



**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO
DE MESQUITA FILHO”
FACULDADE DE MEDICINA**

Thálitta Hetamaro Ayala Lima

**ESTUDO DA VARIABILIDADE DA REGIÃO CODIFICADORA E
3’NÃO TRADUZIDA DO GENE *HLA-F* NO BRASIL**

Dissertação apresentada à Faculdade de Medicina, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Botucatu, como requisito para obtenção do título de Mestre em Patologia.

Orientador: Prof. Dr. Erick da Cruz Castelli

**Botucatu
2015**

Thálitta Hetamaro Ayala Lima

**ESTUDO DA VARIABILIDADE DA REGIÃO CODIFICADORA E
3'NÃO TRADUZIDA DO GENE *HLA-F* NO BRASIL**

Dissertação apresentada à Faculdade de
Medicina, Universidade Estadual
Paulista “Júlio de Mesquita Filho”,
Campus de Botucatu, como requisito
para obtenção do título de Mestre em
Patologia.

Orientador: Prof. Dr. Erick da Cruz Castelli

Botucatu

2015

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Lima, Thálitta Hetamaro Ayala.

Estudo da variabilidade da região codificadora e 3'Não
traduzida do Gene HLA-F no Brasil / Thálitta Hetamaro
Ayala Lima. - Botucatu, 2015

Dissertação (mestrado) - Universidade Estadual Paulista
"Júlio de Mesquita Filho", Faculdade de Medicina de
Botucatu

Orientador: Erick da Cruz Castelli
Capes: 40105008

1. Genética. 2. Patologia. 3. Variação genética. 4.
Sequenciamento genético. 5. Antígenos de
histocompatibilidade HLA. 6. Histocompatibilidade.

Palavras-chave: HLA-F; Sequenciamento de Nova Geração;
Variabilidade.

Epígrafe

"Foi o tempo que dedicaste à tua rosa que a fizeste tão importante"

Antoine de Saint-Exupéry

Dedicatória

“Quando eu cheguei a este mundo, não sabia ao certo o que estava fazendo aqui, até que percebi que haviam pessoas que eu poderia contar para me orientarem na jornada.

Vocês são firmes, quando é da firmeza que eu preciso...

Vocês são ternos, quando é da ternura que eu necessito...

Vocês são amigos, quando é direção que eu busco...

Vocês são o abrigo, quando é da companhia que estou em busca...

Vocês são os ensinamentos necessários, quando decisões preciso tomar...

Quando eu cheguei a este mundo, não sabia ao certo o que estava fazendo ...

Hoje, bem... eu sei claramente o que estou fazendo aqui porque vocês fizeram mais que apenas me orientar; caminharam ao meu lado muitas vezes, me seguiram de perto outras tantas, e andaram à minha frente muitas outras, deixando rastros de luz, como diretrizes seguras que eu pudesse seguir.

Hoje eu sei muito bem o papel que me cabe na construção de um mundo melhor porque isso eu aprendi com vocês.

E peço a Deus por vocês.”

Em especial ao meu querido e amado avô/pai Sebastião Rommel Ferreira de Alencar, por sempre me incentivar e acreditar que esse dia fosse possível, por me acompanhar e educar. Sem palavras para agradecer a diferença que fez na minha vida, eu peço a Deus que do Alto receba meu muito obrigada e ‘aquele amplexo’.

À minha amada avó/mãe Heloisa Helena Melo de Alencar, pelo incomparável carinho, amor e dedicação. Umas das mulheres mais inteligentes que já conheci, obrigada por sempre ‘pegar no pé’ e acreditar que a educação é a alternativa na qual devemos apostar para a construção de um futuro melhor. A senhora é

minha inspiração, exemplo de mulher e profissional, visto a postura e moral que possuí e a exímia professora que fui durante 26 anos de profissão.

A minha querida e amável mãe Bianca Ayala Melo Di Alencar, por acreditar que sou capaz, por apostar nas minhas decisões e me acompanhar passo a passo.

Por ser meu porto seguro, calma e amparo nos momentos difíceis e por ser a fonte dos meus sorrisos mais sinceros. Agradeço a Deus todos os dias por ter me dado essa oportunidade de ser sua filha. Aprendo com você todos os dias.

Ao meu namorado Bruno Ribeiro Darros, pelo amor, companheirismo, incentivo, compreensão e fraternidade e também por ser uma pessoa que traz consigo as virtudes de um homem de bem. Obrigada pelas palavras de ânimo e cafés que sempre me acalmam quando a saudade de casa e ‘correria’ batem à porta. Agradeço muito a Deus por ter colocado você em minha vida.

*Agradecimentos
Especiais*

Ao Prof. Dr. Erick da Cruz Castelli, pela oportunidade, orientação, empenho e disponibilidade; pelo zelo e comprometimento que sempre demonstra ter para com sua profissão.

À Universidade Estadual Paulista “Júlio Mesquita Filho”, Campus Botucatu, que sempre me recebeu prontamente para o trabalho ao qual me propus.

Ao programa de Pós-Graduação em Patologia da Faculdade de Medicina de Botucatu, pelo acolhimento e infraestrutura.

A todos os doadores das amostras de sangue, que contribuíram sobretudo para que este estudo fosse realizado.

Agradecimientos

Primeiramente a Deus, que a cada amanhecer renova minhas forças e esperança, guiando-me na direção correta, por ser meu porto seguro em todos os momentos, por ter me abençoado nessa jornada a quilômetros de casa.

Ao professor Dr. Deilson Elgui Oliveira, pela educação de qualidade que compartilha com seus alunos, pela exímia disciplina e inteligência. Sinto-me honrada pela oportunidade de ter sido sua aluna na pós-graduação.

A todos mestres e doutores da pós-graduação, cada qual com seu modo singular de ensinar e compartilhar experiências. Meus sinceros agradecimentos! Sinto-me privilegiada de ter sido aprendiz desse seleto 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 meu crescimento profissional e pessoal.

Aos secretários da pós-graduação em Patologia, Vânia Soler e Diego Cezario Bovolim de Oliveira, pelo apoio e esclarecimentos que surgiram nessa jornada. Pelo excelente trabalho que desempenham, facilitando sempre nosso dia-a-dia.

À Dra. Denise Fecchio, pela calma e dedicação com as quais coordena a pós-graduação em Patologia e sobretudo, pela pessoa amável, dedicada e presente.

À Dra. Luciane Alarcão Dias-Melício, pela troca de conhecimentos científicos, disciplinas adoráveis e pelas imensas boas risadas que demos juntas.

Ao Dr. Celso Teixeira Mendes-Júnior, pelo trabalho de qualidade que realiza, pela exímia dedicação e comprometimento com a ciência.

Ao Dr. Ramon Kaneno, pelas contribuições e sugestões a este trabalho, por sempre se prontificar a ajudar-me. Pela educação de qualidade que compartilha com seus alunos e também com comprometimento com o saber científico de qualidade.

A Dra. Maria Inês de Moura Campos Pardini, pelas contribuições dirigidas a este trabalho durante o Exame Geral de Qualificação e por sempre se mostrar disponível para nos ajudar.

Aos familiares que com palavras de ânimo sempre me incentivaram fazer o melhor e depositaram em mim total confiança; em especial agradeço aos meus tios Pablo Ayala Melo Di Alencar, Diego Ayala Melo Di Alencar e Thiago Ayala Melo Di Alencar e suas respectivas esposas Luciana Teixeira, Renata Teles Silva e Karinna Ferreira de Sousa Matias.

A querida Dra. Carla Silva, pela amizade que construímos ao longo desse tempo, por me receber tão bem e fazer com que eu não me sentisse tão longe de casa.

Ao amigo Marcos Franchi, pela amizade e risadas que compartilhamos.

Aos “avós” que ganhei Dona Helecine Cresti Ribeiro e Seu Adhemar Gonçalves Ribeiro, pela amizade, conselhos e carinho dedicados a mim e também por me fazerem sentir parte da família.

Ao “Tio” Silvio César Gonçalves Ribeiro e sua esposa Daniela Polo Camargo da Silva, pela amizade fiel e companheirismo, pelas boas risadas e momentos divertidos que sempre passamos juntos.

À amiga Iane de Oliveira Pires Porto, que além de colegas de profissão, tornamos grandes amigas, pelo intercâmbio e discussões acadêmicas que sempre desenvolvemos, pelas risadas, almoços e quebra-cabeças que fizeram esses dias mais leves e divertidos.

Ao amigo Leandro Prado Felício, pela amizade e comprometimento com meu crescimento, por me ajudar a encontrar soluções diante de dúvidas que surgiram durante essa caminhada e por sempre me ensinar.

Aos colegas de trabalho Luciana Caricati Veiga-Castelli, Jaqueline Ramalho, Renato Buturra, Michelle Paz, Neilton Paulo Bezerra pelas ‘nerdices’, cafés, bolos que sempre compartilhamos. Vocês são os melhores!

A Nathália Souza, por ser companheira de todas as horas, por ceder inúmeras vezes o ombro quando o coração e a saudade de casa apertaram. Muito obrigada!

As novas e boas amigas que fiz, Lígia Maria, Andréia Souza e Paula M. Assis, por tornarem meus dias mais alegres e a certeza de saber que posso contar com vocês.

Aos pequenos Nícolas Ayala Melo Teixeira, Laura Ayala Teles e Lorena Darros por representarem a esperança de um futuro melhor.

As amigas Fernanda Dias, Ariane Melo, Grazielle Guimarães e Karla Cristina pela amizade sincera e credibilidade.

À minha bicicleta e às corridas que serviram de conforto e escape nos momentos de estresse e tensão.

A todos que contribuíram direta ou indiretamente para este dia.

A todos vocês, meu muito obrigada!

Resumo

RESUMO

HLA-F é um gene de classe I não clássico do Complexo Principal de Histocompatibilidade humano. Este *locus* difere-se dos genes clássicos pelo baixo polimorfismo e diferente padrão de expressão. A exata função do gene *HLA-F* ainda permanece desconhecida. Acredita-se que o *HLA-F* possua uma função tolerogênica e imunomodulatória. Atualmente, há pouca informação acerca da variabilidade do gene *HLA-F* e os estudos disponíveis avaliaram apenas pequenos segmentos do gene e/ou buscaram por variantes já conhecidas. Neste estudo apresentamos uma estratégia para a avaliação completa da variabilidade do gene *HLA-F*, incluindo suas regiões regulatórias, usando sequenciamento de segunda geração (ou nova geração). Esta estratégia foi aplicada para descrever a variabilidade de uma amostra de 196 indivíduos oriundos do estado de São Paulo. Os resultados indicam que o gene *HLA-F* é muito conservado, considerando-se as diferentes moléculas codificadas, sendo que todas as sequências encontradas codificam apenas 4 moléculas HLA-F distintas. Uma dessas moléculas, associada ao grupo de alelos F*01:01, representa 82,45% das proteínas codificadas pelas sequências de *HLA-F* encontradas no Brasil. No entanto, a variabilidade nucleotídica e haplotípica encontrada no presente estudo mostrou-se muito superior a já descrita na literatura, embora a maioria das sequências detectadas apresenta mutações sinônimas ou intrônicas. A região 3' não traduzida do gene *HLA-F* mostrou-se pouco variável, com haplótipos bem definidos associados aos alelos de região codificadora. Esta baixa variabilidade proteica está provavelmente associada ao papel crítico do *HLA-F* na fisiologia do sistema imunitário.

Palavras-chave: *HLA-F*, Sequenciamento de Nova Geração, variabilidade.

Abstract

ABSTRACT

HLA-F is a non-classical HLA class I gene and is distinguished from its classical counterparts by low allelic polymorphism and distinctive expression patterns. The exact function of *HLA-F* remains unknown. It is believed that *HLA-F* has tolerogenic and immune modulatory features. Currently, there is little information regarding the *HLA-F* allelic variation among humans and the available studies have evaluated only a fraction of the *HLA-F* gene segment and/or have searched for known alleles only. Here we present a strategy to evaluate the complete *HLA-F* variability including its regulatory segments (promoter and 3'UTR) by using massive parallel sequencing (or second generation sequencing) procedures. *HLA-F* variability was surveyed on 196 individuals from the Southeast of Brazil. The results indicate that the *HLA-F* gene is indeed conserved at the protein level, in which the three coding haplotypes detected encode for only four different HLA-F molecules, and one of these molecules represent 82.45% of all. However, *HLA-F* worldwide haplotype variability is much higher than our current knowledge. The 3'UTR presented few variable sites and well-defined haplotypes that are usually associates with the same coding alleles. This protein conservation is probably a consequence of the *HLA-F*'s key role in the immune system physiology.

Keywords: *HLA-F*, Next Generation Sequencing, variability.

Sumário

Sumário

Capítulo I.....	1
Revisão de Literatura.....	2
Referências	15
Capítulo II.....	22
ABSTRACT	24
Introduction	25
Methods	27
DNA samples and amplification.....	27
Library preparation and sequencing	28
Raw data processing (mapping).....	28
Genotype calling and processing	29
Results	31
Discussion.....	34
Acknowledgements	39
References	39
Conclusão	54
Anexos.....	55

Lista de Figuras da Revisão de Literatura

Figura 1. A) Estrutura das moléculas MHC de classe I. B) Molécula HLA de classe I (magenta) com peptídeo acoplado (esferas verdes), vista superior.	3
Figura 2 - Processamento e apresentação antigênica envolvendo uma molécula MHC de classe I	7

Lista de Tabelas da Revisão de Literatura

Tabela 1. Número de Alelos identificados para os genes de classe I (The IMGT/HLA Database, versão 3.21.1)	8
--	---

Lista de Abreviaturas e Siglas

VCF	(do inglês, <i>Variant Call Format</i>)
μL	Microlitro (10^{-6} litro)
APC	Célula Apresentadora de Antígeno (do inglês, <i>Antigen Presenting Cell</i>)
DNA	Ácido Desoxirribonucleico (do inglês, <i>Deoxyribonucleic acid</i>)
EDTA	Ácido Etileno Diamino Tetracético
HLA	Antígenos Leucocitários Humanos (do inglês, <i>Human Leukocyte Antigens</i>)
IMGT/HLA	International Immunogenetics Database
Kb	Kilobase (10^3 bases)
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
NT	Não traduzida
°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	Estado de São Paulo
TCR	Receptor de Célula T
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>)
B2M	β2- microglobulina

Capítulo I

Revisão de Literatura

O Complexo Principal de Histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*) é composto por mais de 220 genes e localiza-se em um segmento de aproximadamente 3,6 Mb do braço curto do cromossomo seis (6p21.3) (Klein e Sato, 2000a; b). Didaticamente, os genes que compõem o MHC são agrupados em três grupos, denominados de genes de classe I, II ou III. Em humanos, em especial dentro das regiões de classe I e II, está o sistema conhecido como Antígenos Leucocitários Humanos (*HLA*, do inglês *Human Leucocytes Antigens*) (Shiina *et al.*, 2009), cujos genes codificam moléculas associadas ao processamento e apresentação antigênica.

Os genes do complexo *HLA* estão envolvidos na resposta imunitária a patógenos, pois os principais genes deste complexo atuam na apresentação de antígenos aos linfócitos T. Essa função é desempenhada por glicoproteínas transmembrânicas, que se ligam a peptídeos de origem extra e intracelular (oriundos de patógenos ou do próprio organismo) e os apresentam às células do sistema imunitário, principalmente os linfócitos T (Klein e Sato, 2000a).

Os genes de classe I codificam estas proteínas transmembrânicas e estão envolvidos diretamente na apresentação de antígenos intracelulares aos linfócitos durante a resposta imunitária adaptativa. Entre estes genes estão o *HLA-A*, *HLA-B* e *HLA-C*, também chamados de genes clássicos de classe I (Klein e Sato, 2000a; b). Entretanto, na mesma região genômica encontram-se os genes não clássicos de classe I (*HLA-G*, *HLA-E* e *HLA-F*), envolvidos em menor grau com a apresentação de antígenos e em maior grau com modulação do sistema imunitário (Geraghty *et al.*, 1992). Esses últimos são responsáveis, principalmente, pela função de tolerância imunológica por ligarem-se a receptores inibitórios nos linfócitos T e células NK (NK, do inglês, *Natural Killer*) (Donadi *et al.*, 2011).

A estrutura de uma molécula de classe I clássica e não clássica é semelhante (Figura 1), sendo composta por uma cadeia α pesada codificada no cromossomo 6 e uma cadeia β_2 -microglobulina codificada no cromossomo 15, ou seja, fora do complexo *HLA* (Klein e Sato, 2000a). A cadeia α possui cinco domínios principais: dois envolvidos na ligação a peptídeos ($\alpha 1$ e $\alpha 2$), um domínio de imunoglobulina ($\alpha 3$), um domínio transmembrana e um domínio citoplasmático (Figura 1) (Klein e Sato, 2000a). Possivelmente essa semelhança estrutural

ocorre devido ao fato de os genes de classe I terem se originado por duplicações imperfeitas de blocos gênicos (Kulski *et al.*, 2001).

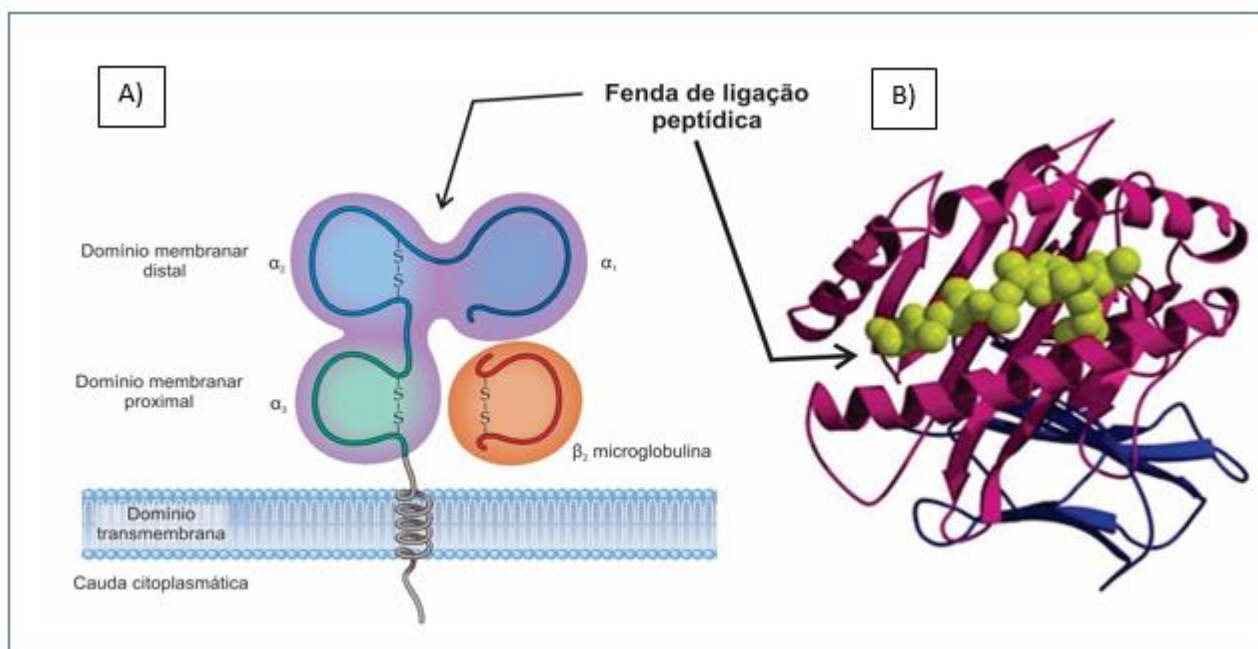


Figura 1. A) Estrutura das moléculas MHC de classe I. Disponível em: https://lookfordiagnosis.com/mesh_info.php?term=HLA-B8+Antigen&lang=1. Acesso em: 21 de Setembro de 2015. B) Molécula HLA de classe I (magenta) com peptídeo acoplado (esferas verdes), vista superior. Fonte: Adaptado de CARRINGTON, M. High Marks for Destruction - Genetic variations boost HIV-killing immune response to slow disease progression. **CCR Connections**, v. 4, n°1, p. 5, 2010. Disponível em: http://home.ccr.cancer.gov/connections/2010/Vol4_No1/docs/news_2.pdf. Acesso em: 22 de Setembro de 2015

Após a tradução, as moléculas MHC de classe I são montadas no retículo endoplasmático, ou seja, ocorre a união das cadeias alfa e beta. Este processo envolve a atuação das chaperonas calnexina, calreticulina, p57 do retículo endoplasmático e tapasina (proteína ligante do heterodímero denominado *Transportador Associado a Processamento de Peptídeos* - TAP). O TAP também é importante para a estabilização das moléculas MHC de classe I recém-sintetizada, que aguardam a ligação de peptídeos, permitindo com que a molécula MHC agora em sua versão completa possa ser exportada à superfície celular (Figura 2) (Klein e Sato, 2000b; Abbas *et al.*, 2007).

Os peptídeos apresentados pelas moléculas MHC de classe I são derivados de proteínas citosólicas, cuja maioria é sintetizada endogenamente em células nucleadas ou são antígenos que entram no citosol por meio de microrganismos fagocitados, microrganismos intracelulares ou de vírus (Murphy, 2010). Portanto, antígenos exógenos também podem ser apresentados via moléculas de classe I. Este mecanismo é possível porque alguns microrganismos conseguem

romper as membranas celulares dos fagossomos e atingirem o citosol, podendo então serem processados da mesma forma que antígenos originados do meio intracelular (Murphy, 2010). O mecanismo para geração desses peptídeos envolve a participação de um complexo proteico, denominado proteassomo (Figura 2), que possui uma ampla faixa de atividade proteolítica. As proteínas destinadas a esse complexo proteico são marcadas através da ligação de um pequeno polipeptídeo, as ubiquitinas. Desta forma, os peptídeos gerados no citosol são transportados ao retículo endoplasmático por meio do TAP e associados a moléculas MHC (Yang *et al.*, 1996; York e Rock, 1996; Abbas *et al.*, 2007). Esses peptídeos associados a moléculas de classe I são apresentados aos linfócitos T CD8+ citotóxicos (Yang *et al.*, 1996).

A apresentação de antígenos também ocorre via moléculas MHC de classe II, contudo a fonte de antígenos proteicos apresentados por essas moléculas são proteínas endossômicas e lisossômicas, internalizadas predominantemente do meio extracelular (Vyas *et al.*, 2008). A geração dos peptídeos a serem apresentados pelas moléculas MHC de classe II ocorre pela ação de enzimas presentes nesses compartimentos, como por exemplo as catepsinas. No entanto, moléculas de classe II apresentam antígenos as células TCD4+. Alguns microrganismos são interiorizados em vesículas denominadas de fagossomos, que possuem a capacidade de se ligarem aos lisossomos, originando desta forma estruturas conhecidas como fagolisossomos, responsáveis pela degradação de antígenos (Abbas *et al.*, 2007; Vyas *et al.*, 2008; Murphy, 2010).

Assim que a molécula MHC de classe I associada ao peptídeo está na superfície celular, o complexo interage com receptores TCR (TCR, do inglês *T cell receptor*) e CD8 dos linfócitos T citotóxicos (CTLs) (Figura 2), ou ainda, com diversos outros receptores em células NK, que podem ser ativadores ou inibitórios. Assim que ocorre a interação entre MHC e TCR, uma série de proteínas estabilizam essa interação, permitindo com que ligante e receptor fiquem unidos tempo o suficiente para o desencadeamento de uma resposta imunológica. A ativação dos linfócitos e células NK por antígenos não próprios em geral resulta na destruição das células apresentando estes antígenos, que em geral representam células infectadas ou anormais (Gao *et al.*, 2000; Klein e Sato, 2000b).

Os tecidos e órgãos linfóides estão distribuídos por todo o nosso corpo. A denominação órgãos linfóides dá-se pelo comprometimento desses tecidos e órgãos com a produção, crescimento e desenvolvimento dos linfócitos. Os órgãos linfóides podem ser divididos em primários ou centrais (medula óssea e timo), responsáveis pela produção e maturação de linfócitos (Thompson, 2012) e secundários ou periféricos (tonsilas, baço, apêndice, *placas de*

peyer, vasos linfáticos e linfonodos) nos quais os linfócitos encontram os estímulos antigênicos, iniciando as respostas imunes adaptativas (Ruddle e Akirav, 2009).

Os linfócitos são as principais células do sistema imune adaptativo, encontradas apenas nos vertebrados, proporcionando meios de defesa mais eficientes frente a infecções causadas pelo mesmo agente infeccioso (Litman *et al.*, 2010).

Cada linfócito virgem, presente na circulação possui um receptor de antígenos com uma única especificidade, contudo, essa especificidade é determinada por um mecanismo de rearranjo gênico especial (mecanismo que acontece durante o desenvolvimento dos linfócitos na medula óssea, afim de gerar diversas variantes dos genes codificadores das moléculas receptoras). Desta forma, embora cada linfócito apresente um receptor com especificidade única, cada linfócito possui uma especificidade distinta, permitindo com que a gama de antígenos estranhos reconhecido por essas células de defesa seja muito grande (Lederberg, 1988; Abbas *et al.*, 2007).

Somente os linfócitos que encontraram um antígeno com o qual seu receptor é capaz de interagir poderão ser ativados afim de proliferarem e se diferenciarem em células efectoras (Pardigon *et al.*, 1998). A ativação das células T necessita mais do que a ligação do TCR ao complexo MHC/peptídeo, sendo também necessários sinais co-estimulatórios (Pardigon *et al.*, 1998). A ausência ou presença desses sinais dita se a interação entre TCR e complexo MHC/peptídeo irá desencadear à ativação ou à indução de tolerância (Sakaguchi *et al.*, 2008).

O número de linfócitos capazes de ligarem-se a um determinado antígeno torna-se restrito à medida que cada célula apresenta um padrão distinto de receptor antigênico. Dessa forma para produzir células efectoras específicas em quantidade suficiente para combater uma infecção, um linfócito deve ser capaz de se proliferar (expansão clonal), antes de sofrer diferenciação em célula efectora. A expansão clonal estabelece que apenas aquela célula capaz de reconhecer um antígeno específico irá se proliferar, desta forma, sendo selecionada em detrimento das outras (Denizot *et al.*, 1986; Murphy, 2010).

Na resposta imune adaptativa três aspectos são essenciais: a) gerar uma diversidade suficiente para combater a gama de antígenos não-próprios aos quais somos expostos; b) distinção próprio/não próprio e c) memória imunológica. Com base no exposto acima sobre expansão clonal, temos que o desenvolvimento dessa memória imunológica seria originada pela expansão no número de um determinado clone ao estímulo antigênico (Denizot *et al.*, 1986).

Quanto a distinção do próprio/ não-próprio, é essencial que o sistema imunitário consiga fazer essa distinção (Kappler *et al.*, 1987). Quando essa distinção não acontece, uma resposta imune

é desencadeada contra antígenos próprios, levando ao desenvolvimento de doenças autoimunes; por outro lado, quando essa distinção é feita corretamente, a ausência dessa resposta imune contra um antígeno próprio é denominada tolerância ao próprio (Sakaguchi *et al.*, 2008).

Como citado anteriormente somente a interação do complexo MHC/peptídeo ao receptor linfocitário não é o suficiente para desencadear a resposta imunológica, podendo ocasionar também morte ou anergia celular. Desta forma, a seleção negativa pode atuar, eliminando ou alterando linfócitos durante seu desenvolvimento, devido a alta afinidade de ligação entre os receptores de antígenos e auto-antígenos, localizados nos órgãos linfoides geradores, quando essa alta afinidade acontece os linfócitos são eliminados por apoptose, processo conhecido como deleção clonal (Kazansky, 2008). 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). Contudo, em algumas situações, as moléculas MHC apresentam antígenos próprios aos linfócitos, podendo levar ao desenvolvimento de doenças autoimunes (Caillat-Zucman, 2009).

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 MHC do hospedeiro nas APCs nos tecidos periféricos (Janeway Ca Jr, 2001).

As moléculas MHC em algumas situações não conseguem realizar uma apresentação antigênica adequada, permitindo a ação de agentes infecciosos (Jobim e Jobim, 2008). Ainda, a ausência dessas moléculas é ocorrência comum em células malignas (Abbas *et al.*, 2007), configurando um possível mecanismo de escape imunológico.

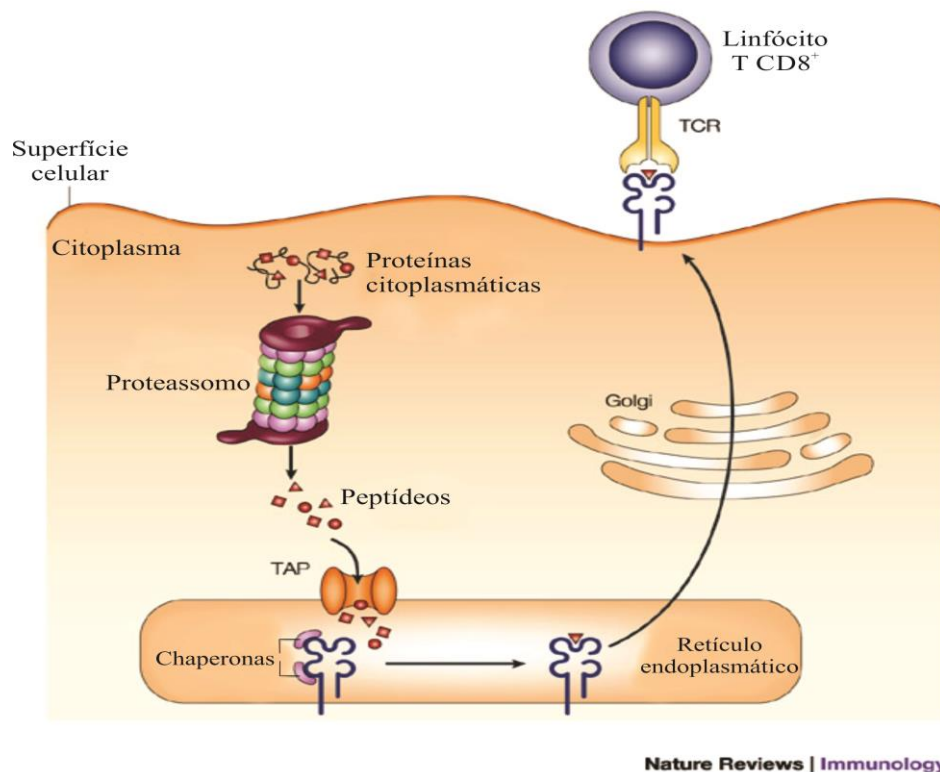


Figura 2 - Processamento e apresentação antigênica envolvendo uma molécula MHC de classe I.

Fonte: Adaptado de 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.

O MHC compreende os genes mais polimórficos do genoma da maioria dos vertebrados. Esta variabilidade está principalmente contida nos genes de classe I clássicos (tabela 1), em especial nos segmentos que codificam a fenda de ligação ao peptídeo (Figura 1). A variação nos sítios de ligação de peptídeos tem sido considerada a principal responsável pela habilidade dos genes de classe I apresentarem antígenos de diversos patógenos e milhares de diferentes peptídeos. Desta forma, muitos estudos apoiam a hipótese de que essa diversidade alélica seja mantida por meio de seleção balanceadora mediada por microorganismos, levando a um aumento da heteroziguidade. A seleção balanceadora é aquela que aumenta a variação genética em relação ao esperado sob neutralidade. 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, configurando uma “vantagem do heterozigoto” (Doherty e Zinkernagel, 1975; Hughes e Yeager, 1998; Penn *et al.*, 2002; Bernatchez e Landry, 2003).

Por serem os genes mais variáveis do genoma humano, com 10.041 alelos distintos já descritos (tabela 1), 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 de adaptabilidade e resistência a patógenos, por outro lado, a necessidade ocasional de realização de enxertos alogênicos (transplantes) em humanos torna essa variabilidade problemática, pois a compatibilização desses polimorfismos é necessária para uma boa aceitação do enxerto. Estudos clínicos demonstraram que a diferença de um único aminoácido em uma molécula HLA de classe I clássica pode causar rejeição aguda de transplante de medula óssea. Muitos polimorfismos desses genes foram associados com doenças autoimunes, degenerativas 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).

Tabela 1. Número de Alelos identificados para os genes de classe I (The IMGT/HLA Database, versão 3.21.1)

CLASSE I CLÁSSICOS		CLASSE I NÃO CLÁSSICOS	
Gene	Alelos	Gene	Alelos
<i>HLA-A</i>	3192	<i>HLA-E</i>	17
<i>HLA-B</i>	3977	<i>HLA-F</i>	22
<i>HLA-C</i>	2740	<i>HLA-G</i>	50

Os genes *HLA* não clássicos (*HLA-E*, *HLA-F* e *HLA-G*) apresentam baixa variabilidade quando comparados aos genes clássicos de classe I (*HLA-A*, *HLA-B* e *HLA-C*). O *HLA-G* é o gene não clássico melhor caracterizado funcionalmente e estruturalmente. A molécula *HLA-G* foi identificada na interface materno-fetal, especificamente nas células do trofoblasto da placenta, local onde deve ocorrer uma tolerância imunológica adequada permitindo o desenvolvimento de um feto semialogênico. Fora do contexto da gravidez, a transcrição desse gene acontece a nível basal em praticamente todas as células e tecidos de indivíduos saudáveis (Carosella *et al.*, 2008). No entanto, a expressão aumentada da molécula *HLA-G* foi também identificada na matriz da unha (Carosella *et al.*, 2003), córnea (Le Discorde *et al.*, 2003), timo (Lefebvre *et al.*, 2000), trofoblasto (Hunt *et al.*, 2006), precursores eritróides e endoteliais (Menier *et al.*, 2004). Em situações patológicas, a expressão aumentada de *HLA-G* foi identificada durante episódios de rejeição enxerto-hospedeiro em pacientes transplantados (Creput *et al.*, 2003; Crispim, Duarte, *et al.*, 2008; Twito *et al.*, 2011; Mociornita *et al.*, 2013;

Seliger, 2013; Xiao *et al.*, 2013), transformações malignas (Dunker *et al.*, 2008; Cao *et al.*, 2011; Dong *et al.*, 2012), infecções virais (Cordero *et al.*, 2009; Simoes *et al.*, 2009; Haddad *et al.*, 2011; Kim *et al.*, 2013; Da Silva *et al.*, 2014; Segat *et al.*, 2014), doenças autoimunes (Brenol *et al.*, 2012) e processos inflamatórios (Carosella *et al.*, 2001; Khosrotehrani *et al.*, 2001; Graebin *et al.*, 2012).

O processo de imunomodulação relacionado ao *HLA-G*, principalmente durante a gestação, deve-se principalmente a interação dessas moléculas com receptores inibitórios presentes nas células T, NK e células apresentadoras de antígenos, em especial os receptores ILT2/CD85j/LILRB1 (ILT2), expresso em todos os monócitos, algumas linhagens de células B e T (Kamishikiryo e Maenaka, 2009); ILT4/CD85d/LILRB2 (ILT4), expresso somente em monócitos e células dendríticas (Shiroishi *et al.*, 2003) ; e KIR2DL4/CD158d (KIR2DL4) que possui uma expressão restrita a células NK (Rajagopalan e Long, 1999). As interações com esses receptores evita o mecanismo de lise celular que seria ativado por essas células, na ocasião da ativação das células efectoras pela presença de antígenos não próprios. Assim, as moléculas de *HLA-G* passam a ter grande importância em contextos onde há a necessidade de uma modulação fina do sistema imune, como por exemplo, durante a gestação, em doenças autoimunes, transplantes e no desenvolvimento de tumores (Donadi *et al.*, 2011). Ainda, a molécula *HLA-G* pode se associar a um pequeno grupo de peptídeos, porém sua principal função não é de apresentação antigênica.

O gene *HLA-G* está ainda relacionado a doenças autoimunes, um campo relativamente recente das pesquisas envolvendo expressão dessa molécula (Brenol *et al.*, 2012). Os primeiros estudos avaliaram a expressão do gene *HLA-G* nas fibras musculares em diversas miopatias inflamatórias (Wiendl *et al.*, 2000), dermatite atópica (Khosrotehrani *et al.*, 2001) e psoríase (Aractingi *et al.*, 2001). Os achados iniciais sugeriram que o gene *HLA-G* poderia desviar a resposta imune para um perfil Th2, que possui ação anti-inflamatória e sua baixa expressão estaria associada ao surgimento de células autorreativas, facilitando a instalação de uma doença autoimune (Brenol *et al.*, 2012). Estudos posteriores avaliando a variabilidade genética nas regiões regulatórias dessa molécula e sua expressão, estabeleceram relações entre o gene *HLA-G* e a doença de Kawasaki (Kim *et al.*, 2008), doença de Behcet (Park *et al.*, 2007), artrite reumatóide, lúpus eritematoso sistêmicos (Verbruggen *et al.*, 2006; Veit *et al.*, 2009; Brenol *et al.*, 2012) e sarcoidose (Hviid *et al.*, 2006).

Outro gene não clássico bem caracterizado é o gene *HLA-E*. Este gene apresenta uma expressão que ocorre na maioria dos tecidos em níveis mais baixos do que as moléculas

clássicas. *HLA-E* é o menos polimórfico de todos os genes *HLA* de classe I (Veiga-Castelli *et al.*, 2012). Apesar da baixa variabilidade descrita, estudos mostraram que o produto dos dois alelos mais frequentes, que codificam moléculas que diferem em apenas um aminoácido, apresentam atividades distintas. Um desses alelos estaria relacionado com níveis maiores de expressão de *HLA-E*, maior afinidade por peptídeos e uma maior estabilidade do complexo *HLA-E*/peptídeo (O'callaghan *et al.*, 1998; Ulbrecht *et al.*, 1999; Strong *et al.*, 2003; Sullivan *et al.*, 2008; Pietra *et al.*, 2010; Di Cristofaro *et al.*, 2011; Veiga-Castelli *et al.*, 2012). A molécula *HLA-E* interage principalmente com os receptores CD94/NKG2A, presente nas células NK (Braud *et al.*, 1998), inibindo a função destas.

Apesar da menor influência dos genes não clássicos *HLA-G* e *HLA-E* na apresentação antigênica aos linfócitos T, sua função na resposta imune tem sido intensamente estudada (Geraghty *et al.*, 1992). Os genes *HLA-E* e *HLA-G* são muito conservados quando comparados aos genes clássicos (Tabela 1). A baixa variabilidade apresentada por esses dois genes está provavelmente associada ao papel crucial que desempenham na modulação da resposta imunológica (Castelli *et al.*, 2014). A presença de polimorfismos específicos ou expressão desregulada desses genes tem sido identificada em uma série de tumores, incluindo câncer de cólon (Cao *et al.*, 2011), câncer de mama (Dong *et al.*, 2012), melanoma (Cao *et al.*, 2011), câncer renal (Dunker *et al.*, 2008), câncer de pulmão (Cao *et al.*, 2011), glioblastoma (Kren *et al.*, 2011) e carcinoma de bexiga (Castelli *et al.*, 2008). Aparentemente, a expressão destes genes em sítios tumorais facilita o escape dos tumores do sistema imune, devido à sua ação inibitória de células NK, linfócitos T e células dendríticas (Castelli *et al.*, 2008). Da mesma forma, a expressão destes genes em tecidos transplantados está associada com menor rejeição (Creput *et al.*, 2003; Crispim, Duarte, *et al.*, 2008; Hosseini *et al.*, 2013).

Dentre os genes não clássicos de classe I, o *HLA-F* é o menos estudado. A molécula possui um padrão de expressão diferenciado dos demais genes de classe I. O *HLA-F* parece estar associado ao domínio $\beta 2$ microglobulina, TAP e a calreticulina dentro da célula, contudo ainda não se sabe se essa molécula é capaz de se associar a peptídeos assim como as demais moléculas de classe I e promover a apresentação antigênica na superfície celular (Allan *et al.*, 2002). Sabe-se que o *HLA-F* possui um padrão muito restrito de expressão nos tecidos e parece não ser expresso na superfície das células, com exceção da superfície de células infectadas pelo vírus Epstein-Barr (EBV, do inglês, *Epstein Barr Virus*) (Lee *et al.*, 2010), células B, uma linhagem de células T (HUT78) e tecidos linfóides como timo, pâncreas, tonsilas e fígado fetal (Allan *et al.*, 2002).

O gene *HLA-F* apresenta muitas características compartilhadas com os genes *HLA-G* e *HLA-E*. Dessa forma, embora a função exata da molécula *HLA-F* não seja bem definida, acredita-se que ela possua atividade tolerogênica e imunomodulatória. As evidências que suportam esta associação são: (a) o *HLA-F* é pouco polimórfico, assim como os demais genes não clássicos, (b) é um gene muito conservado entre primatas (incluindo homólogos entre humanos, bonobo, gorila, orangotango e chimpanzé), sugerindo uma função crítica na resposta imunitária (Goodridge *et al.*, 2010) e (c) análises de predição das sequências de aminoácidos e a comparação com proteínas *MHC-F* de outros primatas, demonstraram que o *MHC-F* é uma molécula com estrutura típica de classe I. O fato de ser um gene conservado ao longo da evolução dos primatas sugere que este gene possui importante papel na fisiologia celular. Pode ser que proteínas do *MHC-F* (ou *HLA-F* em humanos) possam estar envolvidas juntamente com o *MHC-G* e *MHC-E* (*HLA-G* e *HLA-E* em humanos), na regulação da atividade das células T e NK (Moscoso *et al.*, 2007).

Como visto anteriormente, o gene *HLA-F* foi descrito até o momento apresentando uma variabilidade tão reduzida quanto a dos genes *HLA-G* e *HLA-E*, apesar de sua localização na região de maior variabilidade do genoma humano. Da mesma forma, *HLA-F* parece apresentar uma expressão restrita a células B, células T quiescentes, pele, pâncreas, fígado fetal, porém sua função ainda não está totalmente compreendida (Pyo *et al.*, 2006). A molécula *HLA-F* interage com os receptores ILT2 e ILT4, presente nos monócitos (Lepin *et al.*, 2000) e não há relatos até o momento sobre a interação dessa moléculas com células NK, como evidenciado para as outras moléculas não clássicas.

A molécula não clássica *HLA-F* ainda carece de estudos que esclareçam sua função e estrutura. A expressão desta molécula na superfície celular é regulada positivamente após a ativação de monócitos e na maioria dos subconjuntos de linfócitos, incluindo as células NK, células B e todos os subconjuntos de células T exceto as reguladoras (Lee *et al.*, 2010). Diferentemente dos demais genes de classe I, alguns estudos apontam que *HLA-F* parece ser expresso como uma proteína independente da ligação de peptídeos e/ou β 2M, gerando uma molécula em conformação aberta e que não apresenta a estrutura descrita na Figura 1 (Wainwright *et al.*, 2000; Goodridge *et al.*, 2010).

Em geral, moléculas de classe I desprovidas de β 2M ou peptídeo, i.e., em conformação aberta (Schnabl *et al.*, 1990), são instáveis e não permanecem na superfície celular. No entanto, parece que a maioria das proteínas de classe I, quando em conformação aberta, ligam-se à *HLA-F*. Desta forma, a interação física entre essas moléculas parece desempenhar algum papel na

função das moléculas de classe I em conformação aberta. Provavelmente, as moléculas de classe I em conformação aberta, estabilizadas pela HLA-F, conservam a especificidade de ligação do peptídeo conferida pela fenda de ligação ao peptídeo. Adicionalmente, podem ligar-se a epítomos sem limitação de tamanho (Goodridge *et al.*, 2013), que, considerando as moléculas MHC de classe I na sua estrutura clássica definida na Figura 1, seria de oito a onze aminoácidos em uma conformação flexível e prolongada (Abbas *et al.*, 2007), além de permitir a formação de homodímeros (Arosa *et al.*, 2007).

O fato de *HLA-F* interagir fisicamente com moléculas de classe I de conformação aberta, porém não com moléculas de classe I tradicionais, implica em uma possível relação funcional entre moléculas de classe I e *HLA-F*. Uma provável função do *HLA-F* poderia ser na estabilização e transporte das moléculas de classe I para a superfície celular (Goodridge *et al.*, 2013). A interação física entre a HLA-F e moléculas clássicas de classe I em conformação aberta, juntamente com a expressão dessas moléculas na superfície celular de linfócitos ativos, sugere uma interdependência dessas moléculas no transporte de peptídeos até superfície celular. A HLA-F provavelmente estabiliza as moléculas clássicas em conformação aberta, formando um heterodímero que então transita até a superfície celular (Goodridge *et al.*, 2013). Outro possível papel do *HLA-F* seria auxiliar no processo de internalização de antígenos e moléculas de classe I (reciclagem) (Goodridge *et al.*, 2013).

A molécula *HLA-F* é expressa em uma gama de tumores e aparentemente pode influenciar o prognóstico do paciente, porém os achados são contraditórios ou dependentes da lesão avaliada. Estudos indicam que sua expressão no adenocarcinoma gástrico está relacionado com um mau prognóstico do paciente, visto que sua expressão promove um comportamento agressivo do tumor, provavelmente relacionado com imunossupressão das células antitumorais (Ishigami *et al.*, 2013). Alterações na expressão de *HLA-F* em pacientes com carcinoma esofágico de células escamosas e câncer de pulmão também relaciona-se com o prognóstico desse paciente, uma vez que a expressão exacerbada desta molécula parece fazer com que esses pacientes possuam uma sobrevida menor quando comparado aos grupos que possuem baixa expressão desse gene (Lin *et al.*, 2011; Zhang, X. *et al.*, 2013). Ainda, a expressão da molécula HLA-F foi associada com um mau prognóstico de pacientes com carcinoma hepatocelular e correlacionado com invasão celular e metástase (Xu *et al.*, 2015). No entanto, Zhang e colaboradores não encontraram significado clínico para expressão do gene *HLA-F* em câncer gástrico, sua expressão parece não estar relacionada ao prognóstico do paciente (Zhang, J. G.

et al., 2013). Ainda, baixos níveis dessas duas moléculas estão associados a um pior prognóstico em pacientes com neuroblastoma (Morandi *et al.*, 2013).

O gene *HLA-F* pode estar relacionado com processos de imunomodulação durante gestação e após transplantes. Não se sabe ao certo se sua expressão acontece na superfície celular ou no trofoblasto extraviloso durante a gravidez. Os resultados são contraditórios, uma vez que Ishitani *et al.* (2003) e Nagamatsu *et al.* (2006) detectaram a presença da molécula *HLA-F* somente no trofoblasto extraviloso, (Ishitani *et al.*, 2003; Nagamatsu *et al.*, 2006) enquanto Shobu e colaboradores (2006) detectaram a molécula em superfície celular (Shobu *et al.*, 2006). Entretanto Apps e colaboradores não conseguiram detectar a presença da molécula *HLA-F* em trofoblasto extraviloso, bem como na superfície celular, utilizando imunohistoquímica e citometria de fluxo (Apps *et al.*, 2008). Quanto aos transplantes, diversos trabalhos apontam para a influência dos genes não clássicos *HLA-G* e *HLA-E* na aceitação do enxerto (Pabon *et al.*, 2014; Biedron *et al.*, 2015; Hosseini *et al.*, 2015; Krongvorakul *et al.*, 2015; Mossallam *et al.*, 2015), contudo nenhuma associação significativa foi observada entre o gene *HLA-F* e pacientes transplantados. Não se pode excluir a participação desta molécula nos desfechos dos transplantes e gestação, uma vez que essa molécula pode atuar em conformação aberta como visto anteriormente, auxiliando outras moléculas *HLA(s)* em sua função (clássicos e não clássicos).

O gene *HLA-F*, assim como membros do grupo de genes não clássicos, parece estar associado com mecanismos de tolerância imunológica, escape de células tumorais, tolerância a aloenxertos, susceptibilidade a doenças autoimunes e infecções, por mecanismos inibitórios de células T e células NK. A variabilidade inerente ao gene *HLA-F* foi minimamente explorada, uma vez que apenas dois trabalhos avaliaram polimorfismos do gene *HLA-F*, um conduzido na China (Pan *et al.*, 2013) e outro no Brasil (Manvailer *et al.*, 2014). No entanto, apesar de ambos terem utilizado metodologias sensíveis, exploraram apenas um pequeno segmento do gene ou buscaram por polimorfismos já reconhecidos pelo banco de dados IMGT/HLA. Ainda, o mecanismo subjacente à baixa variabilidade detectada neste gene ainda não foi esclarecido, tornando este gene virtualmente invariável quando comparado aos seus vizinhos *HLA-A*, *HLA-B* e *HLA-C*, que apresentam estrutura molecular semelhante, ou até mesmo quando comparados aos demais genes clássicos. Considerando a região onde este gene insere-se (a mais variável do genoma) e a baixa variabilidade detectada até o momento, é provável que esta baixa variabilidade seja consequência do papel chave do *HLA-F* nas respostas imunitárias. No entanto, não sabemos se esse gene é pouco variável porque até o momento não foi realizada

nenhuma análise que possa detectar toda a variabilidade do gene ou se o *HLA-F* é de fato pouco polimórfico nas diversas populações mundiais. Neste estudo, avaliamos a variabilidade do gene *HLA-F* em 196 amostras brasileiras do sudeste do Brasil, em um segmento que compreende todo o gene (promotora e 3'NT).

Até o momento não há estudos que caracterizaram a variabilidade do gene *HLA-F* na profundidade aqui proposta, em especial em uma população como a brasileira, que é um reconhecido repositório de variação genética, com contribuições Europeia, Africana e Ameríndia. O estado de São Paulo, alvo do presente estudo é o mais populoso do Brasil e contém uma considerável miscigenação entre os grupos citados acima, contudo para essa região do país a maior contribuição é Européia (79%), seguido da Africana (14%) e Ameríndia (7%) (Ferreira *et al.*, 2006).

Este estudo permitirá a criação de um banco de dados da variabilidade inerente a este gene em uma população normal do estado de São Paulo, que poderá ser usado para estudos futuros de associação com doenças, além de detectar potenciais polimorfismos funcionais na região codificadora deste gene, que poderiam influenciar a expressão de *HLA-F*. Por fim, a variabilidade da região codificadora e das regiões regulatórias do *HLA-F* foi avaliada. Os haplótipos da região codificadora encontrados neste estudo foram apresentados no capítulo II utilizando a nomenclatura padrão para alelos HLA (anexo 1). Os haplótipos das regiões regulatórias, por não haver um padrão vigente de nomenclatura, foram apresentados utilizando-se uma nomenclatura criada para este estudo.

Referências

- ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. **Cellular and molecular immunology**. 6th. Philadelphia: Saunders Elsevier, 2007. viii, 566 p.
- ALLAN, D. S. et al. Tetrameric complexes of HLA-E, HLA-F, and HLA-G. **J Immunol Methods**, v. 268, n. 1, p. 43-50, Oct 1 2002.
- APPS, R. et al. Natural-killer cell ligands at the maternal-fetal interface: UL-16 binding proteins, MHC class-I chain related molecules, HLA-F and CD48. **Hum Reprod**, v. 23, n. 11, p. 2535-48, Nov 2008.
- ARACTINGI, S. et al. HLA-G and NK receptor are expressed in psoriatic skin: a possible pathway for regulating infiltrating T cells? **Am J Pathol**, v. 159, n. 1, p. 71-7, Jul 2001.
- AROSA, F. A.; SANTOS, S. G.; POWIS, S. J. Open conformers: the hidden face of MHC-I molecules. **Trends Immunol**, v. 28, n. 3, p. 115-23, Mar 2007.
- 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.
- BIEDRON, M. et al. The role of soluble HLA-G and HLA-G receptors in patients with hematological malignancies after allogeneic stem cell transplantation. **Med Oncol**, v. 32, n. 8, p. 664, Aug 2015.
- BRAUD, V. M. et al. HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. **Nature**, v. 391, n. 6669, p. 795-9, Feb 19 1998.
- BRENOL, C. V. et al. The role of the HLA-G gene and molecule on the clinical expression of rheumatologic diseases. **Rev Bras Reumatol**, v. 52, n. 1, p. 82-91, Jan-Feb 2012.
- CAILLAT-ZUCMAN, S. Molecular mechanisms of HLA association with autoimmune diseases. **Tissue Antigens**, v. 73, n. 1, p. 1-8, Jan 2009.
- CAO, M. et al. Plasma soluble HLA-G is a potential biomarker for diagnosis of colorectal, gastric, esophageal and lung cancer. **Tissue Antigens**, v. 78, n. 2, p. 120-8, Aug 2011.
- CAROSELLA, E. D. et al. HLA-G: a shield against inflammatory aggression. **Trends Immunol**, v. 22, n. 10, p. 553-5, Oct 2001.
- CAROSELLA, E. D. et al. HLA-G molecules: from maternal-fetal tolerance to tissue acceptance. **Adv Immunol**, v. 81, p. 199-252, 2003.
- 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. HLA-G polymorphism and transitional cell carcinoma of the bladder in a Brazilian population. **Tissue Antigens**, v. 72, n. 2, p. 149-57, Aug 2008.
- CASTELLI, E. C. et al. Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity. **Front Immunol**, v. 5, p. 476, 2014.
- CORDERO, E. A. et al. HLA-G polymorphism influences the susceptibility to HCV infection in sickle cell disease patients. **Tissue Antigens**, v. 74, n. 4, p. 308-13, Oct 2009.

- CREPUT, C. et al. Human leukocyte antigen-G (HLA-G) expression in biliary epithelial cells is associated with allograft acceptance in liver-kidney transplantation. **J Hepatol**, v. 39, n. 4, p. 587-94, Oct 2003.
- 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. K. et al. Influence of HLA-G polymorphisms in human immunodeficiency virus infection and hepatitis C virus co-infection in Brazilian and Italian individuals. **Infect Genet Evol**, v. 21, p. 418-23, Jan 2014.
- DENIZOT, F. et al. Clonal expansion of T cells: a cytotoxic T-cell response in vivo that involves precursor cell proliferation. **Proc Natl Acad Sci U S A**, v. 83, n. 16, p. 6089-92, Aug 1986.
- DI CRISTOFARO, J. et al. Linkage disequilibrium between HLA-G*0104 and HLA-E*0103 alleles in Tswa Pygmies. **Tissue Antigens**, v. 77, n. 3, p. 193-200, Mar 2011.
- 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, D. D. et al. Importance of HLA-G expression and tumor infiltrating lymphocytes in molecular subtypes of breast cancer. **Human Immunology**, v. 73, n. 10, p. 998-1004, Oct 2012.
- DUNKER, K. et al. Expression and regulation of non-classical HLA-G in renal cell carcinoma. **Tissue Antigens**, v. 72, n. 2, p. 137-48, Aug 2008.
- FERREIRA, L. B. et al. Genomic ancestry of a sample population from the state of Sao Paulo, Brazil. **Am J Hum Biol**, v. 18, n. 5, p. 702-5, Sep-Oct 2006.
- 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.
- 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.
- GOODRIDGE, J. P. et al. HLA-F complex without peptide binds to MHC class I protein in the open conformer form. **J Immunol**, v. 184, n. 11, p. 6199-208, Jun 1 2010.
- GOODRIDGE, J. P. et al. HLA-F and MHC-I open conformers cooperate in a MHC-I antigen cross-presentation pathway. **J Immunol**, v. 191, n. 4, p. 1567-77, Aug 15 2013.
- GRAEBIN, P. et al. Polymorphic variants in exon 8 at the 3' UTR of the HLA-G gene are associated with septic shock in critically ill patients. **Crit Care**, v. 16, n. 5, p. R211, 2012.
- HADDAD, R. et al. HLA-G 14-bp insertion/deletion polymorphism is a risk factor for HTLV-1 infection. **AIDS Res Hum Retroviruses**, v. 27, n. 3, p. 283-8, Mar 2011.

- HOSSEINI, E.; SCHWARER, A. P.; GHASEMZADEH, M. Do human leukocyte antigen E polymorphisms influence graft-versus-leukemia after allogeneic hematopoietic stem cell transplantation? **Exp Hematol**, v. 43, n. 3, p. 149-57, Mar 2015.
- HOSSEINI, E. et al. The impact of HLA-E polymorphisms on relapse following allogeneic hematopoietic stem cell transplantation. **Leuk Res**, v. 37, n. 5, p. 516-9, May 2013.
- 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.
- HUNT, J. S. et al. The role of HLA-G in human pregnancy. **Reprod Biol Endocrinol**, v. 4 Suppl 1, p. S10, 2006.
- HVIID, T. V. et al. HLA-G polymorphisms and HLA-G expression in sarcoidosis. **Sarcoidosis Vasc Diffuse Lung Dis**, v. 23, n. 1, p. 30-7, Mar 2006.
- ISHIGAMI, S. et al. Clinical-pathological implication of human leukocyte antigen-F-positive gastric adenocarcinoma. **J Surg Res**, Apr 26 2013.
- ISHITANI, A. et al. Protein expression and peptide binding suggest unique and interacting functional roles for HLA-E, F, and G in maternal-placental immune recognition. **J Immunol**, v. 171, n. 3, p. 1376-84, Aug 1 2003.
- 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.
- JOBIM, M.; JOBIM, L. F. Natural killer cells and immune surveillance. **J Pediatr (Rio J)**, v. 84, n. 4 Suppl, p. S58-67, Aug 2008.
- KAMISHIKIRYO, J.; MAENAKA, K. HLA-G molecule. **Curr Pharm Des**, v. 15, n. 28, p. 3318-24, 2009.
- KAPPLER, J. W.; ROEHM, N.; MARRACK, P. T cell tolerance by clonal elimination in the thymus. **Cell**, v. 49, n. 2, p. 273-80, Apr 24 1987.
- KAZANSKY, D. B. MHC restriction and allogeneic immune responses. **J Immunotoxicol**, v. 5, n. 4, p. 369-84, Oct 2008.
- KHOSROTEHRANI, K. et al. HLA-G expression in atopic dermatitis. **J Invest Dermatol**, v. 117, n. 3, p. 750-2, Sep 2001.
- KIM, J. J. et al. Genetic variants in the HLA-G region are associated with Kawasaki disease. **Hum Immunol**, v. 69, n. 12, p. 867-71, Dec 2008.
- KIM, S. K. et al. Association between HLA-G 14-bp insertion/deletion polymorphism and hepatocellular carcinoma in Korean patients with chronic hepatitis B viral infection. **Hepatogastroenterology**, v. 60, n. 124, p. 796-8, Jun 2013.
- 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.
- _____. The HLA system. Second of two parts. **N Engl J Med**, v. 343, n. 11, p. 782-6, Sep 14 2000b.
- KREN, L. et al. Expression of immune-modulatory molecules HLA-G and HLA-E by tumor cells in glioblastomas: an unexpected prognostic significance? **Neuropathology**, v. 31, n. 2, p. 129-34, Apr 2011.

- KRONGVORAKUL, J. et al. Association of soluble human leukocyte antigen-G with acute tubular necrosis in kidney transplant recipients. **Asian Pac J Allergy Immunol**, v. 33, n. 2, p. 117-22, Jun 2015.
- KULSKI, J. K. et al. The association between HLA-A alleles and an Alu dimorphism near HLA-G. **J Mol Evol**, v. 53, n. 2, p. 114-23, Aug 2001.
- LE DISCORDE, M. et al. Expression of HLA-G in human cornea, an immune-privileged tissue. **Hum Immunol**, v. 64, n. 11, p. 1039-44, Nov 2003.
- 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.
- LEE, N.; ISHITANI, A.; GERAGHTY, D. E. HLA-F is a surface marker on activated lymphocytes. **Eur J Immunol**, v. 40, n. 8, p. 2308-18, Aug 2010.
- LEFEBVRE, S. et al. Modulation of HLA-G expression in human thymic and amniotic epithelial cells. **Hum Immunol**, v. 61, n. 11, p. 1095-101, Nov 2000.
- 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.
- LIN, A. et al. HLA-F expression is a prognostic factor in patients with non-small-cell lung cancer. **Lung Cancer**, v. 74, n. 3, p. 504-9, Dec 2011.
- LITMAN, G. W.; RAST, J. P.; FUGMANN, S. D. The origins of vertebrate adaptive immunity. **Nat Rev Immunol**, v. 10, n. 8, p. 543-53, Aug 2010.
- MANVAILER, L. F. et al. HLA-F polymorphisms in a Euro-Brazilian population from Southern Brazil. **Tissue Antigens**, v. 84, n. 6, p. 554-9, Dec 2014.
- MENIER, C. et al. Erythroblasts secrete the nonclassical HLA-G molecule from primitive to definitive hematopoiesis. **Blood**, v. 104, n. 10, p. 3153-60, Nov 15 2004.
- 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.
- MOCIORNITA, A. G. et al. Can HLA-G polymorphisms predict the development of cardiac allograft vasculopathy? **Hum Immunol**, v. 74, n. 4, p. 464-7, Apr 2013.
- MORANDI, F. et al. Plasma levels of soluble HLA-E and HLA-F at diagnosis may predict overall survival of neuroblastoma patients. **Biomed Res Int**, v. 2013, p. 956878, 2013.
- MOSCOSO, J.; SERRANO-VELA, J. I.; ARNAIZ-VILLENA, A. MHC-F polymorphism and evolution. **Tissue Antigens**, v. 69 Suppl 1, p. 136-9, Apr 2007.
- MOSSALLAM, G. I. et al. HLA-E polymorphism and clinical outcome after allogeneic hematopoietic stem cell transplantation in Egyptian patients. **Hum Immunol**, v. 76, n. 2-3, p. 161-5, Mar 2015.
- MURPHY, K., TRAVERS, P. , WALPORT, M. . **Imunobiologia de Janeway**. 7. Porto Alegre: 2010.
- NAGAMATSU, T. et al. Human leukocyte antigen F protein is expressed in the extra-villous trophoblasts but not on the cell surface of them. **Am J Reprod Immunol**, v. 56, n. 3, p. 172-7, Sep 2006.

- O'CALLAGHAN, C. A. et al. Structural features impose tight peptide binding specificity in the nonclassical MHC molecule HLA-E. **Mol Cell**, v. 1, n. 4, p. 531-41, Mar 1998.
- PABON, M. A. et al. Impact of human leukocyte antigen molecules E, F, and G on the outcome of transplantation. **Transplant Proc**, v. 46, n. 9, p. 2957-65, Nov 2014.
- PAN, F. H.; LIU, X. X.; TIAN, W. Characterization of HLA-F polymorphism in four distinct populations in Mainland China. **Int J Immunogenet**, Apr 2 2013.
- PARDIGON, N. et al. Role of co-stimulation in CD8+ T cell activation. **Int Immunol**, v. 10, n. 5, p. 619-30, May 1998.
- PARK, K. S. et al. HLA-E*0101 and HLA-G*010101 reduce the risk of Behcet's disease. **Tissue Antigens**, v. 69, n. 2, p. 139-44, Feb 2007.
- 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.
- PIETRA, G. et al. The emerging role of HLA-E-restricted CD8+ T lymphocytes in the adaptive immune response to pathogens and tumors. **J Biomed Biotechnol**, v. 2010, p. 907092, 2010.
- PYO, C. W. et al. HLA-E, HLA-F, and HLA-G polymorphism: genomic sequence defines haplotype structure and variation spanning the nonclassical class I genes. **Immunogenetics**, v. 58, n. 4, p. 241-51, May 2006.
- RAJAGOPALAN, S.; LONG, E. O. A human histocompatibility leukocyte antigen (HLA)-G-specific receptor expressed on all natural killer cells. **J Exp Med**, v. 189, n. 7, p. 1093-100, Apr 5 1999.
- RUDDLE, N. H.; AKIRAV, E. M. Secondary lymphoid organs: responding to genetic and environmental cues in ontogeny and the immune response. **J Immunol**, v. 183, n. 4, p. 2205-12, Aug 15 2009.
- SAKAGUCHI, S. et al. Regulatory T cells and immune tolerance. **Cell**, v. 133, n. 5, p. 775-87, May 30 2008.
- SCHNABL, E. et al. Activated human T lymphocytes express MHC class I heavy chains not associated with beta 2-microglobulin. **J Exp Med**, v. 171, n. 5, p. 1431-42, May 1 1990.
- SEGAT, L. et al. HLA-G 14 bp deletion/insertion polymorphism and mother-to-child transmission of HIV. **Tissue Antigens**, v. 83, n. 3, p. 161-7, Mar 2014.
- SELIGER, B. The non-classical antigens of HLA-G and HLA-E as diagnostic and prognostic biomarkers and as therapeutic targets in transplantation and tumors. **Clin Transpl**, p. 465-72, 2013.
- 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.
- SHIROISHI, M. et al. Human inhibitory receptors Ig-like transcript 2 (ILT2) and ILT4 compete with CD8 for MHC class I binding and bind preferentially to HLA-G. **Proc Natl Acad Sci U S A**, v. 100, n. 15, p. 8856-61, Jul 22 2003.
- SHOBU, T. et al. The surface expression of HLA-F on decidual trophoblasts increases from mid to term gestation. **J Reprod Immunol**, v. 72, n. 1-2, p. 18-32, Dec 2006.

- SIMOES, R. T. et al. HLA-G polymorphisms in women with squamous intraepithelial lesions harboring human papillomavirus. **Mod Pathol**, v. 22, n. 8, p. 1075-82, Aug 2009.
- STRONG, R. K. et al. HLA-E allelic variants. Correlating differential expression, peptide affinities, crystal structures, and thermal stabilities. **J Biol Chem**, v. 278, n. 7, p. 5082-90, Feb 14 2003.
- SULLIVAN, L. C. et al. The major histocompatibility complex class Ib molecule HLA-E at the interface between innate and adaptive immunity. **Tissue Antigens**, v. 72, n. 5, p. 415-24, Nov 2008.
- THOMPSON, E. C. Focus issue: Structure and function of lymphoid tissues. **Trends Immunol**, v. 33, n. 6, p. 255, Jun 2012.
- TWITO, T. et al. The 14-bp deletion in the HLA-G gene indicates a low risk for acute cellular rejection in heart transplant recipients. **J Heart Lung Transplant**, v. 30, n. 7, p. 778-82, Jul 2011.
- ULBRECHT, M. et al. Cell surface expression of HLA-E: interaction with human beta2-microglobulin and allelic differences. **European Journal of Immunology**, v. 29, n. 2, p. 537-47, Feb 1999.
- 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.
- VEIT, T. D. et al. Association of the HLA-G 14 bp polymorphism with systemic lupus erythematosus. **Lupus**, v. 18, n. 5, p. 424-30, Apr 2009.
- VERBRUGGEN, L. A. et al. Soluble HLA-G in rheumatoid arthritis. **Hum Immunol**, v. 67, n. 8, p. 561-7, Aug 2006.
- VYAS, J. M.; VAN DER VEEN, A. G.; PLOEGH, H. L. The known unknowns of antigen processing and presentation. **Nat Rev Immunol**, v. 8, n. 8, p. 607-18, Aug 2008.
- WAINWRIGHT, S. D.; BIRO, P. A.; HOLMES, C. H. HLA-F is a predominantly empty, intracellular, TAP-associated MHC class Ib protein with a restricted expression pattern. **J Immunol**, v. 164, n. 1, p. 319-28, Jan 1 2000.
- WIENDL, H. et al. Muscle fibers in inflammatory myopathies and cultured myoblasts express the nonclassical major histocompatibility antigen HLA-G. **Ann Neurol**, v. 48, n. 4, p. 679-84, Oct 2000.
- XIAO, L. et al. HLA-G expression in the peripheral blood of live kidney transplant recipients. **Chin Med J (Engl)**, v. 126, n. 14, p. 2652-5, Jul 2013.
- XU, Y. et al. Lesion human leukocyte antigen-F expression is associated with a poor prognosis in patients with hepatocellular carcinoma. **Oncol Lett**, v. 9, n. 1, p. 300-304, Jan 2015.
- 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, J. G. et al. Lesion HLA-F expression is irrelevant to prognosis for patients with gastric cancer. **Hum Immunol**, v. 74, n. 7, p. 828-32, Jul 2013.

ZHANG, X. et al. Alteration of HLA-F and HLA I antigen expression in the tumor is associated with survival in patients with esophageal squamous cell carcinoma. **Int J Cancer**, v. 132, n. 1, p. 82-9, Jan 1 2013.

Capítulo II

HLA-F CODING AND REGULATORY SEGMENTS VARIABILITY OF A BRAZILIAN POPULATION SAMPLE BY USING MASSIVE PARALLEL SEQUENCING PROCEDURES

Thálitta H.A. Lima^{1,2}, Renato V. Butura^{1,2}, Eduardo A. Donadi³, Luciana C. Veiga-Castelli³, Celso T. Mendes-Junior⁴, Erick C. Castelli^{1,2}

1 Post-Graduate in Pathology, School of Medicine, Unesp – Univ. Estadual Paulista, Botucatu, State of São Paulo, Brazil.

2 Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPLEX), School of Medicine, Unesp – Univ. Estadual Paulista, Botucatu, State of São Paulo, Brazil.

3 School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, State of São Paulo, Brazil.

4 Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto - SP, Brazil.

- **To be submitted to publication**

ABSTRACT

HLA-F is a non-classical HLA class I gene distinguished from its classical counterparts by low allelic polymorphism and distinctive expression patterns. Its exact function remains unknown. It is believed that *HLA-F* has tolerogenic and immune modulatory properties. Currently, there is little information regarding the *HLA-F* allelic variation among human populations and the available studies have evaluated only a fraction of the *HLA-F* gene segment and/or have searched for known alleles only. Here we present a strategy to evaluate the complete *HLA-F* variability including its regulatory segments (promoter and 3'UTR) by using massive parallel sequencing procedures. *HLA-F* variability was surveyed on 196 individuals from the Brazilian Southeast. The results indicate that the *HLA-F* gene is indeed conserved at the protein level (the three coding haplotypes detected encode only four different HLA-F molecules, and one of these molecules represents 82.45% of all), but the *HLA-F* haplotype variability is much higher than our current knowledge at least in the present sample. This protein conservation is probably a consequence of the key role of *HLA-F* in the immune system physiology.

Keywords: *HLA-F*, Next Generation Sequencing, variability.

Introduction

The *HLA-F* locus is a non-classical gene of the Human Leukocyte Antigen Complex (HLA) located within the Major Histocompatibility Complex (MHC). Most of the HLA genes are related to antigen presentation, mainly those known as classical HLA genes. HLA genes are considered the most polymorphic ones among vertebrates (1), but this variability is usually explored and reported within the classical HLA genes.

Although surrounded by the most variable human genes, *HLA-F* is described presenting low allelic and protein polymorphism and distinctive expression patterns. While the pivotal role of classical class I HLA genes (such as *HLA-A*, *HLA-B* and *HLA-C*) in antigen presentation is well understood and documented, the importance and function of non-classical genes (such as *HLA-G*, *HLA-E* and *HLA-F*) are poorly understood. Most of the studies so far evaluated *HLA-G* and several pieces of evidence points to its important role in immune tolerance (2-12).

The exact function of *HLA-F* remains unknown, but evidences indicate that its encoded molecule has tolerogenic and immune modulatory features following the same pattern observed for other non-classical genes. This is based mostly on the following *HLA-F* features. Firstly, *HLA-F* is reported presenting low variability, with only 22 different coding sequences already described by the International ImmunoGenetics Database (IMGT/HLA, version 3.21.1). This same pattern of variation was observed for other non-classical genes such as *HLA-G* and *HLA-E* (10, 12-17) which have strong evidences suggesting their roles on immune modulation (18-25). Secondly, *HLA-F* (or MHC-F in primates) presents sequence conservation among primates, suggesting a critical role in the immune response (26). The conservation of this molecule throughout primates' evolution suggests an important role in cellular physiology (2). Thirdly, analyses of the predicted amino acid sequences and their comparison among MHC-F molecules of different primates indicate that it presents a typical HLA class I structure, i.e., a heavy chain (α domain) with five domains: two forming the peptide binding groove ($\alpha 1$ and $\alpha 2$), an immunoglobulin-like domain ($\alpha 3$), a transmembrane domain and a cytoplasmic domain (26, 27). This heavy domain may be associated with a beta-2-microglobulin (B2M) molecule. Fourthly, the HLA-F molecule was reported interacting with inhibitory receptors ILT2 and ILT4, usually detected at monocytes, thus, it might influence immune effector cells activation (28).

HLA class I molecules lacking $\beta 2M$ or their associated peptides usually do not present the typical HLA class I molecule structure. Instead, they might be expressed as open conformers (29). These atypical HLA molecules are unstable and do not remain on the cell surface. However, recent evidences indicate that the HLA-F molecule might bind to these HLA open conformers, and physical interaction appears to play a key role on the stabilization and transportation of these class I molecules in open conformation to the cell surface (26, 27, 30). *HLA-F* expression was detected so far at B cells, T cell line HUT37 and lymphoid tissues such as thymus, spleen and tonsil (27, 28, 31). Although this molecule apparently presents a typical HLA class I structure, it is not clear whether they are able to associate with peptides and present them on the cell surface (2), once this molecule was mainly found as an intracellular protein (27, 28, 31-33).

Furthermore, the *HLA-F* gene may be related to immune response modulation during pregnancy and transplantation. Until now, the HLA-F molecule seems to be mainly intracellular, although some studies found conflicting results, considering pregnancy. For instance, HLA-F was detected on the trophoblast cell surface (34), only on the cytoplasm on decidual trophoblasts (35) and no expression at all on trophoblasts (36). Regarding transplantation, several studies reported the influence of non-classical genes such as *HLA-G* and *HLA-E* on graft acceptance considering both expression and specific polymorphisms (19, 25, 37-40), but no study was conducted regarding *HLA-F*. However, we cannot rule out the *HLA-F* involvement in transplant and pregnancy outcomes, since this molecule may act stabilizing HLA open conformers as previously described, helping other HLA molecules in their function (classical and non-classical ones).

It has been considered that *HLA-F* presents few polymorphisms. As previously introduced, only 22 different sequences are presented on the IMGT/HLA database (version 3.21.1). Nevertheless, *HLA-F* variability was poorly explored in worldwide populations. In fact, only two studies were performed evaluating *HLA-F* variability, the first one conducted in China (41), searching for only known variable sites described at that moment, and the second in Brazil (42) evaluating only a fraction of the *HLA-F* gene segment. Thus, it is not clear whether HLA-F truly presents a low rate of polymorphisms, or it was poorly explored. Considering the region where *HLA-F* is located, i.e., the most variable in the human genome, and the low allelic polymorphism detected so far, this low variability might be a consequence of the *HLA-F* key role in the regulation of immune responses.

Here we present a strategy to evaluate the variability of the complete *HLA-F* gene segment, including regulatory sequences, by using massive parallel sequencing (or Next Generation Sequencing – NGS) procedures. This strategy was applied to evaluate the *HLA-F* variability on Brazilian samples from the State of São Paulo (Southeaster Brazil), which is a very admixed population (43) and considered a large repository of genetic variation.

Methods

DNA samples and amplification

HLA-F variability was evaluated in 196 non-related individuals from the state of São Paulo, Southeastern Brazil. All participants signed an informed consent before blood withdraw, and this study protocol was reviewed and approved by the local Human Research Ethics Committee. The estimates of admixture components based on the gene identity method for a similar sample of this same geographic region resulted in 79% European, 14% African, and 7% Amerindian contributions (44).

The *HLA-F* segment was amplified encompassing its 5' upstream segment (that includes 5' upstream segment, the proximal promoter and the 5' untranslated region), its complete coding region (including introns), the 3' untranslated region and several nucleotides downstream *HLA-F*. In general, an amplicon of approximately 6,245 nucleotides was produced, from nucleotides 29,689,321 to 29,695,535 considering the sequence available for chromosome 6 (human genome assembly hg19). However, only the variability between nucleotides 29,690,685 (position -556 considering nucleotide +1 as the Adenine of the first translated ATG) and 29,695,490 (position +4252) is reported. Amplification was carried out using primers HFPR.F1 (5'-GGAGAGAACAACACTCAGGTGGC-3') and HFUT.R1 (5'-CCACTAAACACCCAGCCCAT-3') in a final volume of 50 uL, containing 0.2 mM of dNTPs (Invitrogen - Carlsbad, CA USA), 15 pmol of each primer, 1.50 units of DNA polymerase (Long PCR Enzyme Mix, Thermo Fisher Scientific Inc. Waltham, MA USA) and 0.8X the PCR buffer supplied with the DNA polymerase. Cycling conditions were performed as follows: 94°C for 5 minutes, 10 cycles of 94°C for 30 seconds, 63°C for 45 seconds and 68°C for 7 minutes, 22 cycles of 94°C for 30 seconds, 61°C for 45 seconds and 68°C for 7 minutes. Amplification was evaluated on 1% agarose gel stained with GelRed® (Biotium™, Hayward, USA),

quantified using Qubit dsDNA High-Sensitivity assays and normalized to the recommended concentration for sequencing library preparation.

Library preparation and sequencing

Amplicons were sequenced by massive parallel sequencing (or Next Generation Sequencing - NGS) procedures using the MiSeq Platform (Illumina, Inc. San Diego, CA USA). Sequencing libraries were prepared using Nextera XT Sample Preparation Kit and multiplexed with the Nextera XT Index Kit (both from Illumina, Inc.). Libraries were quantified by qPCR using Kapa (Kapa Biosystems, Wilmington, USA) and normalized to the recommended concentration. Fragmentation size was estimated using Bioanalyzer High Sensitivity DNA chips (Agilent Technologies, Santa Clara – CA, USA). Sequencing was performed using MiSeq Reagent Kit (V2, 500 cycles). The paired-end sequencing data was evaluated using freely available and locally developed software, as described further.

Raw data processing (mapping)

Prior to mapping, all short DNA segments produced by NGS, referenced to as ‘reads’, were trimmed on both ends for primer and adapter sequences. A major goal when performing HLA sequencing by NGS procedures is to get a reliable mapping of the produced reads. As previously introduced in a former manuscript by our group addressing *HLA-E* variability (15), and only briefly described here, this is particularly challenging for HLA genes. To achieve a reliable read mapping, we used *hla-mapper* (15), available at www.castelli-lab.net.

By using *hla-mapper*, reads were filtered to keep only pairs in which both sequences were larger than 80 nucleotides and presented at least 80% of bases with a minimum quality of 25 (Phred $\geq$ 25). After the removal of small or low quality sequences, PCR duplicates, i.e., pairs with the same forward and reverse sequences, were removed. Then, *hla-mapper* addressed each sequence to its specific HLA gene (in this case, *HLA-F*). The strategy used by this software is described in details elsewhere (15). The final step of data processing by *hla-mapper* generates BAM files with all sequences mapped to the reference genome version hg19.

Genotype calling and processing

The Genome Analysis Toolkit (GATK, version 3.3) UnifiedGenotyper and HaplotypeCaller routines (45, 46) (47) were used to infer genotypes and VCF (variant call format) files were generated considering all samples together, using the following parameters: *dcov* set to 1000 (only for UnifiedGenotyper), *stand_call_conf* set to 200, *stand_emit_conf* set to 50, and the chromosome 6 sequence (hg19) as reference. A mixed VCF file was then manually generated, in which variable sites presenting insertions or deletions (indel) came from HaplotypeCaller and others came from UnifiedGenotyper. This procedure was applied because HaplotypeCaller is better to infer genotypes in regions presenting indels, but UnifiedGenotyper is better to detect low frequency variants.

Some level of miss-mapped reads is expected on HLA genes and might bias genotype inference. This is particularly true when more than one HLA gene is sequenced at the same time (as it was on this study), mainly because of the polymorphic nature and the high level of sequence similarity among HLA genes. Although this phenomenon was minimized by *hla-mapper*, this must be considered when analyzing any HLA gene by NGS procedures. To circumvent these issues, the VCF file generated by GATK was handled by *vcfx* (available at www.castelli-lab.net), which applied the following rules in order to get a reliable genotype calling and assure that only high quality heterozygous genotypes are passed forward to the phasing and imputation procedures to be applied afterwards.

- For homozygous genotypes inferred in segments with less than 8 available reads, a missing allele was introduced (*vcfx* *alpha*=8). The definitive status of this kind of genotype (homozygous or heterozygous) will be inferred during a final imputation step.
- For heterozygous genotypes in which one allele was extremely underrepresented (proportion of reads under 1%), they were considered as homozygous for the most represented allele (*vcfx* *beta*= 0.01). This procedure minimizes the influence of miss-mapped reads to the *HLA-F* region and the high level of sequencing errors that characterizes NGS data. Such correction can only be applied in situations characterized by high depth of coverage (100 or more reads available for the evaluated position).
- For heterozygous genotypes in which one allele was mildly underrepresented (with a proportion of reads between 1% and 20%), a missing allele was introduced representing this

underrepresented allele (vcfx delta= 0.20). The definitive status of this kind of genotype (homozygous or heterozygous) will be inferred during a final imputation step.

– For heterozygous genotypes inferred at segments with less than 5 available reads in which one allele was underrepresented (with a proportion of reads between 20% and 50%), a missing allele was introduced representing this underrepresented allele (vcfx gamma=0.50). The definitive status of this kind of genotype will be inferred during a final imputation step.

After applying the rules described above, the *HLA-F* variation data presented approximately 0.55% of missing alleles that were imputed latter by the PHASE algorithm (48) as described further.

The *HLA-F* gene presents some short tandem repeats (STR) along the regulatory and coding regions. One of these STRs lays on the *HLA-F* 3'UTR, at the approximate position +3096 within the segment that is tracked by the IMGT/HLA database. The genotypes regarding such STR were inferred manually by calculating the rate of stutter formation comparing known genotypes and observing the number of reads detected for each allele by the GATK UnifiedGenotyper routine.

The association between each variable site was inferred using GATK (routine ReadBackedPhasing) using a minimal Phase Quality Threshold of 1000, dcov set to 1000 and minimum base quality set to 25. This assures that only alleles from variable sites that are close enough to be present in a same fragment (same read) should be phased. Considering the low *HLA-F* polymorphism described so far and the fact that, sometimes, variable sites are quite distant from each other in a same sample, not all variable sites were straightforwardly phased. In the present series, 61.7% of the heterozygous sites were phased using GATK. The remaining 38.3% were phased using the PHASE algorithm (48), which also imputed the 0.55% of missing alleles.

To proceed with the PHASE algorithm analysis, all singletons were removed, i.e., variable sites detected in only one individual and in a heterozygous state. Then, the partially GATK-phased VCF file was converted into an input file for PHASE and a file containing the known phases between variable sites. In some situations, blocks of known phase, but with unknown phase among them, were generated and the PHASE algorithm was used considering each of these blocks and the resultant haplotypes were compared. In the final PHASE run, all inferred haplotype pairs were compatible with the known phases (the ones determined by

GATK). The most probable haplotype pair of each sample was then converted into complete *HLA-F* sequences using the hg19 reference sequence as a draft and replacing the correct nucleotide in each position. In addition, singletons that were properly phased by GATK were introduced at their proper sequences. By using a local BLAST server with databases containing all known HLA alleles described so far (files in Fasta format downloaded from IMGT/HLA database, version 3.21.1), the closest known *HLA-F* coding allele was defined for each haplotype.

Results

The strategy proposed in the previous section revealed the presence of 70 variable sites on the evaluated *HLA-F* segment considering the 196 samples from Brazil (Table 1). Since there is no consensus regarding where the *HLA-F* transcription starts, and which segment characterizes the *HLA-F* promoter, and also considering that the IMGT/HLA database indicates the segment between positions -300 and -1 as a 5' untranslated region (5'UTR), here we considered all variable sites detected before nucleotide -300 as variable sites encompassing the 5' upstream segment. The variable sites between nucleotide -300 and -1 were considered within the 5'UTR. The same was applied to variable sites after the *HLA-F* stop codon. They were classified as 3'UTR when these variable sites are within the segment tracked by the IMGT/HLA database, 3'UTR or 3'downstream in segments that are not tracked by the IMGT/HLA but are considered by the hg19 annotations as 3'UTR, or only 3' downstream for variable sites located after the transcription end site described by the hg19 annotations (Table 1).

Altogether, we detected ten variable sites located upstream the first translated ATG between segment -556 and -1. Of those, 5 variable sites presented a minor allele frequency (MAF) higher than 10%. In addition, considering the segment tracked by the IMGT/HLA database, i.e, from -300 to -1, two variable sites detected here may be considered as new mutations since there is no description of them in this database. Considering the segment between the first translated codon until the stop codon, 45 variable sites were detected (table 1). These variable sites were distributed as follows: 10 sites on exon segments, in which 50% represent non-synonymous mutations located at exon 2 (position +342), which has been previously associated with the F*01:04 allele, two variable sites at exon 3 (position +923 and +982, both configuring new variable sites), and two at exon 4 (positions +1570 and +1771), in which the first configures a new variable site and the second has been previously associated

with alleles F*01:03:01:01 or F*01:03:01:02. At least 15 variable sites were detected downstream the *HLA-F* stop codon, some of them presenting high frequencies for the minor allele (table 1). Finally, 60 out of 70 variable sites (85.7%) detected in this Brazilian population sample has also been detected and reported by the 1000Genomes consortium when evaluating 2504 individuals from worldwide populations (table 1).

These 70 variable sites were found arranged into 37 different extended haplotypes. In order to characterize these extended haplotypes, we will present each *HLA-F* segment separately, starting from the coding sequence (table 2), the 5'upstream (table 3) and 3'UTR/3' downstream haplotypes (table 4), and finally, the extended ones (table 5).

The haplotypes for the coding segments were named according to the closest known *HLA-F* allele already reported by the IMGT/HLA database followed by any differences eventually detected. For that, we considered the segment between nucleotide -300 and up to nucleotide 3250, which is the region followed by the IMGT/HLA database.

The list of coding *HLA-F* haplotypes is presented at table 2. The segment considered for these coding haplotypes is the same region tracked by the IMGT/HLA database, which includes 300 bases as 5'UTR (from -300 to -1), the segment between the nucleotide +1 and the stop codon, and the 3'UTR up to nucleotide +3250. This segment was selected in order to allow us name the haplotype following the pattern proposed by the IMGT/HLA database and also to compare our data with the one already in the IMGT/HLA database.

Altogether, 30 different coding haplotypes (or coding alleles) were detected on this Brazilian population sample. Of those, five were identical to known *HLA-F* alleles, including alleles F*01:01:03:01, F*01:01:01:08, F*01:01:01:01 and F*01:01:03:03, listed in order of frequency (Table 2). The remaining 25 coding haplotypes represents new *HLA-F* sequences or new *HLA-F* alleles. Although only five haplotypes were identical to known *HLA-F* sequences, some of the remaining 25 haplotypes carry variable sites associated with specific known alleles, but the complete sequence does not match the ones presented at the IMGT/HLA database. For instance, all five copies of the sequence carrying adenine at position +342 (rs17875380, table 1), which is associated with the IMGT/HLA allele F*01:04, also carried guanine at position +1943 (rs17184813), which does not match the described IMGT/HLA sequence. Other coding haplotypes followed this same pattern, such as F*01:01:01:05 and F*01:01:02:03.

New variable sites, i.e., variable sites that have not been described at the IMGT/HLA database yet, were also detected in this present series and configure new *HLA-F* alleles. They were placed at table 2 with the “new” mutations indicated after the closest known *HLA-F* allele.

Some of these haplotypes presented high frequencies, such as F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, new3205G, +3TG) and F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, 2811delG, new3205G, +3TG), carrying three new mutations and a summed frequency higher than 5% in this Brazilian population sample. However, besides these new variable sites, only one new coding haplotype encodes an HLA-F molecule different from the ones already described (i.e., F*01:01, F*01:03 and F*01:04) and it was indicated as “unnamed” at table 2 and detected in just one individual. Nevertheless, besides possible errors regarding the genotype call at the microsatellite (as described in the Methods section), the number of new *HLA-F* sequences reported here (25 new sequences, table 2) is higher than the number of reported *HLA-F* sequences so far (22 known HLA-F sequences considering release 3.21.1). Since most of the variable sites were detected at intron segments or represent synonymous mutations, only four different *HLA-F* molecules are encoded by the coding sequences here described, and F*01:01 represents more than 95% of all molecules.

Considering the 5' upstream segment evaluated (between nucleotides -556 and -301), five haplotypes were detected (table 3). These haplotypes were named sequentially according to their frequencies as F*Upstream-A to F*Upstream-E. Two haplotypes were quite frequent, the first one (F*Upstream-A) found with a frequency of 58.9%, and the second one (F*Upstream-B) with a frequency of 37.5% and the only one carrying guanine at position -547 (rs3998799).

The 10 haplotypes detected for the *HLA-F* 3' downstream segment (variable sites after the last position tracked by the IMGT/HLA database, i.e., position +3250) are illustrated at table 4. As for the 5' upstream segment, haplotypes were named following their frequencies. The most frequently haplotypes found for this segment were named as F*UTR-1 (frequency = 0.4158), F*UTR-2 (frequency = 0.2372), F*UTR-3 (frequency = 0.1531) and F*UTR-4 (frequency = 0.1097), followed by other less frequent haplotypes.

It should be mentioned that we detected 14 variable sites that were found just in one individual in heterozygosis, therefore, they were considered as singletons (Table 1). Of those, only two singletons had their association with other variable sites inferred by the GATK and were included on the haplotype analysis (positions +1225 and +1570). The *HLA-F* segment between the first translated ATG and the stop codon presented 11 singletons, in which three represent non-synonymous mutations (positions +923, +982 and +1570).

The upstream, coding and downstream haplotypes were found arranged into 37 extended haplotypes as presented in table 5. Usually, each different HLA-F coding allele is

associated with an unique upstream or downstream haplotype, with rare exceptions. For instance, the coding allele F*01:01:01:01 and new sequences that were probably originated from it by single mutations are mainly associated with F*Upstream-B and F*UTR-1, and the F*01:04 sequence detected in Brazil is associated with F*Upstream-A and F*UTR-2. In this context, the mutation that characterizes de F*01:04 allele might be in strong linkage disequilibrium (LD) with position +3942 that is exclusively associated with F*UTR-2 (table 4). In fact, the HLA-F entire segment presented a strong LD, which justify the pattern of association observed among upstream, coding and downstream haplotypes (Figure 1).

Discussion

The technology behind NGS is similar to Sanger Sequencing in many ways, providing reliable and fast results even when several genes and samples are evaluated at the same time. DNA polymerase catalyzes the incorporation of fluorescent nucleotides into a DNA template strand during sequential cycles of DNA synthesis. However, this procedure is done on millions of templates at the same time. One of the major advances in NGS was the development of paired-end sequencing, which reveal the sequences of both ends of a given template, enhancing accuracy and improving indel detection (49). Library preparation for NGS is a straightforward and faster procedure when compared with traditional sequencing methods, and usually requires lower DNA inputs (49). After sequencing, NGS procedures generates several sequences (named as reads), in which the size may vary according to the method used, which must be processed and aligned to a reference in order to get the complete template sequence or to infer genotypes.

NGS procedures are useful to characterize the variability of any DNA segment, but its use is particularly challenging when HLA genes are the main focus. This issue arises mainly due the high level of sequence similarity among paralogous HLA genes. Since NGS techniques produce several small sequences that must be placed (or mapped) at the correct position of a reference, but small sequences derived from HLA genes might be mapped at more than one HLA gene or even in more than one position of a same HLA gene, some level of miss-mapped reads is expected for HLA genes when conventional mapping methods are used. In addition, the high level of polymorphisms that characterizes HLA genes might bias the read mapping procedure (50), mainly overestimating variants that are close to the reference used. In addition, polymorphisms might turn a sequence derived from one HLA gene similar (or close to) sequences that characterize other HLA genes. This is particularly complex when more than one

HLA gene is processed together (as it was in this study) and an unreliable reading mapping would bias genotype inference.

To circumvent this issue, we used two independent softwares to achieve a reliable mapping of the *HLA-F* sequences. The first one, hla-mapper, applies a series of filters to address each read to its proper gene, as described elsewhere (15). The second, vcfx, applies rules to accept (or change) homozygosity or heterozygosity that were inferred by the genotype call method described earlier. This strategy allowed a reliable reading mapping and an improved genotype inference.

Sequencing libraries were generated by using the Nextera® XT Sample Preparation and Nextera® Index kits (Illumina Inc.), and sequencing was performed by using the MiSeq Personal Sequencer platform (Illumina Inc.). Although the use of Nextera Kits is straightforward, we detected some difficulties regarding its use and bias in the template fragmentation pattern produced by them. This bias was mainly characterized by a subrepresentation of reads on certain GC-rich gene segments, typical of regulatory regions and some introns (51, 52). In the case of *HLA-F*, we detected such bias on two short segments, the first at the upstream segment around position -291, and the second at intron 2 around position +650. The coverage detected was calculated considering the number of reads covering a specific base. Despite we aimed to achieve a minimum coverage of 500 for the entire segment, and in most of the cases the average coverage was even higher than 1500, the coverage varied widely considering all samples for *HLA-F* gene and the segments mentioned above presented coverage reduction reaching sometimes a coverage around 20 reads, even if the average was higher than 1000. This appears to be a bias of DNA enzymatic fragmentation produced by the Nextera Kit (52). Those segments presenting low coverage would be prone to genotyping errors, especially when a homozygous genotype is inferred at a position presenting low coverage. This issue has been dealt with hla-mapper and vcfx, and also by the imputation steps that were performed and described in the methods section. Nevertheless, approaches such as this one should be considered when analyzing HLA genes by NGS and when using enzymatic fragmentation kits such as Nextera.

The use of vcfx introduced only 0.55% of missing alleles and the program PHASE (48) was used to impute such missing alleles. In addition, the strategy used here to infer haplotypes took advantage of the NGS data inferring the association between the heterozygous sites present on a same read by the GATK application, but also using probabilistic inference by the PHASE

algorithm. Both strategies were combined in order to produce reliable haplotypes, in which both fragment-inferred and probabilistic-inferred haplotypes were the same.

The overall *HLA-F* variability has been minimally explored among human populations. Currently, only two population samples were evaluated regarding it. The first one described the *HLA-F* variability detected in a Chinese population sample considering two different ethnic groups (Southwestern and Northeastern China) (41). The results indicated that 94% and 5% of the encoded HLA-F molecules are F*01:01 and F*01:03, respectively, in which the first presents serine and the second proline at position 251 (codon 251). Pan and colleagues also noted that geographical and ethnic factors were related to the *HLA-F* allelic distribution (41). However, this study searched for known *HLA-F* variable sites only by using sequence-specific primers. The second study, conducted in a Brazilian population sample from Curitiba (Parana State, South Brazil), used sequencing-base typing to characterize the *HLA-F* variability in a total of 1673-bp comprehending exons 1, 2 and 4, introns 1, 2 and 4 and a short segment of exon 3 and intron 5 (42). The results indicate that the alleles detected at the South Brazil encode mainly the F*01:01 (87.19%) and F*01:03 (12.31%) (42). Notwithstanding that, considering the limitations of these two studies, i.e., the search for known variable sites only or the characterization of only a segment of the *HLA-F* gene, the *HLA-F* variability might have been underestimated. Therefore, it is not clear whether *HLA-F* is really conserved at the nucleotide and protein level, or it was not explored properly. To date, there are 22 coding HLA-F alleles (or haplotypes) described by the IMGT/HLA database version 3.21.1. These alleles encode only 4 different HLA-F molecules, defined as HLA-F*01:01, F*01:02, F*01:03 and F*01:04.

This is the first study to evaluate the entire *HLA-F* variability using a methodology capable of detecting every variable site present in the sample. In addition, the regulatory regions, including a 5' upstream segment (which might include the *HLA-F* promoter), the 5'UTR and the 3'UTR, were also evaluated. As observed in China and Brazil on the former studies, the *HLA-F* gene is associated with low protein polymorphism. Only four different HLA-F encoded molecules were found in the present sample, including F*01:01 (82.45%), F*01:03 (16.08%) and F*01:04 (1.28%), following the same pattern observed in China (41) and in the South Brazil (42). We did not detect the F*01:02 allele in the present sample, corroborating the former *HLA-F* study from Brazil (42). In addition, a new HLA-F molecule due to a novel non-synonymous mutation was also detected, but in only one individual.

Besides this protein conservation, we detected 70 variable sites comprehending the entire *HLA-F* gene considering about 5000 nucleotides long, encompassing the 5' upstream

(from -547) to the HLA-F 3'UTR and a downstream segment (to +4228). These variable sites were arranged into 37 extended haplotypes, which includes 30 coding haplotypes considering the segment tracked by the IMGT/HLA database. The number of different coding haplotypes was much higher than the current knowledge considering the IMGT/HLA database, since only five (16.6%) were identical to known *HLA-F* haplotypes (F*01:01:01:01, F*01:01:01:08, F*01:01:03:01, F*01:01:03:03 and F*01:03:01:01). Altogether, these five haplotypes represent 75% of all the *HLA-F* sequences in this sample. Therefore, the *HLA-F* haplotype variability (or number of alleles) is much higher than the one already described by the IMGT/HLA.

The remaining 25 coding haplotypes did present nucleotide differences from a known IMGT/HLA haplotype, but most of them were found at least twice in the present sample. These haplotypes configure new *HLA-F* alleles, increasing the number of alleles known for this gene. However, it should be noted that the existence of such new *HLA-F* alleles must be confirmed by cloning and Sanger sequencing, especially those that diverge only by the number of dinucleotide repeats on the STR at position +3096, and a proper name must be assigned for them by the IMGT/HLA database. Therefore, these new sequences described here were named considering the closest known allele followed by the differences detected.

Here we detected only five coding haplotypes that are identical to the ones already described. Some of the new sequences detected were very close to known *HLA-F* sequences, differing by few variable sites. One example is the sequence encoding the F*01:04 molecule, which was different from the F*01:04 allele described in the IMGT/HLA database. In all five cases, the F*01:04 allele detected in Brazil lacks the guanine deletion located at position +1943, as described by the IMGT/HLA database. In fact, this deletion is associated with several *HLA-F* alleles as described by the IMGT/HLA database, but we did not detect it in the present sample. Another example is the eighteen copies of the F*01:03:01:02 allele detected in Brazil, which differ from the IMGT/HLA allele in at least two positions, including the lack of the guanine deletion at position +1943 and a different nucleotide at position +3189 (table 2). It should be noted that the 1000Genomes Consortium (phase 3) did not detect this guanine deletion either, leading us wondering whether this deletion really exists and how accurate are the IMGT/HLA sequences regarding the *HLA-F* gene available so far. Nevertheless, despite this increased number of *HLA-F* coding alleles, exception made to just two sequences, these new alleles present synonymous mutations in exons or variable sites in introns, thus encoding the same F*01:01 or F*01:03 as previously described. This protein conservation is probably a consequence of the key role of *HLA-F* in the immune system physiology.

Until now, it was not clear if *HLA-F* is indeed conserved, as we demonstrated here. It is possible that this low variability is associated with its critical role in the immune system. The mechanisms underlying this lack of protein variability detected for the *HLA-F* gene are not understood, mainly because *HLA-F* is located within the most variable region of the human genome. This phenomenon has already been observed for other non-classical class I genes such as *HLA-G* and *HLA-E* (13-16, 53-56) which present tolerogenic and immune modulatory properties (18-21, 23, 57, 58). The *HLA-F* gene is also very conserved among primates evolution (including bonobo, gorilla, orangutan and chimpanzee), suggesting its critical role in the immune response (26). It is possible that the MHC-F molecule (or *HLA-F* in humans) is involved together with the MHC-G and MHC-E (*HLA-G* and *HLA-E* in humans) in the regulation of T and NK cells activity (59). According to analyses of the predicted HLA-F amino acid sequences and their comparison with other primate MHC-F proteins, HLA-F is a molecule with typical classical class I structure, i.e, the presence of heavy chain formed by the peptide binding ($\alpha 1$ and $\alpha 2$), immunoglobulin like ($\alpha 3$), transmembrane and cytoplasmic domains, and its association with $\beta 2$ microglobulin (27). Finally, HLA-F has a restricted tissue expression pattern (28, 35, 60). Because of these characteristics and the similarities with other non-classical genes such as *HLA-G* and *HLA-E*, it is believed that the HLA-F molecule has tolerogenic properties and has a key role in immune modulation.

Although the expression of HLA-F on grafted tissues, during pregnancy and on pathological conditions such as tumors are poorly explored, we can not exclude the *HLA-F* role in the outcome of such situations, since this molecule may interact with other HLA molecules presenting open conformation, stabilizing them on the cell surface (61), thus influencing the function of other HLA classical and non-classical molecules.

To the best of our knowledge, this is the first study to fully characterize the *HLA-F* gene variability considering both regulatory segments and the entire coding sequence. This characterization was performed in a Brazilian population sample, which is considered one of the most admixed populations in the world and a significant repository of genetic variation (44). Whatever the *HLA-F* function is, acting on the stabilization of class I molecules presenting an open conformation, or acting on the recycling of class I molecules, interacting with other non-classical molecules or even interacting with inhibitory receptors such as ILT2 and ILT4, this function might be critical to the immune system and cellular physiology considering the protein conservation observed here.

Acknowledgements

This work was supported by FAPESP/Brazil (Grant# 2013/17084-2). E.C.C, E.A.D. and C.T.M.J. are supported by CNPq/Brazil (Grants# 304471/2013-5, 304753/2009-2, and 309572/2014-2).

References

1. Geraghty DE, Koller BH, Hansen JA, Orr HT. The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. *Journal of immunology*. 1992;**149**:1934-46.
2. Rojo R, Castro MJ, Martinez-Laso J, et al. MHC-F DNA sequences in bonobo, gorilla and orangutan. *Tissue antigens*. 2005;**66**:277-83.
3. Bjorkman PJ, Parham P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annual review of biochemistry*. 1990;**59**:253-88.
4. Carosella ED, Favier B, Rouas-Freiss N, Moreau P, Lemaoult J. Beyond the increasing complexity of the immunomodulatory HLA-G molecule. *Blood*. 2008;**111**:4862-70.
5. Cao M, Yie SM, Liu J, Ye SR, Xia D, Gao E. Plasma soluble HLA-G is a potential biomarker for diagnosis of colorectal, gastric, esophageal and lung cancer. *Tissue antigens*. 2011;**78**:120-8.
6. Dunker K, Schlaf G, Bukur J, Altermann WW, Handke D, Seliger B. Expression and regulation of non-classical HLA-G in renal cell carcinoma. *Tissue antigens*. 2008;**72**:137-48.
7. Twito T, Joseph J, Mociornita A, Rao V, Ross H, Delgado DH. The 14-bp deletion in the HLA-G gene indicates a low risk for acute cellular rejection in heart transplant recipients. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation*. 2011;**30**:778-82.
8. Brown R, Kabani K, Favaloro J, et al. CD86+ or HLA-G+ can be transferred via trogocytosis from myeloma cells to T cells and are associated with poor prognosis. *Blood*. 2012;**120**:2055-63.
9. Rouas-Freiss N, Goncalves RM, Menier C, Dausset J, Carosella ED. Direct evidence to support the role of HLA-G in protecting the fetus from maternal uterine natural killer cytotoxicity. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;**94**:11520-5.
10. Hviid TV, Rizzo R, Melchiorri L, Stignani M, Baricordi OR. Polymorphism in the 5' upstream regulatory and 3' untranslated regions of the HLA-G gene in relation to soluble HLA-G and IL-10 expression. *Human immunology*. 2006;**67**:53-62.

11. Hviid TV, Milman N, Hylenius S, Jakobsen K, Jensen MS, Larsen LG. HLA-G polymorphisms and HLA-G expression in sarcoidosis. *Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG / World Association of Sarcoidosis and Other Granulomatous Disorders*. 2006;**23**:30-7.
12. Hviid TV, Hylenius S, Rorbye C, Nielsen LG. HLA-G allelic variants are associated with differences in the HLA-G mRNA isoform profile and HLA-G mRNA levels. *Immunogenetics*. 2003;**55**:63-79.
13. Felicio LP, Porto IO, Mendes-Junior CT, et al. Worldwide HLA-E nucleotide and haplotype variability reveals a conserved gene for coding and 3' untranslated regions. *Tissue antigens*. 2014;**83**:82-93.
14. Veiga-Castelli LC, Castelli EC, Mendes CT, Jr., 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*. 2012;**79**:15-24.
15. Castelli EC, Mendes-Junior CT, Sabbagh A, et al. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. *Human immunology*. 2015.
16. Castelli EC, Mendes-Junior CT, Veiga-Castelli LC, Roger M, Moreau P, Donadi EA. A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. *Mol Biol Evol*. 2011;**28**:3069-86.
17. Antoun A, Jobson S, Cook M, Moss P, Briggs D. Ethnic variability in human leukocyte antigen-E haplotypes. *Tissue antigens*. 2009;**73**:39-45.
18. Aractingi S, Briand N, Le Danff C, et al. HLA-G and NK receptor are expressed in psoriatic skin: a possible pathway for regulating infiltrating T cells? *The American journal of pathology*. 2001;**159**:71-7.
19. Biedron M, Rybka J, Wrobel T, Prajs I, Poreba R, Kuliczowski K. The role of soluble HLA-G and HLA-G receptors in patients with hematological malignancies after allogeneic stem cell transplantation. *Medical oncology*. 2015;**32**:664.
20. Brenol CV, Veit TD, Chies JA, Xavier RM. The role of the HLA-G gene and molecule on the clinical expression of rheumatologic diseases. *Revista brasileira de reumatologia*. 2012;**52**:82-91.
21. Carosella ED, Moreau P, Le Maoult J, Le Discorde M, Dausset J, Rouas-Freiss N. HLA-G molecules: from maternal-fetal tolerance to tissue acceptance. *Advances in immunology*. 2003;**81**:199-252.
22. da Silva GK, Vianna P, Veit TD, et al. Influence of HLA-G polymorphisms in human immunodeficiency virus infection and hepatitis C virus co-infection in Brazilian and Italian individuals. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2014;**21**:418-23.
23. Dong DD, Yie SM, Li K, et al. Importance of HLA-G expression and tumor infiltrating lymphocytes in molecular subtypes of breast cancer. *Human immunology*. 2012;**73**:998-1004.

24. Hviid TV. HLA-G in human reproduction: aspects of genetics, function and pregnancy complications. *Hum Reprod Update*. 2006;**12**:209-32.
25. Hosseini E, Schwarer AP, Jalali A, Ghasemzadeh M. The impact of HLA-E polymorphisms on relapse following allogeneic hematopoietic stem cell transplantation. *Leukemia research*. 2013;**37**:516-9.
26. Goodridge JP, Burian A, Lee N, Geraghty DE. HLA-F complex without peptide binds to MHC class I protein in the open conformer form. *Journal of immunology*. 2010;**184**:6199-208.
27. Wainwright SD, Biro PA, Holmes CH. HLA-F is a predominantly empty, intracellular, TAP-associated MHC class Ib protein with a restricted expression pattern. *Journal of immunology*. 2000;**164**:319-28.
28. Lepin EJ, Bastin JM, Allan DS, et al. Functional characterization of HLA-F and binding of HLA-F tetramers to ILT2 and ILT4 receptors. *European journal of immunology*. 2000;**30**:3552-61.
29. Arosa FA, Santos SG, Powis SJ. Open conformers: the hidden face of MHC-I molecules. *Trends in immunology*. 2007;**28**:115-23.
30. Goodridge JP, Burian A, Lee N, Geraghty DE. HLA-F and MHC class I open conformers are ligands for NK cell Ig-like receptors. *Journal of immunology*. 2013;**191**:3553-62.
31. Allan DS, Lepin EJ, Braud VM, O'Callaghan CA, McMichael AJ. Tetrameric complexes of HLA-E, HLA-F, and HLA-G. *Journal of immunological methods*. 2002;**268**:43-50.
32. Braud VM, Allan DS, McMichael AJ. Functions of nonclassical MHC and non-MHC-encoded class I molecules. *Current opinion in immunology*. 1999;**11**:100-8.
33. O'Callaghan CA, Bell JI. Structure and function of the human MHC class Ib molecules HLA-E, HLA-F and HLA-G. *Immunological reviews*. 1998;**163**:129-38.
34. Ishitani A, Sageshima N, Lee N, et al. Protein expression and peptide binding suggest unique and interacting functional roles for HLA-E, F, and G in maternal-placental immune recognition. *Journal of immunology*. 2003;**171**:1376-84.
35. Shobu T, Sageshima N, Tokui H, et al. The surface expression of HLA-F on decidual trophoblasts increases from mid to term gestation. *Journal of reproductive immunology*. 2006;**72**:18-32.
36. Apps R, Gardner L, Moffett A. A critical look at HLA-G. *Trends in immunology*. 2008;**29**:313-21.
37. Hosseini E, Schwarer AP, Ghasemzadeh M. Do human leukocyte antigen E polymorphisms influence graft-versus-leukemia after allogeneic hematopoietic stem cell transplantation? *Experimental hematology*. 2015;**43**:149-57.
38. Krongvorakul J, Kantachuvesiri S, Ingsathit A, et al. Association of soluble human leukocyte antigen-G with acute tubular necrosis in kidney transplant recipients. *Asian Pacific*

journal of allergy and immunology / launched by the Allergy and Immunology Society of Thailand. 2015;**33**:117-22.

39. Mossallam GI, Fattah RA, El-Haddad A, Mahmoud HK. HLA-E polymorphism and clinical outcome after allogeneic hematopoietic stem cell transplantation in Egyptian patients. *Human immunology*. 2015;**76**:161-5.

40. Crispim JC, Mendes-Junior CT, Wastowski IJ, et al. HLA polymorphisms as incidence factor in the progression to end-stage renal disease in Brazilian patients awaiting kidney transplant. *Transplantation proceedings*. 2008;**40**:1333-6.

41. Pan FH, Liu XX, Tian W. Characterization of HLA-F polymorphism in four distinct populations in Mainland China. *Int J Immunogenet*. 2013.

42. Manvailer LF, Wowk PF, Mattar SB, da Siva JS, da Graca Bicalho M, Roxo VM. HLA-F polymorphisms in a Euro-Brazilian population from Southern Brazil. *Tissue antigens*. 2014;**84**:554-9.

43. Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SD. Color and genomic ancestry in Brazilians. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;**100**:177-82.

44. Ferreira LB, Mendes-Junior CT, Wiezel CE, Luizon MR, Simoes AL. Genomic ancestry of a sample population from the state of Sao Paulo, Brazil. *American journal of human biology : the official journal of the Human Biology Council*. 2006;**18**:702-5.

45. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;**43**:491-8.

46. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. 2010;**20**:1297-303.

47. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*. 2013;**11**:11 0 1- 0 33.

48. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 2001;**68**:978-89.

49. Grada A, Weinbrecht K. Next-generation sequencing: methodology and application. *The Journal of investigative dermatology*. 2013;**133**:e11.

50. Brandt DY, Aguiar VR, Bitarello BD, Nunes K, Goudet J, Meyer D. Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data. *G3*. 2015;**5**:931-41.

51. Shiina T, Suzuki S, Ozaki Y, et al. Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. *Tissue antigens*. 2012;**80**:305-16.

52. Wang W, Wei Z, Lam TW, Wang J. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Scientific reports*. 2011;**1**:55.
53. Castelli EC, Ramalho J, Porto IO, et al. Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity. *Frontiers in immunology*. 2014;**5**:476.
54. Mendes-Junior CT, Castelli EC, Meyer D, Simoes AL, Donadi EA. Genetic diversity of the HLA-G coding region in Amerindian populations from the Brazilian Amazon: a possible role of natural selection. *Genes and immunity*. 2013;**14**:518-26.
55. Sabbagh A, Luisi P, Castelli EC, et al. Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: balancing selection influencing genetic diversity. *Genes and immunity*. 2014;**15**:95-106.
56. van der Ven K, Skrablin S, Engels G, Krebs D. HLA-G polymorphisms and allele frequencies in Caucasians. *Human immunology*. 1998;**59**:302-12.
57. Zheng H, Lu R, Xie S, et al. Human leukocyte antigen-E alleles and expression in patients with serous ovarian cancer. *Cancer science*. 2015;**106**:522-8.
58. Zanetti BR, Carvalho-Galano DF, Feitosa NL, et al. Differential expression of immune-modulatory molecule HLA-E in non-neoplastic and neoplastic lesions of the thyroid. *International journal of immunopathology and pharmacology*. 2013;**26**:889-96.
59. Moscoso J, Serrano-Vela JJ, Arnaiz-Villena A. MHC-F polymorphism and evolution. *Tissue antigens*. 2007;**69 Suppl 1**:136-9.
60. Ishitani A, Sageshima N, Hatake K. The involvement of HLA-E and -F in pregnancy. *Journal of reproductive immunology*. 2006;**69**:101-13.
61. Goodridge JP, Lee N, Burian A, et al. HLA-F and MHC-I open conformers cooperate in a MHC-I antigen cross-presentation pathway. *Journal of immunology*. 2013;**191**:1567-77.

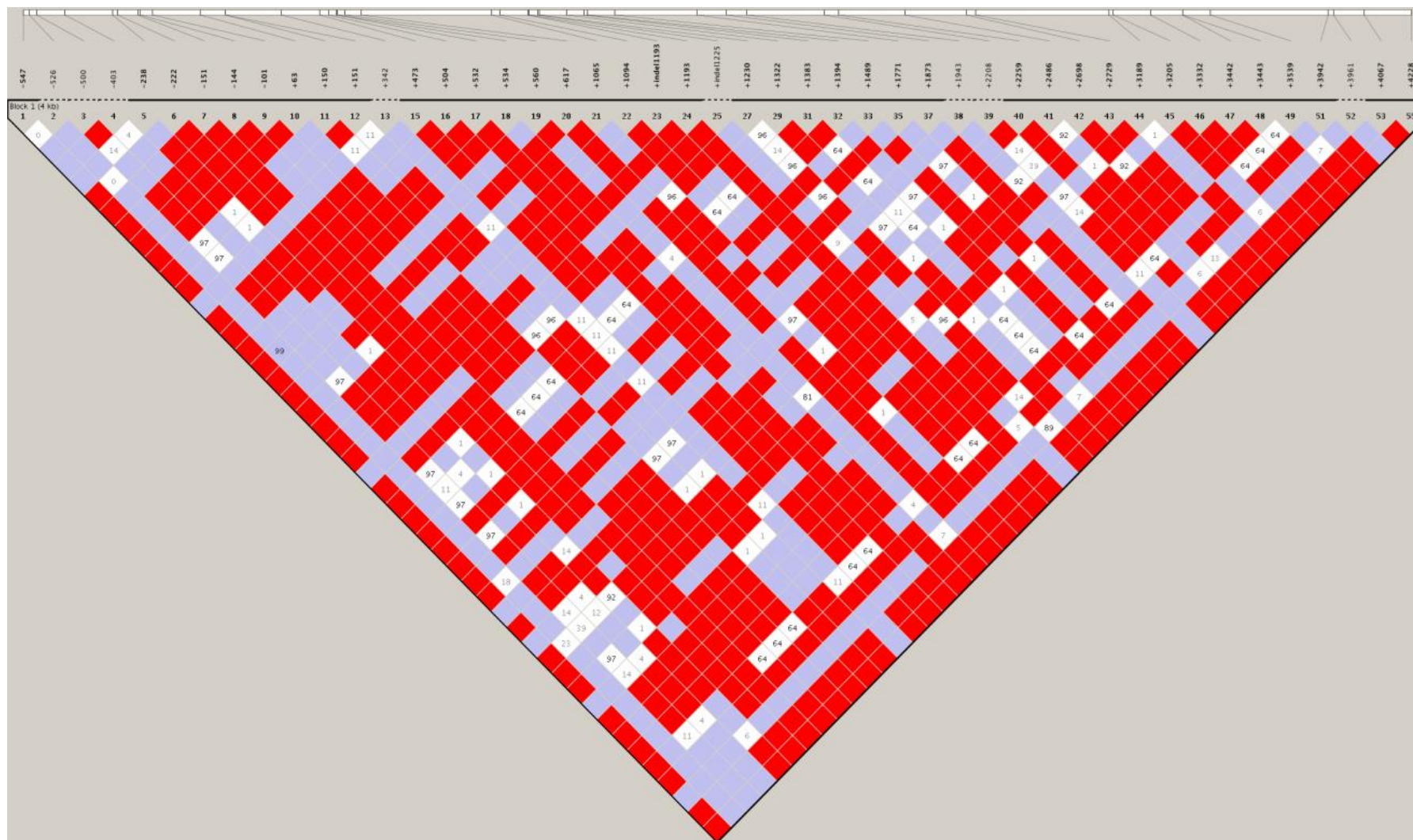


Figure 01: Linkage Disequilibrium patterns considering variable sites at the *HLA-F* gene and surrounded sequences in a Brazilian population sample. LD plot generated by Haploview. Areas in dark red or dark gray indicate strong LD ($\text{LOD} \geq 2$, $D' = 1$), shades of pink or shades of gray indicate moderate LD ($\text{LOD} \geq 2$, $D' < 1$), blue or light gray indicates weak LD ($\text{LOD} < 2$, $D' = 1$), and white indicates no LD ($\text{LOD} < 2$, $D' < 1$). D' values different from 1.00 are represented inside the squares as percentages. LOD, log of the odds; D' , pairwise correlation between single-nucleotide polymorphisms. The LD plot was generated using SNPs with a minimum allele frequency of 1%.

Table 1: List of variable sites detected at the *HLA-F* gene considering the segment between nucleotide -556 and +4252, considering the Adenine of the first translated ATG as +1

Chr6 Pos	SNPid	HLA-F segment ^a	Notes ^b	IMGT/HLA	Reference allele	Frequency	First alternative	Frequency	Second alternative	Frequency
29690694	rs3998799	5' Upstream	F	-547	C	0.6250	G	0.3750		
29690715	rs114795943	5' Upstream	F	-526	T	0.9898	A	0.0102		
29690741	rs17875377	5' Upstream	F	-500	G	0.9745	A	0.0255		
29690838	rs17875378	5' Upstream	F	-403	G	0.9821	A	0.0179		
29691003	rs56044823	5'UTR	C,F	-238	C	0.9464	T	0.0536		
29691019	rs1362126	5'UTR	F	-222	G	0.6250	A	0.3750		
29691090	rs1362125	5'UTR	F	-151	T	0.5612	A	0.4388		
29691097	rs2075682	5'UTR	F	-144	A	0.8367	T	0.1633		
29691140	rs2072896	5'UTR	F	-101	C	0.8367	G	0.1633		
29691197	.	5'UTR	C,E	-44	A	0.9974	AT	0.0026		
29691303	rs2076183	Exon 1	A,F	63	G	0.8367	A	0.1633		
29691390	rs563604282	Intron 1	F	150	C	0.1607	G	0.8393		
29691391	rs530577894	Intron 1	F	151	T	0.1607	C	0.8393		
29691445	rs61739958	Exon 2	A,C,E,F	205	C	0.9974	A	0.0026		
29691582	rs17875380	Exon 2	B,F	342	C	0.9872	A	0.0128		
29691634	rs117540842	Exon 2	A,C,F	394	C	0.9923	G	0.0077		
29691713	rs2076182	Intron 2	F	473	A	0.8367	C	0.1633		
29691744	rs2076181	Intron 2	F	504	C	0.8367	G	0.1633		
29691772	rs2076179	Intron 2	F	532	T	0.5612	C	0.4388		
29691774	rs2076178	Intron 2	F	534	A	0.8010	C	0.1990		
29691800	rs17875381	Intron 2	F	560	C	0.9490	G	0.0510		
29691857	rs2072895	Intron 2	F	617	C	0.5612	G	0.4388		
29691897	.	Intron 2	C,E	657	G	0.9974	C	0.0026		
29692163	.	Exon 3	B,C,E	923	G	0.9974	A	0.0026		
29692222	.	Exon 3	B,C,E	982	G	0.9974	T	0.0026		
29692305	rs1632953	Intron 3	F	1065	A	0.1607	G	0.8393		
29692334	rs9258186	Intron 3	F	1094	A	0.8367	G	0.1633		
29692382	.	Intron 3	C,E	1142	G	0.9974	C	0.0026		
29692429	rs140848774	Intron 3	F	1193-1204	AAATTCTGAGGG	0.8367	A	0.1633		
29692433	rs534575062	Intron 3	F	1193	T	0.3240	C	0.6760		
29692462	rs142275427	Intron 3	F	1225-1239	TGGAATACCGATCCGC	0.9694	T	0.0306		
29692465	rs368156595	Intron 3	C,D,F	1225	A	0.9974	C	0.0026		
29692470	rs563176182	Intron 3	F	1230	C	0.8673	A	0.1327		

29692553	rs193250743	Intron 3	C,F	1313	A	0.9949	G	0.0051		
29692562	rs1736926	Intron 3	F	1322	A	0.1607	G	0.8393		
29692618	rs144376525	Intron 3	C,F	1378	C	0.9949	T	0.0051		
29692622	rs2072899	Intron 3	F	1382	C	0.3980	A	0.4388	G	0.1633
29692623	rs17178385	Intron 3	F	1383	C	0.9490	G	0.0510		
29692634	rs1736925	Intron 3	F	1394	T	0.1607	G	0.8393		
29692710	.	Intron 3	C,E	1470	A	0.9974	G	0.0026		
29692729	rs2072898	Intron 3	F	1489	T	0.8367	G	0.1633		
29692737	rs181298082	Intron 3	C,E,F	1497	T	0.9974	C	0.0026		
29692810	rs142311857	Exon 4	B,C,D,F	1570	C	0.9974	A	0.0026		
29693011	rs1736924	Exon 4	B,F	1771	C	0.1607	T	0.8393		
29693019	rs201844059	Exon 4	A,C,F	1779	G	0.9949	A	0.0051		
29693113	rs2076177	Intron 4	F	1873	C	0.8367	T	0.1633		
29693183	rs17184813	Intron 4	F	1943	G	0.9566	A	0.0434		
29693448	rs17875383	Intron 5	F	2208	C	0.9745	T	0.0255		
29693499	rs2235383	Intron 5	F	2259	A	0.8367	G	0.1633		
29693579	.	Intron 5	C,E	2339	C	0.9974	G	0.0026		
29693726	rs114325231	Intron 5	C,F	2486	G	0.9464	T	0.0536		
29693938	rs2735061	Intron 6	F	2698	G	0.8240	A	0.1760		
29693969	rs1736922	Exon 7	A,F	2729	C	0.7628	T	0.2372		
29694044	rs1736922	Intron 7	F	2811	T	0.8546	TG	0.1454		
29694046	rs201715782	Intron 7	C,E,F	2805	G	0.9974	GA	0.0026		
29694336 ^c	rs56321148	3'UTR ^c	F	3096 ^c	ATGTGTG	0.2602	ATG	0.2372	ATGTGTGTG	0.3878
29694427	rs1059174	3'UTR	F	3189	C	0.1607	T	0.8393		
29694443	rs62391801	3'UTR	C,F	3205	A	0.9464	G	0.0536		
29694526	.	3'UTR or 3' downstream	E	3288	C	0.9974	T	0.0026		
29694570	rs3734813	3'UTR or 3' downstream	F	3332	A	0.8367	G	0.1633		
29694680	rs3734814	3'UTR or 3' downstream	F	3442	A	0.8367	C	0.1633		
29694681	rs3734815	3'UTR or 3' downstream	F	3443	A	0.8367	T	0.1633		
29694777	rs17875384	3'UTR or 3' downstream	F	3539	G	0.9490	A	0.0510		
29694832	rs111228498	3'UTR or 3' downstream	E,F	3594	C	0.9974	T	0.9974		
29695176	rs371631016	3' downstream	F	3938	TTA	0.9949	T	0.0051		
29695180	rs1419696	3' downstream	F	3942	G	0.7628	A	0.2372		
29695199	rs143726784	3' downstream	F	3961	A	0.9847	G	0.0153		
29695305	rs2523405	3' downstream	F	4067	T	0.5612	G	0.4388		
29695412	rs17875386	3' near gene	-	4174	C	0.9923	T	0.0077		
29695466	rs2735060	3' near gene	-	4228	T	0.5612	A	0.4388		

^a **HLA-F segment**

* [5' upstream gene] corresponds to a variable site considered at this segment by both the IMGT/HLA database and the human genome draft (version hg19). Since the HLA-F promoter segment is not yet complete characterized, these variable sites were considered at 5' upstream gene.

* [5' upstream gene or 5'UTR] corresponds to a variable site considered at the 5'UTR segment by the IMGT/HLA database, but as 5' upstream gene by the human genome draft (version hg19). Since the HLA-F promoter segment is not yet complete characterized, these variable sites were considered at 5' upstream gene or 5'UTR.

* [5'UTR] corresponds to a variable site considered at the 5'UTR segment by both the IMGT/HLA database and the human genome draft (version hg19).

* [3' downstream gene or 3'UTR] corresponds to a variable site considered at the 3'UTR segment by the IMGT/HLA database, but as 3' downstream gene by at list one of the HLA-F transcripts annotated at the human genome draft (version hg19).

* [3' downstream gene] corresponds to a variable site considered at the 3' downstream gene by both the IMGT/HLA database.

^bNOTES

A) Synonymous mutation on exon

B) Non-synonymous mutation on exon

C) New variable site on a region covered by the IMGT/HLA database

D) Singleton with a defined haplotype

E) Singleton with haplotype not inferred

F) This variable site was also detected by the 1000Genomes Project, Phase 3

^c Variable site at position 29694336 (+3096)

This variable site corresponds to an STR at position +3094. It presented five alternative alleles, differing from each other by the number of dinucleotide TG repeats. Only two alternative alleles (besides de reference one at the hg19) are presented. Others alternative alleles are (1) ATGTG with a frequency of 0.1020, (2) ATGTGTGTGTG with a frequency of 0.0102 and only A with a frequency of 0.0026.

Table 2: List of *HLA-F* haplotypes found in Brazil, considering the segment that is considered by the IMGT/HLA database

HLA-F coding haplotype *	HLA-F encoded molecule	Occurrences	Frequency (2n=392)
F*01:01:03:01	F*01:01	68	0.1735
F*01:01:01:08	F*01:01	67	0.1709
F*01:01:01:01	F*01:01	57	0.1454
F*01:03:01:01	F*01:03	40	0.1020
F*01:03:01:02 (1943G, 3189C, +2TG)	F*01:03	18	0.0459
F*01:01:01:01 (-1TG)	F*01:01	18	0.0459
F*01:01:01:05 (1943G, +2TG)	F*01:01	17	0.0434
F*01:01:03:03	F*01:01	16	0.0408
F*01:01:02:03 (1943G, 2208C, +1TG)	F*01:01	16	0.0408
F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, new3205G, +3TG)	F*01:01	15	0.0383
F*01:01:02:04 (3189T)	F*01:01	12	0.0306
F*01:01:02:03 (1943G, +1TG)	F*01:01	7	0.0179
F*01:04 (1943G)	F*01:04	5	0.0128
F*01:01:01:08 (-222G, 2698G, -2TG)	F*01:01	5	0.0128
F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, 2811delG, new3205G, +3TG)	F*01:01	5	0.0128
F*01:01:03:01 (new394G)	F*01:01	3	0.0077
F*01:01:02:03 (1943G, +3TG)	F*01:01	3	0.0077
F*01:01:01:05 (1943G, +4TG)	F*01:01	3	0.0077
F*01:01:02:03 (1943G, 2208C, +2TG)	F*01:01	3	0.0077
F*01:03:01:02 (new1378T, 1943G, 3189C, +1TG)	F*01:03	2	0.0051
F*01:01:01:08 (new1779A)	F*01:01	2	0.0051
F*01:01:01:01 (new1313G)	F*01:01	2	0.0051
F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, 2811delG, 3205G)	F*01:01	1	0.0026
F*01:01:02:03 (new1570A, 1943G, 2208C, +1TG)	Unnamed	1	0.0026
F*01:01:02:02 (1943G, 2208C, +1TG)	F*01:01	1	0.0026
F*01:03:01:01 (-1TG)	F*01:03	1	0.0026
F*01:03:01:01 (new1225C)	F*01:03	1	0.0026
F*01:01:01:01 (+1TG)	F*01:01	1	0.0026
F*01:03:01:01 (+1TG)	F*01:03	1	0.0026
F*01:01:03:03 (-1TG)	F*01:01	1	0.0026

* Haplotype names were given according to the closest known HLA-F allele followed by the differences/divergences that were observed for this given haplotype. The segment that is considered by the IMGT/HLA database starts on nucleotide -300 up to nucleotide +3250. The word "new" refers to variable sites that are not currently described at the IMGT/HLA database, version 3.21.0. -1TG, +1TG, +2TG, +3TG and +4TG refers to the number of dinucleotides TG on the *HLA-F* STR at position 3096 (IMGT/HLA) that are different from the mentioned closest known allele. The mutation at position 1570 generates a new HLA-F molecule, referred were as "unnamed" because only the IMGT/HLA database can proper name this new allele and its consequent product. Haplotypes are listed in order of frequency.

Table 3: List of *HLA-F* 5' upstream gene haplotypes found in Brazil.

Chr6 position	29690694	29690715	29690741	29690838	Frequency	Occurrences
SNPid	rs3998799	rs114795943	rs17875377	rs17875378	(2n = 392)	(2n = 392)
IMGT/HLA	-547	-526	-500	-403		
F*upstream-A	C	T	G	G	0.5893	231
F*upstream-B	G	T	G	G	0.3750	147
F*upstream-C	C	T	A	A	0.0179	7
F*upstream-D	C	A	G	G	0.0102	4
F*upstream-E	C	T	A	G	0.0077	3

Minor alleles are marked in shades of gray. Since the HLA-F promoter is not yet complete characterized, haplotypes were named as Upstream-A to -E, according to their frequencies.

Table 4: List of *HLA-F* variable sites at the 3' downstream gene segment found in Brazil.

Chr6 position	29694570	29694680	29694681	29694777	29695176	29695180	29695199	29695305	29695412	29695466		
SNPID	rs3734813	rs3734814	rs3734815	rs17875384	rs371631016	rs1419696	rs143726784	rs2523405	rs17875386	rs2735060		
IMGT/HLA	3332	3442	3443	3539	3938	3942	3961	4067	4174	4228	Occurrences	Frequency (2n=392)
F*UTR-01	A	A	A	G	TTA	G	A	G	C	A	163	0.4158
F*UTR-02	A	A	A	G	TTA	A	A	T	C	T	93	0.2372
F*UTR-03	G	C	T	G	TTA	G	A	T	C	T	60	0.1531
F*UTR-04	A	A	A	G	TTA	G	A	T	C	T	43	0.1097
F*UTR-05	A	A	A	A	TTA	G	A	T	C	T	20	0.0510
F*UTR-06	A	A	A	G	TTA	G	G	G	C	A	6	0.0153
F*UTR-07	G	C	T	G	TTA	G	A	T	T	T	3	0.0077
F*UTR-08	A	A	A	G	T	G	A	G	C	A	2	0.0051
F*UTR-09	G	C	T	G	TTA	G	A	T	C	A	1	0.0026
F*UTR-10	A	A	A	G	TTA	G	A	G	C	T	1	0.0026

Minor alleles are marked in shades of gray. Since the HLA-F 3'UTR is not yet complete characterized, haplotypes were named as UTR-01 to -10.

Table 5: *HLA-F* extended haplotypes considering the segment between nucleotide -556 and +4252, considering the Adenine of the first translated ATG as +1

Occurrences	Freq	5' near gene haplotype	HLA-F haplotype	3' UTR or 3' near gene
56	0.1429	F*upstream-B	F*01:01:01:01	F*UTR-01
13	0.0332	F*upstream-B	F*01:01:01:01 (-1TG)	F*UTR -01
5	0.0128	F*upstream-B	F*01:01:01:01 (-1TG)	F*UTR -06
2	0.0051	F*upstream-B	F*01:01:01:01 (new1313G)	F*UTR -01
1	0.0026	F*upstream-B	F*01:01:01:01 (+1TG)	F*UTR -01
1	0.0026	F*upstream-B	F*01:01:01:01	F*UTR -06
17	0.0434	F*upstream-A	F*01:01:01:05 (1943G, +2TG)	F*UTR -01
3	0.0077	F*upstream-A	F*01:01:01:05 (1943G, +4TG)	F*UTR -01
66	0.1684	F*upstream-B	F*01:01:01:08	F*UTR -01
2	0.0051	F*upstream-D	F*01:01:01:08 (-222G, 2698G, -2TG)	F*UTR -01
2	0.0051	F*upstream-D	F*01:01:01:08 (-222G, 2698G, -2TG)	F*UTR -08
2	0.0051	F*upstream-B	F*01:01:01:08 (new1779A)	F*UTR -01
1	0.0026	F*upstream-B	F*01:01:01:08	F*UTR -10
1	0.0026	F*upstream-A	F*01:01:01:08 (-222G, 2698G, -2TG)	F*UTR -01
15	0.0383	F*upstream-A	F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, new3205G, +3TG)	F*UTR -03
5	0.0128	F*upstream-A	F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, 2811delG, new3205G, +3TG)	F*UTR -03
1	0.0026	F*upstream-A	F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, 2811delG, new3205G)	F*UTR -03
1	0.0026	F*upstream-A	F*01:01:02:02 (1943G, 2208C, +1TG)	F*UTR -03
13	0.0332	F*upstream-A	F*01:01:02:03 (1943G, 2208C, +1TG)	F*UTR -03
7	0.0179	F*upstream-C	F*01:01:02:03 (1943G, +1TG)	F*UTR -03
3	0.0077	F*upstream-E	F*01:01:02:03 (1943G, +3TG)	F*UTR -03

3	0.0077	F*upstream-A	F*01:01:02:03 (1943G, 2208C, +1TG)	UTR-07
2	0.0051	F*upstream-A	F*01:01:02:03 (1943G, 2208C, +2TG)	UTR-04
1	0.0026	Upstream-A	F*01:01:02:03 (1943G, 2208C, +2TG)	UTR-09
1	0.0026	Upstream-A	F*01:01:02:03 (new1570A, 1943G, 2208C, +1TG)	UTR-04
12	0.0306	Upstream-A	F*01:01:02:04 (3189T)	UTR-04
68	0.1735	Upstream-A	F*01:01:03:01	UTR-02
3	0.0077	Upstream-A	F*01:01:03:01 (new394G)	UTR-02
16	0.0408	Upstream-A	F*01:01:03:03	UTR-02
1	0.0026	Upstream-A	F*01:01:03:03 (-1TG)	UTR-02
40	0.1020	Upstream-A	F*01:03:01:01	UTR-01
1	0.0026	Upstream-A	F*01:03:01:01 (-1TG)	UTR-01
1	0.0026	Upstream-A	F*01:03:01:01 (+1TG)	UTR-01
1	0.0026	Upstream-A	F*01:03:01:01 (new1225C)	UTR-01
18	0.0459	Upstream-A	F*01:03:01:02 (1943G, 3189C, +2TG)	UTR-05
2	0.0051	Upstream-A	F*01:03:01:02 (new1378T, 1943G, 3189C, +1TG)	UTR-05
5	0.0128	Upstream-A	F*01:04 (1943G, 3189C)	UTR-02

Haplotypes named following a specific IMGT/HLA allele are marked in shades of gray.

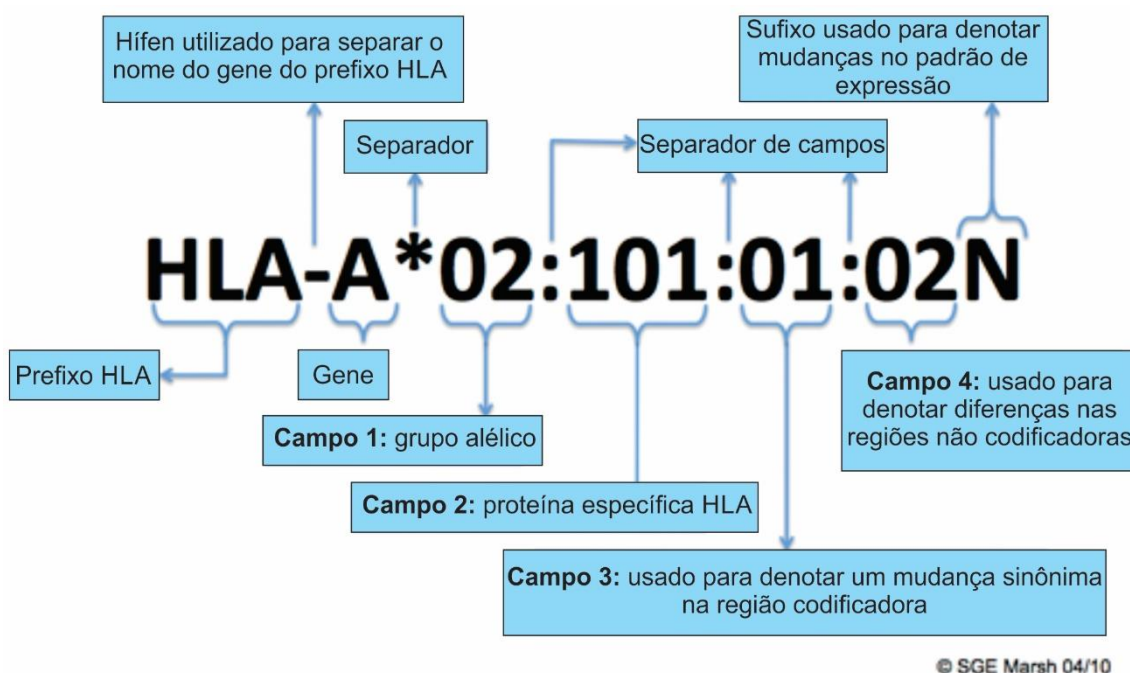
Conclusão

Conclusão

Conclui-se que, a variabilidade nucleotídica e haplotípica encontrada no presente estudo mostrou-se muito superior à já descrita na literatura e embora a maioria dos haplótipos encontrados sejam compatíveis com sequências já conhecidas, algumas sequências de *HLA-F* encontradas no Brasil caracterizam novos haplótipos, porém, a maioria dessas novas sequências apresentam mutações sinônimas ou intrônicas, desta forma esses novos haplótipos codificam as mesmas duas moléculas HLA-F encontradas mundialmente (F*01:01 e F*01:03). Independentemente da função desempenhada pela molécula HLA-F, seja estabilização de moléculas de classe I em conformação aberta, seja na reciclagem de moléculas de classe I, a conservação encontrada neste gene sugere que este papel é crítico para a fisiologia celular. Muitos estudos ainda precisam ser realizados para se entender o *HLA-F* funcionalmente. Contudo, estudos como este, que descrevem a variabilidade genética e proteica do *HLA-F*, fornecem subsídios para o entendimento da biologia desse gene e sua dinâmica no processo de regulação da resposta imunitária.

Anexos

Anexo 1. Denotação do significado de cada campo na nomenclatura dos alelos pelo IMGT/HLA. Fonte: <<http://hla.alleles.org/nomenclature/naming.html>>, acesso em 24 de set. de 2015, modificado pelo autor.



Nomenclatura	Indicação
<i>HLA</i>	A região e o prefixo que indicam um gene do complexo HLA.
<i>HLA-F</i>	Um loco HLA em particular, i.e., F.
<i>HLA-F*01</i>	Um grupo de alelos que codifica a proteína 01 do gene HLA-E.
<i>HLA-F*01:01</i>	Um alelo específico HLA.
<i>HLA-F*01:01:02</i>	Um alelo que difere por uma mutação sinônima na região codificadora do gene <i>HLA-E*01:01:01</i> .
<i>HLA-F*01:01:01:02</i>	Um alelo que contém uma mutação fora da região codificadora que o difere do alelo <i>HLA-E*01:01:01:01</i>
<i>HLA-A*24:09N</i>	Um alelo que codifica para uma proteína nula e, portanto, um alelo não expresso.