

RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta tese
será disponibilizado somente a partir
de 07/10/2021.

ESTUDO DA VARIABILIDADE DO GENE *HLA-A* E DETECÇÃO
DE ASSINATURAS DE SELEÇÃO NATURAL ATUANDO EM
CADA SEGMENTO DO GENE

THÁLITTA HETAMARO AYALA LIMA

Botucatu, SP

2019

UNIVERSIDADE ESTADUAL PAULISTA

“Júlio de Mesquita Filho”

INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

ESTUDO DA VARIABILIDADE DO GENE *HLA-A* E DETECÇÃO
DE ASSINATURAS DE SELEÇÃO NATURAL ATUANDO EM
CADA SEGMENTO DO GENE

THÁLITTA HETAMARO AYALA LIMA

PROF. DR. ERICK DA CRUZ CASTELLI

Tese apresentada ao Instituto de Biociências, Campus de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Ciências Biológicas (Genética).

Botucatu, SP

2019

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

Lima, Thálitta Hetamaro Ayala.

Estudo da variabilidade do gene *HLA-A* e detecção de assinaturas de seleção natural atuando em cada segmento do gene / Thálitta Hetamaro Ayala Lima. - Botucatu, 2019

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Erick da Cruz Castelli
Capes: 20205007

1. Antígenos *HLA*. 2. Haplotipos. 3. Seleção natural. 4. Polimorfismo (Genética).

Palavras-chave: *HLA-A*; Sequenciamento de Nova Geração; haplótipos; seleção balanceadora; seleção natural.

Dedicatória

Dedico esta tese aos dois maiores incentivadores que tive na vida...

Ao meu querido e grande “Avohai” – avô e pai - Sebastião Romel Ferreira de Alencar, por assegurar ao mundo que eu sempre tive e terei um lar e que aos “trancos e barrancos” estaríamos sempre juntos lutando pelos sonhos da nossa família. Sentados à mesa muitas vezes discutimos a necessidade de exercitar nosso lado racional diante à emoção, ainda quando criança exercitamos muitas vezes as situações hipotéticas onde tinha eu que tomar uma decisão, mal sabia eu que em alguns anos essas situações se tornariam uma realidade e fazer escolhas passariam a fazer parte do dia a dia. Ao seu lado aprendi lições sobre honestidade e retidão, sobre nunca dar prejuízos a ninguém e sobre um amor inabalável à família. Quando partiu em regresso à nossa verdadeira casa, percebi que sua missão havia findado aqui na Terra, e pedi a Deus que ele continuasse a me guiar agora do “outro lado”, não duvidei por nenhum momento que você está por aí sempre a me proteger. Ouço sempre na mente sua voz dizendo: “Isso aí é *conditio sine qua non*”, expressão em latim que significa “condição sem a qual não”, que de forma simplificada se traduz pela necessidade de elementos essenciais para a resolução de questões - problemas. Passo meus dias a indagar diante aos meus, qual é a *conditio sine qua non*? Hoje preste a realizar um sonho “nosso”, te vejo sentado, calmo, feliz e um semblante que expressa: - “É disdaí, minha filha” (É isto aí, minha filha).

À minha querida avó (“Pachuquinha”) Heloisa Helena Melo De Alencar, que sonhou com este momento como ninguém, professora por 26 anos me ensinou a arte de amar ensinar aquilo que aprendemos. Companheira fiel e fraterna, agradeço todos os dias com um saudoso “Bom dia, Sol” a possibilidade de ter compartilhado nossa afinidade intensamente nessa oportunidade que nos foi concedida. Ao seu lado desfrutei meus mais sinceros sorrisos e um sentimento de gratidão que ultrapassa esta vida. Obrigada por me receber em seus braços como uma filha, por não ter medido esforços na realização dos meus sonhos e por me educar na certeza que o investimento na educação e ciência são os recursos que temos para melhorar nosso país. Levo você na memória sempre com aquele jeito fantástico de lidar com os problemas, sempre forte, guerreira, disposta e grata à vida. Que exemplo de mulher, inteligência, mãe, filha, esposa, avó e amiga. Você lutou bravamente contra uma doença cruel, mas nem mesmo o câncer foi motivo para te entristecer. Daqui sinto uma saudade irreparável e ainda dolorida. Peço a Deus que continue a te guiar nos passos da evolução espiritual, daqui sigo certa que nosso amor invadirá milênios e milênios neste espaço físico e sideral.

Agradecimentos especiais

Ao Prof. Dr. Erick da Cruz Castelli, pela oportunidade e credibilidade no desempenho do projeto e por me confiar a responsabilidade deste trabalho, ainda pela amizade e possibilidade de uma formação de qualidade.

À Universidade Estadual Paulista “Júlio Mesquita Filho”, Campus de Botucatu, a qual reconheço como berço e casa da minha formação científica, aqui vivenciei as melhores experiências da trajetória acadêmica.

Ao programa de Pós-Graduação em Ciências Biológicas (Genética), pelo respaldo e infraestrutura para o desenvolvimento da pesquisa a qual me dediquei.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de doutoramento, imprescindível para a realização da pós-graduação.

À Fundação de Amparo e Apoio a Pesquisa (FAPESP) pelo financiamento do projeto de pesquisa realizado pelo grupo, que viabilizou o desenvolvimento em sua inteireza - FAPESP/Brasil (2013/17084-2).

Agradecimentos

Agradeço primeiramente à Deus por me conceder a oportunidade da vida e saúde e sobretudo pelo bom ânimo para desempenhar as tarefas do dia a dia. Sempre acreditei em uma força que nos mantém despertos para vivenciar as oportunidades maravilhosas que a vida nos concede, atribuo essa força ao plano espiritual que me serviu de acalento e paz durante o desenvolvimento da tese.

A minha querida e admirável mãezinha Bianca Ayala Melo Di Alencar, por ser meu porto seguro em qualquer momento, por ser a pessoa a quem confio meus sonhos e expectativas, por ser quem se curva em oração pedindo proteção e me desejando o melhor que este plano possa oferecer. Por me apoiar em qualquer decisão, por sempre terminar o discurso com a frase: “Filha, estamos juntas”. Em um mundo onde cada vez mais as pessoas se distanciam das responsabilidades emocionais, esse suporte faz com que eu me sinta abençoada de maneira especial.

Aos meus queridos tios Thiago Ayala Melo Di Alencar e Diego Ayala Melo Di Alencar por acreditarem nesse sonho e sobretudo por acreditarem que podemos mudar o mundo com a educação, pelo incentivo e amor de irmãos que sempre tiveram comigo. As frases: “Parabéns, Thaty!”, “É isso aí, Nega!” serviram de incentivo a cada conquista compartilhada com vocês.

As queridas tias Karinna Matias, Luciana Teixeira e Renata Teles pelas palavras de incentivo e carinho sempre direcionadas a mim. Sou feliz por ter pessoas como vocês ao meu lado, um bom combate se faz bem amparado por pessoas que acreditam e que torcem por nós.

Aos meus queridos primos Nicholas Ayala, Laura Ayala, Yana Ayala e Nathan Ayala, pelo amor gratuito e olhar fraternal que me fazem querer ser a cada amanhecer um ser humano melhor e servir de inspiração para que sigam sempre na trilha do amor, semeando o bem sempre.

Ao meu querido padrinho Pablo Ayala Melo Di Alencar, mais conhecido por “Blito” por ser minha referência de alma boa e forte, por sempre me ouvir com zelo e atenção nos maiores dilemas que enfrentei nos últimos anos, por ser a pessoa que sempre acreditou que estou no caminho certo e que não devo desistir, por ser a palavra certa que busco nos momentos de incerteza. Blito, você é “fera”.

Ao meu amável namorado Marcelo Ribeiro Begnini, por ser fonte de inspiração e compromisso, por cuidar de mim com tanto zelo, proteção e amor. Por me incentivar sempre

quando o assunto são meus sonhos, por sempre responder com firmeza quando a indagação é: “Você acha que vou conseguir? ”. A resposta é sempre “Claro que vai! ”. As pitadas de bom humor e gargalhadas que compartilhamos deixam os meus dias mais divertidos e alegres. Me sinto privilegiada por ter você ao meu lado, obrigada por toda força e carinho a mim dispensados. Você merece todo o sucesso do mundo!

Ao Laboratório de Genética Molecular e Bioinformática pela estrutura e por servir de terreno fértil às ideias ao longo do desenvolvimento do projeto.

Aos colegas de trabalho Michelle de Almeida Paz, Emiliana Weiss, Heloisa de Souza Andrade, Marília Rodrigues Silva Passos, Nayane dos Santos Brito e Silva e Camila Ferreira Bannwart pelo intercâmbio de conhecimentos científicos, pelo bom humor e companheirismo imprescindíveis para um bom trabalho em equipe. Vocês são demais e continuarão a brilhar sempre.

A minha querida e inesquecível amiga Andréia da Silva Souza, você foi o maior presente que ganhei no curso do meu doutoramento, sem você compartilhando esse caminho lado a lado afirmo com plena certeza que não teria conseguido me manter de cabeça erguida em diversas situações. Você foi pela palavra o que me amparou nos momentos de fraqueza, o empurrão quando foi a atitude que precisei tomar em muitas situações e o abraço silencioso que me fortaleceu quando não eram nem de as palavras nem de atitudes a necessidade. Agradeço a Deus todos os dias pela sua existência, e tenho certeza que o mundo está de braços abertos a te esperar e felizes serão os que estiverem ao seu lado. Mantenha essa sua fé inabalável e inquestionável. No Departamento de Genética ficamos conhecidas como “Dupla Dinâmica” e nos experimentos como “B1 e B2”, esse é o sentimento que sempre terei a seu respeito: alguém que posso contar e que não abandona nunca aqueles que estão ao seu lado.

A todos os amigos que a UNESP e Botucatu me concederam e que participaram do meu crescimento profissional e pessoal diretamente, servindo de ombro amigo, ou compartilhando as melhores risadas do mundo, Nathalia Souza, Luciana Bonome, Raquel Bezerra, Ana Paula Carvalheira, Carol Vaccari, Luana Sacramento e Paula Michaela, com vocês tive a certeza que o mundo pode ser sim o lugar que sonhamos e que nunca devemos desistir do que acreditamos.

Aos queridos amigos Iane de Oliveira Pires Porto, Leandro Prado Felício e Kaison Ernane, que concluíram seus trabalhos anteriormente no grupo e que se tornaram os melhores amigos que alguém poderia ter. Vocês são a prova viva que nada na vida acontece ao acaso,

estava programado para compartilharmos um pedaço da caminhada juntos e sigo firme que continuaremos compartilhando nossos melhores e piores momentos juntos, uma gratidão eterna é o que eu sinto pela amizade de vocês.

A todos mestres, doutores e docentes da pós-graduação, cada qual com seu modo singular de ensinar e compartilhar experiências, meu privilégio e gratidão são imensos e seria incapaz de descrever o quão feliz eu sou por ter sido e continuar a ser aprendiz desse seletor grupo.

A todos os profissionais com os quais tive contato no desenvolvimento desse trabalho, incluindo, técnicos, secretaria da pós-graduação, departamento de limpeza, que contribuíram cada um ao seu modo para que esse sonho se realizasse.

Aos alunos que pude exercitar a prática docente, pela receptividade e atenção ao longo das aulas, com vocês compartilhei o meu melhor como profissional e pessoa. Obrigada por me permitirem aprender e exercitar a arte do ensino ao lado de vocês.

A todos meus familiares, colegas e amigos que fizeram parte desta trajetória, todos possuem participações singulares e essenciais para a realização deste sonho. A todos que participaram direta ou indiretamente desta conquista, meu muito obrigada!

“Bom dia, Sol, é minha *conditio sine qua non*”

Sumário

Capítulo I	14
Introdução	15
Estrutura e função das moléculas HLA	16
O gene <i>HLA-A</i> e suas funções	22
Seleção natural e o complexo HLA	25
Entendendo os testes de assinatura de seleção natural	29
Referências	31
Justificativa	37
Hipótese do estudo	39
Objetivo geral	40
Objetivos específicos	40
Capítulo II	41
Artigo científico	42
Abstract	43
Introduction	44
Methods	46
Population sample	46
<i>HLA-A</i> amplification	46
Library preparation and sequencing	46
Raw data processing (mapping)	47
Genotype calling and processing	47
Diversity indexes, neutrality deviation tests and other analyses	49
Results	50
<i>HLA-A</i> coding region haplotypes	51
<i>HLA-A</i> promoter haplotypes.....	51
<i>HLA-A</i> 3' untranslated region haplotypes.....	52
Nucleotide diversity and natural selection signatures at the <i>HLA-A</i> locus.....	52
Discussion	53
Legends	59
Acknowledgements	60
References	68
Conclusões	72
Material suplementar	73
Figure S1. Network of the extended <i>HLA-A</i> haplotypes (associated promoters, coding and 3'UTR haplotypes).....	74

Figure S2. Linkage Disequilibrium across the <i>HLA-A</i> locus.....	75
Figure S3. <i>HLA-A</i> promoter haplotypes detected in Brazil.....	76
Table S1: List of variable sites at the <i>HLA-A</i> gene detected in Brazil.	77
Table S2: List of all <i>HLA-A</i> protein molecules detected in a Brazilian population sample from Southeast Brazil.	89

Lista de Abreviaturas e Siglas

VCF	Variant Call Format
µL	Microlitro (10 ⁻⁶ litro)
APC	Célula Apresentadora de Antígeno (APC, do inglês, <i>Antigen Presenting Cell</i>)
DAA	Vantagem do Alelo Divergente (DAA, do inglês <i>Divergent Allele Advantage</i>)
DNA	Ácido Desoxirribonucleico (DNA, do inglês, <i>Deoxyribonucleic acid</i>)
EDTA	Ácido Etileno Diamino Tetracético
HLA	Antígeno Leucocitário Humano (do inglês, <i>Human Leukocyte Antigens</i>)
IMGT	International Immunogenetics Database
Kb	Kilobase (10 ³ bases)
KIR	Receptor do tipo imunoglobulina (KIR, do inglês <i>Killer Cell Immunoglobulin Like Receptor</i>)
LD	Desequilíbrio de ligação (LD, do inglês <i>Linkage disequilibrium</i>)
Mb	Megabase (10 ⁶ bases)
MHC	Complexo Principal de Histocompatibilidade (do inglês, <i>Major Histocompatibility Complex</i>)
mL	Mililitro (10 ⁻³ Litro)
ng	Nanograma (10 ⁻⁹ grama)
NGS	Sequenciamento de Nova Geração (do inglês, <i>Next Generation Sequencing</i>)
NK	Célula Natural Killer (NK, do inglês <i>Natural Killer</i>)
UTR	Região não traduzida (do inglês, <i>Untranslated Region</i>)
°C	Graus Celsius
Pb	Pares de bases
PCR	Reação em Cadeia da Polimerase (PCR, do inglês, <i>Polymerase Chain Reaction</i>)
SNP	Polimorfismo de Base Única (SNP, do inglês, <i>Single Nucleotide Polymorphism</i>)
SP	São Paulo
TCR	Receptor de Célula T (TCR, do inglês <i>T Cell Receptor</i>)
TAP	Transportador Associado a Peptídeos (do inglês, <i>Transporter Associated with Antigen Processing</i>)
CTL	Linfócito T Citotóxico (CTL, do inglês <i>Cytotoxic T lymphocyte</i>)
B2-M	Beta-2 microglobulina

Resumo

O Complexo HLA (*Human Leukocyte Antigen Complex*) é a região mais variável do genoma humano. O gene *HLA-A* é o segundo mais polimórfico do grupo e responsável por codificar moléculas envolvidas no processo de apresentação antigênica e modulação das células *NK* (*Natural Killer*). Diversos estudos avaliaram a variabilidade do gene *HLA-A* em populações mundiais. No entanto, estes estudos focaram-se principalmente na variabilidade dos éxons, ignorando a variabilidade de íntrons e regiões regulatórias. A variabilidade do gene *HLA-A* é particularmente elevada no segmento que codifica a fenda de ligação a peptídeos (éxons 2 e 3), relacionado com a função de apresentação antigênica e alvo de seleção balanceadora. No presente estudo avaliamos a diversidade genética do gene *HLA-A* considerando um segmento contínuo que compreende o promotor estendido (1.5 Kb upstream do primeiro ATG traduzido), todos os éxons (incluindo a região 3' não-traduzida) e íntrons por meio de sequenciamento paralelo massivo. As estratégias computacionais empregadas no estudo utilizam-se de programas de livre acesso ou desenvolvidos localmente, viabilizando um mapeamento adequado das sequências produzidas e a genotipagem/haplotipagem para genes HLA. A variabilidade detectada para o gene *HLA-A* em uma amostra brasileira demonstrou que os segmentos regulatórios apresentam poucas sequências diferentes, porém as sequências existentes são bastante divergentes entre si. As variantes regulatórias apresentam alto desequilíbrio de ligação com variantes da região codificadora. Assim como os éxons, os íntrons do gene *HLA-A* são bastante variáveis. Além disso, padrões de diversidade molecular sugerem que a região promotora, juntamente com a região codificadora, estão sob o mesmo tipo de pressão seletiva, mas um cenário diferente é observado para o éxon 4 e para a região 3' não-traduzida.

Palavras-chave: seleção balanceadora, haplótipos, alelos, *HLA-A*, seleção natural, Sequenciamento de Nova Geração, polimorfismos, evolução.

Capítulo I

Introdução

O Complexo Principal de Histocompatibilidade (MHC, do inglês, *Major Histocompatibility Complex*) consiste de um grupo de genes localizados no braço curto do cromossomo seis (6p21.3), em um segmento de aproximadamente 4Mb (0.13% do genoma humano). Este complexo apresenta cerca de 200 genes (Klein e Sato, 2000a; b), representando ao todo cerca de 0.5% dos genes codificadores de proteínas descritos no genoma humano (Shiina *et al.*, 2009). Em sua maioria os produtos gênicos codificados pelo MHC estão de alguma forma envolvidos com a resposta imune, sendo eles: ligantes, receptores, fatores de sinalização celular, moléculas reguladoras do processo transcricional. Os produtos gênicos codificados pelo MHC estão relacionados a contextos imunes diversos, tais como: inflamação, processamento e apresentação antigênica (resposta imune adaptativa) e interações com receptores nas células *NK* (do inglês, *Natural Killer*), importantes na resposta imune inata (Shiina *et al.*, 2009). A região do MHC é considerada a região de maior densidade gênica do genoma humano, com uma média de 1 gene a cada 16Kb (Meyer *et al.*, 2018). Além disso, o MHC é a região mais variável do genoma humano e o elevado polimorfismo deve-se principalmente aos genes HLA (*HLA*, do inglês *Human Leucocyte Antigens*). O complexo HLA codifica moléculas responsáveis pelo processamento e apresentação de antígenos. Os genes do complexo HLA estão envolvidos nos processos de combate a patógenos, pois as moléculas que codificam apresentam antígenos aos linfócitos T (Klein e Sato, 2000a).

Didaticamente, os genes do complexo HLA estão divididos em dois grupos principais, denominados genes de classe I e II. Os genes de classe I, *HLA-A*, *HLA-B* e *HLA-C*, são também conhecidos como genes HLA de **classe I clássicos** e codificam glicoproteínas transmembrana expressas na maioria das células nucleadas. Essas moléculas atuam na apresentação de antígenos extracelulares aos linfócitos TCD8 durante a resposta imunitária adaptativa (Klein e Sato, 2000a; b; Goldberg e Rizzo, 2015). Os genes *HLA-G*, *HLA-E* e *HLA-F* são considerados genes **de classe I não clássicos** (Geraghty *et al.*, 1992; Klein e Sato, 2000a; b). Embora apresentem semelhança estrutural e molecular aos genes clássicos, possuem padrões de expressão distinto e estão relacionados principalmente com o processo de tolerância imunológica (Barrow e Trowsdale, 2008; Carosella *et al.*, 2008; Crispim, Duarte, *et al.*, 2008; Harada *et al.*, 2015; Martinez *et al.*, 2017). Em geral, as moléculas não clássicas ligam-se a receptores presentes em linfócitos T e células *NK*, modulando a ação destas células (Donadi *et al.*, 2011). Os genes de classe II, *HLA-DPA1*, *HLA-DPBI*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1* e *HLA-DRA1* também estão envolvidos com o processo de apresentação antigênica,

porém suas moléculas apresentam antígenos aos linfócitos TCD4 na superfície celular (Touraine *et al.*, 1989).

Os genes HLA são considerados os genes mais polimórficos do genoma humano, possuindo uma ampla diversidade alélica, ou seja, diversas sequências de DNA específicas para um determinado gene HLA. O banco de dados IPD-IMGT/HLA constitui um banco especializado em catalogar sequências HLA, contendo informações detalhadas sobre a origem e validação dessas sequências (Robinson *et al.*, 2016). Este banco de dados conta atualmente com um total de 17.344 sequências de alelos descritos. A gama de sequências HLA descritas extrapola mais de 16.000 sequências para genes HLA de classe I (Tabela 1) e por volta de 4.000 sequências para os genes HLA de classe II (Robinson *et al.*, 2016).

Tabela 1. Número de alelos identificados para os genes HLA de classe I (banco de dados IPD-IMGT/HLA, versão 3.36.0.1/20-02-2019).

HLA DE CLASSE I CLÁSSICOS		
Gene	Alelos	Proteínas
<i>HLA-A</i>	5018	3398
<i>HLA-B</i>	6096	4248
<i>HLA-C</i>	4852	3193
HLA DE CLASSE I NÃO CLÁSSICOS		
Gene	Alelos	Proteínas
<i>HLA-E</i>	30	9
<i>HLA-F</i>	44	8
<i>HLA-G</i>	68	19

Estrutura e função das moléculas HLA

As moléculas HLA de classe I clássicas são expressas na maioria das células nucleadas, ligam peptídeos de origem endógena e os apresentam na superfície celular aos receptores de linfócitos TCD8. Caso sejam reconhecidos como não-próprios, uma resposta citotóxica poderá ocorrer. No entanto, caso peptídeos próprios sejam reconhecidos como estranhos ao organismo, há possibilidade de desenvolvimento de doenças autoimunes (Matzaraki *et al.*, 2017). As moléculas de classe II são expressas em células apresentadoras de antígenos e em geral apresentam peptídeos de origem exógena às células TCD4 na superfície celular, levando a um processo de sinalização celular que estimula a multiplicação das células *T-helper*, levando, por

consequência, a estimulação das células B, que são responsáveis pela produção de anticorpos contra o antígeno que desencadeou a resposta (Rock *et al.*, 2016).

As moléculas HLA de classe I e II apresentam fragmentos protéicos às células TCD8 e TCD4, respectivamente. Essas moléculas são essenciais para a imunidade mediada por células e surgiram no início do sistema imune adaptativo cerca de 500 milhões de anos atrás (Flajnik e Kasahara, 2010). A estrutura de uma molécula de classe I, independentemente do gene que a codifica, compreende uma cadeia α pesada codificada no cromossomo 6, associada com uma cadeia β_2 -microglobulina codificada no cromossomo 15 (Figura 1) (Klein e Sato, 2000a). A cadeia α possui cinco domínios principais. Os domínios $\alpha 1$ e $\alpha 2$ dão origem à ‘fenda de ligação a peptídeos’ e estão envolvidos com o acoplamento de peptídeos. Esta fenda provavelmente evoluiu a partir da estrutura de chaperonas, uma vez que a) chaperonas ligam-se a fragmentos proteicos, b) diversas chaperonas possuem estrutura similar as moléculas MHC; c) existem diversas moléculas que se assemelham estruturalmente ao MHC e que atuam como chaperonas específicas no processo de apresentação antigênica restrita ao MHC (*i.e* tapasina) (Dong *et al.*, 2009; Rock *et al.*, 2016). As cadeias α possuem também um domínio semelhante à imunoglobulina ($\alpha 3$), que interage com co-receptores presentes nas células T, um domínio transmembrana e um domínio citoplasmático (Figura 1) (Klein e Sato, 2000a).

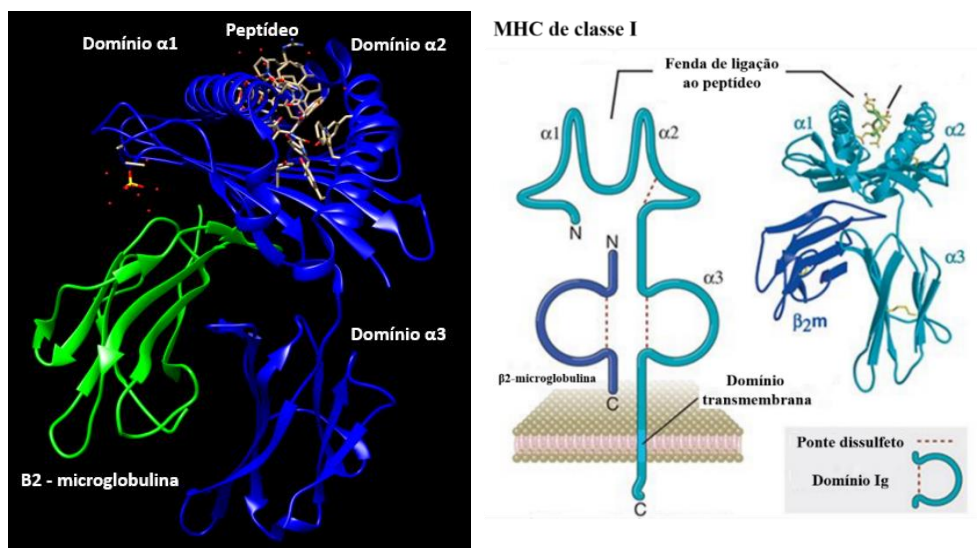


Figura 1. À esquerda, representação gráfica computadorizada da molécula HLA-A2, contendo todos os domínios estruturais e um peptídeo acoplado na fenda ligadora de peptídeos. **Fonte:** Protein Data Bank (<http://www.rcsb.org/structure/5E00>). À direita, ilustração da estrutura da molécula HLA de classe I, composta por três domínios alpha, região transmembrana acoplada a membrana plasmática, cauda citoplasmática e o domínio de β_2 -microglobulina, ligado por meio de ligações não-covalentes à cadeia alpha. **Fonte:** Adaptado de Abbas, *et. al.* Cellular and Molecular Immunology, 2008, p. 193.

Esta semelhança estrutural entre as diferentes moléculas de classe I se deve ao fato de que os genes desse complexo originaram-se por duplicações imperfeitas de blocos gênicos (Kulski *et al.*, 2000). O MHC tem sido descrito como uma região polimórfica composta por blocos conservados. Na região de classe I temos os blocos gênicos alfa e beta. O bloco alfa contém os genes *HLA-F*, *HLA-G*, *HLA-A* e pseudogenes como *HLA-H* e *HLA-J*. Dentro do bloco beta existem duas grandes duplicações de segmentos contendo os genes *HLA-B*, *HLA-C*, *MICA* e *MICB*. Essas duplicações contêm segmentos que compartilham aproximadamente 30 kb de sequências, incluindo sequências codificadoras e intergênicas, sendo que a maior parte das diferenças entre as duplicações estão relacionadas com a alta quantidade de inserções e deleções, contendo retroelementos, pseudogenes e outras sequências repetitivas inseridas nesses segmentos (Dawkins *et al.*, 1999; Gaudieri *et al.*, 1999; Kulski *et al.*, 2000). A origem comum evidenciada para os genes do MHC acarreta uma grande similaridade de sequência entre os vários genes HLA, dificultando abordagens para análise individual da estrutura e variabilidade de cada um desses genes.

No presente trabalho, avaliamos a variabilidade completa do gene *HLA-A*, incluindo suas regiões regulatórias e íntrons, por meio de sequenciamento paralelo massivo. Por conta da alta similaridade entre os genes do complexo HLA e do elevado polimorfismo, desenvolvemos um workflow para análise de dados utilizando ferramentas de bioinformática. Os métodos aplicados para a análise de dados minimizam os custos de genotipagem e definição dos haplótipos, uma vez que (a) utiliza bancos de dados e softwares gratuitos, (b) utiliza primers desenvolvidos localmente e kits de preparação de biblioteca de DNA comuns, sem custos adicionais além dos reagentes de amplificação e sequenciamento, e (c) permite a customização dos experimentos, uma vez que outros alvos (genes) podem ser avaliados em conjunto com o loci HLA. Além disso, a metodologia aqui desenvolvida avalia todo o gene, incluindo segmentos regulatórios, ao invés de focar apenas nas sequências codificadoras como comumente realizado na prática clínica (Sheldon e Poulton, 2006). Para minimizar os problemas de mapeamento errôneo de sequências de genes HLA, utilizamos o software *hla-mapper*, que foi desenvolvido localmente por ocasião da realização deste estudo e de estudos envolvendo outros genes HLA (Castelli *et al.*, 2018).

As moléculas HLA de classe I e classe II são consideradas únicas no proteoma, uma vez que apresentam elevado polimorfismo (>16.000 diferentes alelos já catalogados) (Tabela 1). Essa elevada variabilidade possui consequências interessantes, pois a presença de polimorfismo no topo das *alpha* hélices está relacionada com modificação no padrão de interação entre a molécula

HLA e o TCR (do inglês, *T Cell Receptor*). Além disso, variações no sítio de ligação ao peptídeo estão relacionadas com o perfil de peptídeos apresentados (Rock *et al.*, 2016).

Os resíduos polimórficos presentes na fenda de ligação ao peptídeo são capazes de gerar mudanças na localização dos ‘bolsões’ (*pockets*). Esses *pockets* são preenchidos por cadeias laterais de aminoácidos variáveis complementares, também conhecidos como ‘resíduos âncoras’, com a consequência de que diferentes fragmentos de um determinado antígeno podem ser apresentados por diferentes moléculas HLA (Falk *et al.*, 1994; 2006). Próximo aos resíduos âncoras, outros aminoácidos preenchem o espaço livre, podendo ser qualquer um dos 20 aminoácidos existentes. Uma vez que essas moléculas possuem *pockets* com especificidade para somente algumas poucas cadeias laterais, permitindo que haja variação de 6 a 10 aminoácidos dentre todas as possibilidades, cada molécula HLA pode apresentar um alto repertório de peptídeos. Em teoria as moléculas HLA de classe I podem apresentar cerca de 6×20^6 peptídeos diferentes e HLA de classe II cerca de 12×20^{10} . Contudo, todo esse repertório de peptídeos não pode ser apresentado na superfície celular, uma vez que há uma estimativa de que cerca de 200.000 moléculas HLA de classe I e 20.000 de classe II seriam encontradas na superfície de uma célula T ou B (Eisen *et al.*, 2012).

Importante ressaltar que diferentes moléculas HLA apresentam um conjunto diferente de antígenos. Portanto, quando há mutações em um determinado patógeno que altera um de seus resíduos âncoras críticos para interação adequada com as moléculas HLA, isso prejudica a apresentação desse antígeno por um indivíduo portador das moléculas HLA que originalmente apresentariam este antígeno. Porém, essa apresentação não é prejudicada em indivíduos portadores de outras moléculas HLA (Schmid *et al.*, 2010). Desta forma, o polimorfismo das moléculas HLA é de extrema importância para a adaptabilidade de uma população.

As moléculas clássicas de classe I (HLA-A, HLA-B e HLA-C) associam-se a peptídeos antigênicos oriundos principalmente do meio intracelular. No entanto, alguns microrganismos são capazes de adentrar o meio intracelular e suas proteínas são processadas pelos mesmos mecanismos que processam as proteínas próprias do organismo. Portanto, nessas situações, as moléculas de classe I clássicas também podem se associar a antígenos virais, de microrganismos intracelulares ou fagocitados (Murphy *et al.*, 2012).

As moléculas MHC de classe I são montadas no retículo endoplasmático, onde ocorre a união das cadeias α e β em um processo que requer a ação de chaperonas, calreticulina, p57, calnexina e de uma proteína ligante muito importante na estabilização das moléculas HLA de

classe I recém-sintetizadas, denominada β 2-M (Beta-2 microglobulina). Após a montagem, as moléculas HLA associam-se a um peptídeo e sua versão completa (complexo HLA + peptídeo) é exportada à superfície celular, podendo então interagir com os receptores de células T (Yang *et al.*, 1996; Klein e Sato, 2000a; Abbas *et al.*, 2007). A geração dos peptídeos que serão apresentados pelas moléculas de classe I envolve a participação de um complexo multiproteico conhecido como proteassomo. As proteínas destinadas a este complexo, de alta capacidade proteolítica, são marcadas com ubiquitinas e os peptídeos gerados no citosol são transportados ao retículo endoplasmático por meio do TAP. No retículo, esses peptídeos podem associar-se com moléculas HLA de classe I (Figura 2) (Yang *et al.*, 1996; York e Rock, 1996; Abbas *et al.*, 2007).

Na superfície celular, o complexo HLA + peptídeo interage com receptores TCR e CD8 dos linfócitos T citotóxicos (CTLs), ou ainda, com diversos outros receptores de células NK, que podem ser ativadores ou inibitórios. Na ocasião da interação entre MHC e TCR, uma série de proteínas estabilizam essa interação, permitindo com que ligante e receptor fiquem unidos tempo o suficiente para uma eventual ativação de um linfócito antígeno-específico (Gao *et al.*, 2000; Klein e Sato, 2000b) quando do encontro deste antígeno.

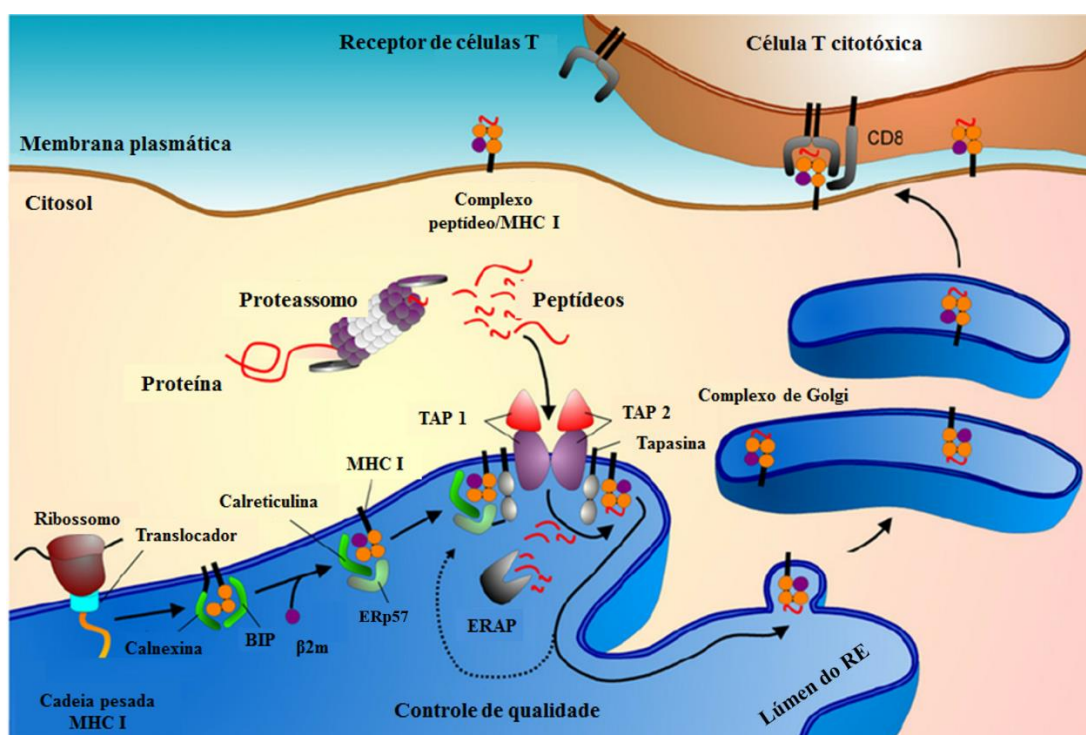


Figura 1. Via de processamento de antígenos das moléculas MHC de classe I. **Fonte:** Adaptado de SCHOLZ, C.; TAMPE, R. The intracellular antigen transport machinery TAP in adaptive immunity and virus escape mechanisms. *J Bioenerg Biomembr*, v. 37, n. 6, p. 509-15, Dec 2005.

Deve-se ressaltar que cada linfócito inativo presente na circulação possui um TCR com uma única especificidade, que foi determinada por um mecanismo de rearranjo gênico e que foi selecionado durante as fases de maturação dos linfócitos e seleção tímica. Em situações normais, esta seleção permite que apenas linfócitos que apresentam uma afinidade (porém baixa) pelas moléculas HLA do indivíduo e que não se associam fortemente com moléculas HLA do indivíduo carregando antígenos próprios são liberados na circulação. Desta forma, cada linfócito possui uma especificidade distinta, porém sua diversidade permite com que a gama de antígenos estranhos reconhecido por essas células do sistema imunitário seja muito grande (Lederberg, 1988; Abbas *et al.*, 2007; Lima *et al.*, 2016). Somente os linfócitos que encontraram um antígeno com o qual seu receptor é capaz de interagir poderão ser ativados a fim de proliferarem e se diferenciarem em células efetoras (Pardigon *et al.*, 1998). Além disso, a ativação das células T necessita também de sinais co-estimulatórios (Pardigon *et al.*, 1998). A presença ou ausência desses sinais pode acarretar na ativação do linfócito T ou a indução de tolerância (Sakaguchi *et al.*, 2008).

Embora a seleção negativa atue eliminando linfócitos auto-reativos durante seu desenvolvimento no timo (Kazansky, 2008), em algumas situações, as moléculas HLA apresentam antígenos próprios aos linfócitos, podendo levar ao desenvolvimento de doenças autoimunes (Caillat-Zucman, 2009; Lima *et al.*, 2016). O processo de seleção negativa consiste em um dos mecanismos mais importantes no processo de tolerância a auto-antígenos (Janeway Ca Jr, 2001; Abbas *et al.*, 2007). Por outro lado, a seleção positiva garante o desenvolvimento das células T cujos receptores interagem fracamente com moléculas MHC do hospedeiro. As células T maduras, cujos precursores sofreram seleção positiva no timo, agora são capazes de reconhecer antígenos não-próprios, apresentados pelas mesmas moléculas HLA do hospedeiro nas APCs nos tecidos periféricos (Janeway Ca Jr, 2001; Lima *et al.*, 2016).

As células *NK* pertencem ao grupo de células que compõe a imunidade inata, eliminando células tumorais ou infectadas por vírus sem a necessidade de sensibilização prévia. A redução ou ausência de moléculas HLA de classe I na superfície da célula alvo é um dos mecanismos que ativam a lise celular mediada pelas células *NK*. Contudo, as células *NK* apresentam uma variedade de receptores em sua superfície celular com capacidade de ligação com moléculas HLA de classe I, inibindo ou aumentando a citotoxicidade mediada por essas células (Davis *et al.*, 1999). O reconhecimento das moléculas HLA de classe I pelos receptores KIR (do inglês, *Killer-cell Immunoglobulin-like Receptors*) presente nas células *NK* é central no controle da função das células *NK*, bem como na modulação da ativação dos linfócitos T citotóxicos. Dentre

os receptores KIR, o *KIR3DL2* parece interagir com variantes do gene *HLA-A* conhecidas como *HLA-A3* e *HLA-A11*, contudo mais estudos ainda precisam ser realizados para a confirmação desses achados (Hansasuta *et al.*, 2004).

O gene *HLA-A* e suas funções

Considerando a alta variabilidade encontrada nos genes clássicos, é improvável que dois indivíduos sejam idênticos quanto a estes polimorfismos considerando-se todos os genes de classe I. Se por um lado esta variabilidade é interessante do ponto de vista populacional, pois aumenta a capacidade de apresentação antigênica, por outro lado, a eventual necessidade de realização de transplantes em humanos torna essa variabilidade problemática, uma vez que a compatibilização desses polimorfismos é crucial para uma boa aceitação do enxerto. A rejeição ao enxerto ocorre quando há um transplante de tecidos de um indivíduo para um receptor geneticamente diferente, levando a uma resposta imunológica direcionada ao órgão transplantado.

A montagem de respostas imunes adaptativas aos tecidos enxertados são o principal impedimento contra o sucesso de um transplante. A rejeição é causada por respostas imunes contra aloantígenos do órgão enxertado, que consistem em proteínas que variam entre indivíduos da mesma espécie e que passam a ser reconhecidas pelo sistema imune do hospedeiro como estranhas (resposta imune alorreativa). Os principais alvos no processo de compatibilização de órgãos para transplantes são as moléculas HLA de classe I clássicas e de classe II. Quando o doador e o receptor diferem quanto ao MHC, a resposta imune alorreativa é dirigida contra moléculas MHC não-próprias alogênicas presentes no órgão enxertado. Existem dois modos possíveis de apresentação dos aloantígenos para as células T do receptor. Na via conhecida como alorreconhecimento direto, as células apresentadoras de antígenos derivadas do doador podem migrar pelos linfonodos localizados próximos a região do enxerto e ativar as células T do hospedeiro. Desta forma, as células T alorreativas efetoras ativadas são direcionadas ao enxerto. O segundo mecanismo que pode desencadear o processo de alorreconhecimento envolve a captação de proteínas alogênicas por APCs do próprio receptor e sua posterior apresentação às células T, incluindo Treg(s), por moléculas MHC próprias, processo conhecido como alorreconhecimento indireto (Bradley, 1991; Sheldon e Poulton, 2006; Abbas *et al.*, 2007; Hansen *et al.*, 2008; Chinen e Buckley, 2010; Ayala Garcia *et al.*, 2012; Murphy *et al.*, 2012)

Alguns estudos já demonstraram que variações em um único aminoácido em uma dessas moléculas HLA clássicas pode levar a rejeição aguda em situações de transplante de medula óssea, pelos mecanismos expostos anteriormente. A análise dos segmentos das porções que acoplam peptídeos (domínios $\alpha 1$ e $\alpha 2$) justifica-se pela importância dessa região na compatibilização com receptores TCR que passaram pelo processo de seleção tímica (reconhecimento próprio e não-próprio) (Choo, 2007; Ayala Garcia *et al.*, 2012).

O gene *HLA-A*, alvo do presente estudo, apresenta atualmente um total de 5.018 alelos (Tabela 1), que codificam 3.398 proteínas distintas. Desta forma, o gene *HLA-A* é o segundo mais variável do genoma humano em termos do número de alelos diferentes, superado apenas pelo gene *HLA-B* (Tabela 1). Grande parte da variabilidade encontrada atualmente para esse gene concentra-se nos éxons 2 e 3, responsáveis pela codificação da fenda de ligação de peptídeos (domínios $\alpha 1$ e $\alpha 2$). Contudo, a variabilidade nesses segmentos justifica-se pela grande importância dada a esses domínios nos diversos contextos patológicos e fisiológicos. Em geral, estes são os segmentos avaliados pelos diversos laboratórios de histocompatibilidade que avaliam a compatibilidade entre doadores e receptores.

A estrutura da molécula HLA-A é de fundamental importância para o desempenho da sua função. Neste contexto o domínio $\alpha 3$, codificado pelo éxon 4, mostra-se importante para a manutenção da resposta imune adaptativa. A interação entre linfócitos TCD8 e seu alvo (moléculas HLA) requerem a interação física entre CD8 e porções conservadas no domínio $\alpha 3$ das moléculas HLA (Fayen *et al.*, 1995). A perda ou dificuldade de interação entre o domínio $\alpha 3$ e CD8 de linfócitos T reduz a sinalização e ativação de CTL (Wesley *et al.*, 1993). Contudo, polimorfismos no segmento $\alpha 3$ parecem influenciar não só a interação com o CD8, mas também com outros receptores. Um estudo de associação de variantes no domínio $\alpha 3$ do gene *HLA-B* encontrou uma associação entre um polimorfismo no códon 194 (acarreta troca de Isoleucina por Valina) e progressão lenta do vírus HIV, por alterar de forma significativa a interação da HLA-B com receptores LIR1 (ILT2/LILRB1/CD85j) (Grifoni *et al.*, 2013). Alterações nas sequências do éxon 5 (domínio transmembrana) do gene *HLA-A* podem modificar a conformação da molécula, uma vez que mutações nesse segmento podem originar moléculas solúveis (não aderidas a membrana plasmática) ou prejudicar sua ancoragem, com consequências na estabilidade dessas moléculas na superfície celular e no desempenho de uma resposta imunológica adequada. Alterações nas sequências dos éxons 6 e 7 podem levar a alterações na regulação do padrão de reciclagem e degradação de moléculas HLA-A. A variabilidade desses éxons pode influenciar no nível de sinais de regulação positiva ou negativa

destas moléculas na superfície celular e, dependendo da situação patológica, pode prejudicar a apresentação de antígenos aos linfócitos, bem como inibir a ação das células NK, importantes para a resolução de determinados quadros patológicos (Schaefer *et al.*, 2008).

Embora a variabilidade nos éxons 2 e 3 seja bastante explorada, os demais segmentos do gene *HLA-A* (incluindo suas regiões regulatórias) são comumente negligenciados. Não está claro se a variabilidade nestes segmentos acompanha a variabilidade observada para os éxons 2 e 3, acarretando modificações funcionais como as descritas acima. Desta forma, a variabilidade genética do gene *HLA-A* (fora dos éxons 2 e 3) pode estar subestimada, apesar do elevado número de alelos registrado para o gene (tabela 1). Ainda, caso a variabilidade das regiões regulatórias do gene *HLA-A* seja tão elevada quanto a observada para os éxons 2 e 3, é provável que diferentes haplótipos encontrados nestes segmentos estejam relacionados com perfis transcricionais distintos. Polimorfismos presentes na região promotora poderiam influenciar a ligação de fatores de transcrição ou influenciar o perfil de metilação do gene. Polimorfismos presentes na região 3' não traduzida poderiam influenciar a ligação de microRNAs, modificando o perfil de transcrição de determinados haplótipos. Por fim, considerando a importância do gene *HLA-A* para a manutenção de uma reposta imunitária adaptativa, não está claro se o perfil de seleção balanceadora, que atua no segmento que codifica a fenda de ligação de peptídeo, também atua nos demais segmentos do gene.

Muitos polimorfismos dos genes clássicos foram associados com doenças autoimunes, degenerativas, infecciosas e susceptibilidade a determinadas neoplasias (Hughes e Yeager, 1998; Meyer e Thomson, 2001; Penn *et al.*, 2002; Bernatchez e Landry, 2003; Crispim, Mendes-Junior, *et al.*, 2008). Por exemplo, determinados alelos *HLA-B* relacionam-se com a progressão da infecção relacionada ao HIV, na qual se encontrou uma forte associação entre o alelo *HLA-B*58:01* e baixa carga viral e o alelo *HLA-B*58:02* e alta carga viral em uma população do Sul da África. Essas variantes *HLA-B* diferem-se por apenas três aminoácidos, destacando que pequenas diferenças entre moléculas HLA de classe I podem refletir em diferenças na especificidade de ligação de peptídeos (Kiepiela *et al.*, 2004). Um trabalho realizado com uma população de Mbeya, Tanzânia, demonstrou que o alelo *HLA-A*74:01* está associado com a proteção contra infecção pelo vírus HIV-1, porém progressão acelerada da doença (Koehler *et al.*, 2010). Os mecanismos de susceptibilidade a infecção e progressão não estão bem caracterizados e precisam ser melhor estudados, contudo a ausência ou presença de determinadas moléculas HLA podem estar relacionadas com diferentes capacidades de apresentação antigênica, influenciando assim os desfechos nos diversos quadros patológicos, a

exemplo das doenças infecciosas. O grupo de alelo *HLA-A*24* parece relacionar-se com a predisposição ao Diabetes mellitus do tipo 1 e os grupos *HLA-A*01*, *HLA-A*11* e *HLA-A*31* parecem desempenhar um papel protetor (Nejentsev *et al.*, 2007). Ainda, o alelo *HLA-A*03:01* parece estar relacionado com a predisposição (Fogdell-Hahn *et al.*, 2000) e *HLA-A*02:01* com a proteção à esclerose múltipla (Harbo *et al.*, 2004).

Em situações de câncer, onde há um descontrole nas vias que regulam a proliferação e morte celular, o desenvolvimento de uma célula tumoral acontece em função de uma sucessão de alterações de origem genética (Virani *et al.*, 2012; Ramsuran *et al.*, 2015) em genes críticos, entre eles aqueles responsáveis pelo controle do ciclo celular (Smyth *et al.*, 2001; Abbas *et al.*, 2007). As respostas imunológicas capazes de destruir células tumorais consistem em linfócitos T CD8 específicos a antígenos tumorais, células *NK* e macrófagos ativados. A ativação dos linfócitos T possui um papel essencial no desencadeamento da resposta imune, pois estas células podem desempenhar uma função de vigilância ao reconhecer e destruir células potencialmente malignas que expressam peptídeos derivados de proteínas celulares mutantes ou proteínas virais oncogênicas. Estes peptídeos são apresentados aos linfócitos T CD8+ associados a moléculas HLA de classe I, ou seja, peptídeos oriundos da célula tumoral são levados à superfície celular pelas moléculas HLA e o complexo HLA-peptídeo é reconhecido pelo receptor de células T. Neste contexto as moléculas HLA clássicas tem papel crucial (Cordon-Cardo *et al.*, 1991; Da Silva *et al.*, 2013). A ausência das moléculas HLA é ocorrência comum em transformações malignas (Abbas *et al.*, 2007).

Seleção natural e o complexo HLA

Darwin propôs em seus estudos que a seleção natural possui como princípio que características hereditárias que aumentam/promovem a aptidão do indivíduo tendem a aumentar em frequência nas populações com o passar do tempo e características que diminuem a aptidão tendem a ser eliminadas (Darwin, 1859). A manutenção ou eliminação dessas características operam por mecanismos distintos, podendo ser por meio da ação de seleção direcional (positiva ou negativa) ou de seleção balanceadora.

A seleção direcional positiva atua sobre uma mutação adaptativa aumentando sua frequência na população. A do tipo direcional negativa, também conhecida como purificadora, reduz a frequência da mesma ou elimina mutações deletérias na população. A seleção balanceadora é responsável pela manutenção dos polimorfismos na população e pode operar

por mecanismos diferentes (vantagem do heterozigoto, flutuação e seleção dependente de frequência). A vantagem do heterozigoto considera que alelos específicos são favorecidos em função da sua habilidade de promover proteção contra determinado patógeno, ou contrariamente quando resultam na susceptibilidade a um determinado patógeno (Penn *et al.*, 2002). A seleção dependente de frequência prediz que os valores de *fitness* não são fixados, mas são capazes de modificarem-se em proporção a frequência de alelos ou genótipos na população, *i.e* quando patógenos mutam determinada sequência em seu material genético, reduzem as chances de seus peptídeos serem apresentados pelas moléculas HLA que possuem elevada frequência, resultando em uma diminuição do *fitness* adaptativo, uma vez que a molécula HLA frequente não será capaz de reconhecer e apresentar esse antígeno. Na sequência, moléculas HLA de menor frequência capazes de reconhecer esse peptídeo aumentarão de frequência, até que uma nova onda de adaptação do patógeno aconteça e a frequência desse alelo diminua novamente, permitindo o aumento de outras moléculas menos frequentes (Meyer e Thomson, 2001). O mecanismo de flutuação considera que a intensidade de seleção natural varia conforme espaço e tempo. Se a frequência de patógenos muda constantemente, diferentes moléculas HLA serão selecionadas em tempos distintos, diferente do modelo dependente de frequência onde o valor do *fitness* muda conforme a frequência alélica, no modelo de flutuação o *fitness* varia em função da frequência de patógenos (Meyer e Thomson, 2001; Meyer *et al.*, 2018; Bitarello, 2011).

A expansão geográfica recente dos humanos, incluindo processos de colonização, expõe nossa espécie constantemente ao contato com diferentes agentes patogênicos. Patógenos potencialmente letais influenciaram sobremaneira as frequências alélicas em nosso genoma. Dessa forma, genes envolvidos com a imunidade apresentam assinaturas moleculares desse processo de adaptação da população humana a essas condições recentes (Ferrer-Admetlla *et al.*, 2008). Atualmente, destaca-se na literatura cerca de 91 genes que apresentam assinaturas de seleção positiva ou balanceadora, ou as duas formas, sendo que diversos desses estão envolvidos com a função imunológica e susceptibilidade a patógenos (Sabeti *et al.*, 2006). Por isso, estudos que esclareçam as assinaturas de seleção natural atuando nos genes HLA, podem trazer informações importantes acerca do passado evolutivo dessas moléculas, acrescentando informações importantes quanto a função desses genes.

A alta variabilidade associada aos genes HLA, principalmente nos éxons 2 e 3, que codificam a fenda de ligação ao peptídeo, levaram alguns autores a estudarem o perfil de seleção nesses genes (Garrigan e Hedrick, 2003; Meyer *et al.*, 2006; Buhler *et al.*, 2016). Uma das

hipóteses para a manutenção da alta diversidade alélica observada nesses segmentos parece ser a seleção balanceadora mediada por microorganismos, que acarreta em um aumento da heteroziguidade nesses segmentos. Diversos estudos têm buscado genes sob ação de seleção balanceadora (capaz de gerar um *fitness* adaptativo), regime que mantém diversas variantes na população com frequências intermediárias, fazendo com que cada alelo persista na população por tempo maior do que aqueles sob ação de seleção neutra (Andres *et al.*, 2009; Bitarello, 2011).

Neste caso, em uma população exposta a diversos patógenos, indivíduos heterozigotos são favorecidos quando comparados aos homozigotos, uma vez que o repertório de peptídeos apresentados pelos heterozigotos é maior (Doherty e Zinkernagel, 1975; Hughes e Yeager, 1998; Penn *et al.*, 2002; Bernatchez e Landry, 2003). Um estudo realizado por Solberg e colaboradores em 2008 considerando 497 populações, identificou um perfil de seleção balanceadora para o segmento codificador do gene *HLA-A*, sendo que as frequências alélicas observadas nas diferentes populações diferiram bastante. Estas diferenças podem ser justificadas pela ação de forças seletivas atuando nessas regiões, devido sua grande importância no processo de apresentação antigênica, sendo influenciado pelo contato com diversos microorganismos atuando nessas populações. No entanto, o perfil de seleção avaliado considerou basicamente polimorfismos presentes nos segmentos codificadores (Solberg *et al.*, 2008) e não considerou os segmentos relacionados com a regulação da expressão das moléculas HLA, tais como regiões promotoras e 3' não traduzidas. Além disso, padrões de migração, deriva genética e efeito fundador são outros fatores que podem levar a modificações nas frequências alélicas (Hartl e Clark, 2007).

Os genes do complexo HLA estão entre os alvos da seleção natural em humanos. Em especial destaca-se seleção balanceadora atuando nos genes clássicos, elevando a variabilidade nestes genes (Solberg *et al.*, 2008) e seleção purificadora atuando nos genes não clássicos. O mecanismo subjacente à baixa variabilidade detectada nas regiões codificadoras destes últimos (genes não clássicos) ainda não foi esclarecido, tornando-os virtualmente invariáveis quando comparados aos seus vizinhos *HLA-A*, *HLA-B* e *HLA-C* (Castelli *et al.*, 2011; Veiga-Castelli *et al.*, 2012; Felicio *et al.*, 2014). Também foram encontradas evidências de seleção balanceadora atuando nas regiões regulatórias 5' e 3' não traduzidas do gene *HLA-G*, mantendo uma alta heterozigose entre os haplótipos encontrados (Castelli *et al.*, 2011). Esta assinatura de seleção balanceadora atuando nas regiões regulatórias do *HLA-G*, porém não na região codificadora, foi encontrada em três diferentes populações Africanas: Senegal, Bênin e Congo (Gineau *et al.*,

2015). Embora seja consenso a atuação de seleção balanceadora na região que codifica a fenda de ligação ao peptídeo das moléculas clássicas (assim como nas regiões regulatórias do gene *HLA-G*), não está claro se esse mesmo perfil de seleção também ocorre nos demais segmentos dos genes clássicos ou em suas regiões regulatórias.

Levando em consideração a elevada variabilidade na fenda de ligação de peptídeos, alguns estudos discutem a possibilidade de Vantagem do Alelo Divergente (DAA, do inglês, *Divergent Allele Advantage*). De acordo com esta proposta, indivíduos heterozigotos que carregam alelos divergentes, teriam uma habilidade aumentada de apresentar uma maior gama de peptídeos comparado aos indivíduos carregando alelos próximos (Buhler *et al.*, 2016). No entanto, muitos estudos populacionais foram desenvolvidos nas últimas décadas e detectou-se uma observação contraditória a essas hipóteses: uma alta proporção de homozigotos em um ou mais genes HLA em populações pequenas e isoladas como as Ameríndias, Tailandesas aborígenes e ilhas do Pacífico, como resultado do rápido efeito da deriva genética durante suas histórias de migração (Black, 1992; Buhler *et al.*, 2016). Neste contexto, a homogeneidade genética dos Ameríndios poderia ser associada a uma alta taxa de mortalidade dessa população após a colonização européia, o que os deixariam susceptíveis a uma série de doenças (Black, 1992). Contudo, esse raciocínio não se confirma nessas populações, *i.e.*, populações tailandesas são extremamente populosas, apesar de serem consideradas as mais homogêneas quando levamos em consideração o *loci* HLA (Lin *et al.*, 2000).

Além disso, dados demográficos apontam que a população Ami e Paiwan exibem 43% e 76% de homogeneidade, respectivamente, para o *loci* *HLA-A*, com o alelo *HLA-A* 24:02* atingindo frequência de 64.0% e 86.3%, respectivamente (Buhler *et al.*, 2016). Os dados encontrados não indicam uma possibilidade de extinção dessa tribo, sugerindo então que outros mecanismos estão envolvidos para assegurar uma resposta imune adequada nessas populações (Buhler *et al.*, 2016). Quando outras espécies são analisadas, observa-se um decréscimo na expectativa de vida associado a homogeneidade das moléculas HLA. No entanto, isso não acontece para todas espécies, destacando o envolvimento de outras forças de seleção natural (Froeschke e Sommer, 2005; Huchard *et al.*, 2010; Worley *et al.*, 2010)

Recentemente um estudo demonstrou uma elevada frequência de variantes deletérias em genes não-HLA localizados no complexo MHC, sendo que a frequência dessas variantes aumenta conforme a proximidade com genes HLA, indicando uma dependência de ligação genética. Os resultados revelaram um custo indireto dessa diversidade genética mantida por meio da ação de seleção balanceadora, o que na maioria das vezes é visto como algo positivo,

trazendo implicações para o processo de recombinação, bem como para a epidemiologia de diversas doenças associadas ao complexo HLA (Lenz *et al.*, 2016).

Entendendo os testes de assinatura de seleção natural

Os testes para detecção das assinaturas de seleção natural baseiam-se em testes de desvio da neutralidade. Entende-se por neutralidade a condição em que o alvo da análise (*i.e* um determinado gene) sob observação não é afetado pela ação da seleção natural. Um ponto importante a se destacar é que os modelos neutros admitem a possibilidade de intensa ação purificadora contra mutações fortemente deletérias (Bitarello, 2011;2016). Assim que a teoria neutra foi proposta, permitiu-se o desenvolvimento de uma série de testes para a teoria nula da neutralidade, que considera que as mudanças evolutivas resultam exclusivamente por conta da ocorrência de mutações e deriva genética, quando a teoria neutra é rejeitada, considera-se hipóteses alternativas incluindo seleção positiva, purificadora ou balanceadora (Meyer *et al.*, 2018). Uma vez que a hipótese nula é baseada no modelo neutro da evolução, esses testes são chamados de testes de neutralidade. Entre eles, destacamos os testes baseados na *no desequilíbrio de ligação e diversidade, testes baseados no padrão de distribuição de frequência alélica, testes de contraste de variação intra e inter espécies e polimorfismos compartilhados entre espécies* (Meyer *et al.*, 2018).

Os testes baseados em **LD e diversidade** consideram que, sob forte ação de seleção positiva, as frequências aumentam rapidamente, reduzindo a diversidade de *locus* neutros ao redor. Para seleção balanceadora, onde duas ou mais variantes são mantidas, a variação genética local aumenta na região. Esses testes são excelentes para detectar seleção positiva recente e menos específicos para detectar seleção balanceadora antiga (Meyer *et al.*, 2018).

Os testes baseados nos **padrões de distribuição de frequência alélica** comparam a distribuição das frequências alélicas observadas com aquelas esperadas sob neutralidade. Incluem-se aqui o teste *D de Tajima, D de Fu e Li, F de Fu e Li*, dentre outros. Quando consideramos a seleção positiva, espera-se observar o aumento no número de variantes de baixa frequência, enquanto a seleção balanceadora aumenta o número de variantes com frequência intermediária acerca do sítio selecionado. Esses testes podem detectar a presença de seleção natural atuando ao longo de muitos períodos de forma comparativa, são mais eficientes quando comparados aos testes de *LD* e *Fst*, porém são influenciados por parâmetros demográficos (Fu e Li, 1993; Zhang *et al.*, 2015; Meyer *et al.*, 2018). No presente trabalho os testes descritos a

seguir foram alternativas empregadas para avaliar o desvio da neutralidade e determinar as assinaturas de seleção presentes no gene *HLA-A*, sendo eles (a) teste de Ewens-Watterson (Ewens, 1972; Watterson, 1978), que indica a possibilidade de neutralidade em determinado segmento, onde pode acontecer uma seleção direcionada (na qual identificamos um excesso de homozigose), ou atuação de seleção balanceadora, prevalecendo uma heterozigose maior que o esperado; (b) teste *D de Tajima* (Tajima, 1989), que compara as estimativas Θ derivado do número médio de diferenças observadas nas sequências (π) e o do número de sítios de segregação (S), utilizando um algoritmo de simulação de coalescência (Tajima, 1989; Excoffier e Lischer, 2010) e (c) teste de Fu e Li (Fu e Li, 1993), que compara a estimativa Θ derivado do número de sítios de segregação (S) e o número de mutações em singleton (alelos que aparecem apenas uma vez na amostra), bem como compara as diferenças nos pareamentos (π) e o número de mutações singleton. Este teste permite a avaliação de pontos de variação muito antigos (Fu e Li, 1993). O uso de diferentes testes de seleção evita resultados falso positivos, uma vez que variações nas frequências alélicas podem acontecer em função de efeitos demográficos, como deriva genética e efeito fundador (Gineau *et al.*, 2015).

Os testes contrastando **variação inter e intra espécies** incluem os testes de *dN/dS* (taxa de proporções de mutações não-sinônimas para sinônimas) e os testes *HKA* e *MK* (comparam polimorfismos e níveis de divergência). As assinaturas de seleção positiva e balanceadora são compatíveis com elevados valores de *dN/dS*, porém apresentam diferentes perfis quando os testes são aplicados. A seleção positiva apresenta baixo nível de polimorfismo comparado a divergência dentro de uma determinada espécie. A seleção balanceadora apresenta elevado polimorfismo em relação a divergência. Esses testes são excelentes para detectar seleção antiga, superior a 1 milhão de anos (Shapiro e Alm, 2008; Meyer *et al.*, 2018; Bitarello, 2011; 2016).

Referências

- ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. **Cellular and molecular immunology**. 6th. Philadelphia: Saunders Elsevier, 2007. viii, 566 p.
- ANDRES, A. M. et al. Targets of balancing selection in the human genome. **Mol Biol Evol**, v. 26, n. 12, p. 2755-64, Dec 2009.
- AYALA GARCIA, M. A. et al. The major histocompatibility complex in transplantation. **J Transplant**, v. 2012, p. 842141, 2012.
- BARROW, A. D.; TROWSDALE, J. The extended human leukocyte receptor complex: diverse ways of modulating immune responses. **Immunol Rev**, v. 224, p. 98-123, Aug 2008.
- BERNATCHEZ, L.; LANDRY, C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? **J Evol Biol**, v. 16, n. 3, p. 363-77, May 2003.
- BITARELLO, B. D. **Natural selection on HLA genes: a molecular investigation of the location and timing of selection events**. 2011. (Mestrado). Instituto de Biociências, Universidade de São Paulo.
- BITARELLO, B. D. **Seleção balanceadora no genoma humano: relevância biológica e consequências deletérias**. 2016. (Doutorado). Instituto de Biociências, Universidade de São Paulo
- BLACK, F. L. Why did they die? **Science**, v. 258, n. 5089, p. 1739-40, Dec 11 1992.
- BRADLEY, B. A. The role of HLA matching in transplantation. **Immunol Lett**, v. 29, n. 1-2, p. 55-9, Jul 1991.
- 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. 401-416, Jul 2016.
- CAILLAT-ZUCMAN, S. Molecular mechanisms of HLA association with autoimmune diseases. **Tissue Antigens**, v. 73, n. 1, p. 1-8, Jan 2009.
- CAROSELLA, E. D. et al. HLA-G: from biology to clinical benefits. **Trends Immunol**, v. 29, n. 3, p. 125-32, Mar 2008.
- CASTELLI, E. C. et al. A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. **Mol Biol Evol**, v. 28, n. 11, p. 3069-86, Nov 2011.
- CASTELLI, E. C. et al. Hla-mapper: An application to optimize the mapping of HLA sequences produced by massively parallel sequencing procedures. **Hum Immunol**, v. 79, n. 9, p. 678-684, Sep 2018.
- CHINEN, J.; BUCKLEY, R. H. Transplantation immunology: solid organ and bone marrow. **J Allergy Clin Immunol**, v. 125, n. 2 Suppl 2, p. S324-35, Feb 2010.
- CHOO, S. Y. The HLA system: genetics, immunology, clinical testing, and clinical implications. **Yonsei Med J**, v. 48, n. 1, p. 11-23, Feb 28 2007.

- CORDON-CARDO, C. et al. Expression of HLA-A,B,C antigens on primary and metastatic tumor cell populations of human carcinomas. **Cancer Res**, v. 51, n. 23 Pt 1, p. 6372-80, Dec 1 1991.
- CRISPIM, J. C. et al. Human leukocyte antigen-G expression after kidney transplantation is associated with a reduced incidence of rejection. **Transpl Immunol**, v. 18, n. 4, p. 361-7, Feb 2008.
- CRISPIM, J. C. et al. HLA polymorphisms as incidence factor in the progression to end-stage renal disease in Brazilian patients awaiting kidney transplant. **Transplantation Proceedings**, v. 40, n. 5, p. 1333-1336, Jun 2008.
- DA SILVA, G. B. et al. Expression of the Classical and Nonclassical HLA Molecules in Breast Cancer. **Int J Breast Cancer**, v. 2013, p. 250435, 2013.
- DARWIN, C. **On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life**. London: J. Murray, 1859. ix, 502 p., 1 folded leaf of plates
- DAVIS, D. M. et al. The transmembrane sequence of human histocompatibility leukocyte antigen (HLA)-C as a determinant in inhibition of a subset of natural killer cells. **J Exp Med**, v. 189, n. 8, p. 1265-74, Apr 19 1999.
- DAWKINS, R. et al. Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. **Immunol Rev**, v. 167, p. 275-304, Feb 1999.
- DOHERTY, P. C.; ZINKERNAGEL, R. M. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. **Nature**, v. 256, n. 5512, p. 50-2, Jul 3 1975.
- 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.
- DONG, G. et al. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. **Immunity**, v. 30, n. 1, p. 21-32, Jan 16 2009.
- EISEN, H. N. et al. Promiscuous binding of extracellular peptides to cell surface class I MHC protein. **Proc Natl Acad Sci U S A**, v. 109, n. 12, p. 4580-5, Mar 20 2012.
- EWENS, W. J. The sampling theory of selectively neutral alleles. **Theor Popul Biol**, v. 3, n. 1, p. 87-112, Mar 1972.
- EXCOFFIER, L.; LISCHER, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Mol Ecol Resour**, v. 10, n. 3, p. 564-7, May 2010.
- FALK, K. et al. Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. **Immunogenetics**, v. 39, n. 4, p. 230-42, 1994.
- FALK, K. et al. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. 1991. **J Immunol**, v. 177, n. 5, p. 2741-7, Sep 01 2006.
- FAYEN, J. et al. Class I MHC alpha 3 domain can function as an independent structural unit to bind CD8 alpha. **Mol Immunol**, v. 32, n. 4, p. 267-75, Mar 1995.

- 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, Feb 2014.
- FERRER-ADMETLLA, A. et al. Balancing selection is the main force shaping the evolution of innate immunity genes. **J Immunol**, v. 181, n. 2, p. 1315-22, Jul 15 2008.
- FLAJNIK, M. F.; KASAHARA, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. **Nat Rev Genet**, v. 11, n. 1, p. 47-59, Jan 2010.
- FOGDELL-HAHN, A. et al. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. **Tissue Antigens**, v. 55, n. 2, p. 140-8, Feb 2000.
- FROESCHKE, G.; SOMMER, S. MHC class II DRB variability and parasite load in the striped mouse (*Rhabdomys pumilio*) in the Southern Kalahari. **Mol Biol Evol**, v. 22, n. 5, p. 1254-9, May 2005.
- FU, Y. X.; LI, W. H. Statistical tests of neutrality of mutations. **Genetics**, v. 133, n. 3, p. 693-709, Mar 1993.
- GAO, G. F. et al. Classical and nonclassical class I major histocompatibility complex molecules exhibit subtle conformational differences that affect binding to CD8alpha. **J Biol Chem**, v. 275, n. 20, p. 15232-8, May 19 2000.
- GARRIGAN, D.; HEDRICK, P. W. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. **Evolution**, v. 57, n. 8, p. 1707-22, Aug 2003.
- GAUDIERI, S. et al. Different evolutionary histories in two subgenomic regions of the major histocompatibility complex. **Genome Res**, v. 9, n. 6, p. 541-9, Jun 1999.
- GERAGHTY, D. E. et al. The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. **J Immunol**, v. 149, n. 6, p. 1934-46, Sep 15 1992.
- 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, Jan-Feb 2015.
- GOLDBERG, A. C.; RIZZO, L. V. MHC structure and function - antigen presentation. Part 1. **Einstein (Sao Paulo)**, v. 13, n. 1, p. 153-6, Jan-Mar 2015.
- GRIFONI, A. et al. Role of HLA-B alpha-3 domain amino acid position 194 in HIV disease progression. **Mol Immunol**, v. 53, n. 4, p. 410-3, Apr 2013.
- HANSASUTA, P. et al. Recognition of HLA-A3 and HLA-A11 by KIR3DL2 is peptide-specific. **Eur J Immunol**, v. 34, n. 6, p. 1673-9, Jun 2004.
- HANSEN, J. A. et al. Genetics of allogeneic hematopoietic cell transplantation. Role of HLA matching, functional variation in immune response genes. **Immunol Res**, v. 41, n. 1, p. 56-78, 2008.
- HARADA, A. et al. Clinical implication of human leukocyte antigen (HLA)-F expression in breast cancer. **Pathol Int**, v. 65, n. 11, p. 569-74, Nov 2015.
- HARBO, H. F. et al. Genes in the HLA class I region may contribute to the HLA class II-associated genetic susceptibility to multiple sclerosis. **Tissue Antigens**, v. 63, n. 3, p. 237-47, Mar 2004.

- HARTL, D. L.; CLARK, A. G. **Principles of population genetics**. 4th. Sunderland, Mass.: Sinauer Associates, 2007. xv, 652 p.
- HUCHARD, E. et al. MHC, mate choice and heterozygote advantage in a wild social primate. **Mol Ecol**, v. 19, n. 12, p. 2545-61, Jun 1 2010.
- HUGHES, A. L.; YEAGER, M. Natural selection and the evolutionary history of major histocompatibility complex loci. **Front Biosci**, v. 3, p. d509-16, May 26 1998.
- JANEWAY CA JR, T. P., WALPORT M, ET AL. . Immunobiology: The Immune System in Health and Disease. **New York: Garland Science.**, v. 5th edition 2001.
- KIEPIELA, P. et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. **Nature**, v. 432, n. 7018, p. 769-75, Dec 9 2004.
- KLEIN, J.; SATO, A. The HLA system. First of two parts. **N Engl J Med**, v. 343, n. 10, p. 702-9, Sep 7 2000a.
- KLEIN, J.; SATO, A. The HLA system. Second of two parts. **N Engl J Med**, v. 343, n. 11, p. 782-6, Sep 14 2000b.
- KOEHLER, R. N. et al. Class I HLA-A*7401 is associated with protection from HIV-1 acquisition and disease progression in Mbeya, Tanzania. **J Infect Dis**, v. 202, n. 10, p. 1562-6, Nov 15 2010.
- KULSKI, J. K.; GAUDIERI, S.; DAWKINS, R. L. Using alu J elements as molecular clocks to trace the evolutionary relationships between duplicated HLA class I genomic segments. **J Mol Evol**, v. 50, n. 6, p. 510-9, Jun 2000.
- LEDERBERG, J. Ontogeny of the clonal selection theory of antibody formation. Reflections on Darwin and Ehrlich. **Ann N Y Acad Sci**, v. 546, p. 175-82, 1988.
- LENZ, T. L. et al. Excess of Deleterious Mutations around HLA Genes Reveals Evolutionary Cost of Balancing Selection. **Mol Biol Evol**, v. 33, n. 10, p. 2555-64, Oct 2016.
- LEWONTIN, R. C.; KRAKAUER, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. **Genetics**, v. 74, n. 1, p. 175-95, May 1973.
- 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.
- LIN, M. et al. Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals. **Tissue Antigens**, v. 55, n. 1, p. 1-9, Jan 2000.
- MARTINEZ, J. et al. Expression of HLA-E molecules in the placental tissue of women infected with HIV-1 and uninfected women. **Placenta**, v. 49, p. 33-36, Jan 2017.
- MATZARAKI, V. et al. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. **Genome Biol**, v. 18, n. 1, p. 76, Apr 27 2017.
- MEYER, D. et al. Signatures of demographic history and natural selection in the human major histocompatibility complex Loci. **Genetics**, v. 173, n. 4, p. 2121-42, Aug 2006.
- MEYER, D.; THOMSON, G. How selection shapes variation of the human major histocompatibility complex: a review. **Annals of Human Genetics**, v. 65, p. 1-26, Jan 2001.

- MEYER, D. et al. A genomic perspective on HLA evolution. **Immunogenetics**, v. 70, n. 1, p. 5-27, Jan 2018.
- MURPHY, K. et al. **Janeway's immunobiology**. 8th. New York: Garland Science, 2012. xix, 868 p.
- NEJENTSEV, S. et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. **Nature**, v. 450, n. 7171, p. 887-92, Dec 6 2007.
- PARDIGON, N. et al. Role of co-stimulation in CD8+ T cell activation. **Int Immunol**, v. 10, n. 5, p. 619-30, May 1998.
- PENN, D. J.; DAMJANOVICH, K.; POTTS, W. K. MHC heterozygosity confers a selective advantage against multiple-strain infections. **Proc Natl Acad Sci U S A**, v. 99, n. 17, p. 11260-4, Aug 20 2002.
- RAMSURAN, V. et al. Epigenetic regulation of differential HLA-A allelic expression levels. **Hum Mol Genet**, v. 24, n. 15, p. 4268-75, Aug 1 2015.
- ROBINSON, J. et al. The IPD-IMGT/HLA Database - New developments in reporting HLA variation. **Hum Immunol**, v. 77, n. 3, p. 233-237, Mar 2016.
- 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.
- SABETI, P. C. et al. Positive natural selection in the human lineage. **Science**, v. 312, n. 5780, p. 1614-20, Jun 16 2006.
- SAKAGUCHI, S. et al. Regulatory T cells and immune tolerance. **Cell**, v. 133, n. 5, p. 775-87, May 30 2008.
- SCHAEFER, M. R. et al. A novel trafficking signal within the HLA-C cytoplasmic tail allows regulated expression upon differentiation of macrophages. **J Immunol**, v. 180, n. 12, p. 7804-17, Jun 15 2008.
- SCHMID, B. V.; KESMIR, C.; DE BOER, R. J. Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway. **Epidemics**, v. 2, n. 3, p. 99-108, Sep 2010.
- SHAPIRO, B. J.; ALM, E. J. Comparing patterns of natural selection across species using selective signatures. **PLoS Genet**, v. 4, n. 2, p. e23, Feb 2008.
- SHELDON, S.; POULTON, K. HLA typing and its influence on organ transplantation. **Methods Mol Biol**, v. 333, p. 157-74, 2006.
- SHIINA, T. et al. The HLA genomic loci map: expression, interaction, diversity and disease. **J Hum Genet**, v. 54, n. 1, p. 15-39, Jan 2009.
- SMYTH, M. J.; GODFREY, D. I.; TRAPANI, J. A. A fresh look at tumor immunosurveillance and immunotherapy. **Nat Immunol**, v. 2, n. 4, p. 293-9, Apr 2001.
- SOLBERG, O. D. et al. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. **Hum Immunol**, v. 69, n. 7, p. 443-64, Jul 2008.
- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, v. 123, n. 3, p. 585-95, Nov 1989.

- TOURAINÉ, J. L. et al. HLA class II antigens: structure, function, and expression in immunodeficiencies, autoimmune diseases, and allograft rejection. **Adv Nephrol Necker Hosp**, v. 18, p. 325-34, 1989.
- 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, Jan 2012.
- VIRANI, S. et al. Cancer epigenetics: a brief review. **ILAR J**, v. 53, n. 3-4, p. 359-69, 2012.
- WATTERSON, G. A. The homozygosity test of neutrality. **Genetics**, v. 88, n. 2, p. 405-17, Feb 1978.
- WESLEY, P. K. et al. The CD8 coreceptor interaction with the alpha 3 domain of HLA class I is critical to the differentiation of human cytotoxic T-lymphocytes specific for HLA-A2 and HLA-Cw4. **Hum Immunol**, v. 36, n. 3, p. 149-55, Mar 1993.
- WORLEY, K. et al. MHC heterozygosity and survival in red junglefowl. **Mol Ecol**, v. 19, n. 15, p. 3064-75, Aug 2010.
- YANG, Y.; SEMPE, P.; PETERSON, P. A. Molecular mechanisms of class I major histocompatibility complex antigen processing and presentation. **Immunol Res**, v. 15, n. 3, p. 208-33, 1996.
- YORK, I. A.; ROCK, K. L. Antigen processing and presentation by the class I major histocompatibility complex. **Annu Rev Immunol**, v. 14, p. 369-96, 1996.
- ZHANG, Q.; TYLER-SMITH, C.; LONG, Q. An extended Tajima's D neutrality test incorporating SNP calling and imputation uncertainties. **Stat Interface**, v. 8, n. 4, p. 447-456, Oct 1 2015.

Conclusões

Em conclusão, nós desenvolvemos uma metodologia altamente eficaz para analisar a variabilidade dos genes HLA, considerando todo seu segmento, incluindo o promotor distal e o segmento completo da região 3' não-traduzida. A região do promotor estendido do gene *HLA-A* é altamente polimórfica, com a presença de poucas sequências, mas altamente divergentes (Figura S3). Os sítios de variação detectado ao longo de todo o segmento gênico estão altamente associados em função do alto LD observado. Como consequência do alto LD, cada grupo de alelos codificadores do gene *HLA-A* (ou grupo de alelos próximos relacionados) constitui uma linhagem *HLA-A* distinta, cada uma caracterizada por um único promotor e uma única sequência 3' não-traduzida. Como esperado, uma elevada diversidade nucleotídica foi observada para os éxons 2 e 3, mas o mesmo padrão também foi observado para o segmento promotor completo, outros éxons e regiões intrônicas.

Quanto as assinaturas de seleção natural atuando neste gene, foi surpreendente detectar o perfil de seleção balanceadora atuando em todo segmento, uma vez que apostávamos na possibilidade de perfis de seleção diferentes atuando nos segmentos regulatórios do gene *HLA-A*. Avaliações funcionais são necessárias para o entendimento completo da biologia e padrão de expressão deste gene, contudo aqui supomos que o *HLA-A* possui um padrão de regulação da expressão muito fino o que garante que seu padrão de expressão seja ubíquo. Essa regulação fina provavelmente se explica pela conservação das regiões do core do promotor, mantendo os elementos chaves para a ação da maquinaria de transcrição, enquanto o promotor estendido apresenta muitos sítios de variação, o que pode modular a expressão deste gene, possibilitando a ligação de diferentes fatores de transcrição a depender do tecido e do microambiente em que se encontra. A região 3' não-traduzida é o segmento mais conservado, provavelmente relacionado com a manutenção de níveis de expressão adequados da molécula HLA-A. Além disso, o domínio $\alpha 3$ do *HLA-A*, que interage com CD8, também parece apresentar certo grau de conservação, de certa forma garantindo uma adequada interação entre HLA-A/CD8, e a porção 3'NT é altamente conservada, o que provavelmente favorece a expressão diferencial do gene *HLA-A* nos tecidos com microRNAs e microambientes diferentes.