

RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta tese
será disponibilizado somente a partir
de 21/02/2022.



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU – IBB



Programa de Pós-Graduação em Ciências Biológicas (GENÉTICA)

**Estudo da variabilidade das regiões promotora e codificadora do
gene *HLA-C* e assinaturas de seleção natural atuando nestes
segmentos**

ANDRÉIA DA SILVA SOUZA

Botucatu - SP

2020



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU – IBB



Programa de Pós-Graduação em Ciências Biológicas (GENÉTICA)

**Estudo da variabilidade das regiões promotora e codificadora do
gene *HLA-C* e assinaturas de seleção natural atuando nestes
segmentos**

ANDRÉIA DA SILVA SOUZA

ORIENTADOR: PROF. DR. ERICK DA CRUZ CASTELLI

Tese apresentada ao Instituto de Biociências,
Universidade Estadual Paulista “Júlio de Mesquita
Filho”, *Campus* de Botucatu, para obtenção do
título de Doutora pelo Programa de Pós-
Graduação em Ciências Biológicas (Genética).

Botucatu – SP

2020

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Souza, Andréia da Silva.

Estudo da variabilidade das regiões promotora e codificadora do gene *HLA-C* e assinaturas de seleção natural atuando nestes segmentos / Andréia da Silva Souza. - Botucatu, 2020

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Erick da Cruz Castelli
Capes: 20200005

1. Antígenos *HLA-C*. 2. Genes MHC Classe I. 3. Variação genética. 4. Seleção natural. 5. Benin. 6. Brasil.

Palavras-chave: Benin; Brasil; *HLA-C*; seleção natural; variabilidade.

Andréia da Silva Souza

**Estudo da variabilidade das regiões promotora e codificadora do gene
HLA-C e assinaturas de seleção natural atuando nestes segmentos**

English Title: *HLA-C* promoter and coding genetic diversity and evolutionary aspects

Orientador: Prof^o. Dr. Erick C. Castelli

Comissão examinadora

Prof. Dr. Erick C. Castelli
Universidade Estadual Paulista (UNESP)

Prof. Dr. Diogo Meyer
Universidade de São Paulo (USP)

Prof. Dr. Celso Teixeira Mendes-Junior
Universidade de São Paulo (USP)

Profa. Dra. Norma Lucena Cavalcanti Licinio da Silva
Fundação Oswaldo Cruz, Centro de Pesquisas Aggeu Magalhaes

Profa. Dra. Camila Ferreira Bannwart
Universidade Estadual Paulista (UNESP)

Botucatu, 21 de fevereiro de 2020

Dedicatória

Dedico este trabalho...

*Aos meus pais **Antônio** e **Lourdes** que são meu porto seguro e sempre apoiam minhas decisões. Obrigada por toda dedicação e amor.*

Agradecimentos

- ❖ *Agradeço primeiramente a **Deus** por ter guiado os meus passos até aqui e pela sua constante presença em minha vida.*
- ❖ *Aos meus pais **Antônio e Lourdes**, ao meu irmão **Tony**, minha cunhada **Luana** e meu sobrinho **Anthony** pelo amor e apoio incondicional em todos os momentos.*
- ❖ *Às minhas tias **Rose e Lêda** por cuidar de mim, sempre me incentivar, aconselhar, apoiar minhas decisões e pela frase “Qualquer coisa, estamos aqui, viu!”.*
- ❖ *A todos familiares e amigos que mesmo distantes estão sempre torcendo e ajudando em minha caminhada.*
- ❖ *Ao meu namorado **Marlon Jocimar**, por todo companheirismo, cuidado e zelo comigo e por me incentivar sempre. Obrigada por tornar essa jornada mais leve.*
- ❖ *À minha querida amiga **Thálitza Ayala**, por estar sempre ao meu lado e por ser a mão amiga a me ajudar sempre que precisei. Obrigada pelas conversas, pelos conselhos, por sempre me animar, pelo cuidado e pela amizade que nos sustentou em tantos momentos de dificuldade e que também nos trouxe tanta alegria e sorrisos durante essa jornada.*
- ❖ *À minha querida amiga **Leticia Pastore Mendonça** pela companhia, pelos conselhos, pelas boas conversas, risadas e pelos cafezinhos da tarde sempre cheios de histórias de vida, de família, de força e de superação que nos ajudou muitas vezes a aguentar a saudade de casa e as dificuldades do dia a dia. Muito obrigada.*
- ❖ *Às minhas amigas **Rosemary Cristina e Luciana Pizzani** pelo acolhimento, carinho e amizade que sempre encontro em vocês.*
- ❖ *Aos amigos especiais que fizeram parte dessa jornada **Iane de Oliveira Pires Porto, Michelle de Almeida Paz e Neilton Paulo Bezerra**. Aprendi a amar cada um com seu jeito único e foi maravilhoso contar com a amizade de vocês durante esse período.*
- ❖ *Aos amigos e colegas de trabalho do GemBio **Heloisa de Souza Andrade, Marília Rodrigues Silva Passos, Nayane dos Santos Brito Silva, Emiliana Weiss, Hector Sebastian Baptista, Arielle Lima da Rocha, Gabriela Pereira de Carvalho e Camila Ferreira Bannwart** que fizeram parte dessa jornada e trouxeram tantos aprendizados,*

estiveram ao meu lado em momentos difíceis e momentos de muita descontração e boas risadas. Obrigada pelos conhecimentos e sonhos compartilhados, pela amizade e principalmente pelas sessões de terapia.

- ❖ *Aos amigos de Botucatu que me acolheram aqui e se tornaram a família de Botucatu.*
- ❖ *Aos professores que fizeram parte da banca de qualificação deste trabalho, Dr. Ivan de Godoy Maia, Dra. Cláudia Aparecida Rainho e Dr. Ramon Kaneno pelas contribuições científicas.*
- ❖ *Aos professores que compuseram a comissão examinadora desta tese, pelas correções e contribuições.*
- ❖ *Ao meu orientador Prof. Dr. **Erick C. Castelli**, por contribuir com o meu aprendizado e com minha formação acadêmica, pelos conselhos, pela oportunidade e por inspirar seus alunos com sua dedicação e profissionalismo. Muito obrigada, Professor!*
- ❖ *À Universidade Estadual Paulista Júlio Mesquita Filho, Campus de Botucatu, pela infraestrutura oferecida e pela oportunidade de construir minha formação científica nessa instituição.*
- ❖ *A todos os profissionais da UNESP com os quais tive contato durante esse período de formação.*
- ❖ *Ao programa de Pós-Graduação em Ciências Biológicas (Genética), pela oportunidade e credibilidade no desempenho do projeto.*
- ❖ *À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de doutoramento.*
- ❖ *À Fundação de Amparo e Apoio a Pesquisa (FAPESP) pelo financiamento do projeto de pesquisa realizado pelo grupo, processo n° 2013/17084-2, do qual faz parte a pesquisa apresentada nesta tese.*
- ❖ *E a todos que fizeram parte dessa caminhada: muito obrigada!*

“Quem acende uma luz é o primeiro a se beneficiar da claridade.”

G. K. Chesterton

SUMÁRIO

LISTA DE FIGURAS	4
LISTA DE TABELAS	5
LISTA DE ABREVIATURAS E SIGLAS	6
RESUMO.....	7
ABSTRACT.....	8
CAPÍTULO I-REVISÃO DE LITERATURA	9
INTRODUÇÃO	10
Estrutura e função das moléculas HLA de classe I	11
<i>HLA-C</i> e suas características particulares	17
Seleção Natural atuando sobre os genes HLA	21
Testes de desvio de neutralidade e demografia das populações estudadas	25
OBJETIVOS	28
Objetivo Geral	28
Objetivos específicos	28
REFERÊNCIAS.....	29
CAPÍTULO II-ARTIGO	37
ARTICLE.....	38
ABSTRACT.....	39
1. INTRODUCTION.....	40
2. MATERIALS AND METHODS	42
2.1. Brazilian samples	42
2.2. Beninese samples	42
2.3. <i>HLA-C</i> gene amplification and sequencing libraries	43
2.4. Raw data processing, mapping, and genotyping	43
2.5. Phasing and <i>HLA-C</i> allele calling	44
2.6. Other analyses	45
3. RESULTS	46
3.1. <i>HLA-C</i> genetic diversity	46
3.2. <i>HLA-C</i> evolutive aspects	48
4. DISCUSSION	49
SUPPLEMENTAL MATERIAL	62
REFERENCE.....	81
CONSIDERAÇÕES FINAIS	86
ANEXOS	87

Lista de Figuras

Lista de Figuras – Capítulo I

Figura 1. Representação esquemática da estrutura de uma molécula de MHC de classe I e apresentação de antígeno às células T CD8.....	11
Figura 2. Processamento de antígeno pela via de moléculas MHC de classe I.....	12
Figura 3. Visão geral esquemática da fenda de ligação a peptídeos.....	14
Figura 4. Representação esquemática das características distintas da molécula HLA-C.....	19

Lista de Figuras – Capítulo II

Figure 1. Nucleotide diversity and Tajima's D at the HLA-C promoter, CDS and 3'UTR.....	61
---	----

Lista de Figuras - Material Suplementar

Figure S1. <i>Linkage Disequilibrium</i> (LD) between pair of single nucleotide polymorphisms (SNPs) of the HLA-C locus, considering genomic positions from 3'UTR to Promoter (31268757- 31273596, from hg38).	63
Supplemental alignment 1. <i>HLA-C</i> promoter sequences in two population samples (Brazil and Benin).	64
Supplemental alignment 2. HLA-C 3'UTR sequences in two population samples (Brazil and Benin).	68

Lista de Tabelas

Lista de Tabelas - Capítulo I

Tabela 1. Número de alelos identificados para os genes de classe I (The IPD-IMGT/HLA Database, versão 3.38.0.	10
--	----

Lista de Tabelas - Capítulo II

Table 1. List of <i>HLA-C</i> promoter sequences in two population samples from Brazil and Benin, their frequencies and the HLA-C molecules associated with them.	55
Table 2. List of <i>HLA-C</i> CDS sequences detected in two population samples from Brazil and Benin, and their frequencies.	56
Table 3. List of HLA-C encoded proteins detected in two population samples from Brazil and Benin, and their frequencies.	57
Table 4. List of <i>HLA-C</i> 3'UTR sequences detected in two population samples from Brazil and Benin, and their frequencies.	57
Table 5. List of <i>HLA-C</i> extended haplotypes detected in two population samples from Brazil and Benin, and their frequencies.	58
Table 6. Nucleotide diversity and Neutrality tests across the <i>HLA-C</i> locus.	60
Table 7. dN/dS ratio test of the <i>HLA-C</i> exons and CDS.	60

Lista de Tabelas - Material Suplementar

Table S1. List of all variants observed across HLA-C and the reference allele frequency in Brazil and Benin, as genotyped by the GATK HaplotypeCaller.	70
Table S2. List of the HLA-C genomic alleles and their frequencies in Brazil and Benin.	78
Table S3. Amino acid exchanges and their frequencies in Brazil and Benin.	79

Lista de Abreviaturas e Siglas

APC	Célula Apresentadora de Antígeno (do inglês, <i>Antigen Presenting Cell</i>)
ATP	Adenosina Trifosfato (do inglês, <i>Adenosine triphosphate</i>)
BiP	Proteína Ligação de imunoglobulina (do inglês, <i>Binding Immunoglobulin Protein</i>)
CTL	Linfócito T Citotóxico
DNA	Ácido Desoxirribonucleico (do inglês, <i>Deoxyribonucleic acid</i>)
ERAAP	Aminopeptidases Associadas ao Processamento de antígeno do Retículo Endoplasmático
HIV	Vírus da Imunodeficiência Humana (do inglês, <i>Human Immunodeficiency Virus</i>)
HLA	Antígeno Leucocitário Humano (do inglês, <i>Human Leukocyte Antigen</i>)
IFN	Interferon
IMGT	International Immunogenetics Database
KIR	Receptor do tipo imunoglobulina (KIR, do inglês <i>Killer Cell Immunoglobulin Like Receptor</i>)
KIRV	Resíduo formado pelos aminoácidos Lisina (K), Isoleucina (I), Arginina (R) e Valina (V)
Kb	Kilobase (10^3 bases)
LD	Desequilíbrio de Ligação (<i>Linkage Disequilibrium</i>)
Mb	Megabase (10^6 bases)
MHC	Complexo Principal de Histocompatibilidade (do inglês, <i>Major Histocompatibility complex</i>)
mRNA	Ácido Ribonucleico mensageiro ou RNA mensageiro
NK	Célula Assassina Natural (do inglês, <i>Natural Killer</i>)
NT	Não Traduzida
Pb	Pares de bases
PCR	Reação em Cadeia da Polimerase (do inglês, <i>Polymerase Chain Reaction</i>)
RE	Retículo Endoplasmático
SNP	Polimorfismo de base única (do inglês, <i>Single Nucleotide Polymorphism</i>)
TAP	Transportador associado ao processamento de antígeno (do inglês, <i>Transporter Associated with Antigen Processing</i>)
TCR	Receptor de célula T
TNF	Fator de Necrose Tumoral
UTR	Região não traduzida (do inglês, <i>Untranslated region</i>)

RESUMO

O gene *HLA-C* está localizado dentro do Complexo Principal de Histocompatibilidade (MHC), a região mais variável do genoma humano. *HLA-C* codifica moléculas que participam principalmente do processo de apresentação de antígenos intracelulares aos linfócitos T citotóxicos e interage com receptores presentes nas células NK, modulando sua atividade. A variabilidade das moléculas HLA permite a apresentação de diferentes peptídeos por um mesmo indivíduo e aumenta o repertório de peptídeos apresentados por uma população. Além disso, *HLA-C* é expresso na interface materno-fetal pelo trofoblasto, onde possui uma importante função imunomodulatória por meio da interação com receptores KIR das células NK maternas. Devido a sua dupla função, alguns alelos de *HLA-C* ou combinações de alelos *HLA-C/KIR* têm sido associados a diversos contextos fisiológicos e patológicos. Contudo, a variabilidade desse gene tem sido explorada principalmente para os éxons, em especial os éxons 2 e 3, que codificam os domínios de ligação ao peptídeo, enquanto a variabilidade de outros segmentos gênicos importantes (como outros éxons, íntrons e regiões regulatórias) foram pouco estudadas. Este estudo avaliou a variabilidade genética e assinaturas de seleção natural ao longo do gene *HLA-C* em uma amostra de 418 indivíduos do Brasil e 108 indivíduos do Benin. Considerando as duas populações, detectamos 359 sítios de variação ao longo da região analisada. Os haplótipos promotores, codificadores e de 3'NT apresentaram uma correlação direta, com alelos que codificam proteínas semelhantes (grupos alélicos) associados a sequências específicas de promotores e de 3'NT. Observamos alta diversidade nucleotídica ao longo de todo o gene, com destaque para o segmento que codifica o domínio transmembrana. Na região promotora foi encontrada a menor diversidade nucleotídica em uma região próxima à posição -700, um segmento monomórfico de 115pb que é compartilhado entre diferentes primatas. Para a 3'NT foi encontrada alta diversidade na segunda metade deste segmento, porém baixa diversidade no trecho inicial. Nas duas populações foram detectados valores positivos de *D de Tajima* e valores negativos de *F de Ewens-Watterson* para quase todos os segmentos do gene *HLA-C*, compatíveis com seleção balanceadora atuando em todo o gene. Além disso, ambas as populações apresentaram evidência de seleção positiva para o éxon 1, que é responsável por codificar o peptídeo de sinal. As frequências dos motivos *HLA-C* previamente associados à interação com KIR e regulação da expressão são semelhantes entre as duas populações, porém há mudanças na frequência de aminoácidos que influenciam o repertório de peptídeo ligantes, indicando que o repertório peptídico apresentado por essas populações pode ser diferente. Essas populações apresentam grandes diferenças na frequência de aminoácidos no segmento transmembrana e também no primeiro aminoácido do domínio alfa-1, mas não está claro se essas modificações alteram a função e a estabilidade do *HLA-C*. Estudos funcionais ainda são necessários para o melhor entendimento sobre a relação entre o padrão de variabilidade de *HLA-C* e sua expressão, capacidade de apresentação antigênica, estabilidade e perfil de ligação KIR/*HLA-C*.

Palavras-chave: *HLA-C*, variabilidade, seleção natural, NGS, Brasil, Benin.

ABSTRACT

Human Leucocyte antigen-C (HLA-C) is a classical HLA class I molecule that binds and presents peptides to cytotoxic T lymphocytes in the cell surface. HLA-C has a dual function since it also interacts with KIR receptors expressed in NK and T cells, modulating their activity. The structure and diversity of the *HLA-C* regulatory regions, as well as the relationship among variants along the *HLA-C* locus, is poorly addressed, and no population-based study explored the complete *HLA-C* variability in different population samples. Here we first present a new molecular and bioinformatics method to evaluate the full HLA-C segment, including regulatory sequences. Then, we applied this method to survey the *HLA-C* diversity in two geographically distinct population samples, one admixed from Brazil (São Paulo State) and one less admixed from Benin. Our results indicate that the *HLA-C* promoter and 3'UTR are very polymorphic, with the presence of few but highly divergent haplotypes. However, both these regulatory regions present conserved segments that are shared among different primates. Nucleotide diversity was higher in other exonic segments rather than exons 2 and 3, and also higher in the second half of the 3'UTR region. We detected evidence of balancing selection on the entire *HLA-C* locus and positive selection in exon 1, for both populations. HLA-C motifs previously associated with KIR interaction and expression regulation are similar between both populations, but the frequency of amino acids influencing peptide-binding is different between samples. Each allele group is associated with specific regulatory sequences, supported by the high linkage disequilibrium along the entire *HLA-C* locus in both populations.

Key words: *HLA-C*, variability, natural selection, NGS, Brazil, Benin.

Capítulo I - Revisão de Literatura

INTRODUÇÃO

Os Antígenos Leucocitários Humanos (HLA, do inglês *Human Leucocytes Antigens*) são moléculas envolvidas no controle da resposta imunitária, principalmente no processo de apresentação de peptídeos aos linfócitos T (Shiina *et al.*, 2004; Shiina *et al.*, 2009). Os genes *HLA* estão localizados dentro do Complexo Principal de Histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*), no braço curto do cromossomo seis, em 6p21.3. Esse complexo possui aproximadamente 250 genes distribuídos em ~4Mb e está dividido didaticamente em três grupos ou classes, denominados classe I, II, ou III (Shiina *et al.*, 2004; Shiina *et al.*, 2009). Os genes *HLA* de classe I são responsáveis pela codificação de proteínas ligadas à membrana celular e possuem um papel essencial no desempenho da resposta imune inata e adaptativa, podendo interagir com receptores dos linfócitos T citotóxicos ou CD8+ e de células *Natural Killer* (NK). Essa classe de genes pode ainda ser subdividida em genes *HLA* de classe I clássicos ou Ia (*HLA-A*, *HLA-B* e *HLA-C*), associados principalmente a apresentação antigênica, e não clássicos ou Ib (*HLA-E*, *HLA-F* e *HLA-G*) que desempenham funções imunomodulatórias (Klein and Sato, 2000; Blais *et al.*, 2011; Donadi *et al.*, 2011). Os genes de classe I, principalmente os loci *HLA-A*, *HLA-B* e *HLA-C*, constituem os genes mais variáveis do genoma humano (Klein and Sato, 2000; Bernatchez and Landry, 2003; Abbas *et al.*, 2011). Atualmente, cerca de 18.441 diferentes alelos para estes genes estão depositados no banco de dados oficial para HLA (The IPD-IMGT/HLA Database) (Robinson *et al.*, 2015) (Tabela 1).

Tabela 1. Número de alelos identificados para os genes de classe I (The IPD-IMGT/HLA Database, versão 3.39.0 de 20-01-2020).

CLASSE I CLÁSSICOS			CLASSE I NÃO CLÁSSICOS		
Gene	Alelos	Proteínas	Gene	Alelos	Proteínas
<i>HLA-A</i>	5.907	3.720	<i>HLA-E</i>	84	15
<i>HLA-B</i>	7.126	4.604	<i>HLA-F</i>	44	6
<i>HLA-C</i>	5.709	3.470	<i>HLA-G</i>	69	19
Total	18.742	11.794	Total	197	40

A região da classe II abrange aproximadamente 0,7 Mb de DNA e contém os genes que codificam as cadeias α e β dos HLA de classe II clássicos, *HLA-DP*, *HLA-DQ* e *HLA-DR*. Essas moléculas são expressas na superfície de células apresentadoras de antígenos (APC, do inglês *Antigen Presenting Cell*) e apresentam antígenos exógenos às células T auxiliares ou CD4+. A região da classe III, localizada entre as regiões de classe I e classe II, apresenta alta densidade

gênica, com aproximadamente um gene a cada 14.516 nucleotídeos, e contém genes do sistema complemento (*e.g.*, *C2* e *C4*) genes de citocinas (*e.g.*, *TNF*, *LTA* e *LTB*) e muitos genes sem relação direta conhecida com a função imunitária ou inflamação (Horton *et al.*, 2004; Shiina *et al.*, 2004; Shiina *et al.*, 2009).

Estrutura e função das moléculas HLA de classe I

Estruturalmente, as moléculas de classe I clássicas e não clássicas são semelhantes, sendo compostas por uma cadeia pesada α associada não covalentemente a uma cadeia leve de β_2 -microglobulina (β_2m), codificada por um gene localizado no cromossomo 15 (Abbas *et al.*, 2011). A cadeia α codificada pelos genes HLA de classe I é composta por cinco domínios: domínios α_1 e α_2 , que constituem o sítio de ligação a peptídeos; domínio α_3 semelhante à imunoglobulina, que interage com o co-receptor CD8 dos linfócitos; domínio transmembrana e domínio citoplasmático (Klein and Sato, 2000; Alberts *et al.*, 2015) (Figura 1).

No entanto, moléculas clássicas e não clássicas são diferencialmente expressas e possuem funções diferentes. As moléculas HLA de classe I clássicas (HLA-A, HLA-B e HLA-C) são expressas pela maioria das células somáticas e atuam principalmente no processo de apresentação de antígenos intracelulares aos linfócitos T citotóxicos ou TCD8+. As moléculas HLA de classe I não clássicas (HLA-E, HLA-F e HLA-G) possuem expressão celular mais restrita e desempenham principalmente um papel imunomodulatório, por meio de ligações com receptores específicos em células do sistema imune, tais como monócitos, células T, B e NK (*Natural Killer*) (Klein and Sato, 2000; Lepin *et al.*, 2000; Pietra *et al.*, 2009; Donadi *et al.*, 2011). No entanto, há evidências de que as moléculas não clássicas também participam da apresentação antigênica (D'souza *et al.*, 2019).

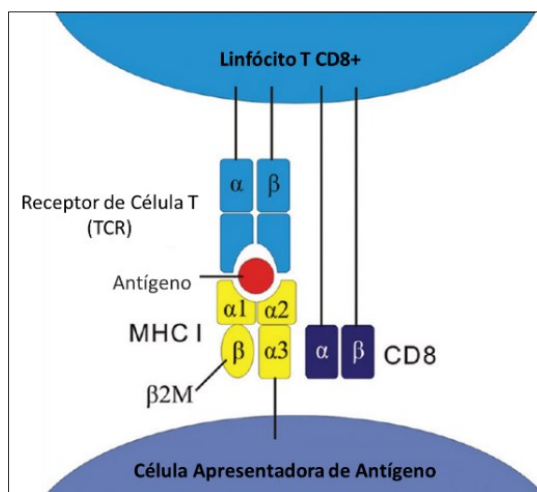


Figura 1. Representação esquemática da estrutura de uma molécula de MHC de classe I e apresentação de antígeno às células T CD8 (Adaptado de Li *et al.*, 2016).

As duas cadeias polipeptídicas das moléculas HLA de classe I (α e β_2m), são sintetizadas separadamente e são associadas na superfície luminal do retículo endoplasmático (RE) pela ação de chaperonas, como a calreticulina, BiP (do inglês, *Binding immunoglobulin Protein*), p57 do retículo endoplasmático e calnexina (Yewdell *et al.*, 2003; Parcej and Tampé, 2010; Abbas *et al.*, 2011; Eggensperger and Tampé, 2015). Os peptídeos que serão associados às moléculas de HLA de classe I são gerados por um complexo multiprotéico denominado proteassoma, responsável pela degradação de proteínas citoplasmáticas, em um processo dependente de ubiquitina e ATP (ATP, do inglês, *Adenosine triphosphate*) (Yang *et al.*, 1996). Esses peptídeos são transportados ao lúmen do RE por meio do Transportador Associado ao Processamento de Antígeno (TAP), onde se associam as moléculas de HLA de classe I com o auxílio do complexo de carregamento de peptídeos (formado por TAP e chaperonas). A tapasina é uma chaperona fundamental para a ligação do peptídeo e interage diretamente com a molécula de MHC I (Hulpke *et al.*, 2012; Blees *et al.*, 2017) (Figura 2).

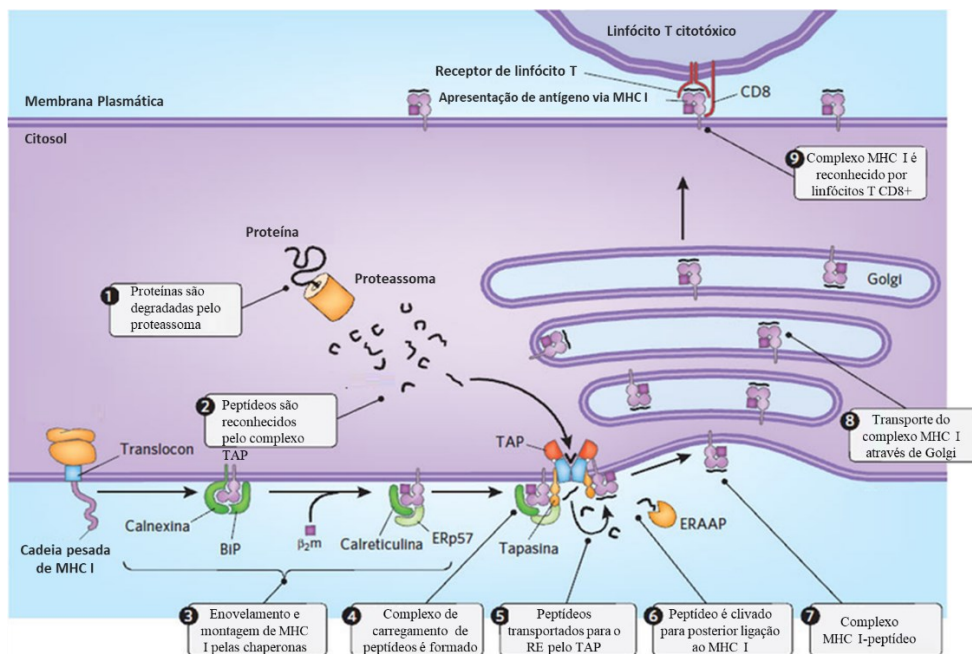


Figura 2. Processamento de antígeno pela via de moléculas MHC de classe I (Adaptado de Parcej e Tampé, 2010).

As moléculas HLA de classe I associam-se a peptídeos de aproximadamente 8 a 9 aminoácidos (AA), dando origem aos complexos HLA-peptídeo (Yang *et al.*, 1996). Peptídeos maiores precisam ser clivados pelas aminopeptidases associadas ao processamento de antígeno do RE (ERAAP) e assim adquirir o tamanho adequado para ligar a fenda de ligação a peptídeo da molécula HLA (Yewdell *et al.*, 2003; Parcej and Tampé, 2010; Blees *et al.*, 2017). Dada a ligação peptídica bem-sucedida, o complexo HLA-peptídeo se dissocia do complexo de

carregamento de peptídeos para ser exportado à superfície celular através da via secretora padrão. Na superfície celular, o complexo HLA-peptídeo será reconhecido pelos Receptores das Células TCD8+ (TCR, do inglês, *T-cell Receptor*) (Parcej and Tampé, 2010) (Figura 2).

Os peptídeos apresentados pelas moléculas HLA de classe I são, em sua maioria, citosólicos, oriundos de proteínas virais, de microrganismos intracelulares, da degradação de proteínas próprias ou ainda proteínas de células transformadas (Alberts *et al.*, 2015; Rock *et al.*, 2016). Uma vez que os receptores dos linfócitos TCD8+ tenham reconhecido peptídeos não-próprios via MHC, eles serão ativados, se diferenciarão em linfócitos efetores e podem então reconhecer qualquer célula alvo expressando os mesmos complexos HLA-peptídeo (Alberts *et al.*, 2015). Em conjunto as interações HLA/peptide-TCR, HLA-CD8, moléculas de adesão e proteínas de sinalização intracelular, na interface entre as duas células (célula infectada e linfócito T), formam uma sinapse imunológica. A resposta citotóxica dos linfócitos T contra a célula infectada/alterada ocorre por meio da liberação de proteases no espaço sináptico ou por induzir a célula alvo à apoptose (Yang *et al.*, 1996; Alberts *et al.*, 2015).

O conjunto HLA-peptídeo interage com os TCRs, porém, os resíduos polimórficos no topo das alfas hélices das moléculas HLA são a base para especificidade de TCRs para uma forma alélica particular de MHC associada a um peptídeo antigênico (fenômeno chamado restrição de MHC) (Rock *et al.*, 2016). Isso ocorre porque as moléculas de HLA estão diretamente envolvidas na seleção do repertório de células T durante a fase de maturação no timo. O processo de seleção de linfócitos (timócitos) garante um repertório de células T funcionais, ou seja, capazes de interagir com moléculas de HLA do indivíduo. Apenas linfócitos portando TCRs que interagem com MHC-peptídeo próprio são selecionados (processo conhecido como seleção positiva) (Klein *et al.*, 2014). Além disso, o processo de maturação tímica fornece um repertório de linfócitos autotolerantes, impedindo que linfócitos portando TCRs com forte afinidade por MHC-peptídeo próprio sejam liberados na circulação periférica (seleção negativa) (Sebzda *et al.*, 1999; Wieggers *et al.*, 2011; Klein *et al.*, 2014). A falha nesse processo é considerada um dos principais mecanismos associados ao desenvolvimento de doenças autoimunes. Entretanto linfócitos com potencial autorreativo que escapam à seleção negativa e adentram a circulação periférica podem ainda ser inativados, suprimidos ou eliminados pela tolerância periférica para não desencadear autoimunidade (Ohashi, 2002; Abbas *et al.*, 2011; Wieggers *et al.*, 2011).

Assim, além da capacidade de apresentar antígenos, as moléculas HLA também são responsáveis pela seleção de células T aptas a distinguir adequadamente entre antígenos

próprios e não próprios. Deste modo, as moléculas HLA tem importante implicação em contextos fisiológicos e patológicos, relacionadas com autotolerância e autoimunidade, defesa contra patógenos, tumores e rejeição a transplantes alógenos.

A habilidade das moléculas HLA de classe I clássicas de apresentar diversos peptídeos diferentes é devido à grande variabilidade existente nos domínios de ligação ao peptídeo (codificada pelos éxons 2 e 3). De fato, os resíduos polimórficos da fenda de ligação modificam a especificidade de ligação a peptídeos, porque modificam os bolsões (do inglês, *pockets*) que são preenchidos pelas cadeias laterais de aminoácidos dos peptídeos antigênicos (os chamados resíduos âncoras). Cada peptídeo antigênico tem poucos resíduos âncoras, que são aminoácidos específicos para ligação aos bolsões da molécula HLA. Os demais aminoácidos próximos aos resíduos âncoras, que interagem com a fenda, podem variar entre todas as possibilidades dos 20 aminoácidos conhecidos (Rock *et al.*, 2016) (Figura 3). Isso permite que o repertório de peptídeos apresentados por uma molécula de HLA seja bastante vasto.

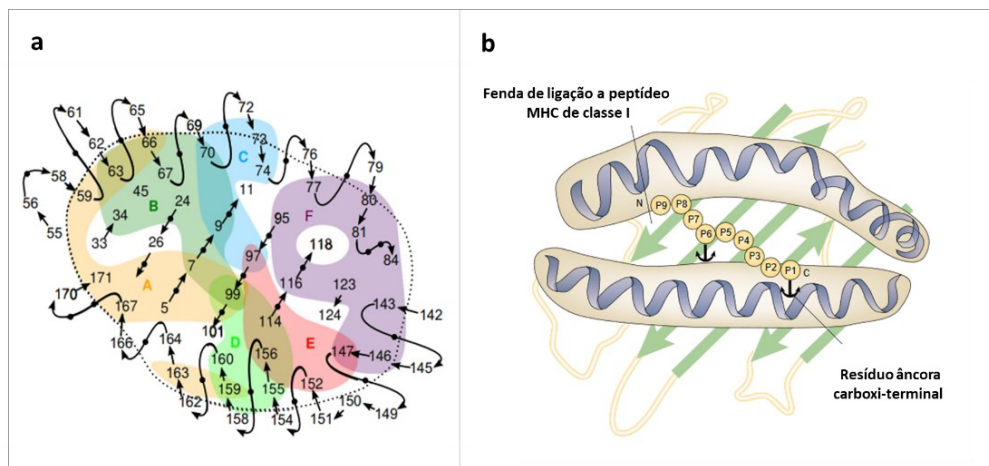


Figura 3. Visão geral esquemática da fenda de ligação a peptídeos. **a)** Painel apresentando os 'pockets' de A a F na fenda de ligação a peptídeo de uma molécula de MHC de classe I (diferenciados por um mapa de cores), com as posições dos AA na molécula (Adaptado de Van Deutekom e Keşmir, 2015). **b)** Fenda de ligação do MHC de classe I com peptídeo antigênico do 'pocket', o peptídeo é um nonâmero com dois resíduos âncoras nas posições 1 e 6 (Adaptado de (Kloetzel, 2001).

De tal modo, a substituição de um único aminoácido na fenda de ligação ao peptídeo pode ter efeito no repertório de peptídeos compatíveis com uma determinada molécula HLA, porém depende da posição substituída e das propriedades físico-químicas dos aminoácidos envolvidos (Van Deutekom and Keşmir, 2015; Kaur *et al.*, 2017). De forma semelhante, a mudança de um resíduo âncora de um epítipo antigênico pode impedir sua ligação e apresentação por uma determinada molécula HLA (Rock *et al.*, 2016).

Van Deutekom e Keşmir (2015) usaram um método *in silico* para estudar o efeito de mutações únicas no repertório de peptídeos de várias moléculas HLA-A e HLA-B e mostraram

que as posições que mais alteram a ligação peptídica são as mais polimórficas. No entanto, as posições que raramente variam entre as moléculas HLA parecem não interferir no repertório de peptídeos apresentados. O mesmo estudo demonstrou que o efeito, no repertório de peptídeo ligantes, causado por uma substituição na fenda peptídica é significativamente maior em moléculas de HLA-B em comparação com as moléculas de HLA-A. Considerando que uma nova molécula HLA com um repertório de peptídeos alterado seja mantida na população, caso proporcione vantagens adaptativas, novas moléculas de HLA-B poderiam evoluir facilmente através de mutações pontuais e, portanto, alcançar uma maior diversidade a nível populacional (Van Deutekom and Keşmir, 2015).

Além da ampla gama de peptídeos que podem ser apresentados por cada molécula HLA de classe I individualmente, especialmente os genes clássicos que são muito polimórficos (ver Tabela 1), o repertório é ainda maior quando consideramos a redundância dessas moléculas na superfície celular. Cada indivíduo pode ter de três a seis moléculas HLA de classe I clássicas, considerando a co-expressão dessas moléculas e dependendo dos alelos herdados (Rock *et al.*, 2016).

A maior diversidade alélica dos genes HLA aumenta a capacidade de apresentação antigênica, favorecendo uma melhor adaptabilidade em um contexto de resposta a patógenos e combate a células tumorais. Todavia, essa elevada diversidade alélica pode representar um problema para a compatibilização genotípica entre dois indivíduos (doador e receptor) no contexto do transplante alogênico. Rejeições a enxertos já foram associados a pequenas diferenças (variações) na fenda ligadora de peptídeos das moléculas HLA dos doadores e receptores (Choo, 2007; Ayala García *et al.*, 2012; Tiercy, 2014).

Ainda, é válido ressaltar que apesar da elevada diversidade nucleotídica dos genes HLA estar associada a uma melhor capacidade de apresentação antigênica no contexto de resposta a patógenos e células tumorais, os vírus e as células tumorais podem evadir a resposta imunológica por meio de mecanismos que afetem o processo de apresentação antigênica (Seliger *et al.*, 2006).

Além da modificação de resíduos âncoras para evitar a ligação à molécula HLA, como já citado, os patógenos podem também: (a) escapar da degradação proteasomal, evitando assim a produção de epítopos imunogênicos (Seliger *et al.*, 2006); (b) codificar proteínas que inibem o transportador de peptídeos (TAP) (Van Hall *et al.*, 2007); (c) codificar proteínas que induzem a retenção de moléculas de MHC I no RE-Golgi (Ziegler *et al.*, 2000); (d) codificar proteínas que deslocam as moléculas MHC de classe I do RE ao citosol, onde são rapidamente degradadas

pelo proteassoma (Wiertz, Jones, *et al.*, 1996; Wiertz, Tortorella, *et al.*, 1996) e/ou (e) reduzir a expressão MHC classe I por induzir endocitose dessas moléculas, acúmulo em vesículas endossomais e degradação (Schwartz *et al.*, 1996).

A ausência ou redução da expressão das moléculas HLA de classe I em células tumorais também tem sido frequentemente observada em muitas neoplasias malignas humanas e representa um importante mecanismo de escape da resposta imune (Campoli *et al.*, 2002; Aptsiauri *et al.*, 2007). A diminuição da expressão de componentes da maquinaria de processamento de antígenos da via HLA classe I (*e.g.*, chaperonas e TAP) geralmente têm um impacto negativo no curso clínico dos tumores. Em um número substancial de neoplasias prostáticas, a maioria dos componentes da maquinaria de processamento de antígenos foi encontrada com expressão reduzida, com conseqüente redução ou ausência de HLA de classe I na superfície celular. Este fenômeno estaria relacionado com progressão e recorrência da doença (Seliger *et al.*, 2010).

Nesse sentido, as células NK são fundamentais na resistência do hospedeiro a infecções virais ou células tumorais, porque estão entre as primeiras células a detectar a liberação de citocinas pró-inflamatórias, bem como a redução da expressão de moléculas MHC de classe I na superfície de células infectadas (Jonjic *et al.*, 2008). O reconhecimento de moléculas HLA por receptores KIR (do inglês, *Killer cell Immunoglobulin like Receptor*) ativatórios e inibitórios promove um equilíbrio de sinais que regulam a atividade de células NK (Parham, 2004; Parham, 2005; Augusto and Petzl-Erler, 2015). Assim, a expressão reduzida de HLA na superfície celular pode induzir a ativação de células NK por ‘*missing self*’ (ausência do próprio) (Parham, 2004; Parham, 2005; Jonjic *et al.*, 2008). Porém, os vírus adquiriram numerosos mecanismos destinados a subverter ou evadir a vigilância imune das células NK (Jonjic *et al.*, 2008; Béziat *et al.*, 2017). As células tumorais também podem aumentar a expressão de HLA classe I não clássicos como mecanismo de escape da resposta imune, visto que estas moléculas desempenham funções imunomodulatórias por meio de ligações a receptores específicos em várias células do sistema imunológico (Paul *et al.*, 1998; Ibrahim *et al.*, 2001; Wiendl *et al.*, 2002; Bossard *et al.*, 2012).

Este cenário evidencia uma constante co-evolução entre patógenos e sistema imune, que no caso das moléculas HLA precisa favorecer a constante mudança no repertório de peptídeos que podem ser apresentados, o escape à regulação da expressão mediada por proteínas dos patógenos e manutenção de certa conservação em domínios importantes para interação com co-receptores, receptores imunomodulatórios e moléculas da via de processamento de antígenos.

HLA-C e suas características particulares

O gene *HLA-C*, alvo deste estudo, pertence à classe de genes mais polimórficos entre os genes *HLA*. Apesar disso, o estudo da variabilidade desse gene foi negligenciado por um longo tempo e *HLA-C* foi considerado pouco polimórfico comparado aos outros genes de classe I clássicos. Atualmente, conhecemos cerca de 5.709 alelos que codificam 3.470 proteínas *HLA-C* diferentes (Tabela 1), variabilidade que acompanha a de outros genes clássicos como *HLA-A* e *HLA-B*. Porém, algumas características específicas de *HLA-C*, como a restrição peptídica e menor expressão na superfície celular, ainda o diferencia dos seus correlatos (Blais *et al.*, 2011).

A menor expressão celular de *HLA-C* tem sido associada a vários fatores, tais como: (a) instabilidade do RNA mensageiro (Mccutcheon *et al.*, 1995); (b) retenção prolongada da molécula *HLA-C* no interior do retículo endoplasmático por associação ineficiente da cadeia α pesada de *HLA-C* com a $\beta 2m$ (Neefjes and Ploegh, 1988), por interações estáveis da molécula *HLA-C* com TAP ou tapasina ou ainda por causa da restrição do repertório de peptídeos (Neisig *et al.*, 1998; Sibilio *et al.*, 2008), que é consequência da maior conservação nos domínios de ligação de peptídeos de *HLA-C* (Zemmour and Parham, 1992; Neisig *et al.*, 1998; Turner *et al.*, 1998) (Figura 4).

O motivo KIRV (resíduos nas posições 66, 67, 69 e 76) na α hélice de *HLA-C* é extremamente conservado na maioria de seus grupos alélicos, exceto para algumas proteínas do grupo *HLA-C*07* e *C*15*, que tem uma mudança de aminoácido na posição ^{Lys66}Asn (IMGT database, 3.38.0) (Robinson *et al.*, 2015). Esse motivo conservado é característico apenas de *HLA-C*, não está presente em outros genes *HLA* de classe I (exceto *HLA-B*46*) e parece favorecer uma rigorosa seleção/restrrição do repertório de peptídeos. Além disso, KIRV confere maior associação de *HLA-C* às chaperonas TAP e tapasina e acúmulo intracelular por deficiência em montagem ou instabilidade pós montagem da molécula (Sibilio *et al.*, 2008).

Um estudo recente com dois alelos de *HLA-C* (*HLA-C*05:01* e *HLA-C*07:02*) mostrou a influência dos domínios de ligação a peptídeo na expressão de superfície celular. *HLA-C*05*, mesmo com um promotor mais fraco que *HLA-C*07* (avaliado por expressão do gene repórter luciferase), apresentou uma expressão duas vezes maior que *HLA-C*07* na superfície celular. Ainda, a maior expressão de *C*05* em relação a *C*07* não foi consequente à regulação diferencial mediada pela porção 3' não traduzida (3'NT), uma vez que para descartar essa influência fora utilizada uma região do gene H-2K^b (MHC de classe I murino) do íntron 3 até a porção 3'NT, tornando esse segmento idêntico para ambos os alelos. O mesmo estudo mostrou que a fenda peptídica de *HLA-C*05* é mais plana enquanto a de *HLA-C*07* é mais profunda e

estreita. Essa característica permitia *HLA-C*05* ligar um número de peptídeos três vezes maior que *HLA-C*07*, indicando que quanto maior a restrição a peptídeos, maior o acúmulo intracelular e menor a expressão na superfície celular (Kaur *et al.*, 2017). Estes achados fornecem informações sobre a complexidade da regulação da expressão de *HLA-C* que depende de fatores pré-transcricionais (importante influência da região promotora), pós-transcricionais (regulação diferencial por ligação de microRNAs a região 3'NT) e pós-traducionais (montagem, estabilidade e ligação a peptídeos), sendo a última, dependente da sequência do alelo codificador.

Contudo, outra característica de *HLA-C* é a imunomodulação, função característica de genes não-clássicos (Klein and Sato, 2000; Lepin *et al.*, 2000; Pietra *et al.*, 2009; Blais *et al.*, 2011; Donadi *et al.*, 2011). O gene *HLA-C* é o único entre os genes *HLA* de classe I clássicos expresso na interface materno-fetal, junto com os genes não clássicos *HLA-E*, *HLA-F* e *HLA-G* (Hackmon *et al.*, 2017). Vários estudos tem mostrado que combinações particulares de alelos *HLA-C* fetais e *KIR* inibitórios presentes nas células *NK* maternas são fundamentais para imunotolerância materno-fetal e sucesso da gravidez (Trowsdale and Moffett, 2008; Chazara *et al.*, 2011).

O *HLA-C* é considerado o mais importante entre os genes clássicos para a biologia das células *NK* (Parham, 2005). A variabilidade encontrada em algumas porções da molécula *HLA-C*, em especial nas α hélices, pode modificar a afinidade com receptores *KIR* (Boyington and Sun, 2002). A presença de um dimorfismo na posição 80 da sequência de aminoácidos na molécula *HLA-C* define dois grupos de ligantes a receptores *KIRs*: *HLA-C1* (Asn80), ligante dos *KIRs* inibitórios *KIR2DL2* e *KIR2DL3*, e *HLA-C2* (Lys80), ligante do inibitório *KIR2DL1* e do ativador *KIR2DS1*. O *KIR* ativador *KIR2DS4* pode interagir com ambos os alótipos de *HLA-C* (Parham, 2005; Blais *et al.*, 2011; Parham and Moffett, 2013). A capacidade de interagir com vários *KIRs* inibitórios e ativadores pode ser um dos motivos por que determinados alelos de *HLA-C* ou combinações específicas de alelos de *HLA-C* e *KIR* têm sido associados a um grande número de doenças autoimunes, alérgicas, inflamatórias, cânceres, abortos recorrentes, pré-eclâmpsia (Kulkarni *et al.*, 2008; Kuśnierczyk, 2013), aloreatividade pós-transplantes e doença do enxerto *versus* hospedeiro (GVHD, do inglês *Graft versus Host disease*) (Parham, 2005).

A função imunomodulatória de *HLA-C* também aumentou as especulações sobre a importância dos níveis de expressão de *HLA-C* em infecções. A observação de que a proteína *Nef* do Vírus da Imunodeficiência Humana (HIV, do inglês, *Human Immunodeficiency Virus*)

reduz a expressão de HLA-A e HLA-B, mas não de HLA-C e HLA-E (Le Gall *et al.*, 1998; Cohen *et al.*, 1999; Williams *et al.*, 2002) (Figura 4) foi considerada um mecanismo de escape do vírus. A diminuição nos níveis de expressão das moléculas HLA-A e HLA-B permitiria ao vírus HIV escape da vigilância imunológica, neste caso, do reconhecimento pelas CTL, enquanto que a manutenção da expressão de HLA-C e HLA-E poderia favorecer a inibição de células *NK* (Williams *et al.*, 2002; Blais *et al.*, 2011).

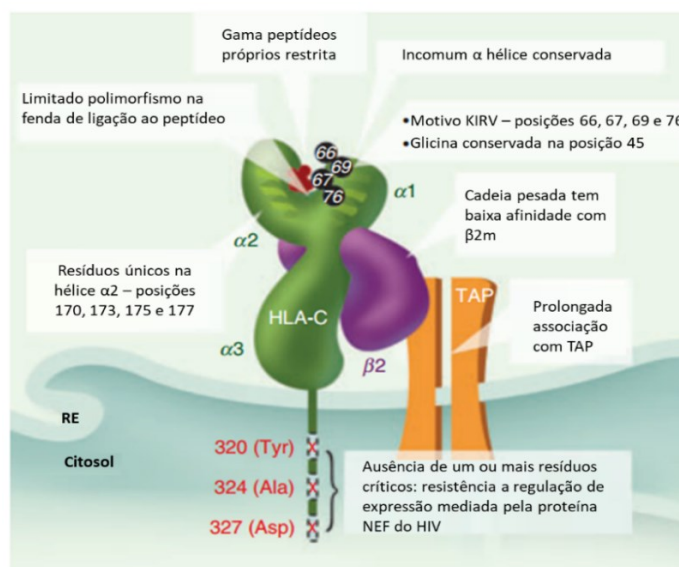


Figura 4. Representação esquemática das características distintas da molécula HLA-C (Adaptado de Blais; Dong and Rowland-Jones, 2011).

Contrariando a visão de que a expressão de HLA-C favoreceria a infecção por HIV, estudos de variabilidade genética em porções regulatórias de *HLA-C* mostrou que variantes favorecendo a expressão do gene estão associadas ao controle da infecção por HIV. Um polimorfismo de nucleotídeo único (SNP) a -35 Kilobases (Kb) de distância do gene *HLA-C* (rs9264942 C) está associada a maior expressão de HLA-C e a uma menor carga viral (Fellay *et al.*, 2007; Thomas *et al.*, 2009). Esse polimorfismo foi ainda encontrado em desequilíbrio de ligação com outra variante na região 3'NT (rs67384697), que também favorece a expressão de HLA-C, uma variante do tipo inserção/deleção, com a deleção afetando a ligação do microRNA miR-148a (Thomas *et al.*, 2009; Kulkarni *et al.*, 2011; Blais *et al.*, 2012; Chen *et al.*, 2012). Ainda, Apps e colaboradores (2013) mostraram que o aumento da expressão de HLA-C em pacientes infectados pelo HIV, independentemente do alótipo, pode aumentar a resposta citotóxica de linfócitos T, apontando um efeito protetor da alta expressão de HLA-C (Apps *et al.*, 2013).

A contribuição de determinados alelos de *HLA-C* no controle de doenças infecciosas, como nos casos de infecções pelo vírus da hepatite C (HCV, do inglês *Hepatitis C Virus*),

citomegalovírus (Kondo *et al.*, 2004; Lauer *et al.*, 2004; Bihl *et al.*, 2005), HIV (Bihl *et al.*, 2005; Makadzange *et al.*, 2010) e vírus Epstein–Barr (Ito *et al.*, 2007), também tem sido descrita. Estes achados parecem estar relacionados a uma apresentação antígeno-específica por determinadas variantes de *HLA-C*. Além disso, respostas eficazes contra infecções virais podem ser mediadas por interações HLA-C/KIR. O receptor ativatório KIR2DS2 de células *NK* é capaz de ligar a moléculas HLA-C*01:02 associadas a peptídeos virais derivados de regiões conservadas de RNA helicases da superfamília dos flavivírus (vírus da dengue, Zika e da febre amarela) e do HCV, levando a ativação de células *NK* (Naiyer *et al.*, 2017). É possível que todas essas respostas sejam impactadas positivamente pela maior expressão celular de HLA-C.

De modo oposto, a maior expressão de HLA-C em doenças autoimunes, como doença de Crohn (Apps *et al.*, 2013), ou maior expressão de um alótipo específico como HLA-C*06 em psoríase, parece ter um efeito deletério (Majorczyk *et al.*, 2014). Nesses casos, provavelmente, por aumento na expressão de alelos comprometidos com a apresentação antigênica e ativação de CTLs contra peptídeos próprios. A hipótese é reforçada pelo achado de efeitos epistáticos entre *HLA-C* e *ERAP1* (ou *ERAAP*) na psoríase. A susceptibilidade a psoríase é aumentada em indivíduos que possuem polimorfismos em *ERAP1* somente quando também apresentam o alelo de risco *HLA-C* (Strange *et al.*, 2010). A ERAP1 desempenha um papel importante no processamento de peptídeos de MHC classe I que serão apresentados a CTLs. De modo semelhante, também foi identificada relação entre os genes ERAP, KIR e HLA-C na suscetibilidade ao aborto espontâneo recorrente (Wilczyńska *et al.*, 2019). Isso porque a mudança no repertório de peptídeos gerados por ERAP pode alterar o complexo HLA-peptídeo e em consequência, alterar tanto a interação HLA-peptídeo-KIR como a interação HLA-peptídeo-TCR.

Em transplantes alogênicos de medula óssea com incompatibilidades permissivas para o gene *HLA-C*, entre doador e receptor, a aloreatividade pós-transplante pode ser impactada pelo nível de expressão da molécula HLA-C. Desta forma, como o nível de expressão do alótipo de HLA-C incompatível pode influenciar o aloreconhecimento por CTL, alelos *HLA-C* de baixa expressão podem representar melhores chances para transplantes com incompatibilidades permissivas (Tiercy, 2014). Assim, é possível que a complexidade dos efeitos de HLA-C, em transplantes e em diversas patogenias, exceda a especificidade de ligação a peptídeo e inclua os diferentes níveis de expressão do gene, que podem modular o potencial da resposta imune.

Entretanto, os mecanismos relacionados com a regulação da expressão não estão completamente elucidados, assim como a variabilidade inerente as regiões regulatórias

(promotora e 3' não-traduzida) do gene *HLA-C* e o impacto desta variabilidade na modulação da expressão gênica. Para a região 3'NT, os poucos estudos existentes apontam para uma influência de um polimorfismo na ligação do miR-148a-3p (Thomas *et al.*, 2009; Chen *et al.*, 2012; Apps *et al.*, 2013). Um outro estudo explorou a influência de variações na 3'NT relacionadas com os alelos *C*07:02* e *C*06:02* na regulação do *HLA-C*, em linhagens celulares homocigotas (Kulkarni *et al.*, 2011).

Em relação ao promotor do *HLA-C*, alguns estudos mostraram variantes que podem afetar a ligação de fatores de transcrição, como rs2395471 A/G, que está localizado em um motivo consenso de ligação do fator de transcrição Oct1 (Vince *et al.*, 2016) e duas variantes regulatórias que influenciam a resposta ao TNF- α (-196A>G, rs2524094) e IFN- γ (-166_-163delTCT, rs10657191) (Hundhausen *et al.*, 2012). Outro estudo explorando a variabilidade na região promotora de *HLA-C*, porém em linhagens celulares homocigóticas para HLA classe I, mostrou que alguns grupos de alelos *HLA-C* tais como *HLA-C *03*, *-C *07* e *-C *17* têm variações em um ou mais motivos do 'core' promotor (Ramsuran *et al.*, 2017). Esses três grupos alélicos *HLA-C* também apresentam a variante na região 3'NT com afinidade para o miR-148a-3p e o alelo rs2395471G na região promotora, ambos associados a baixos níveis de expressão de HLA-C na superfície celular (Vince *et al.*, 2016; Ramsuran *et al.*, 2017).

Embora alguns estudos tenham demonstrado associações entre variantes regulatórias, alelos codificadores e perfis diferenciais de expressão, nenhum estudo explorou a variabilidade das regiões regulatórias e codificadora de *HLA-C* em uma amostra populacional. Geralmente, os segmentos regulatórios e íntrons de *HLA-C* são desconsiderados na maioria dos estudos abordando sua variabilidade. Ainda, o banco de dados IPD-IMGT/HLA, por exemplo, não contempla a variabilidade do *HLA-C* além do promotor proximal e 5' não traduzida (5'NT), e a maioria dos alelos já descritos para *HLA-C* carecem de sequências de íntrons, promotor e 3'NT.

Diante do exposto, é notória a importância dos níveis de expressão da molécula HLA-C para a realização de suas funções, apresentação antigênica e imunomodulação, e que a regulação da expressão é dependente da variabilidade tanto das regiões regulatórias quanto da região codificadora do gene. Portanto, a avaliação da variabilidade de todo o gene é de extrema importância para o entendimento da biologia e função deste gene paradoxal, que combina características de genes HLA clássicos e não clássicos.

Seleção Natural atuando sobre os genes HLA

Estudos de variabilidade genética permitem que o grau de polimorfismo do genoma seja usado como um indicador para identificação de regiões gênicas submetidas a diferentes regimes

de seleção natural (Satta *et al.*, 1998; Meyer *et al.*, 2017). Seleção natural atuando sobre uma região gênica pode, em nível populacional, elevar a frequência de alelos que aumentam o *fitness* de um indivíduo (seleção positiva ou balanceadora). Mas, diminuir a frequência ou eliminar variantes que reduzam o *fitness* do indivíduo (seleção negativa ou purificadora). Regiões gênicas sob seleção direcional (positiva ou negativa/purificadora) apresentam uma taxa reduzida de polimorfismos (Fu and Akey, 2013), enquanto aquelas submetidas à seleção balanceadora tendem a manter um grande número de variantes adaptativas na população, em frequências intermediárias, aumentando o grau de variabilidade e a heterozigose (Key *et al.*, 2014). Sem a ação da seleção natural, esses *loci* evoluem segundo a teoria da neutralidade, em que mutações que não afetam o *fitness* do organismo podem aumentar ou diminuir sua frequência em uma população por meio de deriva genética (Satta *et al.*, 1998).

Seleção balanceadora é um termo amplo que engloba vários regimes seletivos responsáveis pela manutenção de variantes genéticas adaptativas, aumentando a diversidade genética, entre eles: (a) Vantagem do heterozigoto (ou sobredominância), que ocorre quando o '*fitness*' do genótipo heterozigoto é maior que de ambos os alelos em homozigose, mantendo dois ou mais alelos indefinidamente na população; (b) a seleção variável ou flutuação (em tempo e espaço) mantém níveis de diversidade semelhantes ao esperado sobre vantagem do heterozigoto, porém, como o valor adaptativo dos genótipos podem variar ao longo do tempo e espaço, as frequências genotípicas também podem mudar; e (c) vantagem do alelo raro ou seleção dependente de frequência negativa, que pode ser explicada pela co-evolução antagônica entre patógenos e hospedeiros, uma vez que há forte seleção em patógenos para superar a resistência dos alelos MHC mais comuns, os alelos raros que podem ainda oferecer resistência ao patógeno tem uma vantagem seletiva e tendem a aumentar sua frequência (Meyer and Thomson, 2001; Spurgin and Richardson, 2010; Meyer *et al.*, 2017).

Vários autores apontaram para uma incontestável assinatura de seleção balanceadora em genes HLA de classe I clássicos, que seria a explicação para o alto grau de polimorfismos observados, excesso de variantes não-sinônimas e elevado desequilíbrio de ligação nesses genes (Hughes and Yeager, 1998; Meyer and Thomson, 2001; Garrigan and Hedrick, 2003; Sanchez-Mazas, 2007; Spurgin and Richardson, 2010; Meyer *et al.*, 2017). De fato, somente a ação de seleção natural poderia explicar a extrema diversidade genética de genes HLA de classe II e classe I clássicos, como também a alta conservação dos genes não-clássicos.

Seleção balanceadora atuando sobre determinado segmento de DNA pode se estender, por efeito carona, para regiões próximas até uma distância de pelo menos 10kb (Satta *et al.*,

1998). Porém, nos genes HLA, as assinaturas de seleção podem ser diferentes em segmentos muito próximos. No gene *HLA-G*, por exemplo, uma provável seleção purificadora atua sobre a região codificadora, limitando a variabilidade deste segmento (Mendes-Junior *et al.*, 2013). No entanto, as regiões regulatórias do gene *HLA-G* (promotora e 3' *NT*) parecem estar sob ação de seleção balanceadora, elevando a heterozigose de haplótipos divergentes (Tan *et al.*, 2005; Castelli *et al.*, 2011; Castelli *et al.*, 2014; Sabbagh *et al.*, 2014; Gineau *et al.*, 2015).

Estudos sobre a variabilidade do gene *HLA-E* também apontam para perfis de seleção diferentes atuando em diferentes segmentos gênicos. O gene *HLA-E* apresenta um perfil de seleção balanceadora na região codificadora da fenda peptídica, especialmente em um ponto de variação não sinônima no éxon 3, mas o éxon 4 (que codifica $\alpha 3$) e regiões não-codificadoras apresentaram assinaturas de seleção purificadora. Adicionalmente, a região 3' *NT* de *HLA-E* também é muito conservada (Veiga-Castelli *et al.*, 2012; Felicio *et al.*, 2014). De forma semelhante, alta conservação foi descrita para o gene *HLA-F* (Lima *et al.*, 2016).

Mesmo HLA-A que apresenta assinatura de seleção balanceadora em quase todo segmento gênico (região codificadora e regiões regulatórias), apresenta alta conservação e sinais de seleção purificadora no éxon 4, que codifica $\alpha 3$ (Lima *et al.*, 2019).

Em geral, a assinatura de seleção balanceadora observada para genes HLA de classe I clássicos foi detectada na região que codifica a fenda de ligação aos peptídeos (éxons 2 e 3), cuja manutenção da alta diversidade alélica aumentaria a capacidade de apresentar um maior repertório de peptídeos e conseqüentemente, o 'fitness' na resistência contra patógenos (Meyer and Thomson, 2001; Sanchez-Mazas, 2007; Cagliani and Sironi, 2013; Dos Santos Francisco *et al.*, 2015; Bitarello *et al.*, 2016; Buhler *et al.*, 2016).

Ainda, levando em conta a vantagem do heterozigoto, a maior divergência entre os alelos (vantagem do alelo divergente, DAA) de um heterozigoto aumentaria a capacidade de apresentar um conjunto maior de peptídeos que aqueles heterozigotos com alelos mais próximos do ponto de vista de sequência (Buhler *et al.*, 2016). Porém, a grande proporção de homozigotos observada em um ou mais *loci* HLA em estudos populacionais, em pequenas e grandes populações, indica que mecanismos adicionais, além da vantagem do heterozigoto, estão envolvidos para garantir uma adequada proteção imune nessas populações (Buhler *et al.*, 2016).

Esses achados elevaram o interesse em abordagens que investiguem os padrões de variação genética de HLA observados em diferentes populações, em relação à funcionalidade das moléculas (especificidade de ligação ao peptídeo). Os resultados desse tipo de estudo confirmaram que os resíduos polimórficos não são distribuídos aleatoriamente dentro da fenda

de ligação a peptídeo, mas seguem um padrão de maior diversidade em sítios que definem o repertório de peptídeos (Dos Santos Francisco *et al.*, 2015; Van Deutekom and Keşmir, 2015; Bitarello *et al.*, 2016; Buhler *et al.*, 2016). Assim, Buhler e colaboradores (2016) propuseram o modelo de seleção assimétrica divergente dos genes HLA de classe I, que sugere que a falta de diversidade em um gene HLA, como observado em algumas populações, é contrabalanceada por propriedades complementares de ligação ao peptídeo das moléculas codificadas por outros genes. No entanto, Buhler observou que este modelo parece ser robusto para os genes *HLA-A* e *HLA-B*, porém *HLA-C* não aumenta significativamente o potencial de ligação de peptídeo dos HLA de classe I, sugerindo que este *locus* assume um papel mais importante em suas funções relacionadas a KIR (Buhler *et al.*, 2016). Além disso, é possível que a adaptação envolvendo genes HLA seja poligênica e que haplótipos (estendidos) carregando combinações de alelos HLA que apresentam maior número de peptídeos sejam favorecidos (Meyer *et al.*, 2017). Em suporte a essa hipótese, já foi demonstrado que alelos HLA em forte desequilíbrio de ligação, em média, tem menor sobreposição do repertório de peptídeo (Penman *et al.*, 2013).

Embora seja consenso que a alta diversidade do MHC é fundamental para defesa imune contra patógenos, é bem conhecida a associação entre diversos genes localizados nesse complexo e doenças humanas, entre elas, doenças autoimunes como diabetes tipo I, psoríase, doença de Crohn, doença celíaca, esclerose múltipla, lúpus, artrite reumatoide e espondilite anquilosante (Trowsdale and Knight, 2013; Matzaraki *et al.*, 2017).

O delicado equilíbrio entre variantes genéticas que favoreçam o controle de infecções, mas que podem aumentar a suscetibilidade à doença autoimune, pode ser considerado um importante paradigma em genética evolutiva (Chen *et al.*, 2012). De fato, à partir de uma perspectiva evolutiva, é intrigante a existência de condições auto-imunes, que reduzem as chances de sobrevivência e reprodução de um indivíduo, associadas a alelos HLA relativamente comuns (Meyer *et al.*, 2017). Contudo, isso pode ser resultado da seleção de alelos que conferem resistência a doenças infecciosas e ao mesmo tempo estão associadas a condições auto-imunes (Corona *et al.*, 2010; Abadie *et al.*, 2011; Sams and Hawks, 2014; Meyer *et al.*, 2017). Por exemplo, alelos *HLA-C* relacionados com psoríase e doença de Crohn também estão relacionados a proteção contra doenças infecciosas, como infecção pelo HIV (Blais *et al.*, 2012; Chen *et al.*, 2012; Apps *et al.*, 2013; Kulkarni *et al.*, 2013; Majoreczyk *et al.*, 2014).

Algumas interações entre HLA e KIR (especialmente *HLA-C/KIR*) também têm sido associadas a diversas doenças autoimunes (Parham, 2005; Blais *et al.*, 2011; Augusto and Petzl-Erler, 2015). Há evidências de co-evolução entre os complexos KIR e HLA, e é possível que

seleção balanceadora esteja mantendo a diversidade excepcional do sistema KIR-HLA, aparentemente evoluindo sob a pressão proporcionada pelas exigências concorrentes da reprodução e da imunidade inata e adaptativa (Augusto and Petzl-Erler, 2015).

O gene *HLA-C* é altamente polimórfico e ao mesmo tempo mantém vários motivos conservados em sua fenda de ligação a peptídeo, que restringe seu repertório de peptídeos e regula a sua expressão na superfície celular. Adicionalmente, a molécula HLA-C desempenha a função de apresentação antigênica a CTLs (como HLA-A e HLA-B), mas também tem um papel imunomodulatório importante semelhante aos genes não clássicos, ambas situações sendo impactadas pelos níveis de expressão do gene supracitado. Levando em consideração o papel crucial que o gene *HLA-C* desempenha na resposta imunitária, ressalta-se a importância dos estudos de variabilidade no intuito de esclarecer de que forma variações em sua sequência se relacionam com diferentes perfis de expressão. Além disso, estudos que avaliem os perfis de seleção natural em cada segmento de *HLA-C* é de fundamental importância para esclarecer de que maneira a variabilidade desse gene tem sido moldada evolutivamente. Desta forma, o presente estudo busca esclarecer se o perfil de seleção balanceadora se estende por toda região codificadora e segmentos regulatórios de *HLA-C*, ou se estes estão sob diferentes regimes de seleção natural, permitindo assim um melhor entendimento acerca da biologia do gene *HLA-C*.

Testes de desvio de neutralidade e demografia das populações estudadas

Para testar seleção natural atuando a nível de DNA em amostras populacionais, é possível utilizar testes estatísticos de desvio de neutralidade. Esses testes utilizam uma abordagem teórica baseada em simulações de coalescência para detectar desvios significativos de um modelo de neutralidade (Excoffier and Lischer, 2010; Gineau *et al.*, 2015).

Neste estudo as assinaturas de seleção presentes no gene *HLA-C* foram avaliadas por meio de três testes de desvio da neutralidade: (a) teste de Ewens-Watterson (Ewens, 1972; Watterson, 1978), que compara a homozigose observada na população em estudo com a homozigose esperada sob expectativas de neutralidade e em equilíbrio de Hardy-Weinberg, para um mesmo tamanho amostral ($2n$) e com o mesmo número de alelos únicos. O excesso de homozigose indica a possibilidade de uma seleção direcionada, enquanto um excesso de heterozigose pode indicar atuação de seleção balanceadora favorecendo genótipos heterozigotos; (b) teste D de Tajima (Tajima, 1989), que é calculado considerando a diferença entre as estimativas Θ (theta), que se baseia no número médio de diferenças observadas entre pares de sequências (π) e o número de sítios de segregação (S). A significância estatística do D de Tajima é testada através da geração de amostras aleatórias sob a hipótese de neutralidade

seletiva em uma população com tamanho constante e em equilíbrio, utilizando um número determinado de simulações (Excoffier and Lischer, 2010); e (c) Razão dN/dS , que compara as taxas de substituições não-sinônimas (dN) e sinônimas (dS) e permite inferir os regimes de seleção que estão atuando sobre a região codificadora do gene em estudo (Kumar *et al.*, 2016). Segmentos exônicos em que a razão dN/dS é superior a 1 podem estar sob seleção direcional positiva (Kumar *et al.*, 2016). Porém esse critério é considerado conservador, visto que apenas alguns códons podem estar sob seleção positiva (Bitarello *et al.*, 2016), enquanto a maior parte das mutações não-sinônimas são deletérias e são eliminadas da população sem contribuir com a evolução ou o polimorfismo (Kimura, 1991) (seleção purificadora). Assim, muitas vezes esse critério ($dN/dS > 1$) não é atendido quando analisamos genes ou segmentos exônicos inteiros, mesmo quando há presença de alguns códons sob seleção positiva, tornando necessária a análise de subconjuntos de códons. Razões dN/dS inferiores a 1 podem ser um indicativo de seleção direcional negativa (Kumar *et al.*, 2016).

O uso de mais de um teste avaliando as assinaturas de seleção evita resultados falso-positivos, uma vez que a variação das frequências alélicas pode sofrer influência de efeitos demográficos, tais como efeito fundador, migração, miscigenação, contração ou expansão populacional e deriva genética (Stajich and Hahn, 2005; Gineau *et al.*, 2015). Os valores positivos de D de Tajima, por exemplo, são indicativos de um excesso de alelos de frequência intermediária (seleção balanceadora) ou contração recente da população (gargalo), enquanto os valores D negativos de Tajima indicam um excesso de alelos raros, que pode ser resultado de expansão recente da população, varredura seletiva, seleção negativa fraca ou amostra proveniente de uma população miscigenada (Stajich and Hahn, 2005).

Além do uso de três testes distintos, nós também optamos por estudar populações de dois continentes diferentes. A população brasileira é sabidamente uma população miscigenada, formada por 5 séculos de mistura interétnica entre europeus, africanos e nativos americanos (Pena *et al.*, 2011), além das migrações do leste asiático ocorridas no último século, o que poderia levar a resultados tendenciosos nos testes de desvio de neutralidade devido à sua história demográfica. Por isso, nós inserimos nesse estudo uma população do Benin, país localizado no continente africano. Essa amostra é composta por indivíduos da etnia Toffin (etimologicamente “pessoas da água”). Os Toffin estão presentes em mais de 10 aldeias, representando pequenas comunidades autônomas, fugitivos das lutas encabeçadas pelos soldados franceses durante o período pré-colonial e localizados na mesma região há séculos (Sonon *et al.*, 2018). A população do Benin é menos miscigenada que a população brasileira,

mas também tem uma história demográfica que poderia influenciar uma análise de seleção natural de sua população quando estudada individualmente. Não obstante a história demográfica de cada população, temos duas populações com diferentes histórias demográficas, localizadas em continentes distintos e ambas apresentaram o mesmo perfil de assinaturas de seleção natural atuando no gene *HLA-C*.

OBJETIVOS

Objetivo Geral

Avaliar a variabilidade e os haplótipos encontrados na região promotora, codificadora e 3' não-traduzida do gene *HLA-C* em uma amostra brasileira e beninense e detectar assinaturas de seleção natural ao longo do gene.

Objetivos específicos

- Consolidar uma metodologia para avaliação da variabilidade do gene *HLA-C* por sequenciamento massivo paralelo (ou sequenciamento de nova geração);
- Caracterizar os pontos de variação e haplótipos encontrados para o gene *HLA-C* em uma amostra do Brasil e do Benin;
- Caracterizar as amostras brasileiras utilizadas quanto a seus perfis de ancestralidade;
- Avaliar o perfil de seleção atuando ao longo de todos os segmentos do gene *HLA-C*.

REFERÊNCIAS

- ABADIE, V. et al. Integration of genetic and immunological insights into a model of celiac disease pathogenesis. **Annu Rev Immunol**, v. 29, p. 493-525, 2011. ISSN 0732-0582.
- ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. H. I. V. Moléculas do Complexo Principal de Histocompatibilidade e Apresentação do Antígeno aos Linfócitos T. In: (Ed.). **Imunologia Celular e Molecular**. 7ª. Rio de Janeiro: Elsevier, 2011. p.109-138.
- ALBERTS, B. et al. The Innate and Adaptive Immune Systems: T cells and MHC proteins. In: (Ed.). **Molecular Biology of the Cell**. sixth edition. 711 Tird Avenue, New York, NY 10017, US: Garland Science, 2015. chap. 24, p.1297-1342.
- APPS, R. et al. Influence of HLA-C expression level on HIV control. **Science**, v. 340, n. 6128, p. 87-91, Apr 05 2013. ISSN 0036-8075.
- APTSIAURI, N. et al. Role of altered expression of HLA class I molecules in cancer progression. **Adv Exp Med Biol**, v. 601, p. 123-31, 2007. ISSN 0065-2598 (Print)0065-2598.
- AUGUSTO, D. G.; PETZL-ERLER, M. L. KIR and HLA under pressure: evidences of coevolution across worldwide populations. **Hum Genet**, v. 134, n. 9, p. 929-40, Sep 2015. ISSN 0340-6717.
- AYALA GARCÍA, M. A. et al. The Major Histocompatibility Complex in Transplantation. **Journal of Transplantation**, v. 2012, p. 842141-842141, 2012.
- BERNATCHEZ, L.; LANDRY, C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? **Journal of evolutionary biology**, v. 16, n. 3, p. 363-377, 2003.
- BIHL, F. K. et al. Simultaneous assessment of cytotoxic T lymphocyte responses against multiple viral infections by combined usage of optimal epitope matrices, anti- CD3 mAb T-cell expansion and "RecycleSpot". **Journal of Translational Medicine**, v. 3, p. 20-20, 2005.
- BITARELLO, B. D.; FRANCISCO RDOS, S.; MEYER, D. Heterogeneity of dN/dS Ratios at the Classical HLA Class I Genes over Divergence Time and Across the Allelic Phylogeny. **J Mol Evol**, v. 82, n. 1, p. 38-50, Jan 2016. ISSN 0022-2844.
- BLAIS, M.-E.; DONG, T.; ROWLAND-JONES, S. HLA-C as a mediator of natural killer and T-cell activation: spectator or key player? **Immunology**, v. 133, n. 1, p. 1-7, 2011.
- BLAIS, M. E. et al. High frequency of HIV mutations associated with HLA-C suggests enhanced HLA-C-restricted CTL selective pressure associated with an AIDS-protective polymorphism. **J Immunol**, v. 188, n. 9, p. 4663-70, May 01 2012. ISSN 0022-1767.
- BLEES, A. et al. Structure of the human MHC-I peptide-loading complex. **Nature**, v. 551, n. 7681, p. 525-528, Nov 2017. ISSN 1476-4687.
- BOSSARD, C. et al. HLA-E/beta2 microglobulin overexpression in colorectal cancer is associated with recruitment of inhibitory immune cells and tumor progression. **Int J Cancer**, v. 131, n. 4, p. 855-63, 2012 Aug 15 2012. ISSN 0020-7136.
- BOYINGTON, J. C.; SUN, P. D. A structural perspective on MHC class I recognition by killer cell immunoglobulin-like receptors. **Molecular immunology**, v. 38, n. 14, p. 1007-1021, 2002.
- BUHLER, S.; NUNES, J. M.; SANCHEZ-MAZAS, A. HLA class I molecular variation and peptide-binding properties suggest a model of joint divergent asymmetric selection. **Immunogenetics**, v. 68, n. 6-7, p. 401-16, Jul 2016. ISSN 0093-7711.

- BÉZIAT, V. et al. Deciphering the killer-cell immunoglobulin-like receptor system at super-resolution for natural killer and T-cell biology. **Immunology**, v. 150, n. 3, p. 248-264, 03 2017. ISSN 1365-2567.
- CAGLIANI, R.; SIRONI, M. Pathogen-driven selection in the human genome. **Int J Evol Biol**, v. 2013, p. 204240, 2013. ISSN 2090-8032 (Print)2090-052x.
- CAMPOLI, M.; CHANG, C. C.; FERRONE, S. HLA class I antigen loss, tumor immune escape and immune selection. **Vaccine**, v. 20 Suppl 4, p. A40-5, 2002 Dec 19 2002. ISSN 0264-410X (Print)0264-410x.
- CASTELLI, E. C. et al. A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. **Molecular biology and evolution**, v. 28, n. 11, p. 3069-86, 2011.
- CASTELLI, E. C. et al. Transcriptional and posttranscriptional regulations of the HLA-G gene. **J Immunol Res**, v. 2014, p. 734068, 2014. ISSN 2314-7156.
- CHAZARA, O.; XIONG, S.; MOFFETT, A. Maternal KIR and fetal HLA-C: a fine balance. **J Leukoc Biol**, v. 90, n. 4, p. 703-16, Oct 2011. ISSN 0741-5400.
- CHEN, H. et al. Psoriasis patients are enriched for genetic variants that protect against HIV-1 disease. **PLoS genetics**, v. 8, n. 2, p. e1002514-e1002514, 2012.
- CHOO, S. Y. The HLA System: Genetics, Immunology, Clinical Testing, and Clinical Implications. **Yonsei Medical Journal**, v. 48, n. 1, p. 11-23, 2007.
- COHEN, G. B. et al. The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. **Immunity**, v. 10, n. 6, p. 661-71, 1999 Jun 1999. ISSN 1074-7613 (Print)1074-7613.
- CORONA, E.; DUDLEY, J. T.; BUTTE, A. J. Extreme evolutionary disparities seen in positive selection across seven complex diseases. **PLoS One**, v. 5, n. 8, p. e12236, Aug 17 2010. ISSN 1932-6203.
- D'SOUZA, M. P. et al. Casting a wider net: Immunosurveillance by nonclassical MHC molecules. **PLoS Pathog**, v. 15, n. 2, p. e1007567, 02 2019. ISSN 1553-7374.
- DONADI, E. A. et al. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. **Cell Mol Life Sci**, v. 68, n. 3, p. 369-95, Feb 2011. ISSN 1420-9071.
- DOS SANTOS FRANCISCO, R. et al. HLA supertype variation across populations: new insights into the role of natural selection in the evolution of HLA-A and HLA-B polymorphisms. **Immunogenetics**, v. 67, n. 11-12, p. 651-63, Nov 2015. ISSN 0093-7711.
- EGGENSPERGER, S.; TAMPÉ, R. The transporter associated with antigen processing: a key player in adaptive immunity. **Biol Chem**, v. 396, n. 9-10, p. 1059-72, Sep 2015. ISSN 1437-4315.
- EWENS, W. J. The sampling theory of selectively neutral alleles. **Theoretical population biology**, v. 3, n. 1, p. 87-112, 1972.
- EXCOFFIER, L.; LISCHER, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Molecular ecology resources**, v. 10, n. 3, p. 564-567, 2010.

- FELICIO, L. P. et al. Worldwide HLA-E nucleotide and haplotype variability reveals a conserved gene for coding and 3' untranslated regions. **Tissue antigens**, v. 83, n. 2, p. 82-93, 2014.
- FELLAY, J. et al. A whole-genome association study of major determinants for host control of HIV-1. **Science**, v. 317, n. 5840, p. 944-7, Aug 17 2007. ISSN 0036-8075.
- FU, W.; AKEY, J. M. Selection and adaptation in the human genome. **Annu Rev Genomics Hum Genet**, v. 14, p. 467-89, 2013. ISSN 1527-8204.
- GARRIGAN, D.; HEDRICK, P. W. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. **Evolution**, v. 57, n. 8, p. 1707-22, Aug 2003. ISSN 0014-3820 (Print)0014-3820.
- GINEAU, L. et al. Balancing immunity and tolerance: genetic footprint of natural selection in the transcriptional regulatory region of HLA-G. **Genes Immun**, v. 16, n. 1, p. 57-70, 2015 Jan-Feb 2015. ISSN 1476-5470.
- HACKMON, R. et al. Definitive class I human leukocyte antigen expression in gestational placentation: HLA-F, HLA-E, HLA-C, and HLA-G in extravillous trophoblast invasion on placentation, pregnancy, and parturition. **American Journal of Reproductive Immunology**, v. 77, n. 6, p. e12643-n/a, 2017. ISSN 1600-0897.
- HORTON, R. et al. Gene map of the extended human MHC. **Nat Rev Genet**, v. 5, n. 12, p. 889-99, Dec 2004. ISSN 1471-0056 (Print)1471-0056.
- HUGHES, A. L.; YEAGER, M. Natural selection at major histocompatibility complex loci of vertebrates. **Annu Rev Genet**, v. 32, p. 415-35, 1998. ISSN 0066-4197 (Print)0066-4197.
- HULPKE, S.; BALDAUF, C.; TAMPÉ, R. Molecular architecture of the MHC I peptide-loading complex: one tapasin molecule is essential and sufficient for antigen processing. **FASEB J**, v. 26, n. 12, p. 5071-80, Dec 2012. ISSN 1530-6860.
- HUNDHAUSEN, C. et al. Allele-specific cytokine responses at the HLA-C locus: implications for psoriasis. **J Invest Dermatol**, v. 132, n. 3 Pt 1, p. 635-41, Mar 2012. ISSN 0022-202x.
- IBRAHIM, E. C. et al. Tumor-specific up-regulation of the nonclassical class I HLA-G antigen expression in renal carcinoma. **Cancer Res**, v. 61, n. 18, p. 6838-45, 2001 Sep 15 2001. ISSN 0008-5472 (Print)0008-5472.
- ITO, Y. et al. Full-length EBNA1 mRNA-transduced dendritic cells stimulate cytotoxic T lymphocytes recognizing a novel HLA-Cw*0303- and -Cw*0304-restricted epitope on EBNA1-expressing cells. **J Gen Virol**, v. 88, n. Pt 3, p. 770-80, 2007 Mar 2007. ISSN 0022-1317 (Print)0022-1317.
- JONJIC, S. et al. Immune evasion of natural killer cells by viruses. **Curr Opin Immunol**, v. 20, n. 1, p. 30-8, 2008 Feb 2008. ISSN 0952-7915 (Print)1879-0372 (Electronic).
- KAUR, G. et al. Structural and regulatory diversity shape HLA-C protein expression levels. **Nat Commun**, v. 8, p. 15924, Jun 2017. ISSN 2041-1723.
- KEY, F. M. et al. Advantageous diversity maintained by balancing selection in humans. **Curr Opin Genet Dev**, v. 29, p. 45-51, Dec 2014. ISSN 0959-437x.
- KIMURA, M. The neutral theory of molecular evolution: A review of recent evidence. **The Japanese Journal of Genetics**, v. 66, n. 4, p. 367-386, 1991.
- KLEIN, J.; SATO, A. The HLA system. First of two parts. **The New England journal of medicine**, v. 343, n. 10, p. 702-709, 2000.

- KLEIN, L. et al. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). **Nat Rev Immunol**, v. 14, n. 6, p. 377-91, Jun 2014. ISSN 1474-1741.
- KLOETZEL, P. M. Antigen processing by the proteasome. **Nat Rev Mol Cell Biol**, v. 2, n. 3, p. 179-87, Mar 2001. ISSN 1471-0072.
- KONDO, E. et al. Identification of novel CTL epitopes of CMV-pp65 presented by a variety of HLA alleles. **Blood**, v. 103, n. 2, p. 630-8, 2004 Jan 15 2004. ISSN 0006-4971 (Print)0006-4971.
- KULKARNI, S.; MARTIN, M. P.; CARRINGTON, M. The Ying and Yang of HLA and KIR in Human Disease. **Semin Immunol**, v. 20, n. 6, p. 343-52, 2008 Dec 2008. ISSN 1044-5323 (Print)1096-3618 (Electronic).
- KULKARNI, S. et al. Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 51, p. 20705-20710, 2013.
- KULKARNI, S. et al. Differential microRNA regulation of HLA-C expression and its association with HIV control. **Nature**, v. 472, n. 7344, p. 495-498, 2011.
- KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Mol Biol Evol**, v. 33, n. 7, p. 1870-4, 07 2016. ISSN 1537-1719.
- KUŚNIERCZYK, P. Killer Cell Immunoglobulin-Like Receptor Gene Associations with Autoimmune and Allergic Diseases, Recurrent Spontaneous Abortion, and Neoplasms. **Front Immunol**, v. 4, 2013 Jan 29 2013. ISSN 1664-3224 (Electronic).
- LAUER, G. M. et al. High resolution analysis of cellular immune responses in resolved and persistent hepatitis C virus infection. **Gastroenterology**, v. 127, n. 3, p. 924-36, 2004 Sep 2004. ISSN 0016-5085 (Print)0016-5085.
- LE GALL, S. et al. Nef interacts with the mu subunit of clathrin adaptor complexes and reveals a cryptic sorting signal in MHC I molecules. **Immunity**, v. 8, n. 4, p. 483-95, 1998 Apr 1998. ISSN 1074-7613 (Print)1074-7613.
- LEPIN, E. J. et al. Functional characterization of HLA-F and binding of HLA-F tetramers to ILT2 and ILT4 receptors. **Eur J Immunol**, v. 30, n. 12, p. 3552-61, Dec 2000. ISSN 0014-2980.
- LI, L.; DONG, M.; WANG, X. G. The Implication and Significance of Beta 2 Microglobulin: A Conservative Multifunctional Regulator. **Chin Med J (Engl)**, v. 129, n. 4, p. 448-55, Feb 2016. ISSN 0366-6999.
- LIMA, T. H. et al. HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. **Hum Immunol**, v. 77, n. 10, p. 841-53, Oct 2016. ISSN 0198-8859.
- LIMA, T. H. A. et al. HLA-A promoter, coding, and 3'UTR sequences in a Brazilian cohort, and their evolutionary aspects. **HLA**, v. 93, n. 2-3, p. 65-79, Feb 2019. ISSN 2059-2310.
- MAJORCZYK, E. et al. A single nucleotide polymorphism -35 kb T>C (rs9264942) is strongly associated with psoriasis vulgaris depending on HLA-Cw(*)06. **Human immunology**, v. 75, n. 6, p. 504-507, 2014. ISSN 1879-1166 (Electronic)r0198-8859 (Linking).

- MAKADZANGE, A. T. et al. Characterization of an HLA-C-restricted CTL response in chronic HIV infection. **Eur J Immunol**, v. 40, n. 4, p. 1036-41, 2010 Apr 2010. ISSN 0014-2980.
- MATZARAKI, V. et al. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. **Genome Biol**, v. 18, n. 1, p. 76, Apr 2017. ISSN 1474-760X.
- MCCUTCHEON, J. A. et al. Low HLA-C expression at cell surfaces correlates with increased turnover of heavy chain mRNA. **The Journal of experimental medicine**, v. 181, n. 6, p. 2085-95, 1995.
- MENDES-JUNIOR, C. T. et al. Genetic diversity of the HLA-G coding region in Amerindian populations from the Brazilian Amazon: a possible role of natural selection. **Genes Immun**, v. 14, n. 8, p. 518-26, Dec 2013. ISSN 1466-4879.
- MEYER, D. et al. A genomic perspective on HLA evolution. **Immunogenetics**, Jul 2017. ISSN 1432-1211.
- MEYER, D.; THOMSON, G. How selection shapes variation of the human major histocompatibility complex: a review. **Ann Hum Genet**, v. 65, n. Pt 1, p. 1-26, Jan 2001. ISSN 0003-4800 (Print)0003-4800.
- NAIYER, M. M. et al. KIR2DS2 recognizes conserved peptides derived from viral helicases in the context of HLA-C. **Sci Immunol**, v. 2, n. 15, 2017 Sep 15 2017. ISSN 2470-9468.
- NEEFJES, J. J.; PLOEGH, H. L. Allele and locus-specific differences in cell surface expression and the association of HLA class I heavy chain with beta 2-microglobulin: differential effects of inhibition of glycosylation on class I subunit association. **European journal of immunology**, v. 18, n. 5, p. 801-810, 1988.
- NEISIG, A.; MELIEF, C. J.; NEEFJES, J. Reduced cell surface expression of HLA-C molecules correlates with restricted peptide binding and stable TAP interaction. **Journal of immunology (Baltimore, Md. : 1950)**, v. 160, n. 1, p. 171-179, 1998.
- OHASHI, P. S. T-cell signalling and autoimmunity: molecular mechanisms of disease. **Nat Rev Immunol**, v. 2, n. 6, p. 427-38, Jun 2002. ISSN 1474-1733.
- PARCEJ, D.; TAMPÉ, R. ABC proteins in antigen translocation and viral inhibition. **Nat Chem Biol**, v. 6, n. 8, p. 572-80, Aug 2010. ISSN 1552-4469.
- PARHAM, P. Killer cell immunoglobulin-like receptor diversity: balancing signals in the natural killer cell response. **Immunol Lett**, v. 92, n. 1-2, p. 11-3, Mar 29 2004. ISSN 0165-2478 (Print)0165-2478.
- PARHAM, P. MHC class I molecules and kirs in human history, health and survival. **Nat Rev Immunol**, v. 5, n. 3, p. 201-214, 2005.
- PARHAM, P.; MOFFETT, A. Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. **Nat Rev Immunol**, v. 13, n. 2, p. 133-44, Feb 2013. ISSN 1474-1733.
- PAUL, P. et al. HLA-G expression in melanoma: a way for tumor cells to escape from immunosurveillance. **Proc Natl Acad Sci U S A**, v. 95, n. 8, p. 4510-5, Apr 1998. ISSN 0027-8424.
- PENA, S. D. et al. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. **PLoS One**, v. 6, n. 2, p. e17063, 2011. ISSN 1932-6203.

- PENMAN, B. S. et al. Pathogen selection drives nonoverlapping associations between HLA loci. **Proc Natl Acad Sci U S A**, v. 110, n. 48, p. 19645-50, Nov 26 2013. ISSN 0027-8424.
- PIETRA, G. et al. HLA-E and HLA-E-bound peptides: recognition by subsets of NK and T cells. **Curr Pharm Des**, v. 15, n. 28, p. 3336-44, 2009. ISSN 1873-4286.
- RAMSURAN, V. et al. Sequence and Phylogenetic Analysis of the Untranslated Promoter Regions for HLA Class I Genes. **J Immunol**, v. 198, n. 6, p. 2320-2329, Mar 15 2017. ISSN 0022-1767.
- ROBINSON, J. et al. The IPD and IMGT/HLA database: allele variant databases. **Nucleic Acids Res**, v. 43, n. Database issue, p. D423-31, Jan 2015. ISSN 1362-4962.
- ROCK, K. L.; REITS, E.; NEEFJES, J. Present Yourself! By MHC Class I and MHC Class II Molecules. **Trends Immunol**, v. 37, n. 11, p. 724-737, Nov 2016. ISSN 1471-4981.
- SABBAGH, A. et al. Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: balancing selection influencing genetic diversity. **Genes Immun**, v. 15, n. 2, p. 95-106, Mar 2014. ISSN 1476-5470.
- SAMS, A.; HAWKS, J. Celiac disease as a model for the evolution of multifactorial disease in humans. **Hum Biol**, v. 86, n. 1, p. 19-36, Winter 2014. ISSN 0018-7143.
- SANCHEZ-MAZAS, A. An apportionment of human HLA diversity. **Tissue Antigens**, v. 69 Suppl 1, p. 198-202, Apr 2007. ISSN 0001-2815 (Print)0001-2815.
- SATTA, Y.; LI, Y. J.; TAKAHATA, N. The neutral theory and natural selection in the HLA region. **Frontiers in bioscience : a journal and virtual library**, v. 3, p. d459-67, 1998.
- SCHWARTZ, O. et al. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. **Nat Med**, v. 2, n. 3, p. 338-42, 1996 Mar 1996. ISSN 1078-8956 (Print)1078-8956.
- SEBZDA, E. et al. Selection of the T cell repertoire. **Annu Rev Immunol**, v. 17, p. 829-74, 1999. ISSN 0732-0582.
- SELIGER, B.; RITZ, U.; FERRONE, S. Molecular mechanisms of HLA class I antigen abnormalities following viral infection and transformation. **Int J Cancer**, v. 118, n. 1, p. 129-38, 2006 Jan 1 2006. ISSN 0020-7136 (Print)0020-7136.
- SELIGER, B. et al. Association of HLA class I antigen abnormalities with disease progression and early recurrence in prostate cancer. **Cancer Immunol Immunother**, v. 59, n. 4, p. 529-40, 2010 Apr 2010. ISSN 0340-7004 (Print)1432-0851 (Electronic).
- SHIINA, T. et al. The HLA genomic loci map: expression, interaction, diversity and disease. **Journal of Human Genetics**, v. 54, n. 1, p. 15-39, 2009.
- SHIINA, T.; INOKO, H.; KULSKI, J. K. An update of the HLA genomic region, locus information and disease associations: 2004. **Tissue Antigens**, v. 64, n. 6, p. 631-649, 2004.
- SIBILIO, L. et al. A single bottleneck in HLA-C assembly. **J Biol Chem**, v. 283, n. 3, p. 1267-74, Jan 2008. ISSN 0021-9258.
- SONON, P. et al. HLA-G, -E and -F regulatory and coding region variability and haplotypes in the Beninese Toffin population sample. **Mol Immunol**, v. 104, p. 108-127, 12 2018. ISSN 1872-9142.

- SPURGIN, L. G.; RICHARDSON, D. S. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. **Proc Biol Sci**, v. 277, n. 1684, p. 979-88, Apr 07 2010. ISSN 0962-8452.
- STAJICH, J. E.; HAHN, M. W. Disentangling the effects of demography and selection in human history. **Mol Biol Evol**, v. 22, n. 1, p. 63-73, Jan 2005. ISSN 0737-4038.
- STRANGE, A. et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. **Nat Genet**, v. 42, n. 11, p. 985-90, 2010 Nov 2010. ISSN 1061-4036.
- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, v. 123, n. 3, p. 585-595, 1989.
- TAN, Z.; SHON, A. M.; OBER, C. Evidence of balancing selection at the HLA-G promoter region. **Hum Mol Genet**, v. 14, n. 23, p. 3619-28, Dec 2005. ISSN 0964-6906.
- THOMAS, R. et al. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. **Nat Genet**, v. 41, n. 12, p. 1290-1294, 2009.
- TIERCY, J.-M. HLA-C Incompatibilities in Allogeneic Unrelated Hematopoietic Stem Cell Transplantation. **Frontiers in immunology**, v. 5, p. 216-216, 2014.
- TROWSDALE, J.; KNIGHT, J. C. Major histocompatibility complex genomics and human disease. **Annu Rev Genomics Hum Genet**, v. 14, p. 301-23, 2013. ISSN 1527-8204.
- TROWSDALE, J.; MOFFETT, A. NK receptor interactions with MHC class I molecules in pregnancy. **Seminars in Immunology**, v. 20, n. 6, p. 317-320, 2008/12/01/ 2008. ISSN 1044-5323.
- TURNER, S. et al. Sequence-based typing provides a new look at HLA-C diversity. **J Immunol**, v. 161, n. 3, p. 1406-13, 1998 Aug 1 1998. ISSN 0022-1767 (Print)0022-1767.
- VAN DEUTEKOM, H. W. M.; KEŞMİR, C. Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? **Immunogenetics**, v. 67, n. 8, p. 425-436, 2015.
- VAN HALL, T. et al. The Varicellovirus-Encoded TAP Inhibitor UL49.5 Regulates the Presentation of CTL Epitopes by Qa-1b1. 2007-01-15 2007.
- VEIGA-CASTELLI, L. C. et al. Non-classical HLA-E gene variability in Brazilians: a nearly invariable locus surrounded by the most variable genes in the human genome. **Tissue antigens**, v. 79, n. 1, p. 15-24, 2012.
- VINCE, N. et al. HLA-C Level Is Regulated by a Polymorphic Oct1 Binding Site in the HLA-C Promoter Region. **Am J Hum Genet**, v. 99, n. 6, p. 1353-1358, Dec 01 2016. ISSN 0002-9297.
- WATTERSON, G. A. The homozygosity test of neutrality. **Genetics**, v. 88, n. 2, p. 405-417, 1978.
- WIEGERS, G. J. et al. Shaping the T-cell repertoire: a matter of life and death. **Immunol Cell Biol**, v. 89, n. 1, p. 33-9, Jan 2011. ISSN 1440-1711.
- WIENDL, H. et al. A functional role of HLA-G expression in human gliomas: an alternative strategy of immune escape. **J Immunol**, v. 168, n. 9, p. 4772-80, May 2002. ISSN 0022-1767.

- WIERTZ, E. J. et al. The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. **Cell**, v. 84, n. 5, p. 769-79, 1996 Mar 8 1996. ISSN 0092-8674 (Print)0092-8674.
- WIERTZ, E. J. et al. Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. **Nature**, v. 384, n. 6608, p. 432-8, 1996 Dec 5 1996. ISSN 0028-0836 (Print)0028-0836.
- WILCZYŃSKA, K. et al. ERAP, KIR and HLA-C gene interaction in susceptibility to recurrent spontaneous abortion in the Polish population. **Hum Immunol**, v. 80, n. 5, p. 344-348, May 2019. ISSN 1879-1166.
- WILLIAMS, M. et al. Direct Binding of Human Immunodeficiency Virus Type 1 Nef to the Major Histocompatibility Complex Class I (MHC-I) Cytoplasmic Tail Disrupts MHC-I Trafficking. **Journal of Virology**, v. 76, n. 23, p. 12173-12184, 2002.
- YANG, Y.; SEMPÉ, P.; PETERSON, P. A. Molecular mechanisms of class I major histocompatibility complex antigen processing and presentation. **Immunologic research**, v. 15, n. 3, p. 208-33, 1996.
- YEWDELL, J. W.; REITS, E.; NEEFJES, J. Making sense of mass destruction: quantitating MHC class I antigen presentation. **Nat Rev Immunol**, v. 3, n. 12, p. 952-61, Dec 2003. ISSN 1474-1733.
- ZEMMOUR, J.; PARHAM, P. Distinctive polymorphism at the HLA-C locus: implications for the expression of HLA-C. **The Journal of experimental medicine**, v. 176, n. 4, p. 937-50, 1992.
- ZIEGLER, H. et al. The luminal part of the murine cytomegalovirus glycoprotein gp40 catalyzes the retention of MHC class I molecules. In: (Ed.). **EMBO J**, v.19, 2000. p.870-81. ISBN 0261-4189 (Print)1460-2075 (Electronic).