNAs in neurodegenerative diseases. In: International review of cell and molecular biology. **Academic Press**, 2017. p. 309-343.

RAMASAMY, Adaikalavan et al. Key issues in conducting a meta-analysis of gene expression microarray datasets. **PLoS medicine**, v. 5, n. 9, p. e184, 2008.

RAN, Xi et al. Regulation of embryonic stem cell self-renewal and differentiation by microRNAs. **Cellular reprogramming**, v. 19, n. 3, p. 150-158, 2017.

RAUCH, Daniel A.; RATNER, Lee. Targeting HTLV-1 activation of NF $\kappa$ B in mouse models and ATLL patients. **Viruses**, v. 3, n. 6, p. 886-900, 2011.

REDDEL, Roger R. Senescence: an antiviral defense that is tumor suppressive?. **Carcinogenesis**, v. 31, n. 1, p. 19-26, 2009.



REIMAND, Jüri et al. Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and EnrichmentMap. **Nature protocols**, v. 14, n. 2, p. 482, 2019.

ROBERTS, Christopher J. et al. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. **Science**, v. 287, n. 5454, p. 873-880, 2000.

RODGERS-JOHNSON, Pamela et al. HTLV-I and HTLV-III antibodies and tropical spastic paraparesis. **The Lancet**, v. 326, n. 8466, p. 1247-1248, 1985.

RODRIGUEZ-ESTEBAN, Raul; JIANG, Xiaoyu. Differential gene expression in disease: a comparison between high-throughput studies and the literature. **BMC medical genomics**, v. 10, n. 1, p. 59, 2017.

ROTTENBERG, Martin E.; CAROW, Berit. SOCS3, a major regulator of infection and inflammation. **Frontiers in immunology**, v. 5, p. 58, 2014.

RUGGERO, Katia et al. Role of microRNAs in HTLV-1 infection and transformation. **Molecular aspects of medicine**, v. 31, n. 5, p. 367-382, 2010.

RUPAIMOOLE, Rajesha; SLACK, Frank J. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. **Nature reviews Drug discovery**, v. 16, n. 3, p. 203, 2017.

SABARIMURUGAN, Shanthi et al. Systematic review and meta-analysis of prognostic microRNA biomarkers for survival outcome in nasopharyngeal carcinoma. **PloS one**, v. 14, n. 2, p. e0209760, 2019.

SAFARI-ALIGHIARLOO, Nahid et al. Protein-protein interaction networks (PPI) and complex diseases. **Gastroenterology and Hepatology from bed to bench**, v. 7, n. 1, p. 17, 2014.

SAMPEY, Gavin C. et al. Complex role of microRNAs in HTLV-1 infections. **Frontiers**

in **genetics**, v. 3, p. 295, 2012.

SANTOS, Mariana M. et al. Aspectos epidemiológicos da infecção por HTLV-1 e HTLV-2. **Cadernos Hemominas: HTLV**. 6ª ed. v. XVI, p. 117-139. 2015.

SATTENTAU, Quentin. Avoiding the void: cell-to-cell spread of human viruses. **Nature Reviews Microbiology**, v. 6, n. 11, p. 815, 2008.

SCHLEIFER, Robert J. et al. KLHL5 knockdown increases cellular sensitivity to anticancer drugs. **Oncotarget**, v. 9, n. 100, p. 37429, 2018.

SCHMID, Sandra L. Reciprocal regulation of signaling and endocytosis: Implications for the evolving cancer cell. **J Cell Biol**, v. 216, n. 9, p. 2623-2623, 2017.

SEIKI, Motoharu et al. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. **Proceedings of the national academy of sciences**, v. 80, n. 12, p. 3618-3622, 1983.

SHANNON, Paul et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome research**, v. 13, n. 11, p. 2498-2504, 2003.

SHAROV, Alexei A.; SCHLESSINGER, David; KO, Minoru SH. ExAtlas: An interactive online tool for meta-analysis of gene expression data. **Journal of bioinformatics and computational biology**, v. 13, n. 06, p. 1550019, 2015.

SHEEDY, Frederick J. Turning 21: induction of miR-21 as a key switch in the inflammatory response. **Frontiers in immunology**, v. 6, p. 19, 2015.

SHI, Dingding; GROSSMAN, Steven R. Ubiquitin becomes ubiquitous in cancer: emerging roles of ubiquitin ligases and deubiquitinases in tumorigenesis and as therapeutic targets. **Cancer biology & therapy**, v. 10, n. 8, p. 737-747, 2010.

SHIMOYAMA, Masanori; LYMPHOMA STUDY GROUP (1984–87). Diagnostic criteria

and classification of clinical subtypes of adult T-cell leukaemia-lymphoma: a report from the Lymphoma Study Group (1984–87). **British journal of haematology**, v. 79, n. 3, p. 428-437, 1991.

SIANGPHOE, Uma; ARCHER, Kellie J.; MUKHOPADHYAY, Nitai D. Classical and Bayesian random-effects meta-analysis models with sample quality weights in gene expression studies. **BMC bioinformatics**, v. 20, n. 1, p. 18, 2019.

SILIC-BENUSSI, Micol et al. HTLV-1 p13, a small protein with a busy agenda. **Molecular aspects of medicine**, v. 31, n. 5, p. 350-358, 2010.

SNEL, Berend et al. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. **Nucleic acids research**, v. 28, n. 18, p. 3442-3444, 2000.

STEVENS, John R.; DOERGE, Rebecca W. Combining Affymetrix microarray results. **BMC bioinformatics**, v. 6, n. 1, p. 57, 2005.

STICHT, Carsten et al. miRWalk: An online resource for prediction of microRNA binding sites. **PLoS One**, v. 13, n. 10, p. e0206239, 2018.

STRIMBU, Kyle; TAVEL, Jorge A. What are biomarkers?. **Current Opinion in HIV and AIDS**, v. 5, n. 6, p. 463, 2010.

SUN, Xu et al. FBXO2, a novel marker for metastasis in human gastric cancer. **Biochemical and biophysical research communications**, v. 495, n. 3, p. 2158-2164, 2018.

TAGAYA, Yutaka; GALLO, Robert C. The exceptional oncogenicity of HTLV-1. **Frontiers in microbiology**, v. 8, p. 1425, 2017.

TANOSAKI, Ryuji et al. Allogeneic hematopoietic stem cell transplantation using reduced-intensity conditioning for adult T cell leukemia/lymphoma: impact of antithymocyte globulin on clinical outcome. **Biology of Blood and Marrow**

**Transplantation**, v. 14, n. 6, p. 702-708, 2008.

TATTERMUSCH, Sonja et al. Systems biology approaches reveal a specific interferon-inducible signature in HTLV-1 associated myelopathy. **PLoS pathogens**, v. 8, n. 1, p. e1002480, 2012.

TSENG, George C.; GHOSH, Debashis; FEINGOLD, Eleanor. Comprehensive literature review and statistical considerations for microarray meta-analysis. **Nucleic acids research**, v. 40, n. 9, p. 3785-3799, 2012.

UCHIYAMA, Takashi et al. Adult T-cell leukemia: clinical and hematologic features of 16 cases. **Blood**, v. 50, n. 3, p. 481-492, 1977.

VANDAMME, Anne-Mieke; SALEMI, Marco; DESMYTER, Jan. The simian origins of the pathogenic human T-cell lymphotropic virus type I. **Trends in microbiology**, v. 6, n. 12, p. 477-483, 1998.

VIANA, Graca Maria de Castro et al. Seroprevalence of HTLV-1/2 among blood donors in the state of Maranhão, Brazil. **Revista brasileira de hematologia e hemoterapia**, v. 36, n. 1, p. 50-53, 2014.

WATANABE, Toshiki. HTLV-1-associated diseases. **International journal of hematology**, v. 66, n. 3, p. 257-278, 1997.

WU, Chuanyue. Focal adhesion: a focal point in current cell biology and molecular medicine. **Cell adhesion & migration**, v. 1, n. 1, p. 13-18, 2007.

XU, Weijin et al. Meta-analysis of gene expression profiles indicates genes in spliceosome pathway are up-regulated in hepatocellular carcinoma (HCC). **Medical oncology**, v. 32, n. 4, p. 96, 2015.

YAMADA, Yasuaki et al. A new G-CSF-supported combination chemotherapy, LSG15, for adult T-cell leukaemia-lymphoma: Japan Clinical Oncology Group Study 9303. **British journal of haematology**, v. 113, n. 2, p. 375-382, 2001.

YAMAGISHI, Makoto et al. Polycomb-mediated loss of miR-31 activates NIK-dependent NF- $\kappa$ B pathway in adult T cell leukemia and other cancers. **Cancer cell**, v. 21, n. 1, p. 121-135, 2012.

YAMAMOTO-TAGUCHI, Nanae et al. HTLV-1 bZIP factor induces inflammation through labile Foxp3 expression. **PLoS pathogens**, v. 9, n. 9, p. e1003630, 2013.

YAMAO, Fumiaki. Ubiquitin system: selectivity and timing of protein destruction. *The Journal of Biochemistry*, v. 125, n. 2, p. 223-229, 1999.

YAMAUCHI, Junji et al. Use of cerebrospinal fluid CXCL10 and neopterin as biomarkers in HTLV-1-associated myelopathy/tropical spastic paraparesis treated with steroids. **Journal of Neurology, Neurosurgery & Psychiatry**, p. jnnp-2019-321955, 2019.

YASUNAGA, Jun-ichirou; MATSUOKA, Masao. Human T-cell leukemia virus type I induces adult T-cell leukemia: from clinical aspects to molecular mechanisms. **Cancer control**, v. 14, n. 2, p. 133-140, 2007.

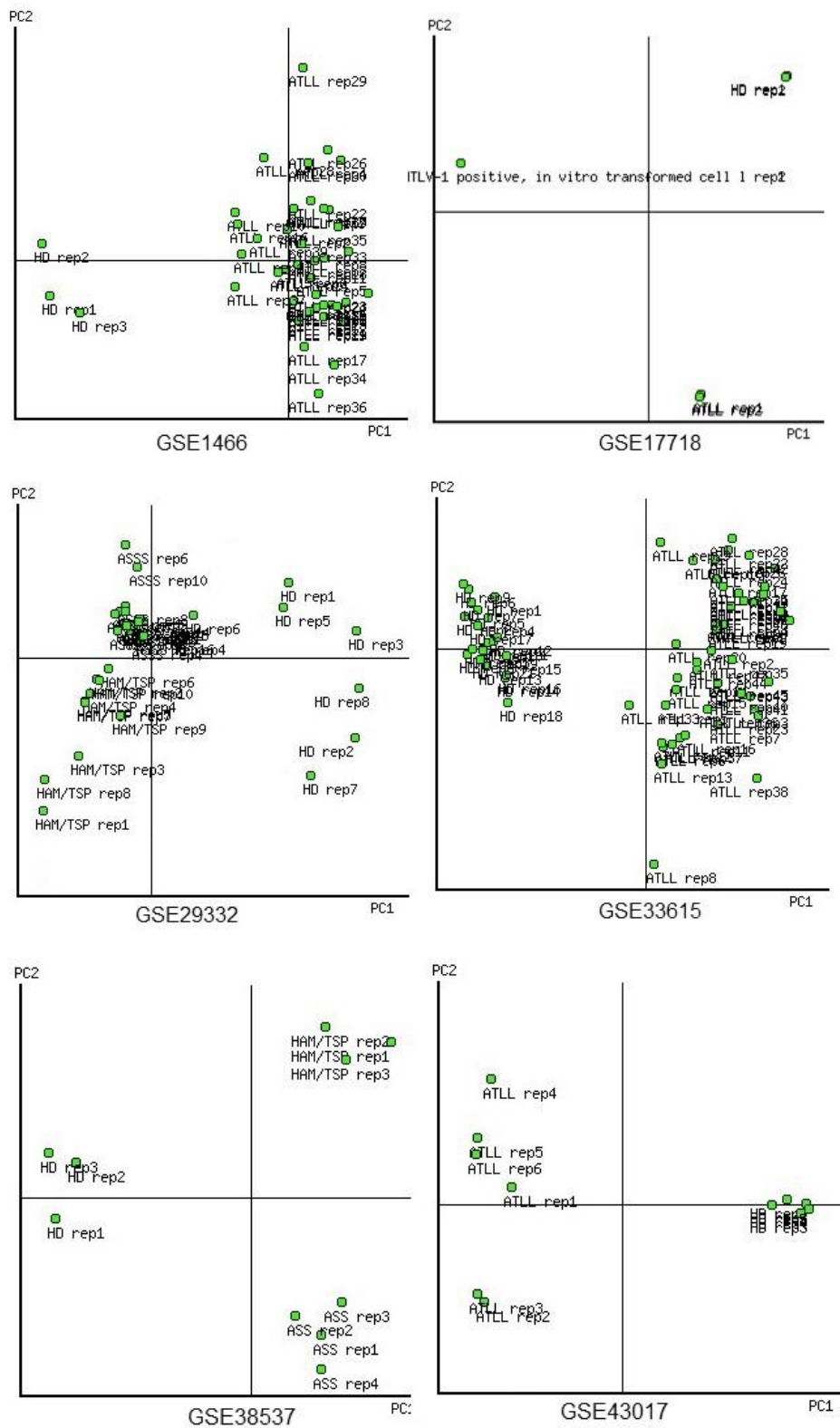
YOSHIDA, Noriaki et al. CCR4 frameshift mutation identifies a distinct group of adult T cell leukaemia/lymphoma with poor prognosis. **The Journal of pathology**, v. 238, n. 5, p. 621-626, 2016.

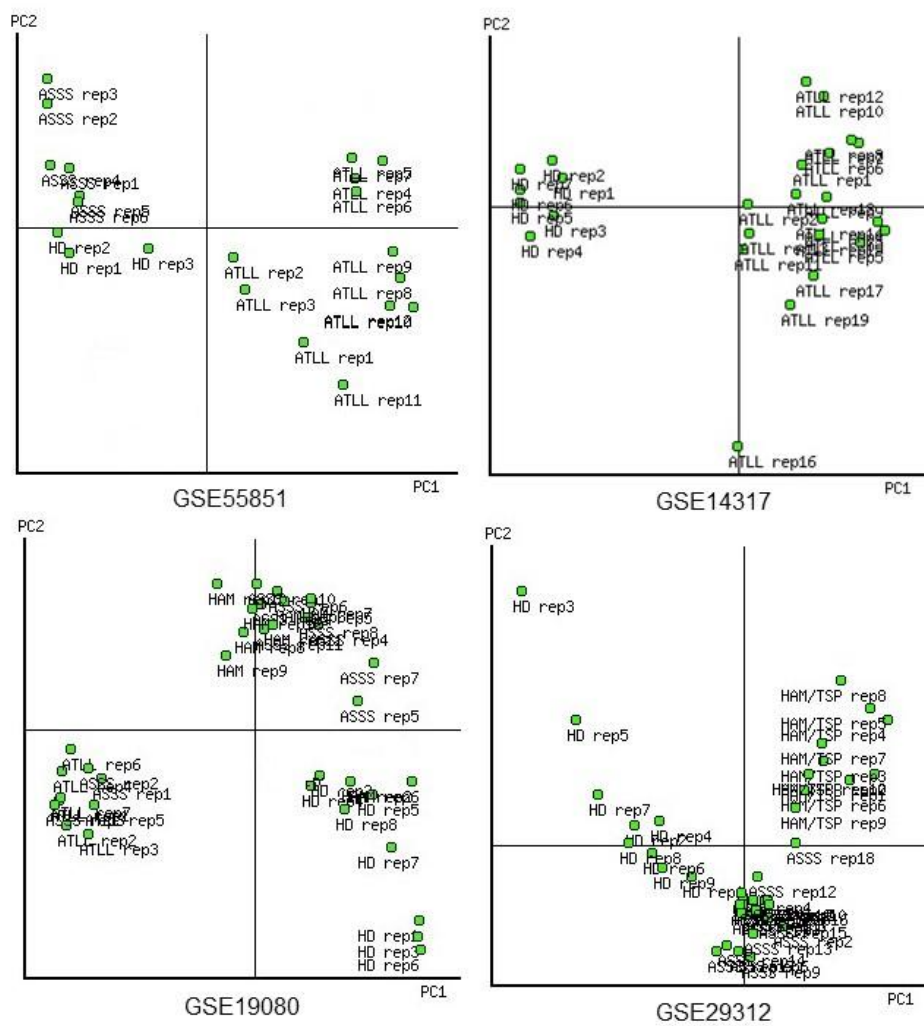
YUEN, Sze Chung; ZHU, Hongmei; LEUNG, Siu-wai. Building Molecular Interaction Networks from Microarray Data for Drug Target Screening. In: **Computational Drug Discovery and Design**. Humana Press, New York, NY, 2018. p. 179-197.

ZHAO, Tiejun. The role of HBZ in HTLV-1-induced oncogenesis. **Viruses**, v. 8, n. 2, p. 34, 2016.

ZHU, Hongmei; LEUNG, Siu-wai. Identification of Potential MicroRNA Biomarkers by Meta-analysis. In: **Computational Drug Discovery and Design**. Humana Press, New York, NY, 2018. p. 473-484.

## APÊNDICE A – PCAs DOS ESTUDOS INDIVIDUAIS





## APÊNDICE B – GDES COMUNS ENTRE OS STATUS CLÍNICOS

| <b>GDEs em comum</b>  |                   |                      |                            |
|-----------------------|-------------------|----------------------|----------------------------|
| <b>ATLL e HAM/TSP</b> | <b>ASS e ATLL</b> | <b>HAM/TSP e ASS</b> | <b>ATLL, HAM/TSP e ASS</b> |
| BATF3                 | POF1B             | NHLRC1               | ANK1                       |
| AIM2                  | P2RY1             | DUSP9                |                            |
| LMO4                  | PARVB             | ZNF474               |                            |
| KIF3B                 | TSPAN5            |                      |                            |
| MT1E                  | COL6A1            |                      |                            |
| LEPROT                |                   |                      |                            |
| HNF4G                 |                   |                      |                            |
| FDFT1                 |                   |                      |                            |
| HEG1                  |                   |                      |                            |
| LRRN3                 |                   |                      |                            |
| SNPH                  |                   |                      |                            |
| DNMT3A                |                   |                      |                            |
| RAI1                  |                   |                      |                            |
| YME1L1                |                   |                      |                            |
| HIC2                  |                   |                      |                            |
| CHMP6                 |                   |                      |                            |
| ZNF407                |                   |                      |                            |
| SFT2D3                |                   |                      |                            |



### APÊNDICE C – GENES HUBS E SEUS SCORES

| Genes Hubs   |      |          |          |
|--------------|------|----------|----------|
| Manifestação | Rank | Nome     | Score    |
| ATLL         | 1    | SOCS3    | 1.22E+48 |
|              | 2    | CUL3     | 1.22E+34 |
|              | 3    | KLHL5    | 1.22E+32 |
|              | 3    | KLHL22   | 1.22E+32 |
|              | 3    | FBXO2    | 1.22E+32 |
| HAMTSP       | 1    | PRPF8    | 2.10E+26 |
|              | 2    | SNRNP200 | 2.10E+26 |
|              | 3    | PRPF6    | 2.10E+25 |
|              | 4    | SNRNP40  | 2.10E+25 |
|              | 5    | PRPF19   | 2.10E+26 |
| ASS          | 1    | CDK1     | 3.63E+05 |
|              | 2    | CKS1B    | 3.63E+05 |
|              | 3    | CDC20    | 3.63E+05 |
|              | 4    | CKS2     | 3.63E+05 |
|              | 4    | CCNB1    | 3.63E+05 |