



**UNIVERSIDADE FEDERAL DA BAHIA - UFBA  
INSTITUTO DE CIÊNCIAS DA SAÚDE - ICS  
DEPARTAMENTO DE BIOQUÍMICA E BIOFÍSICA**



**PROGRAMA DE PÓS-GRADUAÇÃO MULTICÊNTRICO EM BIOQUÍMICA E  
BIOLOGIA MOLECULAR - PMBqBM**

**JÉSSICA LAIS ALMEIDA DOS SANTOS**

**INFECÇÃO PELO HTLV-1 E PROGRESSÃO PARA HAM/TSP:  
CONTRIBUIÇÃO DE MARCADORES MOLECULARES HUMANOS E VIRAIS**

**Tese de Doutorado**

**SALVADOR – BA**

**2020**

**JÉSSICA LAIS ALMEIDA DOS SANTOS**

**INFECÇÃO PELO HTLV-1 E PROGRESSÃO PARA HAM/TSP:  
CONTRIBUIÇÃO DE MARCADORES MOLECULARES HUMANOS E VIRAIS**

**ORIENTADORA: PROFA. DRA. ALINE CRISTINA ANDRADE  
MOTA MIRANDA MASCARENHAS**

Tese apresentada ao Programa de Pós-Graduação  
Multicêntrico em Bioquímica e Biologia Molecular  
para a obtenção do grau de Doutor.

**SALVADOR – BA**

**2020**

## TERMO DE APROVAÇÃO DA DEFESA DE TESE

### INFECÇÃO PELO HTLV-1 E PROGRESSÃO PARA HAM/TSP: CONTRIBUIÇÃO DE MARCADORES MOLECULARES HUMANOS E VIRAIS

Salvador, Bahia, 16 de Dezembro de 2020.

COMISSÃO EXAMINADORA:



---

Profa. Dra. Aline Cristina Andrade Mota Miranda Mascarenhas



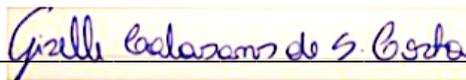
---

Prof. Dra. Maria Luiza Saraiva Pereira



---

Profa. Dra. Joana Paixão Monteiro Cunha



---

Profa. Dra. Giselle Calasans de Souza Costa



---

Profa. Dra Maria Fernanda de Castro Amarante

Dedico esse trabalho a Deus por ser o guia do meu destino  
e a fonte da minha força.  
E aos meus amados pais,  
Gilson e Simone, que são a luz da minha vida  
e meus maiores incentivadores.

## AGRADECIMENTOS

- ❖ À minha orientadora, Dra. Aline Cristina Andrade Mota Miranda Mascarenhas, a minha eterna gratidão. Seu profissionalismo, companheirismo e caráter fazem dela não somente uma orientadora, como também um ser humano especial. Foi um privilégio trabalhar ao seu lado por todos esses anos. “Todo o nosso carinho e gratidão àqueles que nos ensinam muito mais que teorias, que nos preparam também para a vida”;
- ❖ Agradeço ao nosso grupo de Bioinformática, Jéssica Duarte Souza (Jelzinha), Thiago Almeida (Thi), Esther Carvalho Chiaretto, Caroline Nascimento dos Anjos e as iniciantes científicas do trabalho, Beatriz Rolim Lobo (Bea) e Cinthia Cristina de Oliveira Costa (Cin) por todo apoio e contribuição para que o trabalho acontecesse;
- ❖ Ao Dr. Bernardo Galvão Castro Filho (Coordenador do CHTLV-Escola Bahiana de Medicina e Saúde Pública) e a Dra. Thessika Hialla Almeida Araújo pela constante colaboração e suporte no manejo da coorte de indivíduos infectados bem como nas demandas laboratoriais;
- ❖ Ao Dr. Ricardo Khouri e ao doutorando Thiago Cerqueira, pela parceria e apoio fundamentais para a execução e análise do Exoma;
- ❖ Ao Dr. Eduardo Tarazona, e ao Laboratório de Diversidade Genética Humana (LDGH) da Universidade Federal de Minas Gerais, sob sua liderança, pela preciosa contribuição científica, generosidade, disponibilidade, suporte e atenção nas análises do Exoma. Em especial a Camila Zolini pela generosidade e pelo serviço gratuito em muito ajudar;
- ❖ Às plataformas tecnológicas do IGM (Instituto Gonçalo Muniz) e Illumina do Lab. Multiusuário de Sequenciamento em Larga Escala e Expressão Gênica da UNESP, pelo suporte, em especial à Dra. Camila Cesario Fernandes, que foi nossa referência na realização do exoma;
- ❖ Ao Dr. Amancio José de Souza, coordenador do Nucleo de Pesquisa e Inovação (NUPI) da Escola Bahiana de Medicina e Saúde Pública, pela colaboração na disponibilização do espaço para a realização das etapas moleculares e por toda atenção cuidadosa sempre dispensada;
- ❖ À Dra. Fernanda Khouri Barreto (UFBA – Campus Vitória da Conquista) e a Biomédica Marina Cucco pelo suporte e presença constante nas etapas moleculares

de estudo do vírus;

❖ Aos professores, colegas e funcionários do Programa de Pós-Graduação Multicêntrico em Bioquímica e Biologia Molecular;

❖ À FAPESB, pelo apoio financeiro;

❖ Aos meus familiares, especialmente meus irmãos Igor, Tico e Josi;

❖ Aos todos os meus amigos que me sustentaram e me incentivaram durante toda a trajetória;

❖ A todos que direta ou indiretamente ajudaram na realização deste trabalho.

Muito Obrigada!

## RESUMO

O vírus linfotrófico de células T humana do tipo 1 (HTLV-1) é um retrovírus humano sexualmente transmissível, cujas principais manifestações clínicas são a Mielopatia Associada ao HTLV/Paraparesia Espástica Tropical (HAM/TSP) e a Leucemia/Linfoma de Células T do Adulto (ATLL). A HAM/TSP é uma manifestação inflamatória do sistema nervoso central que pode resultar na perda parcial dos movimentos dos membros inferiores. Embora muitos aspectos dos eventos que levam ao desenvolvimento de doença não estejam esclarecidos, a resposta imune do hospedeiro, principalmente a resposta celular desencadeada por células T CD8<sup>+</sup> específicas, é reconhecida como um evento crucial. O objetivo desse trabalho é investigar marcadores moleculares no vírus e no hospedeiro que possam estar relacionados à manifestação de doença, principalmente a HAM/TSP. No Capítulo I, foi realizada a busca por marcadores virais, no que diz respeito à apresentação diferencial de epítomos, para os diferentes perfis clínicos. Para tanto, fizemos um levantamento (Genbank) de 46 genomas completos do HTLV-1, e separamos as regiões codificantes das proteínas Tax, HBZ, p12, gp21 e gp46. Tais regiões gênicas foram avaliadas quanto a presença de SNPs (*Single Nucleotide Polymorphism*) e em seguida traduzidas para proteína. As variações moleculares foram representadas em sequências de AA que foram submetidas ao IEDB (*Immune Epitope Database a Analysis Resource*) para predição de epítomos lineares ligantes de moléculas de HLA-I. Foi possível identificar 7 “alelos protéicos” para a proteína gp21, 23 para gp46, 20 para HBZ, 13 para a proteína Tax, e 15 para a proteína p12, todos distribuídos entre indivíduos ASS, fHAM/TSP (Familiar), sHAM/TSP (Esporádico) e ATLL. Na predição, foram identificados mais de 15.741 epítomos para p12, menor quantidade de epítomos preditos e mais de 50.000 para Tax, com a maior quantidade de epítomos preditos. A proteína p12 teve a maior proporção de epítomos específicos, também seguida pela proteína Tax. A proteína HBZ teve os menores percentuais em todas as faixas de especificidade. Os resultados encontrados para as proteínas estruturais gp46 e gp21 foram semelhantes. Não houve nenhuma especificidade/exclusividade dos alelos de HLA entre as proteínas, “alelos protéicos” e/ou perfis clínicos do hospedeiro. Cada proteína apresentou um alelo de HLA como sendo o mais frequente entre os “alelos protéicos”, não havendo similaridade nessa ocorrência. E por fim, o alelo de HLA\*A 32:01 foi citado entre os mais frequentes em todas as proteínas, e as proteínas gp21 e HBZ apresentaram perfis semelhantes no que se refere à ocorrência dos alelos de

HLA mais frequentes. O perfil clínico HAM/TSP foi o que apresentou o maior número de variações moleculares entre as proteínas estudadas: foram 42 “alelos protéicos” contra 19 de indivíduos com ATLL e 16 de indivíduos ASS (Não HAM/TSP). Os achados de afinidade revelaram que dos 9 epítomos que apresentaram diferenças na afinidade às moléculas de HLA, 6 deles estavam associados a indivíduos ATLL, 4 a indivíduos sHAM/TSP e 2 a indivíduos ASS. Concluímos não haver um perfil de apresentação de epítomos que esteja associado ao status clínico do hospedeiro, uma vez que houve grande semelhança quanto aos alelos de HLA e quanto à posição dos epítomos apresentados entre os “alelos protéicos”. No entanto, identificamos variações moleculares importantes na predição dos epítomos especialmente nos genomas de indivíduos com ATLL e HAM/TSP. No Capítulo II, buscamos por marcadores no hospedeiro, em um grupo de indivíduos (HAM/TSP e assintomáticos) (N=44) infectados. Para tanto, realizamos a determinação do Exoma através da Plataforma Illumina e utilizando um chip para 551.004 SNPs. A taxa total de genotipagem das amostras foi de 0,997001, com perda de 35.163 SNPs (~6%), identificadas 515.841 variantes. A determinação ancestral revelou uma proporcionalidade de contribuição genética mais marcante para a origem africana na população estudada. Observamos predominância de mulheres no grupo caso, como já foi documentado na literatura. A média de idade dos indivíduos infectados foi de 59 anos (27-89 anos), sendo que a carga proviral média destes é de  $55.253,41/10^6$  células. Dentre os SNPs identificados, quatro deles se destacaram quanto às frequências alélicas nos grupos caso e controle, e são eles: rs2857596\_C, rs7917905\_A, rs1265564\_C e rs376863\_A. Após análise de regressão multivariada, ajustada para ancestralidade, apenas os SNPs “rs1265564” ( $p=7,2*10^{-4}$ ) e “rs376863” ( $p=1,25*10^{-3}$ ) se apresentam associados ao desfecho. O SNP rs1265564, foi identificado no gene que codifica para a proteína CUX-2 que está associada ao reparo de DNA oxidado por ação das espécies reativas do oxigênio produzidas nos neurônios. Os quatro SNPs iniciais mostram estar em desequilíbrio de ligação com outros SNPs, porém nenhum deles dos já descritos na literatura e citados neste trabalho. Sugerimos uma busca minuciosa dos SNPs indicados neste trabalho, bem como a associação de marcadores moleculares do hospedeiro com marcadores virais, como forma de entender o processo de infecção e manifestação de doença como resultado de uma integração complexa e bem sucedida.

**Palavras-chave:** HTLV-1, HAM/TSP, SNPs, epítomos, vírus, hospedeiro,

## ABSTRACT

Human T-cell lymphotropic virus type 1 (HTLV-1) is a sexually transmitted human retrovirus. Its main clinical manifestations are HTLV-Associated Myelopathy/Tropical Spastic Paraparesis (HAM/TSP) and T-Cell Leukemia / Lymphoma of the Adult (ATLL). HAM/TSP is an inflammatory disorder of the central nervous system that can result in partial loss of lower limb movements. Although many events aspects that lead to the disease development are unclear, the host's immune response, especially the cellular response triggered by specific CD8<sup>+</sup> T cells, is recognized as a crucial event. The objective of this work is to investigate molecular markers in the virus and in the host that may be related to the manifestation of the disease, mainly HAM/TSP. In Chapter I we have search for viral markers, with regard to the differential epitope presentation, for the different clinical profiles. We carried out a survey (Genbank) of 46 HTLV-1 complete genomes, and separated the coding regions of Tax, HBZ, p12, gp21 and gp46 proteins. Such gene regions were evaluated for the presence of SNPs (Single Nucleotide Polymorphism), and then translated. These molecular variations were represented in AA sequences that were submitted to the IEDB (Immune Epitope Database Analysis Resource) to predict linear epitopes linking HLA-I molecules. It was possible to identify 7 "protein alleles" for the gp21 protein, 23 for gp46, 20 for HBZ, 13 for the Tax protein, and 15 for the p12 protein, all distributed among individuals ASS, fHAM/TSP (Familiar), sHAM/TSP (Sporadic) and ATLL. In prediction results, more than 15,741 epitopes were identified for p12, the lowest amount of predicted epitopes and more than 50,000 for Tax, with the highest amount of predicted epitopes. The p12 protein had the highest proportion of specific epitopes, also followed by the Tax protein. The HBZ protein had the lowest percentages in all specificity ranges. The results found for structural proteins gp46 and gp21 were similar. There was no specificity/exclusivity of HLA alleles among proteins, "protein alleles" and/or host clinical profiles. Each protein presented an HLA allele as being the most frequent among the "protein alleles", with no similarity in this occurrence. Finally, the HLA \* A 32:01 allele was cited as the most frequent in all proteins, and the gp21 and HBZ proteins showed similar profiles with regard to the occurrence of the most frequent HLA alleles. The HAM/TSP clinical profile showed the highest number of molecular variations among the studied proteins: there were 42 "protein alleles" against 19 of individuals with ATLL and 16 of ASS individuals (Non-HAM/TSP). The affinity findings revealed that 9 epitopes showed differences in affinity to HLA molecules, 6 of them were associated with ATLL

individuals, 4 with sHAM/TSP individuals and 2 with ASS individuals. We conclude that there is no epitope presentation profile that is associated with the host clinical status, since there was great similarity in the HLA alleles and in the predicted epitopes positions among the “protein alleles”. However, we have identified important molecular variations in the epitopes prediction, especially in the genomes of individuals with ATLL and HAM/TSP. In Chapter II, we looked for markers in the host, in a group of infected individuals (HAM/TSP and asymptomatic) (N = 44). For this, we have performed the determination of Exoma through the Illumina Platform and using a chip for 551,004 SNPs. The overall samples genotyping rate was 0.997001, with a loss of 35,163 SNPs (~ 6%), identified 515,841 variants. The ancestral determination revealed a proportionality of genetic contribution more marked for the African origin in the studied population. We observed a predominance of women in the case group, as has already been documented in the literature. The average age of infected individuals was 59 years (27-89 years), with an average proviral load of 55,253.41/10<sup>6</sup> cells. Among the SNPs identified, four of them stood out regarding the allele frequencies in the case and control groups, and they are: rs2857596\_C, rs7917905\_A, rs1265564\_C and rs376863\_A. After multivariate regression analysis, adjusted for ancestry, only the SNPs “rs1265564” ( $p = 7.2 * 10^{-4}$ ) and “rs376863” ( $p = 1.25 * 10^{-3}$ ) are associated with the outcome. The SNP rs1265564, was identified in the gene that codes for the CUX-2 protein that is associated with the repair of oxidized DNA by the action of reactive oxygen species produced in neurons. The initial four SNPs show to be in linkage disequilibrium with other SNPs, but none of those already described in the literature and cited in this work. We suggest a thorough search for the SNPs indicated in this work, as well as the association of host molecular markers with viral markers, as a way of understanding the process of infection and disease manifestation as a result of a complex and successful integration.

**Keywords:** HTLV-1, HAM/TSP, SNPs, epitopes, virus, host

## LISTA DE FIGURAS

<b>Figura 1</b>	Representação esquemática da estrutura morfológica do HTLV-1.	18
<b>Figura 2</b>	Diagrama esquemático do genoma do HTLV-1 com destaque para as moléculas de mRNA produzidas no processo de transcrição viral.	19
<b>Figura 3</b>	Epidemiologia do HTLV-1 ao redor do mundo.	24
<b>Figura 4</b>	Curso da infecção pelo HTLV-1 e suas formas de disseminação.	28
<b>Figura 5</b>	Esquema ilustrativo do principal mecanismo associado ao desenvolvimento da HAM/TSP.	32
<b>Figura 6</b>	Esquema ilustrativo da Hipótese para a Apresentação diferencial de epítomos de proteínas estruturais e regulatórias do HTLV-1 em indivíduos com diferentes perfis clínicos.	38
<b>Figura 7</b>	Desenho do Estudo da Apresentação diferencial de epítomos de proteínas estruturais e regulatórias do HTLV-1 em indivíduos com diferentes perfis clínicos.	40
<b>Figura 8</b>	Avaliação Quantitativa dos Epítomos mais específicos entre as proteínas estudadas.	49
<b>Figura 9</b>	Balanço quantitativo dos epítomos preditos, levando em consideração a proteína e seus "alelos".	52
<b>Figura 10</b>	Hipótese para o estudo de Caracterização do exoma de indivíduos infectados pelo HTLV-1: investigação de biomarcadores para manifestação de HAM/TSP.	73
<b>Figura 11</b>	Análise de Componentes Principais da população de estudo.	84
<b>Figura 12</b>	Modelo Teórico Explicativo de associação entre SNPs e a manifestação de HAM/TSP, construído através de Gráfico Acíclico Direcionado (DAG).	87
<b>Figura 13</b>	Identificação dos SNPs associados à HAM/TSP através do <i>Manhattan Plott</i>	89

## LISTA DE TABELAS

<b>Tabela 1</b>	Caracterização dos genomas virais utilizados para a predição de epítomos.	44
<b>Tabela 2</b>	Descrição dos "alelos protéicos" identificados e submetidos à predição de epítomos.	45
<b>Tabela 3</b>	Alelos de HLA previamente descritos na literatura com alguma associação com a infecção pelo HTLV-1 ou sintomatologia.	47
<b>Tabela 4</b>	Avaliação Quantitativa do total de predições identificadas por proteína viral.	48
<b>Tabela 5</b>	Avaliação da ocorrência dos alelos de HLA mais frequentes por proteína viral.	51
<b>Tabela 6</b>	Avaliação da frequência dos epítomos mais comuns, identificados na gp46 com especificidade de 99 a 100%, entre os "alelos protéicos".	53
<b>Tabela 7</b>	Avaliação da frequência dos epítomos mais comuns, identificados na gp21 com especificidade de 99 a 100%, entre os "alelos protéicos".	54
<b>Tabela 8</b>	Avaliação da frequência dos epítomos mais comuns, identificados na proteína Tax com especificidade de 99 a 100%, entre os "alelos protéicos".	56
<b>Tabela 9</b>	Avaliação da frequência dos epítomos mais comuns, identificados no HBZ com especificidade de 99 a 100%, entre os "alelos protéicos".	57
<b>Tabela 10</b>	Avaliação da frequência dos epítomos mais comuns, identificados na proteína p12 com especificidade de 99 a 100%, entre os "alelos protéicos".	58
<b>Tabela 11</b>	Avaliação Descritiva das diferenças encontradas nas afinidades dos epítomos mais comuns entre os diferentes "alelos protéicos".	60
<b>Tabela 12</b>	Descrição do <i>Infinium Core Exome-24 v1.3 BeadChip</i> .	76
<b>Tabela 13</b>	Caracterização da População submetida à determinação do exoma.	81
<b>Tabela 14</b>	Proporção dos Três pilares da Ancestralidade Brasileira na população de estudo.	85
<b>Tabela 15</b>	Análise Descritiva das Covariáveis nos Grupos Caso e Controle.	87
<b>Tabela 16</b>	Descrição dos SNPs e as frequências alélicas que mostraram ser diferentes entre os grupos caso e controle.	88
<b>Tabela 17</b>	Análise bivariada e Multivariada dos genótipos identificados em relação aos grupos Caso e Controle.	90
<b>Tabela 18</b>	Avaliação do desequilíbrio de ligação dos SNPs encontrados em associação com a HAM/TSP.	92
<b>Tabela 19</b>	Relação dos SNPs, já descritos na literatura, com associação à HAM/TSP.	93
<b>Tabela 20</b>	Frequência alélica da análise direcionada por SNPs da literatura.	93
<b>Tabela 21</b>	Breve levantamento bibliográfico sobre os possíveis efeitos das covariáveis (Idade, gênero e carga proviral) sobre desenvolvimento da HAM/TSP.	97

## LISTA DE ABREVIATURAS E SIGLAS

<b>AA</b>	Aminoácido
<b>AFR</b>	Africano
<b>AMR</b>	População Miscigenada
<b>APC</b>	<i>Antigen Presenting Cell</i> - Células Apresentadoras de Antígeno
<b>ASS</b>	Assintomático
<b>ATF</b>	<i>Activating Transcription factor</i> – Fator de Ativação Transcricional
<b>ATK1</b>	Cepa referência do genoma completo do HTLV-1
<b>ATLL</b>	<i>Adult T-cell Leukemia/Lymphoma</i> - Leucemia/Linfoma de Células T do Adulto
<b>cAMP</b>	<i>Cyclic Adenosine Monophosphate</i> Adenosina Monofosfato Cíclico
<b>CA</b>	Capsídeo
<b>CD</b>	Célula Dendrítica
<b>CBP</b>	<i>Binding protein</i> - co-ativador transcricional
<b>CHTLV</b>	Centro Integrativo de HTLV
<b>CREB/p300</b>	<i>CREB-binding protein/p300</i> – Proteína de Ligação à CREB/p300
<b>CTL</b>	<i>Cytotoxic T lymphocytes</i> - Linfócitos T citotóxicos
<b>DAG</b>	<i>Directed Acyclic Graph</i> <i>Gráfico Acíclico Direcionado</i>
<b>DC</b>	<i>Dendritic Cell</i> Célula Dendrítica
<b>DL</b>	Desequilíbrio de ligação
<b>DNA</b>	<i>Deoxyribonucleic acid</i> Ácido Desoxirribonucléico
<b>ELISA</b>	<i>Enzyme-Linked Immunosorbent Assay</i> – Ensaio imuno enzimático
<b>Env</b>	Envelope viral
<b>RE</b>	Retículo endoplasmático
<b>EUR</b>	Europeu
<b>FP</b>	Fusion Peptide
<b>gp46</b>	Glicoproteína de superfície (46kDa)
<b>gp21</b>	Glicoproteína Transmembrana (21kDa)
<b>GLUT-1</b>	<i>Glucose Transporter</i> - Transportador de glicose tipo 1
<b>HAM/TSP</b>	<i>HTLV-1 Associated Myelopathy/Tropical Spastic Paraparesis</i> Mielopatia associada ao HTLV/ Paraparesia Espástica Tropical
<b>HBZ</b>	<i>HTLV-1 Basic Leucine Zipper</i> Zíper de Leucina Básico do HTLV-1
<b>sHBZ</b>	<i>spliced HBZ</i> - Proteína HBZ que sofreu processamento
<b>usHBZ</b>	<i>unspliced HBZ</i> - Proteína HBZ que não sofreu processamento
<b>HBZAD</b>	<i>HBZ activation domain</i> - Domínio de ativação do HBZ

<b>Hc</b>	<i>Heavy chain - Cadeia Pesada</i>
<b>HLA</b>	<i>Human Leukocyte Antigen</i> Antígeno leucocitário humano
<b>HSPG</b>	<i>Heparan sulfate proteoglycans</i> – Proteoglicano heparan sulfato
<b>HTLV-1</b>	<i>Human T-cell Lymphotropic Virus Type 1</i> Vírus linfotrópico de células T humanas do tipo 1
<b>HTLV-2</b>	<i>Human T-cell Lymphotropic Virus Type 2</i> Vírus Linfotrófico de Célula T Humana do tipo 2
<b>IC<sub>50</sub></b>	Concentração Inibitória Mínima
<b>IEDB</b>	<i>Immune Epitope Database</i> – Banco de Dados de Epítomos
<b>ISD</b>	<i>Immune Suppressor Domain</i> - Domínio imunossupressor
<b>IFN-<math>\gamma</math></b>	Interferon- $\gamma$ – Citocina do tipo II
<b>IL-1/IL-4/ IL-6/IL-10</b>	Interleucina – 1/2/6/10
<b>IN</b>	Integrase
<b>LTR</b>	<i>Long Terminal Repeats</i> - Longas Repetições Terminais
<b>LCR</b>	Líquido cefalorraquidiano
<b>LDHG</b>	Laboratório Diversidade Genética Humana
<b>MA</b>	Matriz
<b>MHC-I</b>	<i>Major Histocompatibility Complex class I</i> Complexo principal de histocompatibilidade de classe I
<b>MTOC</b>	<i>Microtubule Organizing Center</i> Centro de organização microtubular
<b>NAT</b>	Nativo Americano
<b>NC</b>	Nucleocapsídeo
<b>NCBI</b>	<i>National Center for Biotechnology Information</i> Centro Nacional para informação de Biotecnologia
<b>NFAT</b>	<i>Nuclear factor of activated T-cells</i> Fatores Nucleares de ativação de células T
<b>NF<math>\kappa</math>B</b>	<i>Nuclear Factor kappa B</i> – Fator nuclear Kappa B
<b>NK</b>	<i>Natural Killer</i> - Células do tipo NK (Exterminadoras)
<b>NRP-1</b>	<i>Neurophilin-1</i> - Neurofilina-1
<b>NUPI</b>	Núcleo de Pesquisa e Inovação
<b>ORF</b>	<i>Open reading frame</i> - Fase de leitura aberta
<b>Pb</b>	Pares de base
<b>PBMC</b>	<i>Peripheral blood mononuclear cells</i> Células mononucleares do sangue periférico
<b>PCA</b>	<i>Principal Component Analysis</i> Análise de Componentes Principais
<b>PCR</b>	Polimerase Chain Reaction - Reação em Cadeia da Polimerase
<b>PF</b>	Peptídeo de fusão
<b>p19</b>	Proteína da matriz (19kDa)
<b>p24</b>	Proteína do capsídeo (24kDa)

<b>p15</b>	Proteína do nucleocapsídeo (15kDa)
<b>RE</b>	Retículo endoplasmático
<b>RNA</b>	<i>Ribonucleic acid</i> - Ácido ribonucleico
<b>RNAseH</b>	Ribonuclease H (Endonuclease não específica)
<b>iRNA</b>	<i>Interference RNA</i> - RNA de interferência
<b>mRNA</b>	Messenger <i>Ribonucleic acid</i> Ácido ribonucleico mensageiro
<b>SD</b>	<i>Standard deviation</i> – Desvio Padrão
<b>SNC</b>	<i>Sistema Nervoso Central</i>
<b>SNP</b>	<i>Single Nucleotide Polymorphism</i> Polimorfismos de Nucleotídeo Único
<b>SQL</b>	<i>Structured Query Language</i> Linguagem de programação (Linguagem de pesquisa declarativa padrão para banco de dados )
<b>STLV</b>	<i>Simian T-cell Lymphotropic Virus</i> Vírus Linfotrópico de Célula T de Símio
<b>SU</b>	Glicoproteína de superfície
<b>TB</b>	Tuberculose
<b>TCR</b>	<i>T-cell Receptor</i> – Receptor de Célula T
<b>Th1/2</b>	<i>T helper cell Type 1/2</i> - Células T auxiliares do tipo 1/2
<b>TLR</b>	<i>Toll Like Receptor</i> - Receptor do tipo Toll
<b>TNF</b>	<i>Tumoral Necrosis Factor</i> - Fator de necrose tumoral
<b>TNF-a</b>	<i>Tumor necrosis factor-alpha</i> -Fator de necrose tumoral alfa
<b>TM</b>	Proteína transmembrana
<b>TR</b>	Transcriptase reversa

## SUMÁRIO

<b>I. INTRODUÇÃO</b> .....	<b>17</b>
I.1 O vírus .....	17
I.1.1 Características Estruturais .....	17
I.1.2 Características Genômicas e Funcionais .....	18
I.2 A infecção .....	22
I.2.1 Epidemiologia .....	23
I.2.2 Ciclo de Infecção e Persistência .....	26
I.2.3 Resposta imune .....	29
I.2.4 Manifestações Clínicas Associadas .....	30
I.3 Progressão para HAM/TSP .....	31
I.3.1 Marcadores Moleculares Virais e do Hospedeiro .....	33
<b>II. CAPÍTULO I: Apresentação diferencial de epítomos de proteínas estruturais e regulatórias do HTLV-1 em indivíduos com diferentes perfis clínicos</b> .....	<b>35</b>
II.1 Justificativa .....	36
II.2 Hipótese .....	38
II.3 Objetivos .....	39
II.3.1 Objetivo Geral .....	39
II.3.2 Objetivos Específicos .....	39
II.4 Metodologia.....	40
II.4.1 Seleção e Organização do Dataset .....	40
II.4.2 Predição de epítomos ligantes de moléculas de HLA classe 1..	41
II.5 Resultados .....	43
II.5.1 Seleção e Organização do Dataset .....	43
II.5.2 Predição de epítomos ligantes de moléculas de HLA classe 1..	48
II.5.2.1 Avaliação Geral das predições .....	48
II.5.2.2 Caracterização dos epítomos preditos: frequência dos epítomos com especificidade entre 99-100%.....	52
II.5.2.2.1 gp46 .....	53
II.5.2.2.2 gp21 .....	54
II.5.2.2.3 Tax .....	55
II.5.2.2.4 HBZ .....	57
II.5.2.2.5 p12 .....	58
II.5.2.3 Caracterização dos epítomos preditos: afinidade dos epítomos com especificidade entre 99-100%.....	59
II.6 Discussão .....	62

## SUMÁRIO

<b>III. CAPÍTULO II: Caracterização do exoma de indivíduos infectados pelo HTLV-1: investigação de biomarcadores para manifestação de HAM/TSP</b>	<b>70</b>
III.1 Justificativa .....	71
III.2 Hipótese .....	73
III.3 Objetivos .....	74
III.3.1 Objetivo Geral .....	74
III.3.2 Objetivos Específicos .....	74
III.4 Metodologia.....	75
III.4.1 Seleção da Amostra e Detecção Molecular do HTLV-1.....	75
III.4.2 Quantificação da Carga Proviral .....	76
III.4.3 Determinação do Exoma.....	76
III.4.3.1 Controle de qualidade e limpeza de dados para genotipagem .....	77
III.4.3.2 Estruturação da população .....	77
III.4.3.3 Ancestralidade genética .....	78
III.4.3.4 Genotipagem e Análises estatísticas .....	78
III.4.3.5 Avaliação do desequilíbrio de ligação dos SNPs .....	79
III.4.3.6 Levantamento Bibliográfico: busca por SNPs de susceptibilidade ou associação com a HAM/TSP.....	79
III.5 Resultados .....	80
III.5.1 Caracterização da amostra .....	80
III.5.2 Análise dos Dados Moleculares do Exoma .....	82
III.5.2.1 Controle de Qualidade.....	82
III.5.2.2 Estruturação da população .....	83
III.5.2.3 Ancestralidade genética .....	84
III.5.2.4 Genotipagem e Análises Estatísticas .....	86
III.5.2.5 Avaliação do desequilíbrio de ligação dos SNPs	91
III.5.2.6 Levantamento Bibliográfico: busca por SNPs de susceptibilidade ou associação com a HAM/TSP	92
III.6 Discussão .....	94
<b>IV. CONCLUSÃO</b> .....	<b>99</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>100</b>
<b>ANEXOS</b> .....	<b>117</b>

## I. INTRODUÇÃO

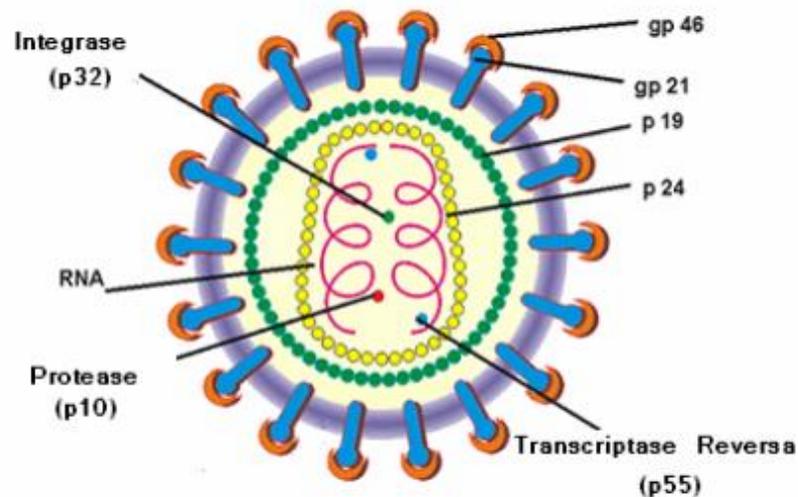
### I.1 O Vírus

O vírus linfotrópico de células T humanas tipo 1 (HTLV-1) foi o primeiro retrovírus humano descrito. Foi isolado no Japão em 1980 de um paciente com linfoma cutâneo de células T (POIESZ et al., 1980; MIYOSHI et al., 1981; YOSHIDA et al., 1982). O HTLV-1 pertence à família Retroviridae e ao gênero Deltaretrovirus (GESSAIN et al., 1992; FRANCHINI, 1995; SLATTERY et al., 1999). A principal hipótese de transmissão viral sugere ter ocorrido a partir de primatas não humanos infectados por STLV (*Simian T-cell Lymphotropic Virus*) através de atividades de caça e uso da carne destes animais para alimentação (SLATTERY et al., 1999; MOSSOUN et al., 2017).

#### I.1.1 Características Estruturais

As partículas virais possuem tamanho médio de 80–120nm (ISHAK et al. 2020), e apresentam estrutura morfológica similar à de outros retrovírus, sendo constituída, basicamente, por um envelope, uma matriz protéica e um nucleocapsídeo (**Figura 1**). O envelope viral é composto por uma glicoproteína de superfície (SU) extracelular, gp46, e uma proteína transmembrana (TM) que está ancorada na SU, chamada de gp21. Junto à membrana do envelope se encontra a proteína p19 da matriz (MA). O capsídeo (CA) de simetria icosaédrica composto pela proteína p24, abriga, no seu interior, o genoma viral definido por duas fitas de RNA diméricas de polaridade positiva, às quais estão associadas a várias pequenas proteínas básicas, chamadas de proteínas do nucleocapsídeo (NC). Outras proteínas também estão presentes no interior do CA, como a transcriptase reversa (TR) e a integrase (IN), essenciais no processo de integração do DNA no genoma da célula hospedeira (**Figura 1**) (RAYNE et al., 2004; MARTIN et al., 2018).

**Figura 1** – Representação esquemática da estrutura morfológica do HTLV-1.



Glicoproteínas do envelope viral (gp21 e gp26), Ácido Ribonucléico: RNA, proteínas estruturais do capsídeo viral (p19) e nucleocapsídeo viral (p24) (Fonte: [www.HTLV.com.br](http://www.HTLV.com.br)).

### I.1.2 Características Genômicas e Funcionais

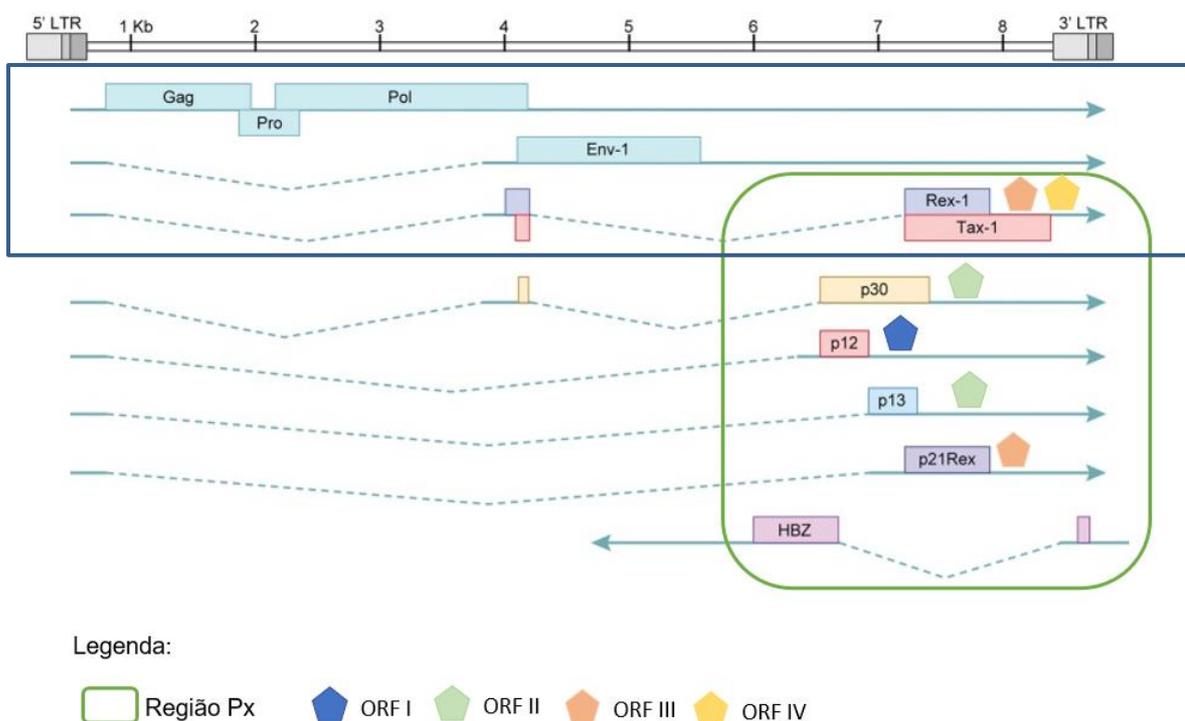
O HTLV tem um genoma de RNA de fita simples, com extensão de 9032 pb, e organização similar a de outros retrovírus (SEIKI et al., 1984). Apresenta as extremidades (5' e 3') flanqueadas por duas regiões repetidas, chamadas de “*Long Terminal Repeats*” (LTR), constituídas por três subunidades: U3, R e U5 (**Figura 2**).

No processo de transcrição, três moléculas de mRNA são produzidas: o mRNA genômico, utilizado para a síntese dos produtos dos genes *gag* e *pol*, transcrito da extremidade LTR 5' até a junção R-U5 na extremidade LTR3'; o mRNA sub-genômico, sintetizado a partir de uma única etapa de processamento e codificante do produto do gene *env*; e um segundo mRNA sub-genômico, duplamente processado através da remoção de dois íntrons, que codifica as proteínas regulatórias Tax e Rex com, pelo menos, quatro fases de leitura aberta (*Open Reading Frame-ORF*) (GHEZELDASHT et al., 2013).

O comprimento total do mRNA do HTLV-1 codifica a proteína *gag* (p55) que é então clivada pela protease viral que forma as proteínas da matriz (p19), capsídeo (p24) e nucleocapsídeo (p15). A repetição de terminal longo de HTLV-I (LTR) nas extremidades 5' e 3' do genoma contém o promotor viral e os elementos reguladores. Além disso, o mRNA viral codifica a proteína *pol*, o mRNA de *splicing* único codifica a proteína Env e um mRNA de *splicing* duplo codifica as proteínas regulatórias Tax e Rex (GHEZELDASHT et al., 2013).

Desta forma, o genoma viral possui os genes *gag*, *pro*, *pol* e *env*, além de uma sequência próxima à extremidade 3' conhecida como região pX, a qual codifica a principal proteína transativadora, Tax e o regulador de splicing de mRNA, Rex que compreende as ORFs de I a IV (**Figura 2**). Tais ORFs produzem formas alternativas de mRNA, as quais codificam as quatro proteínas acessórias p12<sup>I</sup>, p27<sup>I</sup>, p13<sup>II</sup> e p30<sup>II</sup>. O gene *gag* codifica a poliproteína Gag, que consiste em três principais domínios, matriz (MA), capsídeo (CA) e nucleocapsídeo (NC). Os genes *pro* e *pol*, codificam as enzimas virais, como é o caso da protease e das enzimas fundamentais para replicação viral, integrase e transcriptase, respectivamente. E por fim o gene *env* que origina as glicoproteínas de envelope (**Figura 2**) (MICHAEL et al., 2004; ROSADAS & PUCCIONI-SOHLER, 2015; BARRETO et al., 2016; WEIXIN et al., 2018, HARROD, 2019).

**Figura 2** – Diagrama esquemático do genoma do HTLV-1 com destaque para as moléculas de mRNA produzidas no processo de transcrição viral.



O retângulo azul destaca as três principais moléculas de mRNA produzidas no processo de transcrição viral. Fonte: Adaptado de Martinez et al., 2019.

O gene *gag* é inicialmente traduzido como um precursor poliprotéico, cuja subsequente clivagem dá origem às proteínas estruturais maduras: a proteína da

matriz de 19kDa (p19), a proteína do capsídeo de 24 kDa (p24) e a proteína do nucleocapsídeo de 15kDa (p15). Tais produtos protéicos estão frequentemente presentes nas estruturas formadas pelas células infectadas pelo HTLV-1, para a interação à distância entre células e conseqüente disseminação do vírus (OMSLAND et al., 2018). A protease, por sua vez, é codificada por desvio de leitura dos genes *gag* e *pol*, e atua sobre as cadeias poliprotéicas, clivando-as para formação das proteínas estruturais maduras encontradas na partícula viral. O gene *pol* codifica a transcriptase reversa e a integrase, sendo a primeira enzima responsável pela síntese do DNA viral a partir do RNA de fita simples e, portanto, fundamental na fase inicial do ciclo de multiplicação do retrovírus.

O gene *env* codifica a proteína precursora do envelope viral (Env), que é clivada, gerando os produtos maduros: a glicoproteína de superfície de 46kDa (gp46-SU) e a proteína transmembrana de 21kDa (gp21-TM), que estão diretamente associadas ao reconhecimento celular, e à conseqüente entrada do vírus na célula.

As proteínas regulatórias Tax (ORF-IV) e Rex (ORF-III), por sua vez, são codificadas pela transcrição do gene *pX*, e têm função importante na dinâmica viral, de modo que cabe à proteína Tax a ativação de genes celulares que codificam fatores de transcrição, como NF- $\kappa$ B, fatores ligantes dos elementos responsivos ao cAMP, IL-2 e receptor de IL-2. Além disso, Tax abriga epítomos (11-19aa) importantes que fazem dela um alvo antigênico dominante no reconhecimento por linfócitos T citotóxicos na maioria dos indivíduos. Na patogênese da infecção pelo HTLV-1, a proteína regulatória viral, Tax, funciona como um agente fundamental no desenvolvimento das diferentes patologias (KANNAGI et al., 2019; MARTINEZ et al., 2019).

Além das proteínas Tax e Rex, o gene *pX* codifica, através de diferentes ORFs (I e II), e portanto, diferentes formas alternativas de mRNA outras três proteínas acessórias: p12, p13 e p30. Tais proteínas contribuem com a infectividade viral, manutenção de altas cargas provirais, ativação e regulação da transcrição gênica (MICHAEL et al., 2004). A ORF-I é responsável pela síntese da proteína p12 que pode sofrer proteólise na extremidade aminoterminal e gerar a proteína p8, enquanto as proteínas p13 e p30 são resultados do *splicing* alternativo da ORF-II. A proteína p12 localiza-se no sistema celular de endomembranas, particularmente, no RE e complexo de Golgi, enquanto que p8 é dirigida para as balsas lipídicas da membrana celular e é recrutada para a sinapse imunológica através da ligação com o receptor de célula T-TCR (Do inglês, *T-cell Receptor*) (KORALNIK et al., 1993; DING et al., 2001;

FUKUMOTO et al., 2009). Através da depleção dos estoques de cálcio do RE e do aumento de cálcio citoplasmático, p12 modula processos de proliferação de células T e replicação viral, além de ativação de fatores nucleares de ativação de células T (NFAT), que é dependente da ativação por proteínas cálcio-dependente (KIM et al., 2003). É sabido ainda que p12 liga-se às cadeias pesadas recém sintetizadas de MHC-I (Do inglês, *Major Histocompatibility Complex class I*) prevenindo sua associação com a  $\beta$ 2-microglobulina, o que conduz tais cadeias à degradação pelo proteossoma, diminuindo sua expressão (JOHNSON et al., 2001; EDWARDS et al., 2011).

A proteína p30 localiza-se no núcleo e nucléolo sugerindo que sua função relacione-se com a mediação de processos como progressão do ciclo celular, reparo do DNA e exportação de moléculas de mRNA (BOISVERT et al., 2007; EDWARDS et al., 2011). Além disso, p30 se liga especificamente às moléculas de mRNA, duplamente processado (Tax/Rex) retendo-os no núcleo, favorecendo a propagação silenciosa dos genomas provirais do HTLV-1 por divisão celular através da expansão clonal de células infectadas (NICOT et al., 2004; EDWARDS et al., 2011).

Mais recentemente, em relação aos demais produtos gênicos, estudos despertaram um grande interesse por um transcrito produzido a partir do promotor 3' LTR na fita negativa do provírus, este transcrito codifica a proteína conhecida como Fator Básico de Zíper de Leucina do HTLV-1 (HBZ, do Inglês HTLV-1 *Basic Leucine Zipper Factor*). De modo que dois transcritos são gerados na codificação do HBZ: o *spliced* (sHBZ) e o *unspliced* (usHBZ) (**Figura 2**) (GREEN et al., 2001; MATSUOKA et al., 2009; RUSHING et al., 2018).

Compreende-se que seja função do HBZ manter a infecção viral crônica e promover a leucemogênese, em um mecanismo ainda não muito bem esclarecido, mas que envolve interações com seu domínio KIX (Domínios de ativação ordenados por fatores de transcrição) (YANG et al., 2018). O domínio KIX é alvo de fatores de transcrição viral que competem com fatores de transcrição celular, resultando na perturbação de processos celulares importantes. Sabe-se que após a infecção da célula pelo HTLV-1, o vírus codifica a proteína Tax que ativa a transcrição de genes virais nos estágios iniciais da leucemogênese por meio do recrutamento de uma proteína nuclear conhecida como um co-ativador transcricional-CBP (Do Inglês, *Co-Activator Binding Protein*). Tax, por sua vez, pode atuar na regulação negativa do HBZ o que reduz a imunodeteção dessa proteína viral no hospedeiro (YANG et al., 2018).

Esse mecanismo de regulação acontece uma vez que, especificamente, o domínio de ativação do HBZ (HBZAD- Resíduos 1-77) compete com a proteína Tax pelo domínio alvo de fatores de transcrição (KIX) (YANG et al., 2018).

HBZ pode agir através de duas formas funcionais. Em sua forma de RNA, promovendo a proliferação celular através da regulação positiva de uma classe de fatores de transcrição (E2F1) importantes para a regulação do ciclo celular (BASBOUS et al., 2003; THÉBAULT et al., 2004, SATOU & MATSUOKA, 2010). Em sua forma proteica, HBZ suprime a transcrição viral mediada por Tax, através da competição direta pelo recrutamento da proteína de ligação ao elemento responsivo ao monofosfato cíclico de adenosina - cAMP (CREB, do inglês, *Cyclic Adenosine Monophosphate-cAMP- response element-binding protein*) ou da proteína de ligação ao CREB/homólogo p300 (CBP/p300, do inglês, *CREB-binding protein/homologue p300*) (BASBOUS et al., 2003; THÉBAULT et al., 2004; SATOU & MATSUOKA, 2011), e inibe seletivamente a via clássica do fator de transcrição nuclear kappa B do hospedeiro, evento crítico para a transformação, proliferação e sobrevivência das células infectadas pelo HTLV-1 (MATSUOKA & JEANG, 2011).

HBZ é também uma proteína imunogênica reconhecida por clones de linfócitos T citotóxicos (CTL, do inglês, *Cytotoxic T lymphocytes*), sendo considerado um imunógeno mais fraco do que Tax quando reconhecido pelos CTLs. A menor imunogenicidade do HBZ pode permitir que as células infectadas pelo HTLV-1 escapem da resposta imune do hospedeiro (ROWAN et al., 2014).

## **I.2 A Infecção**

As vias de transmissão do HTLV-1 se encontram em três eixos: sexual, parenteral (transmissões horizontais) e entre mãe e filho (transmissão vertical) (MANNNS et al., 1992; PROIETTI et al., 2005; ISHAK et al., 2020). A transmissão do HTLV-1 requer a transferência de células infectadas com o provírus, presentes no sêmen, no leite humano ou no sangue (ISHAK et al., 2020). De modo que, para facilitar a sua transmissão, o HTLV-1 aumenta a população clonal de células infectadas pelas ações pleiotrópicas das proteínas virais, especialmente a proteína Tax (TAYLOR et al., 2005).

A transmissão vertical do HTLV-1 pode ocorrer durante o período intrauterino ou durante o parto, porém, foi demonstrado que a amamentação é a maior via de

transmissão, à medida que células infectadas pelo vírus entram no corpo da criança por via oral (HINO et al., 1985). Portanto, a política de evitar a amamentação em mães que carregam o vírus é geralmente seguida em áreas urbanas e é um procedimento eficiente pelo qual a transmissão é reduzida. A transmissão vertical é comum e provavelmente é a rota mais importante para a manutenção do vírus em comunidades fechadas epidemiologicamente, como visto com o HTLV-2c entre comunidades indígenas da Amazônia do Brasil (PERCHER et al., 2016).

O risco associado à transfusão de sangue (Transmissão horizontal) diminuiu acentuadamente com a introdução de rígidas políticas regulatórias relacionadas à triagem de sangue no Brasil e em outros lugares (CARNEIRO-PROIETTI et al., 2012; MURPHY, 2016). Em doadores de sangue no Brasil a triagem para HTLV acontece de forma obrigatória nos bancos de sangue, porém nenhum teste é feito para diferenciar e caracterizar os tipos e subtipos desses vírus (RIBEIRO et al., 2018). Muitos países adotaram essa triagem do HTLV-1 no sangue doado, o que reduziu efetivamente o risco de transmissão do HTLV-1 por transfusão. Porém, essa triagem ainda não é realizada em todo o mundo (TAGAYA et al., 2019).

E a transmissão sexual por sua vez, certamente é a via de transmissão mais importante para a transmissão de ambos os tipos virais: o HTLV-1 e o HTLV-2. Sendo descrita como um mecanismo eficiente para a disseminação do HTLV-2c entre os grupos indígenas nativos por exemplo. De acordo com o ambiente geográfico, e fatores de risco comportamentais, o aumento do risco de transmissão do vírus aumenta a prevalência e incidência de infecção e doença (ISHAK et al., 2020). A transmissão sexual é igualmente importante para a epidemiologia e acontece a partir do sexo sem proteção, pela transferência de linfócitos infectados presentes no sêmen e secreção vaginal (TAGAYA et al., 2019).

### **I.2.1 Epidemiologia**

A infecção pelo HTLV-1 tem distribuição mundial (KANZAKI, 2018), no entanto, algumas regiões são consideradas áreas endêmicas: sudoeste do Japão, países no Caribe, África sub-Saara e áreas localizadas no Irã e Melanésia (PROIETTI et al., 2005; VERDONCK et al., 2007; GESSAIN et al., 1996; ALESSIO et al., 2018; YANG et al., 2018; MARTINEZ et al., 2019) (**Figura 3**). Assim, pelo menos 5-10 milhões de pessoas estão infectadas HTLV-1 em todo o mundo (BARRETO et al., 2016; RIBEIRO et al., 2018; MARTINEZ et al., 2019).

A soroprevalência do HTLV-1 em áreas de endemicidade é estimada em 1 a 2%, levando em consideração a variável idade, onde dados revelaram resultado positivo para cerca de 20 a 40% dos indivíduos com mais de 50 anos de idade. Revelando então o fator idade como um dos principais determinantes epidemiológicos, independente do ponto de vista socioeconômico e cultural, para a soroprevalência do HTLV-1.

Além desse, outros determinantes importantes em regiões endêmicas podem ser citados como é o caso do gênero, especificamente mulheres, e do status econômico (GESSAIN & CASSAR, 2012). Nas áreas urbanas, a infecção pelo HTLV também é mais comum entre as mulheres (HANANIYA et al., 2019; ISHAK et al., 2020).

O HTLV-1 também é encontrado em países da América do Sul, como no Brasil onde estima-se que 2,5 milhões de indivíduos estejam infectados por este vírus (CARNEIRO-PROIETTI et al., 2002; CATALAN-SOARES et al., 2004) (**Figura 3**).

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



Os círculos retratam a prevalência de indivíduos infectados pelo HTLV-1. O tamanho do círculo é diretamente proporcional a taxa de prevalência da região. Fonte: GESSAIN & CASSAR, 2012.

No Brasil, as maiores prevalências são observadas no Norte e Nordeste (DOURADO et al., 2003; CATALAN-SOARES et al., 2005). No entanto, também se pode observar o fenômeno de distribuição heterogênea do vírus, sendo as maiores prevalências registradas nos estados do Maranhão (São Luis - 10.0/1000 doadores

de sangue), Bahia (Salvador - 1,8% na população geral, ou 9.4/1000 doadores de sangue), Pará (Belém - 9.1/1000 doadores de sangue) e Pernambuco (Recife - 7.5/1000 doadores de sangue) (DOURADO et al., 2003; CATALAN-SOARES et al., 2005).

Salvador é considerada a cidade brasileira com a maior prevalência (1,8% na população geral) global do HTLV-1, e com prevalência de aproximadamente 0,84% entre mulheres grávidas; 1,3% entre doadores de sangue (GALVAO-CASTRO et al., 1997; DOURADO et al., 2003; BITTEN///COURT et al., 2007). O contato sexual também foi associado com a principal forma de transmissão do HTLV-1 na população geral de Salvador, podendo estar relacionado com a elevada prevalência de sífilis (1,3%: duas vezes maior que a prevalência do HIV-1-0,6%), por já ter sido documentada forte co-infecção desses patógenos e por compartilharem desta mesma via de transmissão entre hospedeiros (NUNES et al., 2017).

Ribeiro e cols. (2018), em estudo no Piauí, com doadores de sangue, revelaram prevalência do HTLV-1 de 0,59 por 1.000 (IC95%: 0,38-0,87), estando de acordo com os achados na literatura para taxas demonstradas em outros doadores de sangue no Brasil (0,2 para 1,1 por 1.000). Neste mesmo estudo, percebe-se que alguns comportamentos de risco, como é o caso do uso de drogas injetáveis e relações sexuais desprotegidas, são decisivos para a transmissão viral, bem como, estão associados com a alta frequência deles em doadores de sangue (RIBEIRO et al., 2018).

Indivíduos infectados pelo HTLV-1 comumente possuem co-infecções com outros patógenos, como é o caso de infecções pelo *Mycobacterium tuberculosis*. Essa co-infecção pode influenciar a transmissão da tuberculose (TB), a epidemiologia da doença e até o perfil clínico dos pacientes, uma vez que foi demonstrada relação entre a infecção pelo HTLV-1 e aumento da taxa de mortalidade em pacientes com TB. Inclusive, altas taxas de prevalência do HTLV-1 entre pacientes com TB, mais especificamente na cidade de Salvador, Bahia já foram relatadas na literatura (KOZLOWSKI et al., 2014).

Em populações indígenas, a distribuição de anticorpos contra o HTLV mostra que o aumento da prevalência, nessa comunidade, se dá juntamente com o aumento da idade e de forma contrária ao que já foi documentado, não apresenta diferença na prevalência entre homens e mulheres. Isso permite perceber que para populações fechadas o vírus infecta ambos os sexos igualmente, mesmo nos casos de

transmissão vertical, onde se apresenta como a rota mais importante para a manutenção do vírus em comunidades epidemiologicamente fechadas. Contudo esse cenário não é comumente observado em áreas urbanas (ISHAK et al., 2020).

No norte do país foi detectada baixa soroprevalência geral do HTLV1/2 na população de doadores de sangue, sendo de 0,13%, em Manaus, no Amazonas, por exemplo. Outros estados do Norte do Brasil revelam taxas maiores que as encontradas neste estado, a exemplo do Amapá (0,71%) e Pará (0,91%) (MORAIS et al., 2017). A região do Amazonas é endêmica para o HTLV-2, uma vez que esse vírus é comum em indígenas e esta é uma área predominante rica em grupos indígenas brasileiros (MORAIS et al., 2017). Apesar do número estimado de 800.000 indivíduos infectados em todo o mundo, o HTLV-2 é muito menos prevalente que o HTLV-1.

### **I.2.2 Ciclo de infecção e Persistência**

O entendimento de como se processa o reconhecimento, e consequentemente, a replicação viral passa pelo conhecimento de quais são os tipos celulares infectados pelo HTLV-1. Neste aspecto, embora o tropismo seja por células T periféricas, predominantemente linfócitos TCD4<sup>+</sup> de memória (Linfócitos T auxiliares), como descrito inicialmente, em 1990, por Richardson e colaboradores, o HTLV-1 também tem sido encontrado, *in vivo*, em outros tipos celulares, a exemplo de células T CD8<sup>+</sup> (HANON et al., 2000), monócitos e células B (KOYANAGI et al., 1993), macrófagos (NATH et al., 2003), células dendríticas (KNIGHT et al., 1993) e endoteliais (SETOYAMA et al., 1998).

O HTLV -1 apresenta um ciclo de replicação onde a primeira etapa é dependente dos componentes virais, e a segunda do metabolismo celular.

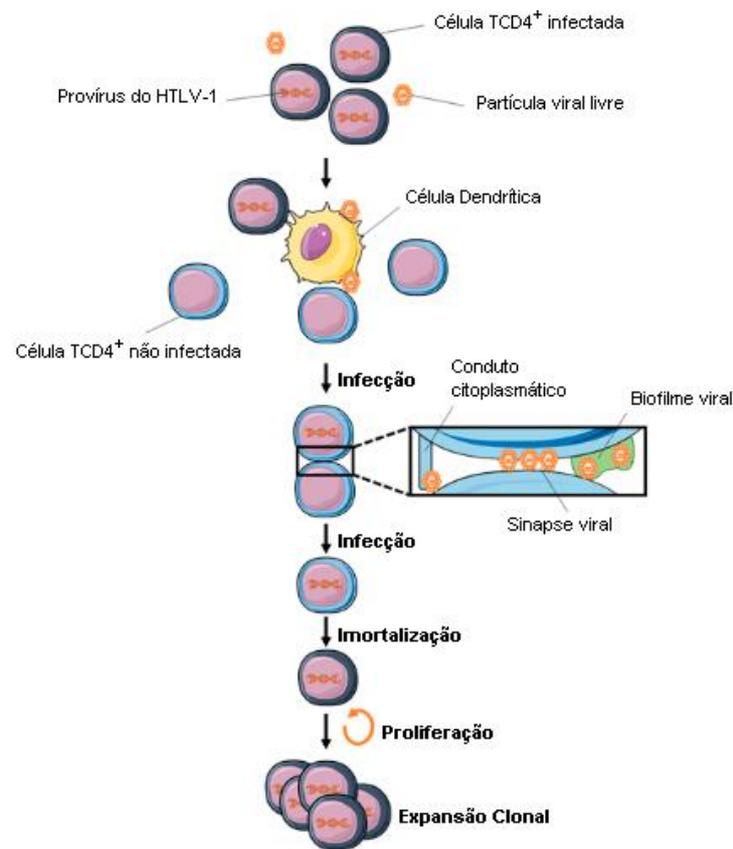
A primeira etapa do ciclo de replicação é o reconhecimento e consequente interação do vírus com a célula hospedeira, interação que ocorre entre o domínio de ligação do receptor, na glicoproteína de superfície do vírus, e o complexo receptor na superfície da célula hospedeira. Atualmente, sabe-se da participação e interação de pelo menos três moléculas distintas na entrada do vírus na célula: GLUT-1, que é um transportador de glicose, na superfície da membrana celular (BATTINI et al., 2003; MANEL et al., 2004; COSKUN & SUTTON, 2005); neurofilina-1 (NRP-1), que é uma glicoproteína de membrana (GHEZ et al., 2006); e heparan sulfato, que são proteoglicanos (HSPG) presentes na superfície da célula (PINON et al., 2003).

A mais recente concepção de reconhecimento e ligação do HTLV-1 às células hospedeiras envolve os seguintes passos: ligação da gp46 às moléculas de HSPG na superfície da célula, aumentando a probabilidade de ligação da gp46 à molécula de NRP-1; em seguida, há a formação do complexo gp46/HSPG/NRP-1 que possibilita a exposição dos domínios de ligação em GLUT-1; finalmente gp46 liga-se à GLUT-1 disparando o processo de fusão das membranas celular e viral, permitindo a entrada do vírus na célula, fusão que envolve também modificações conformacionais na glicoproteína transmembrana (gp21) (JONES et al., 2011). Após a fusão das membranas e consequente internalização do material genético, ocorrerá a transcrição do genoma viral de RNA para DNA, pela enzima TR que ocorrerá dentro do cerne viral usando como oligonucleotídeo iniciador o tRNA<sup>Pro</sup>, que será, tardiamente, degradado pela RNaseH. Subsequentemente, o DNA viral entra no núcleo, e é inserido no genoma da célula hospedeira formando o provírus, função exercida pela integrase viral. O processo de integração do provírus marca o final da fase precoce do ciclo de multiplicação do vírus e inicia a fase tardia que é mediada por enzimas do hospedeiro (SEIKI et al., 1984).

Nesta última etapa, ocorre a síntese do RNA viral tendo como DNA molde o provírus integrado. A síntese do RNA viral leva à formação de um longo transcrito primário, que é processado para formar as moléculas de mRNA e RNA genômico, de forma que as proteínas virais são, então, traduzidas nos ribossomos celulares e em seguida seguem para maturação (PROIETTI, 2006).

O HTLV-1 é incomum entre os retrovírus em seu modo de disseminação. Em modelos clássicos de replicação dos retrovírus, as proteínas recém-sintetizadas e montadas em novas partículas virais infectivas, por transcrição reversa, brotam da superfície celular para reiniciar o ciclo em outra célula hospedeira. A disseminação do HTLV-1, entretanto, baseia-se na mobilidade e contato de células infectadas com células não infectadas para uma transmissão eficiente (COOK et al., 2017). A expansão clonal, replicação passiva do provírus integrado dentro do genoma através de divisões mitóticas, é a principal forma de propagação do HTLV-1 durante o estágio crônico da infecção, contribuindo para o aumento da carga proviral (BANGHAM et al., 2015) (**Figura 4**).

**Figura 4** - Curso da infecção pelo HTLV-1 e suas formas de disseminação.



Fonte: Adaptado de FUTSCH; MAHIEUX; DUTARTRE, 2017.

Além disso, o vírus induz eventos de polarização das células facilitando a passagem viral, num fenômeno conhecido como sinapse viral (BANGHAM, 2003). Quando uma célula infectada entra em contato com outra célula, porém, não infectada, há a formação de um centro de organização microtubular (MTOC- *Microtubule Organizing Center*) que é polarizado na junção célula-célula, formando a interface onde ocorrerá a sinapse virológica (**Figura 4**). A formação desta estrutura permite o acúmulo de proteínas do gene gag e de material genético (RNA), na interface da sinapse, culminado com a passagem desse material para a célula não infectada (MATSUOKA & JEANG, 2007; MAJOROVITS et al., 2008).

Manel e cols. (2005) demonstraram, inclusive, que nestas áreas de “sinapse viral” há um acúmulo de moléculas de GLUT 1, conhecidas por seu papel no reconhecimento celular. É possível identificar também o papel das moléculas de HSPG, na retenção do material genômico do vírus próximo à matriz extracelular de uma célula infectada, facilitando a sinapse virológica.

### I.2.3 Resposta imune

A infecção pelo HTLV-1 provoca uma alteração no ambiente intracelular levando à ativação espontânea das células infectadas. A proteína regulatória Tax é a principal responsável por essa ativação, através da interação com o fator de transcrição NFκB do hospedeiro, evento crítico para a transformação, proliferação e sobrevivência das células infectadas pelo HTLV-1. A proteína Tax estimula a ativação tanto da via canônica quanto da não canônica de NFκB, envolvidas na regulação da inflamação e apoptose, e na regulação da organogênese linfóide, respectivamente, induzindo uma expressão aumentada das proteínas dessas vias, como citocinas estimulatórias e receptores, por exemplo IL-2, e mantendo a ativação e proliferação celular (CURREN et al., 2012).

Linfócitos T citotóxicos CD8<sup>+</sup> são ativados cronicamente e são capazes de reconhecer a proteína viral Tax; sendo este o principal mecanismo de defesa do hospedeiro frente a infecção pelo HTLV-1 (SOUZA MACHADO et al., 2003).

O perfil de citocinas em resposta à infecção pelo HTLV-1 mostra padrão de resposta mediada por células T, com perfil Th1, com aumento da produção de TNF, IFN-γ, IL-6 e redução nos níveis de IL-4. IFN-γ é essencial para a função citotóxica adequada das células T e modulação negativa da resposta Th2, enquanto IL-4 e IL-10 suprimem a resposta Th1 e regulam negativamente a ação do IFN-γ sobre a resposta mediada por células T, com perfil Th2 (SOUZA MACHADO et al., 2003). A alta produção de IL-10 em pacientes assintomáticos poderia explicar o estado clínico desses indivíduos (SOUZA et al., 2012).

A progressão da infecção viral é associada ao aumento da expressão do gene HBZ. De forma curiosa, o HBZ regula negativamente a expressão de Tax na célula, levando a uma menor eficácia de CTL contra o vírus. Entretanto, mesmo em um ambiente de baixa expressão de Tax, as células CTLs ainda são capazes de produzir altos níveis de citocinas pró-inflamatórias (SOUZA et al., 2012).

O processo inflamatório, característico da HAM/TSP, depende da ativação da célula T que, por sua vez, depende da ligação de seu receptor com o complexo MHC (Do inglês, *Major Histocompatibility Complex*) e suas moléculas apresentando os epítopos virais na superfície de células chamadas apresentadoras de antígenos. Além desse reconhecimento entre tais referidas unidades celulares, há também a necessidade da interação por parte das chamadas moléculas co-estimulatórias que se encontram na superfície tanto das células apresentadoras de antígeno (CD80 e

CD86) quanto da célula T (CD28) (HARRIS et al., 1999).

#### **I.2.4 Manifestações Clínicas Associadas**

A infecção pelo HTLV-1 não necessariamente implica no desencadeamento de processos patogênicos em seus portadores, uma vez que os escassos dados disponíveis sobre o tema, apontam para a ocorrência de manifestações clínicas associadas ao HTLV-1 em 2-4% da população infectada (COOK et al., 2013; KANNAGI et al., 2019), o que nem sempre reflete a realidade observada. Diferentes fatores estão envolvidos na interação vírus/hospedeiro, e o modo como esta interação se desenvolve determinará o estado do portador como indivíduo assintomático ou sintomático.

As complicações clínicas associadas à infecção são organizadas em três grupos: (1) doenças malignas, grupo no qual está incluída a ATLL (Do Inglês - *Adult T-cell leukemia/lymphoma*); (2) síndromes inflamatórias, grupo que abriga a Paraparesia Espástica Tropical/Mielopatia Associada ao HTLV (HAM/TSP), uveíte e síndrome do olho seco; e por fim, (3) complicações infecciosas, como estreptocidose e dermatite infecciosa (MOTA-MIRANDA, 2012).

A primeira manifestação clínica que foi associada à infecção pelo HTLV-1 foi a Leucemia/Linfoma de Células T de adulto (POIESZ et al., 1980), cujas primeiras ocorrências foram relatadas em indivíduos infectados no Japão. Posteriormente, foram descritas em 1985, a Paraparesia Espástica Tropical (*Tropical Spastic Paraparesis-TSP*), e a Mielopatia Associada ao HTLV (*HTLV Associated Myelopathy-HAM*), em 1986, como sendo manifestações associadas ao HTLV-1 (GESSAIN et al., 1985; OSAME et al., 1986). Porém anos mais tarde, concluiu-se que se tratava da mesma etiologia e desta forma, atualmente, o HTLV-1 é conhecido como o agente etiológico de uma síndrome neurológica denominada HAM/TSP. Outras manifestações como as artropatias (NISHIOKA, 1996; FRENZEL et al., 2014), polimiosites (MORGAN et al., 1989), uveítes (KAMOI & MOCHIZUKI, 2012), e dermatites infecciosas (LA GRENADE, 1996; HARROD, 2019) também podem estar relacionadas à infecção por este vírus.

O efeito patogênico associado à infecção pode estar relacionado à fatores virais, fatores do hospedeiro, fatores ambientais e, sobretudo, à interação dos três.

A proteína viral Tax, por exemplo, é um potente ativador das vias de transcrição e demonstra ser suficiente para alterar o ciclo celular e imortalizar as células T, *in vitro*,

com conseqüente transformação celular, o que corrobora com o fato de o HTLV-1 ser o agente etiológico da ATLL. Isso porque, a proteína Tax contém domínios que permitem interagir com fatores celulares que afetam grande número de funções celulares e produtos de reações (CURRER et al., 2012). Ela impulsiona a transformação celular através de sua capacidade de alterar a expressão celular de genes, vias de sinalização e funcionamento do ciclo celular (ANUPAM et al., 2013).

Várias proteínas derivadas do HTLV-1 além de interagir com fatores de transcrição podem manipular o ciclo de vida deste vírus pela regulação positiva ou negativa de moléculas de RNA de interferência (iRNA). O mecanismo de desregulação do ciclo celular por iRNA pode gerar entre outros efeitos, o silenciamento do mRNA e silenciamento de genes de transcrição, tendo como consequência alterações que podem levar a um aumento da sobrevivência celular, o que pode aumentar a capacidade de invasão, proliferação e diferenciação celular (SAMPEY et al., 2012).

Na maioria das vezes que ocorre o desenvolvimento da HAM/TSP é de forma esporádica (sHAM/TSP), onde já foi associado a manifestação clínica mais em indivíduos adultos. No entanto, em alguns casos ela pode ter caráter familiar (fHAM/TSP); ocasionalmente ocorre na infância e predominante associada ao gênero (BENEVIDES, 2019).

### **I.3 Progressão para HAM/TSP**

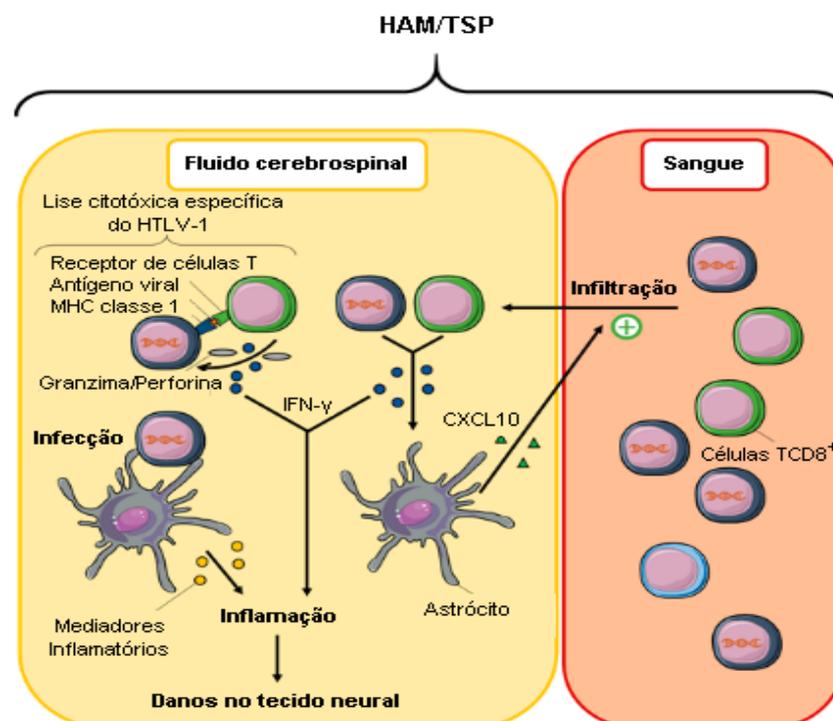
A prevalência da HAM/TSP é de 0,25 a 3,8% em pacientes infectados pelo vírus, sendo mais prevalente em mulheres que em homens, em proporção que varia de 2:1 a 3:1 (MORENO-CARVALHO et al., 1992; EDWARDS et al., 2011; BARRETO et al., 2016). Estudo realizado na Espanha que avaliou as condições clínicas mais frequentes registradas juntamente com o diagnóstico de HTLV-1 em um banco de dados hospitalares público nacional, revelou a HAM/TSP (61 [53%] de 115 pacientes) como mais frequente entre os indivíduos infectados, seguida pela ATLL (29 [25%] de 115 pacientes) (RAMOS, 2020).

A HAM/TSP caracteriza-se por uma mielopatia lentamente progressiva, tendo como consequência a paraparesia dos membros inferiores, quadro clínico que acomete indivíduos, predominantemente, na quarta e quinta décadas de vida (ROMAN et al, 1988; OSAME, 2002). Na HAM/TSP, a patogênese tem relação com a

desmielinização local atribuída à presença da célula infectada, e o desencadeamento de uma resposta inflamatória crônica.

Células infectadas pelo HTLV-1 e células citotóxicas específicas contra o vírus, infiltradas no líquido cefalorraquidiano, estimulam a liberação de citocinas pró-inflamatórias, IFN- $\gamma$ , que levará a danos no tecido neural. Este dano decorre tanto diretamente da exacerbação das respostas inflamatórias quanto indiretamente pelo recrutamento de mais IFN- $\gamma$  produzido pelas células no LCR (Líquido cefalorraquidiano). A expressão de quimiocina CXCL10 por astrócitos pode aumentar a infiltração de células TCD4<sup>+</sup> infectadas pelo HTLV-1 e CTL específicos para o HTLV-1 no LCR, desencadeando um ciclo de *feedback* positivo inflamatório (FUTSCH; MAHIEUX; DUTARTRE, 2017) (**Figura 5**). A atividade de Tax é importante para a capacitação das células infectadas em transpor a barreira hemato-encefálica, ao mesmo tempo em que é o principal alvo da resposta imune celular (CRAVOIS et al., 2000).

**Figura 5** – Esquema ilustrativo do principal mecanismo associado ao desenvolvimento da HAM/TSP.



IFN- $\gamma$  (Citocina pró-inflamatória), CXCL10 (Quimiocina).

Fonte: Adaptado de FUTSCH; MAHIEUX; DUTARTRE, 2017.

### I.3.1 Marcadores Moleculares Virais e do Hospedeiro

Atualmente, sabe-se que a depender de fatores do hospedeiro a infecção pelo HTLV-1 pode ter diferentes desfechos. Alguns fatores já foram associados com desenvolvimento de HAM/TSP, incluindo os antígenos leucocitários humanos (HLA). Moléculas de 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>, determinando a especificidade e eficácia de respostas T CD8<sup>+</sup> ao vírus (DENDROU et al., 2018).

No Japão, foi sugerida uma associação entre a presença do HLA-A\*02 ou HLA-Cw\*08, com uma menor carga proviral e uma menor prevalência de HAM/TSP (JEFFERY et al., 1999; JEFFERY et al., 2000), o mesmo efeito protetor também foi observado do alelo HLA-A\*02 em populações brasileiras (CATALAN-SOARES et al., 2009). Esses alelos provavelmente exercem sua função protetiva devido a uma ligação mais forte aos peptídeos do HTLV-1, principalmente os peptídeos originários do HBZ. Por outro lado, os alelos HLA-DRB1\*0101, HLA-B\*5401 e HLA-B\*07 apresentam associação com um aumento na susceptibilidade para HAM/TSP em indivíduos no Japão (JEFFERY et al., 1999; JEFFERY et al., 2000). Além desses, os receptores de quimiocinas CXCL10, CXCL9 e a Neopterin, presente no fluido cerebrospinal, já foram associados, na literatura, com o desenvolvimento de HAM/TSP, podendo ser fatores utilizados para classificação da severidade da doença a depender de suas presenças e concentrações (YAMAUCHI et al., 2019) assim como a subexpressão de FOXP3 (Proteína associada à função das células T) induzida pelo HBZ em células T CD4<sup>+</sup> avaliadas em camundongos (YAMAMOTO-TAGUCHI et al., 2013).

Outro tipo de fator, que pode influenciar na manifestação clínica, são os polimorfismos de nucleotídeo único (SNP, do inglês: *Single Nucleotide Polymorphism*). A maioria dos SNPs associados à HAM/TSP é encontrada em genes que codificam citocinas celulares inflamatórias. Um polimorfismo no promotor (-634 GC) do gene da citocinas IL-6 foi encontrado em um número significativo em pacientes com HAM/TSP em um estudo no Brasil em comparação com indivíduos assintomáticos (GADELHA et al., 2008). No Japão, outro estudo associou a presença de um polimorfismo no promotor (-592 A) do gene da IL-10 com uma redução de até duas vezes as chances de desenvolvimento de HAM/TSP (SABOURI et al., 2004). Polimorfismos (rs8099917 GG e rs12979618 CT) no gene IL28B (Interferon lambda

3), conhecido por ser um importante adjuvante na patogênese de algumas infecções virais, foram associados com a HAM/TSP, pelo fato de seus portadores apresentarem uma distinta resposta imune ao HTLV-1 (ASSONE et al., 2014). No Peru, pacientes com HAM/TSP apresentaram um polimorfismo (FAS-670 AA) no gene FAS, um receptor celular expresso em células infectadas por vírus, levando à apoptose, sendo portanto esse achado associado à uma maior carga proviral (ROSADO et al., 2017).

Estudo recente (NAITO et al., 2018) sobre assinaturas de expressão gênica distintas induzidas por transativadores virais de diferentes subgrupos do HTLV-1 que conferem um risco diferente para HAM/TSP, demonstrou que diferentes subgrupos do HTLV-1 são caracterizados por diferentes padrões de expressão dos genes do hospedeiro. Para investigar o papel dos subgrupos do HTLV-1 na patogênese viral, por exemplo, foi estudada a diferença funcional nos reguladores transcricionais virais específicos para subgrupos das proteínas Tax e HBZ usando análise de microarranjos, ensaios de genes repórteres e avaliação da interação proteína-proteína do hospedeiro viral. A expressão diferencial de genes relacionados à patogênese regulados direta ou indiretamente por Tax-HBZ ou subgrupo específico delas pode estar associada ao aparecimento da HAM/TSP. Eles propõem a quimiocina CXCL10 como um biomarcador de prognóstico para HAM/TSP, uma vez que esta foi mais eficientemente induzida pelo subgrupo A da proteína Tax (Tax-A) quando comparada ao subgrupo B (Tax-B), *in vitro*, bem como em PBMCs não manipulados (*ex vivo*) obtidos de pacientes com HAM/TSP.

## II. CAPÍTULO I

**Apresentação diferencial de epítomos  
de proteínas estruturais e regulatórias do HTLV-1  
em indivíduos com diferentes perfis clínicos.**

## II.1 Justificativa

Embora muitos aspectos dos eventos que levam ao desenvolvimento das doenças associadas ao HTLV-1, e portanto da HAM/TSP também, não estejam esclarecidos, a resposta imune do hospedeiro frente à infecção viral, principalmente a resposta celular desencadeada por células T CD8<sup>+</sup> específicas anti-HTLV, e os eventos efetores desta resposta são reconhecidos como cruciais, determinando o desfecho da infecção (BANGHAM & OSAME, 2005). Esta resposta celular parece ser influenciada pela via de infecção do hospedeiro, carga proviral, fatores genéticos individuais, como polimorfismos em genes do sistema Antígeno Leucocitário Humano-HLA e em outros genes também envolvidos na resposta imune (VINE et al., 2002). A correlação inversa entre o crescimento tumoral e a imunidade celular, por exemplo, indicou que a inibição da resposta CTL HTLV-1 específica, através do bloqueio com anticorpos anti CD80/CD86, tem papel fundamental contra o desenvolvimento da ATLL, e que as vias de sinalização co-estimulatórias participam ativamente deste papel (HANABUCHI et al., 2000). Sabe-se que respostas de células T específicas, se apresentam de forma diferentemente acentuada em indivíduos com HAM/TSP, uma vez que a resposta CTL direcionada contra a proteína Tax é elevada em pacientes com esta manifestação clínica, enquanto que é prejudicada naqueles indivíduos com ATLL. No entanto, a razão para as diferentes respostas dos linfócitos T citotóxicos nas duas doenças não é bem conhecida (KANNAGI, 2019).

Esse entendimento da importância das células apresentadoras de antígeno e da ativação celular posterior, sobretudo via CTL, ajuda também a explicar que outras proteínas acessórias virais incluindo p12, p8, p30 e p13 potencialmente podem contribuir para a persistência viral, uma vez que degradam moléculas de MHC-I (Do inglês, *Major Histocompatibility Complex class I*) (EDWARDS et al., 2011). p12, particularmente, liga-se às cadeias pesadas recém sintetizadas de MHC-I prevenindo sua associação com a  $\beta$ 2-microglobulina, o que conduz tais cadeias à degradação pelo proteassoma, diminuindo sua expressão (JOHNSON et al., 2001; EDWARDS et al., 2011). A consequente diminuição da expressão de moléculas de MHC-I previne a apresentação de peptídeos imunogênicos, e, portanto, poderia relacionar a atuação de p12 à resposta imune efetiva, portanto, persistência viral. Duas formas de p12 têm sido identificadas, sendo uma com um resíduo de lisina na posição 88, e esta é comumente encontrada em indivíduos HAM/TSP, enquanto que a segunda variante

da proteína abriga um resíduo de arginina nesta mesma posição e é encontrada em isolados de indivíduos ATLL e carreadores assintomáticos.

Adicionalmente, é sabido que a diversidade e a frequência do reservatório de linfócitos T CD8<sup>+</sup>, específicos para Env, podem estar relacionados com a resposta imune exacerbada, que é verificada em indivíduos HAM/TSP (KOSAKO et al., 2011). Por isso substituições nucleotídicas, na região gênica que codifica para as glicoproteínas do envelope (gp21 e gp46), podem influenciar na infectividade viral, no tropismo celular, na taxa de replicação e na latência ou resposta aos mecanismos efetores, como produção de anticorpos (SZUREK et al., 1988; PAQUETTE et al., 1989).

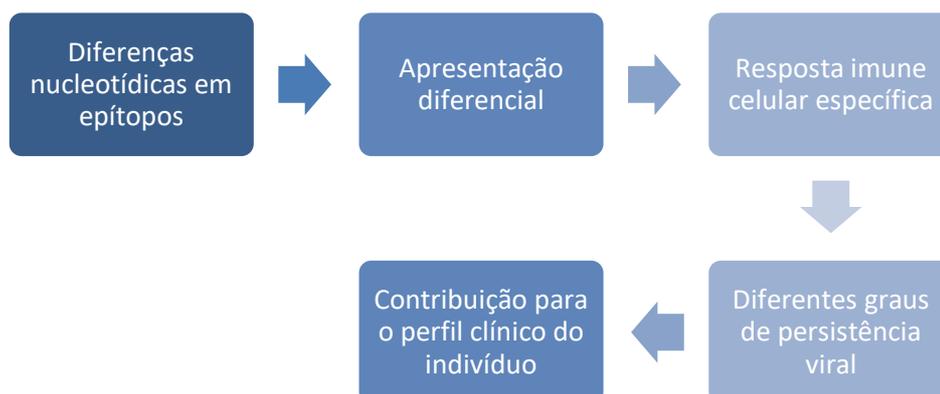
Por fim, ensaios quantitativos revelaram um aumento, em cerca de 2 vezes, no log da carga proviral em modelos animais infectados com variantes do HBZ viral, ressaltando a importância da atuação dessa proteína em eventos iniciais da infecção viral e mesmo depois da persistência do vírus (BARTOE et al., 2000; SILVERMAN et al., 2004). A expressão continuada de HBZ poderia conduzir linfócitos T citotóxicos a mediar a resposta imune contra células T infectadas com HTLV-1 no sistema nervoso central (SNC), intensificando assim a síntese de citocinas Th1 como IFN  $\gamma$  e TNF- $\alpha$ , resultando na morte destas células, sendo este um dos mecanismos de patogênese da HAM/TSP (ARAUJO & SILVA, 2006; LEPOUTRE et al., 2009; MOZHGANI et al., 2017). A regulação da transcrição viral por HBZ, e a expressão do seu mRNA, podem ser um dos mecanismos moleculares envolvidos na latência viral, que permite a prevenção na detecção do sistema imune do indivíduo infectado.

A identificação de mecanismos específicos da resposta ao HTLV-1 pode esclarecer a disfunção imune durante o desenvolvimento de HAM/TSP, especialmente a resposta celular citotóxica, uma vez que ela está diretamente relacionada com o dano tecidual associado à manifestação desta condição clínica.

## II.2 Hipótese

A hipótese desse estudo é que polimorfismos (mesmo em número reduzido) localizados em sítios de apresentação (epítomos) de proteínas regulatórias e estruturais do HTLV-1, possam estar associados às alterações estruturais e funcionais relacionadas a diferentes perfis de apresentação ao sistema imune. Essa apresentação diferencial dos epítomos protéicos aos alelos de HLA do hospedeiro, correlaciona-se com uma resposta imune específica, nestes indivíduos, sendo então este cenário importante na definição do perfil clínico do hospedeiro (**Figura 6**).

**Figura 6** - Esquema ilustrativo da Hipótese para a Apresentação diferencial de epítomos de proteínas estruturais e regulatórias do HTLV-1 em indivíduos com diferentes perfis clínicos.



## **II. 3 Objetivos**

### **II.3.1 Objetivo Geral**

Identificar um perfil diferencial de apresentação de epítomos em proteínas estruturais e regulatórias do HTLV-1 isolados de indivíduos com diferentes perfis clínicos.

### **II.3.2 Objetivos Específicos**

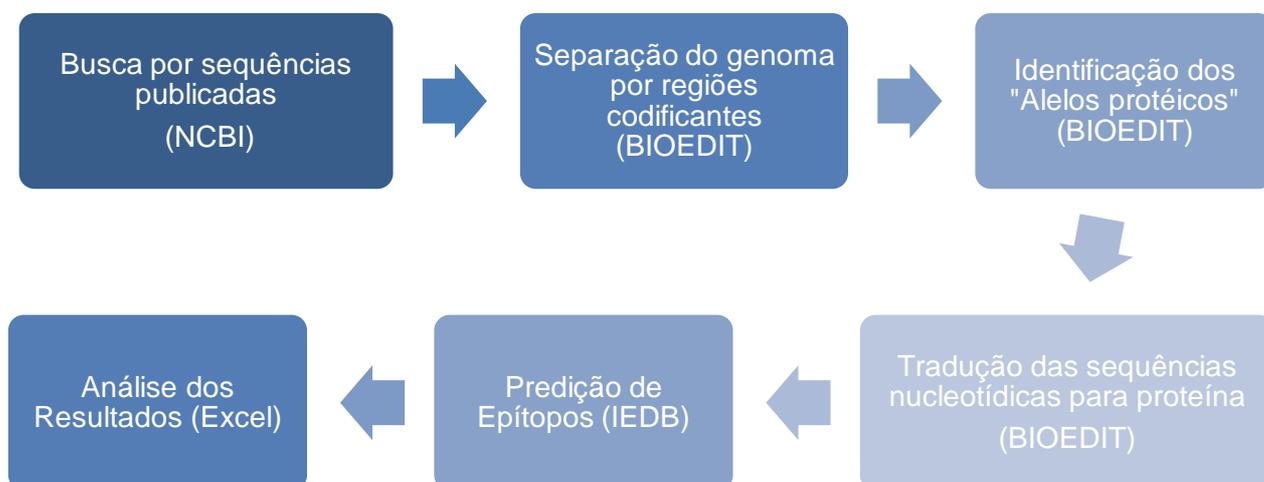
- Realizar a predição de epítomos lineares em proteínas estruturais (gp46 e gp21) e regulatórias (Tax, HBZ e p12) do HTLV-1 isoladas de indivíduos com diferentes perfis clínicos;
- Identificar os principais alelos de HLA utilizados para a apresentação de epítomos lineares, pelas proteínas estudadas, entre os indivíduos com diferentes perfis clínicos;
- Estabelecer as principais diferenças na apresentação de epítomos lineares entre as proteínas regulatórias e as proteínas estruturais do HTLV-1;
- Caracterizar os epítomos mais frequentes e de maior especificidade entre as proteínas estudadas, diferenciando entre os perfis clínicos avaliados;
- Explicar quais são as diferenças em relação à afinidade e especificidade dos epítomos ligantes dos alelos de HLA;
- Sugerir um perfil de apresentação de epítomos lineares das proteínas virais que possa estar relacionado com o perfil clínico do hospedeiro.

## II.4 Metodologia

### II. 4.1 Seleção e Organização do dataset

Este estudo foi realizado a partir de sequências de genomas completos do HTLV-1 que já estavam publicados e, portanto, disponíveis em bancos de dados públicos (**Figura 7**). A busca por tais genomas foi realizada no período de 27 de agosto a 10 de setembro de 2018, no utilizados os seguintes termos de busca: “*Complete genome AND HTLV-1*”. Durante o processo de triagem de informações para a construção do conjunto de sequências a ser utilizado, retiramos do estudo sequências que contemplavam um ou mais dos seguintes critérios de exclusão: sequências que não estavam completas, sequências que foram obtidas de organismos não humanos e sequências que não tinham a informação sobre o perfil clínico do hospedeiro. Por fim, foi realizado o *download* das sequências que se enquadraram nos critérios de elegibilidade.

**Figura 7** - Desenho do Estudo da Apresentação diferencial de epítomos em proteínas estruturais e regulatórias de isolados do HTLV-1 de indivíduos com diferentes perfis clínicos.



Inicialmente as sequências selecionadas foram alinhadas usando o programa Bioedit (HALL, 2011). Para tanto, a sequência ATK1 (Número de Acesso - JCP02029) foi utilizada como genoma referência apenas para a localização nucleotídica das regiões codificantes das proteínas gp21, gp46, p12, Tax e HBZ nos genomas em análise. Selecionamos duas proteínas estruturais (gp21 e gp46) por serem as inicialmente expostas ao sistema imune do hospedeiro, além das principais proteínas regulatórias imunogênicas (Tax e HBZ), associadas às mais importantes estratégias de patogênese associada ao vírus, e por fim, p12 que também é uma proteína regulatória com funções na manutenção do vírus.

Após a definição das regiões gênicas de interesse, os representantes protéicos foram gerados, por tradução (Bioedit), a partir das sequências nucleotídicas. Dessa forma, as sequências protéicas parciais com variações moleculares (a qual denominamos de “alelos protéicos”) foram identificadas e nomeadas para cada perfil de variação de aminoácido encontrado (seja ele fruto de 1 polimorfismo ou mais de um associado). Desse modo, tais “alelos protéicos” (N=78) contemplaram todos os padrões de troca de aminoácidos encontrados nas cinco proteínas virais avaliadas neste estudo.

Simultaneamente ao gerenciamento e organização das sequências a serem analisadas, foi realizada uma busca nas bases de dados PubMed/MEDLINE, sem restrições de idioma ou de tempo, utilizando os termos “HLA AND HTLV-1”, a fim de rastrear o que já havia sido documentado na literatura sobre possíveis associações entre alelos de HLA e aspectos da infecção pelo HTLV-1. O objetivo dessa triagem foi fornecer suporte adicional para melhor correlacionar os achados a cerca das mudanças na apresentação dos epítomos aos alelos de HLA e o desfecho clínico da infecção.

#### **II. 4.2 Predição de epítomos ligantes de molécula de HLA classe 1**

A predição de epítomos foi realizada através da ferramenta *online* denominada IEDB 2.21 (*Immune Epitope Database*) (VITA et al., 2018) ([www.iedb.org](http://www.iedb.org)). Esta avaliação foi realizada para o total de 78 “alelos protéicos” previamente identificados entre os perfis clínicos. Todas estas sequências protéicas foram testadas para as 27 moléculas de HLA classe I disponíveis na plataforma (**ANEXO 1**). Portanto, os epítomos ligantes a tais moléculas de HLA apresentam o tamanho variando de 9 a 10

aminoácidos.

Esta ferramenta fornece um catálogo de epítomos de células T e células B, caracterizados experimentalmente, como ligantes a diferentes moléculas de HLA. Para tanto, a ferramenta foi desenhada a partir de dados coletados ao longo de 4 anos de experimentos publicados, na área de mapeamento de peptídeos em agentes infecciosos, sendo portanto, possível identificar epítomos ligantes em humanos, e não humanos (primatas não humanos, roedores, porcos e gatos). Em sua última atualização (2018), a ferramenta cobria cerca de 92% de todos os experimentos publicados em mapeamento de epítomos de agentes infecciosos. O desenho da ferramenta utiliza a linguagem SQL, para comunicar diferentes domínios do banco de dados e alcançar a busca que se pretende.

Desta forma, a ferramenta faz uma busca por toda a sequência da proteína submetida, para encontrar epítomos, já testados, que possam ter afinidade de ligação, com a respectiva molécula de HLA, cuja informação molecular também já se encontra armazenada na *database* da ferramenta. A cada busca realizada pela ferramenta, são fornecidas informações sobre a sequência do epítomo, o alelo de HLA, ao qual o epítomo é específico e o *score* de ligação de cada epítomo ao alelo de HLA estudado. O *score* fornecido, pela ferramenta, baseia-se no IC<sub>50</sub> (Metade da Concentração Inibitória Máxima) que é a medida de eficiência de inibição de uma substância sobre um processo. Sendo que quanto menor o IC<sub>50</sub> (mM) de um peptídeo, maior a afinidade dele com a molécula de HLA.

Estas informações, geradas no IEDB, foram por fim, exportadas para uma planilha Excel e os dados analisados e gerenciados. Após a predição dos epítomos, e para seguir com a análise e interpretação dos mesmos, foi importante estabelecer que epítomos que apresentassem a localização variando um aminoácido (antes ou depois) deveriam ser considerados o mesmo (Exemplo: o epítomo da posição 50aa é o mesmo da posição 51aa).

## II.5 Resultados

### II. 5.1 Seleção e Organização do Dataset

A busca no GenBank, por sequências do genoma completo do HTLV-1, revelou um total de 255 registros. Porém, desse total, ao aplicar os critérios de exclusão, já mencionados na sessão Metodologia, apenas 46 sequências completas do genoma do HTLV-1 foram incluídas no estudo, as demais 209 sequências foram consideradas inelegíveis. Esses 46 genomas completos do HTLV-1 estão distribuídos nos seguintes perfis clínicos do hospedeiro: 22 genomas virais isolados de indivíduos classificados como HAM/TSP esporádico (sHAM/TSP), 10 genomas virais isolados de indivíduos classificados como HAM/TSP familiar (fHAM/TSP), 6 genomas virais isolados de indivíduos assintomáticos (ASS) e 8 genomas virais isolados de indivíduos com Leucemia/linfoma de células T do adulto (ATLL) (**Tabela 1**).

Após o alinhamento das sequências genômicas obtidas e geração das sequências parciais referentes as regiões codificantes de cada proteína a ser estudada, foi possível identificar: - 23 “alelos protéicos” para a região do gene que codifica a proteína gp46; - 7 “alelos protéicos” para a região do gene que codifica a proteína gp21; 13 “alelos protéicos” para a região codificante da proteína Tax; 20 “alelos protéicos” para a região codificante da proteína HBZ; e 15 “alelos protéicos” para a região do gene que codifica para a proteína p12 (**Tabela 2**). Portanto, foram submetidas para a predição de epítomos um total de 78 sequências protéicas para identificação de ligantes em 27 moléculas de HLA.

**Tabela 1 - Caracterização dos genomas virais utilizados para a predição de epítomos.**

Número de Acesso	Perfil Clínico	Tamanho (nt)	Ano de Publicação
LC192536.1	sHAM/TSP	9021 bp	2017
LC192535.1		9026 bp	
LC192534.1		9010 bp	
LC192533.1		9033 bp	
LC192532.1		9032 bp	
LC192531.1		9032 bp	
LC192530.1		9031 bp	
LC192529.1		8964 bp	
LC192528.1		8957 bp	
LC192527.1		9031 bp	
LC192526.1		8958 bp	
LC192525.1		8986 bp	
LC192524.1		9032 bp	
LC192523.1		9032 bp	
LC192522.1		9032 bp	
LC192521.1		9031 bp	
LC192520.1		9032 bp	
LC192519.1		9028 bp	
LC192518.1		9029 bp	
LC192517.1		9033 bp	
LC192516.1	9031 bp		
KU214243.1	9040 bp	2016	
LC192515.1	fHAM/TSP	9019 bp	2017
LC192514.1		9032 bp	
LC192513.1		9033 bp	
LC192512.1		8952 bp	
LC192511.1		8999 bp	
LC192510.1		9031 bp	
LC192509.1		9009 bp	
LC192508.1		8944 bp	
LC192507.1		8993 bp	
LC192506.1		9032 bp	
LC192505.1	ASS	9032 bp	2017
LC192504.1		9032 bp	
LC192503.1		9032 bp	
LC192502.1		9032 bp	
LC192501.1		9036 bp	
LC192500.1		9030 bp	
LC378575.1	ATL	8980 bp	2018
LC183873.1		9034 bp	
D13784.1 e D00294		8400 bp	1988
AB979451		8941 bp	2015
U19949.1		9036 bp	1995
AF042071.1		8868 bp	1998
AB513134		9034 bp	2010
L03561		9043 bp	1993

pb (Pares de base), HAM/TSP (Mielopatia Associada ao HTLV/Paraparesia Espástica Tropical), sHAM/TSP (HAM/TSP esporádico), fHAM/TSP (HAM/TSP familiar), ASS (Assintomático-Não HAM/TSP), ATLL (Leucemia/Linfoma de Células T de Adulto)

**Tabela 2 - Descrição dos “alelos protéicos” identificados e submetidos à predição de epítomos.**

Número de Acesso	Código	Descrição do "alelo protéico"	Perfil Clínico
<b>gp46</b>			
LC192502	alelo_selvagem_gp46	Selvagem	sHAM/TSP-ATLL-ASS-fHAM/TSP
LC192500	alelo1_gp46	12F→S/305T→A	ASS
LC192501	alelo2_gp46	20G→S	ASS
LC192503	alelo3_gp46	55P→L	ASS
LC192504	alelo4_gp46	245V→I	ASS
D13784	alelo5_gp46	31I→V/55P→L/147R→H/160P→S/192P→S/250S→P	ATLL
AB513134	alelo6_gp46	19L→F/137Q→H	ATLL
LC192517	alelo7_gp46	307G→R	sHAM/TSP
KU214243	alelo8_gp46	7T→S/9I→T/55P→L/71V→I/72S→G/150I→L/160P→S 181S→N/290H→L	sHAM/TSP
LC192531	alelo9_gp46	137Q→H/245V→I	sHAM/TSP
LC192535	alelo10_gp46	59A→T	sHAM/TSP
LC192534	alelo11_gp46	14F→L	sHAM/TSP
LC192533	alelo12_gp46	55L→P/72S→G/188P→L	sHAM/TSP
LC192510	alelo13_gp46	55P→L/215Q→R	fHAM/TSP
LC192515	alelo14_gp46	55P→L/72S→G/187A→T	fHAM/TSP-sHAM/TSP-ATLL
LC192514	alelo15_gp46	12F→S/55P→L/72S→G	fHAM/TSP-sHAM/TSP
LC192513	alelo16_gp46	33V→I/55P→L	fHAM/TSP
LC192505	alelo17_gp46	204L→V	ASS
LC192521	alelo18_gp46	12F→S	sHAM/TSP
LC192523	alelo19_gp46	19A→T	sHAM/TSP
LC192509	alelo20_gp46	72S→G	fHAM/TSP
U19949	alelo21_gp46	19L→F/55P→L	ATLL
J02029	alelo22_gp46	19L→F/55P→L/89I→T/137Q→H	ATLL
<b>gp21</b>			
LC192500	alelo_selvagem_gp21	Selvagem	ASS-fHAM/TSP-ATLL-sHAM/TSP
LC192501	alelo1_gp21	158 L→P	ASS
LC192502	alelo2_gp21	164 Y→N	ASS
KU214243	alelo3_gp21	18 V→M/32K→R/147A→V	sHAM/TSP
LC192529	alelo4_gp21	154C→R/164Y→H	sHAM/TSP
D13784	alelo5_gp21	21G→R/89R→C/91P→L/99P→S/171K→N	ATLL
J02029	alelo6_gp21	21R→G/89C→R/91L→P/99S→P/171N→K	ATLL
<b>Tax</b>			
LC192506	alelo_selvagem_Tax	Selvagem	ASS-sHAM/TSP-ATLL-fHAM/TSP
LC192510	alelo1_Tax	119Y→C/334Q→E	fHAM/TSP
LC192536	alelo2_Tax	323E→K/334Q→E	sHAM/TSP
LC192530	alelo3_Tax	334Q→E	sHAM/TSP-ASS
LC192519	alelo4_Tax	231Q→H	sHAM/TSP-ASS
KU214243	alelo5_Tax	1M→T/77S→F/94I→T/168L→P/191I→M/221V→A 279H→N/327D→N/329N→T/340L→S/343P→H	sHAM/TSP
LC183873	alelo6_Tax	14G→R	ATLL
U19949	alelo7_Tax	8G→R/221A→V/263K→E/304S→N	ATLL
AB979451	alelo8_Tax	253G→S	ATLL
AB513134	alelo9_Tax	221V→A/304N→S	ATLL
LC192502	alelo10_Tax	147W→R	ASS
LC192501	alelo11_Tax	97T→A/325E→G	ASS
D13784ATL	alelo12_Tax	2A→X/192E→K/220V→A/341S→L	ATLL

HBZ			
LC192514	alelo_selvagem1_HBZ	Selvagem	ASS-sHAM/TSP-ATLL-fHAM/TSP
LC192515	alelo1_HBZ	79D→E	fHAM/TSP-sHAM/TSP-ATLL
LC192513	alelo2_HBZ	73R→C/78R→C	fHAM/TSP
LC192507	alelo3_HBZ	15V→M/83R→Q	fHAM/TSP
LC192536	alelo4_HBZ	15V→G/63R→Q/182 R→Q	sHAM/TSP
LC192535	alelo5_HBZ	9E→P	sHAM/TSP
LC192534	alelo6_HBZ	87K→R/90R→Q	sHAM/TSP
LC192533	alelo7_HBZ	90R→Q/115R→Q/181 E→Q	sHAM/TSP
LC192532	alelo8_HBZ	58R→C	sHAM/TSP
LC192531	alelo9_HBZ	15V→A	sHAM/TSP
LC192528	alelo10_HBZ	15V→M	sHAM/TSP
LC192526	alelo11_HBZ	123K→R	sHAM/TSP
KU214243	alelo12_HBZ	25E→D/35E→K/65P→L/66P→L/69K→G/92K→R 96K→E/100K→R/107K→R/110E→K/118R→K/122K→E 127V→S/141K→Q/151Q→R/156K→R/159M→I/164I→V 170Q→E/184K→R/205A→V	sHAM/TSP
D13784	alelo13_HBZ	45A→T/68E→G/118R→K 122K→E/127V→E/164I→V/184K→R	ATLL
U19949	alelo14_HBZ	7V→A/16P→S/102I→T	ATLL
AB513134	alelo15_HBZ	7V→A/16P→S	ATLL
LC192504	alelo16_HBZ	164I→V	ASS
LC192502	alelo17_HBZ	63R→Q	ASS
LC192500	alelo18_HBZ	159M→I	ASS
LC192523S	alelo19_HBZ	9A→P	sHAM/TSP
p12			
LC192514	alelo_selvagem_p12	Selvagem	ASS-sHAM/TSP-ATLL-fHAM/TSP
LC192507	alelo1_p12	24P→S/70P→L	fHAM/TSP
LC192513	alelo2_p12	26D→N/85L→F	fHAM/TSP
LC192515	alelo3_p12	74I→L	fHAM/TSP
LC192536	alelo4_p12	90P→L	ASS
LC192529	alelo5_p12	42L→F/63S→P/91S→P/93P→S	sHAM/TSP
LC192534	alelo6_p12	63S→L/66L→P	sHAM/TSP
LC192533	alelo7_p12	38A→L/63S→L	sHAM/TSP
LC192526	alelo8_p12	30L→P	sHAM/TSP
LC192519	alelo9_p12	24P→S/70P→L/71S→P	sHAM/TSP
KU214243	alelo10_p12	26D→G/35P→L/39C→S/46F→S/53L→P/61F→S 80A→S/84F→S/87*/88R→K/90P→S/95A→T	sHAM/TSP
U19949	alelo11_p12	23S→P/25G→S/51N→G	ATLL
AB513134	alelo12_p12	23S→P/69S→G	ATLL
D13784	alelo13_p12	26D→E/35P→L/85L→P	ATLL
LC192501	alelo14_p12	29G→S	ASS

HAM/TSP (Mielopatia Associada ao HTLV/Paraparesia Espástica Tropical), sHAM/TSP (HAM/TSP esporádico), fHAM/TSP (HAM/TSP familiar), ASS (Assintomático-Não HAM/TSP), ATLL (Leucemia/Linfoma de Células T de Adulto). G (Glicina), A (Alanina), L (Leucina), V (Valina), I (Isoleucina), P (Prolina), F (Fenilalanina), S (Serina), T (Treonina), C (Cisteína), Y (Tirosina), N (Asparagina), Q (Glutamina), D (Aspartato), E (Glutamato), R (Arginina), K (Lisina), H (Histidina), W (Triptofano), M (Metionina).

Na literatura, foi possível identificar 11 estudos que associaram alelos de HLA à infecção pelo HTLV-1: essa relação poderia ser de susceptibilidade ou proteção às manifestações clínicas ocasionadas pelo vírus, mais precisamente ATLL e HAM/TSP ou com a maior susceptibilidade a infecção pelo vírus. Deste total, quatro estudos associaram sete alelos de HLA à suscetibilidade ao desenvolvimento de ATLL. Quanto à manifestação de HAM/TSP, alguns estudos revelaram associação e outros susceptibilidade. E apenas dois estudos relataram de forma direta uma relação entre alelos de HLA e susceptibilidade à infecção pelo vírus (**Tabela 3**).

**Tabela 3** - Alelos de HLA previamente descritos na literatura com alguma associação com a infecção pelo HTLV-1 ou sintomatologia.

Alelo de HLA	Referência	Associação	Perfil Clínico/ Infecção
HLA-A*26	Yashiki et al., 2001	Susceptibilidade	ATLL
	Matsushita et al., 2006		
HLA-B *40:02	Yashiki et al., 2001		
HLA-B * 40:06	Yashiki et al., 2001		
HLA-B * 48:01	Yashiki et al., 2001		
HLA-A, HLA-B and B2M	Kogure & Kataoka, 2017		
HLA-A*02	Jeffery et al., 1998	Proteção	HAM/TSP
	Jeffery et al., 1999		
	Talledo et al., 2010		
	Deschamps et al., 2010	Sem Associação	
	Taghaddosi et al., 2012		
HLA-B*54:01	Jeffery et al., 2000	Susceptibilidade	
HLA-B*07	Trevino et al., 2013		
HLA-Cw*08	Jeffery et al., 2000	Proteção	
	Rafatpanah et al., 2007	Susceptibilidade	
	Taghaddosi et al., 2012	Sem Associação	
HLA-G	Haddad et al., 2011	Susceptibilidade	
HLA-G3'-UTR	Cilião Alves et al., 2016		

ATLL: Leucemia/Linfoma de célula-T do Adulto; HAM/TSP: Mielopatia Associada ao HTLV-1/ Paraparesia Espástica Tropical; HTLV-1: *Human T-Lymphotropic Virus Type 1*

## II.5.2 Predição de Epítomos ligantes de moléculas de HLA classe 1

### II.5.2.1 Avaliação Geral das Predições

O número total de predições para cada “alelo protéico” e/ou proteína variou de 15.741 até 56.889 epítomos (**Tabela 4**).

**Tabela 4** – Avaliação Quantitativa do total de predições identificadas por proteína viral.

Proteína	“Alelos protéicos”	Total de predições
gp21	Todos	28.215
gp46	Alelos 1-18, 20, 21, 22 e 23	50.085
	Alelo 19	33.453
p12	Todos	15.741
Tax	Alelos 1-11, e 13	56.889
	Alelo 12	56.673
HBZ	Todos	33.453

Desse total de predições e epítomos, realizamos uma triagem pelos epítomos mais específicos, aqueles que apresentam menor *score* (RANK) na avaliação da ligação à molécula de HLA e que portanto, tem maior interesse biológico. E algumas diferenças foram observadas entre as proteínas (**Figura 8**).

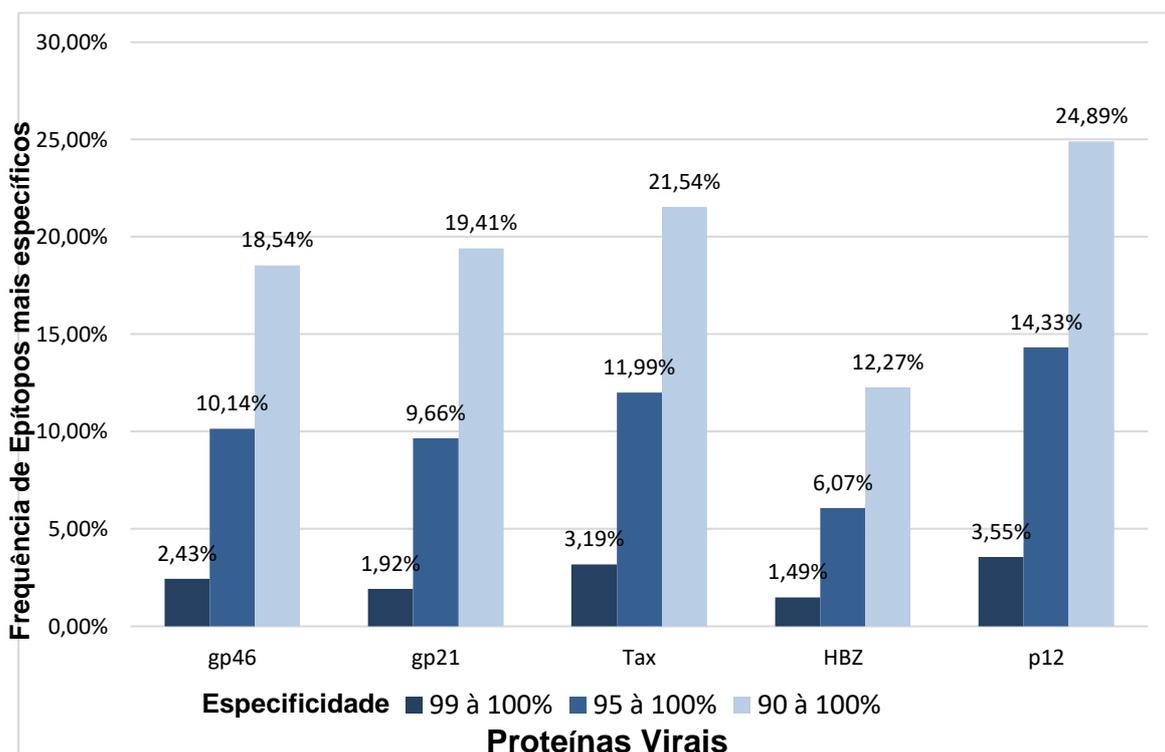
Com especificidade de 99%-100%, os percentuais variaram de 1,49% a 3,55%. A proteína p12 se destacou na maior proporção de epítomos específicos em todas as faixas de especificidade, seguida da proteína Tax. A proteína HBZ, por sua vez, se destacou pelo fenômeno contrário: menores percentuais em todas as faixas de especificidade. Os resultados encontrados para as proteínas estruturais gp46 e gp21 foram semelhantes (**Figura 8**).

No que diz respeito aos alelos de HLA, foi possível identificar epítomos ligantes para todos os alelos de HLA disponíveis na plataforma (27). Esses alelos de HLA, por sua vez, foram observados em todos os “alelos protéicos”, ou seja, não houve nenhuma especificidade/exclusividade dos alelos de HLA entre as proteínas, “alelos protéicos” e/ou perfis clínicos do hospedeiro.

No entanto, quando se observa o perfil dos alelos de HLA mais frequentes entre os “alelos protéicos” é possível descrever que para algumas proteínas as variações moleculares (de aminoácidos) não repercutiram na frequência das moléculas de HLA

ligantes, a exemplo da gp46, no entanto, algumas proteínas revelaram uma diversidade na ocorrência dos alelos de HLA mais frequentes entre os “alelos protéicos” (ANEXO II).

**Figura 8 – Avaliação Quantitativa dos Epítomos mais Específicos entre as proteínas estudadas.**



Cada proteína apresentou um alelo de HLA como sendo o mais frequente entre os “alelos protéicos”, não havendo similaridade nessa ocorrência. A gp46 apresentou o alelo de HLA B\*53:01 como sendo o mais frequente entre todos os seus “alelos protéicos”. Esta proteína não apresentou nenhum padrão, no que se refere à ocorrência das moléculas de HLA mais frequentes, que pudesse ser específico de algum status clínico, pois apresentou um perfil muito semelhante entre os “alelos protéicos”. A gp21, por sua vez, apresentou o perfil de maior diversidade na ocorrência dos alelos de HLA mais frequentes, não sendo possível reproduzir em nenhum dos “alelos protéicos” a ordem de aparecimento dos alelos de HLA, e sendo o HLA mais frequente o A\*02:03 para 6 dos 7 “alelos protéicos”. A proteína Tax revelou o HLA A\*32:01 como sendo o mais frequente em todos os “alelos protéicos” assim como também revelou o perfil mais conservador na ocorrência desses alelos: alta

similaridade na ordem de alelos de HLA mais frequentes entre os “alelos protéicos”. O HBZ destacou o alelo HLA\*B 44:02 como sendo o mais frequente para todos os “alelos protéicos”. Quanto ao perfil de ocorrência desses alelos de HLA, os perfis mais divergentes foram dos “alelos protéicos” 12 e 13 que sobressaem ao perfil dos demais. Por fim a proteína p12, que depois da gp21, foi a proteína com maior divergência na identificação dos alelos de HLA mais frequentes, não sendo possível associar tal ocorrência com o perfil clínico. Entre os alelos de HLA mais frequentes nesta proteína destaca-se o alelo B\*07:02 que foi o mais frequente em 12 dos 15 “alelos protéicos” (**ANEXO II**).

Correlacionando a ocorrência dos alelos de HLA ligantes dos epítomos preditos com os achados de associação já estabelecidos da literatura (**Tabela 3**), no que diz respeito à relação entre os primeiros e susceptibilidade/resistência à manifestação clínica ou infecção pelo HTLV-1, identificamos a presença do grupo de HLA A\*02 entre “alelos protéicos” de indivíduos com diferentes perfis clínicos. O alelo de HLA-A\*26:01 que já foi relacionado à susceptibilidade à ATLL apenas ocorreu para “alelos protéicos” do HBZ, sendo encontrado como o sétimo mais frequente em todos os grupos clínicos (**ANEXO II**). O alelo HLA-B\*07:02 que já foi associado ao desenvolvimento de HAM/TSP também foi identificado em nossas predições, da seguinte maneira: - na gp46 como sendo o quarto alelo de HLA mais frequente porém encontrado em todos os grupos clínicos; - na proteína gp21, que apesar de frequente, apresentou baixa ocorrência; em Tax, como o quinto mais frequente, sem diferenciação quanto ao perfil clínico; - e por fim, na proteína p12 como primeiro mais frequente, em todos os grupos clínicos; (**ANEXO II**).

Levando em consideração a Avaliação dos 10 alelos de HLA mais frequentes entre os “alelos protéicos” (**ANEXO II**), foi possível perceber que 26 alelos de HLA apareceram entre os 10 mais frequentes nas 5 proteínas, revelando uma diversidade alta na ocorrência dos alelos de HLA (**Tabela 5**).

Desses alelos de HLA, um deles (HLA\*A 30:02) apenas foi listado como mais frequente para a gp46 e outros 4 deles (B\*44:02, B\*44:03, B\*40:01 e A\*26:01) apenas foram listados como mais frequentes na proteína HBZ (**Tabela 5**). Oito alelos de HLA apenas apareceram como mais frequentes nas proteínas gp21 e HBZ juntas e 4 deles apareceram como mais frequentes exclusivamente nas proteínas gp46, p12 e Tax. E por fim, o alelo de HLA\*A 32:01 foi citado entre os mais frequentes em todas as proteínas (**Tabela 5**).

**Tabela 5** – Avaliação da ocorrência dos alelos de HLA mais frequentes por proteína viral.

Alelo de HLA	Proteína Viral				
	Estrutural		Regulatória		
	gp46	gp21	HBZ	p12	Tax
B*53:01	X			X	X
A*32:01	X	X	X	X	X
B*58:01	X	X		X	X
B*07:02	X	X		X	X
B*35:01	X			X	X
A*23:01	X			X	X
A*24:02	X			X	X
A*30:02	X				
B*57:01	X	X		X	X
B*51:01	X			X	X
A*02:01	X	X	X	X	
A*02:03		X		X	
A*02:06		X	X	X	X
A*68:02		X	X		
B*08:01		X	X		
B*15:01		X		X	
A*30:01		X	X		
A*11:01		X			
A*31:01		X	X		
A*03:01		X	X		
A*68:01			X	X	
A*33:01			X	X	
B*44:02				X	
B*44:03				X	
B*40:01				X	
A*26:01				X	

Desta avaliação foi possível observar que as proteínas gp21 e HBZ apresentaram perfis semelhantes no que se refere à ocorrência dos alelos de HLA mais frequentes. De modo que as proteínas gp46, Tax e p12 compartilharam semelhanças entre elas (**Tabela 5**).

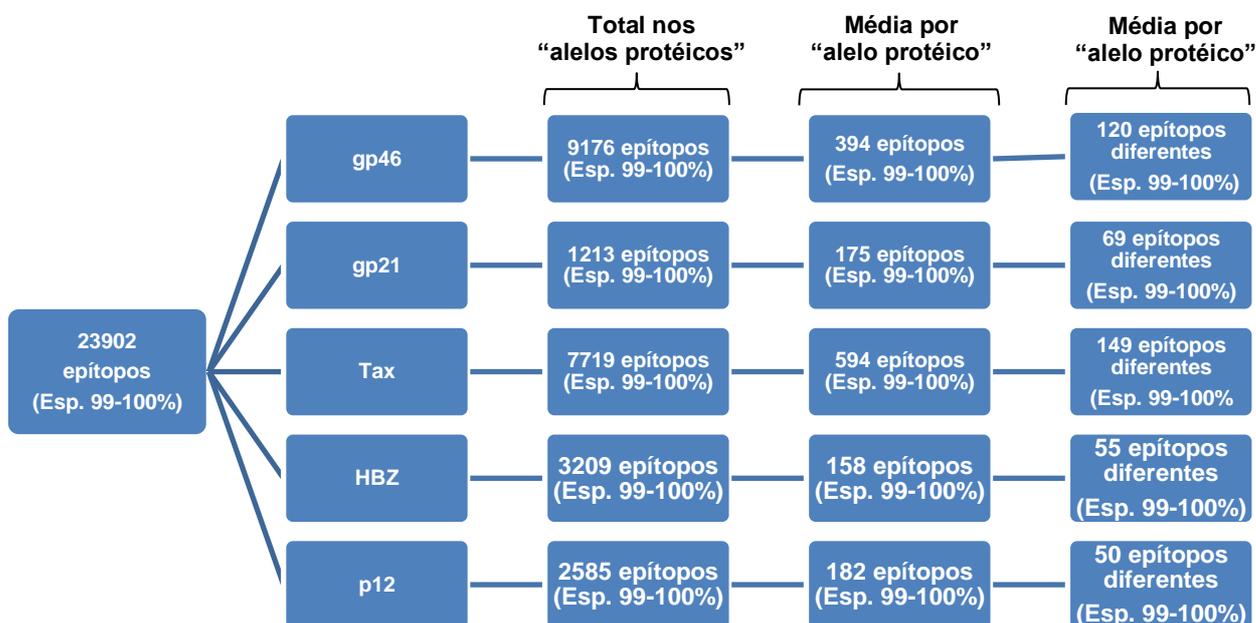
No entanto, as maiores diferenças observadas nas predições dizem respeito à afinidade e frequência dos epítomos apresentados, para as proteínas virais.

## II. 5.2.2 Caracterização dos Epítomos preditos: frequência dos epítomos com especificidade entre 99% e 100%

Considerando que quanto menor o IC<sub>50</sub> (RANK ou SCORE) maior a afinidade dos epítomos aos alelos de HLA, e por isso, mais importantes na resposta imune, analisamos as características quanto à frequência e afinidade de todos os vinte e três mil novecentos e dois (23902) epítomos (gp46=9176, gp21=1213, Tax=7719, HBZ=3209, p12=2585) preditos com especificidade entre 99 e 100% (**Figura 9**).

Com a predição foi possível perceber que a gp46 apresenta 120 epítomos diferentes que se repetem como ligantes de diferentes alelos de HLA. Esse número foi bem superior ao da gp21 que apresentou apenas 69 epítomos diferentes. Entre as proteínas regulatórias, a proteína Tax apresentou o maior número de epítomos (149), seguida pelo HBZ (55) e pela proteína p12 (50) (**Figura 9**).

**Figura 9** – Balanço quantitativo dos epítomos preditos, levando em consideração a proteína e seus “alelos”.



Nessa avaliação, alguns aspectos foram levados em conta: - identificação do epítopo (localização do aminoácido inicial); - “alelo protéico” onde o epítopo foi predito e informação sobre o status clínico que deu origem ao “alelo protéico” e; - frequência de ocorrência do epítopo entre os epítomos mais específicos. Os resultados foram organizados por proteína.

### II.5.2.2.1 gp46

Como é possível observar nos resultados tabulados para esta proteína (**Tabela 6**), identificamos a presença de vinte epítomos, mais frequentes, de um total de 120 diferentes epítomos distribuídos entre os aminoácidos 3 e 298.

**Tabela 6** – Avaliação da frequência dos epítomos mais comuns, identificados na gp46, com especificidade de 99 a 100%, entre os “alelos proteicos”.

"Alelo Protéico"	Perfil Clínico	Posição inicial do epítopo na proteína gp46 (AA)																			
		3	28	62	71	74	77	81	91	102	125	167	179	191	198	232	238	248	261	292	298
Selvagem	ASS-ATLL-fHAM/TSP-sHAM/TSP																				9
1	ASS	0								6				9	15						8
2	ASS								6	6		14		11	15				11		
3	ASS		9							6				11	15						9
4	ASS								6	6		6									
5	ATLL		8									21		0				19			9
6	ATLL																				
7	sHAM/TSP																				21
8	sHAM/TSP																				
9	sHAM/TSP				11																
10	sHAM/TSP																				
11	sHAM/TSP								8												
12	sHAM/TSP				0								6		11						
13	fHAM/TSP																				
14	fHAM/TSP-sHAM/TSP-ATLL				0										11						
15	fHAM/TSP-sHAM/TSP	0													11						
16	fHAM/TSP		13																		
17	ASS																				
18	sHAM/TSP	0																			
19	sHAM/TSP														11						
20	fHAM/TSP				0																
21	ATLL				0																
22	ATLL				0																
<b>Moda da Frequência</b>		<b>10</b>	<b>10</b>	<b>6</b>	<b>13</b>	<b>22</b>	<b>9</b>	<b>22</b>	<b>7</b>	<b>8</b>	<b>27</b>	<b>10</b>	<b>0</b>	<b>7</b>	<b>12</b>	<b>17</b>	<b>11</b>	<b>23</b>	<b>13</b>	<b>7</b>	<b>11</b>
<b>Percentual da Moda</b>		<b>8%</b>	<b>8%</b>	<b>5%</b>	<b>11%</b>	<b>18%</b>	<b>8%</b>	<b>18%</b>	<b>6%</b>	<b>7%</b>	<b>23%</b>	<b>8%</b>	<b>0%</b>	<b>6%</b>	<b>10%</b>	<b>14%</b>	<b>9%</b>	<b>19%</b>	<b>11%</b>	<b>6%</b>	<b>9%</b>

Os valores de frequência que foram discordantes da moda estão sinalizados na Tabela, e os locais em branco identificam a frequência igual à moda. sHAM/TSP (HAM/TSP esporádico), fHAM/TSP (HAM/TSP familiar), ASS (Assintomático-Não HAM/TSP), ATLL (Leucemia/Linfoma de Células T de Adulto).

O epítopo menos frequente (5%) está localizado na posição 62aa tendo se repetido seis vezes entre os “alelos protéicos”, enquanto o epítopo mais frequente (18%) se inicia na posição 125aa, com repetição de 27 vezes em cada “alelo protéico”. O epítopo situado no aminoácido 79 está presente exclusivamente, no “alelo protéico” 12 originado de um indivíduo sHAM/TSP, com frequência de 4% neste “alelo protéico” ou está em frequência muito baixa nos demais “alelos protéicos” (**Tabela 6**).

É possível perceber que as maiores variações na frequência dos epítomos ocorreram entre os “alelos protéicos” de indivíduos assintomáticos, seguido dos

indivíduos sHAM/TSP. Apesar das variações observadas, estas não representaram grandes diferenças na ocorrência dos epítomos, a exceção do epítomo que se inicia na posição 298aa cuja frequência no “alelo protéico” 7 (sHAM/TSP) é quase 2 vezes maior que a frequência desse mesmo epítomo nos demais “alelos protéicos” (**Tabela 6**).

### II.5.2.2.2 gp21

Analisando os epítomos identificados para os 7 “alelos protéicos” da gp21, notamos que essa proteína apresentou o segundo menor número (N=69) de epítomos com especificidade de 99%-100%. Desse total apenas doze epítomos apresentaram frequência superior a 9%, e se distribuíram entre os aminoácidos 18 e 168 (**Tabela 7**).

**Tabela 7** – Avaliação da frequência dos epítomos mais comuns identificados na gp21, com especificidade de 99 a 100%, entre os “alelos proteicos”.

"Alelo Protéico"	Perfil Clínico	Posição do epítomo na proteína gp21 (AA)												
		18	24	27	49	59	93	111	129	144	154	161	168	
Selvagem	ASS-fHAM/TSP-sHAM/TSP- ATLL													
1	ASS													
2	ASS												15	
3	sHAM/TSP	9		10							0			
4	sHAM/TSP									14	5	13		
5	ATLL													
6	ATLL						0							8
<b>Moda da Frequência</b>		<b>0</b>	<b>6</b>	<b>0</b>	<b>6</b>	<b>9</b>	<b>7</b>	<b>8</b>	<b>18</b>	<b>10</b>	<b>0</b>	<b>14</b>	<b>7</b>	
<b>Percentual da Moda</b>		0%	9%	0%	9%	13%	10%	12%	26%	14%	0%	20%	10%	

Os valores de frequência que foram discordantes da moda estão sinalizados na Tabela, e os locais em branco identificam a frequência igual à moda; aa – Aminoácido; sHAM/TSP (HAM/TSP esporádico), fHAM/TSP (HAM/TSP familiar), ASS (Assintomático-Não HAM/TSP), ATLL (Leucemia/Linfoma de Células T de Adulto).

Entre os epítomos mais comuns, os epítomos iniciados nas posições 24aa e 49aa mostraram ser os menos frequentes (9%), enquanto o epítomo situado inicialmente no aminoácido 129 destacou-se como o mais frequente (26%), repetindo-se 18 vezes em cada “alelo protéico”. Os epítomos situados inicialmente nas posições 18aa e 154aa mostraram serem exclusivos dos “alelos protéicos” 3 e 4, respectivamente, ambos originados de um indivíduo sHAM/TSP, ou estão em frequência muito baixa nos demais “alelos protéicos”. Também foi possível descrever que os “alelos protéicos” originados de indivíduos sHAM/TSP foram os que apresentaram maior variação na ocorrência dos epítomos, sendo possível identificar diferenças em cinco dos doze epítomos mais comuns (**Tabela 7**).

### II.5.2.2.3 Tax

A proteína Tax apresentou o maior número de epítomos (149) diferentes sendo apresentados entre os “alelos protéicos” aos alelos de HLA. Entre os 149 epítomos identificados nesta proteína, foi possível identificar 27 como sendo os mais comuns entre eles e ocupando a região que compreende do aminoácido 4 ao aminoácido 312 (**Tabela 8**).

No que diz respeito aos perfis clínicos, foi possível observar que as maiores variações nas frequências referem-se aos “alelos protéicos” originados de indivíduos com ATLL, uma vez que todos eles apresentaram ao menos uma variação em relação à frequência de ocorrência dos epítomos. Observou-se também que o “alelo protéico” 5 originado de um indivíduo com sHAM/TSP foi o mais divergente considerando todos os “alelos protéicos”, sendo possível identificar 7 pontos discordantes na frequência dos 27 epítomos (**Tabela 8**).

As menores frequências foram observadas para os epítomos situados inicialmente nas posições 74aa, 125aa, 139aa e 248aa, que se repetiram 7,6,7 e 8 vezes, respectivamente, entre os “alelos protéicos”. Do contrário, os epítomos mais frequentes (11aa, 62aa, 70aa, 97aa e 157aa) se repetiram entre 17 a 25 vezes entre os “alelos protéicos”. Desta forma pode-se perceber que dos 6 epítomos mais frequentes, 5 deles se concentram no primeiro terço da proteína (entre os aminoácidos 4 e 89 (**Tabela 8**).

Por fim, dos resultados encontrados para a proteína Tax, foram descritos três epítomos exclusivos: são eles os epítomos nas posições iniciais 88aa, 165aa e 312aa. O epítomo posicionado no aminoácido 88 foi exclusivo (ou estava em frequência muito baixa nos demais “alelos protéicos”) do “alelo protéico” 11 originado de um indivíduo ASS. O aminoácido 165 inicia um epítomo exclusivo (ou em frequência muito baixa nos demais “alelos protéicos”) do “alelo protéico” 5 originado de um indivíduo sHAM/TSP. E por fim, o epítomo iniciado no aminoácido 312 mostrou-se exclusivo (ou em frequência muito baixa nos demais “alelos protéicos”) de três dos quatro “alelos protéicos” originados de indivíduos ATLL e um indivíduo fHAM/TSP (“alelo protéico” 1) (**Tabela 8**).

**Tabela 8** – Avaliação da frequência dos epítomos mais comuns identificados na proteína Tax, com especificidade de 99 a 100%, entre os “alelos proteicos”.

"Alelo Protéico"	Perfil Clínico	Posição inicial do epítomo na proteína Tax (AA)																											
		4	11	54	62	70	74	88	97	111	125	139	157	163	165	180	186	219	242	248	273	276	282	284	291	297	300	312	
Selvagem	ASS-fHAM/TSP sHAM/TSP-ATLL																												
1	fHAM/TSP									0																		27	
2	sHAM/TSP																												
3	sHAM/TSP-ASS																												
4	sHAM/TSP-ASS																												
5	sHAM/TSP					14	8							23	6		12				15	12							
6	ATLL		17																								18		
7	ATLL	13																							11	17	27		
8	ATLL																		0										
9	ATLL																								11	17	18		
10	ASS										0																		
11	ASS							10	19																				
12	ATLL										7						15												
<b>Moda da Frequência</b>		14	25	11	19	17	7	0	20	12	6	7	19	15	0	15	16	10	15	8	18	11	10	11	11	10	16	0	
<b>Percentual da Moda</b>		9,4%	16,8%	7,4%	12,8%	11,4%	4,7%	0,0%	13,4%	8,1%	4,0%	4,7%	12,8%	10,1%	0,0%	10,1%	10,7%	6,7%	10,1%	5,4%	12,1%	7,4%	6,7%	7,4%	7,4%	6,7%	10,7%	0,0%	

Os valores de frequência que foram discordantes da moda estão sinalizados na Tabela, e os locais em branco identificam a frequência igual à moda. sHAM/TSP (HAM/TSP esporádico), fHAM/TSP (HAM/TSP familiar), ASS (Assintomático-Não HAM/TSP), ATLL (Leucemia/Linfoma de Células T de Adulto).

### II.5.2.2.4 HBZ

A predição de epítomos realizada para a proteína HBZ apresenta um total de 55 epítomos diferentes com especificidade entre 99% e 100%, sendo a proteína com menor números de epítomos preditos com esta característica (**Tabela 9**).

**Tabela 9** – Avaliação da frequência dos epítomos mais comuns identificados no HBZ, com especificidade de 99 a 100%, entre os “alelos proteicos”.

"Alelo Protéico"	Perfil Clínico	Posição inicial do epítopo na proteína HBZ (AA)									
		1	13	25	45	97	100	123	158	191	198
Selvagem	ASS-fHAM/TSP-sHAM/TSP-ATLL										
1	fHAM/TSP- sHAM/TSP - ATLL										
2	fHAM/TSP										
3	fHAM/TSP		8								
4	sHAM/TSP		0				8				
5	sHAM/TSP	0									
6	sHAM/TSP										
7	sHAM/TSP										
8	sHAM/TSP										
9	sHAM/TSP		0								
10	sHAM/TSP		8								
11	sHAM/TSP										
12	sHAM/TSP			10				0	0		9
13	ATLL				12						
14	ATLL	8	0			10					
15	ATLL	8	0								
16	ASS										
17	ASS										
18	ASS										
19	sHAM/TSP	0									
<b>Moda da Frequência</b>		9	5	13	15	0	0	6	8	12	7
<b>Percentual da Moda</b>		16%	9%	24%	27%	0%	0%	11%	15%	22%	13%

Os valores de frequência que foram discordantes da moda estão sinalizados na Tabela, e os locais em branco identificam a frequência igual à moda. sHAM/TSP (HAM/TSP esporádico), fHAM/TSP (HAM/TSP familiar), ASS (Assintomático-Não HAM/TSP), ATLL (Leucemia/Linfoma de Células T de Adulto).

Desse total de cinquenta e cinco epítomos, apenas dez apresentaram frequência igual ou superior a 9%, o que equivale a cinco repetições em cada “alelo protéico”, ocupando o intervalo compreendido entre os aminoácidos 1 ao 198. Com esta frequência (9%) está o epítopo localizado inicialmente no aminoácido 13, que também mostrou ser o epítopo com maior variação na ocorrência entre os “alelos proteicos”. Os epítomos preditos que iniciaram nos aminoácidos 25 e 45 se

apresentaram com uma frequência de 24% e 27%, sendo portanto os mais frequentes para esta proteína (**Tabela 9**).

Considerando o perfil clínico que deu origem aos “alelos protéicos”, é possível observar que todos os representantes com perfil ATLL apresentaram variações nas frequências dos epítomos preditos, o contrário foi observado para os representantes com perfil ASS cujas frequências foram semelhantes ao panorama geral da proteína (**Tabela 9**).

Os epítomos localizados inicialmente nas posições 97aa e 100aa mostram ser exclusivos aos “alelos protéicos” 14 e 4, cujos perfis clínicos são de indivíduos ATLL e sHAM/TSP, respectivamente, ou apresentaram frequências muito baixas nos demais “alelos protéicos”.

#### II.5.2.2.5 p12

A proteína p12 foi a proteína com maior variabilidade na ocorrência e frequência dos epítomos mais comuns, porém sem nenhuma associação com os perfis clínicos (**Tabela 10**).

**Tabela 10** – Avaliação da frequência dos epítomos mais comuns identificados na proteína p12, com especificidade de 99 a 100%, entre os “alelos proteicos”.

"Alelo Protéico"	Perfil Clínico	Posição inicial do epítomo na proteína p12 (AA)													
		4	10	23	33	35	44	46	58	64	67	69	74	76	88
Selvagem	ASS-fHAM/TSP sHAM/TSP- ATL														
1	fHAM/TSP			0			12			0		29	18		
2	fHAM/TSP													10	
3	fHAM/TSP										5	13	19		
4	ASS									5	9	13	19		
5	sHAM/TSP														12
6	sHAM/TSP				6				8	5					9
7	sHAM/TSP														
8	sHAM/TSP			0											
9	sHAM/TSP			0						0		35			
10	sHAM/TSP			6	12	15	11	0	7				16	0	21
11	ATLL			10			8	10							
12	ATLL			7	9	8						11			
13	ATLL			6										7	
14	ASS			8											
<b>Moda da Frequência</b>		<b>10</b>	<b>5</b>	<b>5</b>	<b>0</b>	<b>0</b>	<b>16</b>	<b>8</b>	<b>6</b>	<b>6</b>	<b>0</b>	<b>15</b>	<b>21</b>	<b>8</b>	<b>10</b>
<b>Percentual da Moda</b>		20%	10%	10%	0%	0%	32%	16%	12%	12%	0%	30%	42%	16%	20%

Valores de frequência discordantes da moda estão sinalizados na Tabela, e os locais em branco identificam a frequência igual à moda. sHAM/TSP (HAM/TSP esporádico), fHAM/TSP (HAM/TSP familiar), ASS (Assintomático-Não HAM/TSP), ATLL (Leucemia/Linfoma de Células T de Adulto).

Como se pode observar (**Tabela 10**), a proteína p12 apresentou 14 epítomos com frequência acima de 10% em pelo menos um “alelo protéico”, de um total de 50 epítomos diferentes identificados. Destes epítomos que ocupam a região do aminoácido 4 ao 88, destacam-se os epítomos cujos aminoácidos iniciais são os das posições 44, 69 e 74, pois apresentaram as maiores frequências sendo elas de 32%, 30% e 42%, respectivamente.

Os epítomos situados inicialmente dos aminoácidos 33, 35 e 67 foram identificados como sendo exclusivos (ou estão em frequência muito baixa nos demais “alelos protéicos”) dos “alelos protéicos” 6, 10 e 12 (epítomo 33aa), 10 e 12 (epítomo 35aa) e 3 e 4 (epítomo 67aa) (**Tabela 10**).

Interessante observar que nesta proteína alguns dos epítomos identificados possuem localização muito próxima: é o caso dos epítomos 33 e 35, 44 e 46, 67 e 69, e 74 e 76.

O epítomo localizado inicialmente na posição 23aa foi o que apresentou a maior variabilidade na frequência entre os “alelos protéicos”, podem estar ausentes em indivíduos com sHAM/TSP ou podendo ter o dobro da frequência em um indivíduo ATLL (**Tabela 10**).

### **II. 5.2.3 Caracterização dos Epítomos preditos: afinidade dos epítomos com especificidade entre 99% e 100%**

Esta avaliação foi realizada através da comparação dos registros de SCORE dos epítomos mais comuns (identificados no item anterior) entre os “alelos protéicos” e alelos de HLA em que eles apareciam.

Em relação à gp46 foram comparados os epítomos situados inicialmente nas posições 74aa (18% de frequência), 81aa (18% de frequência), 125aa (23% de frequência) e 248aa (19% de frequência) entre os 23 “alelos protéicos”. Apenas foi possível observar diferenças na afinidade de ligação dos epítomos 81aa e 248aa para os “alelos protéicos” 22 e 5, respectivamente, quando comparado aos demais vinte e um “alelos protéicos” (**Tabela 11**). Para a posição 248aa o “alelo protéico” 5 apresentou maior afinidade que os demais para os quatro alelos de HLA avaliados. Para a posição 81aa, no entanto, o “alelo protéico” 22 mostrou dois aumentos (Alelos de HLA A\*03:01 e A\*11:01) na afinidade e duas diminuições na afinidade (Alelos de HLA A\*02:03 e A\*02:01). Ambos os “alelos protéicos” que apresentaram diferenças nas afinidades dos epítomos 81aa e 248aa estavam presentes em indivíduos ATLL

(Tabela 11).

**Tabela 11** – Avaliação Descritiva das diferenças encontradas nas afinidades dos epítomos mais comuns entre os diferentes “alelos protéicos”.

Proteína	Epítomo	Alelo de HLA	Perfil comum	Perfil discordante		
			Afinidade	Afinidade	"Alelo protéico" X Perfil Clínico	
gp46	248AA	HLA-B*07:02	0,13	0,07	5 ATLL	
		HLA-B*35:01	0,22	0,13		
		HLA-B*51:01	0,85	0,54		
		HLA-B*53:01	0,84	0,41		
	81AA	HLA-A*02:03	0,17	0,5	22 ATLL	
		HLA-A*03:01	0,27	0,11		
		HLA-A*02:01	0,16	0,42		
		HLA-A*11:01	0,42	0,22		
Tax	11AA	HLA-A*32:01	0,43	0,9	6 ATLL 12 ATLL	
		HLA-A*68:02	0,25	0,75		
		HLA-A*32:01	0,16	0,83		
	97AA	HLA-A*26:01	0,14	0,25	11 ASS	
		HLA-B*58:01	0,4	0,25		
		HLA-B*57:01	0,36	0,23		
		HLA-B*07:02	0,31	0,15		
		HLA-A*32:01	0,31	0,1		
		HLA-B*15:01	0,93	0,48		
	HBZ	25AA	HLA-A*02:03	0,21	0,58	12 sHAM/TSP
HLA-A*02:06			0,1	0,29		
HLA-A*02:01			0,17	0,39		
HLA-A*68:02			0,12	0,27		
HLA-B*08:01			0,97	0,8		
45AA		HLA-A*02:01	0,07	0,16	13 ATLL	
		HLA-A*02:03	0,14	0,36		
		HLA-A*32:01	0,06	0,14		
		HLA-A*68:02	0,42	0,12		
		HLA-B*07:02	0,56	0,76		
		HLA-A*26:01	0,49	0,28		
p12		44AA	HLA-A*02:03	0,58	0,27	10 sHAM/TSP
			HLA-B*51:01	0,48	1	
	HLA-A*02:01		0,89	0,43		
	HLA-B*53:01		0,88	0,59		
	69AA	HLA-B*53:01	0,05	0,13	12 ATLL 9 sHAM/TSP	
		HLA-B*58:01	0,17	0,22		
		HLA-B*35:01	0,15	0,39		
		HLA-A*32:01	0,55	0,68		
		HLA-B*35:01	0,15	0,22		
		HLA-B*07:02	0,13	0,22		
	74AA	HLA-A*32:01	0,55	0,16	3 sHAM/TSP 4 ASS	
		HLA-B*58:01	0,54	0,92		
		HLA-A*31:01	0,36	0,74		
		HLA-B*57:01	0,39	0,7		
			HLA-A*33:01	0,71	0,92	

HAM/TSP (Mielopatia Associada ao HTLV/Paraparesia Espástica Tropical), sHAM/TSP (HAM/TSP esporádico), fHAM/TSP (HAM/TSP familiar), ASS (Assintomático-Não HAM/TSP), ATLL (Leucemia/Linfoma de Células T de Adulto).

Avaliando a gp21, não foram observadas diferenças nas afinidades dos epítomos iniciados nas posições 129aa (26% de frequência) e 161aa (20% de frequência) entre os 7 “alelos protéicos” avaliados.

Quanto à proteína Tax, os epítomos que tiveram os SCORES de afinidade comparados foram os situados nas posições: 11aa (16,8% de frequência), 62aa (12,8% de frequência), 97aa (13,4% de frequência) e 157aa (12,8% de frequência). Desses, apenas dois epítomos apresentaram diferenças nas afinidades: foram eles o epítomo da posição 11aa e o da posição 97aa. O primeiro mostrou diminuição das afinidades em dois alelos de HLA do “alelo protéico” 6 e um alelo de HLA do “alelo protéico” 12, ambos de indivíduos com ATLL. O segundo (97aa) mostrou diferenças na afinidade com 7 alelos de HLA, aumentando a afinidade em 6 deles, sendo todas estas variações observadas no “alelo protéico” 11 de um indivíduo ASS (**Tabela 11**).

Avaliando a proteína HBZ, os epítomos posicionados inicialmente nos aminoácidos 25aa (24% de frequência), 45aa (27% de frequência) e 191aa (22% de frequência) foram comparados quanto a diferenças nas afinidades. O epítomo situado na posição 25aa mostrou-se diferente na capacidade de ligação em cinco alelos de HLA, havendo diminuição da afinidade em quatro deles. O epítomo 45aa, por sua vez, mostrou-se com maior afinidade em três alelos de HLA e menor afinidade em outros três alelos de HLA. Para o primeiro epítomo (25aa) as variações foram observadas num “alelo protéico” de um indivíduo sHAM/TSP (12) e para o segundo epítomo (45aa) as observações foram em relação ao “alelo protéico” 13 (ATLL) (**Tabela 11**).

Por fim, na proteína p12 foram comparadas as afinidades dos epítomos iniciados nos aminoácidos 44 (32% de frequência), 69 (30% de frequência) e 74 (42% de frequência) da proteína, e as três posições revelaram variações. O epítomo da posição 44aa revelou aumento da afinidade em três alelos de HLA do “alelo protéico” 10 de um indivíduo sHAM/TSP. O “alelo protéico” 9 (sHAM/TSP) e o “alelo protéico” 12 (ATLL) tiveram aumento e diminuição da afinidade, respectivamente, para o epítomo na posição 69aa. Por fim o epítomo da posição 74aa revelou diminuição da afinidade em quatro alelos de HLA, quando comparado com os demais, em dois “alelos protéicos”: o 3 (sHAM/TSP) e o 4 (ASS) (**Tabela 11**).

## II.6 Discussão

Das cinco proteínas estudadas quanto à predição de epítomos, a proteína gp46 foi a que apresentou o maior número de “alelos protéicos” (22), seguido pela proteína HBZ (19), demonstrando, por isso, maior variabilidade molecular. A proteína com menor número de variação na composição de aminoácidos foi a gp21, apenas 6 “alelos protéicos” identificados.

Dentre os “alelos protéicos”, destacam-se quanto à maior ocorrência de troca de aminoácidos, os seguintes “alelos”: alelo 8\_gp46, alelo 5\_Tax, alelo 12\_HBZ, e alelo 10\_p12 todos eles presentes em indivíduos com sHAM/TSP, e o alelo 13\_HBZ de um indivíduo com ATLL. O perfil clínico HAM/TSP (somados os de ocorrência familiar e esporádica) foi o que apresentou o maior número de variações moleculares (troca de aminoácidos) entre as proteínas estudadas: foram 42 “alelos protéicos” identificados nesse grupo contra 19 de indivíduos com ATLL e 16 de indivíduos ASS (Não HAM/TSP). Tais observações sugerem que as cepas virais associadas à ocorrência das principais manifestações clínicas (HAM/TSP e ATLL) acumulam mais mutações ao longo do tempo, e em nossos resultados observamos isso mais acentuado em grupos de indivíduos HAM/TSP. Talvez esta variação molecular das cepas virais estejam associadas ao longo tempo de permanência do vírus no hospedeiro para então ocorrer a manifestação da HAM/TSP, que costuma ocorrer entre a 4ª e 5ª décadas de vida. O que pode ser melhor explicado inclusive pela maior pressão seletiva exercida, pelos mecanismos da resposta imune, a que estas cepas virais estão expostas nestes indivíduos com maior tempo de infecção até o desenvolvimento de doença, como é o caso da HAM/TSP.

Estudo que realizou a caracterização molecular de sequências da ORF-I do HTLV-1 de 156 pacientes com diagnóstico negativo ou positivo de HAM/TSP, demonstrou que algumas mutações podem estar associadas a esta manifestação clínica, e elas seriam: C39R, L40F, P45L, S69G e R88K. Além disso, as mutações P34L e F61L (ORF-I) e a mutação P45L, na proteína HBZ, se mostraram associadas com a carga proviral (BARRETO et al., 2016). Dessas mutações, foi possível identificar três delas (39, 61 e 88) num mesmo “alelo protéico” (10) de p12 originado de um indivíduo com sHAM/TSP. Outra mutação localizada no aminoácido 69 foi identificada em um “alelo protéico” (12) de um indivíduo com ATLL. Contudo esse achado é insuficiente para reforçar os dados da literatura, devido a baixa frequência

dessas mutações nos dados do estudo.

É importante relatar que apesar das doenças associadas ao HTLV-1 já terem sido originalmente descritas como entidades esporádicas, algumas apresentações familiares já foram relatadas também. As formas de transmissão, seja por relações sexuais ou pela transmissão vertical (mãe para filho) através da amamentação, explicam a agregação familiar de doenças associadas ao vírus, como é o caso da fHAM/TSP (GOURRAUD et al., 2011; ALVAREZ, 2016).

Em geral, a predição da ligação dos epítomos das cinco principais proteínas virais aos diferentes alelos de HLA, nos permite entender melhor como ocorre a identificação do vírus pelo sistema imune do hospedeiro. Além de tentar entender melhor a contribuição do hospedeiro (através dos alelos HLA) e do vírus (epítomos proteicos) na principal manifestação clínica associada ao HTLV-1.

Neste contexto, é possível observar que no curso da doença (HAM/TSP), o infiltrado inflamatório contém números iguais de células CD4<sup>+</sup> e células T CD8<sup>+</sup>. No entanto, ao longo da infecção as células T CD8<sup>+</sup> são predominantemente detectadas nas lesões inflamatórias crônicas de pacientes com maior duração da doença (AYE et al., 2000; ENOSE-AKAHATA & JACOBSON, 2019). Esse achado sinaliza a importância desta resposta imune para a infecção pelo HTLV-1.

Além disso, apesar de a infecção pelo HTLV-1 ser considerada latente nos indivíduos infectados, a presença de respostas imunes específicas contra o HTLV-1 cronicamente ativadas, sugere que os antígenos virais podem ser continuamente sintetizados (ENOSE-AKAHATA & JACOBSON, 2019).

Na avaliação geral da predição de epítomos foi possível perceber que a proteína Tax, seguida da proteína gp46, apresentou o maior número de predições. Essa observação poderia se justificar pelo tamanho da proteína. A proteína Tax é a que possui uma sequência maior (353 aminoácidos), seguida da gp46 que possui 312 aminoácidos no dataset montado para este estudo. A menor proteína, p12, foi a que também apresentou o menor número de predições (15.741).

Quando avaliamos os epítomos mais específicos, ou seja, aqueles que apresentaram os menores registros do Score de afinidade, observamos que a proteína p12, apesar de ser a de menor tamanho, e a de menor número de predições foi a que apresentou uma maior proporção (3,55%) de epítomos com especificidade entre 99-100% (Score entre 0 e 1). Em segundo lugar se apresentou a proteína Tax com 3,19% dos epítomos totais com esta característica (Score entre 0 e 1), e em terceiro lugar a

gp46. A proteína HBZ foi a que apresentou a menor proporção de epítomos altamente específicos.

Ao contrário dos níveis de expressão de Tax e HBZ, a frequência de linfócitos T citotóxicos específicos de Tax é maior do que os CTLs específicos de HBZ (KANNAGI et al., 2019). Talvez esta evidência ajude a explicar o grande número de epítomos específicos identificados em Tax e o menor número na proteína HBZ.

É interessante observar que o mesmo perfil que une as proteínas gp21 e HBZ, no que se refere ao percentual (1,92%-gp2p1 e 1,49%-HBZ) de epítomos específicos, também pôde ser observado quanto aos alelos de HLA cujos epítomos são específicos. Estas proteínas compartilharam semelhanças na ocorrência dos alelos de HLA, de modo que possuem 8 alelos de HLA comuns nas predições. O mesmo ocorre quando se observa as proteínas, gp46, Tax e p12 juntas. Estas três compartilharam semelhanças no perfil de utilização das moléculas de HLA para a apresentação de epítomos.

O genótipo das moléculas de HLA classe I presentes nos indivíduos infectados auxilia na determinação da especificidade e a eficácia da resposta de células T CD8<sup>+</sup> ao vírus, e por isso controlam a carga proviral no hospedeiro e influencia a suscetibilidade a doenças associadas ao HTLV-1. Os genes que codificam as moléculas de HLA classe I, como HLA-A\*02 e HLA-Cw\*08, foram associadas à redução significativa da carga proviral e portanto, à proteção ao desenvolvimento da HAM/TSP no Japão (JEFFERY et al., 1999; JEFFERY et al., 2000). Semelhante efeito protetor das moléculas de HLA-A\*02 também foi observado no Brasil (CATALAN-SOARES et al., 2009). Além disso, os genes codificantes das moléculas de HLA classe I, HLA-A\*02 e HLA-Cw\*08, mostraram maior afinidade com a proteína HBZ, e menor carga proviral e risco de desenvolver HAM/TSP (MACNAMARA et al., 2010). Em contraste, os alelos da classe I, HLA-B\*07 e HLA-B\*5401, e o alelo de classe II, HLA-DRB1 \* 0101, foram associados a uma maior suscetibilidade ao desenvolvimento da HAM/TSP (ENOSE-AKAHATA & JACOBSON, 2019).

Nos achados deste trabalho, três alelos de HLA classe I A\*02 (A\*02:01, A\*02:03 e A\*02:06) foram identificados como ligantes dos epítomos preditos, porém sem associação com as sintomatologias. O alelo A\*02:01 apresentou epítomos para gp46, gp21, HBZ e p12, quanto o alelo A\*02:06 apresentou epítomos das proteínas gp21, HBZ, Tax e p12. O alelo B\*07:02 também foi identificado entre os ligantes dos epítomos preditos pelas proteínas, gp46, gp21, Tax e p12. Os alelos HLA-Cw\*08 e HLA-B\*5401,

por sua vez, não foram identificados como ligantes dos epítomos preditos neste trabalho.

Na predição de epítomos da gp46 foi possível observar, diferente das demais proteínas, uma maior variação molecular dos “alelos protéicos” oriundos de indivíduos ASS (Não-HAM/TSP). Os epítomos sugeridos cobrem toda a sequência protéica utilizada e também estão presentes em regiões que já foram previamente caracterizadas pela prevalência de epítomos lineares, como é o caso da região que compreende os aminoácidos 53-75aa (TALLET et al., 2001), onde se encontra um dos epítomos mais prevalentes dessa proteína. Também foi possível identificar que um segundo epítomo mais prevalente (81aa) entre os “alelos protéicos” está localizado dentro do *Motif* (YSLY) de endereçamento da gp46 para a membrana plasmática (ILINSKAYA et al., 2010). Ficando evidente a necessidade de se estudar um pouco mais sobre essa proteína, levando em consideração que esta se trata do principal alvo dos anticorpos produzidos pelo hospedeiro, frente à infecção.

Algumas experiências vacinais positivas baseadas em peptídeos da gp46 e de Tax elevaram o nível de anticorpos e a produção de IFN- $\gamma$  levando à indução de uma alta resposta celular e uma diminuição significativa na carga proviral, bem como proteção parcial em animais imunizados. Frangione-Beebe e cols. (2000) relataram a capacidade de um protótipo vacinal contra o HTLV-1 compreendendo um peptídeo da gp46 (aa 175-218) ligado a um epítomo promíscuo de células T (aa 288-302), para eliciar a proteção contra o vírus em modelos de camundongos e coelhos. Por outro lado, a construção do peptídeo multivalente a partir de epítomo da proteína Tax (aa 11-19, aa 178-186 e aa 233-241) ligado por meio de resíduos duplos de arginina, foi desenhada por Sundaram e cols. (2003), para avaliar as respostas imunológicas. Outro epítomo de fusão originado da proteína Tax ligado a outros peptídeos únicos foram eficientes para induzir a produção de citocina interferon-gama (IFN- $\gamma$ ) em pacientes com HAM/TSP em camundongos. O referido epítomo da gp46 (aa 175-218) e o peptídeo Tri-Tax (aa 11-19, aa 178-186 e aa 306-315), ambos fundidos ao epítomo promíscuo de células T, foram testados para avaliar a imunogenicidade e eficiência protetora em macacos-esquilo infectados com o HTLV -1.

Na predição de epítomos na gp21, a maior variação na ocorrência dos mesmos foi para os “alelos protéicos” originados de indivíduos sHAM/TSP, como também observado para as proteínas HBZ e p12. Epítomos não tão prevalentes (9%) foram

sugeridos no peptídeo fusão, importante domínio funcional nesta proteína. Em 1985, Cianciollo e colaboradores relataram a presença de um domínio imunossupressor (ISD) dentro da subunidade transmembrana (gp21) do envelope, que é oculto pela subunidade da superfície (gp46) do envelope (CIANCIOLLO et al., 1985). Este domínio imunossupressor possui importante papel inibindo a proliferação de células T, sendo então crucial para a infecção viral e para apoiar as células tumorais no escape imune (ROTEM et al., 2018). É uma região membranotrópica chamada de peptídeo de fusão-FP (Do inglês, *Fusion Peptide*) que está oculto dentro do envelope e que abrange os primeiros 23 aminoácidos da proteína. Uma alteração conformacional, após a ligação da subunidade da superfície ao receptor celular, expõe o PF levando à sua inserção na membrana plasmática e à fusão com a célula hospedeira e ativação do domínio imunossupressor (KOBÉ et al., 1999; ROTEM, 2018). Esse processo facilita então a entrada do vírus na célula-alvo e inibe a ativação das células T, o que induz uma diminuição na resposta imunológica do tipo Th1 e promove uma elevação nas respostas Th2 no hospedeiro, ou seja, esta inibição é caracterizada pela regulação negativa da resposta tipo Th1 da célula T (ROTEM et al., 2018).

A resposta antiviral através de linfócitos T citotóxicos é ativada *in vivo* e direcionada principalmente contra epítopos específicos da proteína regulatória Tax, e em menor proporção aos epítopos em Gag, Env e Pol (KANNAGI et al., 1991; PARKER et al., 1992; ELOVAARA et al., 1993; PIQUE et al., 1996). Desta forma, dependendo do epítipo viral e dos fatores genéticos do hospedeiro, tem se estabelecido a existência de reservatórios distintos de CTL em resposta aos diferentes epítopos, com a mesma função de reconhecimento de células infectadas, via HLA/peptídeo, no entanto, com proliferação e liberação de citocinas diferenciadas (LIM et al., 2000). Bangham & Osame (2005) demonstraram que a taxa de lise, via CTL, de células T CD4<sup>+</sup> expressando Tax foi correlacionada negativamente com a carga proviral tanto em indivíduos assintomáticos como em indivíduos HAM/TSP.

Nos achados relatados aqui, foi possível observar que as maiores variações na ocorrência dos epítopos se deu entre os “alelos protéicos” de indivíduos com ATLL. Esse achado foi único em relação às demais proteínas. Muitas dessas variações ocorreram inclusive no sítio de ativação de ATF/CREB da proteína que se situa entre os aminoácidos finais da mesma. O 97aa, que apresentou a maior frequência entre os “alelos protéicos”, está situado no sítio funcional da proteína de ligação ao DNA do

hospedeiro. Outros dois epítomos também prevalentes (97aa e 157aa) estão presentes no sítio de ativação de NF-KB (88AA-232aa) (HARROD, 2019). Outros epítomos menos frequentes estão distribuídos em outros domínios funcionais de Tax, como o Dedo de Zinco *Motif* (54aa) e Zíper de Leucina (125aa e 139aa).

A proteína HBZ foi documentada como essencial para evasão imunológica, persistência viral e manutenção de altos títulos provirais *in vivo* (HARROD, 2019). O HBZ é o único gene viral detectável em todas as fases da infecção em linhagens celulares e PBMCs infectados cronicamente derivadas de portadores assintomáticos do HTLV-1, pacientes HAM/TSP e com ATLL (HARROD, 2019).

Em 2014, Rowan e cols., em contrapartida, demonstraram que o linfócito citotóxico autólogo reconhece preferencialmente o epítomo HBZ 26-34, que é eficaz contra as células T CD4<sup>+</sup> infectadas com o HTLV-1 e que apresentam esse epítomo através da molécula de HLA-A\*02:01, por apresentar alta afinidade. Este referido epítomo e um segundo localizado na posição 45AA estão presentes na predição de epítomos realizada neste trabalho com prevalência de 24% e 27% entre os “alelos protéicos”, sendo as maiores observadas nesta proteína. Um terceiro epítomo prevalente (191aa) está localizado em um domínio do tipo Zíper de Leucina da proteína.

As variações ocorridas entre os epítomos preditos para a proteína HBZ ocorreram igualmente em “alelos protéicos” de indivíduos sHAM/TSP e ATLL, sendo tais variações mais concentradas no domínio de ativação da proteína (primeiros 50 aminoácidos).

A proteína p12I, por sua vez, interage e regula negativamente a expressão da cadeia pesada (Hc) das moléculas de MHC classe 1. A interação de p12I com as proteínas residentes no ER calreticulina e calnexina mais provavelmente ocorre com a forma não clivada de p12I e essa interação favoreceria a fuga do reconhecimento imunológico de células infectadas com vírus (FUKUMOTO et al., 2009). Em nossos achados houve a predição de epítomos para a localização do primeiro sítio de clivagem de p12 (10AA) com prevalência de 10% entre os “alelos protéicos”. Alvos do sistema imune nesta região podem reforçar a importância funcional da mesma. Interessantemente a proteína que apresentou a maior proporção de epítomos específicos foi também a que apresentou a maior variabilidade na ocorrência dos mesmos. Esta variação é mais frequente entre os “alelos protéicos” de indivíduos com sHAM/TSP, tendo sido encontrada em representantes de todas as

sintomatologias. Assim como esta variação não esteve associada à uma região da proteína e sim apresentou comportamento difuso.

Por fim, os achados de afinidade entre os epítomos preditos revelaram que dos nove epítomos que apresentaram diferenças na afinidade às moléculas de HLA, 6 deles estavam associados a “alelos protéicos” de indivíduos ATLL, 4 de indivíduos sHAM/TSP e 2 de indivíduos ASS. Estes achados corroboram outros resultados já discutidos anteriormente de que as maiores variações moleculares foram discriminadas em indivíduos com manifestação de doença. Desse modo, tais alterações moleculares nas sequências protéicas podem de fato estar repercutindo na conformação dos epítomos e, portanto, na capacidade de ligação à molécula de HLA. Discretamente os achados de variação na afinidade dos epítomos indicam uma tendência maior de diminuição da afinidade entre os “alelos protéicos” de indivíduos com ATLL e sHAM/TSP, sendo necessário outras avaliações para confirmar a existência ou não dessa tendência.

Apesar de a afinidade de ligação de peptídeos da proteína HBZ a moléculas de HLA de classe 1 ter sido considerada significativamente mais fraca do que a de peptídeos de Tax (HILBURN et al., 2011; MACNAMARA et al., 2010), não encontramos diferenças na afinidade dos epítomos quanto à proteína viral.

A persistência do HTLV-1 é mantida não por infecção *de novo* por meio de partículas virais, mas sobretudo pela longa vida útil e/ou proliferação clonal de células infectadas pelo vírus. Como consequência, aproximadamente 5% dos indivíduos infectados desenvolverão doença após um longo período de latência (BANGHAM & MATSUOKA, 2017). Foi demonstrado que clones infectados com provírus defeituosos exibiram maior abundância clonal do que aqueles intactos (KATSYUYA et al., 2019), revelando a importância de alterações genômicas virais. Assim, a caracterização do provírus com alta resolução é fundamental para entender a persistência viral e a patogênese.

Neste trabalho, não foi possível associar nenhum perfil de apresentação dos epítomo com as diferentes manifestações de doença, no entanto, os resultados apresentados aqui revelam existir algumas diferenças no que diz respeito à especificidade dos epítomos e aos alelos de HLA ligantes deles que merecem ser melhor investigados sobretudo nos indivíduos com sHAM/TSP e ATLL. Por isso, acreditamos que essa validação merece ser buscada em outras estratégias científicas.

Embora a heterogeneidade do desfecho dessas doenças ainda não esteja clara, provavelmente este será especificado pela interação entre fatores do vírus, como Tax, HBZ, carga proviral e a genética do hospedeiro e características epigenéticas. Especificamente em relação à resposta de células T específicas contra o HTLV-1, existem diferenças acentuadas: a resposta CTL é elevada em pacientes HAM/TSP, enquanto é prejudicada em pessoas que sofrem de ATLL, uma vez que os CTLs são supostamente essenciais para a vigilância antitumoral na infecção pelo HTLV-1, e seu comprometimento provavelmente favorece a leucemogênese (KANNAGI et al, 2019).

Além disso, motivos de peptídeos comuns homólogos de peptídeos do envelope viral (Env-192-199 e 237-243) e anticorpos contra o HBZ foram frequentemente detectados no líquido de pacientes com HAM/TSP (FUJIMORI et al., 2004). Estudos mostram também que uma medula espinhal mais atrofica em HAM/TSP já foi associada com maior porcentagem de linfócitos T CD8<sup>+</sup> e carga proviral no líquido de HAM/TSP (AZODI et al., 2017; ENOSE-AKAHATA & JACOBSON, 2019).

Desta forma, este compartimento (LCR), tão importante para a manifestação da HAM/TSP, pode ser o melhor microambiente para que novos estudos investigando diferentes epítomos apresentados por isolados virais seja melhor investigado. É possível que a apresentação diferencial de epítomos entre indivíduos HAM/TSP e ATLL, exista de fato, sendo necessário aprimorar a busca pelos epítomos, sobretudo no que diz respeito à melhor priorização de características específicas de cada manifestação clínica.

### III. CAPÍTULO II

**Caracterização do exoma de indivíduos infectados pelo HTLV-1:  
investigação de biomarcadores para manifestação de HAM/TSP.**

### III.1 Justificativa

A infecção pelo HTLV-1 não implica, necessariamente, no desenvolvimento de uma patologia em seus portadores (MARTINS et al., 2015). De modo que o estado assintomático, ou portador de uma patologia associada dependerá de fatores ambientais, virais e do hospedeiro. Estudo recente sugere que particularmente a suscetibilidade de um determinado indivíduo infectado pelo HTLV-1 para desenvolver HAM/TSP está profundamente conectada às duas diferenças nos componentes geneticamente determinados da resposta imune: hospedeiro e subgrupo do HTLV-1 (NOZUMA & JACOBSON, 2019). De forma que pouca ou nenhuma evidência dão força à associação unilateral de genótipos virais à manifestação de doença.

Estudo avaliando possíveis locais de integração do HTLV-1 no genoma de pacientes com HAM/TSP revelou forte viés para integração em regiões ativas do genoma (MEEKINGS et al., 2008), mais especificamente em locais de ligação para fatores de transcrição. Sabe-se também que o HTLV-1 liga-se ao organizador chave da cromatina quando vai inserir seu material genético no genoma do hospedeiro, levantando a hipótese de que o HTLV-1 desregula a expressão de genes do hospedeiro (COOK et al., 2017; MELAMED et al., 2018).

No contexto de fatores que contribuem para a manifestação de doença, a resposta imune do hospedeiro frente à infecção viral, principalmente a resposta celular desencadeada por células T CD8<sup>+</sup> específicas anti-HTLV, é reconhecida como um evento crucial, determinando o desfecho da infecção (BANGHAM & OSAME, 2005). Esta resposta celular parece ser influenciada pela via de infecção do hospedeiro, carga proviral, fatores genéticos individuais, como polimorfismos em genes do sistema HLA e genes envolvidos na resposta imune (VINE et al., 2002). A diminuição da expressão de moléculas de MHC-I, por exemplo, previne a apresentação de peptídeos imunogênicos, e, portanto, poderia relacionar a atuação de p12 à resposta imune efetiva e, portanto, à persistência viral, e também ressaltar a importância de moléculas de HLA.

Além disso indivíduos que desenvolvem HAM/TSP apresentam elevados níveis de carga proviral e exacerbada resposta imune, como se pode 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

vírus, já que o mesmo não é encontrado em pacientes ATLL (HANABUCHI et al., 2000).

Dada a importância da ativação das células T, em indivíduos acometidos com HAM/TSP, variações no genoma do hospedeiro que possam alterar esta via podem ser utilizadas como biomarcadores de progressão para doença. Nesse contexto, na ausência da co-estimulação, as células T que encontram antígenos virais, por exemplo, falham ao responder e morrem por apoptose ou entram em um estado prolongado de não responsividade. Desse modo, polimorfismos em proteínas que participam da cascata de co-estimulação e sinalização das células T podem ser importantes ferramentas no entendimento de que fatores genéticos do hospedeiro podem representar gatilhos para uma resposta imune exacerbada o que precipitaria o desenvolvimento da HAM/TSP.

Analisando os genótipos de 66 indivíduos infectados, entre assintomáticos e HAM/TSP, e 192 indivíduos controle, Vallinoto e cols. 2012 revelam que o polimorfismo FAS -670A/G pode estar associado não apenas com susceptibilidade à infecção, mas também com a progressão à HAM/TSP. Uma vez que a localização desse SNP pode favorecer sua ligação ao fator de transcrição STAT1, o que seria suficiente para desencadear regulação positiva ou negativa na expressão do gene FAS. Esse mesmo polimorfismo já havia sido associado à susceptibilidade, condição clínica e sobrevida de pacientes com ATLL (FARRE et al., 2008). FAS (TNFRSF6/CD95/APO-1) contém um domínio da super família TNFR (Receptor para Fator de Necrose Tumoral) que está envolvido na sinalização da apoptose de diversos tipos celulares, e por isso, está envolvido na patogênese de inúmeras malignidades e doenças do sistema imune (VALLINOTO et al., 2012). No entanto, este é o primeiro relatório científico, até o momento, que investiga todo o exoma de indivíduos infectados pelo HTLV-1 como estratégia de busca desses SNPs.

Em processos evolutivos, a variação genética, seja no hospedeiro ou no vírus, sofre pressões seletivas até ser selecionada positivamente ou negativamente, revelando as mudanças funcionais de cada informação gênica. Agentes infecciosos, por exemplo, podem representar uma poderosa força seletiva, especialmente em moléculas regulatórias, a exemplo das células T. Forni e cols. 2013, testando 15 genes que codificam proteínas associadas a co-estimulação de células T, em 39 espécies de mamíferos, revelaram que 9 desses genes eram alvo de seleção positiva (FORNI et al., 2013). Quando os SNPs avaliados buscavam associação com manifestação de

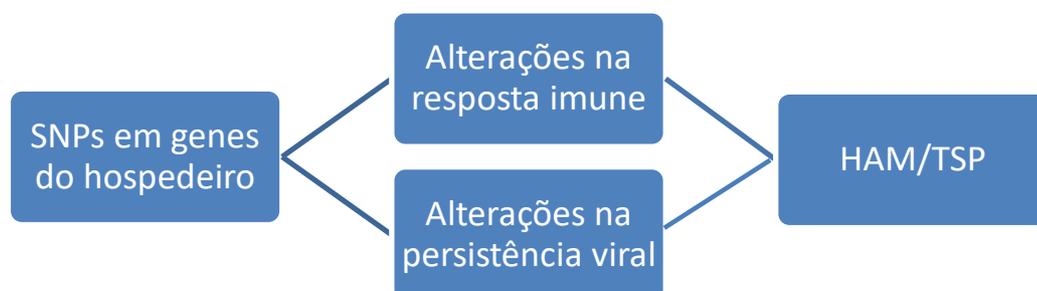
doença ou susceptibilidade à infecção, foi possível demonstrar que as variações associadas a doença em genes de células T são preferencialmente os alvos da seleção dirigida por patógenos (FORNI et al., 2013).

No entanto, está cada vez mais evidente de que não apenas os fatores virais e ou do hospedeiro vistos isoladamente contribuirão para esse entendimento, e sim o modo como a interação vírus/hospedeiro se desenvolve. Sendo importante para determinar o estado do portador como indivíduo assintomático ou paciente sofrendo de patologia seja de natureza neoplásica (ATLL) ou inflamatória (HAM/TSP) (TAGAYA; MATSUOKA; GALLO, 2019).

### III.2 Hipótese

Portanto, a hipótese desse trabalho (**Figura 10**) é de que polimorfismos em genes do hospedeiro podem estar associados às alterações na interação vírus/hospedeiro e no consequente desfecho da infecção. Associando à variação genética do hospedeiro à pressão seletiva exercida pelo HTLV-1 tais variações seriam selecionadas e cronicamente representariam marcadores de progressão para HAM/TSP.

**Figura 10** - Hipótese para o estudo de Caracterização do exoma de indivíduos infectados pelo HTLV-1: investigação de biomarcadores para manifestação de HAM/TSP.



### **III.3 Objetivos**

#### **III.3.1 Objetivo Geral**

- Pesquisar a ocorrência de SNPs em genes do hospedeiro que estejam associados à manifestação da HAM/TSP.

#### **III.3.2 Objetivos Específicos**

- Determinar a carga proviral nos indivíduos infectados pelo HTLV-1 assintomáticos e com HAM/TSP;
- Realizar a caracterização do hospedeiro infectado assintomático e com HAM/TSP, através da identificação de SNPs no exoma do mesmo, bem como identificar a frequência deles;
- Correlacionar os SNPs encontrados no exoma do hospedeiro com possíveis alterações funcionais;
- Avaliar o potencial de variáveis associadas ao hospedeiro, no desfecho clínico;
- Associar a presença do SNP, no genoma humano, com o estado clínico (Assintomático ou HAM/TSP) para avaliação do potencial do mesmo como biomarcador de progressão.

### III.4 Metodologia

#### III.4.1 Seleção da Amostra e Detecção Molecular do HTLV-1

Este estudo, de corte transversal, foi desenvolvido a partir de uma amostra de conveniência, uma vez que foram utilizadas apenas amostras de Sangue total e/ou PBMC (Células Mononucleares de Sangue Periférico) previamente armazenadas no Centro Integrativo Multidisciplinar de HTLV (CHTLV). Por isso, este estudo dispensa a assinatura do Termo de Consentimento Livre Esclarecido-TCLE, devido ao fato de os indivíduos serem anônimos não-vinculados e das amostras já estarem estocadas. No entanto, é importante ressaltar que o projeto foi aprovado pelo Comitê de Ética em Pesquisa do Centro de Pesquisa Gonçalo Moniz/FIOCRUZ (Número do parecer: 388.613).

Inicialmente, foram selecionadas 50 amostras de indivíduos, de ambos os sexos, infectados pelo HTLV-1 com *Western Blot* positivo. Todos os indivíduos, cujas amostras estocadas foram selecionadas, foram devidamente diagnosticados como positivos para o HTLV-1, sendo submetidos à triagem, pelo ELISA (Cambridge Biotech Corporation, Worcester, MA) e as amostras que se mostraram reagentes neste teste foram submetidas à confirmação pelo teste de *Western Blot* (HTLV Blot 2.4, Genelabs Diagnostics, Science Park Drive, Singapore). Como critérios de inclusão foram aplicados: detecção molecular do vírus, apresentar avaliação com neurologista (Para segregação em Assintomático e HAM/TSP), e ter o máximo de material biológico disponível para as análises moleculares. Os critérios de exclusão aplicados foram: casos de co-infecção e amostras que não cumpriam os critérios de inclusão. Todas as informações sociodemográficas dos indivíduos selecionados para o estudo, disponíveis nos prontuários médicos, foram tabulados em planilhas Excel.

A detecção molecular do HTLV-1 nas amostras de sangue total foi realizada nos Laboratórios de Biologia Molecular do NUPI (Núcleo de Pesquisa e Inovação da Escola Bahiana de Medicina e Saúde Pública) que dá suporte às atividades experimentais do CHTLV. As células mononucleares de sangue periférico foram obtidas das amostras de sangue total através de gradiente de centrifugação, e em seguida, o DNA genômico foi extraído utilizando o kit de extração Qiagen (QIAamp® DNA Blood Kit) do PBMC disponível, sendo a concentração final de DNA medida por espectrofotometria (Gene Quantpro RNA/DNA Calculator).

### III.4.2 Quantificação da Carga Proviral

A carga proviral do HTLV-1 em PBMC foi determinada pelo método de PCR quantitativo em tempo real usando o ABI Prism 7700 (*PE-Applied Biosystems*), utilizando  $10^6$  células. O gene da albumina foi utilizado como controle interno. A quantidade de provirus foi calculada pelo número de cópias do HTLV-1 (pol) por  $10^6$  células = [(número de cópias de pol)/(número de cópias de  $\beta$ -actina/2)]  $10^6$  segundo Dehé e cols., 1992.

### III.4.3 Determinação do Exoma

As amostras de DNA dos indivíduos infectados pelo HTLV-1, divididos nos dois grupos de Assintomáticos e HAM/TSP, foram submetidos para a determinação do exoma através da identificação de SNPs utilizando a Plataforma Illumina, sediada no Laboratório Multiusuário de Sequenciamento em Larga Escala e Expressão Gênica do Depto de Tecnologia, da Faculdade de Ciências Agrárias e Veterinárias - UNESP – Jaboticabal. Foi utilizado o kit *Infinium Core Exome-24* (**Tabela 12**) cujo chip rastreia 551.004 marcadores no exoma humano.

**Tabela 12** - Descrição do *Infinium Core Exome-24 v1.3 BeadChip*.

<b>Características</b>	<b>Descrição</b>		
Espécie	Humana		
nº total de marcadores	551,004		
nº de amostras por chip	24 amostras		
Quantidade de DNA/amostra	200 ng		
Genoma de Referência	GRCh37		
<b>Desempenho de dados</b>	<b>Valor <sup>a</sup></b>	<b>Especificações do produto <sup>b</sup></b>	
Call rate	99,80%	> 99% méd.	
Reprodutibilidade	99,99%	> 99,9%	
Desvio do log R	0,09	<0,30c	
<b>Espaçamento</b>			
Espaçamento (kb)	Média	Mediana	90th% <sup>c</sup>
	5.27	1.82	14.30

a. Os valores são derivados da genotipagem de amostras de referência 333 do HapMap.  
b. Marcadores do cromossômico Y para amostras femininas excluídos.  
c. Valor esperado para projetos típicos usando protocolos padrão Illumina.

### III.4.3.1 Controle de qualidade e limpeza de dados para genotipagem

Os procedimentos de controle de qualidade e limpeza de dados começam com o Controle Illumina SNP-Array Quality e a etapa de exportação de dados. Inicialmente os dados brutos recebidos da plataforma Illumina foram submetidos ao programa GenomeStudio onde foi calculado a taxa de genotipagem dos SNPs pesquisados e a taxa de genotipagem geral do exoma dos indivíduos. Posteriormente foram exportados do Genome Studio como arquivos nos formatos PED e MAP (formatos lidos pelo programa PLINK) usando o mesmo *plug-in* Illumina com os seguintes parâmetros: (i) "*Use Forward Strand*" (Fita positiva do DNA) definido como "*True*" e também com a (ii) remoção dos SNPs que não tiveram sinal. Depois deste processo, quatro etapas foram realizadas utilizando o software PLINK:

1. Limpeza inicial do conjunto de dados, onde filtros de dados básicos e procedimentos de "verificação de fios" (Fitas de leitura) foram aplicados;
2. Separação da informação genética referente aos cromossomos autossômicos e mitocondriais, bem como separação de SNPs do cromossomo X e Y em conjuntos de dados distintos;
3. Controle de Qualidade e limpeza de dados dos SNPs autossômicos;
4. Avaliação da taxa de perda genotípica por indivíduo;

Após a realização do controle de qualidade nos dados, os mesmos foram organizados em três conjuntos diferentes para serem submetidos nas diferentes análises subsequentes: estruturação da população de estudo, ancestralidade genética e análise estatísticas.

### III.4.3.2 Estruturação da população

A análise de componentes principais está associada à ideia de redução de massa de dados, com menor perda possível da informação. Sendo ela utilizada neste estudo para avaliar agrupamento de indivíduos. O agrupamento acontece devido à similaridade do conjunto de características que define o indivíduo da população estudada, isso permite entender o comportamento deles neste contexto (VARELLA, 2008). Portanto, para avaliar a estruturação da população em questão, observamos a dispersão desses dados de acordo com suas similaridades nas características genéticas. Para isso foi plotado a dispersão dos dados em uma figura no plano bidimensional, chamado de PCA (*Principal Component Analysis*) gerados pelo software Eigenstrat (PRICE et al., 2006).

### III.4.3.3 Ancestralidade genética

Por se tratar de uma variável dúbia, uma vez que a etnia que o indivíduo se autodeclarava poderia não corresponder a sua ancestralidade real, optamos por avaliar a ancestralidade a partir da comparação genética, homologia de SNPs, presentes no genótipo de três populações referências. A análise de ancestralidade foi estimada avaliando a proporção da frequência dos alelos característicos dos 3 *clusters* de referência da ancestralidade brasileira (Africados, europeus e nativos americanos) utilizando o software Admixture (PARRA et al., 2001).

### III.4.3.4 Genotipagem e Análises estatísticas

O cálculo da frequência alélica da população de estudo foi realizado, utilizando o *software* Plink (PURCELL et al., 2007), após o controle de qualidade, considerando todo os SNPs identificados nas amostras. Inicialmente foi calculada a frequência alélica dos SNPs na população geral e posteriormente foi calculada a frequência alélica dos SNPs separadamente nos grupos caso e controle.

Na avaliação estatística geral dos dados, submetemos as covariáveis etnia, gênero, carga proviral, idade e ancestralidade genética dos pacientes do estudo aos testes estatísticos que buscavam avaliar correlação delas com o desfecho clínico (HAM/TSP). A análise unitária das variáveis também foi realizada. Para isso, o teste de Chi quadrado ( $X^2$ ) ou teste exato de Fisher para variáveis categóricas e um teste U de Mann-Whitney, para variáveis contínuas, foram utilizados respondendo as características e comportamento de cada variável nos dados em questão (distribuição no histograma).

A análise de regressão logística, por sua vez, foi feita para testar a associação entre o status da doença e o genótipo sob a suposição de um modelo genético aditivo. Onde também idade, sexo, carga proviral e um dos três principais marcadores informativos sobre ancestralidade (Ancestralidade africana, por ser mais representativo da população Bahiana) foram usados como covariáveis para ajustar o modelo e verificar possível associação, uma vez que já foram descritas na literatura com esse papel. Desta forma construiu-se um modelo teórico explicativo, através do gráfico acíclico direcionado. Neste modelo, incorporamos os dados das covariáveis para então confirmar a parcela de influência de cada uma delas no desfecho e possíveis correlações entre elas de modo que duas ou mais juntas potencializassem a associação do genótipo com o desfecho.

O pacote estatístico de *software* R versão 4.0.2 (PURCELL et al., 2007) bem como o *software* Plink versão 1.90 (PURCELL et al., 2007) foi utilizado para todas as análises estatísticas desse estudo. Para a avaliação da regressão logística (análise mais densa) contamos com o suporte de um servidor disponibilizado pelo Laboratório de Diversidade Genética Humana (LDHG), alocado na Universidade Federal de Minas Gerais (UFMG), onde todas estas análises foram realizadas.

#### **III.4.3.5 Avaliação do desequilíbrio de ligação dos SNPs**

Em cruzamentos aleatórios os alelos de qualquer gene são combinados em genótipos, também de forma aleatória na frequência estipulada, segundo proporções de *Hardy-Weinberg*. Buscamos assim avaliar se os SNPs no genoma de indivíduos infectados pelo HTLV-1 que mostraram associação com o fenótipo estão se comportando segundo a teoria do equilíbrio de *Hardy-Weinberg*. Para essa investigação, avaliamos os SNPs identificados neste trabalho como associados com a patologia, comparando-os par a par com SNPs referências de três populações ancestrais: as mesmas utilizadas na análise de determinação da ancestralidade genética (africana, europeia e nativos americanos).

Foram calculadas as medidas pareadas de desequilíbrio de ligação e criada uma representação gráfica. Blocos foram gerados e definidos com base em algoritmo já definido, com limites de confiança de 95%, os valores e esquemas de cores baseados no  $r^2$ . Os SNPs que estavam altamente correlacionados com os SNPs de interesse foram identificados pelo  $r^2 \geq$  limiar de 0,8. Para a reprodução dessa avaliação, utilizamos o *software* Haploview (BARRETT et al., 2005).

#### **III.4.3.6 Levantamento Bibliográfico: busca por SNPs de suscetibilidade/associação com a HAM/TSP**

Foi realizada uma busca na literatura pelos SNPs já associados com a manifestação da HAM/TSP, através de artigos científicos já publicados na plataforma NCBI, utilizando os seguintes termos booleanos: “SNP AND HTLV-1” e “susceptibility AND HAM/TSP”; triando estudos publicados até novembro de 2019, que comprovaram associação da variação genética com a doença. Essa seleção possibilitou posteriormente uma busca direcionada/guiada, determinando, por exemplo, a frequência dos alelos desses SNPs triados, verificação da presença de

desequilíbrio de ligação e possível associação com a HAM/TSP dentro dos dados de exoma dos indivíduos investigados.

### III.5 RESULTADOS

#### III.5.1 Caracterização da Amostra

A partir dos critérios de exclusão estabelecidos, foi possível triar cerca de 50 amostras de sangue total de indivíduos infectados pelo HTLV-1 armazenadas no biobanco disponível no CHTLV. No entanto, apenas 48 foram finalmente incluídos no estudo, pela limitação do número de amostras a serem testadas pelo kit Illumina, na análise do Exoma. Das 48 amostras submetidas à determinação do exoma, duas foram excluídas por erro de duplicação. Sendo assim, 46 amostras seguiram para a etapa de tratamento e controle de qualidade dos dados (**Tabela 13**). Essas amostras foram organizadas em dois grupos: grupo caso, composto por indivíduos infectados pelo HTLV-1 com manifestação definida da HAM/TSP e pelo grupo controle, composto por indivíduo infectados pelo vírus, sem HAM/TSP e, chamados aqui de assintomáticos. Na tabela abaixo descrevemos as características da população de estudo, com as variáveis sociais disponíveis no cadastro do centro de acompanhamento desses indivíduos (CHTLV) e as informações laboratoriais a cerca das amostras de sangue total armazenadas no biobanco, vinculadas aos indivíduos do estudo, e que foram utilizadas para as análises moleculares.

**Tabela 13** - Caracterização da População submetida à determinação do exoma.

<b>ID Amostras</b>	<b>Perfil Clínico</b>	<b>Gênero</b>	<b>Etnia</b>	<b>Idade</b>	<b>Carga Proviral (PVL/10<sup>6</sup> células)</b>
TSP_01	HAM/TSP	F	Pardo	69	55245,49
TSP_02	HAM/TSP	F	Pardo	66	52986,86
TSP_03	HAM/TSP	F	Pardo	61	38505,86
TSP_04	HAM/TSP	F	Negro	55	64597,92
TSP_05	HAM/TSP	F	NA	84	44920,85
TSP_06	HAM/TSP	F	NA	81	NA
TSP_07	HAM/TSP	F	Pardo	60	44223,30
TSP_08	HAM/TSP	F	Pardo	63	28867,80
TSP_09	HAM/TSP	F	Pardo	63	36652,99
TSP_10	HAM/TSP	F	Negro	61	34343,05
TSP_11	HAM/TSP	F	Negro	72	29040,82
TSP_12	HAM/TSP	F	Pardo	48	34029,92
TSP_13	HAM/TSP	F	NA	55	70560,90
TSP_14	HAM/TSP	F	NA	90	27909,02
TSP_15	HAM/TSP	F	NA	67	42091,02
TSP_16	HAM/TSP	F	Negro	63	39314,67
TSP_17	HAM/TSP	F	Negro	47	58930,93
TSP_18	HAM/TSP	F	Negro	41	59603,02
TSP_19	HAM/TSP	F	Pardo	54	29090,79
TSP_20	HAM/TSP	M	Negro	65	57092,40
TSP_21	HAM/TSP	F	Pardo	54	31039,05
TSP_22	HAM/TSP	F	NA	58	NA
TSP_23	HAM/TSP	M	Negro	67	42455,02
ASS_24	ASS	M	Negro	56	30814,74
ASS_25	ASS	F	Pardo	44	32023,02
ASS_26	ASS	F	Negro	39	25022,30
ASS_27	ASS	F	Negro	42	28012,01
ASS_28	ASS	M	Negro	41	32092,02
ASS_29	ASS	F	NA	82	62042,01
ASS_30	ASS	M	NA	57	50232,01
ASS_31	ASS	F	Negro	61	44021,01
ASS_32	ASS	F	NA	51	56232,01
ASS_33	ASS	M	Negro	70	39093,12
ASS_34	ASS	M	Negro	58	56023,02
ASS_35	ASS	F	Pardo	55	44023,02
ASS_36	ASS	M	Pardo	50	30435,93
ASS_37	ASS	M	NA	54	25024,02
ASS_38	ASS	M	NA	55	31023,02
ASS_39	ASS	F	Pardo	50	20023,02
ASS_40	ASS	M	Branco	57	52042,03
ASS_41	ASS	F	Pardo	84	45023,20
ASS_42	ASS	M	NA	60	14240,02
ASS_43	ASS	M	Branco	61	35040,01
ASS_44	ASS	M	Pardo	61	23248,98
ASS_45	ASS	F	Negro	71	NA
ASS_46	ASS	F	Negro	27	NA

NA (Não informado), HAM/TSP (Mielopatia Associada ao HTLV/Paraparesia Espástica Tropical), ASS (Assintomático-Não HAM/TSP), F (Feminino), M (Masculino).

### III.5.2 Análise dos Dados Moleculares do Exoma

#### III.5.2.1 Controle de Qualidade

A análise do controle de qualidade dos dados obtidos na determinação do exoma diminuiu em 4 amostras a população a ser estudada: de uma amostragem de 48 indivíduos infectados pelo HTLV-1 para 44 indivíduos. Esse corte foi devido ao fato de que 2 indivíduos foram excluídos por erro de duplicação da amostra e, 2 indivíduos foram excluídos pela baixa taxa de genotipagem.

Desse modo, as análises dados moleculares do exoma foram baseadas em 3 conjuntos de dados com SNPs autossômicos consensos:

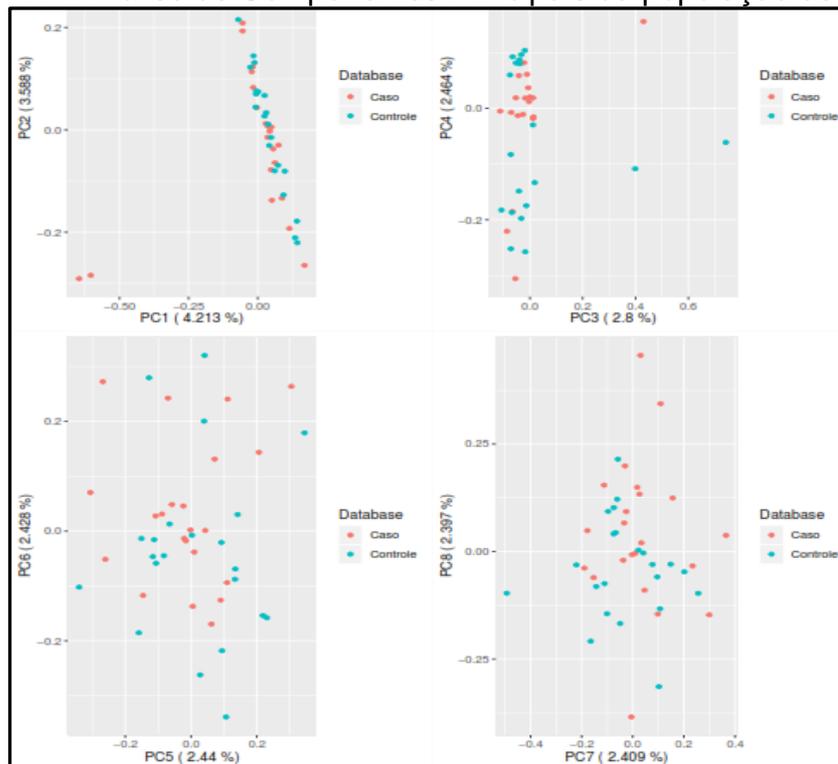
- Data set Final (N=44): gerado após a limpeza dos dados e controle de qualidade sendo então composto por 22 indivíduos ASS e 22 indivíduos HAM/TSP;
- Data set para Análise de regressão (N=40): Foram excluídos 4 genomas do dataset final (ASS\_45, ASS\_46, TSP\_06 e TSP\_22), uma vez que esses não possuíam dados informativos sobre a carga proviral desses pacientes;
- Data set para Avaliação da ancestralidade (N=310): criado a partir de um conjunto de dados mesclado do dataset final, mais um conjunto de dados referência, composto por sequências representativas dos três pilares que sustentam a ancestralidade da população brasileira: africanos, europeus e nativos americanos. Este último grupo populacional foi retirado do dataset final pois não foi representado em nossos dados. O dataset referência foi originado do projeto 100 genomas e é composto por 108 genomas africanos, 99 genomas europeus e 59 genomas nativos americanos.

Na avaliação geral do controle de qualidade, foi possível determinar a taxa total de genotipagem das amostras, sendo esta de 0,997001. Obtivemos uma perda de 35.163 SNPs (~6%) quando comparamos com o número total de SNPs presentes nos dados antes e depois do controle de qualidade. Sendo assim, foi possível identificar 515.841 variantes nos 44 indivíduos incluídos no estudo.

### III.5.2.2 Estruturação da população

Para avaliação da diversidade da população estudada, os dados foram submetidos a uma caracterização com base na similaridade genética, para testar a ocorrência de estruturação de subpopulações. Essas estruturações são visíveis entre indivíduos da mesma família. Essas informações precisam ser estudadas uma vez que já foi documentado que a carga proviral do HTLV-1 é de 7 a 16 vezes maior em pacientes com HAM/TSP do que em indivíduos ASS, porém em caso de parentes genéticos com HAM/TSP, a carga proviral é significativamente maior do que em casos sem histórico familiar (SAITO, 2019), ou seja, o traço familiar pode estar associado ao desfecho.

Os dados obtidos revelaram a não ocorrência de estruturação entre os indivíduos, em nenhum dos 8 PCAs testados, independente do grupo de estudo (**Figura 11**). Apesar de não existir um modelo estatístico para definir quantos PCAs são necessários para cada análise, escolhemos avaliar os 8 primeiros PCAs gerados porque já foram suficientes para diferenciar os indivíduos da amostragem e avaliar a diversidade da população. Como a variável analisada nesta matriz foi a informação genética, percebemos que não existe estruturação de indivíduos, dentro deste universo, devido aos seus genótipos. Isso revela que os dados não possuem traços familiares que os unam. Desta forma os PCAs não precisam ser incluídos na análise de regressão logística.

**Figura 11** - Análise de Componentes Principais da população de estudo.

PC (Componente principal)

### III.5.2.3 Ancestralidade Genética

Avaliando a origem étnica dos dados vimos que o número de marcadores informativos sobre ancestralidade, analisado neste estudo, foi suficiente para detectar diferenças na proporção da miscigenação desses indivíduos em relação aos três *clusters* analisados (**Tabela 14**).

**Tabela 14 - Proporção dos Três pilares da Ancestralidade Brasileira na população de estudo.**

<b>Identificação</b>	<b>NAT</b>	<b>AFR</b>	<b>EUR</b>	<b>Ancestralidade predominante</b>
TSP_01	0.040113	0.569766	0.390121	Africana
TSP_12	0.044067	0.591347	0.364587	Africana
TSP_02	0.030352	0.586636	0.383012	Africana
TSP_13	0.086755	0.402993	0.510252	Europeia
TSP_03	0.025896	0.562087	0.412017	Africana
TSP_04	0.069478	0.506483	0.424039	Africana
TSP_15	0.050059	0.485821	0.464120	Africana
TSP_05	0.009780	0.089930	0.900290	Europeia
TSP_16	0.030719	0.799074	0.170206	Africana
TSP_06	0.023742	0.579637	0.396621	Africana
TSP_17	0.072927	0.639254	0.287819	Africana
TSP_07	0.030507	0.428970	0.540523	Europeia
TSP_18	0.086145	0.404824	0.509030	Europeia
TSP_19	0.031169	0.622574	0.346258	Africana
TSP_08	0.132756	0.421052	0.446192	Europeia
TSP_20	0.019021	0.648692	0.332288	Africana
TSP_09	0.063888	0.685066	0.251046	Africana
TSP_21	0.034818	0.365619	0.599563	Europeia
TSP_10	0.044397	0.618637	0.336965	Africana
ASS_24	0.011966	0.826749	0.161285	Africana
TSP_11	0.015002	0.709329	0.275668	Africana
ASS_25	0.033777	0.434071	0.532152	Europeia
ASS_26	0.069419	0.835215	0.095366	Africana
ASS_37	0.026637	0.472819	0.500544	Europeia
ASS_27	0.055173	0.587619	0.357208	Africana
ASS_38	0.017126	0.737513	0.245361	Africana
ASS_28	0.095801	0.504555	0.399644	Africana
ASS_39	0.056347	0.421266	0.522387	Europeia
ASS_29	0.024586	0.812335	0.163080	Africana
ASS_30	0.045477	0.515374	0.439149	Africana
ASS_41	0.049996	0.627222	0.322782	Africana
ASS_31	0.020838	0.493640	0.485522	Africana
ASS_42	0.015235	0.697597	0.287168	Africana
ASS_32	0.080039	0.288330	0.631631	Europeia
ASS_43	0.055698	0.432390	0.511912	Europeia
ASS_33	0.035088	0.658553	0.306359	Africana
ASS_44	0.122388	0.541135	0.336477	Africana
ASS_45	0.077084	0.560472	0.362444	Africana
ASS_34	0.055068	0.662865	0.282067	Africana
ASS_46	0.019664	0.503162	0.477174	Africana
ASS_35	0.027037	0.561771	0.411192	Africana
TSP_22	0.009709	0.343850	0.646441	Europeia
ASS_36	0.109784	0.331195	0.559022	Europeia
TSP_23	0.024575	0.916471	0.058954	Africana

NAT (Nativo Americano), EUR (Europeu), AFR (Africano)

Assim, a determinação ancestral dos 44 indivíduos do estudo revelou uma proporcionalidade de contribuição genética mais marcante para as origens africana e europeia. Os nativos americanos, por sua vez, contribuíram em menor proporção. Foi observado também que marcadores ancestrais africanos se apresentam em maior proporção na maioria dos indivíduos do estudo (**Tabela 14**).

Observando a ancestralidade geral da população deste estudo, foi demonstrado que 70% dos indivíduos (31) são originados predominantemente de matriz africana e 30% (13 indivíduos) de origem predominantemente europeia (**Tabela 14**). Nenhum indivíduo possui ancestralidade definida para nativos americanos, sendo justificado pela baixa prevalência de marcadores genéticos deste grupo na população estudada.

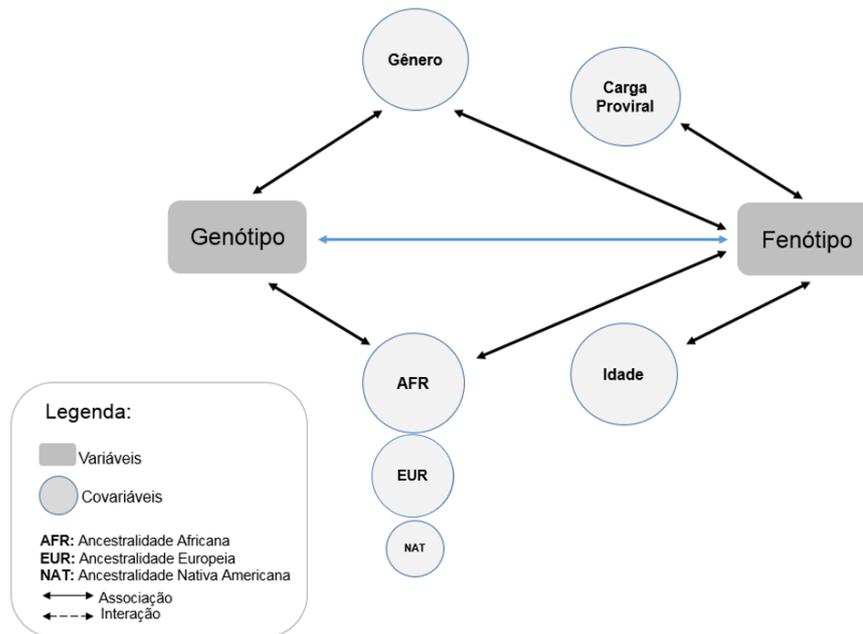
#### III.5.2.4 Genotipagem e Análises Estatísticas

A utilização de um Gráfico Acíclico Direcionado-DAG (Do inglês - *Directed Acyclic Graph*) (**Figura 12**) permitiu realizar uma modelagem causal com o intuito de reforçar a noção de causalidade através da influência, associação e interação entre variáveis e covariáveis (CORTES et al., 2016). Na tentativa de avaliar a associação entre genótipo (SNP) e fenótipo (HAM/TSP), nesse estudo, consideramos o genótipo e o fenótipo como variáveis, e os aspectos sociais e laboratoriais como covariáveis (ancestralidade genética, idade, gênero, etnia autodeclarada e carga proviral).

Observando os dados do estudo, percebemos uma predominância de mulheres no grupo caso. Esse achado corrobora com o que já foi documentado na literatura em relação à presença de um número maior mulheres infectadas pelo HTLV-1 (MOXOTO et al., 2007). A média de idade dos indivíduos infectados é de 59 anos (27-89 anos), sendo que a carga proviral média destes é de  $40.432,0291/10^6$  células (**Tabela 15**). Vale ressaltar que existe uma desproporção entre o número de homens e mulheres no grupo caso e no grupo controle, uma vez que essa variável não foi pareada quando a distribuição dos gêneros nos grupos.

A covariável etnia refere-se à forma como os indivíduos se autodeclaram quanto a sua ancestralidade e foi obtida a partir do questionário de triagem do Centro de HTLV (CHTLV). A maioria se referencia como Negro seguido de Pardo e Branco (39%, 34% e 2% de prevalência, respectivamente). Contudo como existia uma porcentagem elevada de indivíduos que não responderam a esta questão no questionário (25%), essa covariável foi desconsiderada nas análises posteriores (**Tabela 15**).

**Figura 12** - Modelo Teórico Explicativo de associação entre SNPs e a manifestação de HAM/TSP, construído através de Gráfico Acíclico Direcionado (DAG).



**Tabela 15.** Análise Descritiva das Covariáveis nos Grupos Caso e Controle.

Covariáveis\Variáveis		Controle (N=22) N (%)	Caso (N=22) N (%)	Geral	Percentual sem informação	Teste Estatístico
<b>Gênero</b>	Masculino	11 (85%)	2 (15%)	13 (29%)	0%	χ <sup>2</sup>
	Feminino	11 (35%)	20 (65%)	31 (71%)		
<b>Idade (Média)</b>		56	62	59 anos (27-84)	0%	Ttest
<b>Carga proviral* (Mediana)</b>		32.058	42.273	40.432,03	9,10%	Ttest
<b>Etnia</b>	Branco	1 (100%)	0 (0%)	1 (2%)	11 (25%)	Fisher test
	Pardo	6 (40%)	9 (60%)	15 (34%)		
	Negro	9 (53%)	8 (47%)	17 (39%)		
<b>Ancestralidade</b>	EUR	0.38	0.41	13 (30%)	0%	Ttest
	AFR	0.57	0.54	31 (70%)		
	NAT	0.05	0.04	0 (0%)		

\* Em 10<sup>6</sup> células.

Aspectos ancestrais foram incluídos utilizando os dados de ancestralidade genética calculada. A maioria dos indivíduos do estudo possui a ancestralidade africana como predominante (70% dos casos), seguido da europeia (30%) e nenhum

caso de ancestralidade nativa americana (**Tabela 15**). Sobre as referidas covariáveis não foi possível submeter os dados a uma análise bivariada, pela distribuição dos dados ser insuficiente para alcançar poder estatístico.

No que se refere à genotipagem a frequência alélica foi calculada para 515.842 SNPs disponíveis nos dados do exoma dos 44 indivíduos do estudo. Sendo definida para o mesmo número de SNPs independente dos grupos, onde foi observado que a taxa de genotipagem no grupo caso foi de 0.997533. E no grupo controle a frequência alélica foi calculada dentro de uma taxa de genotipagem levemente menor do que no grupo caso: 0.996469.

Dentre os 515.842 SNPs identificados, quatro deles se destacaram quanto às frequências alélicas nos grupos caso e controle, e são eles: rs2857596\_C, rs7917905\_A, rs1265564\_C e rs376863\_A (**Tabela 16**). Nesta avaliação, a tabela de contingência que teve como suporte estatístico o Teste Fisher com correção de Lancaster demonstra os quatro SNPs identificados (rs2857596\_C, rs7917905\_A, rs1265564\_C e rs376863\_A) que foram selecionados quando valor de  $p < 0,01$  no critério de FDR (False Discovery Rate) proposto por Benjamini & Hochberg (1995). Dos quatro SNPs encontrados, dois foram mais fortemente associados à manifestação clínica da HAM/TSP: para o “rs376863” não foi possível identificar uma região gênica que o contenha, enquanto o SNP “rs1265564” já foi previamente caracterizado presente no gene que codifica a proteína CUX-2 e associado a neurodegeneração.

**Tabela 16** – Descrição dos SNPs e as frequências alélicas que mostraram ser diferentes entre os grupos caso e controle.

SNP	Cromossomo	Alelo secundário	Frequência Alélica no grupo Caso	Frequência Alélica no grupo Controle	Alelo Principal	Valor de $p$
rs 1265564	12	C	0.682	0.090	A	6.30 e-09
rs7917905	10	A	0.000	0.500	C	6.90 e-09
rs 376863	15	A	0.182	0.727	C	2.38 e-07
rs 2857596	6	C	0.401	0.000	A	5.69 e-07

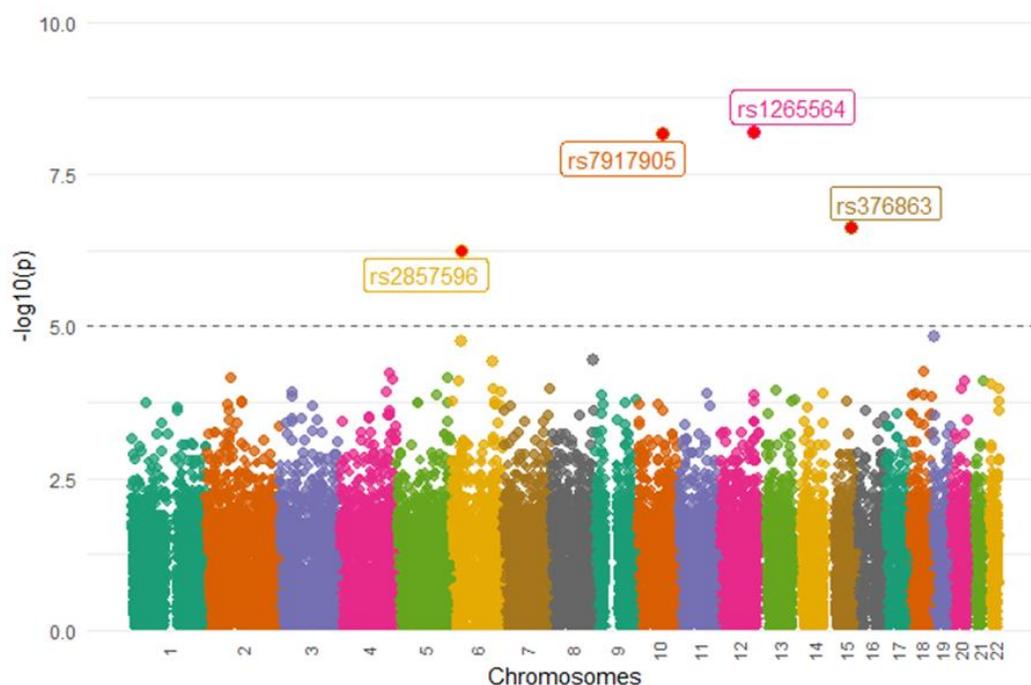
SNP (Single Nucleotide Polimorphism)

Com esses dados foi possível construir o Manhattan Plott dos valores  $p$  de associação do Teste exato de Fisher para as análises de variantes única com  $\text{MAP} \geq 1\%$ . Os cromossomos são identificados no eixo X e no eixo y, o  $-\log^{10}$  do valor

de  $p$ . A linha tracejada representa o limite de significância de todo o genoma de  $5,00 \times 10^{-8}$ , enquanto a linha azul corresponde ao limite sugestivo de significância  $p = 1,00 \times 10^{-5}$ , após a correção de Benjamini & Hochberg para os 515.842 SNPs presentes nos dados. Variantes acima desta linha são consideradas estatisticamente significantes.

Portanto, também através dessa representação é possível perceber a associação dos quatro SNPs referidos anteriormente como associados à HAM/TSP (**Figura 13**).

**Figura 13.** Identificação dos SNPs associados à HAM/TSP através do Manhattan Plott



Identificados os SNPs que revelaram associação com o desfecho (HAM/TSP), submetemos esses dados a uma análise de regressão multivariada. Esta análise permitiu então verificar de forma mais fidedigna se genótipos específicos, mais precisamente SNPs específicos, possuem correlação e influência no fenótipo de estudo e quais possíveis outras variáveis podem interferir nesta relação, a partir da frequência alélica deles nos dados.

Como vimos anteriormente, não foi possível associar as covariáveis incorporadas a este trabalho, com o desfecho da doença, quando analisadas de forma individual. A avaliação do comportamento delas de forma integrada a modelos de regressão linear através de uma análise multivariada possibilitaria uma visão mais eficiente da influência das mesmas com a patologia estudada. Contudo, nossos dados são insuficientes para incorporar todas as covariáveis disponíveis que poderiam exercer influência na relação fenótipo/genótipo como proposto na **Figura 12**, uma vez que o estudo não possui poder amostral para inserir tantas variáveis.

É importante falar também que o não pareamento no número de homens e mulheres entre os grupos, inviabilizou a utilização desta covariável no modelo final. Desta forma, o modelo final foi composto pelo genótipo (SNP), fenótipo e ancestralidade africana.

Desta forma, optamos inicialmente por realizar uma análise bivariada simples, que correlaciona a variável “SNP” no grupo caso e no grupo controle. Pelos dados obtidos, nesta análise, a presença dos SNPs rs7917905 e rs376863 no indivíduo poderia conferir fator de proteção para o desenvolvimento da HAM/TSP, uma vez que os valores de OR são inferiores a 1 (0,01 e 0,08, respectivamente). De forma contrária, os SNPs rs1265564 e rs2857596 se revelam potenciais fatores de risco para o desenvolvimento da HAM/TSP, com valores de OR de 21,43 e 59,34, respectivamente (**Tabela 17**).

**Tabela 17.** Análise bivariada e Multivariada dos genótipos identificados em relação aos grupos Caso e Controle.

SNPs	Análise Bivariada (Modelo não ajustado)			Análise Multivariada (Modelo ajustado)	
	Odds ratio	Valor de <i>p</i>	<i>False Discovery Rate-BH</i> □	Odds ratio***	Valor de <i>p</i> ***
rs1265564	21,43 (6,40 - 71,70)	6,30*10 <sup>-09</sup>	0,00179	4,12 (1,81-9,36)	7,20*10 <sup>-4</sup>
rs7917905	0,01 (6,51 x10 <sup>-4</sup> - 0,19)**	6,90*10 <sup>-09</sup>	0,00179	0,13 (0,03-0,62)	0,01
rs376863	0,08 (0,03—0,23)	2,38*10 <sup>-07</sup>	0,04096	0,31 (0,15-0,63)	1,25*10 <sup>-3</sup>
rs2857596	59,34 (3,43—1026,46)**	5,69*10 <sup>-07</sup>	0,0734	5,28 (1,18-23,80)	0,03

Total de SNPs testados: 515.841;(\*\*) Odds ratio calculado através da correção de Haldane-Ascombe; \*\*\*Regressão logística de Firth ajustada para ancestralidade; SNP (*Single Nucleotide Polimorphism*).

Posteriormente em uma análise multivariada, o modelo foi ajustado apenas pela ancestralidade dos indivíduos, por ser uma variável tradicionalmente utilizadas em modelos de regressão (Conceito epidemiológico) (**Tabela 17**). Os dados foram testados para os três *clusters* de ancestralidade: africana (AFR), europeia (EU) e nativa americana (NAT). Contudo, utilizamos no modelo final da regressão logística multivariada apenas a ancestralidade africana, por ser a ancestralidade mais representada nos dados.

Quando avaliamos os quatro SNPs na análise Bivariada (Modelo não ajustado) foi possível sugerir uma possível associação deles com a HAM/TSP, porém quando aplicamos o Modelo ajustado por ancestralidade na análise de regressão logística multivariada vemos que apenas os SNPs “rs1265564” e “rs376863” se apresentam associados ao desfecho (**Tabela 17**). Desta forma, sugere-se que o efeito de associação dos outros dois SNPs (rs7917905 e rs2857596) seja decorrente do efeito genético de ancestralidade.

#### **III.5.2.5 Avaliação do desequilíbrio de ligação dos SNPs**

Quando em desequilíbrio de ligação (DL) os alelos específicos em locais próximos podem co-ocorrer no mesmo haplótipo com frequência maior do que o esperado por acaso. O DL é importante no mapeamento genético porque é usado na clonagem posicional para rastrear a variação que produziu um sinal de ligação e para estudos de associação nos quais variantes de doenças podem ser detectadas através da presença de associação nas proximidades (WALL & PRITCHARD, 2003).

**Tabela 18** – Avaliação do desequilíbrio de ligação dos SNPs encontrados em associação com a HAM/TSP.

SNPs	População Referência	Frequência Ancestral	SNPs em DL	Média $r^2$	Dados da comparação
<b>Rs376863 (alelo A)</b>	AFR	0.2345	10	0.977	641 SNPs/641 testes
	EUR	0.4642	11	0.968	276 SNPs/276 testes
	AMR	0.4496	8	0.983	505 SNPs/505 testes
<b>Rs1265564 (alelo C)</b>	AFR	0.01362	2	0.974	443 SNPs/443 testes
	EUR	0.4314	1	0.968	162 SNPs/162 testes
	AMR	0.2161	NA		
<b>Rs2857596 (alelo C)</b>	AFR	0.2126	3	0.974	403 SNPs/403 testes
	EUR	0.2535	5	0.976	256 SNPs/256 testes
	AMR	0.1715	5	0.981	418 SNPs/418 testes
<b>Rs7917905 (alelo A)</b>	AFR	0.1513	20	0.976	421 SNPs/421 testes
	EUR	0.07753	51	0.989	226 SNPs/226 testes
	AMR	0.04035	49	0.979	424 SNPs/424 testes

SNP (*Single Nucleotide Polimorphism*); AFR (Africano), EUR (Europeu), AMR (População Miscigenada); NA (Não Avaliada/encontrada); DL (Desequilíbrio de ligação).

Também observamos que dos quatro SNPs identificados, três deles são mais comuns em populações europeias, sendo dois deles os de significância estatística (**Tabela 18**). O SNP “rs1265564”, que se apresentou como mais associado ao desenvolvimento de HAM/TSP nos nossos dados, se apresenta em desequilíbrio de ligação com outros três SNPs. Levando em consideração a população africana: para as populações africana e europeia se apresenta em desequilíbrio com ele mesmo (rs1265564) e apenas para população africana para o SNP “rs1265565”, todos pertencentes ao cromossomo 12. Não foi encontrado na literatura associação do SNP “rs1265565” com aspectos fisiológicos ou patogênicos.

### III.5.2.6 Levantamento Bibliográfico: busca por SNPs de susceptibilidade ou associação com a HAM/TSP

A HAM/TSP vem sendo alvo de alguns estudos moleculares quando se trata da associação genótipo/fenótipo de indivíduos infectados pelo HTLV-1. Seis estudos triados demonstraram associação, correlação ou interação com o desfecho (**Tabela 19**). Nove dos onze SNPs relatados possuíam nomenclatura “rs” disponível, e estes foram investigados no nosso conjunto de dados do exoma que dispunhamos. Das nove variações pesquisadas apenas os SNPs “rs3138045” e “rs12979860” estão presentes nos 44 indivíduos do estudo.

**Tabela 19** – Relação dos SNPs, já descritos na literatura, com associação à HAM/TSP.

SNP	Região genômica	Referência	Outras Informações
rs3138053 rs2233406 rs3138045	Região promotora do gene NFKBIA	TALLEDO, 2012	
rs12979860	Gene IL-28B	VALLINOTO, 2015	
rs8099917	Gene IL-28B	ASSONE, 2014	Genótipo GG
Ausente	Gene IL-6	GADELLA, 2008	Alelo G/174 Alelo C/634
rs1800682	Gene FAS-670		Alelos A/G
rs2234767	Gene FAS-1377	VALLIONOTO et al. 2012	Alelos G/A
rs763110	Gene FAS-844		Alelos C/T
rs1051070	Gene FAS-670	ROSADO, 2017	Genótipo A/A

SNP (*Single Nucleotide Polimorphism*)

Fazendo uma busca direcionada da frequência dos dois SNPs já documentados na literatura e que se apresentam nos dados do estudo (rs3138045 e rs12979860), observamos frequências alélicas maiores de ambos SNPs no grupo caso (**Tabela 20**). Apesar da evidência sobre a maior frequência destes SNPs nos indivíduos que manifestaram a HAM/TSP, eles não aparecem na análise de regressão logística realizada, o que poderia ser explicado pelo número de indivíduos avaliados neste estudo.

**Tabela 20** - Frequência alélica da análise direcionada por SNPs da literatura.

SNP	Região genômica	Alelos	Frequência no grupo Caso	Frequência no grupo Controle	Genoma referência
rs3138045	Região promotora do gene NFKBIA	G/A	0.25	0.2273	GRCh37/hg19
rs12979860	Gene IL-28B	C/T	0.4524	0.4091	GRCh37/hg19

SNP (*Single Nucleotide Polimorphism*)

### III.6 Discussão

Doenças infecciosas exercem uma pressão genética seletiva significativa sobre o hospedeiro. Desta forma, mutações e variações genéticas específicas têm sido fortemente implicadas por conferir suscetibilidade ou resistência a diferentes patologias e/ou infecções (FORNI et al., 2013; KLEBANOV, 2018). Mais especificamente, estudos genômicos têm demonstrado fortes associações entre polimorfismos em genes codificantes de moléculas de HLA com suscetibilidade a doenças infecciosas, reforçando a importância biológica das variações genômicas sobre os desfechos clínicos.

Variações genéticas associadas à resistência a infecções específicas frequentemente envolvem moléculas receptoras utilizadas na entrada do vírus, coreceptores ou enzimas modificadoras de receptores. Da mesma forma, polimorfismos que levam à super ou subprodução de citocinas específicas também podem influenciar a gravidade da doença viral. Defeitos genéticos em outros aspectos das respostas imunes celulares inatas e adaptativas a infecções virais, como detecção do vírus, sinalização em resposta a infecção, atividade de fatores de restrição antivirais ou início adequado de respostas de células T, também foram associados com maior gravidade de inúmeras infecções virais (KENNEY et al., 2017).

Como já citado anteriormente, existem alguns SNPs em genes do hospedeiro já indicados como possíveis marcadores da HAM/TSP. O SNP “rs3138045”, por exemplo, está presente na região promotora do gene NFKBIA que é responsável por codificar o Fator Nuclear kappa beta (NF- $\kappa$ B). Esse promove a expressão de citocina/quimiocina e conseqüentemente a proliferação de vários tipos de células imunes (COTO-SEGURA, 2019). Posadas e cols. (2019), revelaram alta associação de SNPs na região *upstream* deste gene à doença para indivíduos em Lima, no Peru. O SNP “rs12979860” por sua vez, está localizado no gene IL-28B. A interleucina-28B, pertence à família de citocinas-interferon  $\lambda$ . Estudos revelam que essa citocina inibe a replicação viral (VALLINOTO, 2015). Treviño e cols. (2012) retratam associação entre polimorfismos da IL-28B e o desfecho HAM/TSP. Em contrapartida em uma análise mais pontual, com a investigação do do mesmo SNP “rs12979860”, Vallinoto e cols. (2015), não o associam com a HAM/TSP em indivíduos no Pará.

No entanto, até o momento este é o primeiro plano piloto, de nosso conhecimento, de rastreamento do genoma do hospedeiro para a busca de SNPs que

possam estar associados a manifestação de doença. Portanto, sugerimos que essa estratégia de busca extensiva por marcadores no hospedeiro seja replicada em uma coorte mais numerosa, uma vez que este tipo de dado molecular pode ser útil no entendimento da HAM/TSP.

Isso porque, estudos de ampla associação genômica requer três elementos essenciais: (1) suficientemente grandes amostras de estudo de populações que efetivamente forneçam informações genéticas sobre a questão de pesquisa, (2) alelos polimórficos que possam ser genotipados de forma eficiente e cobrir todo o genoma adequadamente, e (3) métodos analíticos estatisticamente poderosos que possam ser empregados para identificar as associações genéticas de uma forma imparcial (CANTOR et al., 2010).

Sobre a caracterização da população desse estudo, a ancestralidade predominantemente Africana dos nossos dados corrobora com o demonstrado pelo projeto EPIGEN Brasil, onde foi visto o parentesco e a consanguinidade interagindo em três populações no Brasil. A região Nordeste apresenta ascendência africana em 50% dos casos, enquanto as regiões Sudeste e Sul apresentam ascendência europeia superior a 70% (KEHDY, 2015).

Como já descrito, a população brasileira é o produto da mistura pós-colombiana entre ameríndios, colonizadores ou imigrantes europeus e escravos africanos (SALZANO, 1967; KEHDY, 2015). Sendo assim, os três *clusters* utilizados como referência na comparação são os três principais pilares da ancestralidade que melhor justificam a origem da população brasileira, segundo o projeto denominado Epigen Brasil. Além de que esses três *clusters* são muito bem documentados no projeto 1000 genomas.

Sendo assim, é importante observar que simulações indicam que o limite de significância do genoma amplamente usado de  $5 \times 10^{-8}$  para estudos sobre populações europeias, controla adequadamente o número de SNPs independentes em todo o genoma, independentemente da densidade de SNPs do estudo (DUDBRIDGE & GUSNANTO, 2008). No entanto, ao testar populações africanas, limites mais rigorosos são necessários devido a maior diversidade genética entre esses indivíduos (MAREES et al., 2018). Deste modo, estudos semelhantes conduzidos em nossa população devem levar em conta essa maior diversidade genética e, portanto, construir uma amostra populacional com o devido tamanho para sofrer as diversas validações estatísticas de forma satisfatória.

Outra característica marcante observada na população de estudo, foi a predominância de mulheres no grupo caso. Esse achado corrobora com os da literatura, onde Romanelli e cols. (2013) descrevem os dados demográficos de 181 indivíduos soropositivos para HTLV-1 com definição para manifestação clínica da HAM/TSP, e associaram esse status a uma maior frequência em mulheres (ROMANELLI et al., 2013), especialmente na faixa etária entre os 40 e 50 anos de idade e com provável infecção por via sexual (PINEDA et al., 2019).

Infelizmente não conseguimos fazer uma avaliação da influência das covariáveis (Idade, gênero e carga proviral), sobre o desfecho da HAM/TSP, devido à distribuição muito heterogênea dos dados em nossa amostra que foi de conveniência. O ideal para estudos futuros é de fato que a população de estudo já tenha uma distribuição numérica nos grupos bem representada em relação às covariáveis, para que os dados possam ser analisados e tenham poder estatístico para conclusões mais acertadas e confiáveis.

Devido à nossa limitação amostral, decidimos aplicar critérios estatísticos mais rígidos, como por exemplo, menor significância do valor de  $p$ . Além disso, é sabido que três alternativas são amplamente aplicadas para determinar o significado biológico de um genoma, e são eles: - uso da correção de Bonferroni; Benjamini – Hochberg, através do cálculo de Taxa de Falsa Descoberta (FDR); - e, teste de permutação. A correção de Bonferroni, que visa controlar a probabilidade de haver pelo menos um achado falso positivo, calcula o limite do valor  $p$  ajustado com a fórmula  $0,05/n$ , (n sendo o número de SNPs testados). No entanto, como muitos SNPs são correlacionados, devido ao desequilíbrio de ligação-LD (Do inglês, *Linkage Disequilibrium*) e, portanto, por definição, não são independentes, este método é muitas vezes muito conservador e leva a um aumento na proporção de resultados falsos negativos (MAREES et al., 2018).

Apesar de não termos verificado a influência dessas covariáveis sobre a HAM/TSP, nesta população, isso já é sugerido na literatura (**Tabela 21**) e, portanto, merece ser melhor investigada em estudos de associação genômica ampla, como este estudo piloto, em coortes mais numerosas de indivíduos infectados, uma vez que investigações como essas podem ser bastante beneficiadas por este tipo de rastreamento genômico.

**Tabela 21.** Breve levantamento bibliográfico sobre os possíveis efeitos das covariáveis (Idade, gênero e carga proviral) sobre desenvolvimento da HAM/TSP.

Variáveis	Associação com fenótipo e/ou doença	Referências
Idade	A variável idade como modificador de efeito (HAM/TSP - Depressão)	BOA-SORTE, 2015
	Média de idade para início da HAM/TSP (40-50 anos de vida)	GESSAIN, 2012
	Relatou a quarta década de vida como início da HAM/TSP	VALLINOTO, 2019
Gênero	Relatou que a HAM/TSP é mais comum em mulheres ( $p=0,0139$ )	PINEDA, 2019
	Demonstrou que a HAM/TSP acomete 2 mulheres a cada homem	VARANDAS, 2018
Carga proviral	Carga proviral (CP) em indivíduos sintomáticos é mais elevada que em pacientes ASS. Pacientes com HAM/TSP com CP=5,03 log10 cópias de HTLV-1/10 <sup>6</sup> PBMCs	PINEDA, 2019
	Forte correlação com a gravidade da doença (HAM/TSP) ( $p=0,05$ )	DERAKHSHAN, 2019
	O genótipo FAZ-670 está associado à alta carga proviral em casos de HAM/TSP	VALLINOTO, 2012
	O genótipo FAZ-670 AA está associado à alta carga proviral em casos de HAM/TSP	ROSADO, 2017

Dos pouco mais de quinhentos mil SNPs testados neste trabalho, apenas 4 deles apresentaram diferença estatística na ocorrência entre os grupos caso e controle e mostraram estar associados com o desfecho. Numa análise mais criteriosa, ajustando os dados com a ancestralidade, esse efeito apenas permaneceu em dois dos quatro SNPs.

Interessantemente, um deles o rs1265564 está localizado no gene que codifica a proteína CUX-2. CUX-2 exerce um importante papel na especificação de células neuronais e por isso, continua a ser expressa em neurônios pós-mitóticos. CUX-2 e CUX-1 pertencem a uma classe de proteínas que apresenta em sua estrutura três regiões altamente semelhantes conhecidas como pontos de corte, 1, 2 e 3 (CUBELOS et al., 2010; LI et al., 2010). Recentemente, foi atribuída a CUX-1 a função de estimular a 8-oxoguanina DNA glicosilase 1 (OGG1), uma enzima que remove os resíduos oxidados das purinas do DNA e introduz uma ruptura de fita única, através de sua atividade apurínica/apirimidínica liase para iniciar o reparo do DNA (RAMDZAN et al., 2014; RAMDZAN et al., 2015). Posteriormente, foi demonstrado que, *in vitro*, as repetições de corte de CUX2 aumentaram a ligação de OGG1 a 7,8-dihidro-8-oxoguanina no DNA e estimulou as atividades glicosilase e liase de OGG1. Esses resultados demonstram que CUX2 funciona como um fator acessório que estimula o reparo de dano oxidativo ocorrido no DNA. Os neurônios de fato produzem um alto nível de espécies reativas de oxigênio devido à sua dependência da oxidação aeróbia da glicose como sua fonte de energia. As espécies reativas de oxigênio (ROS) geradas pelo metabolismo celular representam uma grande ameaça à integridade do DNA. Este acúmulo é considerado uma possível causa para a perda progressiva da

função dos neurônios associada ao envelhecimento e, portanto, potencialmente associa a deficiência do reparo da base oxidada, à neurodegeneração relacionada à idade. Dessa forma, a expressão persistente de CUX2 em neurônios pós-mitóticos contribui para a manutenção da integridade do genoma por meio de sua estimulação do reparo de danos oxidativos ao DNA (RANJANA et al., 2015).

Portanto, esse SNP pode de fato representar um marcador importante para a manifestação da HAM/TSP, que também é caracterizada por uma síndrome cujos sinais clínicos estão associados à neurodegeneração.

Um substancial número de estudos de ampla associação genômica indicam, recentemente que, para a maioria dos distúrbios, apenas algumas variantes comuns estão implicadas e os SNPs associados explicam apenas uma pequena fração do risco genético. Assim, a maioria dos SNPs associados estão em desequilíbrio de ligação com variantes predisponentes que aumentam a doença dos portadores riscos entre 10% e 30% sobre o risco em não portadores (CANTOR et al., 2010).

Estudos de ampla associação genômica são baseados no princípio do desequilíbrio de ligação (LD) ao nível da população. Desequilíbrio de ligação é a associação não aleatória entre alelos em diferentes *loci* e é criado por forças evolutivas, como mutação, deriva e seleção. Quanto maior o tamanho (efetivo) da população, mais fraco o LD para uma determinada distância (VISSCHER et al., 2012).

Em nossos resultados os quatro SNPs identificados mostraram-se em desequilíbrio de ligação com outros SNPs, cujas funções biológicas precisarão ser melhor investigados para esclarecer o significado real desse desequilíbrio. Preliminarmente, detectamos que os SNPs já descritos na literatura e buscados em nossa coorte não estão entre tais referidos SNPs citados na análise de desequilíbrio de ligação.

#### IV. CONCLUSÃO

A realização desse trabalho nos permitiu concluir não haver um perfil de apresentação de epítomos que esteja associado ao status clínico do hospedeiro, uma vez que houve grande semelhança quanto aos alelos de HLA utilizados e quanto à posição dos epítomos apresentados entre os “alelos protéticos”. No entanto, foi possível identificar variações moleculares importantes na predição dos epítomos especialmente nos genomas de indivíduos com alguma sintomatologia: ATLL e HAM/TSP. Tais observações dizem respeito à frequência de apresentação dos epítomos e à afinidade de ligação dos mesmos às moléculas de HLA. Por isso, sugerimos que outras estratégias de investigação *in silico* e *in vitro* sejam desenhadas para verificar se estas mudanças na apresentação de epítomos lineares pode ter repercussões funcionais e contribuir com o desfecho clínico.

Quanto à investigação do exoma do hospedeiro, foi possível identificar quatro SNPs associados com a HAM/TSP, sendo dois deles com associação mais forte. Para um deles, o rs376863, não identificamos que ele esteja presente em região gênica. No entanto, o SNP rs1265564, já foi caracterizado como SNP do gene que codifica para a proteína CUX-2 que está associada ao reparo de DNA oxidado por ação das espécies reativas do oxigênio produzidas nos neurônios, levando estas células à neurodegeneração. Os quatro SNPs iniciais mostram estar em desequilíbrio de ligação com outros SNPs, nenhum deles dos já descritos na literatura e citados neste trabalho. Por isso, sugerimos uma busca mais minuciosa dos SNPs indicados neste trabalho, bem como a associação de marcadores moleculares do hospedeiro com marcadores virais, como forma de entender o processo de infecção e manifestação de doença como resultado de uma integração complexa e bem sucedida.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALVAREZ, C. et al. Family Aggregation of Human T-Lymphotropic Virus 1-Associated Diseases: A Systematic Review. *Front Microbiol.* 2016 Oct 28;7:1674. **eCollection**, 2016.

ALESSIO, L. et al. Low prevalence of HTLV1/2 infection in a population of immigrants living in southern Italy. *PLOS Neglected Tropical Diseases* / <https://doi.org/10.1371/journal.pntd.0006601>, 2018.

ANUPAM, B; et al. The need to accessorize: molecular roles of HTLV-1 p30 and HTLV-2 p28 accessory proteins in the viral life cycle. *Frontiers in microbiology*. doi: 10.3389/fmicb.2013.00275, 2013.

ARAUJO AQC, SILVA MTT. The HTLV-1 neurological complex. *Lancet Neurol.*; 5:1068-76, 2006.

ASSONE, T. et al. IL28B gene polymorphism SNP rs8099917 genotype GG is associated with HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) in HTLV-1 carriers. *PLoS neglected tropical diseases*, v. 8, n. 9, p. e3199, 2014.

AYE, M. et al. Histopathological analysis of four autopsy cases of HTLV-1-associated myelopathy/tropical spastic paraparesis: inflammatory changes occur simultaneously in the entire central nervous system. *Acta Neurophthol*, 100:245-52, 2000.

AZODI S. et al. Imaging spinal cord atrophy in progressive myelopathies: HTLV-1-associated neurological disease HAM/TSP and multiple sclerosis. *Ann Neurol*, 82, 719-28, 2017.

BANGHAM, CR. The Immune control and cell-to-cell spread of human T-lymphotropic virus type 1. *J Gen Virol* 84, 3177-3189, 2003.

BANGHAM C.R.M. et al. HTLV-1-associated myelopathy/tropical spastic paraparesis. *Nature Reviews Disease Primers*. 1(1):15012, 2015.

BANGHAM, C.R. & OSAME, M. Cellular immune response to HTLV-1. *Oncogene*; 24: 6035-46, 2005.

BANGHAM C R M, MATSUOKA, M. Human T-cell leukaemia virus type 1: parasitism and pathogenesis. *Philos Trans R Soc Lond B Biol Sci.*, Oct

19;372(1732):20160272, 2017.

BARRETT JC.et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005.

BARRETO, F. K. et al. Analyses of HTLV-1 sequences suggest interaction between ORF-I mutations and HAM/TSP outcome. *Infect Genet Evol. Nov; 45: 420–425. Aug 21. doi: 10.1016/j.meegid.2016.08.020*, 2016.

BARTOE, JT; ALBRECHT, B; COLLINS, ND; ROBEK, MD; RATNER, L; GREEN, PL; LAIRMORE, MD. Functional role of pX open reading frame II of human T-lymphotropic virus type 1 in maintenance of viral loads in vivo. *J Virol 74, 1094-1100*, 2000.

BASBOUS, J.et al. HBZ factor of HTLV-I dimerizes with transcription factors JunB and c-Jun and modulates their transcriptional activity. *J. Biol. Chem. 278, 43620-43627*, 2003.

BATTINI, J.L. et al. The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV. *Cell 115, 449-459*, 2003.

BENJAMINI Y & HOCHBERG Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), Vol. 57, No. 1, pp. 289-300*, 1995.

BITTENCOURT, AL.et al. Adult T-Cell Leukemia/Lymphoma in Bahia, Brazil. Analysis of Prognostic Factors in a Group of 70 Patients. *Am J Clin Pathol 128, 875-882*, 2007.

CANTOR, RM.et al. Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application. *The American Journal of Human Genetics 86, 6–22, January 8*, 2010.

CARNEIRO-PROIETTI, A.B. F. et al. Infecção e doença pelos vírus linfotrópicos humanos de células T (HTLVI/II) no Brasil. *Rev da Soc Bras Med Trop; 35: 499-508*, 2002.

CARNEIRO-PROIETTI, A.B.F. et al. Human T-lymphotropic virus type 1 and type 2 seroprevalence, incidence, and residual transfusion risk among blood donors in Brazil during 2007–2009. *AIDS Res Hum Retroviruses. 28:1265–72*, 2012.

CATALAN-SOARES, B. et al. Human T-cell lymphotropic virus in family members

of seropositive blood donors: silent dissemination. *Rev Panam Salud Publica*; **16**: 387-94, 2004.

CATALAN-SOARES, B. et al. 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. *Cad Saude Publica*; **21**: 926-31, 2005.

CATALAN-SOARES, B. C. et al. HLA class I alleles in HTLV-1-associated myelopathy and asymptomatic carriers from the Brazilian cohort GIPH. *Medical microbiology and immunology*, v. **198**, n. **1**, p. 1-3, 2009.

CIANCIOLO G.J, et al. Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science*. **230(4724):453-5**. *Epub 1985/10/25. PMID: 2996136*, 1985.

COOK, L.B.; et al. HTLV-1: Persistence and pathogenesis. *J. Virol. [s.l.]*, v. **435**, n. **1**, p.131-140, jan. 2013.

COOK, L. et al. The impact of HTLV-1 on the cellular genome. *Current Opinion in Virology* **2017**, **26**:125-131, 2017.

COSKUN, A.K. & SUTTON, R.E. Expression of glucose transporter 1 confers susceptibility to human T-cell leukemia virus envelope-mediated fusion. *J Virol* **79(7)**, 4150-4158, 2005.

COTO-SEGURA P, COTO E, GONZÁLEZ-LARA L, ALONSO B, GÓMEZ J, CUESTA-LLAVONA E, QUEIRO R. Gene Variant in the NF- $\kappa$ B Pathway Inhibitor NFKBIA Distinguishes Patients with Psoriatic Arthritis within the Spectrum of Psoriatic Disease. *Biomed Res Int*. **Nov 11**; **2019:1030256**, 2019.

CRAVOIS M, GESSAIN A, GOUT O, WAIN-HOSBON S, WATTEL E. Common Human T-cell leukemia virus Type 1 (HTLV-1) integration sites in cerebrospinal fluid and blood lymphocytes of patients with HTLV-1-associated-myelopathy/tropical spastic paraparesis indicate that HTLV-1 crosses the blood brain barrier via oral clonal HTLV-1-infected cells. *J Infect. Dis*. **182 (4)**, 1044-50, 2000.

CUBELOS, B.et al. Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. *Neuron* **66**, 523-535, 2010.

CURRER, R.et al. HTLV tax: a fascinating multifunctional co-regulator of viral and

cellular pathways. *Front. Microbiol.*, v. 3, n. 406, p. 1 – 24, nov. 2012.

DENDROU, C. A. et al. HLA variation and disease. *Nature Reviews Immunology*, v. 18, n. 5, p. 325, 2018.

DING, W. et al. Endoplasmic Reticulum and *cis*-Golgi Localization of Human T-Lymphotropic Virus Type 1 p12<sup>I</sup>: Association with Calreticulin and Calnexin. *J Virol* 75(16), 7672–7682, 2001.

DOURADO, I. et al. HTLV-I in the general population of Salvador, Brazil: a city with African ethnic and sociodemographic characteristics. *J Acquir Immune Defic Syndr*; 34: 527-31, 2003.

DUDBRIDGE F & GUSNANTO A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol, Apr;32(3):227-34*, 2008.

EDWARDS, D. et al. Orf-I and Orf-II-Encoded Proteins in HTLV-1 Infection and Persistence. *Viruses, [s.l.]*, v. 3, n. 6, p. 861-885, 2011.

ELOVAARA, I.; et al. High human T cell lymphotropic virus type 1 (HTLV-1)-specific precursor cytotoxic T lymphocyte frequencies in patients with HTLV-1-associated neurological disease. *J Exp Med* 177, 1567– 1573, 1993.

ENOSE-AKAHATA Y & JACOBSON S. Immunovirological markers in HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP). *Retrovirology*, 16:35, 2019

FARRE L. et al. Fas 670 promoter polymorphism is associated to susceptibility, clinical presentation, and survival in adult T cell leukemia. *J Leukoc Biol. Jan;83(1):220-2*, 2008.

FORNI, D. et al. A 175 million year history of T cell regulatory molecules reveals widespread selection, with adaptive evolution of disease alleles. *Immunity. Jun 27;38(6):1129-41*, 2013.

FRANCHINI, G. Molecular mechanisms of human T-cell leukemia/lymphotropic virus type I infection. *Blood*; 86: 3619-39, 1995.

FRANGIONE-BEEBE, M, ALBRECHT, B, DAKAPPAGARI, ROSE, RT, BROOKS, CL, SCHWENDEMAN S P, LAIRMORE, M D, KAUMAYA P T. Enhanced immunogenicity of a conformational epitope of human Tlymphotropic virus type 1 using a novel chimeric peptide, *Vaccine* 19, 1068-1081, 2000.

FRENZEL, et al. HTLV-1-associated arthropathy treated with anti-TNF-alpha agent. **Joint Bone Spine. Vol 81, Issue 4**, 2014.

FUJIMORI J. et al. Epitope analysis of the cerebrospinal fluid IgG in HTLV-I associated myelopathy patients using phage display method. **J Neuroimmunol.; 152:140-6**, 2004.

FUKUMOTO R, ANDRESEN V, BIALUK I, CECCHINATO V, WALSER JC, VALERI VW, NAUROTH JM, GESSAIN A, NICOT C, FRANCHINI G. In vivo genetic mutations define predominant functions of the human T-cell leukemia/lymphoma virus p12I protein. **Blood 113:3726-3734**, 2009.

GADELHA, S.R. et al. Correlation between polymorphisms at interleukin-6 but not at interleukin-10 promoter and the risk of human T lymphotropic virus type I-associated myelopathy/tropical spastic paraparesis in Brazilian individuals. **Journal of medical virology, v. 80, n. 12, p. 2141-2146**, 2008.

GESSAIN, A. et al. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. **Lancet 2, 407-410**, 1985.

GESSAIN, A. et al. Chronic myelopathy associated with human T-lymphotropic virus type I (HTLV-I). **Ann Intern Med; 117: 933-46**, 1992.

GESSAIN, A. et al. What is the situation of human T cell lymphotropic virus type II (HTLV-II) in Africa? Origin and dissemination of genomic subtypes. **J Acquir Immune Defic Syndr Hum Retrovirol; 13: 228-35**, 1996.

GESSAIN A & CASSAR O. Epidemiological Aspects and World Distribution of HTLV-1 Infection. **Frontiers in Microbiology**. 2012.

GHEZ, D; et al. Neuropilin-1 is involved in Human T-cell Lymphotropic virus type 1 entry. **J Virol 80, 6844-6854**, 2006.

GHEZELDASHT, S.A. et al. Human T Lymphotropic Virus Type I (HTLV-I) Oncogenesis: Molecular Aspects of Virus and Host Interactions in Pathogenesis of Adult T cell Leukemia/Lymphoma (ATL). **Iran J Basic Med Sci. Mar; 16(3): 179-195.**, 2013.

GOURRAUD P. A. et al. Aggregation of multiple sclerosis genetic risk variants in multiple and single case families. **Ann. Neurol. 69, 65-74**, 2011.

GREEN, P.L. et al. Human T-cell leukemia virus type II nucleotide sequences

between env and the last exon of tax/rex are not required for viral replication or cellular transformation. *J Virol*; **69**: 387-94, 2001.

HALL, T. BioEdit: an important software for molecular biology. *GERF Bulletin of Biosciences*. ;**2(1)**:60-61, 2011.

HANABUCHI, S. et al. Development of Human T-Cell Leukemia Virus Type 1-Transformed Tumors in Rats following Suppression of T-Cell Immunity by CD80 and CD86 Blockade. *J. Virol.*, **0022-538X/00/\$04.0010**. p. 428-435, 2000.

HANON, E. et al. Fratricide among CD8(+) T lymphocytes naturally infected with human T cell lymphotropic virus type I. *Immunity* **13**, 657-664, 2000.

HARRIS, N. L.; RONCHESE, F. The role of B7 costimulation in T-cell immunity. *Immunol Cell Biol. Aug*; **77(4)**:304-11, 1999.

HARROD, R. Silencers of HTLV-1 and HTLV-2: the pX-encoded latency-maintenance factors. *Retrovirology*, **6 de set**; **16 (1)**: 25. doi: 10.1186/s12977-019-0487-9, 2019.

HILBURN S. et al. In vivo expression of human T-lymphotropic virus type 1 basic leucine-zipper protein generates specific CD8+ and CD4+ T-lymphocyte responses that correlate with clinical outcome. *J Infect Dis.*; **203**:529-36, 2011.

HINO, S. et al. Mother to child transmission of human T-cell leukemia virus type 1. *Jpn J Cancer Res*; **76**: 474-80, 1985.

ILINSKAYA, A. et al. Opposing Effects of a Tyrosine-based sorting motif and a PDZ-Binding motif regulate Human T-Lymphotropic Virus type 1 envelope Trafficking. *J of Virology* **84**, 6995-7004, 2010.

ISHAK R, et al. The challenge of describing the **epidemiology** of **HTLV** in the Amazon region of Brazil. *Retrovirology. Feb 14*; **17(1)**:4. doi: 10.1186/s12977-020-0512-z, 2020.

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.

JOHNSON, JM. et al. Molecular biology and pathogenesis of the human T-cell

leukaemia/lymphotropic virus Type-1 (HTLV-1). *Int J Exp Pathol* **82**, 135-147, 2001.

JONES, K.S. et al. Molecular Aspects of HTLV-1 Entry: Functional Domains of the HTLV-1 Surface Subunit (SU) and Their Relationships to the Entry Receptors. *Viruses* **3**, 794-810, 2011.

KABIRI, M. et al. Robust mucosal and systemic responses against HTLV-1 by delivery of multi-epitope vaccine in PLGA nanoparticles, *European Journal of Pharmaceutics and Biopharmaceutics*, 2018.

KAMOIK & MOCHIZUKI M. HTLV-1 uveitis. *Front Microbiol.* **3**:270, 2012.

KANNAGI, M. et al. Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD81 cytotoxic T cells directed against HTLV-I-infected cells. *Int. Immunol.* **3**:761– 767, 1991.

KANNAGI, M. et al. Impact of host immunity on HTLV-1 pathogenesis: potential of Tax-targeted immunotherapy against ATL. *Retrovirologia*; **16 (1): 23**. doi: **10.1186 / s12977-019-0484-z**, 2019.

KANZAKI, L.I.B. HTLV-1: A Real Pathogen or a Runaway Guest of a Diseased Cell? *J Biosci, Sep; 43 (4): 785-795*, 2018.

KATSUYA, H. et al. The Nature of the HTLV-1 Provirus in Naturally Infected Individuals Analyzed by the Viral DNA-Capture-Seq Approach. *Cell Reports* **29**, 724–735, Oct 15, 2019

KENNEY A. et al. Yount Human Genetic Determinants of Viral Diseases. *Annu Rev Genet. November 27; 51: 241–263*, 2017.

KIM SJ, DING W, ALBRECHT B, GREEN PL, LAIRMORE MD. A conserved calcineurin-binding motif in human T lymphotropic virus type 1 p12I functions to modulate nuclear factor of activated T cell activation. *J.Biol.Chem.***278**:15550–15557, 2003.

KEHDY FSG et. al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci U S A. Jul 14; 112(28): 8696–8701*, 2015.

KLEBANOV N. The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. Genetic Predisposition to Infectious Disease. *Cureus* **10(8): e3210. DOI 10.7759/cureus.3210**, 2018.

KNIGHT, S.C. et al. Dendritic cells in HIV-1 and HTLV-1 infection. **Adv. Exp. Med. Biol.** **329**, 545-549, 1993.

KOBE, B; CENTER, RJ; KEMP, BE; POUMBOURIOS, P. Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. **Proc. Natl. Acad. Sci. USA** **96**, 4319-4324, 1999.

KORALNIK, IJ. et al. The p12I, p13II, and p30II proteins encoded by human T-cell leukemia/lymphotropic virus type I open reading frames I and II are localized in three different cellular compartments. **J Virol** **67**, 2360-2366, 1993.

KOSAKO T. et al. 2011. Target Epitopes of HTLV-1 Recognized by Class I MHC-Restricted Cytotoxic T Lymphocytes in Patients With Myelopathy and Spastic Paraparesis and Infected Patients With Autoimmune Disorders. **J Medical Virology** **83**, 501-509, 2011.

KOYANAGI, Y. et al. In vivo infection of human T-cell leukemia virus type I in non-T cells. **Virology** **196**, 25-33, 1993.

KOZLOWSKI, A.G. et al. Prevalence and genetic characterisation of HTLV-1 and 2 dual infections in patients with pulmonary tuberculosis in Central-West Brazil. **Mem Inst Oswaldo Cruz. Feb;109(1):118-21. doi: 10.1590/0074-0276130230**, 2014.

LA GRENADE, L. HTLV-I-associated infective dermatitis: past, present, and future. **J Acquir Immune Defic Syndr Hum Retrovirol** **13 Suppl 1**, S46-49, 1996.

LEPOUTRE, V. et al. Role of resident CNS cell populations in HTLV-1-associated neuroinflammatory disease. **Front Biosci** **14**, 1152-68, 2009.

LIM, D.G. et al. Examination of CD81 T Cell Function in Humans Using MHC Class I Tetramers: Similar Cytotoxicity but Variable Proliferation and Cytokine Production Among Different Clonal CD81 T Cells Specific to a Single Viral Epitope. **The J of Immunol** **165**, 6214-6220, 2000.

LI, N. et al. The transcription factor Cux1 regulates dendritic morphology of cortical pyramidal neurons. **PLoS One** **5**, e10596, 2010.

MACNAMARA A. et al. HLA class I binding of HBZ determines outcome in HTLV-1 infection. **PLoS Pathog.**; **6:e1001117**, 2010.

MAJOROVITS, E. et al. Human T-Lymphotropic Virus-1 Visualized at the Virological Synapse by Electron Tomography. **Plos One** **3**, 2008.

MANEL, N. et al. GLUT-1 is the receptor of retrovirus HTLV. *Med Sci*; **20**: 277-9, 2004.

MANEL, N. et al. HTLV-1 tropism and envelope receptor. *Oncogene* **24 (39)**, 6016-6025, 2005.

MANNS, A. et al. A prospective study of transmission by transfusion of HTLV-I and risk factors associated with seroconversion. *Int J Cancer.*, v. **51**, n. **6**, p. **886 – 891**, jul. 1992.

MAREES AT, KLUIVER H, STRINGER S, VORSPAN F et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Methods Psychiatr Res.*; **27**:e1608, 2018.

MARTIN, J.L. et al. Critical Role of the Human T-Cell Leukemia Virus Type 1 Capsid N-Terminal Domain for Gag-Gag Interactions and Virus Particle Assembly. *J. of Virology. Issue 14 e00333-18*, 2018.

MARTINS, M. B.; SOUZA, J. G.; FRANCO, G. M.; SANTOS, D. M. S.; BARBOSA-STANCIOLI, E. F. Capítulo 2 PATOGÊNESE DA INFECÇÃO PELO HTLV. In: CARNEIRO -PROIETTI, A. B. F. **Cadernos Hemominas HTLV** Volume XVI. 6ª ed. Atualizada e Aumentada. Belo Horizonte: Formato eletrônico, 2015, p. 28.

MARTINEZ, M.P. et al. Comparative virology of HTLV-1 and HTLV-2. *Retrovirology Aug 7;16(1):21*.doi: 10.1186/s12977-019-0483-0, 2019.

MATSUOKA, M. et al. The HBZ gene, a key player in HTLV-I pathogenesis. *Comparative Virology of HTLV-1 and HTLV-2 Retrovirology*; **7;16(1):21. 6: 71**, 2009.

MATSUOKA, M & JEANG. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer*; **7: 270-80**, 2007.

MATSUOKA, M.; JEANG, K. Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: viral infectivity, Tax, HBZ and therapy. *Oncogene*, v. **30**, p.1379-1389, 2011.

MEEKINGS, K.N. et al. HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. *PLoS Pathog.*, **4**, e1000027, 2008.

MELAMED A, YAGUCHI H, MIURA M, WITKOVER A, FITZGERALD TW, IRNEY

E, BANGHAM CRM. The human leukemia virus HTLV-1 alters the structure and transcription of host chromatin in cis. **eLife.**; **7**: e36245, 2018.

MICHAEL, B. et al. Role of accessory proteins of HTLV-1 in viral replication, T cell activation and cellular gene expression. **Front. Biosci.** **9**, 2527-2539, 2004.

MIYOSHI, I. et al. Type C virus particles in a cord T-cell line derived by co-cultivating normal human cord leukocytes and human leukaemic T cells. **Nature**; **294**:770-1, 1981.

MORAIS, M.T.M.et al. Perfil socioepidemiológico dos portadores do HTLV em um município do sudoeste baiano. **Rev. de Saúde Coletiva da UEFS**, v. **7**, n.3, dez. 2017.

MORENO-CARVALHO, O.A. et al. Evidence of preferential female prevalence of HTLV-I associated tropical spastic paraparesis in Bahia-Brazil. **Arq. de NeuroPsiquiatria, São Paulo**, v. **50**, n. **2**, p. **183-188**, 1992.

MORGAN, O.et al. HTLV-1 and polymyositis in Jamaica. **Lancet** **18 (2)**, 1184-1187, 1989.

MOSSOUN, A. et al. Bushmeat Hunting and Zoonotic Transmission of Simian T-Lymphotropic Virus 1 in Tropical West and Central Africa. **J Virol.** **91(10)**: e02479-16, 2017.

MOTA-MIRANDA, A C A. Estudo Molecular de Proteínas Estruturais (gp21 e gp46) e Regulatórias (HBZ) do HTLV-1 em Indivíduos com Diferentes Perfis Clínicos. **Tese de Doutorado, Fundação Oswaldo Cruz, Centro de Pesquisa Gonçalo Moniz**, 2012

MOXOTO I. et al: Sociodemographic, epidemiological and behavioral profile of women infected with HTLV-1 in Salvador, Bahia, an endemic area for HTLV. **Rev. Soc. Bras. Med. Trop. vol.40 no.1 Uberaba Jan./Feb.** 2007.

MOZHGANI, SH.et al. Human T-lymphotropic virus 1 (HTLV-1) pathogenesis: A systems virology study. **J. of cellular Biochemistry**, **119 (5)**, 2017.

MURPHY E.L. Infection with human T-lymphotropic virus types-1 and -2 (HTLV-1 and -2): implications for blood transfusion safety. **Transfus Clin Biol.** **23:13–9**, 2016.

NAITO T, YASUNAGA J, MITOBE Y, SHIRAI K, SEJIMA H, USHIROGAWA H, TANAKA Y, NAKAMURA T, HANADA K, FUJII M, MATSUOKA M, SAITO M. Distinct gene expression signatures induced by viral transactivators of different

HTLV-1 subgroups that confer a different risk of HAM/TSP. **Retrovirology**, Nov **6;15(1):72**, 2018.

NATH, M.D. et al. Regulation of the cell surface expression of an HTLV-I binding protein in human T-cells during immune activation. **Blood** **101**, **3085-3092**, 2003.

NICOT, C. et al. HTLV-1-encoded p30<sup>II</sup> is a post-transcriptional negative regulator of viral replication. **Nature Medicine** **10**, **197-201**, 2004.

NISHIOKA, K. HTLV-I arthropathy and Sjogren syndrome. **J Acquir Immune Defic Syndr Hum Retrovirol** **13 Suppl 1**, **S57-62**, 1996.

NOZUMA & JACOBSON. Neuroimmunology of Human T-Lymphotropic Virus Type 1-Associated Myelopathy/Tropical Spastic Paraparesis. **Front. Microbiol.** **10:885**, 2019.

NUNES, D. et al. HTLV-1 is predominantly sexually transmitted in Salvador, the city with the highest HTLV-1 prevalence in Brazil. **PLoS One.** **3;12(2):e0171303**. doi: **10.1371/journal.pone.0171303**. eCollection, 2017.

OMSLAND, M. et al. Inhibition of Tunneling Nanotube (TNT) Formation and Human T-cell Leukemia Virus Type 1 (HTLV-1) Transmission by Cytarabine. **Scientific Reports.** **8:11118**, 2018.

OSAME, M. Pathological Mechanisms of human T-cell lymphotropic virus type-1 associated myelopathy (HAM-TSP). **J Neurovirol** **8 (5)**, **359-364**, 2002.

PARRA E.J, et al. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. **Am J Phys Anthropol.** **114(1):18-29**, 2001.

PARKER, C. E. et al. Activated, HTLV-1-specific cytotoxic T-lymphocytes are found in health seropositives as well as in patients with tropical spastic paraparesis. **Virology** **188:628-636**, 1992.

PERCHER, F. et al. Mother-to-child transmission of HTLV-1 epidemiological aspects, mechanisms and determinants of mother-to-child transmission. **Viruses.** **8:E40**, 2016.

PINEDA, M.V. et al. Relevance of HTLV-1 proviral load in asymptomatic and symptomatic patients living in endemic and non-endemic areas of Argentina. **PLoS One.** **Nov 22;14(11):e0225596**, 2019.

PINON, J.D. et al. Human T-cell leukemia virus type 1 envelope glycoprotein

gp46 interacts with cell surface heparan sulfate proteoglycans. *J Virol* **77**, 9922-9930, 2003.

PIQUE, C; CONNAN, F; LEVILAIN, JP; CHOPPIN, J; DOKHÉLAR, MC. Among all Human T-cell Leukemia Virus Type 1 Proteins, Tax, Polymerase, and Envelope Proteins are Predicted as Preferential targets for the HLA-A2 Restricted Cytotoxic T-cell Response. *J of Virol* **4919-4926**, 1996.

POIESZ, B.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 of the United States of America*, v. **77**, n. **12**, p.7415–9, 1980.

POSADAS AE, ROSADO JJ, GOTUZZO E, TALLEDO MJ. Delimitation of the upstream region of NFKBIA gene associated with HTLV-1-associated myelopathy/tropical spastic paraparesis using candidate Tag-SNPs in Peruvian HTLV-1 infected individuals. *Infect Genet Evol.* **Nov;75:103929**, 2019.

PRICE A, et al. Principal components analysis corrects for stratification in genomewide association studies. *Nat Genet* **38: 904–909**, 2006.

PROIETTI, A. et al. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene*, vol **24**, p. **6058-6068**, 2005.

PROIETTI, A.B.F.C. HTLV. 304 p.; 17 X 23 cm; *Cadernos Hemominias; v.XIII*. 2006.

PURCELL S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.***81(3):559–75**, 2007.

RAMDZAN, Z. M. et al The function of CUX1 in oxidative DNA damage repair is needed to prevent premature senescence of mouse embryo fibroblasts. *Oncotarget* **6, 3613–3626** **53**, 2015.

RAMDZAN, Z. et al. RAS transformation requires CUX1-dependent repair of oxidative DNA damage. *PLoS Biol.* **12, e1001807**, 2014.

RAMOS J.M., et al. HTLV-1 infection and health outcomes. *Lancet Infect Dis.* **Apr;20(4):407-408. doi: 10.1016/S1473-3099(20)30133-X**, 2020.

RANJANA, ZUBAIDAH M. et al. CUX2 Protein Functions as an Accessory Factor in the Repair of Oxidative DNA Damage. *THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 290, NO. 37, pp. 22520 –22531, September 11*, 2015.

RAYNE, F. et al. *In Vivo* Homodimerisation of HTLV-1 Gag and MA Gives Clues

to the Retroviral Capsid and TM Envelope Protein Arrangement. **J molecular Biology. Volume 343, Issue 4, Pages 903-916**, 2004.

RIBEIRO, I. P. et al. HTLV-1 and 2 in a first-time blood donor population in Northeastern Brazil: prevalence, molecular characterization, and evidence of intrafamilial transmission. **J Med Virol. doi: 10.1002/jmv.25231**, 2018.

ROMAN, G. C. et al. Identity of HTLV-I-associated tropical spastic paraparesis and HTLV-I-associated myelopathy. **The Lancet, [s.l.], v. 331, n. 8586, p. 651, mar., 1988.**

ROMANELLI LCF et al. Incidence of Human T Cell Lymphotropic Virus Type 1-Associated Myelopathy/Tropical Spastic Paraparesis in a Long-Term Prospective Cohort Study of Initially Asymptomatic Individuals in Brazil. **AIDS Research and Human Retroviruses Vol. 29, No. 9**, 2013.

ROSADAS, C. & PUCCIONI-SOHLER, M. HTLV-1 ORF-I Encoded Proteins and the Regulation of Host Immune Response: Viral Induced Dysregulation of Intracellular Signaling. **J Immunol Res**, 2015.

ROSADO, J. et al. The FAS-670 AA genotype is associated with high proviral load in Peruvian HAM/TSP patients. **Med. Virology. <https://doi.org/10.1002/jmv.24681>**, 2017.

ROTEM, E. et al. The HTLV-1 gp21 fusion peptide inhibits antigen specific T-cell activation in-vitro and in mice. **PLOS Pathogens, [s. l.], v. 14, n. 5, p. e1007044**, 2018.

ROWAN A.G. et al. A lise de linfócitos T citotóxicos de células infectadas com HTLV-1 é limitada pela fraca expressão da proteína HBZ, mas não especificamente aprimorada na indução da expressão de Tax. **Retrovirology. 11 : 116. doi: 10.1186 / s12977-014-0116-6**, 2014.

RUSHING, A. W. et al. The HTLV -1 basic leucine zipper factor (HBZ) attenuates repair of double - 2 stranded DNA breaks via non -homologous end joining (NHEJ). **J. Virol. doi:10.1128/JVI.00672-18. J Immunol Res. 498054. doi: 10.1155/2015/498054**, 2018.

SABOURI, Amir H. et al. Polymorphism in the interleukin-10 promoter affects both provirus load and the risk of human T lymphotropic virus type I-associated myelopathy/tropical spastic paraparesis. **The Journal of infectious diseases, v. 190, n. 7, p. 1279-1285**, 2004.

SAITO, M. Association Between HTLV-1 Genotypes and Risk of HAM/TSP.

*Front. Microbiol.*, 15 May, 2019.

SAMPEY, G. C.; et al. Complex role of microRNAs in HTLV-1 infections. *Frontiers in genetics*. doi: 10.3389/fgene.2012.00295, 2012.

SATOU, Y & MATSUOKA, M. HTLV-1 and the Host Immune System : How the Virus Disrupts Immune Regulation, Leading to HTLV-1 Associated Diseases. *J Clin Exp Hematopathol* 50, 2010.

SEIKI, M. et al. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc Natl Acad Sci USA*. 80, 3618-3622. PMID: 6304725, 1984.

SETOYAMA, M. et al. Detection of HTLV-1 by polymerase chain reaction in situ hybridization in adult T-cell leukemia/lymphoma. *Am J Pathol* 152, 683–689, 1998.

SILVERMAN, LR. et al. Human T-cell lymphotropic virus type 1 open reading frame II-encoded p30II is required for in vivo replication: evidence of in vivo reversion. *J Virol* 78, 3837-3845, 2004.

SLATERRY, J.P. et al. Genomic Evolution, Patterns of Global Dissemination, and Interspecies Transmission of Human and Simian T-cell Leukemia/Lymphotropic Viruses. *Genome Research*, [s.l.], v. 9, p. 525-540, 1999.

SOUZA, A. et al. Immunopathogenesis and neurological manifestations associated to HTLV- 1 infection. *Revista da Sociedade Brasileira de Medicina Tropical*, v. 45, n. 5, p. 545-552, set/out. 2012.

SOUZA MACHADO, A. et al. Imunopatogênese da infecção pelo HTLV-1: influência sobre a resposta imune tipo 2. *Rev. Bras. Alerg. Immunopatol.*, v. 26, n. 4, p. 159 - 167, 2003.

SUNDARAM, R. et al. A novel multivalent human CTL peptide construct elicits robust cellular immune responses in HLA-A\* 0201 transgenic mice: implications for HTLV-1 vaccine design, *Vaccine* 21 2767-2781, 2003.

SZUREK, PF. et al. Identification of point mutations in the envelope gene of Moloney murine leukemia virus TB temperature-sensitive paralogous mutant ts1: molecular determinants for neurovirulence. *J Virol* 62 (1), 357-360, 1988.

TAGAYA, Y. et al. 40 years of the human T-cell leukemia virus: past, present, and future. *F1000 Research*, v. 8, p. 1-14, 2019.

TALLET, B. et al. Sequence Variations in the Amino- and Carboxy-Terminal Parts of the Surface Envelope Glycoprotein of HTLV Type 1 Induce Specific Neutralizing Antibodies. *AIDS Res. And Hum. Retrov.* **17**, 337-348, 2001.

TAYLOR, G.P. et al. The seroepidemiology of human T-lymphotropic viruses: types I and II in Europe: a prospective study of pregnant women. *J Acquir Immune Defic Syndr*; **38**: 104-9, 2005.

THÉBAULT, S. et al. HBZ interacts with JunD and stimulates its transcriptional activity. *FEBS Lett.* **562**, 165-170, 2004.

TREVIÑO A, LOPEZ M, VISPO E, AGUILERA A, RAMOS JM, BENITO R, ROC L, EIROS JM, MENDOZA C, SORIANO V. Development of tropical spastic paraparesis in human T-lymphotropic virus type 1 carriers is influenced by interleukin 28B gene polymorphisms. *Clin Infect Dis. Jul;55(1):e1-4*, 2012.

VALLINOTO, A.C.R. FAS-670A/G single nucleotide polymorphism may be associated with human T lymphotropic virus-1 infection and clinical evolution to TSP/HAM. *Virus Research* **163 (2012) 178–182**, 2012.

VALLINOTO ACR, SANTANA BB,1 SÁ KSG, FERREIRA TCS, SOUSA RCM, AZEVEDO VN, FEITOSA RNM, MACHADO LFA, ISHAK MOG, ISHAK R. HTLV-1-Associated Myelopathy/Tropical Spastic Paraparesis Is Not Associated with SNP rs12979860 of the IL-28B Gene. *Mediators of Inflammation Volume 2015, Article ID 804167, 7 pages*, 2015.

VARANDAS CMN, et al. Early Juvenile Human T-cell Lymphotropic Virus Type-1-Associated Myelopathy/Tropical Spastic Paraparesis: Study of 25 Patients. *Clin Infect Dis. Oct 15;67(9):1427-1433*, 2018.

VARELLA, C. A. A. Análise Multivariada Aplicada as Ciências Agrárias: Análise de Componentes Principais. *PÓS-GRADUAÇÃO EM AGRONOMIA – CPGA-CS*, 2008.

VERDONCK, K. et al. Human T-lymphotropic virus 1: recent knowledge about an ancient infection. *Lancet Infect Dis.* **7:266–81**, 2007.

VINE, A.M. et al. The role of CTLs in persistent viral infection: cytolytic gene expression in CD8+ lymphocytes distinguishes between individuals with a high or low proviral load of human T cell lymphotropic virus type 1. *J. Immunol*; **173: 5121-9**, 2002.

VISSCHER PM. et al. Five Years of GWAS Discovery. *The American Journal of Human Genetics* **90, 7–24, January 13**, 2012.

VITA R. et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* Oct 24. doi: 10.1093/nar/gky1006, 2018.

WALL JD & PRITCHARD. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 4: 587-597, 2003.

WEIXIN, W.u. et al. Human T-cell leukemia virus type 1 Gag domains have distinct RNA-binding specificities with implications for RNA packaging and dimerization. *J Biol Chem.* Oct 19; 293(42): 16261–16276, 2018.

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

YAMAUCHI, J. et al. Use of cerebrospinal fluid CXCL10 and neopterin as biomarkers in HTLV-1- associated myelopathy/tropical spastic paraparesis treated with steroids. *J Neurol Neurosurg Psychiatry*; 0:1–3. doi:10.1136/jnnp-2019-321955, 2019.

YANG, K. et al. Structural basis for cooperative regulation of KIX-mediated transcription pathways by the HTLV-1 HBZ activation domain. [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1810397115/-](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1810397115/-/DCSupplemental) /DCSupplemental, 2018.

YASHIKI S. et al. HLA-A\*26, HLA-B\*4002, HLA-B\*4006, and HLA-B\*4801 alleles predispose to adult T cell leukemia: the limited recognition of HTLV type 1 tax peptide anchor motifs and epitopes to generate anti-HTLV type 1 tax CD8(+) cytotoxic T lymphocytes. *AIDS Res Hum Retroviruses*, Jul 20;17(11):1047-61, 2001

YOSHIDA, M. et al. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc Natl Acad Sci U S A.*; 79: 2031-5, 1982.

**ANEXO I****Alelos de HLA classe I testados na predição de epítomos**

<b>HLA* A</b>		
HLA-A*01:01	HLA-A*23:01	HLA-A*32:01
HLA-A*02:01	HLA-A*24:02	HLA-A*33:01
HLA-A*02:03	HLA-A*26:01	HLA-A*68:01
HLA-A*02:06	HLA-A*30:01	HLA-A*68:02
HLA-A*03:01	HLA-A*30:02	
HLA-A*11:01	HLA-A*31:01	
<b>N= 16</b>		
<b>HLA* B</b>		
HLA-B*07:02	HLA-B*44:02	HLA-B*58:01
HLA-B*08:01	HLA-B*44:03	
HLA-B*15:01	HLA-B*51:01	
HLA-B*35:01	HLA-B*53:01	
HLA-B*40:01	HLA-B*57:01	
<b>N=11</b>		

### ANEXO II

Descrição dos 10 alelos de HLA mais frequentes entre os “alelos protéicos” de cada proteína viral.

gp46												
"Alelo Protéico"	Perfil Clínico	Alelos de HLA										
		B*53:01	A*32:01	B*58:01	B*07:02	B*35:01	A*23:01	A*24:02	A*30:02	B*57:01	B*51:01	A*02:01
Selvagem	ASS-fHAM/TSP-sHAM/TSP-ATLL	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
1	ASS	Red	Orange	Yellow	Green	Red	Blue	White	Green	Blue	Grey	Purple
2	ASS	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
3	ASS	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
4	ASS	Red	Red	Green	Yellow	Orange	Green	Blue	Grey	Blue	Purple	White
5	ATLL	Red	Orange	Yellow	Green	Red	Green	Blue	Purple	Blue	Grey	White
6	ATLL	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
7	sHAM/TSP	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
8	sHAM/TSP	Red	Red	Green	Yellow	Orange	Green	Blue	Grey	Blue	Purple	White
9	sHAM/TSP	Red	Red	Green	Yellow	Orange	Green	Blue	Grey	Blue	Purple	White
10	sHAM/TSP	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
11	sHAM/TSP	Red	Red	Green	Yellow	Orange	Green	Blue	Blue	Grey	Purple	White
12	sHAM/TSP	Red	Orange	Green	Yellow	Red	Green	Blue	Grey	Purple	Blue	White
13	fHAM/TSP	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
14	fHAM/TSP-sHAM/TSP-ATLL	Red	Red	Orange	Yellow	Green	Green	Blue	Blue	Grey	Purple	White
15	fHAM/TSP-sHAM/TSP	Red	Orange	Yellow	Green	Red	Blue	Grey	Blue	Green	Purple	White
16	fHAM/TSP	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
17	ASS	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
18	sHAM/TSP	Red	Orange	Yellow	Green	Red	Blue	Grey	Blue	Green	Purple	White
19	sHAM/TSP	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
20	fHAM/TSP	Red	Red	Green	Yellow	Orange	Green	Blue	Blue	Grey	Purple	White
21	ATLL	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White
22	ATLL	Red	Red	Yellow	Green	Orange	Green	Blue	Grey	Blue	Purple	White





p12												
"Alelo Protéico"	Perfil Clínico	Alelos de HLA										
		B*07:02	B*53:01	A*02:01	A*32:01	A*23:01	A*02:03	A*24:02	B*35:01	A*02:06	B*58:01	B*57:01
Selvagem	ASS-fHAM/TSP-sHAM/TSP-ATLL											
1	fHAM/TSP											
2	fHAM/TSP											
3	fHAM/TSP											
4	ASS											
5	sHAM/TSP											
6	sHAM/TSP											
7	sHAM/TSP											
8	sHAM/TSP											
9	sHAM/TSP											
10	sHAM/TSP											
11	ATLL											
12	ATLL											
13	ATLL											
14	ASS											

1° Alelo mais frequente	2° Alelo mais frequente	3° Alelo mais frequente	4° Alelo mais frequente	5° Alelo mais frequente	6° Alelo mais frequente	7° Alelo mais frequente	8° Alelo mais frequente	9° Alelo mais frequente	10° Alelo mais frequente	Alelo ausente entre os 10
-------------------------------	-------------------------------	-------------------------------	-------------------------------	-------------------------------	-------------------------------	-------------------------------	-------------------------------	-------------------------------	--------------------------------	------------------------------------