



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

Renato Vidal Buttura

**ESTUDO DA VARIABILIDADE DA REGIÃO PROMOTORA
DO GENE *HLA-F* EM AMOSTRAS BRASILEIRAS DO
ESTADO DE SÃO PAULO**

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

Orientador: Prof. Dr. Erick da Cruz Castelli

**Botucatu
2016**

Renato Vidal Buttura

**ESTUDO DA VARIABILIDADE DA REGIÃO PROMOTORA
DO GENE *HLA-F* EM AMOSTRAS BRASILEIRAS DO
ESTADO DE SÃO PAULO**

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

Orientador: Prof.Dr. Erick da Cruz Castelli

Botucatu
2016

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

Buttura, Renato Vidal.

Estudo da variabilidade da região promotora do gene
HLA-F em amostras brasileiras do estado de São Paulo /
Renato Vidal Buttura. - Botucatu, 2016

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

Orientador: Erick da Cruz Castelli

Capes: 40105008

1. Antígenos de histocompatibilidade HLA. 2. Patologia.
3. Polimorfismo (Genética). 4. Regiões promotoras
(Genética).

Palavras-chave: HLA-F; MHC; Região promotora;
Sequenciamento de nova geração; Variabilidade.

Dedicatória

*Aos meus pais, José Carlos Buttura e Rita de
Cássia Vidal Buttura e à minha namorada, Jaqueline
Ramalho, dedico...*

Epígrafe

“Nunca troque o que mais quer na vida pelo que mais quer no momento. Momentos passam e a vida continua...”

Renato Russo

Agradecimientos

Agradeço primeiramente a Deus pela minha vida e por tudo o que, com suor, consegui conquistar.

Aos meus pais por sempre acreditarem em mim e por serem sempre o meu alicerce, estando comigo em todos os momentos de minha vida.

À minha amiga e namorada Jaqueline pela paciência, pelo carinho, companheirismo e pelos ensinamentos que recebi todo esse período.

Aos meus amigos que aqui conquistei, em especial a Prof^a. Dra. Carla Adriene da Silva, pelo grande carinho que recebi desde o momento em que cheguei à cidade de Botucatu.

Ao meu orientador Prof. Dr. Erick da Cruz Castelli, primeiramente pela oportunidade que me deste, pelos ensinamentos e pela paciência que sempre teve comigo.

Aos meus queridos companheiros de laboratório: Iane, Thálitta, Michelle, pelo grande aprendizado e companheirismo durante esses dois anos.

À CAPES por fornecer todo o apoio financeiro e a bolsa de mestrado. E a todos aqueles que, diretamente ou indiretamente, contribuíram para a realização deste trabalho.

A todos um muito obrigado!

Resumo

Resumo

O *HLA-F* é um gene de classe I não clássico do Complexo de Antígenos Leucocitários Humanos (HLA). A molécula HLA-F parece apresentar atividade tolerogênica e de imunorregulação, uma vez que, assim como os outros genes não clássicos (*HLA-G* e *HLA-E*), apresenta um limitado polimorfismo e relativa conservação entre espécies. Além disso, estudos indicam que a HLA-F pode estabilizar outras moléculas HLA em conformação aberta, influenciando, portanto, a apresentação de antígenos. Pouco se sabe sobre a biologia do HLA-F e o perfil de variabilidade deste gene em populações mundiais. O objetivo deste trabalho foi avaliar a variabilidade e estrutura de haplótipos da região promotora distal do gene *HLA-F*, em 204 amostras brasileiras do estado de São Paulo, utilizando sequenciamento de nova geração. Foram encontrados 15 pontos de variação, dos quais a maioria possui frequência superior a 15% para o alelo alternativo. Esses pontos de variação estão arranjados em pelo menos 6 haplótipos distintos, no entanto mais de 50% das sequências de promotor HLA-F encontradas nesta amostragem correspondem a um único haplótipo. Foi detectado um elevado desequilíbrio de ligação entre a região do promotor distal e a região codificadora do gene *HLA-F*, demonstrado pela presença de apenas poucos haplótipos estendidos diferentes. A conservação da região promotora distal do *HLA-F* pode estar relacionada com o perfil de ligação de fatores de transcrição.

Abstract

Abstract

The *HLA-F* locus is a non classical class I gene of the Human Leukocyte Antigen Complex. The HLA-F molecule might present tolerogenic and immunoregulatory features, mainly because it presents the same characteristics of other non classical genes, such as *HLA-G* and *HLA-E*, i.e., limited polymorphism and conservation among species. Moreover, recent studies do indicate that HLA-F may stabilize other HLA molecules presenting open conformations, thus it might influence antigen presentation. Little is known about the *HLA-F* biology and the pattern of variation and haplotypes of this gene in worldwide populations. This study aims to evaluate the variability and haplotype structure of the *HLA-F* distal promoter segment in 204 Brazilian samples from the state of São Paulo, using next generation sequencing approaches. We detected 15 variable sites, most of them presenting frequencies greater than 15% for the alternative allele. These polymorphisms were arranged into 6 different promoter haplotypes, 50% of all *HLA-F* promoters correspond to a single haplotype. There is a strong linkage disequilibrium between the distal promoter and the coding region, as demonstrated by the presence of only few extended haplotypes. The conservation of the distal promoter region of the *HLA-F* gene may be associated with the transcription factors binding profile.

Sumário

Capítulo I	14
Revisão da Literatura	15
Referências.....	24
Objetivos	31
Capítulo II	32
Artigo científico a ser enviado ao periódico HLA.....	33
1. Introdução	35
2. Materiais e Métodos.....	36
3. Resultados.....	39
4. Discussão	48
Referências	51
Conclusão	57

Capítulo 1

Revisão da Literatura

Descrito pela primeira vez em camundongos, no ano de 1937 por Peter Gorer, o Complexo Principal de Histocompatibilidade ou MHC (do inglês *Major Histocompatibility Complex*) foi uma descoberta de grande valia para a ciência, em especial sua influência na compatibilidade de tecidos transplantados (1). Os genes desse complexo foram associados em sua maioria com a regulação das respostas imunitárias, principalmente atuando na distinção do que é próprio ou não próprio do organismo (histocompatibilidade), a qual lhe rendeu o nome. Em humanos, o MHC só foi descrito na década de 50 por Jean B. Dausset, o que lhe rendeu o Prêmio Nobel na década de 80 (2). Seus estudos foram baseados em experimentos sorológicos de pacientes submetidos a transplantes, nos quais foram observadas algumas diferenças antigênicas entre os leucócitos dos mesmos, moléculas de superfície de membrana que receberam a denominação de Antígenos Leucocitários Humanos ou HLA (do inglês *Human Leukocyte Antigen*) (2).

O surgimento do MHC nos vertebrados foi algo crucial no mecanismo de reconhecimento e diferenciação do que é próprio ou não ao indivíduo e na montagem das repostas imunes adaptativa ou adquirida direcionadas aos antígenos não próprios. As moléculas do MHC, ou HLA, são glicoproteínas de membrana especializadas na apresentação de antígenos, que interagem com receptores de células T (TCR) influenciando sua ativação e desenvolvimento de suas funções efetoras (3–5). Seja com a produção de citocinas, e/ou ativação de linfócitos B específicos e/ou citotoxicidade, a combinação das moléculas HLA apresentando peptídeos ao TCR de linfócitos é o que caracteriza a alta especificidade e diversidade da resposta imune adquirida (3–7).

O MHC humano está localizado no braço curto do cromossomo 6 (6p21.3) e contém cerca de 220 genes dispostos em aproximadamente 4 megabases de DNA (8). Na década de 70, esse complexo foi dividido didaticamente em três classes ou regiões, denominadas classe I, II e III (3). Os genes de classe III codificam algumas citocinas, moléculas do sistema complemento, entre outras, que participam da montagem e regulação da resposta imune. Os genes de classe II codificam moléculas HLA que apresentam antígenos extracelulares aos linfócitos T CD4⁺ e são expressas em células apresentadoras de antígenos, como as células dendríticas, além de macrófagos e linfócitos B (3).

Já os genes do MHC de classe I estão relacionados com a apresentação antigênica aos linfócitos T CD8⁺ e modulação das respostas imunitárias. Esses genes estão divididos em duas subclasses. A primeira, denominada Ia ou genes clássicos de classe I, compreende os

genes *HLA-A*, *HLA-B* e *HLA-C*, que participam diretamente da apresentação antigênica. A segunda, denominada Ib ou genes não clássicos de classe I, compreende os genes *HLA-G*, *HLA-E* e *HLA-F*, responsáveis principalmente por modulação das respostas imunitárias (**Figura 1**).

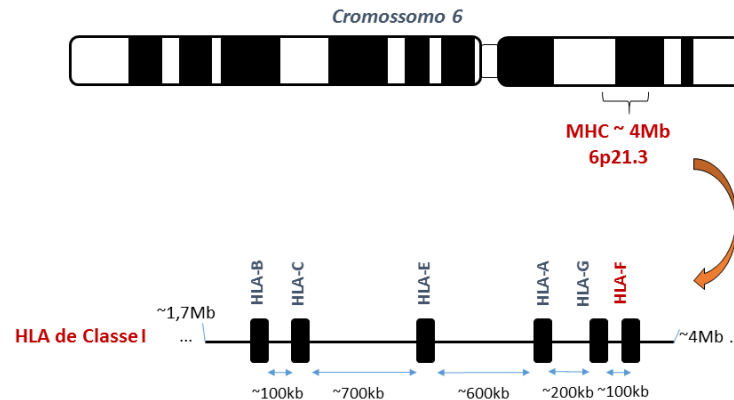


Figura 1: Genes HLA de classe I clássicos e não clássicos com as posições relativas entre os genes e da região de classe I dentro do MHC. Imagem própria baseada em JANEWAY *et al.* e http://labs.icb.ufmg.br/lbcd/prodabi5/grupos/hugo_deise_liza/pagina/mapa.gif.

Os genes *HLA-A*, *HLA-B* e *HLA-C* (Ia) são expressos em todas as células nucleadas, sendo responsáveis pela apresentação de antígenos aos linfócitos T CD8⁺ ou T citotóxico. Estes genes são considerados os mais polimórficos do genoma humano. Diferentemente, os genes de classe Ib possuem distribuição limitada a certos tecidos ou expressão reduzida, e função distinta, além de um polimorfismo limitado quando comparado aos genes de classe Ia, como pode-se observar na **Tabela 1** (3,4,6,7,9).

A alta variabilidade encontrada nos genes MHC (10), tendo como parâmetro os genes clássicos de classe I, é provavelmente resultado de um longo processo evolutivo envolvendo seleção natural, deriva genética, fluxo gênico e principalmente, mutação, que modificaram as frequências alélicas desses genes acarretando em uma alta diversidade intra e inter-populacional (**Tabela 1**). A alta variabilidade característica desses genes é decorrente de um processo no qual a manutenção de polimorfismos é privilegiada. Isso é extremamente importante do ponto de vista funcional, já que essa alta variabilidade de moléculas HLA significaria uma maior diversidade de patógenos reconhecidos por uma população (11). Os mecanismos subjacentes a esse tipo de seleção são a chamada “vantagem do heterozigoto” e a “seleção dependente de frequência”. Neste contexto, os heterozigotos seriam capazes de apresentar antígenos de um número maior de microrganismos devido aos dois alelos HLA diferentes, o que é uma vantagem adaptativa em relação aos homozigotos (12). Já a seleção

dependente de frequência baseia-se na ideia de que alelos raros teriam uma vantagem seletiva prioritária de modo a aumentar sua frequência. Dessa forma, esses alelos mais raros seriam melhores que os mais comuns, pois as moléculas por eles codificadas poderiam apresentar antígenos diferentes de um determinado microrganismo em relação aos alelos comuns, conferindo desta forma uma vantagem a estes alelos (13).

Tabela 1. Número de alelos e proteínas identificados para os genes de classe I

HLA Classe I	Clássicos			Não Clássicos		
Gene	<i>HLA-A</i>	<i>HLA-B</i>	<i>HLA-C</i>	<i>HLA-E</i>	<i>HLA-F</i>	<i>HLA-G</i>
Alelos	3399	4242	2950	21	22	53
Proteínas	2396	3131	2089	8	4	18
Alelos Nulos*	157	134	109	1	0	2

Fonte: The International Immunogenetics Database - IMGT/HLA, release 3.24.0.1

* Alelos que não se expressam ou produzem proteínas truncadas

Já nos genes HLA de classe I não clássicos, que são bem menos variáveis que os genes HLA clássicos, o perfil de seleção atuante é diferente. Por exemplo, no gene *HLA-G* temos evidências de seleção purificadora atuando na região codificadora do gene, ou seja, limitando alterações no produto proteico final e mantendo certos alelos com alta frequência, e seleção balanceadora nas regiões regulatórias, mantendo a presença de haplótipos altamente divergentes e de alta frequência (14–16). O gene *HLA-F*, por sua vez, parece ser muito conservado, pois 82,45% das sequências encontradas em uma amostra de Brasileiros do estado de São Paulo codificam a mesma proteína (17).

De acordo com o banco de dados internacional que regulamenta as sequências conhecidas para estes genes (IPD-IMGT/HLA, versão 3.24.0.1, de 04/maio/2016), atualmente são reconhecidos cerca de 10.591 alelos para os genes clássicos de classe I, enquanto que para os não clássicos são reconhecidos apenas 96 alelos (**Tabela 1**). Por estarem próximos aos genes mais polimórficos do genoma humano (**Figura 1**) e também por possuírem um baixo grau de polimorfismo, os genes não clássicos são interessantes do ponto de vista evolutivo, pelo fato de serem pouco variáveis e conservados entre humanos e entre outros primatas (18,19).

Moléculas dessa subclasse, assim como os genes clássicos, podem também apresentar antígenos, porém, sua função principal está mais direcionada para imunomodulação das respostas imunitárias por meio da interação com receptores inibitórios presentes em células

NK e T. Os genes não clássicos são de grande importância em situações como a gestação, onde há necessidade de modulação do sistema imunitário materno. Ainda, estes genes podem influenciar a susceptibilidade e progressão de condições patológicas, tais como infecções, tumores e doenças autoimunes (6,20–24).

Apesar de ser a região mais variável do genoma humano, o sistema HLA também possui uma alta similaridade entre seus genes. Acredita-se que os genes HLA tenham surgido por meio de inúmeras duplicações e conversões ao longo da evolução dos vertebrados (25,26). Dessa forma, a estrutura dos genes e as moléculas que codificam são semelhantes entre si.

Tanto os genes HLA de classe I clássicos (*HLA-A*, *HLA-B* e *HLA-C*) como os não clássicos (*HLA-G*, *HLA-E* e *HLA-F*) (3) codificam moléculas que formam um heterodímero constituído por uma cadeia alfa e um domínio β 2-microglobulina, esse codificado no cromossomo 15. A cadeia alfa é composta por cinco domínios, sendo eles: α 1 e α 2 (que formam a fenda de ligação do peptídeo), α 3 (o qual liga-se não covalentemente a β 2-microglobulina), um domínio transmembrana e um citoplasmático, responsável pelo envio de sinais intracelulares (**Figura 2**) (3,27).

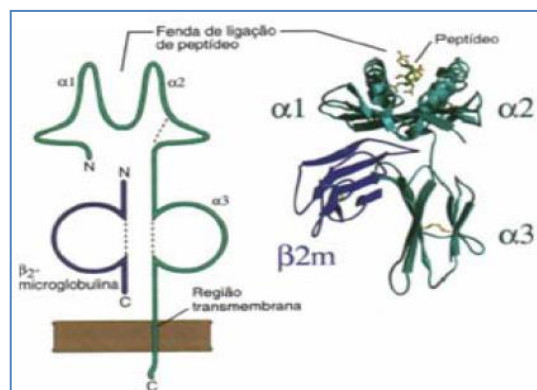


Figura 2: Estrutura linear e tridimensional da molécula HLA de Classe I (27)

As moléculas de HLA de classe I são montadas no retículo endoplasmático, onde há a união não covalente de suas cadeias alfa e beta. Moléculas como p57 do retículo endoplasmático, tapasina (TAP), chaperonas, calnexina e calreticulina são de grande importância para a montagem da molécula de MHC de classe I (27).

Considerando os genes de classe I clássicos, uma mesma molécula de HLA tem a capacidade de apresentar diversos peptídeos antigênicos diferentes, variando conforme o padrão da fenda de ligação da molécula de HLA (28,29). Por terem uma menor variabilidade, os genes HLA de classe I não clássicos codificam moléculas que reconhecem um menor

repertório de peptídeos, considerando que sua principal função é a modulação das respostas imunitárias (7,22–24).

Os peptídeos que se ligam às moléculas HLA de classe I são processados de antígenos maiores por meio de proteassomos no citosol (27,30) e são posteriormente destinados ao retículo endoplasmático com o auxílio da TAP (*Transporter associated with Antigen Processing*) (31). Chaperonas atuam como proteínas acessórias para que o processo, desde a montagem da molécula de HLA de classe I até o acoplamento da mesma com um peptídeo, possa ser estabelecido (31). Após o acoplamento do peptídeo, a molécula de HLA de classe I é transportada para o complexo de Golgi, onde ocorre sua glicosilação e finalmente apresentação na superfície da célula (27,32) (**Figura 3**).

Ao ser apresentado na superfície celular, a molécula de HLA de classe I interage com receptores TCR (*T cell receptor*) de células T CD8⁺ e com uma variedade de receptores (ativadores ou inibidores) das células T e NK. A interação com receptores, somado a outros sinais nos linfócitos, pode resultar na ativação dessas células e desencadear uma resposta específica contra este antígeno, acarretando na destruição das células infectadas ou de células anormais, ou inibir a ação dessas células (3,33).

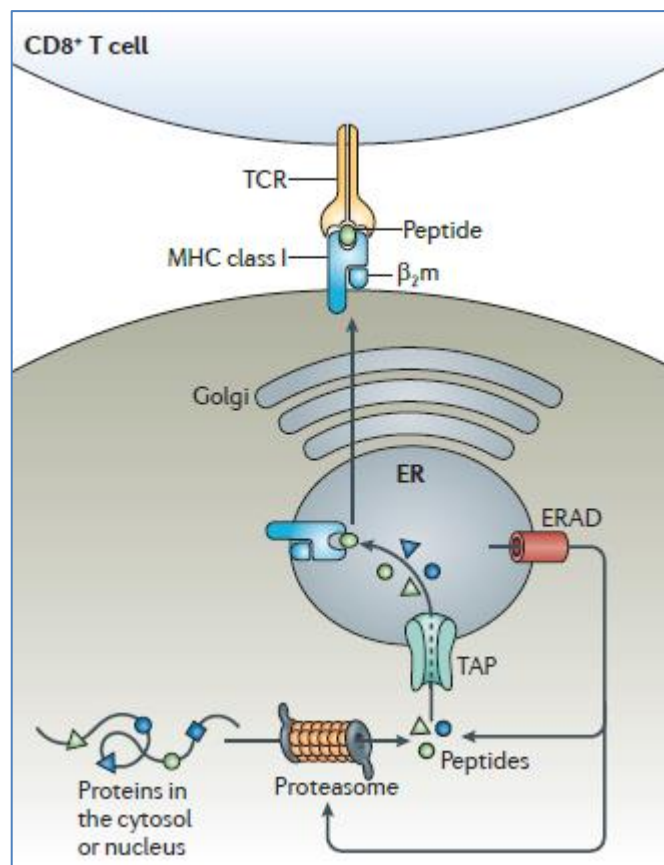


Figura 3: Processamento e apresentação antigênica pelas moléculas MHC de classe I (30).

O *HLA-F*, um dos genes HLA de classe I não clássico, está presente na porção mais telomérica do MHC (**Figura 1**). De acordo com as anotações do genoma humano de referências (versões hg19 ou hg38), o gene *HLA-F* apresenta 7 éxons e 6 íntrons, com uma região 5' não traduzida de aproximadamente 123 bases no éxon 1 e uma região 3' não traduzida que se inicia no éxon 7 (**Figura 4**).

No banco de dados IPD-IMGT/HLA (versão 3.24.0.1) estão descritos 22 alelos que codificam 4 proteínas diferentes de HLA-F. Estes alelos foram caracterizados entre as posições -300 e +3250, considerando a Adenina do primeiro ATG traduzido como nucleotídeo +1. Portanto, o *HLA-F* é o segundo gene menos variável entre os loci de classe I do ponto de vista da presença de alelos diferentes, no entanto, é o mais conservado quando se considera o número de diferentes proteínas codificadas por estes alelos (**Tabela 1**).

Pela escassez de estudos acerca da variabilidade de HLA-F em populações mundiais, o *HLA-F* poderia apresentar uma maior variabilidade do que a conhecida atualmente. O estudo em populações com índices altos de miscigenação, como a brasileira, poderia mostrar se a baixa variabilidade desse gene é uma característica encontrada mundialmente. Historicamente, o Brasil foi palco de diversas colônias providas de várias partes do mundo, como por exemplo a população europeia (principalmente portugueses, espanhóis, italianos e alemães), africana, árabe e orientais, além da população indígena de origem. Por conta desta mistura, a diversidade genética no Brasil seria muito relevante (34). Porém, de fato, a conservação da HLA-F em nível proteico foi demonstrada em um estudo anteriormente realizado no Brasil, onde, apesar dos 25 alelos novos de *HLA-F* aqui detectados, a maioria dos alelos codificava uma mesma proteína HLA-F (35).

Além da variabilidade, a função da molécula HLA-F ainda é pouco explorada. Acredita-se que a molécula de HLA-F possua funções tolerogênicas, semelhante ao já descrito para moléculas HLA-E e HLA-G (7,22–24,36). Muitos trabalhos descrevem a HLA-F como uma molécula expressa em células B, células T quiescentes, monócitos, tecidos linfóides, timo e baço, fígado fetal, pele e bexiga (18,22,36–41). Além disso, a molécula de HLA-F tem sido descrita em células do trofoblasto extraviloso a partir do segundo mês de gestação e em linfócitos ativados.

No entanto, sua função ainda não é bem definida, assim como sua expressão na superfície celular é questionada: é detectado o mRNA de HLA-F, mas a molécula parece ficar retida intracelularmente, sendo expressa na superfície em condições específicas (21,37,38,42–45).

Por ser um gene muito conservado entre humanos e primatas (incluindo bonobo,

gorila, orangotango e chimpanzé), acredita-se que o gene *HLA-F* e a molécula que ele codifica podem ter funções críticas e de grande importância na resposta imunológica (18,37). Trabalhos realizados com esse gene sugerem que proteínas provenientes do *HLA-F* possam estar envolvidas em conjunto com as moléculas HLA-G e HLA-E na regulação da atividade das células T e NK (46).

Já foi descrito também que as moléculas de HLA-F podem formar tetrâmeros e interagir com os receptores inibitórios ILT2 e ILT4 das células T (22,47), assim como a molécula HLA-G (23,48). Isso sugere uma função imunomodulatória da molécula de HLA-F semelhante ao HLA-G, porém, diferente de HLA-G e HLA-E, ainda não é conhecido nenhum peptídeo que se associe à molécula de HLA-F de forma estável (47), o que contribui para sua questionada expressão na superfície celular.

Outra hipótese é que a função de HLA-F é preferencialmente intracelular, estando associado com outras moléculas HLA de classe I na forma de “moléculas abertas” (*open conformers*) (39,46,49). Essas moléculas em conformação aberta não estariam complexadas com peptídeos e por isso apresentariam instabilidade. A função de HLA-F então estaria envolvida na estabilização dessas moléculas ou mesmo serviria como mecanismo de reciclagem de outras moléculas HLA de classe I, além de poder estar associada a um mecanismo de apresentação cruzada de antígenos exógenos (39,46,49). Além disso, esses complexos de HLA-F com moléculas abertas poderiam também interagir com receptores KIR nas células NK influenciando sua função. Essa hipótese deve-se ao fato de evidências da localização comum das moléculas HLA-F e *open conformers* (39,46,49). Dessa forma, embora ainda não definida, a função da molécula HLA-F é sugerida como envolvida com a imunomodulação, mas parece participar de mecanismos diferentes das outras moléculas HLA não clássicas (39,49).

Alguns trabalhos associam a expressão da molécula de HLA-F em tumores com um mau prognóstico do paciente (50,51). Esse fato se dá pela associação desta molécula com possíveis funções de imunossupressão do sistema imunológico, o que acarretaria para essas células neoplásicas a capacidade de escape das células efetivas do sistema imune, auxiliando em sua disseminação (50–53).

Apesar de haver algumas associações com doenças, como carcinoma hepático (53), carcinoma de esôfago (52) e carcinoma mamário (51), e até mesmo sugestões sobre seu papel na resposta imunológica, o *HLA-F* é um gene pouco explorado quanto a sua variabilidade em populações mundiais. Os estudos publicados foram feitos na China, onde há um baixo índice de miscigenação (54), e em uma amostragem brasileira da cidade de Curitiba, estado do

Paraná (41), os quais se descrevem os principais alelos presentes nessas amostras. No entanto, apenas polimorfismos que caracterizam os principais alelos foram considerados. Apenas outros dois estudos descreveram a variabilidade estendida do gene *HLA-F*, com metodologias diferentes (35,36).

Em geral os trabalhos sobre variabilidade de gene HLA concentram sua exploração e discussão nas regiões codificadoras, que irão diretamente influenciar a estrutura da proteína codificada e, portanto, quais peptídeos essas moléculas poderão ligar e com quais receptores essas moléculas poderão interagir, modificando diretamente sua função. No banco de dados IPD-IMGT/HLA estão descritos alelos para cada um dos genes HLA e regulamenta principalmente a variabilidade da região codificadora, informando variações existentes em éxons e íntrons. Aproximadamente 300 pares de base antes do primeiro ATG traduzido (**Figura 4**) também são considerados, além de segmentos da região 3' não traduzida dos genes. Essas sequências não fazem parte da região codificadora, mas são as regiões regulatórias.

A região 3' não traduzida se inicia após o códon de parada da tradução e estende-se até o fim do mRNA. Este segmento pode influenciar a estabilidade do mRNA e sua disponibilidade, por ser alvo para ligação de pequenos RNAs não codificadores, os chamados microRNA. Essa ligação consiste em um mecanismo de regulação da expressão gênica pós transcricional o qual impede a tradução e/ou degrada o mRNA, regulando negativamente a expressão do gene (55,56).

O segmento de DNA que antecede o códon de iniciação da tradução (ATG), cujo tamanho pode variar de gene para gene, é chamada de região promotora do gene. Nessa região estão presentes os elementos que regulam a transcrição de um gene, ou seja, é o local onde o aparato basal de transcrição se monta para iniciar a produção de mRNA. Na região promotora cerne e proximal estão os sítios mais conhecidos de ligação a fatores transcricionais, como TATA box e CCAAT box, onde se liga o aparato basal de transcrição (43). Nessas e na região promotora distal estão sítios chamados de estimuladores e silenciadores, nos quais se ligam fatores de transcrição, proteínas ativadoras ou repressoras da transcrição, que vão influenciar diretamente a expressão do gene (43).

Os primeiros estudos que descreveram a estrutura do gene *HLA-F* mostraram que sua região promotora proximal se assemelha as dos genes HLA de classe I clássicos, fazendo com que a expressão da molécula de HLA-F também fosse similar às das moléculas clássicas (8,42). No caso do *HLA-F*, nessa região promotora proximal há elementos responsivos ao NF- κ B, IRF1 (IFN- γ) e CIITA (43,57).

Diferente dos genes HLA de classe I clássicos, mas semelhante ao *HLA-G*, na promotora cerne do *HLA-F* o elemento denominado Inr (“*Initiator*”) não é conservado, o que pode explicar a relativa limitação da expressão da molécula a alguns tecidos (43). Assim, tendo elementos regulatórios semelhantes na região promotora, tanto comparados aos genes HLA de classe I, como ao *HLA-G*, a expressão de *HLA-F* apresentaria um perfil intermediário (42,58,59).

Como citado anteriormente, a região promotora distal, mesmo que esteja mais distante do início da transcrição do gene, pode conter sítios ligação a proteínas regulatórias que acentuam ou silenciam a expressão gênica. Para o *HLA-F*, quando em 1990 GERAGHTY *et al.* o descreveram primeiramente essa promotora distal, foi encontrado sequências altamente repetitivas configurando a presença de um microssatélite (**Figura 4**). Além da influência estrutural na molécula de DNA, que por si interfere no nível de expressão de um gene, não foi descartado a possível função regulatória dessa região repetitiva (42).

Portanto, além da região codificadora, cuja variabilidade pode influenciar a função devido a alterações no produto proteico final, a quantidade de mRNA transcrito e a quantidade de proteína traduzida são influenciadas pelas regiões regulatórias. Desta forma, o presente trabalho objetivou avaliar a diversidade da região promotora distal do gene *HLA-F* em amostras brasileiras provenientes do estado de São Paulo (indicado na **Figura 4**). Como será abordado mais adiante, a região de microssatélite não foi considerada. Acredita-se que polimorfismos funcionais do *HLA-F* estejam localizados nas regiões regulatórias do gene, influenciando diretamente sua expressão.

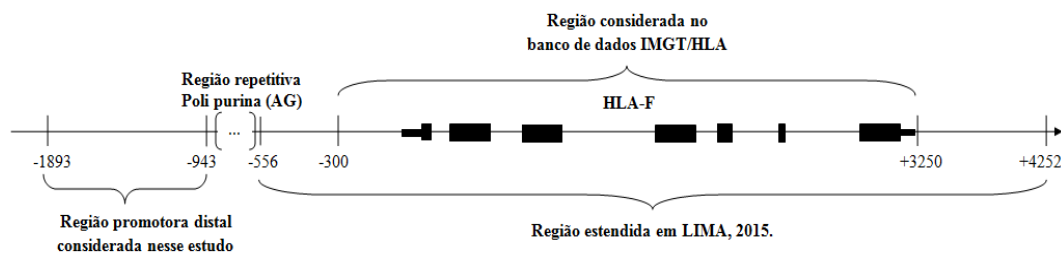


Figura 4 - Estrutura e regiões consideradas para o estudo da variabilidade de gene *HLA-F*. As posições são relativas a primeira adenina do ATG transcrito considerado como nucleotídeo +1. Imagem própria.

Referências

1. Gorer PA. The genetic and antigenic basis of tumour transplantation. *J Pathol Bacteriol* [Internet]. John Wiley & Sons, Ltd.; maio de 1937 [citado 24 de junho de 2016];44(3):691–7. Recuperado de: <http://doi.wiley.com/10.1002/path.1700440313>
2. Dausset J. Iso-Leuko-Antibodies. *Vox Sang* [Internet]. Blackwell Publishing Ltd; janeiro de 1958 [citado 24 de junho de 2016];3(1):40–1. Recuperado de: <http://doi.wiley.com/10.1111/j.1423-0410.1958.tb03559.x>
3. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med* [Internet]. 7 de setembro de 2000 [citado 24 de junho de 2016];343(10):702–9. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/10974135>
4. Klein J, Sato A. The HLA system. Second of two parts. *N Engl J Med* [Internet]. 14 de setembro de 2000 [citado 24 de junho de 2016];343(11):782–6. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/10984567>
5. Li XC, Raghavan M. Structure and function of major histocompatibility complex (MHC) class I antigens. *Curr Opin Organ Transpl*. 2010;15(4):499–504.
6. Goldberg AC, Rizzo LV. MHC structure and function – antigen presentation. Part 1. *Einstein (São Paulo, Brazil)*. 2015;13(1):153–6.
7. Goldberg AC, Rizzo LV. MHC structure and function - antigen presentation. Part 2. *Einstein (Sao Paulo)*. 2015;13(1):157–62.
8. Beck S, Geraghty D, Inoko H, Rowen I. Complete sequence and gene map of a human major histocompatibility complex. *Nature*. 1999;401(October):921–3.
9. Fainardi E, Rizzo R, Melchiorri L, Vaghi L, Castellazzi M, Marzola A, et al. Presence of detectable levels of soluble HLA-G molecules in CSF of relapsing-remitting multiple sclerosis: relationship with CSF soluble HLA-I and IL-10 concentrations and MRI findings. *J Neuroimmunol* [Internet]. setembro de 2003 [citado 24 de junho de 2016];142(1-2):149–58. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/14512174>
10. Markov P V, Pybus OG. Evolution and Diversity of the Human Leukocyte Antigen(HLA). *Evol Med public Heal* [Internet]. Oxford University Press; 2015 [citado 4 de julho de 2016];2015(1):1. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25577609>
11. Hedrick PW. Balancing selection and MHC. *Genetica* [Internet]. [citado 24 de junho de 2016];104(3):207–14. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/10386384>
12. Spencer HG, Marks RW. The maintenance of single-locus polymorphism. I. Numerical

- studies of a viability selection model. *Genetics* [Internet]. outubro de 1988 [citado 24 de junho de 2016];120(2):605–13. Recuperado de:
<http://www.ncbi.nlm.nih.gov/pubmed/3197961>
13. Bodmer WF. Evolutionary significance of the HL-A system. *Nature* [Internet]. 19 de maio de 1972 [citado 24 de junho de 2016];237(5351):139–45 passim. Recuperado de:
<http://www.ncbi.nlm.nih.gov/pubmed/4113158>
 14. Tan Z, Shon AM, Ober C. Evidence of balancing selection at the HLA-G promoter region. *Hum Mol Genet* [Internet]. 1 de dezembro de 2005 [citado 24 de junho de 2016];14(23):3619–28. Recuperado de:
<http://www.ncbi.nlm.nih.gov/pubmed/16236759>
 15. Sabbagh A, Luisi P, Castelli EC, Gineau L, Courtin D, Milet J, et al. Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: balancing selection influencing genetic diversity. *Genes Immun* [Internet]. março de 2014 [citado 24 de junho de 2016];15(2):95–106. Recuperado de:
<http://www.ncbi.nlm.nih.gov/pubmed/24352166>
 16. Mendes-Junior CT, Castelli EC, Meyer D, Simões 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 Immun* [Internet]. dezembro de 2013 [citado 24 de junho de 2016];14(8):518–26. Recuperado de:
<http://www.ncbi.nlm.nih.gov/pubmed/24089150>
 17. Castelli EC, Mendes-Junior CT, Sabbagh A, Porto IOP, Garcia A, Ramalho J, et al. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. *Hum Immunol* [Internet]. dezembro de 2015 [citado 24 de junho de 2016];76(12):945–53. Recuperado de:
<http://www.ncbi.nlm.nih.gov/pubmed/26187162>
 18. Rojo R, Castro MJ, Martinez-Laso J, Serrano-Vela JI, Morales P, Moscoso J, et al. MHC-F DNA sequences in bonobo, gorilla and orangutan. *Tissue Antigens* [Internet]. outubro de 2005 [citado 24 de junho de 2016];66(4):277–83. Recuperado de:
<http://www.ncbi.nlm.nih.gov/pubmed/16185322>
 19. Pratheek BM, Nayak TK, Sahoo SS, Mohanty PK, Chattopadhyay S, Chakraborty NG, et al. Mammalian non-classical major histocompatibility complex I and its receptors: Important contexts of gene, evolution, and immunity. *Indian J Hum Genet* [Internet]. abril de 2014 [citado 24 de junho de 2016];20(2):129–41. Recuperado de:
<http://www.ncbi.nlm.nih.gov/pubmed/25400340>

20. Kovats S, Main EK, Librach C, Stubblebine M, Fisher SJ, DeMars R. A class I antigen, HLA-G, expressed in human trophoblasts. *Science* [Internet]. 13 de abril de 1990 [citado 24 de junho de 2016];248(4952):220–3. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/2326636>
21. Ishitani A, Sageshima N, Hatake K. The involvement of HLA-E and -F in pregnancy. *J Reprod Immunol* [Internet]. abril de 2006 [citado 24 de junho de 2016];69(2):101–13. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16487601>
22. Djuricic S, Hviid TVF. HLA Class Ib Molecules and Immune Cells in Pregnancy and Preeclampsia. *Front Immunol* [Internet]. 2014 [citado 24 de junho de 2016];5:652. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25566263>
23. Amiot L, Vu N, Samson M, Amiot L, Vu N, Samson M. Immunomodulatory Properties of HLA-G in Infectious Diseases. *J Immunol Res* [Internet]. Hindawi Publishing Corporation; 2014 [citado 24 de junho de 2016];2014:1–14. Recuperado de: <http://www.hindawi.com/journals/jir/2014/298569/>
24. Morandi F, Pistoia V. Interactions between HLA-G and HLA-E in Physiological and Pathological Conditions. *Front Immunol* [Internet]. Frontiers Media SA; 2014 [citado 24 de junho de 2016];5:394. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25202308>
25. Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, et al. Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc Natl Acad Sci U S A* [Internet]. 20 de agosto de 1996 [citado 24 de junho de 2016];93(17):9096–101. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/8799160>
26. Kasahara M, Nakaya J, Satta Y, Takahata N. Chromosomal duplication and the emergence of the adaptive immune system. *Trends Genet* [Internet]. março de 1997 [citado 24 de junho de 2016];13(3):90–2. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9066265>
27. Abbas AK, Lichtman AH, Pillai S. *Imunologia Celular e Molecular*. 7º ed. Elsevier; 2012. 560 p.
28. Vyas JM, Van der Veen AG, Ploegh HL. The known unknowns of antigen processing and presentation. *Nat Rev Immunol* [Internet]. agosto de 2008 [citado 24 de junho de 2016];8(8):607–18. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/18641646>
29. Mester G, Hoffmann V, Stevanović S. Insights into MHC class I antigen processing

- gained from large-scale analysis of class I ligands. *Cell Mol Life Sci* [Internet]. maio de 2011 [citado 24 de junho de 2016];68(9):1521–32. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/21387142>
30. Neefjes J, Jongstra ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol*. Nature Publishing Group; novembro de 2011;11(12):823.
 31. Wearsch PA, Cresswell P. The quality control of MHC class I peptide loading. *Curr Opin Cell Biol* [Internet]. dezembro de 2008 [citado 24 de junho de 2016];20(6):624–31. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/18926908>
 32. Saric T, Chang S-C, Hattori A, York IA, Markant S, Rock KL, et al. An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nat Immunol* [Internet]. dezembro de 2002 [citado 24 de junho de 2016];3(12):1169–76. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/12436109>
 33. Gao GF, Willcox BE, Wyer JR, Boulter JM, O’Callaghan CA, Maenaka K, et al. Classical and nonclassical class I major histocompatibility complex molecules exhibit subtle conformational differences that affect binding to CD8alphaalpha. *J Biol Chem* [Internet]. 19 de maio de 2000 [citado 24 de junho de 2016];275(20):15232–8. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/10809759>
 34. SEYFERTH G. COLONIZAÇÃO, IMIGRAÇÃO E A QUESTÃO RACIAL NO BRASIL. *Rev USP* [Internet]. 30 de maio de 2002 [citado 24 de junho de 2016];0(53):117. Recuperado de: <http://www.revistas.usp.br/revusp/article/view/33192>
 35. Lima THA [UNESP]. Estudo da variabilidade da região codificadora e 3’ não traduzida do gene HLA-F no Brasil. *Aleph*. Universidade Estadual Paulista (UNESP); 2015;56 f.
 36. Pyo C-W, Williams LM, Moore Y, Hyodo H, Li SS, Zhao LP, et al. HLA-E, HLA-F, and HLA-G polymorphism: genomic sequence defines haplotype structure and variation spanning the nonclassical class I genes. *Immunogenetics* [Internet]. maio de 2006 [citado 24 de junho de 2016];58(4):241–51. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16570139>
 37. Moscoso J, Serrano-Vela JI, Pacheco R, Arnaiz-Villena A. HLA-G, -E and -F: allelism, function and evolution. *Transpl Immunol* [Internet]. dezembro de 2006 [citado 24 de junho de 2016];17(1):61–4. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/17157219>
 38. Lee N, Ishitani A, Geraghty DE. HLA-F is a surface marker on activated lymphocytes. *Eur J Immunol* [Internet]. agosto de 2010 [citado 24 de junho de 2016];40(8):2308–18.

- Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/20865824>
39. Goodridge JP, Burian A, Lee N, Geraghty DE. HLA-F and MHC class I open conformers are ligands for NK cell Ig-like receptors. *J Immunol* [Internet]. 1 de outubro de 2013 [citado 24 de junho de 2016];191(7):3553–62. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/24018270>
 40. Kochan G, Escors D, Breckpot K, Guerrero-Setas D. Role of non-classical MHC class I molecules in cancer immunosuppression. *Oncoimmunology* [Internet]. Landes Bioscience; 1 de novembro de 2013 [citado 24 de junho de 2016];2(11):e26491. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/24482746>
 41. Manvailer LFS, Wowk PF, Mattar SB, da Siva JS, da Graça Bicalho M, Roxo VMMS. HLA-F polymorphisms in a Euro-Brazilian population from Southern Brazil. *Tissue Antigens* [Internet]. dezembro de 2014 [citado 24 de junho de 2016];84(6):554–9. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25413105>
 42. Geraghty DE, Wei XH, Orr HT, Koller BH. Human leukocyte antigen F (HLA-F). An expressed HLA gene composed of a class I coding sequence linked to a novel transcribed repetitive element. *J Exp Med* [Internet]. 1 de janeiro de 1990 [citado 24 de junho de 2016];171(1):1–18. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/1688605>
 43. Howcroft TK, Singer DS. Expression of Nonclassical MHC Class Ib Genes: Comparison of Regulatory Elements. *Immunol Res* [Internet]. Humana Press; 2003 [citado 24 de junho de 2016];27(1):1–30. Recuperado de: <http://link.springer.com/10.1385/IR:27:1:1>
 44. Shobu T, Sageshima N, Tokui H, Omura M, Saito K, Nagatsuka Y, et al. The surface expression of HLA-F on decidual trophoblasts increases from mid to term gestation. *J Reprod Immunol* [Internet]. dezembro de 2006 [citado 24 de junho de 2016];72(1-2):18–32. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16806485>
 45. Boyle LH, Gillingham AK, Munro S, Trowsdale J. Selective export of HLA-F by its cytoplasmic tail. *J Immunol* [Internet]. 1 de junho de 2006 [citado 24 de junho de 2016];176(11):6464–72. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16709803>
 46. Goodridge JP, Burian A, Lee N, Geraghty DE. HLA-F complex without peptide binds to MHC class I protein in the open conformer form. *J Immunol* [Internet]. 1 de junho de 2010 [citado 24 de junho de 2016];184(11):6199–208. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/20483783>

47. Allan DSJ, Lepin EJM, Braud VM, O'Callaghan CA, McMichael AJ. Tetrameric complexes of HLA-E, HLA-F, and HLA-G. *J Immunol Methods* [Internet]. 1 de outubro de 2002 [citado 24 de junho de 2016];268(1):43–50. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/12213342>
48. Hofmeister V, Weiss EH. HLA-G modulates immune responses by diverse receptor interactions. *Semin Cancer Biol* [Internet]. outubro de 2003 [citado 24 de junho de 2016];13(5):317–23. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/14708711>
49. Wainwright SD, Biro PA, Holmes CH. HLA-F is a predominantly empty, intracellular, TAP-associated MHC class Ib protein with a restricted expression pattern. *J Immunol* [Internet]. 1 de janeiro de 2000 [citado 24 de junho de 2016];164(1):319–28. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/10605026>
50. Lin A, Zhang X, Ruan Y-Y, Wang Q, Zhou W-J, Yan W-H. HLA-F expression is a prognostic factor in patients with non-small-cell lung cancer. *Lung Cancer* [Internet]. dezembro de 2011 [citado 24 de junho de 2016];74(3):504–9. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/21561677>
51. Ishigami S, Arigami T, Setoyama T, Okumura H, Sasaki K, Uchikado Y, et al. Clinical-pathological implication of human leukocyte antigen-F-positive gastric adenocarcinoma. *J Surg Res* [Internet]. outubro de 2013 [citado 24 de junho de 2016];184(2):802–6. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/23706560>
52. Zhang X, Lin A, Zhang J-G, Bao W-G, Xu D-P, Ruan Y-Y, 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* [Internet]. 1 de janeiro de 2013 [citado 24 de junho de 2016];132(1):82–9. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/22544725>
53. Xu Y, Han H, Zhang F, Lv S, Li Z, Fang Z. Lesion human leukocyte antigen-F expression is associated with a poor prognosis in patients with hepatocellular carcinoma. *Oncol Lett* [Internet]. janeiro de 2015 [citado 24 de junho de 2016];9(1):300–4. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25435979>
54. Pan FH, Liu XX, Tian W. Characterization of HLA-F polymorphism in four distinct populations in Mainland China. *Int J Immunogenet* [Internet]. outubro de 2013 [citado 24 de junho de 2016];40(5):369–76. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/23551590>
55. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* [Internet]. 23 de janeiro de 2004 [citado 24 de junho de 2016];116(2):281–97.

- Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/14744438>
56. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* [Internet]. 23 de janeiro de 2009 [citado 24 de junho de 2016];136(2):215–33. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/19167326>
 57. Gobin SJ, van den Elsen PJ. Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F, and HLA-G. *Hum Immunol* [Internet]. novembro de 2000 [citado 24 de junho de 2016];61(11):1102–7. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/11137213>
 58. Wei XH, Orr HT. Differential expression of HLA-E, HLA-F, and HLA-G transcripts in human tissue. *Hum Immunol* [Internet]. outubro de 1990 [citado 24 de junho de 2016];29(2):131–42. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/2249951>
 59. O’Callaghan CA, Bell JL. Structure and function of the human MHC class Ib molecules HLA-E, HLA-F and HLA-G. *Immunol Rev* [Internet]. junho de 1998 [citado 24 de junho de 2016];163:129–38. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9700506>
 60. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. Nature Publishing Group; setembro de 2015;526(7571):75–81.

Objetivos

O objetivo deste trabalho foi avaliar a variabilidade genética inerente à região promotora distal do gene *HLA-F* por meio de sequenciamento de nova geração, buscando a identificação de polimorfismos funcionais nesta região.

Os objetivos específicos foram:

- Caracterizar a variabilidade da região promotora distal do gene *HLA-F* em uma amostra brasileira do estado de São Paulo.
- Caracterizar o perfil de desequilíbrio de ligação ao longo da região promotora, os principais haplótipos existentes e a relação entre essas sequências com haplótipos da região codificadora;
- Elaborar um banco de dados de variação no gene *HLA-F* para futuros estudos de associação com doenças.

Capítulo II

Artigo científico a ser enviado ao periódico HLA

**VARIABILIDADE E ESTRUTURA DE HAPLÓTIPOS DA REGIÃO PROMOTORA
DISTAL DO GENE *HLA-F* EM AMOSTRAS BRASILEIRAS DO ESTADO DE SÃO
PAULO**

Renato Vidal Buttura¹, Thálitta Hetamaro Ayala Lima¹, Eduardo Antônio Donadi², Luciana Caricati Veiga-Castelli², Celso Teixeira Mendes-Junior³, Erick da Cruz Castelli^{1,4*}

¹ Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPLEX), Sector 5, School of Medicine, Unesp – Univ. Estadual Paulista, Botucatu, State of São Paulo, Brazil.

² School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, State of São Paulo, Brazil.

³ Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto - SP, Brazil.

⁴ Pathology Department, School of Medicine, Unesp – Univ. Estadual Paulista, Botucatu, State of São Paulo, Brazil.

* Contact: Erick C. Castelli

Departamento de Patologia, Faculdade de Medicina

Unesp, Botucatu – SP, CEP 18618970, Brazil

castelli@fmb.unesp.br, +55 14 3880 1696

RESUMO

O locus *HLA-F* é um gene de classe I não clássico do Complexo de Antígenos Leucocitários Humanos. Acredita-se que a molécula HLA-F possua funções tolerogênicas e imunomodulatórias, principalmente porque apresenta as mesmas características de outros genes não clássicos, como *HLA-G* e *HLA-E*, i.e., limitado polimorfismo e conservação entre as espécies. Além disso, estudos recentes indicaram que HLA-F pode estabilizar outras moléculas HLA que se apresentam na forma de moléculas abertas, assim HLA-F pode influenciar na apresentação de antígenos. Pouco se sabe sobre a biologia do gene *HLA-F* e o padrão de variabilidade e haplótipos nas populações mundiais. Esse estudo teve como objetivo avaliar a variabilidade e estrutura de haplótipos do segmento da região promotora distal do *HLA-F* em 204 amostras brasileiras do estado de São Paulo, por meio de sequenciamento de nova geração. Foram detectados 15 pontos de variação, sendo que a maioria deles apresentarem frequência maior que 15% para o alelo alternativo. Esses polimorfismos foram arranjados em 6 haplótipos de promotor distal diferentes, sendo que mais de 50% de todas as promotoras de *HLA-F* corresponde a um único haplótipo. Há um forte desequilíbrio de ligação entre a região promotora distal e região codificadora como demonstrado pela presença de poucos haplótipos estendidos. A conservação da região promotora distal do gene *HLA-F* pode estar associada com o padrão de ligação de fatores de transcrição.

Palavras-chave: HLA-F, MHC, variabilidade, polimorfismos, haplótipos, Sequenciamento de Nova Geração; NGS, região promotora.

1. Introdução

O locus *HLA-F* é um gene HLA de classe I não clássico e está presente na porção mais telomérica do Complexo Principal de Histocompatibilidade (MHC), localizado no braço curto do cromossomo 6. A molécula que codifica pode apresentar atividade tolerogênica, semelhante ao que já foi descrito para moléculas HLA-E e HLA-G (1–5). Ainda, o gene *HLA-F* é muito conservado entre humanos e primatas (incluindo bonobo, gorila, orangotango e chimpanzé), e por isso, acredita-se que essa molécula tenha funções críticas e de grande importância na resposta imunológica (6,7).

Além de pertencerem a mesma classe dos HLA de classe I não clássicos e apresentarem de forma semelhante um perfil de variabilidade mais conservado, uma das observações que sustentam as atividades imunomodulatórias da molécula de HLA-F é que esta parece formar tetrâmeros e interagir com os receptores inibitórios ILT2 e ILT4 das células T (2,8), assim como a molécula HLA-G (4). Apesar dessa observação, ainda não se sabe quais peptídeos se associam à molécula de HLA-F de forma estável, diferente das outras moléculas não clássicas (8).

Além dessa função tolerogênica não ser bem definida, a expressão de HLA-F na superfície celular também é questionada. Muitos trabalhos descrevem a HLA-F como uma molécula expressa em células B, células T quiescentes, monócitos, tecidos linfóides, timo e baço, fígado fetal, pele e bexiga (1,2,6,7,9–13). Além disso, a molécula de HLA-F tem sido descrita em células do trofoblasto extraviloso a partir do segundo mês de gestação e em linfócitos ativados. Porém, em muitos casos, é detectado seu mRNA no citoplasma, mas a molécula parece ficar retida intracelularmente, sendo expressa na superfície em condições específicas (7,10,14–17).

Assim, a HLA-F pode estar relacionada também a uma função preferencialmente intracelular, associado com outras moléculas HLA de classe I na forma de “moléculas abertas” (*open conformers*) (11,18,19). Essas moléculas em conformação aberta não estariam complexadas com peptídeos e por isso apresentariam instabilidade. A função de HLA-F então estaria envolvida na estabilização dessas moléculas, ou mesmo serviria como mecanismo de reciclagem de outras moléculas HLA de classe I, além de poder estar associada a um mecanismo de apresentação cruzada de antígenos exógenos (11,18,19). Além disso, esses complexos de HLA-F com moléculas abertas poderiam também interagir com receptores KIR nas células NK influenciando sua função. Essa hipótese deve-se ao fato de evidências da localização comum das moléculas HLA-F e *open conformers* (11,18,19). Com isso, embora a

função exata da molécula HLA-F ainda não esteja definida, é sugerido que também possua um possível papel imunomodulatório, porém participando de mecanismos diferentes das outras moléculas HLA não clássicas.

Apesar do MHC humano ser a região mais variável do genoma, o gene *HLA-F* apresenta limitado polimorfismo. No banco de dados do IPD-IMGT/HLA (versão 3.24.0.1), que caracteriza as sequências de alelos HLA já conhecidas, estão descritos 22 alelos que codificam 4 proteínas HLA-F diferentes. No entanto, devido à escassez de estudos acerca da variabilidade de HLA-F em populações mundiais, o *HLA-F* poderia apresentar uma maior variabilidade do que a conhecida atualmente. O estudo de populações miscigenadas, como a brasileira, poderia mostrar se a conservação desse é uma característica mundial. Historicamente, o Brasil é o resultado da miscigenação de diferentes povos, incluindo europeus (principalmente portugueses, espanhóis, italianos e alemães), africanos, árabes e orientais, além da população indígena de origem. Por conta desta mistura, a diversidade genética no Brasil seria muito relevante (20). No entanto, o limitado polimorfismo de HLA-F, considerando os tipos de alelos que codificam diferentes moléculas, foi demonstrado em um estudo anteriormente realizado no Brasil, onde, apesar dos 25 alelos novos de *HLA-F* aqui detectados, a maioria dos alelos codificava a mesma proteína final (21).

Em geral os trabalhos sobre variabilidade de genes HLA concentram-se na exploração das regiões codificadoras, que irão diretamente influenciar a estrutura da proteína codificada e, portanto, quais peptídeos essas moléculas poderão ligar e com quais receptores essas moléculas poderão interagir, modificando diretamente sua função. No entanto, não se conhece a variabilidade inerente das regiões regulatórias do HLA-F, e variações nestes segmentos poderiam modificar sobremaneira o perfil de expressão do gene.

Desta forma, o presente trabalho teve como proposta avaliar a diversidade da região promotora distal do gene *HLA-F* em amostras brasileiras provenientes do estado de São Paulo. Acredita-se que polimorfismos funcionais do *HLA-F* estejam localizados nas regiões regulatórias do gene, influenciando diretamente sua expressão.

2. Materiais e Métodos

2.1 Definição Das Amostras

No presente estudo foi extraído e avaliado o material genético de 204 amostras de sangue de indivíduos de estado de São Paulo, provenientes principalmente das cidades de Botucatu e Ribeirão Preto. Foram coletados 10ml de sangue periférico e o material genético

foi posteriormente extraído por meio de um procedimento *salting out*. Todos os indivíduos assinaram o Termo de Consentimento Livre e Esclarecido para uso e armazenamento das amostras. Esse trabalho foi fruto de um projeto maior, cujo objetivo foi caracterizar a variabilidade dos genes HLA de classe I. Esse projeto foi aprovado pelo Comitê de Ética (Protocolo CEP 24157413.7.0000.5411).

2.2 Produção dos *Amplicons* e Sequenciamento

A metodologia utilizada neste estudo insere-se dentro de um projeto maior, cujo objetivo foi sequenciar todos os genes MHC de classe I nas mesmas amostras (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-G*, *HLA-E* e *HLA-F*).

O gene *HLA-F* foi amplificado isoladamente em um único fragmento pela Reação em Cadeia da Polimerase (PCR), compreendendo sua região regulatória 5' ou promotora, codificadora e 3' não traduzida, produzindo um fragmento de aproximadamente 6245pb. Os iniciadores utilizados foram desenhados localmente (HFPR.F1: 5'-GGAGAGAACTCAGGTGGC-3' e HFUT.R1: 5'-CCACTAAACACCCAGCCCAT-3'). O protocolo de amplificação está descrito adiante: volume final de 50µL contendo 0,80X de solução tampão, 1,5mM de MgCl₂, 0,20mM de cada dNTP (Invitrogen - Carlsbad, CA USA), 15pmol de cada iniciador, 1,5 unidade de DNA Polimerase (Long PCR Enzyme Mix, Thermo Fisher Scientific Inc., Waltham, MA, USA) e aproximadamente 100ng de amostra de DNA. O perfil de ciclagem utilizado foi 94° por 5 minutos, seguido por 10 ciclos de 94° por 30 segundos, 63° por 45 segundos e 68° por 7 minutos, e 22 ciclos de 94° por 30 segundos, 61° por 45 segundos e 68° por 7 minutos, finalizando com 68° por 4 minutos. Os produtos de amplificação foram testados por eletroforese em gel de agarose 1% corado por Gelred® (Biotium™, Hayward, CA, USA).

As bibliotecas de fragmentos para o sequenciamento de nova geração foram criadas fazendo-se combinações de genes HLA de classe I de um mesmo indivíduo, utilizando-se os kits Nextera® XT Index e Nextera® DNA Sample Preparation (Illumina Inc.). Os sequenciamentos do tipo leitura pareada das duas extremidades de cada fragmento foram realizados por meio da plataforma *MiSeq Personal Sequencer* (Illumina, Inc., San Diego, CA USA).

2.3 Mapeamento das Sequências e Busca de SNPs

As sequências de cada amostra (formato fastq) foram obtidas pelo software *MiSeq Reporter* (Illumina Inc.). Para o mapeamento das sequências foi utilizado o programa *hla-*

mapper, desenvolvido localmente (22) para alinhar sequências de genes HLA com genoma humano de referência, na versão hg19. Nesse processo, as sequências foram filtradas de acordo com o gene de interesse (*HLA-F*), considerando sequências já descritas no banco de dados IPD-IMGT/HLA e no projeto *1000 Genomas*. Após esse processo, foram gerados mapeamentos para a região promotora do gene *HLA-F* (formato BAM), considerando a região promotora distal, entre as posições 29689348 (-1893) a 29690298 (-943), referente ao genoma humano hg19. O segmento do gene *HLA-F* entre as posições 29690685 (-556) e 29695490 (+4252), que compreende a região promotora proximal, codificadora e 3' não traduzida, foi analisada em um estudo anterior do grupo (21). O segmento entre as posições -942 e -557 não foi incluído no estudo por se tratar de uma região de baixa complexidade, em parte devido à presença de microssatélite no qual se repetem bases purínicas (adenina e guanina). Por ser uma extensa região repetitiva, o mapeamento de sequências neste segmento não é confiável.

Os mapeamentos da região promotora distal considerada (arquivo .BAM) foram realinhados nas regiões contendo variações do tipo *indel*, usando a rotina “*Indel Realigner*” do pacote GATK (23,24). Os alinhamentos foram visualizados no Integrative Genomics Viewer (IGV) (25).

Em seguida, os alinhamentos de cada amostra passaram pela rotina de “*BaseRecalibrator*” e “*PrintReads*” do GATK para melhorar a qualidade dos genótipos e minimizar possíveis erros. Logo depois, foi criada uma lista dos pontos de variação encontrados em todas as amostras por meio da rotina “*HaplotypeCaller*” do GATK, originando arquivo .VCF. Genótipos duvidosos foram tratados utilizando um programa local denominado *vcfx*, ou seja, genótipos homozigotos inferidos em regiões de baixa cobertura e/ou genótipos com grandes desproporções entre os alelos inferidos, foram modificados inserindo-se alelos desconhecidos para posterior imputação. A fase dos pontos de variação foi definida utilizando a rotina “*ReadBackedPhasing*” do GATK. Os haplótipos finais foram obtidos conforme especificado no item “Inferência de Haplótipos”, que será descrito a seguir.

2.4 Inferência de haplótipos

Embora a metodologia proposta permita a detecção da fase dos pontos de variação (rotina *ReadBackedPhasing* do GATK), eventualmente, esta fase se perde devido à distância entre alguns desses sítios. O sequenciamento de nova geração sequencia diversos fragmentos da sequência de interesse, os quais são chamados *reads*. Esses *reads* são sequenciados em pares e abrangem um tamanho máximo. Pontos de variação no mesmo *read* estão certamente no mesmo cromossomo e, portanto, na mesma fase. No entanto, quando os pontos de variação

estão muito distantes, a ausência de *reads* que contenham esses pontos de variação impede uma análise direta da fase entre estes pontos. Este fenômeno ocorre principalmente com genes menos polimórficos como já observado em estudos prévios envolvendo os genes *HLA-G* e *HLA-E* (22). Desta forma, uma abordagem computacional foi realizada para inferir a fase dos pontos de variação nestes trechos.

Com o arquivo .VCF gerado pelo “*ReadBackedPhasing*” do GATK, as fases conhecidas foram usadas para complementar a abordagem computacional do método PHASE (26). Os haplótipos foram estimados dentro dessas regiões usando informações de pontos de variação vizinhos, de fase conhecida pelo sequenciamento, melhorando a qualidade da inferência de haplótipos.

De posse das informações dos pontos de variação e suas fases, a sequência completa de cada haplótipo de cada indivíduo foi montada utilizando o vcfx (www.castelli-lab.net), que usou como referência o genoma humano na versão hg19. Considerando o intervalo da região promotora distal e pela informação presente no arquivo .VCF, para cada haplótipo, os pontos de variação eram substituídos na sequência selecionada. As análises estatísticas de diversidade nucleotídica e gênica foram realizadas por meio do software Arlequin versão 3.5.1.2 (27). Os dados para a região promotora distal do gene *HLA-F* foram acrescentados aos dados das demais regiões do gene (trabalho anterior do grupo) (21). Além disso, com o intuito de avaliar o padrão de desequilíbrio de ligação (LD) entre os pontos de variação detectados, um *plot* de desequilíbrio de ligação (LD) foi gerado utilizando o software Haploview 4.2 (28).

Considerando que a região promotora pode influenciar o perfil de expressão de um gene por meio da ligação diferencial de fatores de transcrição em sítios polimórficos, as variações encontradas foram avaliadas quanto a sua influência na função da região promotora por meio do software HaploReg versão 4.1 (29), disponível online.

3. Resultados

Considerando a região promotora distal analisada do gene *HLA-F*, que compreende 950pb e está presente aproximadamente a 1,8kb anteriormente ao início de tradução do gene, e as 204 amostras analisadas, foram encontrados 15 pontos de variação (**Tabela 1**). Dez variantes apresentam frequência superior a 15% para o alelo alternativo. Das quatro variantes encontradas apenas em um único indivíduo (*singletons*), apenas 2 foram inseridas na análise de haplótipos (**Tabela 1**). Isto ocorreu por duas razões: (a) a grande distância entre estes pontos de variação um ponto heterozigoto adjacente não permite a análise direta pelo GATK

ReadBackedPhasing e (b) o método PHASE não é capaz de definir a fase de variantes que ocorreram apenas uma vez na amostra estudada.

Tabela 1. Lista dos pontos de variação detectados na região promotora distal do gene *HLA-F* em amostras brasileiras do estado de São Paulo

Posição no chr6 ^a	SNPid	Obs. ^b	Posição relativa IMGT/HLA ^c	Alelo de Referência ^d	Frequência dos alelos de referência (2n=408)	Alelo Alternativo	Frequência do alelo alternativo (2n=408)
29689423	rs144001783	A	-1818	G	0.9975	A	0.0025
29689532	rs9258169	.	-1709	T	0.8284	G	0.1716
29689590	rs9258170	.	-1651	T	0.8284	G	0.1716
29689631	rs17875375	.	-1610	T	0.9510	C	0.0490
29689683	rs78739406	.	-1558	A	0.8284	AG	0.1716
29689717	rs9258171	.	-1524	G	0.8284	T	0.1716
29689742	rs183540921	B	-1499	T	0.9975	G	0.0025
29689897	.	B	-1344	A	0.9975	G	0.0025
29689973	rs9258172	.	-1268	C	0.8284	T	0.1716
29690002	rs9258173	.	-1239	G	0.8284	T	0.1716
29690026	rs9258174	.	-1215	T	0.8284	C	0.1716
29690037	.	A	-1204	G	0.9975	A	0.0025
29690056	rs17875376	.	-1185	C	0.8015	T	0.1985
29690125	rs9258175	.	-1116	G	0.8284	A	0.1716
29690228	rs9258176	.	-1013	C	0.8284	G	0.1716

^a Posição no cromossomo 6 considerando o genoma humano de referência, versão hg19

^b Observações: A: variação encontrada em apenas um indivíduo (*singleton*), com haplótipo definido

B: variação encontrada em apenas um indivíduo (*singleton*), sem haplótipo definido^c Posição relativa ao banco de dados IPD-IMGT/HLA, considerando a Adenina do primeiro ATG traduzido como nucleotídeo +1

^d Alelo referência no genoma humano versão hg19.

Os 13 pontos de variação aqui descritos formam 6 haplótipos distintos, que foram nomeados de acordo com a frequência encontrada em nossa amostra, conforme mostrado na **Tabela 2**. Três haplótipos, aqui denominados de F*distal-A ao -C, representam aproximadamente 95% das sequências encontradas nesta amostra brasileira, sendo que apenas o haplótipo F*distal-A é responsável por mais de 57% das sequências. O haplótipo F*distal-A está relacionado aos alelos encontrados na referência usada, que neste caso são mais frequentes que os alternativos. A frequência somada dos haplótipos F*distal-A, F*distal-B, F*distal-D e F*distal-E, que apresentam apenas um ponto de variação diferente do primeiro, é de 82.59%. O haplótipo F*distal-B está associado à variante -1185T (MAF=19.85%), de alta frequência em nossa amostra. O haplótipo F*distal-C é o mais divergente de todos e está associado aos alelos alternativos de quase todos os pontos de variação encontrados (-1709G, -1651G, -1558AG, -1524T, -1268T, -1239T, -1215C, -1116A e -1013G), que estão em absoluto desequilíbrio de ligação (**Tabela 2, Figura 1**). Portanto, basicamente há dois perfis

de promotores distais em nossa amostra: haplótipos derivados de F*distal-A e o haplótipo mais divergente, F*distal-C.

Utilizando o software online HaploReg v4.1 (29), os 13 pontos de variação incluídos na análise de haplótipos foram analisados *in silico* quanto a sua influência na ligação de fatores de transcrição. A **Tabela 3** mostra, por ponto de variação, quais motivos de fatores de transcrição seriam afetados, sendo divididos em dois grupos, referência e alternativo, de acordo com a maior afinidade que o sítio possuiria com o alelo referência ou alelo alternativo, respectivamente. Como observado na **Tabela 3**, há diferença de afinidades a fatores de transcrição que se ligam às sequências quando da presença de determinados alelos. Essa diferença de afinidade de ligação (maior para a sequência que possui o alelo de referência ou o alelo alternativo) pode ser melhor visualizada no **Gráfico 1**, que mostra o valor de diferença de afinidade do motivo do fator de transcrição por sítio de variação, considerando os dados fornecidos pela análise do HaploReg (*Position Weight Matrix score* (29)). O cálculo da diferença de afinidade foi feito com base na seguinte equação: diferença = (score para alelo de referência) – (score para alelo alternativo), sendo que quando a diferença é maior que 0 a afinidade é maior para a sequência do alelo referência, e quando a diferença é menor que 0, a afinidade é maior para a sequência do alelo alternativo. Como pode ser melhor observado no **Gráfico 1**, a maioria dos fatores de transcrição elencados tem maior afinidade pela sequência do alelo de referência, porém há fatores com maior afinidade pela sequência do alelo alternativo. Assim, é possível que o perfil transcricional relacionado aos haplótipos F*distal-A (e derivados), caracterizados pela maioria dos alelos de referência, seja diferente do perfil transcricional relacionado ao haplótipo F*distal-C, principalmente quando consideramos que o F*distal-C está associado à maioria dos alelos alternativos descritos na **Tabela 3**.

Tabela 2. Lista dos haplótipos encontrados na região promotora distal do gene *HLA-F* em amostras brasileiras do estado de São Paulo, compreendendo os pontos de variação entre os nucleotídeos -1893 a -943, considerando nucleotídeo +1 como a Adenina do primeiro ATG traduzido.

Haplótipos ^a	Posições relativas de acordo com IMGT/HLA ^b													Frequência Amostras brasileiras (2n=408)
	-1818	-1709	-1651	-1610	-1558	-1524	-1268	-1239	-1215	-1204	-1185	-1116	-1013	
F*distal-A	G	T	T	T	A	G	C	G	T	G	C	G	C	0.5784
F*distal-B	G	T	T	T	A	G	C	G	T	G	T	G	C	0.1961
F*distal-C	G	G	G	T	AG	T	T	T	C	G	C	A	G	0.1716
F*distal-D	G	T	T	C	A	G	C	G	T	G	C	G	C	0.0490
F*distal-E	G	T	T	T	A	G	C	G	T	A	C	G	C	0.0025
F*distal-F	A	T	T	T	A	G	C	G	T	G	T	G	C	0.0025
Diversidade nucleotídica														0.0031 +/- 0.0018
Diversidade gênica														0.5966 +/- 0.0204

Os alelos alternativos estão marcados em cinza.

^a Haplótipos foram listados conforme frequência observada nas amostras da população brasileira

^b Posições calculadas considerando como nucleotídeo +1 a Adenina do primeiro ATG traduzido. As posições cromossômicas estão listadas na Tabela 1.

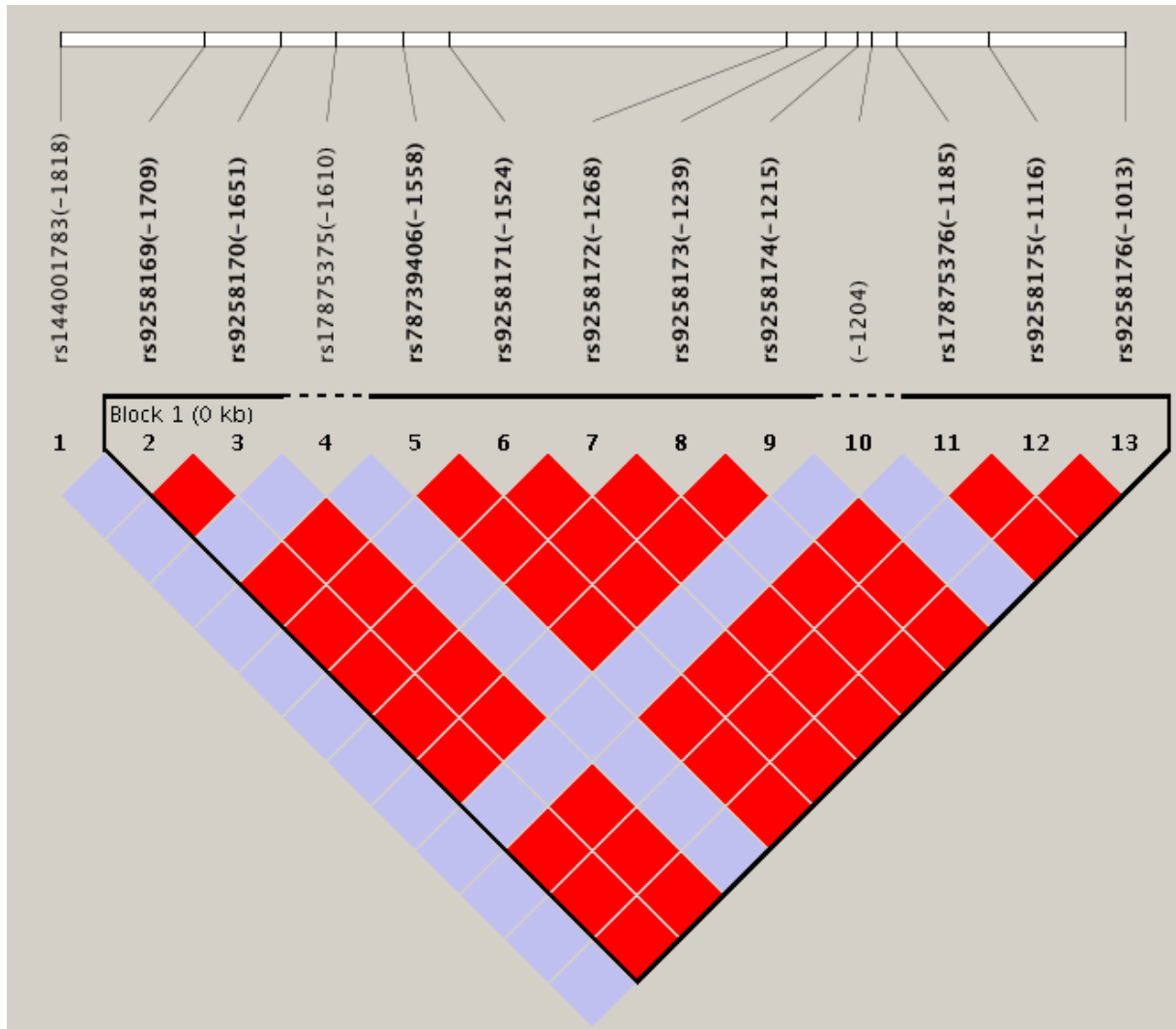


Figura 1. Mapa de desequilíbrio de ligação (LD) entre pares de pontos de variação na região promotora distal do gene *HLA-F*. Essa imagem foi gerada pelo Haploview 4.2 usando pontos de variação com frequência do alelo mínimo (MAF) $\geq 0,1\%$. Áreas em vermelho indicam forte LD [$\log$ de *odds ratio* ($LOD \geq 2$, $D' = 1$)] e áreas em azul/cinza indicam LD fraco ($LOD \leq 2$, $D' = 1$).

Foi possível inferir os haplótipos estendidos do gene *HLA-F* e avaliar a associação entre os haplótipos de promotor distal e os haplótipos de região codificadora do *HLA-F*. Para esta análise, consideramos apenas 196 amostras, que tiveram a região codificadora do *HLA-F* avaliada em estudo anterior (21). Neste contexto, encontramos 39 haplótipos estendidos (**Tabela 4.**), porém apenas 4 haplótipos apresentaram frequência superior a 10% e representaram em torno de 60% de todas as sequências nesta amostragem. Considerando-se os haplótipos estendidos, nota-se que o haplótipos F*distal-C está relacionado exclusivamente com alelos do grupo F*01:01:02. Os demais haplótipos promotores estão associados com alelos do grupo F*01:01:01 e F*01:03 (**Tabela 4.**). Esta associação restrita entre haplótipos promotores e codificadores pode ser explicada pelo elevado desequilíbrio de ligação existente ao longo de todo o segmento do gene *HLA-F*, como indicado pela **Figura 2.**

Tabela 3. Predição *in silico* de motivos de ligação a fatores de transcrição alterados pelos pontos de variação encontrados na região promotora distal do gene *HLA-F*.

Posição Chr6 (hg19)	SNPid	Posição relativa a IMGT/HLA ^a	Alelo de referência ^b	Alelo alternativo	Frequên- cia ^c	Motivos de fatores de transcrição com afinidade pelos diferentes alelos ^d	
						Referência	Alternativo
29689423	rs144001783	-1818	G	A	0.0025	EWSR-FLI1, Zf143_known1	TCF4_disc2
29689532	rs9258169	-1709	T	G	0.1716	Dbx1, Hoxa10, Hoxd10, TATA_known1	Pax-4_4
29689590	rs9258170	-1651	T	G	0.1716	Branchyury_2, Eomes, HNF4_known4, SRF_disc1, SRF_known2, SRF_known4, TBX5_1, TBX5_2	Rhox11
29689631	rs17875375	-1610	T	C	0.0490	HDAC2_disc6, PLZF	GATA_known8, TATA_disc7
29689683	rs78739406	-1558	A	AG	0.1716	Mef2_known5, PU.1_disc1	ELF1_disc1, FEV, PU.1_known1, RXRA_disc4, p300_disc10
29689717	rs9258171	-1524	G	T	0.1716	Ets_known1, Pax-5_known1, Zfp691, p300_known1	-
29689742	rs183540921	-1499	T	G	0.0025	ERalpha-a_disc1	-
29689973	rs9258172	-1268	C	T	0.1716	Pou3f2_2	Foxf1, Foxi1, Foxl1_1, Foxl1_2, Pou2f2_known2, TATA_known1, TATA_known3
29690002	rs9258173	-1239	G	T	0.1716	DMRT4, DMRT5, MIF-1	DMRT1, DMRT3, DMRT7, RFX5_known2
29690026	rs9258174	-1215	T	C	0.1716	AIRE_2, AP-1_disc2, E2F_disc4, Irf_disc, NF- Y_disc1, NF-Y_known1, NF-Y_known3, Nobox_2, Pbx3_disc1, RFX5_disc2, SP1_disc1, SP2_disc1, TATA_disc6	-
29690056	rs17875376	-1185	C	T	0.1985	Pitx2	CEBPD, Dlx3, En-1_1, Msx-1_1
29690125	rs9258175	-1116	G	A	0.1716	SZF1-1	-
29690228	rs9258176	-1013	C	G	0.1716	RXRA_knonw3, ZBRK1	BDP1_disc1, SP1_disc3

^a Posição considerando a Adenina do primeiro ATG traduzido como nucleotídeo +1

^b Alelo referência no genoma humano versão hg19

^c Frequência do alelo alternativo

^d Position Weight Matrix ID (Biblioteca de Kheradpour and Kellis, 2013), (KHERADPOUR *et al.*, 2013; MAHONY *et al.*, 2007; MATYS *et al.*, 2003).

Gráfico 1. Afinidade dos fatores de transcrição reportados pelo software HaploReg por sítio de variação na região promotora distal do gene *HLA-F*

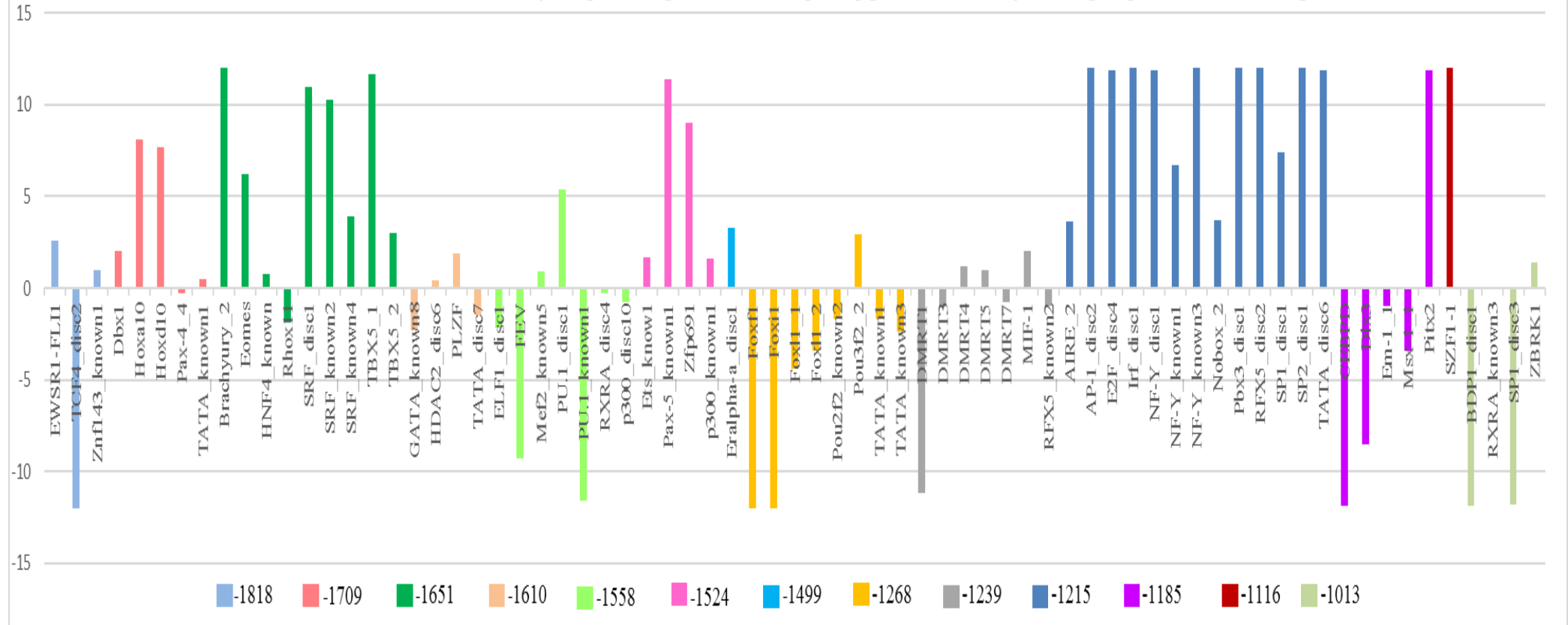


Gráfico 1. Diferença de afinidade dos motivos de fatores de transcrição reportados pelo software HaploReg pela sequência do alelo referência ou alternativo. Valores de diferença positivos indicam maior afinidade pelo alelo de referência, e valores negativos indicam maior afinidade pelo alelo alternativo. As posições dos sítios de variação considerados estão indicadas por cor. A relação dos alelos referência/alternativo por sítio de variação estão indicados na Tabela 3.

Tabela 4. Haplótipos estendidos do gene *HLA-F* considerando segmento entre os nucleotídeos -1893 e +4252, considerando Adenina do primeiro ATG traduzido como +1

Promotora distal ^a	5' upstream ^b	Alelo <i>HLA-F</i> ^b	3' downstream ^b	Frequência nas amostras brasileiras (2n=392)
F*distal-A	F*upstream-A	F*01:01:01:09	F*downstream-B	0.1735
F*distal-A	F*upstream-B	F*01:01:01:08	F*downstream-A	0.1684
F*distal-B	F*upstream-B	F*01:01:01:01	F*downstream-A	0.1403
F*distal-A	F*upstream-A	F*01:03:01:01	F*downstream-D	0.1020
F*distal-D	F*upstream-A	F*01:03:01:01 (1383G)	F*downstream-E	0.0459
F*distal-A	F*upstream-D	F*01:01:01:05 (1943G, [TG] ₁₂)	F*downstream-A	0.0434
F*distal-A	F*upstream-A	F*01:01:01:11	F*downstream-B	0.0383
F*distal-C	F*upstream-E	F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, new3205G, [TG] ₁₃)	F*downstream-C	0.0383
F*distal-C	F*upstream-C	F*01:01:02:03 (1943G, 2208C, [TG] ₁₁)	F*downstream-C	0.0332
F*distal-B	F*upstream-B	F*01:01:01:01 ([TG] ₁₂)	F*downstream-A	0.0332
F*distal-C	F*upstream-C	F*01:01:02:04 (3189T)	F*downstream-C	0.0306
F*distal-C	F*upstream-F	F*01:01:02:03 (1943G, [TG] ₁₁)	F*downstream-C	0.0179
F*distal-A	F*upstream-A	F*01:01:01:09 (342A) ou F*01:04 (1943G, 3189T)	F*downstream-B	0.0128
F*distal-B	F*upstream-B	F*01:01:01:01 ([TG] ₁₂)	F*downstream-F	0.0128
F*distal-C	F*upstream-E	F*01:01:02:02 (new-238T, 1943G, 2208C, new2486T, new3205G, [TG] ₁₃)	F*downstream-C	0.0128
F*distal-A	F*upstream-A	F*01:01:01:09 (new394G)	F*downstream-B	0.0077
F*distal-C	F*upstream-H	F*01:01:02:03 (1943G, [TG] ₁₃)	F*downstream-C	0.0077
F*distal-A	F*upstream-D	F*01:01:01:05 (1943G, [TG] ₁₄)	F*downstream-A	0.0077
F*distal-C	F*upstream-C	F*01:01:02:03 (1943G, 2208C, [TG] ₁₁)	F*downstream-G	0.0077
F*distal-D	F*upstream-A	F*01:03:01:01 (new1378T, 1383G, [TG] ₁₁)	F*downstream-E	0.0051
F*distal-A	F*upstream-G	F*01:01:01:08 (-222G, 2698G, [TG] ₁₁)	F*downstream-A	0.0051
F*distal-A	F*upstream-B	F*01:01:01:08 (new1779A)	F*downstream-A	0.0051
F*distal-A	F*upstream-G	F*01:01:01:08 (-222G, 2698G, [TG] ₁₁)	F*downstream-H	0.0051
F*distal-C	F*upstream-C	F*01:01:02:03 (1943G, 2208C, [TG] ₁₂)	F*downstream-C	0.0051
F*distal-B	F*upstream-B	F*01:01:01:01 (new1313G)	F*downstream-A	0.0051
F*distal-A	F*upstream-D	F*01:01:01:08 (-222G, 2698G, [TG] ₁₁)	F*downstream-A	0.0026
F*distal-C	F*upstream-E	F*01:01:02:02 (new-238T, 1943G, 2208C, new2486T, new3205G)	F*downstream-C	0.0026
F*distal-C	F*upstream-C	F*01:01:02:03 (new1570A, 1943G, 2208C, [TG] ₁₁)	F*downstream-C	0.0026
F*distal-B	F*upstream-B	F*01:01:01:01	F*downstream-F	0.0026
F*distal-C	F*upstream-C	F*01:01:02:02 (1943G, 2208C, [TG] ₁₁)	F*downstream-C	0.0026
F*distal-A	F*upstream-A	F*01:03:01:01 ([TG] ₁₁)	F*downstream-D	0.0026
F*distal-A	F*upstream-B	F*01:01:01:08	F*downstream-J	0.0026
F*distal-C	F*upstream-C	F*01:01:02:03 (1943G, 2208C, [TG] ₁₂)	F*downstream-I	0.0026
F*distal-E	F*upstream-A	F*01:01:01:11	F*downstream-B	0.0026
F*distal-A	F*upstream-A	F*01:03:01:01 (new1225C)	F*downstream-D	0.0026
F*distal-B	F*upstream-B	F*01:01:01:01 ([TG] ₁₄)	F*downstream-A	0.0026
F*distal-A	F*upstream-A	F*01:03:01:01 ([TG] ₁₃)	F*downstream-D	0.0026
F*distal-A	F*upstream-A	F*01:01:01:11 ([TG] ₉)	F*downstream-B	0.0026
F*distal-F	F*upstream-B	F*01:01:01:01	F*downstream-A	0.0026
Diversidade nucleotídica				0.0040 +/- 0.0020
Diversidade gênica				0.9267 +/- 0.0062

^a Haplótipos encontrados no presente estudo, descritos na Tabela 2.

^b Haplótipos encontrados em estudo anterior (21).

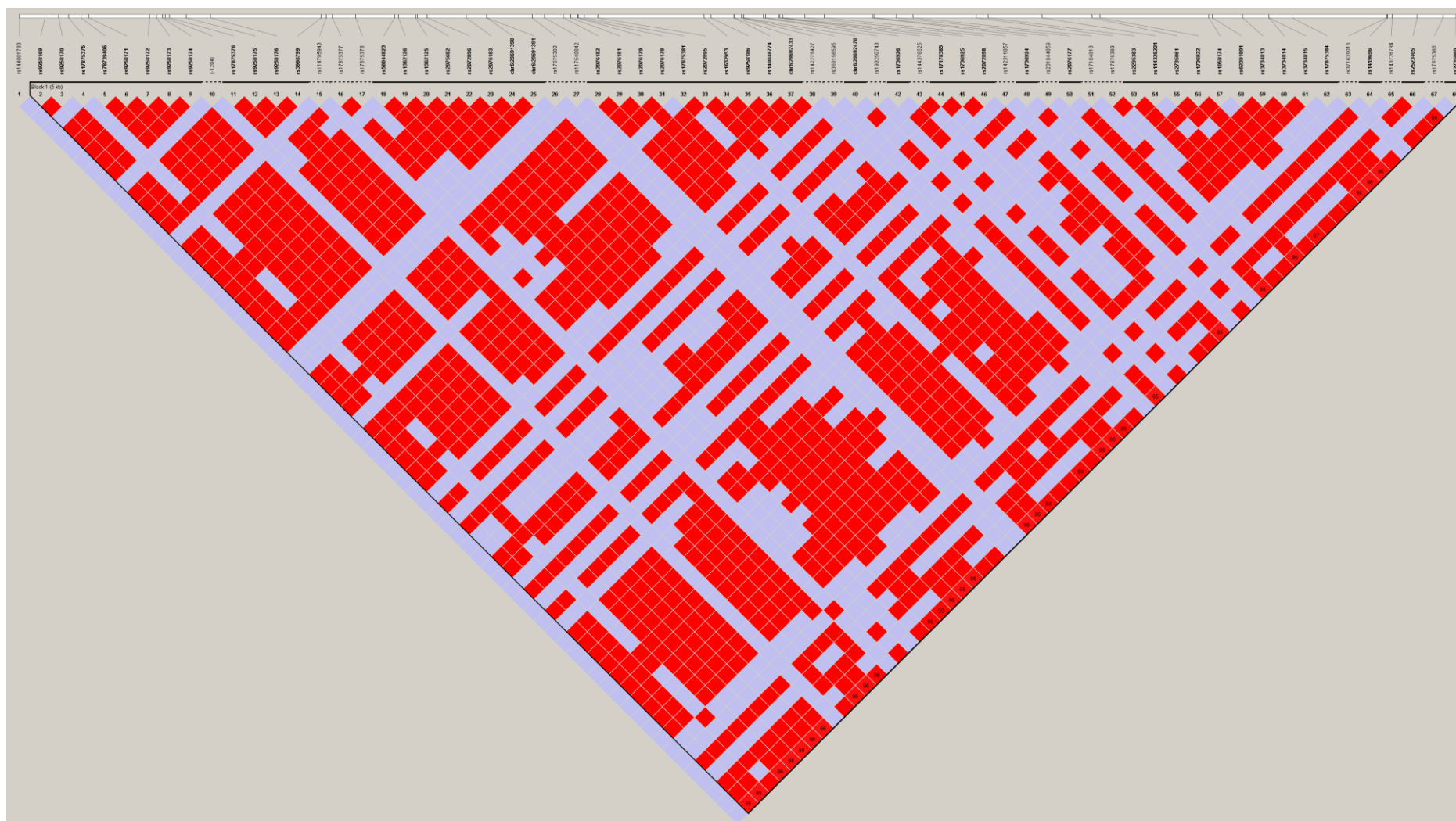


Figura 2. Mapa de desequilíbrio de ligação (LD) entre pares de pontos de variação na região estendida do gene *HLA-F* (de -1893 a +4252). Essa imagem foi gerada pelo Haploview 4.2 usando pontos de variação com frequência do alelo mínimo (MAF) $\geq 0,1\%$. Áreas em vermelho indicam forte LD [$\log$ de odds ratio ($LOD \geq 2$, $D' = 1$)], áreas em azul/cinza indicam LD fraco ($LOD \leq 2$, $D' = 1$) e áreas em branco indicam nenhum LD ($LOD \leq 2$, $D' \leq 1$). Valores de D' diferente de 1 estão representados dentro dos quadrados na forma de porcentagem.

4. Discussão

O presente estudo avaliou aproximadamente 950 bp da região promotora distal do gene *HLA-F*. Dos três trabalhos existentes sobre a variabilidade do *HLA-F*, apenas um deles teve abrangência semelhante a este estudo, considerando uma região promotora distal a aproximadamente 2kb do gene (9). Os outros dois trabalhos nessa linha discutem apenas a variabilidade da região codificadora considerando os tipos de alelos encontrados na população avaliada. Estes trabalhos prévios foram realizados na China (30) e no sul do Brasil (13). Em 2006, PYO et al. analisou o gene estendido em 33 amostras providas da Ásia, África, América e caucasianas, mas por uma metodologia diferente da usada neste estudo. Dos 15 pontos de variação encontrados aqui, o *singleton* na posição -1204G/A (**Tabela 1**) já tinha sido descrito nesse trabalho anterior (9).

O gene *HLA-F* é considerado um dos menos variáveis dentre os genes da família HLA. Os estudos realizados até o momento indicam que apenas duas moléculas HLA-F são codificadas pelos alelos mais frequentemente encontrados em populações mundiais. (9,13,21,30). Estas duas moléculas são denominadas F*01:01 (codificada por alelos dos grupos F*01:01:01 e F*01:01:02) e F* 01:03 (codificada pelos alelos do grupo F*01:03). Apesar desta baixa variabilidade na região codificadora (em termos de proteína produzida), a região promotora distal do *HLA-F* parece ser bastante polimórfica quando comparada à região codificadora.

Neste estudo, verificamos que os 13 pontos de variação encontrados estão arranjos em apenas 6 haplótipos diferentes. Além disso, este segmento apresenta um forte desequilíbrio de ligação, o que explica o perfil dos haplótipos encontrados. A maioria dos alelos alternativos estão associados com o mesmo haplótipo (F*distal-C). A região promotora distal do *HLA-F* pode ser considerada uma região relativamente polimórfica quando comparada à região codificadora, considerando a quantidade e frequência dos pontos de variação encontrados. No entanto, a diversidade de haplótipos é reduzida, concentrando-se em apenas dois grandes grupos: os haplótipos F*distal-A (e derivados) e F*distal-C. Embora a maioria dos haplótipos sejam derivados de F*distal-A, a presença de um haplótipo altamente divergente e com frequência alta (F*distal-C) poderia influenciar o perfil transcricional dos genes *HLA-F* que carregam este haplótipo divergente, por modificar a afinidade de fatores de transcrição pelo DNA do promotor.

Na região promotora estão presentes os elementos que regulam a transcrição de um gene, ou seja, este é o local onde o aparato basal de transcrição se monta para iniciar a

produção de mRNA. Na região promotora cerne e proximal estão os sítios mais conhecidos de ligação a fatores transcricionais, como TATA box e CCAAT box, onde se liga o aparato basal de transcrição (15). Nessas e na região promotora distal estão motivos de DNA com capacidade de ligação à proteínas estimulatórias ou inibidoras da transcrição (15).

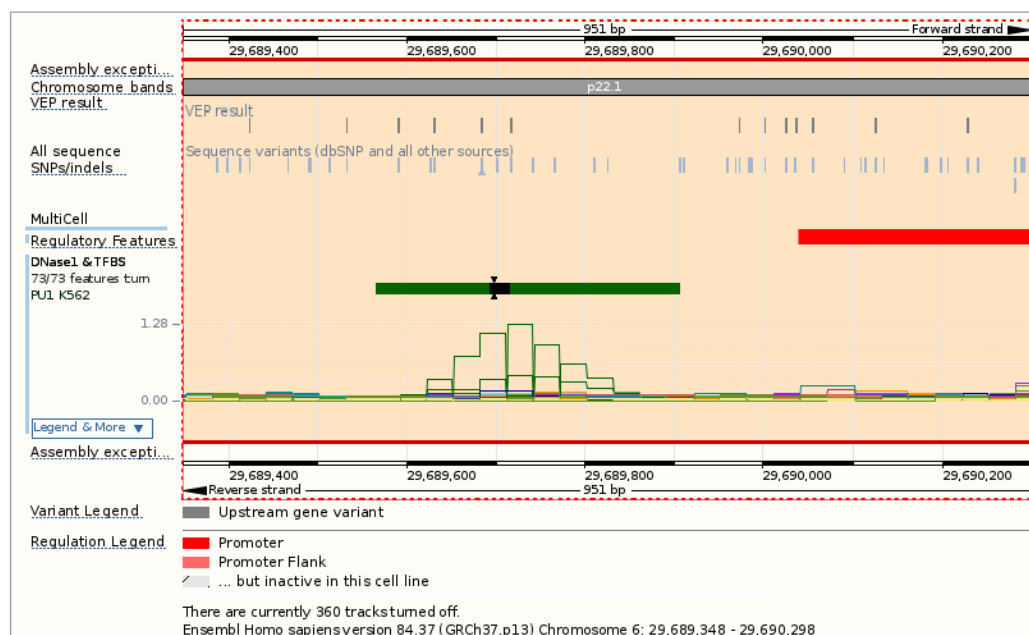
Os primeiros estudos que descreveram a estrutura do gene *HLA-F* mostraram que sua região promotora proximal se assemelha as dos genes HLA de classe I clássicos, fazendo com que a expressão da molécula de HLA-F também fosse similar às das moléculas clássicas, (14,31). No caso do *HLA-F*, nessa região promotora proximal há elementos responsivos ao NF- κ B, IRF1 (IFN- γ) e CIITA (15,32). No entanto, diferente dos genes HLA de classe I clássicos, mas semelhante ao gene *HLA-G*, a promotora cerne do HLA-F não possui um elemento denominado Inr (“*Initiator*”) conservado, o que poderia explicar a relativa limitação da expressão da molécula a alguns tecidos (15). Assim, tendo elementos regulatórios semelhantes na região promotora, tanto comparados aos genes HLA de classe I, como ao *HLA-G*, a expressão de HLA-F apresentaria um perfil intermediário (14,31,33).

Como citado anteriormente, a região promotora distal pode conter sítios de ligação a proteínas regulatórias que acentuam ou silenciam a expressão gênica. Para o *HLA-F*, quando em 1990 GERAGHTY *et al.* o descreveram primeiramente essa promotora distal, foi encontrado sequências altamente repetitivas configurando a presença de um microsatélite. Além da influência estrutural na molécula de DNA, que por si interfere no nível de expressão de um gene, não foi descartado a possível função regulatória dessa região repetitiva (14).

No presente estudo, verificamos que os pontos de variação encontrados no promotor distal do *HLA-F* poderiam influenciar a ligação de dezenas de fatores de transcrição (**Tabela 3**). Ainda, como a maioria das variantes ocorreu associada à um único haplótipo (F*distal-C), este haplótipo poderia ser responsável por um perfil transcricional completamente diferente. Mais precisamente, o haplótipo de F*distal-C, representado pela maioria dos alelos alternativos, parece perder a afinidade por diversos fatores de transcrição, enquanto o haplótipo F*distal-B parece ser o haplótipo com maior afinidade a mais fatores de transcrição. Estes dados apontam para uma importância funcional dessa região em relação à expressão do gene *HLA-F*.

Além dos resultados apontados na **Tabela 3** e no **Gráfico 1**, obtidos pelo programa HaploReg (29), dados do projeto ENCODE (34), referentes aos sítios regulatórios e motivos de fatores de transcrição, destacaram um segmento regulatório dentro da região promotora distal estudada. Como mostra a **Figura 3**, o sítio regulatório relacionado ao fator de transcrição PU.1 aparece em destaque. PU.1 é um fator de transcrição ativador envolvido na

diferenciação e ativação de macrófagos e células B e atua como um *enhancer* linfóide específico (38). Além disso, é possível observar que 4 pontos de variação estão na mesma região de ligação desse fator de transcrição, 3 deles com os alelos alternativos relacionados ao haplótipo de F*distal-C (-1651G, -1558AG e -1524T, **Tabela 2**) e o outro ao haplótipo F*distal-D (-1610C, **Tabela 2**). Dentre esses quatro pontos de variação, a *indel* A/AG anotado na região de -1558 é o mais próximo do motivo de ligação de PU.1, caracterizado por ser uma sequência rica em purina (5'- GAGGAA-3') (38). No **Gráfico 1**, os resultados apontados pelo HaploReg mostram que afinidade de PU.1 é maior para a sequência do alelo alternativo do que para a do alelo de referência, observado pelo valor de diferença de afinidade indicado em “PU.1_known1”, que equivale aos dados do projeto ENCODE, utilizados para o banco de dados do HaploReg. O sítio “PU.1_disc1” equivale a um motivo semelhante descoberto pelas análises do HaploReg, que teria maior afinidade pela sequência de referência. (39). Esse dado reforça a importância funcional dessa região e aponta também para um possível perfil transcricional diferente relacionado a promotora F*distal-C, que, por exemplo, teria afinidade maior por PU.1.



Ensembl GRCh37 release 84 - March 2016 © WTSI / EMBL-EBI
<http://grch37.ensembl.org>

Figura 3. Região promotora distal compreendendo os nucleotídeos de -1893 a -943. Em “VEP result” estão marcados os 13 pontos de variação considerados na inferência de haplótipos (Tabela 2). Em “MultiCell” estão os dados de múltiplas linhagens celulares contidas no Projeto ENCODE associado a anotação de sítios regulatórios e motivos de ligação de fatores de transcrição, ilustrados na forma de barra e picos. Nesse caso, o sítio destacado com maior pico é referente ao fator de transcrição PU.1

A relação entre os haplótipos promotores e codificadores também foi avaliada, como mostra a **Tabela 4**. Os haplótipos F*distal-A (e derivados) estão associados com alelos que codificam ambas as proteínas HLA-F frequentes, F*01:01 e F*01:03, enquanto que o F*distal-C está exclusivamente associada com alelos F*01:01:02 ou derivados, que codificam a proteína F*01:01. Considerando o perfil de desequilíbrio de ligação característico do *HLA-F* (**Figura 2**), os alelos que codificam a proteína F*01:01 (a mais frequentemente encontrada) contam com dois promotores altamente divergentes. Este fenômeno já foi observado anteriormente para outro gene não clássico, o *HLA-G* (40–42), em que promotores divergentes estariam associados com alelos codificadores que codificam a mesma proteína. Assim como observado para o *HLA-G*, é possível que a seleção natural esteja modificando as frequências alélicas no promotor de *HLA-F*. Já foi demonstrado que o promotor do *HLA-G* sofre influência de seleção balanceadora, elevando a heterozigose e a frequência de diversas variantes no promotor (41). A heterozigose nesta região traria uma vantagem evolutiva, pois o indivíduo seria capaz de estimular a expressão de HLA-G quando necessário (gestação, doenças autoimunes) ou diminuir sua expressão quando da necessidade de uma resposta imunológica mais intensa (infecções, tumores). Este fenômeno estaria em parte associado com a presença de haplótipos promotores de alta e baixa produção no mesmo indivíduo, ou haplótipos que respondem a diferentes estímulos, sem modificação da proteína produzida. É possível que o mesmo fenômeno esteja ocorrendo com o gene *HLA-F*, no entanto, os testes de desvio de neutralidade não apontaram seleção natural neste segmento. Por fim, são necessários estudos funcionais para provar que estes promotores podem influenciar a expressão de *HLA-F*.

Referências

1. Ishitani A, Sageshima N, Hatake K. The involvement of HLA-E and -F in pregnancy. *J Reprod Immunol* [Internet]. abril de 2006 [citado 24 de junho de 2016];69(2):101–13. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16487601>
2. Djuricic S, Hviid TVF. HLA Class Ib Molecules and Immune Cells in Pregnancy and Preeclampsia. *Front Immunol* [Internet]. 2014 [citado 24 de junho de 2016];5:652. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25566263>
3. Goldberg AC, Rizzo LV. MHC structure and function – antigen presentation. Part 1. Einstein (São Paulo, Brazil). 2015;13(1):153–6.
4. Amiot L, Vu N, Samson M, Amiot L, Vu N, Samson M. Immunomodulatory

- Properties of HLA-G in Infectious Diseases. *J Immunol Res* [Internet]. Hindawi Publishing Corporation; 2014 [citado 24 de junho de 2016];2014:1–14. Recuperado de: <http://www.hindawi.com/journals/jir/2014/298569/>
5. Morandi F, Pistoia V. Interactions between HLA-G and HLA-E in Physiological and Pathological Conditions. *Front Immunol* [Internet]. Frontiers Media SA; 2014 [citado 24 de junho de 2016];5:394. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25202308>
 6. Rojo R, Castro MJ, Martinez-Laso J, Serrano-Vela JI, Morales P, Moscoso J, et al. MHC-F DNA sequences in bonobo, gorilla and orangutan. *Tissue Antigens* [Internet]. outubro de 2005 [citado 24 de junho de 2016];66(4):277–83. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16185322>
 7. Moscoso J, Serrano-Vela JI, Pacheco R, Arnaiz-Villena A. HLA-G, -E and -F: allelism, function and evolution. *Transpl Immunol* [Internet]. dezembro de 2006 [citado 24 de junho de 2016];17(1):61–4. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/17157219>
 8. Allan DSJ, Lepin EJM, Braud VM, O’Callaghan CA, McMichael AJ. Tetrameric complexes of HLA-E, HLA-F, and HLA-G. *J Immunol Methods* [Internet]. 1 de outubro de 2002 [citado 24 de junho de 2016];268(1):43–50. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/12213342>
 9. Pyo C-W, Williams LM, Moore Y, Hyodo H, Li SS, Zhao LP, et al. HLA-E, HLA-F, and HLA-G polymorphism: genomic sequence defines haplotype structure and variation spanning the nonclassical class I genes. *Immunogenetics* [Internet]. maio de 2006 [citado 24 de junho de 2016];58(4):241–51. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16570139>
 10. Lee N, Ishitani A, Geraghty DE. HLA-F is a surface marker on activated lymphocytes. *Eur J Immunol* [Internet]. agosto de 2010 [citado 24 de junho de 2016];40(8):2308–18. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/20865824>
 11. Goodridge JP, Burian A, Lee N, Geraghty DE. HLA-F and MHC class I open conformers are ligands for NK cell Ig-like receptors. *J Immunol* [Internet]. 1 de outubro de 2013 [citado 24 de junho de 2016];191(7):3553–62. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/24018270>
 12. Kochan G, Escors D, Breckpot K, Guerrero-Setas D. Role of non-classical MHC class I molecules in cancer immunosuppression. *Oncoimmunology* [Internet]. Landes Bioscience; 1 de novembro de 2013 [citado 24 de junho de 2016];2(11):e26491.

- Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/24482746>
13. Manvailer LFS, Wowk PF, Mattar SB, da Siva JS, da Graça Bicalho M, Roxo VMMS. HLA-F polymorphisms in a Euro-Brazilian population from Southern Brazil. *Tissue Antigens* [Internet]. dezembro de 2014 [citado 24 de junho de 2016];84(6):554–9. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25413105>
 14. Geraghty DE, Wei XH, Orr HT, Koller BH. Human leukocyte antigen F (HLA-F). An expressed HLA gene composed of a class I coding sequence linked to a novel transcribed repetitive element. *J Exp Med* [Internet]. 1 de janeiro de 1990 [citado 24 de junho de 2016];171(1):1–18. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/1688605>
 15. Howcroft TK, Singer DS. Expression of Nonclassical MHC Class Ib Genes: Comparison of Regulatory Elements. *Immunol Res* [Internet]. Humana Press; 2003 [citado 24 de junho de 2016];27(1):1–30. Recuperado de: <http://link.springer.com/10.1385/IR:27:1:1>
 16. Shobu T, Sageshima N, Tokui H, Omura M, Saito K, Nagatsuka Y, et al. The surface expression of HLA-F on decidual trophoblasts increases from mid to term gestation. *J Reprod Immunol* [Internet]. dezembro de 2006 [citado 24 de junho de 2016];72(1-2):18–32. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16806485>
 17. Boyle LH, Gillingham AK, Munro S, Trowsdale J. Selective export of HLA-F by its cytoplasmic tail. *J Immunol* [Internet]. 1 de junho de 2006 [citado 24 de junho de 2016];176(11):6464–72. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/16709803>
 18. Goodridge JP, Burian A, Lee N, Geraghty DE. HLA-F complex without peptide binds to MHC class I protein in the open conformer form. *J Immunol* [Internet]. 1 de junho de 2010 [citado 24 de junho de 2016];184(11):6199–208. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/20483783>
 19. Wainwright SD, Biro PA, Holmes CH. HLA-F is a predominantly empty, intracellular, TAP-associated MHC class Ib protein with a restricted expression pattern. *J Immunol* [Internet]. 1 de janeiro de 2000 [citado 24 de junho de 2016];164(1):319–28. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/10605026>
 20. SEYFERTH G. COLONIZAÇÃO, IMIGRAÇÃO E A QUESTÃO RACIAL NO BRASIL. *Rev USP* [Internet]. 30 de maio de 2002 [citado 24 de junho de 2016];0(53):117. Recuperado de: <http://www.revistas.usp.br/revusp/article/view/33192>
 21. Lima THA [UNESP]. Estudo da variabilidade da região codificadora e 3' não traduzida

- do gene HLA-F no Brasil. *Aleph*. Universidade Estadual Paulista (UNESP); 2015;56 f.
22. Castelli EC, Mendes-Junior CT, Sabbagh A, Porto IOP, Garcia A, Ramalho J, et al. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. *Hum Immunol* [Internet]. dezembro de 2015 [citado 24 de junho de 2016];76(12):945–53. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/26187162>
 23. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* [Internet]. Cold Spring Harbor Laboratory Press; setembro de 2010 [citado 24 de junho de 2016];20(9):1297–303. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/20644199>
 24. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* [Internet]. Nature Publishing Group; 10 de abril de 2011 [citado 24 de junho de 2016];43(5):491–8. Recuperado de: <http://www.nature.com/doifinder/10.1038/ng.806>
 25. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* [Internet]. março de 2013 [citado 24 de junho de 2016];14(2):178–92. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/22517427>
 26. Qin ZS, Niu T, Liu JS. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet* [Internet]. novembro de 2002 [citado 24 de junho de 2016];71(5):1242–7. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/12452179>
 27. EXCOFFIER L, LISCHER HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* [Internet]. Blackwell Publishing Ltd; maio de 2010 [citado 24 de junho de 2016];10(3):564–7. Recuperado de: <http://doi.wiley.com/10.1111/j.1755-0998.2010.02847.x>
 28. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* [Internet]. Oxford University Press; 15 de janeiro de 2005 [citado 24 de junho de 2016];21(2):263–5. Recuperado de: <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/bth457>
 29. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation,

- and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* [Internet]. janeiro de 2012 [citado 24 de junho de 2016];40(Database issue):D930–4. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/22064851>
30. Pan FH, Liu XX, Tian W. Characterization of HLA-F polymorphism in four distinct populations in Mainland China. *Int J Immunogenet* [Internet]. outubro de 2013 [citado 24 de junho de 2016];40(5):369–76. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/23551590>
 31. Wei XH, Orr HT. Differential expression of HLA-E, HLA-F, and HLA-G transcripts in human tissue. *Hum Immunol* [Internet]. outubro de 1990 [citado 24 de junho de 2016];29(2):131–42. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/2249951>
 32. Gobin SJ, van den Elsen PJ. Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F, and HLA-G. *Hum Immunol* [Internet]. novembro de 2000 [citado 24 de junho de 2016];61(11):1102–7. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/11137213>
 33. O’Callaghan CA, Bell JI. Structure and function of the human MHC class Ib molecules HLA-E, HLA-F and HLA-G. *Immunol Rev* [Internet]. junho de 1998 [citado 24 de junho de 2016];163:129–38. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9700506>
 34. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* [Internet]. 6 de setembro de 2012 [citado 9 de julho de 2016];489(7414):57–74. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/22955616>
 35. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR, Freedman M, et al. The Ensembl Regulatory Build. *Genome Biol* [Internet]. BioMed Central; 24 de dezembro de 2015 [citado 9 de julho de 2016];16(1):56. Recuperado de: <http://genomebiology.com/2015/16/1/56>
 36. Zerbino DR, Johnson N, Juetteman T, Sheppard D, Wilder SP, Lavidas I, et al. Ensembl regulation resources. *Database* [Internet]. Oxford University Press; 17 de fevereiro de 2016 [citado 9 de julho de 2016];2016:bav119. Recuperado de: <http://database.oxfordjournals.org/lookup/doi/10.1093/database/bav119>
 37. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. *Nucleic Acids Res* [Internet]. Oxford University Press; 4 de janeiro de 2016 [citado 9 de julho de 2016];44(D1):D710–6. Recuperado de: <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkv1157>

38. Hornbeck P V, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 2015;43.
39. Kheradpour P, Kellis M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res [Internet]*. Oxford University Press; março de 2013 [citado 18 de setembro de 2016];42(5):2976–87. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/24335146>
40. Castelli EC, Ramalho J, Porto IOP, Lima THA, Felício LP, Sabbagh A, et al. Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity. *Front Immunol [Internet]*. 2014 [citado 8 de julho de 2016];5:476. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25339953>
41. 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 [Internet]*. novembro de 2011 [citado 24 de junho de 2016];28(11):3069–86. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/21622995>
42. Santos KE, Lima THA, Felício LP, Massaro JD, Palomino GM, Silva ACA, et al. Insights on the HLA-G evolutionary history provided by a nearby Alu insertion. *Mol Biol Evol [Internet]*. novembro de 2013 [citado 8 de julho de 2016];30(11):2423–34. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/23945036>

Conclusão

O promotor distal do gene *HLA-F* parece ser bastante polimórfico quando comparado à região codificadora. No entanto, embora diversos pontos de variação apresentem elevada frequência, a maioria dos alelos alternativos estão associados com um único haplótipo altamente divergente. Estes dois conjuntos de haplótipos estão associados com um perfil de ligação de fatores de transcrição diferentes e portanto, poderiam influenciar a expressão de *HLA-F* de forma diferente. No entanto, estes dois grupos de promotores codificariam as mesmas duas proteínas *HLA-F* encontradas mundialmente. São necessários estudos funcionais para provar que estes promotores podem influenciar a expressão de *HLA-F*.