



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

Jaqueline Ramalho

REGIÃO PROMOTORA E CODIFICADORA NO GENE
***HLA-E*: ESTRUTURA, VARIABILIDADE E HAPLÓTIPOS**

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

Orientador: Prof. Dr. Erick da Cruz Castelli

Botucatu
2017

Jaqueline Ramalho

REGIÃO PROMOTORA E CODIFICADORA NO GENE
***HLA-E*: ESTRUTURA, VARIABILIDADE E**
HAPLÓTIPOS

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

Orientador: Prof. Dr. Erick da Cruz Castelli

Botucatu
2017

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

Ramalho, Jaqueline.

Região promotora e codificadora no gene *HLA-E* : estrutura, variabilidade e haplótipos / Jaqueline Ramalho. - Botucatu, 2017

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu
Orientador: Erick da Cruz Castelli
Capes: 21103003

1. Antígenos de histocompatibilidade *HLA*. 2. MicroRNAs. 3. Haplótipos. 4. Expressão gênica. 5. Sequenciamento de nucleotídeo.

Palavras-chave: *HLA-E*; Região codificadora; Regiões regulatórias; Sequenciamento de nova geração; Variabilidade.

Dedicatória

À minha família, meus pais Waldir Luiz Ramalho e Roselei Aparecida R. Ramalho e meus irmãos Gisele Ramalho e Júlio Ramalho, à minha amiga e companheira em Botucatu, Franciane Ribeiro e ao meu namorado, amigo e companheiro de laboratório, Renato Vidal Buttura, dedico...

Agradecimientos

Agradeço primeiramente a Deus pelo presente de viver e pelas oportunidades de crescimento e de estudo, pelas pessoas que colocou e coloca no meu caminho.

Aos meus pais Waldir e Roselei, pela educação dada, pelo investimento, pelos conselhos, pelo apoio e fé depositados em mim desde o início de minha formação. Aos meus irmãos Gisele e Julio, pelo carinho, mimos, abraços e pela presença sempre. Aos meus avós e tios também pelo apoio e por sempre torcerem por mim.

À minha amiga e companheira de apartamento, Franciane Ribeiro, pelo ombro amigo, pelas risadas, pelas conversas e broncas, pela paciência, pelos chás de madrugada, pela companhia, pelo apoio, pela torcida e pela Belinha Gordinha! Agradeço não apenas por esses dois anos de Mestrado, mas pelo afeto desde o início do curso de Graduação.

Ao meu namorado, que começou como amigo e companheiro de laboratório e se tornou meu amigo e companheiro de todas as horas. Pelo imenso carinho e apoio, pela torcida, pelas freadas no ânimo desenfreado, pelo lazer para aliviar as tensões e pela paciência.

Ao meu orientador, Prof. Dr. Erick da Cruz Castelli, que me aceitou como aluna de iniciação científica e fez eu me apaixonar por estudar Imunogenética e Bioinformática. Pelo aprendizado sobre ter olhar crítico e dinâmica de apresentações e seminários. Obrigada pela oportunidade!

Aos companheiros do Laboratório de Genética Molecular e Bioinformática: Iane, Thálitta e Andreia, pelos ensinamentos e boas conversas, e Michelle, pupila, pela amizade, pelo apoio, pela visão diferenciada das coisas e discussões filosóficas e divertidas depois do almoço.

Ao Programa de Pós-Graduação em Patologia da Faculdade de Medicina de Botucatu (FMB-UNESP), por todo seu corpo docente e o ensinamento em disciplinas que forneceram. Ao CNPq pelo auxílio financeiro e bolsa de mestrado concedida nesses dois anos.

E aos membros da banca examinadora que aceitaram participar da avaliação desse trabalho bem como contribuir para minha formação e construção de conhecimento. A todos aqueles que direta ou indiretamente contribuíram para a concretização de mais essa etapa.

A todos, muito obrigada!

Epígrafe

***“Disciplina é liberdade,
Compaixão é fortaleza,
Ter bondade é ter coragem”***

Renato Russo – Legião Urbana

Resumo

Resumo

Os genes do Complexo Principal de Histocompatibilidade (MHC) codificam moléculas envolvidas com a regulação do sistema imune, participando do reconhecimento do que é próprio ou estranho ao indivíduo. O gene *HLA-E*, parte do MHC, é caracterizado por apresentar baixa, porém, ampla expressão pelos tecidos do corpo. O *HLA-E* é extremamente conservado e um dos menos polimórficos entre os genes *HLA* de classe I. A principal função da molécula *HLA-E* está relacionada ao mecanismo de imunovigilância, através da interação com receptores de células *Natural Killer* e linfócitos T. Pontos de variação nas regiões regulatórias e codificadora de *HLA-E* podem alterar sua função através da modificação da expressão gênica ou estrutura da molécula codificada, influenciando a interação com peptídeos e receptores. O presente trabalho propôs a avaliação da variabilidade do segmento estendido do gene *HLA-E*, incluindo promotor e íntrons, e estrutura de haplótipos por meio de sequenciamento de nova geração ou sequenciamento massivo paralelo. A metodologia foi aplicada para avaliação da variabilidade de 420 amostras do Estado de São Paulo. Considerando um segmento de mais de 7kb, o gene *HLA-E* mostrou-se conservado, apresentando poucas sequências diferentes e frequentes. Ao todo, 63 pontos de variação foram encontrados e caracterizados em 75 haplótipos estendidos. Foram encontrados 37 haplótipos de região promotora, porém apenas 10 apresentam frequência superior a 1%. Para a região codificadora, foram encontrados 27 haplótipos, sendo que 15 caracterizam novos alelos *HLA-E*. No entanto, duas das moléculas codificadas, representadas pelos alelos do tipos E*01:01 e E*01:03, correspondem a mais de 90% de todas as sequências. Essa característica reforçou a conservação da região codificadora, na qual a maioria dos pontos de variação ocorreram em segmentos regulatórios e íntrons, ou caracterizavam mutações sinônimas em éxons. Na região 3' não traduzida, 12 haplótipos foram encontrados, porém cerca de 90% das sequências são representadas por apenas dois haplótipos, 3UTR-1 e 3UTR-2. A diversidade nucleotídica foi maior no segmento promotor do que na região codificadora e 3' não traduzida. Também foi observado fraco desequilíbrio de ligação ao longo de *HLA-E*, provavelmente devido a presença de diversos elementos *Alu* que pode influencia a taxa de recombinação entre os segmentos e, por causa disso, um mesmo haplótipo regulatório pode ocorrer associado a diferentes sequências de codificadora. Análises futuras relacionados a ligação de fatores de transcrição e RNAs não codificadores são necessárias para averiguar se essas variações influenciariam o perfil de expressão do gene.

Abstract

Abstract

The Major Histocompatibility Complex (MHC) genes encode molecules involved mainly in recognition of self and non self antigens. The *HLA-E* gene is characterized by low but wide expression among tissues. Thus far, *HLA-E* is considered a conserved gene and one of the least polymorphic class I *HLA* genes. The main *HLA-E* function is related to immune surveillance by the interaction with natural killer cell receptors and T lymphocytes. Variable sites within the *HLA-E* regulatory and coding segments may influence gene function by modifying its expression pattern or encoded molecule, thus, influencing its interaction with receptors and the peptide. Here we propose an approach to evaluate the complete *HLA-E* variability, including promoter and introns segments, and the evaluation of the *HLA-E* haplotype structure by using massively parallel sequencing, or next-generation sequencing. The methodology was applied to survey the variability of a very admixed population such as Brazilians counting 420 samples of São Paulo State. Considering a segment of about 7kb, the *HLA-E* gene is indeed conserved with few different and frequent sequences. In general, 63 variable sites were detected, arranged 75 extended haplotypes. We found 37 promoter haplotypes, but only 10 presented a frequency greater than 1%. For the coding region, 27 haplotypes were detected, with 15 representing new *HLA-E* alleles. Nevertheless, the *HLA-E* coding alleles did encode mainly two different full-length molecules, known as alleles E*01:01 and E*01:03, which corresponds to about 90% of all. This feature reinforced the conservation at the coding region, where most of the variable sites detected are intronic mutations, synonymous mutations in exons, or mutations at regulatory elements. In the 3' untranslated region, 12 haplotypes were found, but only two different sequences (3UTR-1 and 3UTR-2) presented a summed frequency of 90%. Nucleotide diversity was higher in the distal promoter segment than in the coding and 3' untranslated segments. *HLA-E* presents a weak linkage disequilibrium along the entire segment, mainly because of the presence of many *Alu* elements that may influence recombination within the gene. Because of that, the same regulatory haplotype may occur associated with different coding sequences. Further analysis involving the binding of transcription factors and non-coding RNAs are necessary to evaluate whether these variable sites at regulatory segments (or even in the coding sequence) may influence the gene expression profile.

Lista de ilustrações

Figura 1	Esquema genômico detalhado da região do Complexo Principal de Histocompatibilidade (MHC) humano, localizado no braço curto do cromossomo 6.....	20
Figura 2	Estrutura das moléculas HLA de classe I (esquerda) e classe II (direita).....	22
Figura 3	Apresentação de antígenos endógenos via HLA de classe I.....	23
Figura 4	Apresentação de antígenos via HLA de classe II.....	24
Figura 5	Interação da célula T com célula apresentadora de antígeno (APC).....	25
Figura 6	Elementos regulatórios dos genes HLA de classe I, de classe II e elementos regulatórios conservados entre as duas classes.....	27
Figura 7	Estrutura molecular do complexo HLA de classe I/peptídeo.....	28
Figura 8	Representação gráfica da região de classe I do MHC humano (mais telomérica) indicando os genes <i>HLA</i> clássicos e não clássicos de classe I e a distância aproximada em kilobases (kb) entre eles.....	29
Figura 9	Esquema correlacionando estrutura do gene <i>HLA-E</i> com a molécula codificada.....	30
Figura 10	Interação de peptídeo hidrofóbico nonamérico na fenda de ligação na molécula de HLA-E.....	34
Figura 11	Peptídeos âncora na fenda de ligação ao peptídeo da molécula de HLA-E.....	34
Figura 12	Montagem do complexo HLA-E/peptídeo dependente de TAP no retículo endoplasmático...	35
Figura 13	Interação HLA-E/peptídeo com receptor CD94/NKG2A.....	36
Figura 14	Interação HLA-E/peptídeo com os receptores de células NK: CD94/NKG2A (inibitório), à direita, e CD94/NKG2C (ativador), à esquerda.....	36
Figura 15	Papel duplo da molécula HLA-E.....	38
Figura 16	Regulação pós transcricional por microRNAs.....	40
Figura 17	Elementos <i>Alu</i> presentes na região do gene <i>HLA-E</i>	41
Figura 18	Esquema da região promotora típica de um gene HLA de classe I com as posições aproximadas dos elementos regulatórios conhecidos em relação ao início da tradução.....	42
Figura 19	Regulação transcricional de <i>HLA-E</i>	42
Figure 1	<i>HLA-E</i> segments for haplotype analysis.....	67
Figure 2	Linkage disequilibrium (LD) between pairs of single nucleotide polymorphisms (SNPs) at the extended <i>HLA-E</i> gene segment (from -2143 to +4776 positons).....	75
Figure 3	<i>Alu</i> elements present at <i>HLA-E</i> segment.....	81

Lista de tabelas

Tabela 1	Número de alelos identificados para os genes de classe I (IMGT/HLA, release 3.27.0).....	31
Table 1	List of variable sites detected at the <i>HLA-E</i> gene on a Brazilian population sample considering the segment between nucleotide -2143 and +4776 and the Adenine of the first translated ATG as nucleotide +1.....	63
Table 2	List of variable sites without defined haplotype detected at the <i>HLA-E</i> gene on a Brazilian population sample considering the segment between nucleotide -2143 and +4776 and the Adenine of the first translated ATG as nucleotide +1.....	65
Table 3	List of the <i>HLA-E</i> 5' upstream and 5' untranslated region haplotypes found in a Brazilian population sample.....	67
Table 4	List of <i>HLA-E</i> coding haplotypes or coding alleles found in Brazilian population sample.....	69
Table 5	List of haplotypes at the <i>HLA-E</i> 3' untranslated region (exon 8) detected on Brazilian population sample.....	71
Table 6	<i>HLA-E</i> extended haplotypes detected on a Brazilian population sample, considering the segment between nucleotide -2143 and +4776.....	72

Lista de abreviaturas e siglas

3'NT	3' não traduzida
β 2M	Beta-2-microglobulina
bp ou pb	<i>base pairs</i> (pares de base)
CIITA	<i>Class II transcription factor</i>
CLIP	<i>Class II-associated invariant chain peptide</i>
DNA	Ácido desoxirribonucleico
EBV	Epstein-Barr vírus
<i>enhA/B</i>	<i>Enhancer A/B</i>
GAS	<i>Gamma-interferon-activation sites</i>
HCMV	Citomegalovírus humano
HCV	Vírus da hepatite C
HIV	Vírus da imunodeficiência humana
HLA	<i>Human Leukocyte Antigen</i> (Antígeno Leucocitário Humano)
HPV	Herpes vírus
Hsp60	<i>Heat shock protein 60</i>
ICAM	<i>Intercellular Adhesion Molecule</i>
IFN- γ	Interferon gama
Ig	Imunoglobulina
IL-1 β	Interleucina 1 beta
IL-10	Interleucina 10
IL-2	Interleucina 2
IMGT	<i>ImMunoGeneTics information system</i>
InflM	<i>Influenza matrix protein</i>
IRF-1	<i>Interferon regulatory factor-1</i>
ISRE	<i>Interferon-Stimulated Response Element</i>
ITAM	<i>Immunoreceptor tyrosine-based activation motif</i>
ITIM	<i>Immunoreceptor tyrosine-based inhibition motif</i>
Kb	Kilobases (10^3 bases)
MAF	<i>Minimum allele frequency</i>
Mb	Megabases (10^6 bases)

MHC	<i>Major Histocompatibility Complex</i> (Complexo Principal de Histocompatibilidade)
mRNA	RNA mensageiro
NF-κB	<i>Nuclear factor kappa B</i>
NGS	<i>Next generation sequencing</i>
NK	Células <i>Natural Killer</i>
NK-CTL	Células <i>Natural Killer</i> citotóxicas
PCR	Reação em Cadeia da Polimerase
Prdx5	Peroxiredoxina 5
RE	Retículo endoplasmático
RFX	<i>Regulatory fator X</i>
RNA	Ácido ribonucleico
SHP-1	<i>Src Homology Phosphatase 1</i>
SINE	<i>Short interspered element</i>
SNP	<i>Single nucleotide polymorphism</i> (Polimorfismo de base única)
STAT1	<i>Signal transducer and activator of transcription 1</i>
TAP	<i>Transporter associated with antigen processing</i>
TBP	<i>TATA binding protein</i>
TCR	Receptor de célula T
TFIID	<i>Transcriptional factor II D</i>
TNF	Fator de necrose tumoral
UTR	<i>Untranslated region</i>
VCF	<i>Variant Call Format</i>

Sumário

Capítulo I	19
1. REVISÃO DE LITERATURA	20
1.1 Complexo Principal de Histocompatibilidade.....	20
1.2 Sistema HLA	21
1.3 HLA-E.....	28
1.3.1 Região codificadora do gene <i>HLA-E</i>	29
1.3.2 A molécula HLA-E: características e função	32
1.3.3 Regiões regulatórias do gene <i>HLA-E</i>	39
1.3.4 Importância clínica de HLA-E	44
1.4 Referências Bibliográficas	46
2. JUSTIFICATIVA.....	55
3. OBJETIVOS	55
Capítulo II.....	56
4. ARTIGO.....	57
4.1 Introduction	58
4.2 Material and Methods.....	59
4.2.1 Samples	60
4.2.2 <i>HLA-E</i> amplification, library preparation and sequencing.....	60
4.2.3 Data processing and analysis.....	61
4.3 Results	63
4.4 Discussion	76
4.4.1 The <i>HLA-E</i> variability at the 5'upstream segment	76
4.4.2 The <i>HLA-E</i> variability at the coding segment	78
4.4.4 The <i>HLA-E</i> variability at extended <i>HLA-E</i> segment	80
4.4.5 Concluding remarks	80
4.5 References	82
5 CONCLUSÃO	89

Capítulo I

1. REVISÃO DE LITERATURA

1.1 Complexo Principal de Histocompatibilidade

O Complexo Principal de Histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*) foi caracterizado primeiramente em camundongos, como a região do genoma cujos genes eram determinantes no reconhecimento de tecidos e na aceitação ou rejeição de enxertos alogênicos (1–3). Característico de todos os mamíferos, os genes do MHC, no entanto, têm uma função mais ampla: codificam moléculas associadas à indução e regulação da resposta imunitária, principalmente adaptativa, influenciando, além do discernimento entre o que é próprio e não ao organismo, a susceptibilidade e progressão de doenças autoimunes, infecciosas e neoplásicas (1,2,4–7).

Em humanos, o MHC compreende uma região de aproximadamente 4 Megabases (Mb) no braço curto do cromossomo 6, localizado especificamente em 6p21.3. Essa região apresenta cerca de 224 genes (8,9), em sua maioria associados com o controle das respostas imunitárias (Figura 1). O MHC é considerado como uma das regiões com maior densidade gênica e a mais variável do genoma da maioria dos vertebrados, incluindo o homem (4–7).

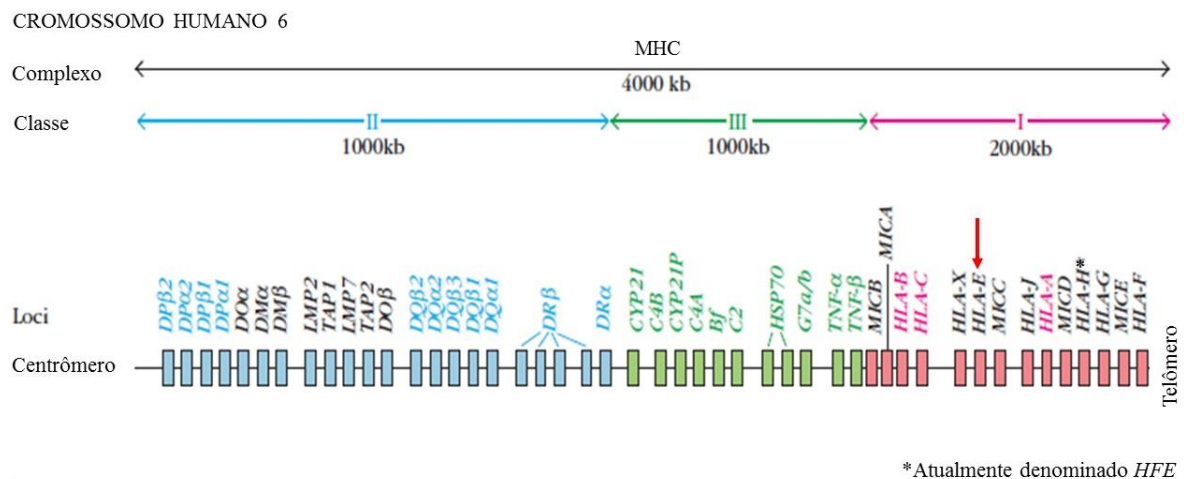


Figura 1. Esquema genômico detalhado da região do Complexo Principal de Histocompatibilidade (MHC) humano, localizado no braço curto do cromossomo 6 (10). Em azul está representada a região dos genes de classe II, na porção mais centromérica do MHC. Os genes escritos em azul representam os genes *HLA* de classe II. Em verde estão representados os genes de classe III, que codificam citocinas, moléculas do sistema complemento, entre outros. Na porção mais telomérica do MHC estão representados os genes de classe I em rosa; os genes escritos em rosa representam os genes *HLA* de classe I clássicos. Em destaque, apontado pela seta vermelha, está o gene *HLA-E*, alvo do presente estudo. Adaptado de Kindt et al., 2007.

O MHC pode ser dividido didaticamente em três regiões, denominadas classe I, II e III (Figura 1) (10,11). Os principais genes de classe I (região mais telomérica) e II (mais

centromérica) compreendem o sistema HLA, detalhado mais adiante, cujos genes codificam glicoproteínas de membrana responsáveis pela apresentação de antígenos proteicos. Outros genes presentes na região de classe I são *HFE* (metabolismo de ferro) e genes *MIC* (similares aos genes *HLA* de classe I com função em células epiteliais e estresse celular). Outros genes de classe II, além dos que codificam as cadeias α e β das moléculas HLA-DP, HLA-DQ, HLA-DR, HLA-DO e HLA-DM, são *LMP* (envolvidos na montagem do proteossoma) e *TAP* (transportador associado ao processamento de antígenos), ambos envolvidos no processamento intracelular de antígenos e montagem de moléculas HLA de classe I, principalmente. Os genes presentes na região de classe III (região intermediária), codificam moléculas importantes ao sistema imunológico, como moléculas do sistema complemento, hormônios, citocinas (*TNF* e *LTA*) e proteínas de choque térmico (4–7,10,11).

1.2 Sistema HLA

As moléculas HLA foram descobertas por métodos sorológicos como resultado da procura de moléculas de superfície celular em um indivíduo, que seriam reconhecidas como estranhas em outro indivíduo (12). Nos camundongos anteriormente citados, homólogos dessas moléculas seriam os fatores determinantes na aceitação/rejeição do enxerto e, portanto, as principais moléculas de histocompatibilidade, característica que originou a nomenclatura do MHC (3). Nos seres humanos essas moléculas foram nomeadas como antígenos leucocitários humanos ou HLA (do inglês, *Human Leukocyte Antigen*) e desempenham função central na resposta imune contra antígenos de origem proteica, os quais são apresentados na superfície celular ligados a essas moléculas. Os complexos HLA-peptídeo interagem com células efectoras do sistema imune (células T e células *Natural Killer* - NK), atuando diretamente no reconhecimento do que é próprio ou estranho ao indivíduo e induzindo, no segundo caso, resposta imune apropriada (13,14).

Os genes que codificam as moléculas HLA estão presentes nas regiões de classe I e de classe II do MHC humano (Figura 1); que codificam, respectivamente, as moléculas HLA de classe I e classe II. Os genes *HLA* de classe I se subdividem em genes clássicos (*HLA-A*, *HLA-B* e *HLA-C*), expressos em praticamente toda célula somática, e não clássicos (*HLA-G*, *HLA-E* e *HLA-F*), expressos em alguns tecidos ou de forma constitutiva em baixos níveis (15). Os genes *HLA* de classe I clássicos codificam moléculas que participam da apresentação de antígenos de origem endógena ou processados intracelularmente, e apresentam estes antígenos aos linfócitos T CD8⁺ (células T citotóxicas). Já os genes *HLA* de

classe I não clássicos codificam moléculas de estrutura semelhante, porém sua principal função é a imunomodulação (15–19). Por sua vez, os genes *HLA* de classe II (*DRA*, *DRB1*, *DQB1*, entre outros) codificam moléculas que apresentam antígenos principalmente de origem exógena (ex. bacterianos) para células T CD4⁺ (células T auxiliares) e são expressos em alguns tipos celulares como macrófagos, linfócitos B e células dendríticas.

As moléculas HLA de classe I e II diferem um pouco na sua estrutura (Figura 2), mas ambas têm como principal função a ligação e apresentação de antígenos interagindo com os receptores de células T (TCR), tendo assim papel central na imunidade adaptativa. As moléculas HLA tem formato de taça ancorada à membrana das células. A parte extracelular das moléculas HLA incluem os domínios $\alpha 1$ e $\alpha 2$, em classe I, e $\alpha 1$ e $\beta 1$, em classe II, que formam a fenda de ligação do peptídeo formada por um assoalho de 8 fitas antiparalelas do tipo β e uma borda em α -hélice (região mais variável da molécula). Os resíduos que formam essa fenda criam um motivo ou padrão, específico para cada alelo HLA, que define qual repertório de peptídeos poderá ser ligado. Os demais domínios extracelulares são semelhantes a Imunoglobulinas (Ig) e interagem com co-receptores dos linfócitos T (10,11,13).

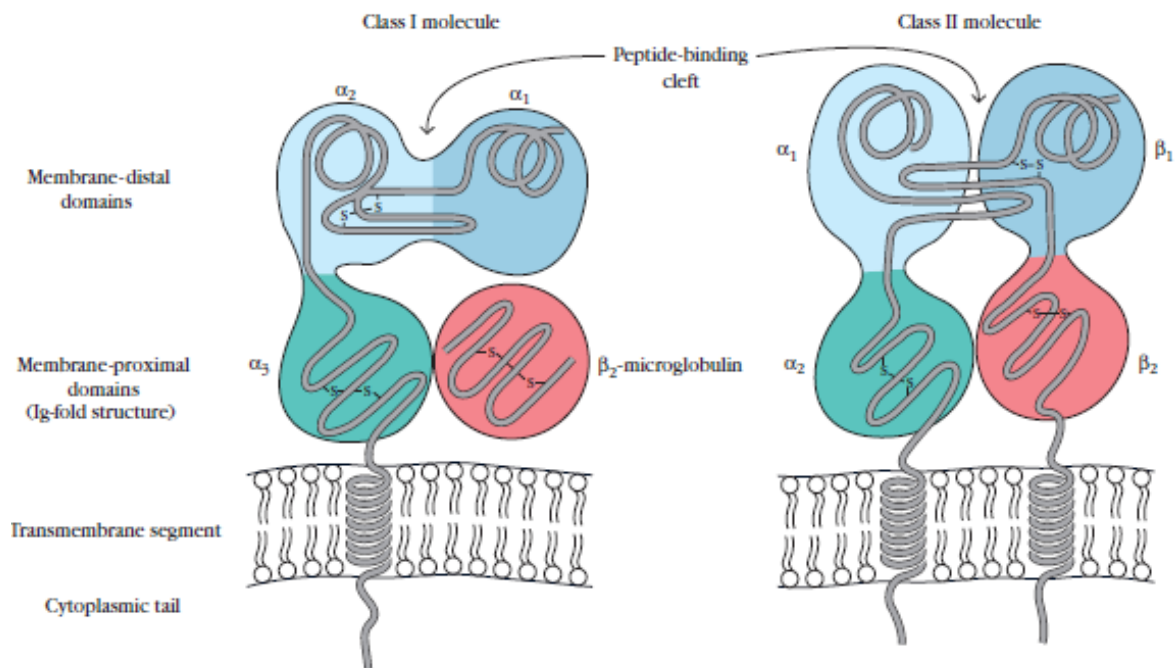


Figura 2. Estrutura das moléculas HLA de classe I (esquerda) e classe II (direita) (10). A molécula HLA de classe I é formada por domínios $\alpha 1$ e $\alpha 2$, que formam fenda de ligação ao peptídeo (em azul), $\alpha 3$ (em verde) que interage com molécula de $\beta 2$ -microglobulina ($\beta 2M$, em rosa, codificada no cromossomo 15), região transmembrana e cauda citoplasmática. A molécula HLA de classe II é formada por domínios $\alpha 1$ e $\alpha 2$ que interagem com $\beta 1$ e $\beta 2$ (codificados por um par de genes *HLA* de classe II, A e B), região transmembrana e cauda citoplasmática. Os domínios $\alpha 1$ e $\beta 1$ formam a fenda de ligação de peptídeo na molécula HLA de classe II (em azul).

A princípio, as moléculas de classe I apresentam antígenos endógenos (Figura 3) e as de classe II, extracelulares (Figura 4), mas há também apresentação cruzada de peptídeos. No citoplasma, o proteossoma degrada proteínas da própria célula ou de patógenos intracelulares, gerando peptídeos que são transportados para o retículo endoplasmático (RE) com o auxílio de proteínas transportadoras de antígeno (TAP). As moléculas HLA de classe I também são encaminhadas ao RE onde, com o auxílio de chaperonas, entram em contato com os peptídeos e então os complexos HLA- β 2M-peptídeo são montados e direcionados à superfície celular (Figura 3). Os complexos HLA de classe II-peptídeo, por sua vez, são montados em vesículas endossômicas, nas quais as proteínas fagocitadas são degradadas por reação enzimática gerando peptídeos. As moléculas de HLA de classe II são direcionadas a essas vesículas a partir do RE, ligadas a moléculas acessórias (CLIP) e depois, uma vez montado um complexo estável, este também é direcionado à superfície celular (Figura 4) (10,11).

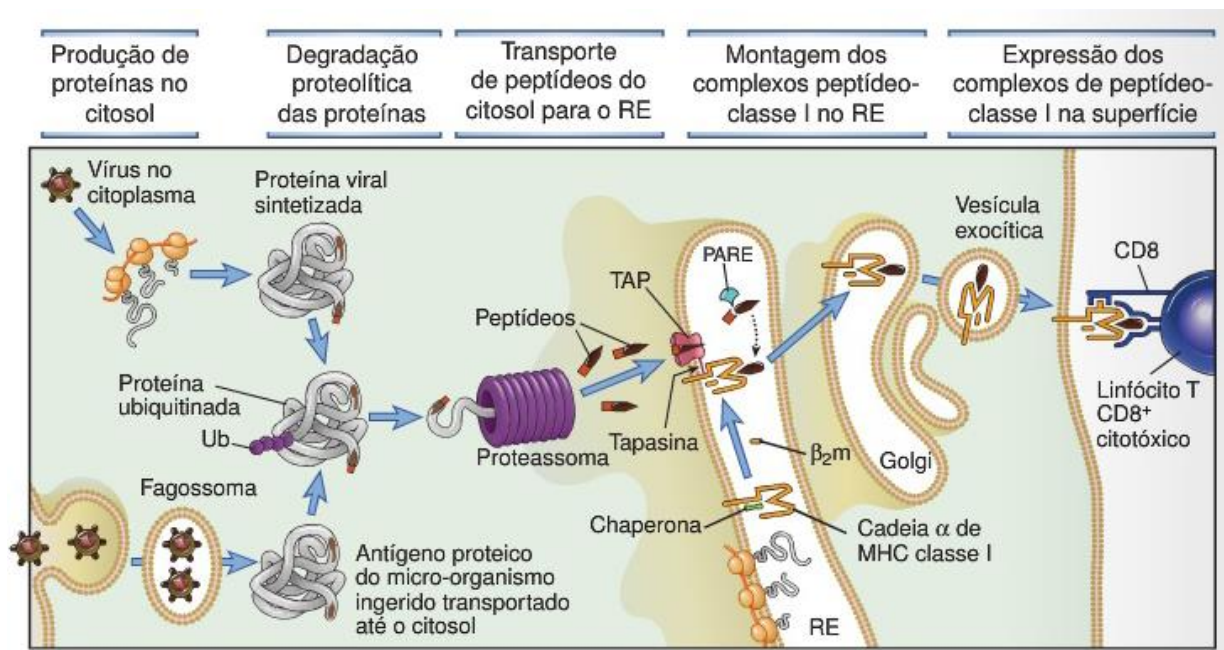


Figura 3. Apresentação de antígenos endógenos via HLA de classe I (11). Proteínas ubiquitinadas presentes no citoplasma são degradadas pelo proteossoma gerando peptídeos menores. Esses peptídeos são captados pelo retículo endoplasmático por meio de TAP. No retículo endoplasmático é montada a molécula de HLA de classe I, com auxílio de chaperonas que estabilizam a molécula HLA até ser montado complexo estável de HLA- β 2M-peptídeo, exportado posteriormente para a membrana celular por meio do complexo de Golgi.

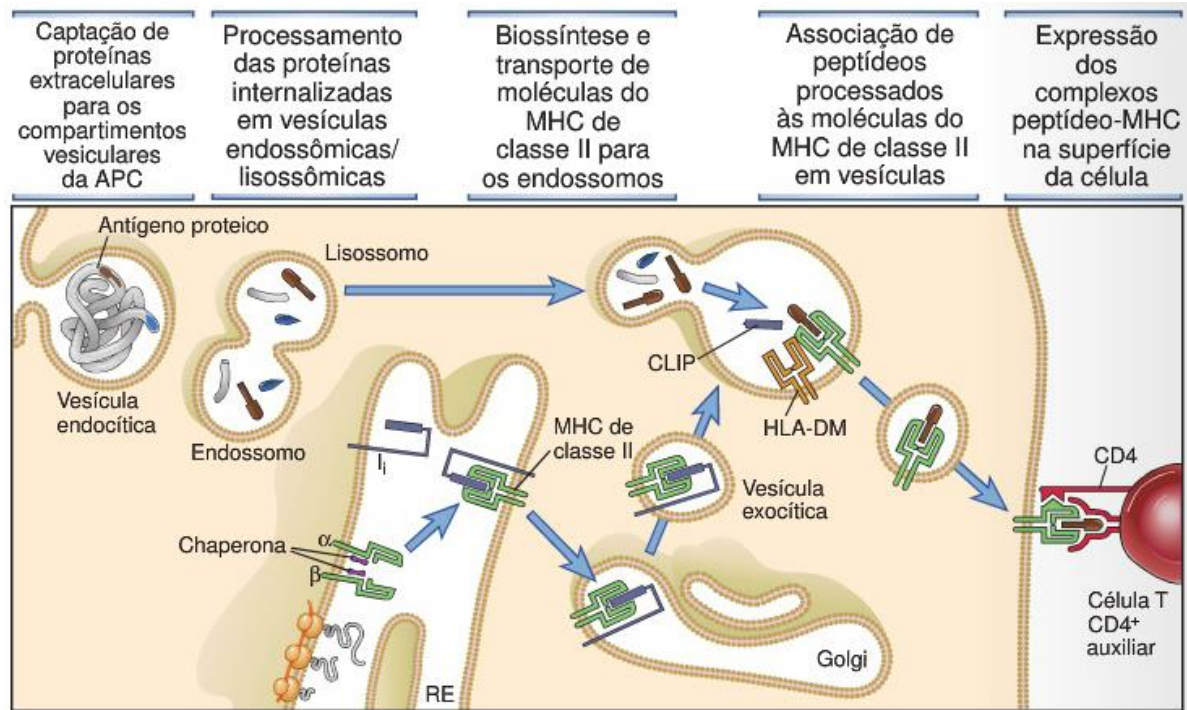


Figura 4. Apresentação de antígenos via HLA de classe II (11). Proteínas extracelulares são fagocitadas em vesículas endocíticas que se combinam com vesículas lisossômicas, cujas enzimas degradam as proteínas em peptídeos menores. Moléculas de HLA de classe II ligadas a CLIP, que protege o sítio de ligação a peptídeo e estabiliza temporariamente a molécula HLA de classe II, são transportadas do retículo endoplasmático por meio do complexo de Golgi para os endossomos, onde ocorre a ligação HLA-peptídeo, sendo posteriormente encaminhadas à membrana celular.

Relacionada a essa função de ligação antigênica, está a alta variabilidade dos genes *HLA*, uma de suas principais características (20), o que proporciona de uma maneira geral também alta diversidade das moléculas HLA. Essa alta diversidade é resultante principalmente de mutações na região codificadora do gene (principalmente na região que codifica a fenda de ligação de peptídeos), que geraram polimorfismos cuja manutenção foi privilegiada. Por codificarem moléculas que devem apresentar um repertório de antígenos aos linfócitos T e modular a resposta imune contra patógenos, principalmente, essa alta variabilidade dos genes *HLA* é uma vantagem em termos evolutivos. Ou seja, com uma diversidade maior de moléculas, o repertório de peptídeos que podem ser apresentados aumenta, proporcionando melhor adaptação à população no combate de diferentes patógenos (4,21).

Nesse mesmo contexto, existem mecanismos de seleção característicos dos genes *HLA* em termos evolutivos: a chamada “vantagem do heterozigoto” e “seleção dependente de frequência”. No primeiro caso, indivíduos heterozigotos seriam capazes de apresentar maior número de peptídeos diferentes, em relação aos homozigotos (22). Além disso, alelos raros

teriam uma vantagem seletiva de modo a aumentar sua frequência, na ideia de que seriam novos alelos com capacidade de apresentação de novos repertórios de antígenos (23). Em termos clínicos, porém, esse cenário de diversidade de alelos *HLA* caracteriza um obstáculo para os transplantes, por exemplo, pois a compatibilidade entre doador e receptor, necessária para uma boa aceitação do enxerto, é dificultada em vários aspectos envolvendo regulação e ativação do sistema imunológico.

A ativação e regulação do sistema imunológico está intimamente relacionado à expressão de moléculas HLA e à apresentação de antígenos. Desde a maturação dos linfócitos T no timo até ativação da resposta imune contra patógenos extra e/ou intracelulares, a interação entre as moléculas HLA apresentando diferentes antígenos, próprios ou não, com os receptores de células T (TCR) (Figura 5), e com outros receptores ativadores ou inibitórios presentes em células imunocompetentes, tem papel primordial na regulação de secreção de citocinas, ação de citotoxicidade, produção de anticorpos e tolerância imunológica. Dependendo do peptídeo apresentado, o complexo HLA presente na superfície celular pode interagir com receptores de linfócitos T e células NK com maior ou menor afinidade, que irão modular a resposta imunológica. Além disso, o perfil de expressão da molécula HLA (quanto é produzido e onde se expressa) também é um aspecto relevante a ser considerado, e as regiões regulatórias do gene ganham importância neste sentido.

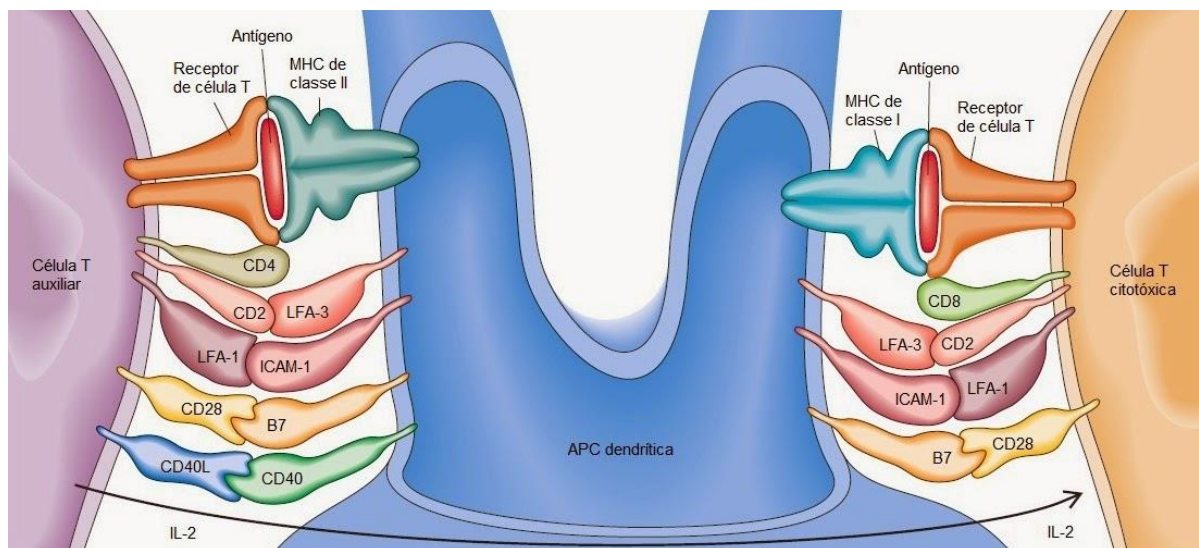


Figura 5. Interação da célula T com célula apresentadora de antígeno (APC). As células T auxiliares (CD4+) reconhecem antígenos associados a HLA de classe II e as células T citotóxicas (CD8+), a de HLA de classe I. Nesse microambiente, a liberação de citocinas como IL-2 serve como sinal acessório a ativação das células imunes. ICAM: *intercellular adhesion molecule*; LFA: *leukocyte function-associated antigen*. Adaptado de Kindt et al, 2007 (24).

Um dos primeiros mecanismos que regulam a expressão gênica ocorre por meio da ligação de fatores de transcrição na região regulatória promotora desses genes, que definem o perfil de expressão transcricional de cada *locus HLA*. Os genes *HLA* de classe I clássicos são expressos constitutivamente em todas as células nucleadas, enquanto os de classe I não clássicos tem um perfil de expressão mais limitado, por exemplo. Porém, de forma geral, esses genes podem ser induzidos por *interferons* e $\text{TNF-}\alpha$. Os genes de classe II, por sua vez, são expressos constitutivamente em níveis baixos, mas são fortemente induzidos por ativação das células apresentadoras de antígeno (APC), estimulados por $\text{IFN-}\gamma$ e sinais do receptor Toll-like. Em outros tipos celulares diferentes de APCs, genes *HLA* de classe II podem ser expressos durante processos inflamatórios (10,11).

Embora haja diferenças, existem sítios conservados na região promotora proximal de genes *HLA* de classe I e de classe II relacionados à expressão desses genes. Os principais fatores de transcrição envolvidos na expressão de genes *HLA* são CIITA (*class II trans-activator*) e RFX (*regulatory fator X*), responsáveis pela expressão constitutiva e induzida por $\text{IFN-}\gamma$, que se ligam em sítios conservados, como mostra a Figura 6 (10,25). Outros elementos presentes nessa região e combinações entre eles, distinguem o perfil transcricional de cada *locus HLA*. Além dessa diferença entre genes, mesmo sabendo que os alelos *HLA* são expressos de forma codominante, o nível de expressão de um mesmo tipo de *HLA* pode ser diferente dependendo dos elementos estimuladores ou silenciadores associados aos seus diferentes alelos.

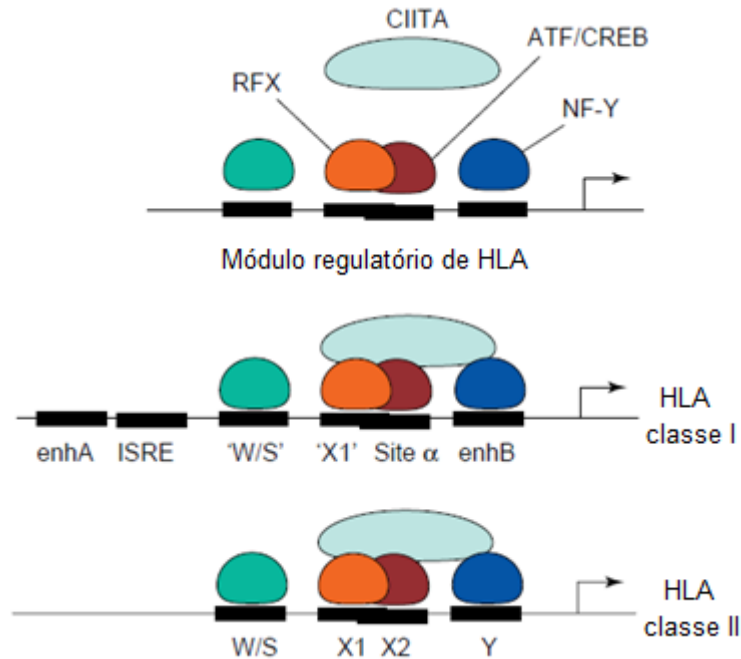


Figura 6. Elementos regulatórios dos genes *HLA* de classe I, de classe II e elementos regulatórios conservados entre as duas classes. CIITA exerce sua atividade principalmente pela interação com elementos X1–X2/site α –Y box e suas proteínas de ligação. Os sítios denominados de *enhancer A* (enhA), ISRE, W/S, X1, X2/site α e *enhancer B* (enhB)/Y box são sequências de DNA conservadas na região promotora proximal dos genes *HLA* de classe I e II. Adaptado de Peter J. van den Elsen et al. 1998 (25).

Todos esses aspectos são importantes para estudo dos genes *HLA*, cujo conhecimento pode auxiliar no entendimento de processos patológicos como autoimunidade, rejeição de transplantes, abortos espontâneos, susceptibilidade e progressão de tumores e processos infecciosos, principalmente considerando a função chave das moléculas *HLA* na montagem das respostas imunológicas adaptativas e regulação fina destas respostas. Também têm importância em estudos evolutivos e antropológicos devido sua alta variabilidade e pressões seletivas específicas. Dessa forma, torna-se crucial entender como as moléculas de *HLA* funcionam e como as variações nos genes que as codificam podem influenciar sua ação e sua regulação.

Embora a principal função associada às moléculas *HLA* seja a apresentação de antígenos aos linfócitos T, o subgrupo de genes *HLA* de classe I não clássicos têm um padrão de expressão, variabilidade e função diferenciados. As moléculas *HLA* de classe I não clássicas, *HLA-E*, *HLA-F* e *HLA-G*, podem apresentar antígenos, porém com um propósito diferente das moléculas clássicas. Interagem com outros receptores que não estão restritos às células T e modulam a resposta imunológica no reconhecimento do que é próprio ou não ao indivíduo (19,26,27). O gene *HLA-E*, foco desse trabalho, faz parte desse subgrupo, cujas características serão apresentadas a seguir.

1.3 HLA-E

O gene *HLA-E* faz parte dos genes HLA de classe I não clássicos, assim como *HLA-F* e *HLA-G*, que foram descobertos entre os anos de 1987 e 1990 (28–30). A estrutura da molécula por eles codificada é similar à estrutura das moléculas de HLA de classe I, mostrada na Figura 2 e na Figura 7.



Figura 7. Estrutura molecular do complexo HLA de classe I/peptídeo. Em verde está representada a molécula HLA, interagindo com $\beta 2M$ (em azul). Na parte superior, está a fenda de ligação com peptídeo ligado (em azul). (31).

A principal função das moléculas HLA não clássicas, até então descrita, é a modulação da resposta imune e tolerância imunológica (15–19). De maneira geral, essas moléculas estão associadas à regulação do sistema imune por meio da interação com receptores inibitórios (27). Por exemplo, o gene *HLA-G* é expresso principalmente em células do trofoblasto extraviloso durante a gestação, atuando na modulação do sistema imune materno para que este não reaja contra o feto semi alogênico (32–34). O gene *HLA-F* parece ser expresso e exercer sua função prioritariamente intracelular, auxiliando o funcionamento das demais genes *HLA* de classe I (35–37).

A baixa variabilidade nesses genes em comparação com os demais genes *HLA* é uma característica marcante, assim como a expressão em menores níveis e/ou distribuição restrita a alguns tecidos específicos (como *HLA-G*) (38–41). Essas características sugerem que esses genes possuem aspectos funcionais e mecanismos de regulação da expressão gênica mais específicos do que os presentes nos genes clássicos (38,39) o que sugere que a função das moléculas de classe I não clássicas é determinante no controle da resposta imune.

O gene *HLA-E*, objeto de estudo desse trabalho, mesmo entre os genes não clássicos, tem particularidades que se destacam como: (a) localização e padrão de variabilidade do gene; (b) estrutura da molécula codificada; (c) padrão de regulação e seleção; todos relacionados (d) à função da molécula *HLA-E* no organismo, revisado a seguir.

O gene *HLA-E* está localizado centralmente na região do MHC de classe I, entre dois dos genes mais variáveis do genoma humano, *HLA-A* e *HLA-C* (Figura 8). Foi descrito pela primeira vez em 1988 por Koller et al. como um gene diferente dos demais genes *HLA* de classe I até então descritos (29) pelos seguintes aspectos: (i) foram achados três elementos *Alu* (elemento repetitivo de transposição (42,43)) associados ao *HLA-E*: um na região promotora, aproximadamente 261 pb antes do códon de início da tradução (ATG), um no íntron 5 e outro na região 3' não traduzida (3'NT); (ii) elementos comuns na região promotora dos genes *HLA* de classe I divergem no *HLA-E* e (iii) sua região codificadora possui baixa variabilidade em comparação com os demais genes *HLA*, além de haver alterações em resíduos de *HLA-E* que seriam conservados em outras moléculas *HLA* de classe I (29).

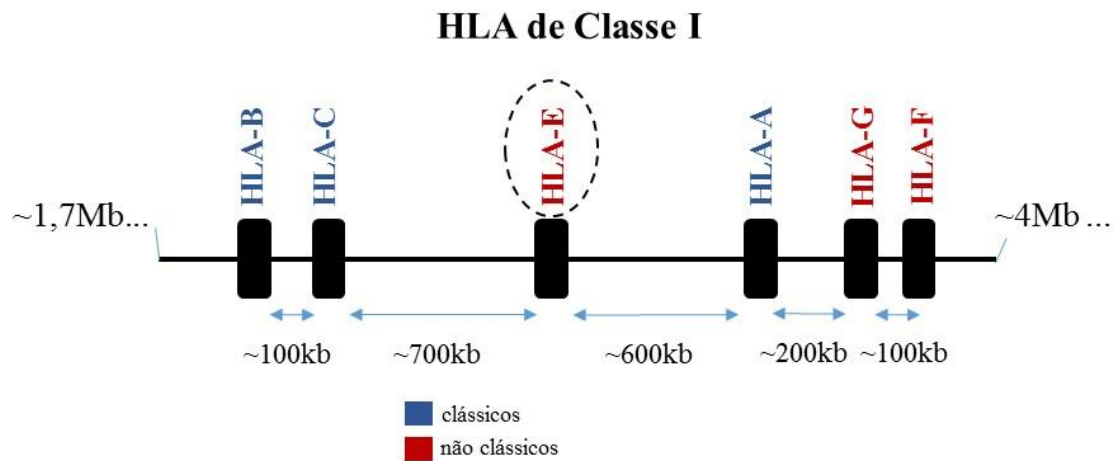


Figura 8. Representação gráfica da região de classe I do MHC humano (mais telomérica) indicando os genes *HLA* clássicos e não clássicos de classe I e a distância aproximada em kilobases (kb) entre eles. Escritos em azul estão os genes *HLA* de classe I clássicos; em vermelho, os não clássicos. Em destaque, circulado, está o gene *HLA-E*. Imagem própria baseada em Janeway et al, 2001 (44) e http://labs.icb.ufmg.br/lbcd/prodabi5/grupos/hugo_deise_liza/pagina/mapa.gif.

A variabilidade nas regiões codificadora e regulatórias do gene, influenciam a molécula codificada e o padrão de expressão do gene, respectivamente. A organização e função dessas regiões do *HLA-E* parece ser uma particularidade do gene entre os demais genes de *HLA* classe I.

1.3.1 Região codificadora do gene *HLA-E*

Semelhante à estrutura dos demais genes *HLA* de classe I, a região codificadora do *HLA-E* possui oito éxons e sete íntrons (Figura 9) e segue também o mesmo padrão com o éxon 1 codificando a sequência líder; éxons 2 e 3 codificam os domínios $\alpha 1$ e $\alpha 2$, onde se ligará o peptídeo; éxon 4 codifica domínio $\alpha 3$, que interage não covalentemente com

molécula de β 2-microglobulina (β 2M); éxon 5 codifica região transmembrana; éxons 6 e 7 codificam cauda citoplasmática; e éxon 8 faz parte da região 3' não traduzida (3'NT) do RNA mensageiro (mRNA) (29,45) (Figura 9).

Por outro lado, diferente dos genes *HLA* de classe I clássicos e semelhante aos não clássicos, o *HLA-E* codifica uma cauda citoplasmática relativamente menor, devido a uma alteração do tipo *frameshift* no éxon 7 (29,45,46). Além disso, outra divergência em relação a classe I é a substituição de dois aminoácidos codificados pelos códons 143 e 147 por resíduos Serina (natureza polar neutra), em vez de resíduos Treonina (natureza apolar neutra) e Triptofano (natureza apolar), respectivamente, considerados como conservados em outras moléculas HLA de classe I; estas substituições influenciam a fenda de ligação do peptídeo (45). Isso também acontece e é conservado nas moléculas homólogas a HLA-E em camundongos (molécula Qa-1), ratos (molécula RTBM-1) e primatas não humanos como orangotango (molécula Popy-E) (45).

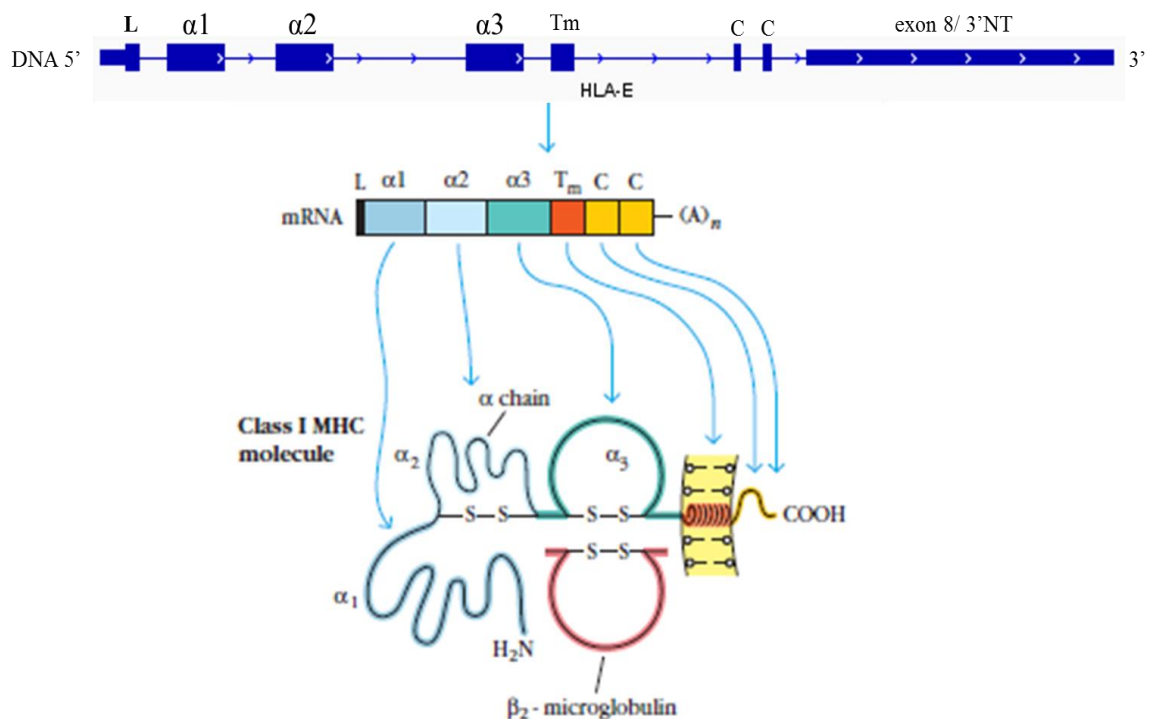


Figura 9. Esquema correlacionando estrutura do gene *HLA-E* com a molécula codificada. Acima, em azul escuro está representada a estrutura do gene *HLA-E*, com 8 éxons (caixas mais altas), 7 íntrons (linhas) e as regiões regulatórias (caixas mais baixas). Abaixo, o RNA mensageiro transcrito com as respectivas porções equivalentes na molécula. L: sequência líder; α 1: domínio alfa 1; α 2: domínio alfa 2; α 3: domínio alfa 3; Tm: região transmembrana; C: cauda citoplasmática. Adaptado de Kindt et al, 2007 (10).

Mesmo presente entre dois dos genes mais variáveis do genoma humano, *HLA-A* e *HLA-C* (Figura 8), o *HLA-E* é considerado historicamente como o mais conservado da classe. Como mostra a Tabela 1, a quantidade de sequências diferentes (alelos) descritas para

o *HLA-E* é uma das menores de classe I. Não apenas a quantidade de alelos, mas também a diversidade de moléculas diferentes que esses alelos codificam mostram o quanto esse gene é pouco variável. Entre os não clássicos, os dados depositados no banco de dados IMGT/HLA, versão 3.27.0, apontam para a menor variabilidade de *HLA-F*, porém, estudo recente encontrou novas mutações que caracterizam novos alelos ainda não depositados neste banco de dados (47).

Tabela 1. Número de alelos identificados para os genes de classe I (IMGT/HLA, release 3.27.0) (20).

CLASSE I CLÁSSICOS			CLASSE I NÃO CLÁSSICOS		
Gene	Alelos	Moléculas	Gene	Alelos	Moléculas
<i>HLA-A</i>	3830	2703	<i>HLA-E*</i>	25	8
<i>HLA-B</i>	4647	3408	<i>HLA-F</i>	22	4
<i>HLA-C</i>	3382	2391	<i>HLA-G*</i>	53	18

*Alguns desses alelos foram descritos em trabalhos prévios do grupo (48–52).

O baixo número de alelos de região codificadora do gene *HLA-E*, se comparado com outros genes *HLA*, foi observado para o *HLA-E* desde os primeiros estudos realizados (53–55). Atualmente, considerando-se todos os alelos conhecidos, o *HLA-E* codifica apenas oito proteínas diferentes. Isto ocorre porque a maioria das variações genéticas conhecidas ocorrem em íntrons ou são mutações sinônimas em éxons. Além disso, é importante destacar que dessas, somente duas proteínas, codificadas pelos alelos do tipo E*01:01 e E*01:03, são frequentemente encontradas em populações mundiais (53,54,56–63). Essas duas moléculas principais diferem estruturalmente em apenas um aminoácido, codificado pelo códon 107: Arginina (de natureza básica) no alelo E*01:01 e Glicina (de natureza apolar) em E*01:03, no domínio $\alpha 2$ da molécula de *HLA-E*. Essa substituição não está associada a fenda de ligação do peptídeo, mas foi sugerido que pode influenciar a interação com receptores, sendo que a molécula E*01:03 apresenta uma maior afinidade pelo peptídeo e maior nível de expressão na superfície celular (53,64).

Embora novos alelos e mutações não sinônimas tenham sido caracterizados, a alta conservação de *HLA-E* é marcante, com uma frequência de aproximadamente 50% para cada um dos grupos de alelos E*01:01 e E*01:03 nas diversas populações já analisadas, enquanto os demais aparecem como alelos raros (60,65–67). Comparando a sequência de *HLA-E* com

homólogo *MHC-E* de primatas não humanos, foi proposto que o alelo ancestral seria o *E*01:03* (40,56); por outro lado, dados recentes sugerem que essa dicotomia já existia desde estes outros primatas (67). De qualquer forma, é sugerido que a mutação que diferencia esses dois alelos principais aconteceu antes da expansão do *Homo sapiens* (56) e também mesmo antes da maioria dos polimorfismos que caracterizam as moléculas HLA clássicas, visto que ambos alelos *HLA-E* estão associados a diversos alelos de *HLA-A* e *HLA-B*, por exemplo (53). A manutenção desses alelos em alta frequência tem ocorrido por seleção balanceadora direcionada para as duas moléculas HLA-E comumente encontradas (67). Em outras regiões do gene, como nos íntrons, por outro lado, a baixa variabilidade parece ser fruto de seleção purificadora (62,67).

Em estudos anteriores, as análises com dados do projeto *1000 Genomes* fase 1 (68) e amostras brasileiras, e outro em duas populações africanas, mostraram uma baixa diversidade da região codificadora e 3'NT do *HLA-E* (59,60,62). Esse fenômeno foi observado não apenas em humanos, mas também entre espécies (40), especialmente primatas não humanos e murinos. Outros estudos também observaram a baixa variabilidade e conservação do gene (58), considerando segmento contínuo de éxons e íntrons, e recentemente da região promotora proximal (69).

Com base nessa conservação, acredita-se que a molécula HLA-E seja um importante elemento fundador do sistema imune nos mamíferos, considerando que esta conservação nos humanos e entre as espécies estaria relacionada com suas propriedades imunomodulatórias, de forma que a baixa variabilidade no gene garante que o mecanismo de tolerância com a qual a molécula está envolvida seja invariável (17,31,45,56,70–73). Cada aminoácido que estrutura a molécula, seja na fenda de ligação do peptídeo, seja os demais domínios, determinam quais peptídeos serão ligados e com qual afinidade, influenciando a expressão do complexo na superfície celular, além de influenciar a interação com receptores. Assim, a estrutura da região codificadora do *HLA-E* determina e está intimamente relacionada à função da molécula codificada e, portanto, variações nessa região podem alterar a função da molécula.

1.3.2 A molécula HLA-E: características e função

As primeiras evidências sobre a função do gene *HLA-E* ocorreram entre 1996 e 1998 quando estudos observaram que a molécula HLA-E liga-se a tipos restritos de peptídeos, que geralmente são derivados de outras moléculas HLA de classe I (19,70,74). Esse repertório

restrito está relacionado a duas características da molécula HLA-E, que a diferencia das demais moléculas HLA de classe I.

Uma dessas particularidades, característico não só de HLA-E mas também de suas homólogas em primatas não humanos e murinos, é a substituição de alguns aminoácidos na fenda de ligação ao peptídeo, discutido anteriormente, além da conservação da estrutura da molécula, principalmente entre as regiões dos domínios $\alpha 1$ e $\alpha 2$, que formam a fenda, diferentemente do que é observado nas moléculas HLA de classe I clássicas (45,53,59,60,62). Semelhante à homóloga Qa-1 em camundongos, a substituição dos resíduos 143 e 147 por Serina, em vez de Treonina e Triptofano, respectivamente, que são conservados nas demais moléculas HLA de classe I (45,53,70), parece afetar a ligação da extremidade C-terminal dos peptídeos na fenda de ligação. Além disso, na posição 67, somente HLA-E possui resíduo de Alanina, que também influencia a ligação de peptídeos à molécula (70). Dessa forma, foi caracterizado que a fenda de ligação de peptídeos na molécula de HLA-E, formada por assoalho de resíduos aromáticos, tem alta afinidade por peptídeos hidrofóbicos, o que restringe o repertório de peptídeos que se ligam à molécula de HLA-E (19). Semelhante às moléculas clássicas, HLA-E acomoda peptídeos com 9 aminoácidos (9-mer), e com menor estabilidade alguns peptídeos de 10 ou 11-mer. Foi então observado que peptídeos derivados da sequência líder ou sinal de outras moléculas HLA de classe I (HLA-A, -B, -C e -G) ligam-se com alta afinidade a HLA-E. Peptídeos derivados da sequência líder da própria HLA-E ou de HLA-F não têm essa afinidade, pois são sequências menores e não geram o mesmo peptídeo que as demais moléculas de classe I (70).

Outra característica que restringe o repertório de peptídeos que se ligam a HLA-E e que torna esta ligação altamente específica está relacionada com o fato de que a interação desses peptídeos hidrofóbicos (como VMAPRTLVL ou VMAPRTVLL) ocorre com pelo menos 5 resíduos âncora, sendo dois principais (os resíduos P2 e P9 do peptídeo) e os outros secundários (os resíduos P3, P6 e P7 do peptídeo) (19,31,70), ilustrado na Figura 10. A fenda de ligação ao peptídeo apresenta em geral 6 *pockets* que interagem com os resíduos do peptídeo (*pockets* A-F). No caso da HLA-E, todos são utilizados para interação de HLA-E com peptídeo. O *pocket* A interage com resíduo 1 do peptídeo (A-P1), B com P2, C com P6, D com P3, E com P7 e F com P9. Os *pockets* B e F são profundos e interagem com os dois resíduos âncoras principais do peptídeo. O resíduo P7 também fica profundamente na fenda e é uma importante âncora secundária (19,31) (Figura 10 e 11).

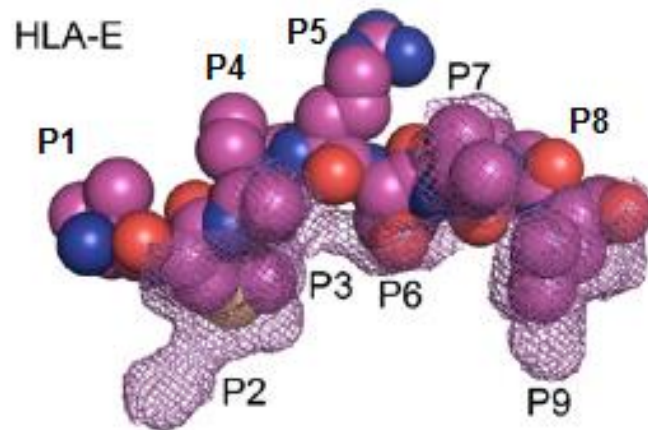


Figura 10. Peptídeos âncora na fenda de ligação ao peptídeo da molécula de HLA-E. Os principais resíduos P2 (*pocket B*) e P9 (*pocket F*) acomodam-se mais profundamente, assim como P7, enquanto que os demais resíduos secundários, P3 e P6, são mais superficiais. Adaptado de Zeng *et al*, 2012 (75).

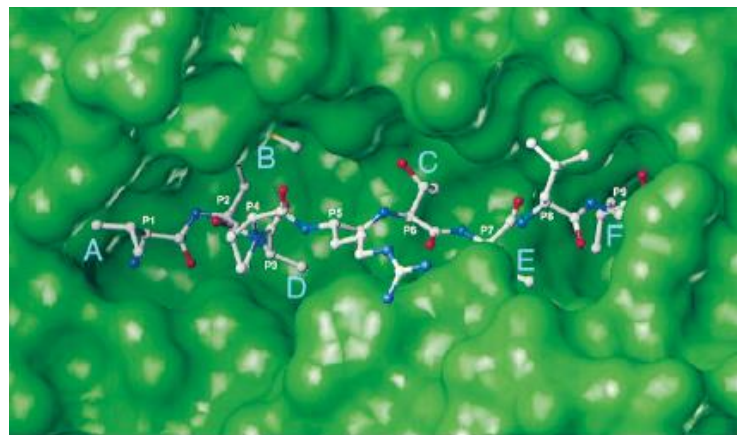


Figura 11. Interação de peptídeo hidrofóbico nonamérico na fenda de ligação na molécula de HLA-E. Os *pockets* de A a F e os resíduos (P1 a P9) do peptídeo estão representados (31).

Assim, primeiramente definiu-se o repertório de antígenos que se ligam a HLA-E como restrito aos peptídeos nonaméricos e hidrofóbicos derivados da sequência sinal de outras moléculas HLA classe I, com exceção da HLA-F e da própria HLA-E (19,31,70). A ligação desses peptídeos na molécula de HLA-E também está relacionada ao mecanismo TAP dependente, ou seja, a sequência líder é clivada da molécula de HLA de classe I pela peptidase sinal. A parte N hidrofílica permanece no citoplasma e é processada pelo proteossoma gerando peptídeos menores que então são carregados por TAP para o RE (Figura 12) (73,74).

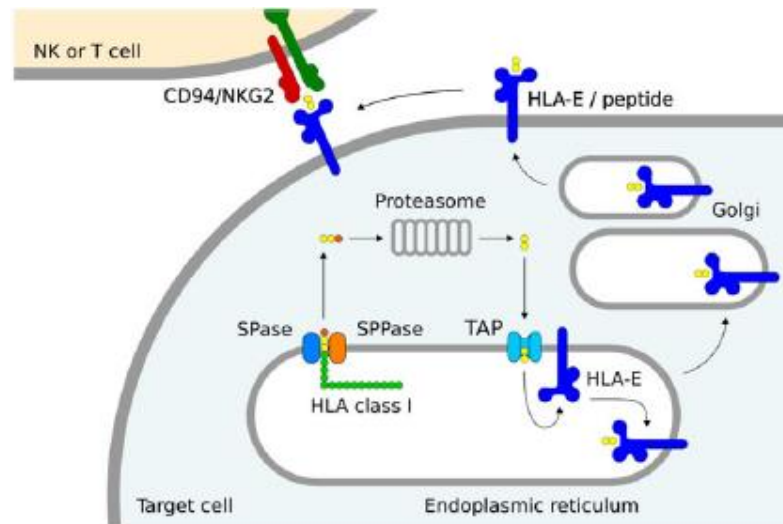


Figura 12. Montagem do complexo HLA-E/peptídeo dependente de TAP no retículo endoplasmático. Moléculas HLA de classe I recém sintetizadas são encaminhadas para o RE em cuja membrana o complexo de peptidase sinal cliva a sequência líder destas moléculas HLA. No citoplasma, a sequência líder passa pela maquinaria de processamento gerando peptídeos menores que entram no RE por meio de TAP, onde o complexo HLA-E/peptídeo poderá ser gerado. SPase: peptidase sinal (73).

Essa característica do repertório de peptídeos apontou para a principal função da molécula de HLA-E na imunovigilância, já que sua expressão na superfície celular depende da formação de complexos estáveis HLA-E/peptídeo, que por sua vez dependem da expressão dos demais genes HLA de classe I e da funcionalidade da maquinaria de processamento que envolve TAP. Observou-se que a expressão de HLA-E/peptídeo sinal confere proteção contra lise mediada por células NK (76), por meio da interação com receptores do tipo CD94/NKG2, presentes em células NK e alguns linfócitos T, principalmente do tipo inibitório. Ora, a presença de moléculas HLA-E na superfície celular é indicativo de que a expressão de MHC de classe I e maquinaria de processamento de antígenos estaria funcionando e, portanto, a atividade citolítica das células NK deve ser inibida (18,76,77).

A interação com os receptores CD94/NKG2 é do tipo chave-fechadura e envolve resíduos carregados: a porção ácida de CD94 interage com a porção básica do domínio $\alpha 1$ de HLA-E e a porção ácida do domínio $\alpha 2$ de HLA-E interage com porção básica de NKG2, que também interage com resíduos do peptídeo ligado à molécula de HLA-E (Figura 13). As principais interações entre o complexo HLA-E/peptídeo são com a molécula invariável de CD94, enquanto a sinalização intracelular é de responsabilidade das moléculas NKG2, porém a interação com esta porção do receptor influencia a afinidade final entre complexo HLA-E e receptor. As moléculas NKG2 dos tipos A e B são inibitórias (com cauda citoplasmática com motivos ITIM) e C, E e H ativadoras (com cauda citoplasmática menor associada a moléculas DAP-12 que possuem motivos ITAM) (78–80) (Figura 14).

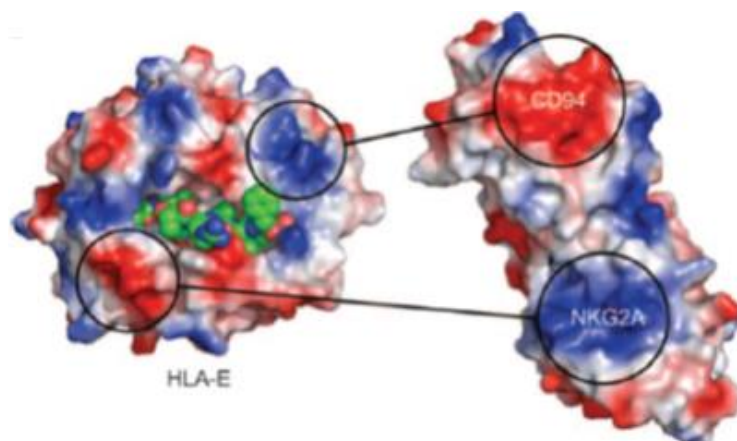


Figura 13. Interação HLA-E/peptídeo com receptor CD94/NKG2A. As porções em vermelho indicam regiões de caráter ácido e as porções em azul indicam caráter básico (72).

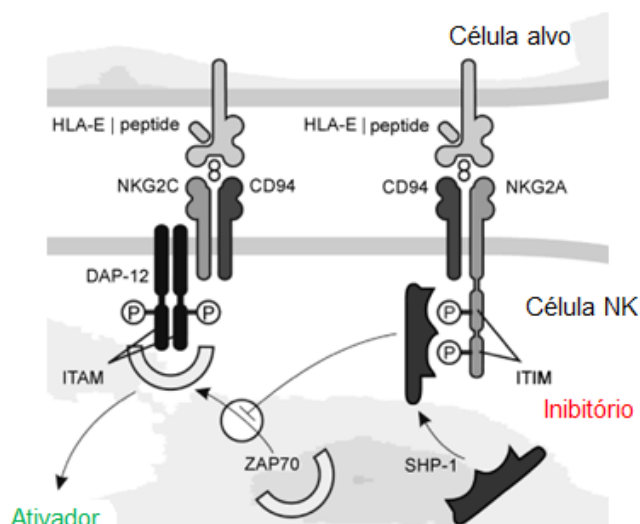


Figura 14. Interação HLA-E/peptídeo com os receptores de células NK: CD94/NKG2A (inibitório), à direita, e CD94/NKG2C (ativador), à esquerda. Os receptores inibitórios possuem cauda citoplasmática com motivos ITIM que ao serem fosforilados recrutam proteínas SHP-1 que atuam desfosforilando as proteínas ZAP70, envolvida na cascata de sinais ativadores. Os receptores ativadores estão associados a moléculas adaptadoras DAP-12 que possuem motivos ITAM e interagem com as proteínas ZAP70 desencadeando uma cascata de sinalização que resulta na ativação da célula NK. Adaptado de Iwasko *et al*, 2012 (81).

Alguns resíduos do peptídeo ligado à molécula de HLA-E também influenciam a afinidade do complexo com o receptor, pois como dito anteriormente, interagem com porções de NKG2 além de influenciarem a conformação estrutural do complexo HLA-E/peptídeo. Entre esses resíduos estão P5, P6 e P8, que ficam mais expostos na fenda de ligação de HLA-E (Figura 10) (31,64,82,83). A maior afinidade dos complexos HLA-E/peptídeo é com o receptor CD94/NKG2A inibitório, principalmente quando ligado a peptídeos derivados da sequência líder da HLA-G (que possui Fenilalanina em P8). Também há interação com receptor ativador CD94/NKG2C, e novamente a afinidade é significativamente maior com

peptídeos da HLA-G (31,82,83). Essa interação é importante, por exemplo, no cenário gestacional, quando a expressão de HLA-G é bem característica, como nas próximas sessões.

Com o avançar dos estudos, diferente do que originalmente se pensava, foi observado que o repertório de peptídeos que podem se ligar a HLA-E não é tão restrito, considerando condições não fisiológicas. Além dos peptídeos derivados da sequência líder das outras moléculas HLA de classe I, outros diferentes peptídeos próprios e não próprios podem estabilizar a molécula de HLA-E e estimular sua expressão na superfície celular. Entre esses diferentes peptídeos estão os derivados de proteínas virais como gpUL40 (Citomegalovírus Humano, HCMV), BZLF-1 (Epstein-Barr, EBV), InflM (*influenza matrix protein*), vírus da hepatite C (*core protein*, HCV) e proteína gag (vírus da Imunodeficiência Humana, HIV). Além disso, peptídeos bacterianos como da *Salmonella enterica* e *Mycobacterium tuberculosis*, e peptídeos originados de proteínas próprias do corpo como aquelas relacionadas ao estresse celular e ao choque térmico (Hsp60), isoformas de peroxiredoxina 5 (Prdx5D2 e Prdx5D2,3), ABC MRP7 (*ATP-binding cassette transporter*, proteína associada a resistência multidrogas), variantes de gliadina e TCR V β 1 e V β 2 (72,73,79,84) também poderiam se ligar a HLA-E.

A ligação de peptídeos derivados de patógenos, como HCMV e HIV, também formam um complexo com afinidade pelo CD94/NKG2A, o que caracteriza um mecanismo de escape imunológico das células infectadas. Outros peptídeos, como o derivado de Hsp60, não tem afinidade por esses receptores e permitem a ação das células NK contra as células em estado de estresse. No entanto, no cenário de infecção crônica, quando HLA-E pode ligar peptídeos derivados de alguns patógenos, como descrito anteriormente, foi observado a interação desta HLA-E com outro receptor em células específicas, que indicou uma segunda função da molécula HLA-E.

Através de estudos posteriores descobriu-se que HLA-E, além de regular a função das células NK, modulando a imunidade inata, também pode atuar na imunidade adaptativa apresentando peptídeos a células T CD8⁺ restritas a HLA-E, também chamadas de NKT ou NK-CTL, em resposta a patógenos e estresse celular, mesmo esta não sendo sua função primária (72,79,84,85). Assim sabe-se hoje que HLA-E é ligante para receptores do sistema imune inato e adaptativo e possui, portanto, um papel duplo (Figura 15). A interação com receptores $\alpha\beta$ TCR nas células T CD8⁺ parece exercer um papel no combate a patógenos.

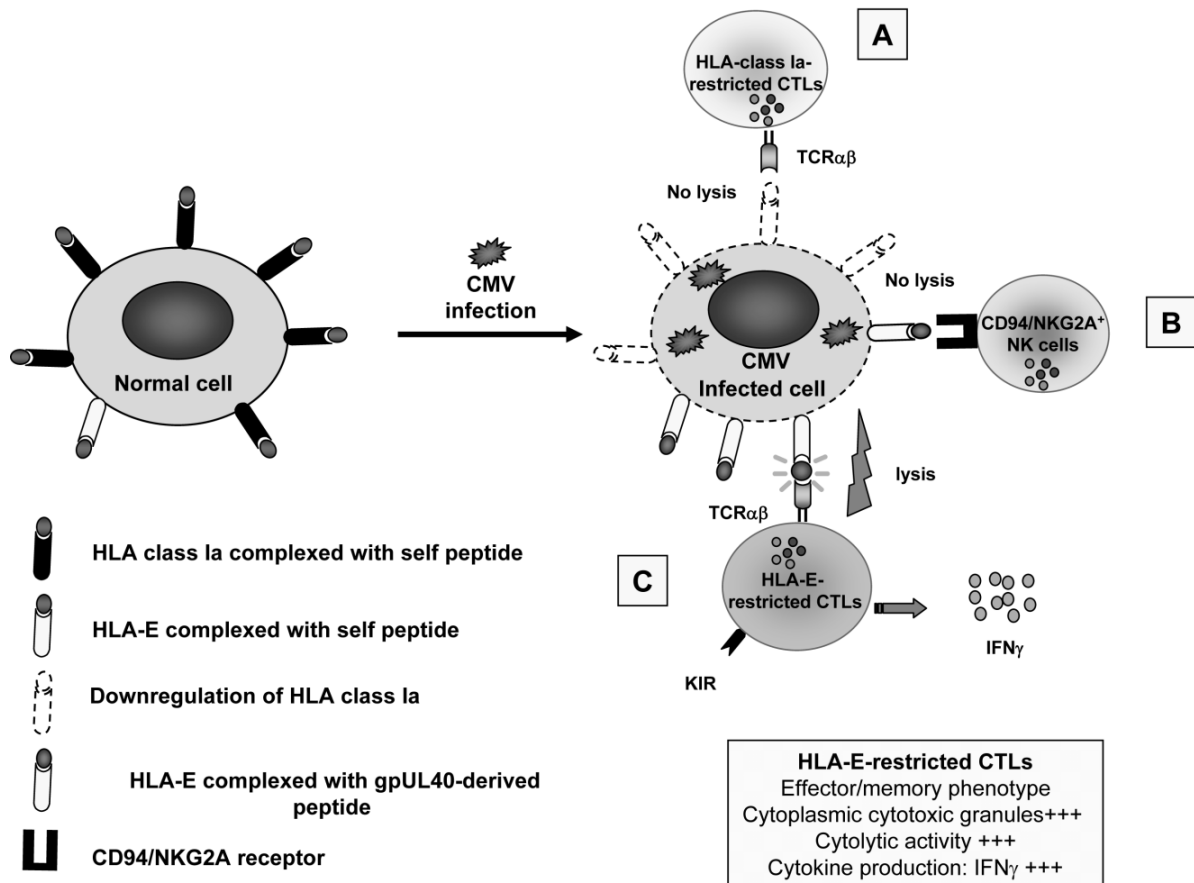


Figura 15. Papel duplo da molécula HLA-E. (A) Durante uma infecção crônica pelo Citomegalovírus humano (HCMV), como mecanismos de escape do patógeno, a expressão de moléculas HLA de classe I clássicas é inibida, e a célula infectada não é reconhecida pelos linfócitos T CD8⁺. (B) Em contrapartida, peptídeos virais são capazes de complexar com moléculas HLA-E que são expressas na superfície das células infectadas e reconhecidos pelo receptor inibitório CD94/NKG2A de células NK, cuja citotoxicidade é também inibida como aconteceria na imunovigilância. (C) Nesse cenário, HLA-E na superfície celular pode também ser reconhecido por linfócitos T CD8⁺ específicos que agem contra a célula infectada (79).

Seja apresentando preferencialmente peptídeos próprios derivados do HLA de classe I, seja outros antígenos, a molécula de HLA-E expressa na superfície celular interage principalmente com receptores CD94/NKG2 presentes nas células NK e T, e interage com o TCR, apresentando um papel tanto na imunidade inata quanto na adaptativa. A interação com receptores CD94/NKG2 inibitórios ou ativadores em células NK produzem transdução de sinal que modulam a atividade destas células, fazendo um controle dinâmico entre citotoxicidade e secreção de citocinas (32,71,73,86). Considerando a interação do tipo chave-fechadura, se houver uma mudança na conformação da molécula, a estabilidade do complexo bem como a afinidade pelo receptor pode ser alterada (para mais ou para menos) (64,82). Além da estrutura molecular, o padrão de expressão do gene *HLA-E*, aumentando ou diminuindo a quantidade de moléculas disponíveis nas células, é outro fator que pode influenciar sua função no organismo. Assim, estudar a variabilidade do gene, tanto

considerando a região codificadora como as regiões regulatórias, que determinam esses fatores, é primordial.

1.3.3 Regiões regulatórias do gene *HLA-E*

Além da alteração da molécula (mutações na região codificadora), outra forma de modificar a ação de uma proteína é alterar o seu padrão de expressão quanti e/ou qualitativamente, ou seja, aumentar ou diminuir o número de moléculas e/ou alterar o local de expressão (célula ou tecido) e quando esta expressão ocorre. Semelhante às moléculas HLA-A, HLA-B e HLA-C, a HLA-E é expressa em tecidos adultos e embrionários de forma ubíqua, principalmente nas células do sistema imune, porém em menores níveis que as moléculas clássicas (29,38,39,41). Sua expressão é estimulada em diferentes condições como após transplantes, durante infecções virais, doenças autoimunes e condições inflamatórias em geral devido a produção de citocinas pró inflamatórias como TNF- α , IL-1 β e principalmente IFN- γ (87).

No nível molecular, a regulação da expressão gênica pode ser transcricional, por meio de fatores de transcrição e estimuladores da transcrição (como citocinas e hormônios), ou pós transcricional, por exemplo, por meio da interação do mRNA de *HLA-E* com microRNAs. A região 3'NT de um gene, i.e., todo o segmento do mRNA após o códon de parada da tradução, é um segmento que pode influenciar a estabilidade do mRNA, além de ser o principal alvo para ligação de pequenos RNAs não codificadores, os chamados microRNAs. Essa ligação consiste em um mecanismo de regulação da expressão gênica que ocorre pós transcionalmente, o qual impede a tradução e/ou degrada o mRNA, regulando negativamente a expressão do gene (Figura 16). A interação com microRNAs ocorre por complementaridade de bases e pode, portanto, ser influenciada pela presença de polimorfismos na sequência de DNA e consequentemente na sequência do mRNA (88–90), ou na sequência do microRNA. No caso de *HLA-E*, em estudos anteriores foi constatado que sua região 3'NT possui baixa taxa de polimorfismo (59,60).

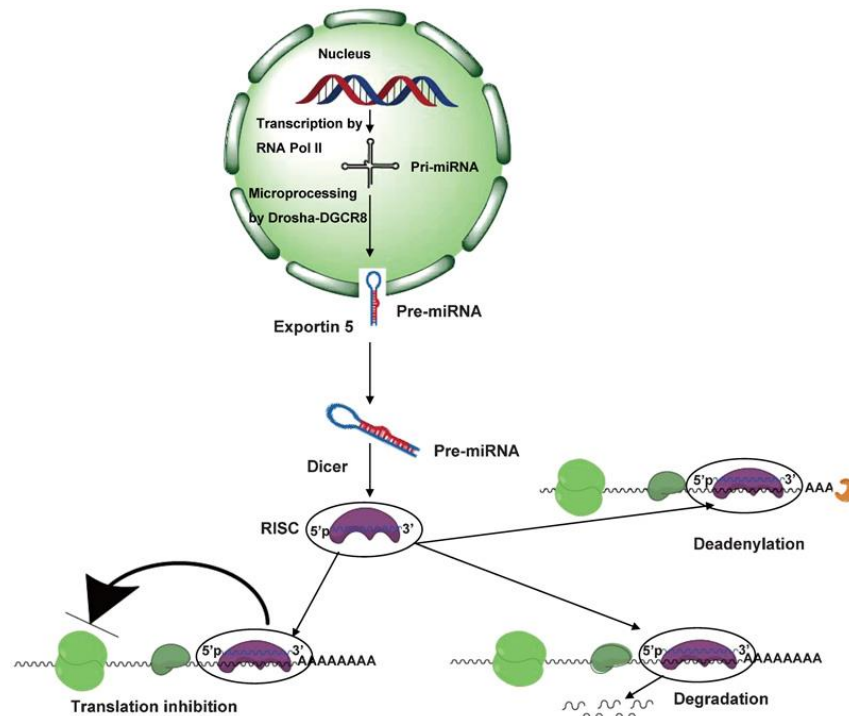


Figura 16. Regulação pós transcricional por microRNAs (91). O microRNA se combina com enzimas formando um complexo denominado RISC. Por complementaridade, o microRNA pode se ligar ao mRNA, principalmente na região 3'NT, ocasionando uma inibição da tradução desse mRNA por degradação, inibição ou desadenilação.

Outra característica de *HLA-E* é a presença de elementos *Alu* associados ao longo do segmento estendido deste gene, inclusive dentro da região de 3'NT (29,45) (Figura 17). Como mencionado anteriormente, desde quando o gene *HLA-E* foi descrito, foi observada a presença desses elementos ao longo do gene, um dos fatores que caracteriza *HLA-E*. *Alu* é um tipo de elemento repetitivo móvel (retrotransposon) da família SINE (*short interspersed elements*) específico de primatas e que apresenta grande número de cópias no genoma humano, somando por volta de 10% do genoma (42,43).

Os elementos *Alu* podem influenciar as taxas de recombinação e sua inserção em um gene (ou perto de um) tem o potencial de influenciar a expressão gênica através de vários mecanismos (42), por exemplo, afetando a maquinaria de regulação por microRNAs quando a inserção acontece dentro da região 3'NT de um gene (ocorrência em 16% dos genes humanos) (92), como no caso do *HLA-E*.

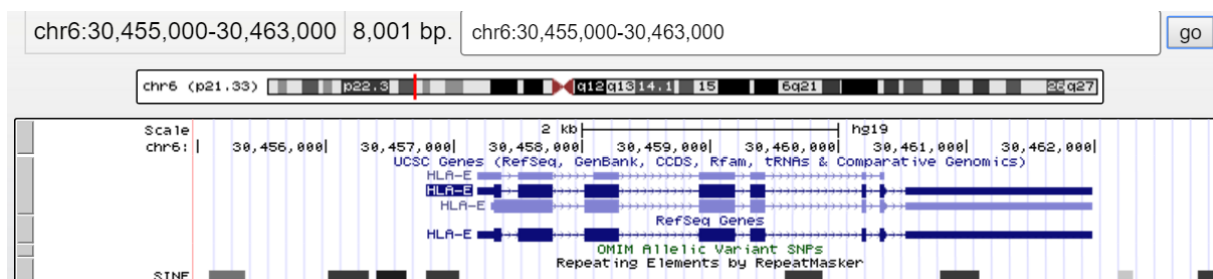


Figura 17. Elementos *Alu* presentes na região do gene *HLA-E*. Imagem obtida de UCSC Genome Browser compreendendo segmento estendido do gene *HLA-E* com os elementos *Alu* representados como família de elementos SINE (caixas pretas e cinzas). Há três elementos *Alu* na região distal 5', um no intron 5 e um dentro da região 3'NT (93).

Semelhante a região 3'NT, a região promotora de um gene participa da regulação da expressão gênica no nível transcricional. A princípio, quanto mais um gene é transcrito e traduzido em proteína, mais moléculas estarão presentes no organismo. A alteração da quantidade de moléculas *HLA-E* pode alterar sua função considerando sua interação com receptores ativadores e inibitórios.

Os genes do MHC de classe I são altamente regulados transcricionalmente e os padrões de expressão estão diretamente relacionados a sequência de DNA de elementos que regulam a transcrição presentes na região promotora (38,39) de cada um dos genes. Mas, outra diferença encontrada no *HLA-E* em relação aos genes *HLA* de classe I clássicos, foi a não conservação de alguns elementos ativadores (*enhancers*) típicos de genes *HLA* (29). Uma comparação feita entre as regiões promotoras dos genes *HLA* (com base em elementos conhecidos e bem caracterizados) mostrou que os principais elementos que regulam a expressão dos genes *HLA* clássicos estão na região proximal do promotor (distantes até 200 bases do início da transcrição), mas a associação com regiões distais também é importante para determinação do padrão de expressão. Em geral, a região proximal do promotor de genes *HLA* está associada com resposta a hormônios e citocinas (39).

As variações entre os elementos regulatórios da região promotora dos diferentes *loci HLA* estão relacionadas com o padrão de expressão de cada gene, ou seja, as diferenças na região proximal da promotor resultam em diferentes afinidades por fatores de transcrição, que então regulam o nível de transcrição de forma específica para cada gene (38,39,94). Cada gene teria um único padrão desses elementos na promotora, o que determinaria sua expressão em níveis e locais apropriados, relacionados a função da molécula (39).

Como discutido anteriormente, os principais elementos conservados em genes de classe I já estudados são: *enhancer A*, ISRE (do inglês, *interferon-stimulated response element*) e bloco SXY (Figura 18). A sequência dos elementos regulatórios da região

promotora do *HLA-E* e *HLA-G* são as mais divergentes em relação aos demais (38,94). Os elementos denominados *enhancer A* e ISRE estão relacionados à expressão induzidas por citocinas e interação com fatores de transcrição constitutivos, e SXY, com transcrição constitutiva e mediada por CIITA (*class II transactivator*). Polimorfismos nesses elementos interferem na interação com fatores de transcrição (respectivamente, NF- κ B – *nuclear factor κ B* –, IRF-1 – *interferon regulatory factor* – e CIITA) (25,38).

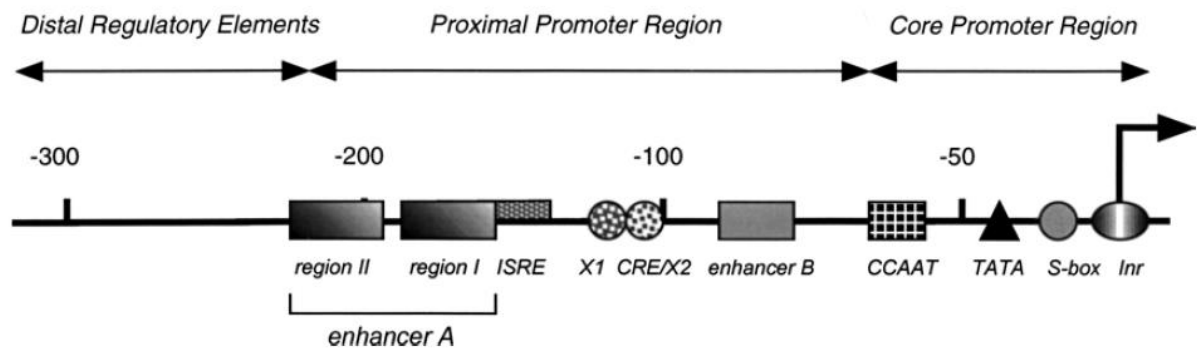


Figura 18. Esquema da região promotora típica de um gene *HLA* de classe I com as posições aproximadas dos elementos regulatórios conhecidos em relação ao início da tradução. Elementos localizados à montante do *enhancer A* são considerados como elementos da região promotora distal. A região promotora cerne está entre o CCAAT box e o início da tradução. Intermediária, até 200 nucleotídeos antes do início da tradução, está a região promotora proximal. Adaptado de Howcroft and Singer, 2003 (39).

No caso do *HLA-E*, o elemento *enhancer A* não é funcional e, portanto não responde ao NF- κ B (sem atividade de ligação) (39). O elemento ISRE não liga IRF-1 (principal fator de transcrição induzido por IFN- γ), mas a indução de IFN- γ ocorre por meio de fatores de transcrição que se ligam em regiões anteriores ao ISRE correspondentes a dois sítios de ligação a STAT1, que se sobrepõe em parte com ISRE do *HLA-E*; é conhecido como elemento GAS (38,39,94). O bloco SXY tem a maior parte conservada e é induzido por CIITA (38,39) (Figura 19).

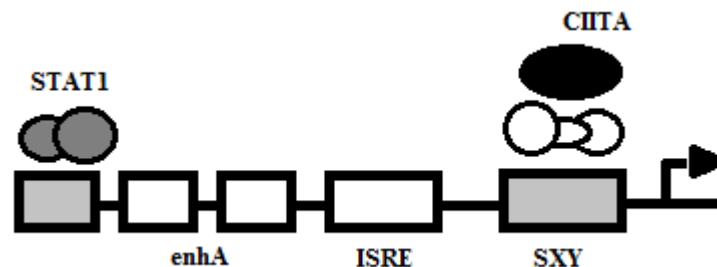


Figura 19. Regulação transcricional de *HLA-E*. Esquema dos elementos regulatórios ativos e ligação de proteínas (fatores de transcrição). As caixas em cinza claro representam os principais elementos regulatórios de *HLA-E*. As caixas em branco representam elementos conservados nos genes *HLA* de classe I, mas que não são funcionais para *HLA-E*. enhA: *enhancer A*; ISRE: *interferon-stimulated response element*; SXY: bloco SXY; CIITA: *class II transactivator*; STAT1: *Signal transducer and activator of transcription 1*. Adaptado de Gobin and van den Elsen, 2000 (38).

Além desses elementos, na região mais próxima do início da transcrição (promotora cerne) estão presentes os elementos *TATA box* do tipo TATAAA, diferente dos outros genes de classe I (TCTAAA) (Figura 18). Essa sequência é um sítio de ligação para TBP (*TATA binding protein*), componente da TFIID (fator de transcrição geral), que recruta RNA polimerase II para o início da transcrição (39).

Outra diferença em relação aos demais genes *HLA* de classe I encontrada foi em relação ao padrão de metilação das ilhas CpG na região promotora, principalmente em células do trofoblasto (46). Foi observado que *HLA-E* é o único gene de classe I livre de metilação no promotor, o que permite ser este o único gene *HLA* transcrito em linhagem celular mimetizando o trofoblasto. Mais que isso, a região de *HLA-E* no MHC e não apenas sua região promotora, mas regiões internas ao gene e 3'NT também tem padrão diferente de metilação. Isso suporta a expressão ampla de *HLA-E* já observada, mesmo na gestação, além de sugerir esse ser um gene *HLA* ancestral que nunca foi metilado, fortalecendo sua grande importância biológica. Além disso, como mencionada anteriormente, há mais elementos *Alu* presentes na região de *HLA-E*: três na região promotora e outro no íntron 5 (Figura 17). Esses elementos podem caracterizar *hotspots* de recombinação, o que pode explicar também a provável formação de *loops* na cromatina e esse escape de metilação (46).

Conhecer a estrutura da região promotora do *HLA-E* auxiliaria no entendimento sobre a expressão desse gene e como polimorfismos nesta região regulatória poderiam influenciar a interação com fatores de transcrição e o consequente padrão de expressão gênica. Assim, variações na codificadora ou promotora são potenciais em alterar a função devido a proteínas novas ou quantidade da molécula expressa, respectivamente (79,83,84). Portanto, polimorfismos poderiam afetar a imunomodulação dependente de *HLA-E*, além dos resultados de infecções virais, implantação e manutenção do feto, transplantes, progressão de tumores e doenças autoimunes.

Embora muitos estudos tenham sido conduzidos para avaliar a variabilidade da região codificadora e definição dos alelos de *HLA-E*, não há avaliação do gene estendido quanto a variabilidade e estrutura de seus haplótipos, principalmente considerando a região promotora. Assim, o conhecimento sobre a variabilidade dessas regiões é importante para o planejamento e tratamento clínico.

1.3.4 Importância clínica de HLA-E

É sabido que pelo menos dois principais alelos de *HLA-E* estão presentes nas diversas populações mundiais (E*01:01 e E*01:03) em alta frequência e que possuem certa diferença funcional considerando a estabilidade do complexo formado com o peptídeo e o nível de expressão na superfície celular, que é maior para os alelos do grupo E*01:03. Além disso, como já foi discutido, a molécula HLA-E expressa na superfície tem um papel duplo na imunidade.

Pensando na condição fisiológica da gestação, a molécula HLA-E, juntamente com *HLA-G* e *HLA-C*, são as únicas expressas pelo trofoblasto na interface materno-fetal (86) e possuem então função importante durante a gestação, quando o sistema imune materno deve permitir o desenvolvimento do feto (32,86). No ambiente materno fetal, HLA-E é expressa continuamente e liga peptídeos derivados de HLA-G (71,73,95). Na decídua materna, principalmente no sítio de implantação, estão presentes grande quantidade de células NK diferenciadas (CD56^{high}CD16^{low}), que expressam grande quantidade de receptores inibitórios CD94/NKG2A, mas também ativadores CD94/NKG2C, além de possuírem menor atividade citolítica e alta produção de citocinas (73). A principal função da HLA-E nessa situação é modular a atividade dessas células, inibindo sua citotoxicidade (por meio da interação com receptor inibitório CD94/NKG2A) e ativar a secreção de citocinas (pelo receptor CD94/NKG2C) (como TGF- β e IL-10) por essas e pelos linfócitos T $\gamma\delta$ (que também tem maior expressão de CD94), que por sua vez auxiliarão na vascularização e formação adequada da placenta (73,86,96). A quantidade de moléculas HLA-E expressa e sua afinidade pelos receptores é importante para que a função moduladora ocorra na medida certa, pois, caso contrário, há predisposição para ocorrência de complicações, tais como aborto espontâneo recorrente e/ou mal formação da placenta. Por exemplo, em indianos, a homozigose para E*01:01, alelo relacionado com uma menor expressão na superfície celular, está associado com aborto espontâneo devido talvez a menor inibição das células NK (32,73).

A expressão de HLA-E em tumores pode estar envolvida em um mecanismo de escape do sistema imunológico (73,84,97,98), assim como ocorre em algumas infecções virais. Quando a expressão de não clássicos está aumentada em tumores, geralmente há inibição da atividade citotóxica das células NK, fazendo com que o tumor não seja percebido pelo sistema imune (87). Assim, essa atividade tolerogênica contribuiria para um mau prognóstico ao paciente com câncer. Por outro lado, alterações na expressão de HLA-E

indicariam um problema na maquinaria de apresentação antigênica, possivelmente acarretando na ativação das funções citotóxicas.

No caso de neoplasias, já foi observado aumento da expressão de HLA-E em linfomas frescos, carcinoma de ovário, gliomas, câncer de colon, melanoma, câncer de mama, câncer de nasofaringe, glioblastoma e câncer invasivo cervical associado ao HPV (73,84,95,99). Também já foi observado infiltrado tumoral de células T com aumento na expressão de CD94/NKG2A, que podem contribuir para esse efeito inibitório e escape das células tumorais (17). Embora em alguns casos a alta expressão tenha sido associada ao alelo E*01:03 e por consequência um pior prognóstico, em outros como glioblastoma, adenocarcinoma cervical e carcinoma colorretal, a alta expressão de HLA-E teve associação com melhor sobrevida do paciente (73,100,101). Nesses casos, é importante considerar a interação de HLA-E com TCR em linfócitos T CD8+. Por exemplo, há formação de complexos HLA-E com peptídeos derivados de isoformas de Prdx5, este não é reconhecido pelo CD94/NKG2A e por isso não protege a célula contra atividade citotóxica das células NK (79). Em melanoma e doenças inflamatórias, níveis altos de Prdx5 são produzidos, o que pode auxiliar no reconhecimento e combate de células tumorais (84).

No caso de infecções, o quadro é semelhante: quando a expressão de HLA de classe I clássicos é inibida, a apresentação de antígenos dos patógenos fica comprometida e a expressão de não clássicos pode auxiliar na tolerância de células infectadas. No entanto, HLA-E nesse cenário pode atuar na ativação da imunidade adaptativa e apresentar antígenos a células T, como no caso de HCMV, HIV, HCV, *Salmonella* e *Mycobacterium*. O balanço entre alta ou baixa expressão considerando os alelos E*01:01 e E*01:03 vai então depender do peptídeo ligado e a maior afinidade pelo tipo de receptor (ativador ou inibitório). Por exemplo, o alelo E*01:03 já foi associado a risco reduzido de infecção por HIV, enquanto para HCV, o genótipo homozigoto para o alelo E*01:01 é que foi associado com proteção (73,84).

No caso de infecções por HCMV, a proteína gpUL40 gera peptídeos que podem se ligar independentemente de TAP à molécula de HLA-E. Esses peptídeos são idênticos a peptídeos líder derivados de alguns alelos *HLA-A* e *HLA-C* (72,73,79,84,85,102). Foi nesse cenário que observou-se a existência das células T restritas ao HLA-E, que tinham um fenótipo de memória e intermediário entre os linfócitos T e células NK, pois além do TCR eram expressos os receptores CD94/NKG2 (79,84). Porém, observou-se também que a resposta contra a infecção e atividade dessas células dependia do genótipo dos alelos de *HLA-A* e/ou *HLA-C*, já que os peptídeos virais eram semelhantes a alguns derivados destes alelos.

Dessa forma, indivíduos com alelos do tipos HLA-C*02 e HLA-C*07 e HLA-A*02 não geravam resposta contra o patógeno, pois HLA-E/peptídeo líder era conhecido como próprio e geravam tolerância (79,84,98,103). Extrapolando, essa situação poderia se reverter no caso de transplantes.

O sucesso de transplantes também está associado aos alelos (polimorfismos), maior ou menor expressão de HLA-E e sua interação com peptídeos e receptores. Para que o enxerto não seja rejeitado pelo hospedeiro, a função tolerogênica do HLA-E é a que deve prevalecer, pois, se peptídeos do doador forem reconhecidos como alogênicos podem surgir complicações (72,73,79,84,98). Conhecendo a função de HLA-E na apresentação de peptídeos líder de outras moléculas HLA de classe I, qualquer incompatibilidade de alelos poderia gerar rejeição mediada por HLA-E. Também, considerando a questão de infecção por HCMV, pacientes com o vírus estão mais susceptíveis a resposta contra enxerto dependendo dos alelos HLA entre receptor e doador (79,84), já que o vírus gera peptídeos idênticos aos peptídeos líder de HLA-A e HLA-C. Em outros casos, em transplante de células tronco hematopoéticas, o alelo E*01:03 parece estar relacionado a melhor sobrevida e menor taxa de doença de enxerto-versus-hospedeiro, assim como o genótipo homozigoto para o alelo E*01:01 está associado a maior risco de infecções. Porém em relação ao transplante de órgãos sólidos, há controvérsias do que seja melhor por causa do papel duplo da molécula HLA-E (72,73,98,104).

O HLA-E também pode influenciar a patogênese de doenças autoimunes como diabetes mellitus, espondilite anquilosante, doença de Behçet, doença de Kawasaki em chineses (vasculite sistêmica pediátrica), esclerose múltipla e psoríase (73,95). Por exemplo, na psoríase é importante a ativação de células NK contra células T autorreativas através do receptor CD94/NKG2C, o que auxilia na prevenção e/ou melhor prognóstico (105). Em outros casos, a baixa expressão de CD94/NKG2A é que foi associada à severidade da doença (17). Na artrite reumatoide, o genótipo E*01:01 foi associado à melhor resposta ao tratamento, além de estar significativamente pouco presente nos pacientes (106). De maneira geral, todos esses aspectos fisiológicos e patológicos dependem do tipo e da quantidade de molécula HLA-E que o indivíduo expressa em cada condição.

1.4 Referências Bibliográficas

1. Klein J, Sato A. The HLA system. First of two parts. N Engl J Med [Internet]. 2000 Sep 7 [cited 2016 Oct 11];343(10):702–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10974135>
2. Klein J, Sato A. The HLA system. Second of two parts. N Engl J Med [Internet]. 2000 Sep 14 [cited

- 2016 Oct 11];343(11):782–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10984567>
3. Gorer PA. The genetic and antigenic basis of tumour transplantation. *J Pathol Bacteriol* [Internet]. John Wiley & Sons, Ltd.; 1937 May [cited 2016 Oct 11];44(3):691–7. Available from: <http://doi.wiley.com/10.1002/path.1700440313>
4. Meyer D, Thomson G. How selection shapes variation of the human major histocompatibility complex: a review. *Ann Hum Genet* [Internet]. 2001;65(1):1–26. Available from: <http://doi.wiley.com/10.1046/j.1469-1809.2001.6510001.x>
5. Shankarkumar U. The Human Leukocyte Antigen (HLA) System. *Int J Hum Genet*. 2004;4(2):91–103.
6. Li XC, Raghavan M. Structure and function of major histocompatibility complex (MHC) class I antigens. *Curr Opin Organ Transpl*. 2010;15(4):499–504.
7. Agrawal S, Sharma RK. The past, present, and future of human leukocyte antigen techniques. *Indian J Transplant* [Internet]. Indian Society of Organ Transplantation; 2012;6(1):2–19. Available from: [http://dx.doi.org/10.1016/S2212-0017\(12\)60074-4](http://dx.doi.org/10.1016/S2212-0017(12)60074-4)
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. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Genet*. 2009;54(1):15–39.
10. Kindt TJ, Goldsby RA, Osborne BA, Kuby J, Kuby J. *Kuby immunology*. W.H. Freeman; 2007. 27 p.
11. Abbas AK, Lichtman AHH, Pillai S, Baker DL (Medical illustrator), Baker A. *Imunologia Celular e Molecular*. Elsevier; 2015. 549 p.
12. Dausset J. Iso-Leuko-Antibodies. *Vox Sang* [Internet]. Blackwell Publishing Ltd; 1958 Jan [cited 2016 Oct 11];3(1):40–1. Available from: <http://doi.wiley.com/10.1111/j.1423-0410.1958.tb03559.x>
13. Goldberg AC, Rizzo LV. MHC structure and function – antigen presentation. Part 1. Einstein (São Paulo, Brazil) [Internet]. [cited 2016 Oct 11];13(1):153–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25807245>
14. Goldberg AC, Rizzo LV. MHC structure and function - antigen presentation. Part 2. Einstein (São Paulo, Brazil) [Internet]. [cited 2016 Oct 11];13(1):157–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25807243>
15. Donadi EA, Castelli EC, Arnaiz-Villena A, Roger M, Rey D, Moreau P. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. *Cell Mol Life Sci* [Internet]. 2011 Feb [cited 2016 Oct 11];68(3):369–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21107637>
16. Castelli EC, Veiga-Castelli LC, Yaghi L, Moreau P, Donadi EA. Transcriptional and posttranscriptional regulations of the HLA-G gene. Vol. 2014, *Journal of Immunology Research*. Hindawi Publishing Corporation; 2014.
17. 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*. 2014;20(2):129–41.
18. Braud VM, Allan DS, McMichael AJ. Functions of nonclassical MHC and non-MHC-encoded class I molecules [Internet]. Vol. 11, *Current Opinion in Immunology*. 1999 [cited 2016 Mar 15]. p. 100–8.

- Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10047540>
19. O'Callaghan CA, Bell JI. Structure and function of the human MHC class Ib molecules HLA-E, HLA-F and HLA-G. Vol. 163, Immunological Reviews. 1998. p. 129–38.
 20. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res [Internet]. 2015 Jan [cited 2016 May 17];43(Database issue):D423–31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25414341>
 21. Hedrick PW. Balancing selection and MHC. Genetica [Internet]. [cited 2016 Oct 11];104(3):207–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10386384>
 22. Spencer HG, Marks RW. The maintenance of single-locus polymorphism. I. Numerical studies of a viability selection model. Genetics [Internet]. 1988 Oct [cited 2016 Oct 11];120(2):605–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3197961>
 23. Bodmer WF. Evolutionary significance of the HL-A system. Nature [Internet]. 1972 May 19 [cited 2016 Oct 11];237(5351):139–45 passim. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/4113158>
 24. Redman BG, Chang AE. MedicinaNET - Tumor Immunology [Internet]. ACP Medicine. 2009. p. 1–10. Available from: http://www.medicinanet.com.br/m/conteudos/acp-medicine/4584/imunologia_de_tumores_%25E2%2580%2593_bruce_g_redman_do_alfred_e_chang_md_fac.htm
 25. van den Elsen PJ, Peijnenburg A, van Eggermond MC, Gobin SJ. Shared regulatory elements in the promoters of MHC class I and class II genes. Immunol Today [Internet]. 1998;19(7):308–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9666603>
http://ac.els-cdn.com/S0167569998012870/1-s2.0-S0167569998012870-main.pdf?_tid=87cc50b2-4d49-11e3-9a4a-00000aab0f6b&acdnat=1384446575_10ec874d65fac4f68d2cfd8f5a39237a
 26. Braud VM, Allan DS, McMichael AJ. Functions of nonclassical MHC and non-MHC-encoded class I molecules. Curr Opin Immunol [Internet]. 1999 Feb [cited 2016 Oct 14];11(1):100–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10047540>
 27. Allan DSJ, Lepin EJM, Braud VM, O'Callaghan CA, McMichael AJ. Tetrameric complexes of HLA-E, HLA-F, and HLA-G. Vol. 268, Journal of Immunological Methods. 2002. p. 43–50.
 28. 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]. 1990;171(1):1–18. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1688605>
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2187653/pdf/je171111.pdf>
 29. Koller BH, Geraghty DE, Shimizu Y, DeMars R, Orr HT. HLA-E. A novel HLA class I gene expressed in resting T lymphocytes. J Immunol [Internet]. 1988 Aug 1 [cited 2016 Oct 11];141(3):897–904. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3260916>
 30. Geraghty DE, Koller BH, Orr HT. A human major histocompatibility complex class I gene that encodes a protein with a shortened cytoplasmic segment. Proc Natl Acad Sci U S A [Internet]. 1987 Dec [cited 2016 Oct 11];84(24):9145–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3480534>
 31. O'Callaghan CA, Tormo J, Willcox BE, Braud VM, Jakobsen BK, Stuart DI, et al. Structural features impose tight peptide binding specificity in the nonclassical MHC molecule HLA-E. Mol Cell [Internet].

- 1998 Mar [cited 2016 Mar 16];1(4):531–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9660937>
32. Djuricic S, Hviid TVF. HLA class Ib molecules and immune cells in pregnancy and preeclampsia. *Front Immunol*. 2014;5(DEC):1–17.
 33. Hviid TVF, Møller C, Sørensen S, Morling N. Co-dominant expression of the HLA-G gene and various forms of alternatively spliced HLA-G mRNA in human first trimester trophoblast. *Hum Immunol*. 1998;59(2):87–98.
 34. Carosella ED. The tolerogenic molecule HLA-G [Internet]. Vol. 138, *Immunology Letters*. Elsevier B.V.; 2011. p. 22–4. Available from: <http://dx.doi.org/10.1016/j.imlet.2011.02.011>
 35. 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]. 2013;191(7):3553–62. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3780715&tool=pmcentrez&rendertype=abstract>
 36. Goodridge JP, Lee N, Burian A, Pyo C, Tykodi SS, Warren EH, et al. HLA-F and MHC-I open conformers cooperate in a MHC-I antigen cross-presentation pathway. *J Immunol* [Internet]. 2013;191(4):1567–77. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3732835&tool=pmcentrez&rendertype=abstract>
 37. Boyle LH, Gillingham AK, Munro S, Trowsdale J. Selective export of HLA-F by its cytoplasmic tail. *J Immunol*. 2006;176(Table I):6464–72.
 38. Gobin SJP, Van Den Elsen PJ. Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F and HLA-G. In: *Human Immunology*. 2000. p. 1102–7.
 39. Howcroft TK, Singer DS. Expression of nonclassical MHC class Ib genes: comparison of regulatory elements. *Immunol Res*. 2003;27(1):1–30.
 40. Moscoso J, Serrano-Vela JI, Pacheco R, Arnaiz-Villena A. HLA-G, -E and -F: Allelism, function and evolution. *Transpl Immunol*. 2006;17(1):61–4.
 41. Wei X, Orr HT. Differential expression of HLA-E, HLA-F, and HLA-G transcripts in human tissue. *Hum Immunol*. 1990;29(2):131–42.
 42. Deininger P. Alu elements: know the SINEs. *Genome Biol* [Internet]. 2011 [cited 2016 Oct 13];12(12):236. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22204421>
 43. Dridi S. Alu mobile elements: from junk DNA to genomic gems. *Scientifica (Cairo)* [Internet]. 2012;2012:545328. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3820591&tool=pmcentrez&rendertype=abstract>
 44. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. *Immunobiology*. Garland Science; 2001;
 45. Le Bouteiller P, Lenfant F. Antigen-presenting function(s) of the non-classical HLA-E, -F and -G class I molecules: the beginning of a story. *Res Immunol* [Internet]. 1996 Jun [cited 2016 Oct 11];147(5):301–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8876058>
 46. Boucraut J, Guillaudeux T, Alizadeh M, Boretto J, Chimini G, Malecaze F, et al. HLA-E is the only class I gene that escapes CpG methylation and is transcriptionally active in the trophoblast-derived

- human cell line JAR. *Immunogenetics* [Internet]. 1993 [cited 2016 Oct 11];38(2):117–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7683306>
47. Lima THA, Buttura RV, Donadi EA, Veiga-Castelli LC, Mendes-Junior CT, Castelli EC. HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. *Hum Immunol* [Internet]. 2016 Oct [cited 2016 Sep 22];77(10):841–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27448841>
 48. Castelli EC, Mendes CT, Donadi EA. HLA-G alleles and HLA-G 14 bp polymorphisms in a Brazilian population. *Tissue Antigens*. 2007;70(1):62–8.
 49. Veiga-Castelli LC, Felicio LP, Georg RC, Castelli EC, Donadi EA. A nonsynonymous mutation at HLA-E defines the new E*01:06 allele in Brazilian individuals. *Tissue Antigens* [Internet]. 2013;82(3):216–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24032735>
<http://onlinelibrary.wiley.com/doi/10.1111/tan.12177/abstract>
<http://onlinelibrary.wiley.com/store/10.1111/tan.12177/asset/tan12177.pdf?v=1&t=ho>
[snk561&s=1d10a668d75751871e62ef454df31f9f91a45b56](http://onlinelibrary.wiley.com/doi/10.1111/tan.12177/asset/tan12177.pdf?v=1&t=ho)
 50. Veiga-Castelli LC, Castelli EC, Beauchemin K, Donadi EA, Roger M. A novel HLA-E allele, E*01:03:05, identified in two Brazilian individuals. *Tissue Antigens* [Internet]. 2012 Aug [cited 2016 Oct 12];80(2):200–1. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22594712>
 51. Castelli EC, Mendes-Junior CT, Wiezel CE V, Peres NTA, Simões AL, Rossi NMM, et al. A novel HLA-G allele, HLA-G*010111, in the Brazilian population. *Tissue Antigens* [Internet]. 2007 Oct [cited 2016 Oct 12];70(4):349–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17767566>
 52. Castelli EC, de Albuquerque RS, Veiga-Castelli LC, Felício LP, Beauchemin K, Faucher MC, et al. Identification of two new HLA-G alleles, G*01:01:03:03 and G*01:01:21, in Brazilian individuals. *Tissue Antigens* [Internet]. 2012 Jul [cited 2016 Oct 12];80(1):70–1. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22449125>
 53. Geraghty DE, Stockscheider M, Ishitani A, Hansen JA. Polymorphism at the HLA-E locus predates most HLA-A and -B polymorphism. *Hum Immunol* [Internet]. 1992 Mar [cited 2016 May 19];33(3):174–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1618657>
 54. Matte C, Lacaille J, Zijenah L, Ward B, Roger M, Chidawanyika H, et al. HLA-G and HLA-E polymorphisms in an indigenous african population. In: *Human Immunology*. 2000. p. 1150–6.
 55. Grimsley C, Kawasaki A, Gassner C, Sageshima N, Nose Y, Hatake K, et al. Definitive high resolution typing of HLA-E allelic polymorphisms: Identifying potential errors in existing allele data. *Tissue Antigens* [Internet]. 2002 Sep [cited 2016 May 19];60(3):206–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12445303>
 56. Grimsley C, Ober C. Population genetic studies of HLA-E: evidence for selection. *Hum Immunol* [Internet]. 1997;52(1):33–40. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9021407
<http://www.ncbi.nlm.nih.gov/pubmed/9021407>
 57. Arnaiz-Villena A, Vargas-Alarcon G, Serrano-Vela JJ, Reguera R, Martinez-Laso J, Silvera-Redondo C, et al. HLA-E polymorphism in Amerindians from Mexico (Mazatecs), Colombia (Wayu) and Chile (Mapuches): evolution of MHC-E gene. *Tissue Antigens* [Internet]. Blackwell Publishing Ltd; 2007 Apr

- [cited 2016 May 19];69(s1):132–5. Available from: http://doi.wiley.com/10.1111/j.1399-0039.2006.763_2.x
58. Pyo CW, 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*. 2006;58(4):241–51.
 59. Felício LP, Porto IOP, Mendes-Junior CT, Veiga-Castelli LC, Santos KE, Vianello-Brondani RP, et al. Worldwide HLA-E nucleotide and haplotype variability reveals a conserved gene for coding and 3' untranslated regions. *Tissue Antigens*. 2014;83(2):82–93.
 60. 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]. 2015 Dec [cited 2016 Mar 16];76(12):945–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26187162>
 61. Antoun A, Jobson S, Cook M, Moss P, Briggs D. Ethnic variability in human leukocyte antigen-E haplotypes. *Tissue Antigens* [Internet]. 2009 Jan [cited 2016 May 19];73(1):39–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19140831>
 62. Veiga-Castelli LC, Castelli EC, Mendes CT, da Silva WA, Faucher M-C, Beauchemin K, et al. Non-classical HLA-E gene variability in Brazilians: a nearly invariable locus surrounded by the most variable genes in the human genome. *Tissue Antigens* [Internet]. 2012 Jan [cited 2016 Oct 12];79(1):15–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22050290>
 63. Liu XX, Pan FH, Tian W. Characterization of HLA-E polymorphism in four distinct populations in Mainland China. *Tissue Antigens* [Internet]. 2012 Jul [cited 2016 May 19];80(1):26–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22486789>
 64. Strong RK, Holmes MA, Li P, Braun L, Lee N, Geraghty DE. HLA-E allelic variants: Correlating differential expression, peptide affinities, crystal structures, and thermal stabilities. *J Biol Chem*. 2003;278(7):5082–90.
 65. Carvalho dos Santos L, Tureck LV, Wowk PF, Mattar SB, Gelmini GF, Magalhães JCM, et al. HLA-E polymorphisms in an Afro-descendant Southern Brazilian population. *Hum Immunol* [Internet]. American Society for Histocompatibility and Immunogenetics; 2013;74(2):199–202. Available from: <http://dx.doi.org/10.1016/j.humimm.2012.10.020>
 66. Veiga-Castelli LC, Castelli EC, Mendes CT, Da Silva WA, Faucher MC, Beauchemin K, et al. Non-classical HLA-E gene variability in Brazilians: A nearly invariable locus surrounded by the most variable genes in the human genome. *Tissue Antigens*. 2012;79(1):15–24.
 67. Felício LP. Análise da variabilidade e perfil de seleção atuando sobre o gene HLA-E em populações mundiais. Universidade Federal de Goiás; 2016.
 68. McVean GA, Altshuler (Co-Chair) DM, Durbin (Co-Chair) RM, Abecasis GR, Bentley DR, Chakravarti A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* [Internet]. 2012;491(7422):56–65. Available from: <http://www.nature.com/doifinder/10.1038/nature11632>
 69. Veiga-Castelli LC, da Silveira Bulcão Bertuol JM, Castelli EC, Donadi EA. Low variability at the HLA-E promoter region in the Brazilian population. *Hum Immunol* [Internet]. American Society for Histocompatibility and Immunogenetics; 2015;77(2):172–5. Available from:

- <http://dx.doi.org/10.1016/j.humimm.2015.11.014>
70. Braud V, Jones EY, McMichael A. The human major histocompatibility complex class Ib molecule HLA-E binds signal sequence-derived peptides with primary anchor residues at positions 2 and 9. *Eur J Immunol* [Internet]. 1997 May [cited 2016 Oct 13];27(5):1164–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9174606>
 71. King A, Allan DS, Bowen M, Powis SJ, Joseph S, Verma S, et al. HLA-E is expressed on trophoblast and interacts with CD94/NKG2 receptors on decidual NK cells. *Eur J Immunol* [Internet]. 2000 Jun [cited 2016 Oct 13];30(6):1623–31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10898498>
 72. Sullivan LC, Clements CS, Rossjohn J, Brooks AG. The major histocompatibility complex class Ib molecule HLA-E at the interface between innate and adaptive immunity. Vol. 72, *Tissue Antigens*. 2008. p. 415–24.
 73. Iwaszko M, Bogunia-Kubik K. Clinical significance of the HLA-E and CD94/NKG2 interaction. *Arch Immunol Ther Exp (Warsz)*. 2011;59(5):353–67.
 74. Lee N, Goodlett DR, Ishitani A, Marquardt H, Geraghty DE. HLA-E surface expression depends on binding of TAP-dependent peptides derived from certain HLA class I signal sequences. *J Immunol* [Internet]. 1998 May 15 [cited 2016 Oct 13];160(10):4951–60. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9590243>
 75. Zeng L, Sullivan LC, Vivian JP, Walpole NG, Harpur CM, Rossjohn J, et al. A Structural Basis for Antigen Presentation by the MHC Class Ib Molecule, Qa-1. *J Immunol* [Internet]. 2012 Jan 1 [cited 2016 Dec 28];188(1):302–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22131332>
 76. Borrego F, Ulbrecht M, Weiss EH, Coligan JE, Brooks AG. Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J Exp Med* [Internet]. 1998 Mar 2 [cited 2016 Mar 15];187(5):813–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9480992>
 77. Lee N, Llano M, Carretero M, Ishitani A, Navarro F, López-Botet M, et al. HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A. *Proc Natl Acad Sci U S A* [Internet]. 1998 Apr 28 [cited 2016 Mar 16];95(9):5199–204. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9560253>
 78. Brooks AG, Borrego F, Posch PE, Patamawenu A, Scorzelli CJ, Ulbrecht M, et al. Specific recognition of HLA-E, but not classical, HLA class I molecules by soluble CD94/NKG2A and NK cells. *J Immunol* [Internet]. 1999 Jan 1 [cited 2016 Oct 13];162(1):305–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9886400>
 79. Pietra G, Romagnani C, Moretta L, Mingari MC. HLA-E and HLA-E-bound peptides: recognition by subsets of NK and T cells. *Curr Pharm Des* [Internet]. 2009;15(28):3336–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19860683>
 80. Braud VM, Allan DS, O’Callaghan CA, Söderström K, D’Andrea A, Ogg GS, et al. HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* [Internet]. 1998 Feb 19 [cited 2016 Oct 14];391(6669):795–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9486650>
 81. Milena I, Katarzyna G, Katarzyna B-K. Non-Classical MHC Class Ib Molecules and their Receptors—Role in Allogeneic Transplantation of Hematopoietic Stem Cells. *J Transplant Technol Res* [Internet]. OMICS International; 2012 [cited 2016 Dec 30];2(3). Available from:

- <http://www.omicsonline.org/2161-0991/2161-0991-2-117.digital/2161-0991-2-117.html>
82. Valés-Gómez M, Reyburn HT, Erskine RA, López-Botet M, Strominger JL. Kinetics and peptide dependency of the binding of the inhibitory NK receptor CD94/NKG2-A and the activating receptor CD94/NKG2-C to HLA-E. *EMBO J* [Internet]. 1999 Aug 2 [cited 2016 Oct 14];18(15):4250–60. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10428963>
 83. Llano M, Lee N, Navarro F, García P, Albar JP, Geraghty DE, et al. HLA-E-bound peptides influence recognition by inhibitory and triggering CD94/NKG2 receptors: preferential response to an HLA-G-derived nonamer. *Eur J Immunol* [Internet]. 1998 Sep [cited 2016 Mar 16];28(9):2854–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9754572>
 84. Pietra G, Romagnani C, Manzini C, Moretta L, Mingari MC. The emerging role of HLA-E-restricted CD8⁺ T lymphocytes in the adaptive immune response to pathogens and tumors. *J Biomed Biotechnol* [Internet]. 2010 [cited 2016 Mar 16];2010:907092. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20634877>
 85. García P, Llano M, de Heredia AB, Willberg CB, Caparrós E, Aparicio P, et al. Human T cell receptor-mediated recognition of HLA-E. *Eur J Immunol* [Internet]. 2002 Apr [cited 2016 Mar 16];32(4):936–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11920559>
 86. Ishitani A, Sageshima N, Hatake K. The involvement of HLA-E and -F in pregnancy. Vol. 69, *Journal of Reproductive Immunology*. 2006. p. 101–13.
 87. Kochan G, Escors D, Breckpot K, Guerrero-Setas D. Role of non-classical MHC class I molecules in cancer immunosuppression. *Oncoimmunology* [Internet]. 2013;2(11):e26491-8. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3894240&tool=pmcentrez&rendertype=abstract>
 88. Bartel DP, Lee R, Feinbaum R. MicroRNAs : Genomics , Biogenesis , Mechanism , and Function. *Cell*. 2004;116:281–97.
 89. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. Vol. 136, *Cell*. 2009. p. 215–33.
 90. Castelli EC, Moreau P, Chiromatzo AO e, Mendes-Junior CT, Veiga-Castelli LC, Yaghi L, et al. In silico analysis of microRNAs targeting the HLA-G 3' untranslated region alleles and haplotypes. *Hum Immunol*. 2009;70(12):1020–5.
 91. Wei B, Pei G. microRNAs: critical regulators in Th17 cells and players in diseases. *Cell Mol Immunol* [Internet]. 2010 May [cited 2016 Oct 13];7(3):175–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20364159>
 92. Hoffman Y, Dahary D, Bublik DR, Oren M, Pilpel Y. The majority of endogenous microRNA targets within Alu elements avoid the microRNA machinery. *Bioinformatics*. 2013;29(7):894–902.
 93. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. *Genome Res* [Internet]. 2002 May 16 [cited 2016 Mar 17];12(6):996–1006. Available from: <http://www.genome.org/cgi/doi/10.1101/gr.229102>
 94. Gobin SJP, Zutphen M Van, Woltman AM, Elsen PJ Van Den. Transactivation of Classical and Nonclassical HLA Class I Genes Through the IFN-Stimulated Response Element. *J Immunol* [Internet]. 1999;163:1428–34. Available from: <http://www.jimmunol.org/content/163/3/1428>
 95. Morandi F, Pistoia V. Interactions between HLA-G and HLA-E in Physiological and Pathological

- Conditions. *Front Immunol* [Internet]. 2014;5(August):394. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4141331&tool=pmcentrez&rendertype=abstract>
96. Djuricic S, Sørensen AE, Hviid TVF. A fast and easy real-time PCR genotyping method for the HLA-G 14-bp insertion/deletion polymorphism in the 3' untranslated region. *Tissue Antigens*. 2012;79(3):186–9.
 97. Marchesi M, Andersson E, Villabona L, Seliger B, Lundqvist A, Kiessling R, et al. HLA-dependent tumour development: a role for tumour associate macrophages? *J Transl Med* [Internet]. *Journal of Translational Medicine*; 2013;11(1):247. Available from: *Journal of Translational Medicine*
 98. Kraemer T, Blasczyk R, Bade-Doeding C, Kraemer T, Blasczyk R, Bade-Doeding C. HLA-E: a novel player for histocompatibility. *J Immunol Res* [Internet]. Hindawi Publishing Corporation; 2014 [cited 2016 Oct 14];2014:352160. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25401109>
 99. Gonçalves MAG, Le Discorde M, Simões RT, Rabreau M, Soares EG, Donadi EA, et al. Classical and non-classical HLA molecules and p16(INK4a) expression in precursors lesions and invasive cervical cancer. *Eur J Obstet Gynecol Reprod Biol* [Internet]. 2008 Nov [cited 2016 Oct 26];141(1):70–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18692948>
 100. Spaans VM, Peters AA, Fleuren G, Jordanova ES. HLA-E expression in cervical adenocarcinomas: association with improved long-term survival. *J Transl Med* [Internet]. *Journal of Translational Medicine*; 2012;10(1):184. Available from: *Journal of Translational Medicine*
 101. Benevolo M, Mottolese M, Tremante E, Rollo F, Diodoro MG, Ercolani C, et al. High expression of HLA-E in colorectal carcinoma is associated with a favorable prognosis. *J Transl Med* [Internet]. BioMed Central Ltd; 2011;9(1):184. Available from: <http://www.translational-medicine.com/content/9/1/184>
 102. Heatley SL, Pietra G, Lin J, Widjaja JML, Harpur CM, Lester S, et al. Polymorphism in human cytomegalovirus UL40 impacts on recognition of human leukocyte antigen-E (HLA-E) by natural killer cells. *J Biol Chem*. 2013;288(12):8679–90.
 103. Halenius A, Gerke C, Hengel H. Classical and non-classical MHC I molecule manipulation by human cytomegalovirus: so many targets—but how many arrows in the quiver? *Cell Mol Immunol* [Internet]. 2015 Mar [cited 2016 Mar 17];12(2):139–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25418469>
 104. Pabón MA, Navarro CE, Osorio JC, Gómez N, Moreno JP, Donado AF, et al. Impact of Human Leukocyte Antigen Molecules E, F, and G on the Outcome of Transplantation. *Transplant Proc* [Internet]. Elsevier Inc.; 2014;46(9):2957–65. Available from: <http://dx.doi.org/10.1016/j.transproceed.2014.07.010>
 105. Patel F, Marusina AI, Duong C, Adamopoulos IE, Maverakis E. NKG2C, HLA-E and their association with psoriasis. Vol. 22, *Experimental Dermatology*. 2013. p. 797–9.
 106. Iwaszko M, Świerkot J, Kolossa K, Jeka S, Wiland P, Bogunia-Kubik K. Polymorphisms within the human leukocyte antigen-E gene and their associations with susceptibility to rheumatoid arthritis as well as clinical outcome of anti-tumour necrosis factor therapy. *Clin Exp Immunol* [Internet]. 2015 Dec [cited 2016 May 19];182(3):270–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26307125>

2. JUSTIFICATIVA

O *HLA-E* é um importante gene que codifica molécula com papel duplo, atuando na resposta imunológica inata e adaptativa. Sua principal função é a imunomodulação, principalmente modulando a resposta de células NK envolvida na imunovigilância, porém também é capaz de apresentar antígenos e interagir com às células T CD8+, participando assim da imunidade contra infecções. Embora considerado conservado, principalmente na região codificadora, variações no gene *HLA-E* têm o potencial de alterar a função da molécula codificada e/ou modificar seu padrão de expressão. Ainda não se conhece o perfil de variabilidade da região promotora do *HLA-E*. Assim, determinar os polimorfismos presentes nessas regiões codificadoras e regulatórias poderá auxiliar no entendimento de como o gene é regulado e qual o padrão de variabilidade de sua molécula em populações miscigenadas como a brasileira, além de servir como informação para estudos funcionais futuros e discussões acerca da história evolutiva desse gene.

3. OBJETIVO

O presente trabalho teve como objetivo avaliar a estrutura, variabilidade e haplótipos do gene *HLA-E*, considerando sua região promotora distal (aproximadamente 2 kb antes do início do gene) até 200 pb após o término do gene, incluindo todos os íntrons, em uma amostra brasileira do estado de São Paulo. Para se atingir estes objetivos, inicialmente foi elaborada uma metodologia que permite a avaliação da variabilidade do *HLA-E* por sequenciamento de nova geração. A metodologia, assim como os resultados e sua discussão estão descritos no artigo a seguir.

Capítulo II

4. ARTIGO

HLA-E REGULATORY AND CODING REGION VARIABILITY AND HAPLOTYPES IN A BRAZILIAN POPULATION SAMPLE

Jaqueline Ramalho¹, Eduardo Antônio Donadi², Celso Teixeira Mendes-Junior³, Erick da Cruz Castelli^{1,4*}

¹ Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPEX), 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

4.1 Introduction

The *HLA-E* gene is a non classical Human Leukocyte Antigen (HLA) class I loci and a member of the human Major Histocompatibility Complex (MHC). The *HLA-E* locus lays between two of the most variable genes of the human genome (*HLA-A* and *HLA-C*), but, thus far, it is quite conserved in terms of nucleotide variability encoded protein molecules compared to other *HLA* genes (1,2). According to the International ImmunoGenetics (IMGT/HLA) database, version 3.27.0, *HLA-E* presents 25 different coding alleles (or coding sequences) encoding eight different full-length protein molecules (3).

The HLA-E molecule may affect both the innate and adaptive immunity (4–9). HLA-E is broadly expressed in small amounts in a variety of tissues (10–13) and interacts mainly with Natural Killer (NK) cells through the CD94/NKG2A inhibitory receptors (14–20). In addition, HLA-E may also interact and activate T CD8⁺ cells through the T cell receptor (TCR) (4,7,8,21). The HLA-E molecule binds nonamers peptides derived from HLA class I signal sequences, mainly from HLA-G, thus, it presents self-antigens in a transporter associated with antigen processing (TAP) dependent way (17,22–24). This peptide binding is related to an immune surveillance mechanism and it indicates that the expression machinery and assembly of HLA molecules is operating properly, thus, the NK cytotoxicity should not operate (14,24,25). However, in situations such as tumors and/or infected cells, in which the HLA class I expression profile might be altered and no HLA-E expressed at the cell surface, NK cells may act against the affected cells, influencing the outcome by this immune surveillance mechanism via HLA-E (9).

In addition, in this same situation, the NK cell cytotoxicity inhibition via the HLA-E and CD94/NKG2A interaction might configure an escape mechanism when HLA-E is expressed at the cell surface associated with non self antigens (from pathogens, for instance) (7–9,26–32). Besides the HLA class I signal sequence-derived peptides, non-self-peptides may stabilize the HLA-E molecule and stimulate its expression on the cell surface. Among them we may find peptides derived from viral proteins such as gpUL40 (HCMV, HCMV), BZLF-1 (Epstein-Barr virus, EBV), InflM (influenza matrix protein), Hepatitis C virus (core protein, HCV) and protein gag (Human Immunodeficiency virus, HIV). In addition, bacterial peptides such as from *Salmonella enterica* and *Mycobacterium tuberculosis* and other self peptides such as those related to cellular stress and thermal shock (Hps60), peroxiredoxin isoforms 5 (Prdx5D2 and Prdx5D2,3), ABC MRP7 (ATP-binding cassette transporter,

multidrug resistance associated protein), gliadin and variants and V β 2 TCR V β 1, also may bind to the HLA-E molecule (4,7–9).

Moreover, mainly in the chronic infection and cellular stress contexts, HLA-E may also interact and present these different peptides to T CD8⁺ lymphocytes restricted to HLA-E, activating them against altered cells (33). Thus, HLA-E operates either modulating the immune response by interacting with CD94/NKG2 receptors of NK cells, either presenting antigens and triggering the cytotoxic T cells by TCR (4,7–9,21). Notwithstanding that, the HLA-E binding peptide repertoire is quite small compared to other HLA molecules.

This low peptide repertoire is consistent with the *HLA-E* low variability observed so far (34–38). Although eight different encoded protein molecules are already described for HLA-E, only two are actually frequent worldwide. These frequent ones are known as E*01:01 and E*01:03, which are different in just one amino acid encoded at exon 3, Arginine for E*01:01 and Glycine for E*01:03. This protein conservation may be associated with its key role modulating immune responses and also during pregnancy, since the fetal HLA-E must interact with the maternal NK cells (39–43).

Although *HLA-E* seems to be conserved in the coding region, since many studies reported only two main protein molecules as addressed earlier, these studies were in general conducted using molecular techniques that allows the evaluation of only known variants, or only few segments of the gene. In fact, many studies reported the variability of only exons 2 and 3 and few ones reported a continuous segment for the coding region or regulatory segments (34–38). In addition, the regulatory segments, specially the promoter region, were not properly explored. In fact, only three studies reported variability in the regulatory segments, two addressing the variability of the 3'untranslated region (UTR) segment (35,38) and another addressing the proximal promoter (44). Variable sites at regulatory segments may influence gene expression levels by different mechanisms, which include differential binding of transcriptional factors, chromatin remodeling and the binding of microRNAs, for instance.

In this study, we present a methodology to evaluate the entire *HLA-E* segment, which includes the extended promoter, complete coding sequence with introns and the complete 3'UTR segment, by using massive parallel sequencing. This methodology was applied to evaluate 420 samples from the Southeast Brazil, in order to characterize extended haplotypes for the *HLA-E* locus.

4.2 Material and Methods

4.2.1 Samples

A total of 420 unrelated individuals from the State of São Paulo, Brazil, were randomly recruited and peripheral blood sample was collected. These samples do not necessarily represent Brazilian population, however, considering the Brazil admixed nature they are heterogeneous samples and were mainly used for methodological goals. All participants gave written informed consent before blood withdraw and the local Human Research Ethics Committee did approve the study protocol (Protocol 24157413.7.000.5411). DNA was obtained by a salting out procedure, quantified using Qubit Broad Range Assays (Thermo Fisher Scientific Inc., Waltham, MA) and normalized to a final concentration of 50 ng/μL.

4.2.2 *HLA-E amplification, library preparation and sequencing*

The *HLA-E* locus was amplified as a unique amplicon of approximately 7,694 bp, encompassing its extended promoter segment (2,775 nucleotides upstream the first translated ATG, excluding primer forward segment), the complete coding sequence including all introns, the complete 3'UTR segment and 206 nucleotides downstream the *HLA-E* locus, excluding primer reverse segment, according to the annotations provided by human genome draft (versions hg19 or hg38). Polymerase Chain Reaction (PCR) was carried out using primers HEPR.F1: 5'- GCTTCGCAGTGAATGTGGCA - 3' and HE3UTR.R1: 5'- GGACTCCCTGGGCTTTCTCACCG - 3', in a final volume of 50 μL, containing 1.50U of DNA polymerase and 0.8X de buffer solution (Long PCR Enzyme Mix, Thermo Fisher Scientific Inc., Waltham, MA, USA), 1.5mM of MgCl₂, 0.20mM of each dNTP (Invitrogen – Carlsbad, CA, USA) and 0.30μM of each primer. PCR cycling conditions are as follow: 94°C by 5 minutes, 35 cycles of 94°C by 30 seconds, 62°C by 45 seconds and 68°C by 8 minutes, finalizing with 68°C by 4 minutes and hold of 4°C. Amplicons were evaluated by electrophoresis on 1% agarose gel stained with GelRed® (Biotium, Inc. Hayward, CA). Although only the *HLA-E* variability and haplotypes are reported here, *HLA-A*, *HLA-B*, *HLA-C*, *HLA-F* and *HLA-G* were also amplified and sequenced together with *HLA-E*, comprising the complete study on the *HLA* class I genes variability.

For library preparation, we followed the conditions recommended by Illumina NGS protocols for amplicon quantification and sample pooling. Libraries were prepared using Nextera XT Sample Preparation Kit and multiplexed with the Nextera XT Index Kit (Illumina, Inc., San Diego, CA). Library quantification was performed by qPCR using Kapa

(Kapa Biosystems, Wilmington, MA) and the fragmentation pattern was evaluated using Agilent High Sensitivity DNA Kits (Agilent Technologies, Santa Clara, CA). Finally, libraries were normalized to the recommended concentration for MiSeq Reagent Kit version V2 (500 cycles, 2 x 250-bp).

4.2.3 Data processing and analysis

All sequences (reads) produced were firstly trimmed for adapters, primer sequences. Sequences were also processed to remove low quality 5' and 3' ends using *seqtk trimfq* (45) with an error rate threshold set to 0.02. After that, to get a reliable mapping of the *HLA*-related sequences, we used ***hla-mapper*** version 0.6 (database version 1.7) (38,46) to generate the *HLA-E*-specific aligned files (BAM format) to the human reference genome version hg19. This strategy was firstly introduced evaluating the *HLA-E* variability in two African population samples (38) and recently also evaluating the *HLA-F* variability in Brazilian samples (46).

After mapping, genotypes were inferred using the Genome Analysis Toolkit (GATK, version 3.6) (47–49): first, we inferred genotypes running the routine HaplotypeCaller per-sample using the GVCF mode with a minimum base quality score set to 20 and not using soft-clipped segments. Genotypes were inferred in the chromosome 6 between nucleotide 30,454,700 and 30,462,100, considering the human genome draft version hg19. Then, a multi-sample VCF file was created joining the g.vcf files using the GATK GenotypeGVCFs routine. Variants were annotated based on the dbSNP database version 146. The VCF file was then processed by *vcfx* (described in details previously (38,46)) version 0.10.1 with default parameters, which guarantees that only high quality genotypes are passed to a further imputation step. The *vcfx*-treated VCF file was manually inspected, mainly in two segments (one at the promoter and one at intron 5) that present low complexity and the presence of repetitive nucleotides. In these segments, many variable sites that were detected by GATK and presented a severe unbalance between alleles were manually removed. This phenomenon was particularly evident at the segment comprising the *HLA-E* nucleotides -400 to -900 (considering the Adenine of the first translated ATG as nucleotide +1), since almost all variable sites detected here by GATK was further removed because of low quality genotypes, exception made to a variable site at position -682. It is not clear whether these removed variable sites really exist or they are artifacts because of the low complexity of this segment, as later discussed.

The association between each variable site (to infer the two *HLA-E* sequences of each individual) was inferred using GATK routine ReadBackedPhasing using a minimal Phase Quality Threshold of 2000, which is 100 times higher than the default value. This assures that only alleles from variable sites that are close enough to be present in a same fragment (same read) should be phased. Considering that, sometimes, variable sites are quite distant from each other in a same sample, not all variable sites were straightforwardly phased. This is particularly evident for non classical *HLA* genes such as *HLA-E*, in which, in general, the variable sites from an individual are quite distant from each other. Moreover, ReadBackedPhasing does not perform phase inference on indels and multiallelic loci. In the present series, 24.46% of the heterozygous sites were phased using GATK. The remaining 75.54% were evaluated using the PHASE algorithm (50), which also imputed the 0.087% of missing alleles after the *vcfx* treatment.

To proceed with the PHASE algorithm analysis, the partially GATK-phased VCF file (obtained with the ReadBackedPhasing algorithm) was converted into an input file for PHASE and an accessory file containing the known phases between variable sites obtained previously by GATK. In some situations, blocks of known phase among variable sites, but with unknown phase among the blocks, were generated and the PHASE algorithm with 1000 iterations was used to infer the phase between these blocks until a complete pair of haplotypes is defined. For the *HLA-E* gene, different from what was observed for others *HLA* class I genes, because of its low variability and the distance between each variable site, the phase between the distal promoter and the coding segment was mainly inferred by the PHASE algorithm and not GATK. After the final PHASE run, it was verified that all inferred haplotype pairs were compatible with the known phases (the ones determined by GATK ReadBackedPhasing algorithm). The output PHASE file was converted using local *Perl* scripts into a final VCF file with all variable sites phased. The variable sites with just once occurrence (singletons) were not considered for the PHASE analysis. However, whenever possible, they were manually introduced in the final VCF file.

Allele, genotype and haplotype frequencies were counted directly. Nucleotide and gene diversities, as well as the adherence of the genotype frequencies under the Hardy-Weinberg expectations, were evaluated using Arlequin 3.5 (51). The linkage disequilibrium was inferred using the software Haploview 4.2 (52), and considering only variable sites with a minor allele frequency higher than 1%.

4.3 Results

Considering the *HLA-E* region evaluated between nucleotide 30,454,700 and 30,462,100 (human genome draft version hg19), which comprises the *HLA-E* 5' distal and proximal promoter, the complete coding sequence with all introns included, the complete 3' UTR and few nucleotides downstream *HLA-E*, we detected 63 variable sites (Table 1), in which all were included in the haplotype analysis. Many of these (23,81%) variable sites occurred just once (as singletons, noted as “D” at Table 1), but they were included in the haplotype analysis because they were properly phased by GATK or they occurred at complete homozygous samples exception made to that singleton. At least 22 variable sites might be considered true polymorphisms at this population sample, presenting a minor allele frequency (MAF) higher than 1% (Table 1). However, some of these sites were detected with a MAF higher than or close to 30%. Among them, we may find positions -2143 and -2015 upstream the *HLA-E* coding sequence, +424 and +756 at exons 2 and 3, respectively, and +3777 and +4776 at the 3'UTR/downstream segments (Table 1). All the variable sites detected and their genotype frequencies are under the Hardy-Weinberg equilibrium expectations.

Table 1. List of variable sites detected at the *HLA-E* gene on a Brazilian population sample considering the segment between nucleotide -2143 and +4776 and the Adenine of the first translated ATG as nucleotide +1

Chr6 Position ^a	SNPid	Notes ^b	HLA-E segment ^c	IMGT/HLA relative position ^d	Reference allele ^e	Alternative allele	Alternative allele frequency (2n=840)
30455166	rs2078675	.	5' upstream	-2143	T	C	0.2536
30455167	rs188844729	.	5' upstream	-2142	G	A	0.0036
30455186	rs139422860	.	5' upstream	-2123	G	A	0.0131
30455203	rs17875359	.	5' upstream	-2106	G	A	0.0690
30455240	rs28780108	.	5' upstream	-2069	G	C	0.0488
30455294	rs12207974	.	5' upstream	-2015	C	G	0.2750
30455321	rs17875360	.	5' upstream	-1988	T	C	0.1345
30455364	rs139555215	.	5' upstream	-1945	G	T	0.0071
30455429	rs147590779	.	5' upstream	-1880	A	G	0.0036
30455436	.	D	5' upstream	-1873	G	A	0.0012
30455532	rs3757335	.	5' upstream	-1777	G	T	0.0024
30455541	.	D	5' upstream	-1768	A	C	0.0012
30455846	rs188359497	D	5' upstream	-1463	C	T	0.0012
30455886	rs762324	.	5' upstream	-1423	G	A	0.0881
30455903	.	D	5' upstream	-1406	C	T	0.0012
30455920	rs1264459	.	5' upstream	-1389	G	A	0.8429
30455970	.	D	5' upstream	-1339	T	C	0.0012
30456063	.	.	5' upstream	-1246	C	T	0.0024
30456080	rs140409752	.	5' upstream	-1229	C	G	0.0048
30456093	rs17875363	.	5' upstream	-1216	A	T	0.0024
30456142	rs116253207	.	5' upstream	-1167	A	G	0.0119

30456150	rs17875364	.	5' upstream	-1159	A	G	0.0750
30456151	rs17875365	.	5' upstream	-1158	T	C	0.0024
30456192	rs574018332	.	5' upstream	-1117	C	G	0.0071
30456230	rs146647219	.	5' upstream	-1079	G	T	0.0071
30456316	rs17875366	.	5' upstream	-993	G	A	0.0024
30456627	rs138267201	.	5' upstream	-682	AG	A	0.0083
30456931	rs186882745	.	5' upstream	-378	G	C	0.0024
30457196	rs141224659	C	5'UTR	-113	T	C	0.0060
30457204	rs9468784	C	5'UTR	-105	A	G	0.0036
30457205	rs61356961	C	5'UTR	-104	A	G	0.0238
30457283	rs76971248	.	5'UTR	-26	G	T	0.0167
30457732	rs1059510	A	Exon 2	424	T	C	0.7167
30457766	rs150949676	B	Exon 2	458	G	A	0.0024
30458064	rs1264457	B	Exon 3	756	G	A	0.5738
30458279	rs145034129	A, C, D	Exon 3	971	G	A	0.0012
30458298	rs370425404	C, D	Intron 3	990	A	C	0.0012
30458322	rs41265828	.	Intron 3	1014	T	A	0.0060
30458502	.	C, D	Intron 3	1194	A	G	0.0012
30458586	rs182627071	C	Intron 3	1278	C	T	0.0024
30458591	rs114425530	.	Intron 3	1283	G	A	0.0071
30458630	rs116563630	C	Intron 3	1322	G	A	0.0024
30458933	rs11548296	A	Exon 4	1625	G	C	0.0238
30458999	rs17875370	D	Exon 4	1691	G	A	0.0012
30459165	rs62621992	B	Exon 4	1857	C	T	0.0060
30459296	rs200270359	.	Intron 4	1988	G	A	0.0024
30459577	rs183165297	C	Intron 5	2269	T	C	0.0024
30459935	rs41543014	.	Intron 5	2627	G	A	0.0024
30460232	rs17875371	.	Intron 6	2924	C	T	0.0393
30460245	rs147326588	C, D	Intron 6	2937	G	A	0.0012
30460253	rs374002055	C, D	Intron 6	2945	T	C	0.0012
30460755	rs17195369	C	Exon 8/3' UTR	3447	C	T	0.0048
30460776	rs74295295	C	Exon 8/3' UTR	3468	A	C	0.0333
30460798	rs765254483	C	Exon 8/3' UTR	3490	C	A	0.0036
30460942	rs17195376	.	Exon 8/3' UTR	3634	G	A	0.0369
30460999	.	D	Exon 8/3' UTR	3691	A	C	0.0012
30461085	rs1059655	.	Exon 8/3' UTR	3777	G	A	0.7429
30461086	rs115051198	.	Exon 8/3' UTR	3778	A	G	0.0226
30461574	.	D	Exon 8/3' UTR	4266	G	A	0.0012
30461605	rs9283	.	Exon 8/3' UTR	4297	G	A	0.0357
30461728	rs566930407	D	Exon 8/3' UTR	4420	C	T	0.0012
30461844	.	D	Exon 8/3' UTR	4536	T	C	0.0012
30462084	rs1264456	.	3' downstream	4776	A	G	0.7429

The region tracked by the IMGT/HLA database is represented with a grey background.

^aChromosome 6 position considering the human genome draft version hg19.

^b**NOTES:** A) Synonymous mutation on exon

B) Non-synonymous mutation on exon

C) New variable site on region covered by the IMGT/HLA database (release 3.27.0, 2017-01-20)

D) Singleton with a defined haplotype

^c *HLA-E* segment: Considering the *HLA-E* mRNA structure annotated both at the human genome drafts hg19 and hg38, variable sites were classified as 5' upstream (from -2143 to -127), 5' untranslated region [5'UTR] (from -126 to -1), as encompassing Exons or on Introns for variable sites between the first translated ATG and the stop codon, as 3' untranslated region [3'UTR]/Exon 8 (from +3218 to +4674), and as 3' downstream (after +4675).

^d IMGT/HLA relative position, considering the Adenine of the first translated ATG as nucleotide +1.

^e The reference allele at the human genome draft (version hg19); it is not necessarily the minor frequent allele.

Beside singletons shown at Table 1, additional 25 singletons were detected (Table 2). However, given the great distance between these singletons and the nearest variable sites, and since they occurred just once in samples carrying other heterozygous sites, or some samples presented missing alleles at this particular variable site, it was not possible to properly phase these singletons. Thus, they were not included in the haplotype analysis, mainly because the PHASE algorithm does not handle singletons properly. Considering these singletons and the segment between nucleotides -300 and +3522, which is the *HLA-E* region tracked by the IMGT/HLA database, 16 sites at Table 2 (in shades of gray) might be considered new variable sites, including two non-synonymous mutations at positions +1644 (exon 4) and +2009 (exon 5) (noted as “B” at Table 2), 11 at intron segments and one at the 3’ untranslated region (UTR). In addition, a known non-synonymous mutation at exon 4 was also detected at position +1822, which characterizes the allele E*01:09 (Table 2).

Table 2. List of variable sites without defined haplotype detected at the *HLA-E* gene on a Brazilian population sample considering the segment between nucleotide -2143 and +4776 and the Adenine of the first translated ATG as nucleotide +1

Chr6 position ^a	SNPId	Notes ^b	HLA-E segment ^c	IMGT/HLA relative position ^d	Reference allele ^e	Alternative allele
30455328	rs555000642	D	5' upstream	-1981	G	A
30455405	.	D	5' upstream	-1904	A	G
30455628	.	.	5' upstream	-1681	T	C
30456121	rs192094585	.	5' upstream	-1188	A	T
30456341	.	.	5' upstream	-968	G	A
30456939	.	.	5' upstream	-370	A	G
30457465	.	C, D	Intron 1	157	G	A
30457469	.	C, D	Intron 1	161	G	A
30457797	.	C, D	Intron 2	489	A	T
30458054	.	A, C	Exon 3	746	G	A
30458952	rs149396632	B (Asp→Asn), C	Exon 4	1644	G	A
30459130	rs141487266	B (E*01:09)	Exon 4	1822	A	C
30459232	rs199731038	C	Intron 4	1924	C	T
30459317	rs148162840	B (Pro→His), C	Exon 5	2009	C	A
30459384	rs146142563	A, C	Exon 5	2076	C	T
30459513	.	C	Intron 5	2205	G	A
30459540	rs193246407	C	Intron 5	2232	C	T
30459553	.	C	Intron 5	2245	C	A
30459848	.	C	Intron 5	2540	G	A
30460038	rs772726027	C	Intron 5	2730	C	T
30460292	rs374005814	C, D	Intron 6	2984	C	T
30460471	rs760930383	C	Intron 7	3163	T	G
30460778	.	C, D	Exon 8/3'UTR	3470	A	T
30461113	.	D	Exon 8/3'UTR	3805	A	G
30461634	rs570363365	(3UTR-6)	Exon 8/3'UTR	4326	G	A

The region tracked by the IMGT/HLA database is represented with a grey background.

^aChromosome 6 position considering the human genome draft version hg19.

^b**NOTES:** A) Synonymous mutation on exon

B) Non-synonymous mutation on exon

C) New variable site on region covered by the IMGT/HLA database (release 3.27.0, 2017-01-20)

D) Variable site detected certainly at once but with missing alleles

Others: The non-synonymous mutation at exon (B) are accompanied with notes about the associated allele at IMGT/HLA or the amino acid change. Asp: Aspartic acid; Asn: Asparagine; Pro: Proline; His: Histidine.

^c *HLA-E* segment: Considering the *HLA-E* mRNA structure annotated both at the human genome drafts hg19 and hg38, variable sites were classified as 5' upstream (from -2143 to -127), 5' untranslated region [5'UTR] (from -126 to -1), as encompassing Exons or on Introns for variable sites between the first translated ATG and the stop codon, as 3' untranslated region [3'UTR]/Exon 8 (from +3218 to +4674), and as 3' downstream (after +4675).

^d IMGT/HLA relative position, considering the Adenine of the first translated ATG as nucleotide +1.

^e The reference allele at the human genome draft (version hg19); it is not necessarily the minor frequent allele.

The *HLA-E* variability and haplotypes will be presented following the segments defined at Figure 1. Firstly, it will be presented the variability and haplotypes detected at the approximately 2kb nucleotides upstream the first translated ATG, from -2143 to -1, which includes the *HLA-E* distal and proximal promoter and the 5' untranslated region (Table 3); secondly, to allow us name the coding haplotypes following the IMGT/HLA nomenclature, it will be presented the variability and haplotypes comprehending the *HLA-E* segment that is tracked by the IMGT/HLA database, i.e., from nucleotide -300 to +3522 (Table 4); thirdly, it will be presented the variability detected at exon 8, because this exon represents 99.86% of the entire mRNA 3'UTR segment (Table 5); finally, the haplotypes defined for each segment will be combined as extended haplotypes, including the most frequent variable site detected at the *HLA-E* 3'downstream segment, +4776 (Table 6).

The 5' upstream region presented 32 variable sites considering the 2.5 kb evaluated (Table 3), in which 12 present frequencies higher than 1% on this Brazilian population sample (Table 1). The *HLA-E* proximal promoter and 5' UTR segments are quite conserved, since only four low-frequency variable sites were detected between nucleotides -300 and -1. When the entire 5' upstream segment is considered, 37 haplotypes were detected (Table 3). Since this region is not entirely considered by the IMGT/HLA database and since there is no nomenclature pattern defined for it, haplotypes were named as Epromo-01 to -37 by descending frequency order detected in the present series (Table 3). Haplotype Epromo-01 is the most frequent (40.48%), followed by Epromo-02 (14.05%). These two haplotypes do differ at only one variable site, at position -1389. In addition, considering all haplotypes, only ten were found presenting frequencies higher than 1%.

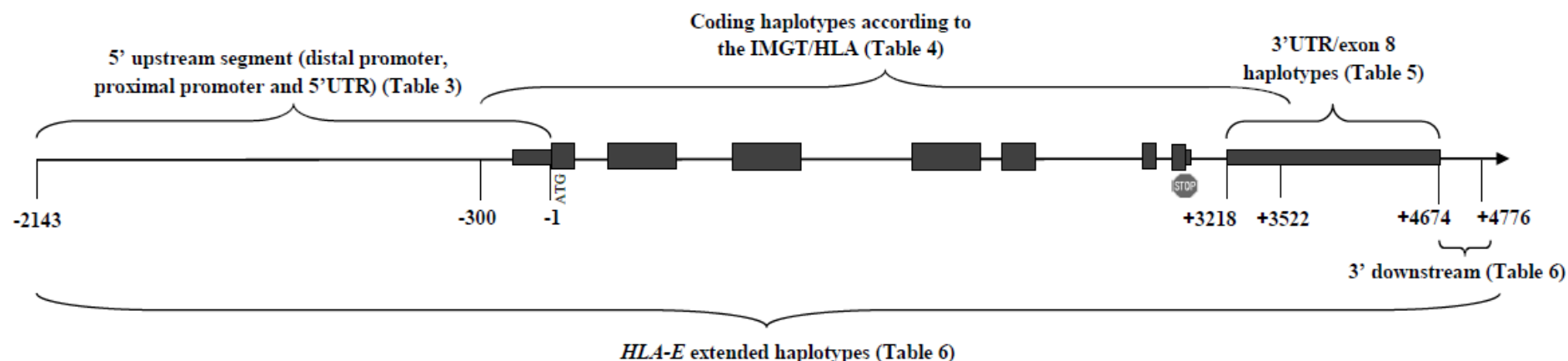


Figure 1. *HLA-E* segments for haplotype analysis. 5' upstream segment (distal promoter, proximal promoter and 5'UTR) variability and haplotypes are represented at Table 3, comprising nucleotides from -2143 to -1 (considering the Adenine of first ATG translated as nucleotide +1). Coding segment is covered by IMGT/HLA database, from nucleotide -300 to +3522, and these haplotypes are represented at Table 4. The 3'untranslated region (UTR)/ exon 8 haplotypes are represented at Table 5, considering nucleotides from +3218 to +4674. The extended *HLA-E* segment haplotypes are represented at Table 6 (considering all segments besides 3' downstream region up to nucleotide +4776).

Table 3. List of the *HLA-E* 5' upstream and 5' untranslated region haplotypes found in a Brazilian population sample.

Haplotypes ^a	Relative HLA-E positions according to the IMGT/HLA ^b																												Frequency (2n=840)				
	-2143	-2142	-2123	-2106	-2069	-2015	-1988	-1945	-1880	-1873	-1777	-1768	-1463	-1423	-1406	-1389	-1339	-1246	-1229	-1216	-1167	-1159	-1158	-1117	-1079	-993	-682	-378		-113	-105	-104	-26
Epromo-01	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.4048
Epromo-02	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	G	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.1405
Epromo-03	C	G	G	G	G	G	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.1000
Epromo-04	C	G	G	G	G	G	C	G	A	G	G	A	C	A	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0810
Epromo-05	T	G	G	A	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0619
Epromo-06	T	G	G	G	C	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0452

Epromo-07	C	G	G	G	G	G	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	G	T	C	G	G	AG	G	T	A	A	G	0.0381
Epromo-08	C	G	G	G	G	G	C	G	A	G	G	A	C	G	C	A	T	C	C	A	A	G	T	C	G	G	AG	G	T	A	G	G	0.0167
Epromo-09	T	G	A	G	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	T	0.0131
Epromo-10	T	G	G	G	G	G	C	G	A	G	G	A	C	G	C	A	T	C	C	A	G	A	T	C	G	G	AG	G	T	A	A	G	0.0119
Epromo-11	T	G	G	G	G	G	C	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	T	G	AG	G	T	A	A	G	0.0071
Epromo-12	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	A	G	T	A	A	G	0.0071
Epromo-13	C	G	G	G	G	G	C	G	A	G	G	A	C	G	C	A	T	C	C	A	A	G	T	C	G	G	AG	G	T	A	A	G	0.0071
Epromo-14	T	G	G	A	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	G	G	G	AG	G	T	A	A	G	0.0071
Epromo-15	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	G	T	C	C	A	A	G	T	C	G	G	AG	G	C	A	A	G	0.0060
Epromo-16	T	G	G	G	G	C	T	T	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0060
Epromo-17	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	G	T	C	G	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0048
Epromo-18	T	G	G	G	G	C	T	G	A	G	G	A	C	A	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	G	G	0.0036
Epromo-19	T	G	G	G	C	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	G	A	G	0.0036
Epromo-20	C	A	G	G	G	G	C	G	A	G	G	A	C	G	C	A	T	C	C	A	A	G	T	C	G	G	AG	G	T	A	G	G	0.0036
Epromo-21	T	G	G	G	G	C	T	G	G	G	A	C	G	C	G	T	C	C	A	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0036
Epromo-22	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	T	0.0036
Epromo-23	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	G	T	C	G	G	AG	G	T	A	A	G	0.0024
Epromo-24	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	G	T	C	C	A	A	A	T	C	G	G	AG	C	T	A	A	G	0.0024
Epromo-25	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	A	T	T	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0024
Epromo-26	T	G	G	G	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	T	A	A	C	C	G	A	AG	G	T	A	A	G	0.0024
Epromo-27	T	G	G	G	G	G	C	G	A	G	T	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0024
Epromo-28	C	G	G	G	G	G	C	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0012
Epromo-29	T	G	G	G	G	G	C	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0012
Epromo-30	C	G	G	G	G	G	C	G	A	G	G	A	C	A	C	A	C	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0012
Epromo-31	C	G	G	G	G	G	T	G	A	A	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0012
Epromo-32	T	G	G	G	G	C	T	G	A	G	G	A	T	G	C	A	T	C	C	A	A	G	T	C	G	G	AG	G	T	A	A	G	0.0012
Epromo-33	T	G	G	G	G	C	T	T	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	A	G	T	A	A	G	0.0012
Epromo-34	C	G	G	G	G	C	T	G	A	G	G	A	C	G	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0012
Epromo-35	C	G	G	G	G	G	C	G	A	G	G	C	C	A	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0012
Epromo-36	C	G	G	G	G	G	T	G	A	G	G	A	C	G	T	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0012
Epromo-37	T	G	G	G	G	C	T	G	A	G	G	A	C	A	C	A	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0012
Nucleotide diversity																											0.0010 +/- 0.0006						
Gene diversity																											0.7925 +/- 0.0115						

The alternative alleles are represented with a grey background.

^a Haplotypes were named considering the frequency in descending order.

^b The position was calculated following the pattern used by the IMGT/HLA database, i.e., nucleotide +1 is the Adenine of the first translated ATG.

Considering the segment covered by the IMGT/HLA database, i.e., from nucleotide -300 to +3522, with approximately 3.8 kb, we detected 26 variable sites (Table 1), in which 4 were detected before the first translated ATG, 7 in exons, 12 in introns and 3 after the stop codon. Of those, 14 might be considered new variable sites according to the IMGT/HLA database version 3.27.0. The variable sites at positions -105, +1194, +2627, +2924, +2937, +2945 and +3490 (Table 1) were detected only on this Brazilian population sample so far (3,35,37,38,44). The new variable sites detected here (noted as “C” at Table 1 and Table 2) occur mostly in intron or untranslated sequences.

The variable sites at the coding segment were found arranged into 27 haplotypes, which were named following the haplotypes available at the IMGT/HLA database, version 3.27.0 (Table 4). Most of the sequences were identical to the ones already described by the aforementioned database, which includes copies of the alleles E*01:01:01:01, E*01:01:01:03, E*01:01:01:04, E*01:01:01:06, E*01:01:01:08, E*01:03:01:01, E*01:03:01:02, E*01:03:02:01, E*01:03:02:02, E*01:03:04, E*01:03:05, E*01:05 and E*01:06, counting for 95.36% of all coding *HLA-E* sequences. The remaining sequences configure new *HLA-E* coding alleles. In these cases, the closest known *HLA-E* allele was identified, followed by the mutations observed. The new *HLA-E* alleles are probably derived from E*01:01:01:01, E*01:03:01:01 and E*01:03:02:01, with novel mutations in different segments (Table 4).

Although 27 haplotypes were detected, they encode only four different *HLA-E* molecules. This phenomenon is related to the fact that most of the variable sites detected are synonymous mutations in exons or mutations at intron and untranslated region segments (Table 1). Two encoded molecules are frequently found, i.e., E*01:01 and E*01:03, with frequencies of 57.38% and 41.79%, respectively. However, it should be noted that an additional encoded molecule with just one occurrence would be detected if the variable sites +1822, related to the E*01:09 allele, would be included in the haplotype analysis.

Table 4. List of *HLA-E* coding haplotypes or coding alleles found in Brazilian population sample

HLA-E allele or HLA-E coding haplotype ^a	Encoded molecule ^b	Frequency (2n=840)
E*01:01:01:01	E*01:01	0.4750
E*01:01:01:01 ^(new +1194G)	E*01:01	0.0012
E*01:01:01:01 ^(new +1278T)	E*01:01	0.0024
E*01:01:01:01 ^(new +2937A, new +2945C)	E*01:01	0.0012
E*01:01:01:01 ^(new +3447T)	E*01:01	0.0048
E*01:01:01:01 ^(new +3468C)	E*01:01	0.0155
E*01:01:01:01 ^(new +3490A)	E*01:01	0.0036

E*01:01:01:01 ^(+424T)	E*01:01	0.0012
E*01:01:01:01 ^(new -105G)	E*01:01	0.0036
E*01:01:01:03	E*01:01	0.0393
E*01:01:01:04	E*01:01	0.0071
E*01:01:01:06	E*01:01	0.0167
E*01:01:01:08	E*01:01	0.0024
E*01:03:01:01	E*01:03	0.1012
E*01:03:01:01 ^(new +1322A)	E*01:03	0.0012
E*01:03:01:01 ^(new +3468C)	E*01:03	0.0012
E*01:03:01:01 ^(new +971A, new +1322A)	E*01:03	0.0012
E*01:03:01:01 ^(new +990C)	E*01:03	0.0012
E*01:03:01:01 ^(new -113C)	E*01:03	0.0036
E*01:03:01:02	E*01:03	0.0024
E*01:03:02:01	E*01:03	0.2726
E*01:03:02:01 ^(new +2269C)	E*01:03	0.0024
E*01:03:02:02	E*01:03	0.0060
E*01:03:04	E*01:03	0.0012
E*01:03:05 ^{compatible}	E*01:03	0.0238
E*01:05 ^{compatible}	E*01:05	0.0024
E*01:06	E*01:06	0.0060
Nucleotide diversity		0.0003 +/- 0.0002
Gene diversity		0.6878 +/- 0.0121

^a Haplotype names were given according to the closest known *HLA-E* allele followed by the differences/divergences that were observed for this given haplotype. The segment that is considered by the IMGT/HLA database starts on nucleotide -300 up to nucleotide +3522. The word "new" refers to variable sites that are not currently described at the IMGT/HLA database, version 3.27.0. The word "compatible" indicates that the sequence found in Brazil is compatible with the sequence defined by the IMGT/HLA database, but these alleles are only partially characterized on the IMGT/HLA database.

^b The encoded *HLA-E* molecule considering the full-length mRNA encoded by this haplotype.

The *HLA-E* exon 8 accounts for 99.86% of the entire mRNA 3'UTR segment, with approximately 1.4 kb. In addition, no variable sites were detected after the stop codon and before position +3447, which is inside exon 8. It was detected 12 haplotypes considering exon 8 (Table 5), and 5 of them were already described in former studies regarding *HLA-E* variability in African samples (3UTR-1, 3UTR-2, 3UTR-3, 3UTR-4 and 3UTR-5) (38). In addition, new haplotypes were detected and named as 3UTR-8 to 3UTR-14 (Table 5) considering the descending frequency order, since haplotypes 3UTR-6 and 3UTR-7 was detected in Africa (38) but not in the present series. Nevertheless, this segment is quite conserved, since almost 90% of all sequences refer to haplotypes 3UTR-1 or 3UTR-2, which differ by only position +3777 (Table 5).

Table 5. List of haplotypes at the *HLA-E* 3' untranslated region (exon 8) detected on Brazilian population sample

Haplotypes ^a	Relative HLA-E positions according to the IMGT/HLA ^b											Frequency (2n=840)
	3447	3468	3490	3634	3691	3777	3778	4266	4297	4420	4536	
3UTR-1	C	A	C	G	A	A	A	G	G	C	T	0.6381
3UTR-2	C	A	C	G	A	G	A	G	G	C	T	0.2560
3UTR-3	C	C	C	G	A	A	A	G	G	C	T	0.0333
3UTR-4	C	A	C	A	A	A	A	G	A	C	T	0.0357
3UTR-5	C	A	C	G	A	A	G	G	G	C	T	0.0226
3UTR-8	T	A	C	G	A	A	A	G	G	C	T	0.0048
3UTR-9	C	A	A	G	A	A	A	G	G	C	T	0.0036
3UTR-10	C	A	C	G	C	G	A	G	G	C	T	0.0012
3UTR-11	C	A	C	A	A	A	A	G	G	C	T	0.0012
3UTR-12	C	A	C	G	A	A	A	G	G	C	C	0.0012
3UTR-13	C	A	C	G	A	A	A	G	G	T	T	0.0012
3UTR-14	C	A	C	G	A	A	A	A	G	C	T	0.0012
Nucleotide diversity											0.0005 +/- 0.0004	
Gene diversity											0.5250 +/- 0.0155	

The alternative alleles are represented with a grey background.

^a Haplotype were named following the same pattern proposed in a previous HLA-E study (Castelli *et al.*, 2015)

^b The position was calculated following the pattern used by the IMGT/HLA database, i.e., nucleotide +1 is the Adenine of the first translated ATG.

The complete *HLA-E* segment explored here presented 75 extended haplotypes, with only 13 presenting a frequency higher than 1% (Table 6). It can be noticed an association mainly between (a) Epromo-01 and derived haplotypes (Epromo-05, -06, -09, -12, -14, -16, -18, -19, -22, -25, -26, -33 and -37 at Table 3), E*01:01:01 and 3UTR-1; (b) Epromo-02 and derived haplotypes (Epromo-15, -17, -21 and -24, at Table 3), E*01:03 and 3UTR-2. However, several different haplotypes can be detected, indicating that the relationship among promoter, coding and 3'UTR haplotypes are not straightforward. Nevertheless, haplotypes carrying -104G (Epromo-08, -18 and -20, at Table 3) was found associated with E*01:03:05^{compatible} and 3UTR-1 and the derived haplotype 3UTR-13, which was previously observed elsewhere (38), and haplotypes with -26T (Epromo-09 and -22, at Table 3) was found associated with E*01:01:01:06 and 3UTR-3.

Table 6. *HLA-E* extended haplotypes detected on a Brazilian population sample, considering the segment between nucleotide -2143 and +4776

5' Upstream/5'UTR Haplotypes ^a	HLA-E coding allele ^b	3'UTR/exon8 Haplotypes ^c	+4776 ^d	Frequency (2n=840)
Epromo-01	E*01:01:01:01	3UTR-1	G	0.3500
Epromo-02	E*01:03:02:01	3UTR-2	A	0.1179
Epromo-03	E*01:03:02:01	3UTR-2	A	0.0917
Epromo-04	E*01:03:01:01	3UTR-1	G	0.0726
Epromo-05	E*01:01:01:01	3UTR-1	G	0.0417
Epromo-06	E*01:01:01:01	3UTR-1	G	0.0381
Epromo-07	E*01:03:02:01	3UTR-4	G	0.0357
Epromo-01	E*01:01:01:03	3UTR-2	A	0.0179
Epromo-01	E*01:01:01:01 (new +3468C)	3UTR-3	G	0.0155
Epromo-08	E*01:03:05 ^{compatible}	3UTR-1	G	0.0155
Epromo-09	E*01:01:01:06	3UTR-3	G	0.0131
Epromo-10	E*01:01:01:01	3UTR-5	G	0.0119
Epromo-02	E*01:03:02:01	3UTR-1	G	0.0107
Epromo-05	E*01:01:01:03	3UTR-1	G	0.0095
Epromo-13	E*01:01:01:01	3UTR-1	G	0.0071
Epromo-14	E*01:01:01:01	3UTR-1	G	0.0071
Epromo-05	E*01:01:01:04	3UTR-1	G	0.0071
Epromo-11	E*01:03:01:01	3UTR-5	G	0.0071
Epromo-12	E*01:03:01:01	3UTR-1	G	0.0071
Epromo-16	E*01:01:01:03	3UTR-1	G	0.0060
Epromo-03	E*01:03:02:02	3UTR-2	A	0.0060
Epromo-04	E*01:06	3UTR-1	G	0.0060
Epromo-02	E*01:01:01:01	3UTR-1	G	0.0048
Epromo-01	E*01:01:01:03	3UTR-1	G	0.0048

Epromo-17	E*01:03:02:01	3UTR-2	A	0.0048
Epromo-06	E*01:01:01:01	3UTR-2	A	0.0036
Epromo-19	E*01:01:01:01 ^(new -105G)	3UTR-1	G	0.0036
Epromo-01	E*01:01:01:01 ^(new +3490A)	3UTR-9	G	0.0036
Epromo-22	E*01:01:01:06	3UTR-3	G	0.0036
Epromo-15	E*01:03:01:01 ^(new -113C)	3UTR-1	G	0.0036
Epromo-21	E*01:03:02:01	3UTR-2	A	0.0036
Epromo-18	E*01:03:05 ^{compatible}	3UTR-1	G	0.0036
Epromo-20	E*01:03:05 ^{compatible}	3UTR-1	G	0.0036
Epromo-25	E*01:01:01:01	3UTR-1	G	0.0024
Epromo-01	E*01:01:01:01 ^(new +1278T)	3UTR-1	G	0.0024
Epromo-05	E*01:01:01:01 ^(new +1988A)	3UTR-1	G	0.0024
Epromo-01	E*01:01:01:01 ^(new +3447T)	3UTR-8	G	0.0024
Epromo-26	E*01:01:01:01 ^(new +3447T)	3UTR-8	G	0.0024
Epromo-24	E*01:03:01:01	3UTR-1	G	0.0024
Epromo-06	E*01:03:01:01	3UTR-1	G	0.0024
Epromo-27	E*01:03:01:02	3UTR-5	G	0.0024
Epromo-02	E*01:03:02:01 ^(new +2269C)	3UTR-2	A	0.0024
Epromo-15	E*01:05 ^{compatible}	3UTR-1	G	0.0024
Epromo-02	E*01:01:01:01	3UTR-2	A	0.0012
Epromo-01	E*01:01:01:01	3UTR-14	G	0.0012
Epromo-34	E*01:01:01:01	3UTR-1	G	0.0012
Epromo-05	E*01:01:01:01	3UTR-2	A	0.0012
Epromo-03	E*01:01:01:01	3UTR-1	G	0.0012
Epromo-28	E*01:01:01:01	3UTR-1	G	0.0012
Epromo-01	E*01:01:01:01	3UTR-12	G	0.0012
Epromo-01	E*01:01:01:01 ^(new +1194G)	3UTR-1	G	0.0012
Epromo-01	E*01:01:01:01 ^(new +2937A, new +2945C)	3UTR-1	G	0.0012
Epromo-01	E*01:01:01:01 ^(+424T)	3UTR-1	G	0.0012
Epromo-06	E*01:01:01:03	3UTR-1	G	0.0012
Epromo-29	E*01:03:01:01	3UTR-5	G	0.0012
Epromo-33	E*01:03:01:01	3UTR-1	G	0.0012
Epromo-01	E*01:03:01:01	3UTR-1	G	0.0012
Epromo-30	E*01:03:01:01	3UTR-1	G	0.0012
Epromo-35	E*01:03:01:01	3UTR-1	G	0.0012
Epromo-04	E*01:03:01:01	3UTR-2	A	0.0012
Epromo-37	E*01:03:01:01	3UTR-1	G	0.0012
Epromo-32	E*01:03:01:01	3UTR-1	G	0.0012
Epromo-23	E*01:03:01:01 ^(new +1322A)	3UTR-1	G	0.0012
Epromo-04	E*01:03:01:01 ^(new +3468C)	3UTR-3	G	0.0012
Epromo-23	E*01:03:01:01 ^(new +971A, new +1322A)	3UTR-1	G	0.0012
Epromo-02	E*01:03:01:01 ^(new +990C)	3UTR-1	G	0.0012
Epromo-02	E*01:03:02:01	3UTR-10	A	0.0012

Epromo-31	E*01:03:02:01	3UTR-2	A	0.0012
Epromo-07	E*01:03:02:01	3UTR-1	G	0.0012
Epromo-36	E*01:03:02:01	3UTR-2	A	0.0012
Epromo-07	E*01:03:02:01	3UTR-11	G	0.0012
Epromo-01	E*01:03:02:01	3UTR-2	A	0.0012
Epromo-03	E*01:03:02:01	3UTR-1	G	0.0012
Epromo-02	E*01:03:04	3UTR-2	A	0.0012
Epromo-08	E*01:03:05 ^{compatible}	3UTR-13	G	0.0012
Nucleotide diversity				0.0006 +/- 0.0003
Gene diversity				0.8445 +/- 0.0101

^a Haplotypes are displayed at Table 3.

^b *HLA-E* coding alleles according to the IMGT/HLA Database and displayed at Table 4.

^c Haplotypes are displayed at Table 5.

^d Variable site at 3' downstream region.

The high number of different extended haplotypes, together with the fact that most of the coding alleles might be associated with more than one upstream or 3'UTR haplotypes, might be a reflection of the low linkage disequilibrium (LD) pattern observed along the entire gene (Figure 2). Besides this low LD, there are blocks of strong LD, the first ones encompassing the distal promoter variable sites, the second including only positions +424 and +756 at the coding segment and the third encompassing the variable site downstream position +3777 (Figure 2). However, apparently, the *HLA-E* locus might include some hotspots of recombination as discussed latter. Finally, considering nucleotide diversity, we may consider the 5' upstream region as the most variable *HLA-E* segment (Table 4), followed by the 3' UTR/exon 8 downstream (Table 5). The most conserved segment (in terms of nucleotide diversity) is the coding region, with nucleotide diversity of only 0.0003+/- 0.0002 (Table 3).



Figure 2. Linkage disequilibrium (LD) between pairs of single nucleotide polymorphisms (SNPs) at the extended *HLA-E* gene segment (from -2143 to +4776 positions). This image was generated by Haploview 4.2 using variable sites with minimum allele frequency (MAF) $\geq 1\%$. Areas in red indicate strong LD [$\log$ de odds ratio ($\text{LOD} \geq 2$, $D' = 1$)]; areas in light red indicate moderate LD ($\text{LOD} \geq 2$, $D' \leq 1$); areas in blue or almost white indicate weak LD ($\text{LOD} \leq 2$, $D' = 1$) and white areas indicated no LD ($\text{LOD} \leq 2$, $D' \leq 1$). D' values different from 1 are represented inside the squares as percentages. There are three blocks of strong LD groups: the first ones at distal 5' segment, one encompassing positions +424 and +756 at coding segment and the last one including the variable site +3777 at 3'UTR segment.

4.4 Discussion

The HLA system is considered as the most polymorphic segment of the human genome, and this polymorphism is related to its antigen presentation function. Although the HLA-E molecule do present peptides, and its gene is located between two of the most variable human genes (*HLA-A* and *HLA-C*) (1,2), here the *HLA-E* locus presented low genetic variability and limited encoded protein diversity. This low polymorphism was already been reported by previous studies in different populations (35,37,38,53), but the complete gene variability has never been assigned in a admixed population such as Brazilian, which reinforce its low protein diversity.

Here, approximately 7 kb of the *HLA-E* segment was sequenced considering 420 Brazilians from the São Paulo state, Southeast Brazil. We have found altogether 30 new variable sites in the coding region (noted as “C” at Table 1 and Table 2), in the segment that is covered by the official HLA database (the IMGT/HLA database) and at least 7 of these are not reported either by the aforementioned database, or in previous studies covering this region (35,37,38). Many of these new variable sites was detected as singletons and most of them were not included in the haplotype analysis (Table 2). However, this great number of singleton is not compatible with what was observed for three *HLA* genes, *HLA-A*, *HLA-G* and *HLA-F*, using the same population sample and the same analysis methodology (data not shown). In addition, it should be mentioned that two singletons that were not included in the haplotype analysis characterize non-synonymous mutations, thus encoding new HLA-E molecules. The 5' upstream region was firstly characterized here and its variability pattern discussed below. The *HLA-E* 3'UTR variability was also evaluated, with the characterization of new alleles and haplotypes.

4.4.1 The *HLA-E* variability at the 5'upstream segment

The segment upstream the translation start site (more than 2 kb upstream the first translated ATG, including the distal and proximal promoter, and the 5'UTR) was deeply characterized regarding its variability and haplotype structure. As far as we known, these is the first characterization of continuous segment surrounding the *HLA-E* locus.

The 5' upstream segment was quite variable when compared to the coding and 3'UTR regions. In fact, it presented the highest nucleotide diversity. Considering previous studies by Pyo and colleagues, who evaluated a similar extended *HLA-E* segment in 33 DNA samples (37), from our group in which African samples were evaluated regarding *HLA-E*

variability by next generation sequencing (38) and Brazilian samples evaluate regarding the proximal promoter variability by conventional sequencing (44), the majority of the variable sites detected for the *HLA-E* upstream segment (Table 1 and Table 2) is here described for the first time. However, as previously introduced, most of these new variable sites were singletons and were not included in the haplotype analysis, exception made to positions -1873, -1768, -1406 and -1339 (Table 1).

Although this *HLA-E* upstream segment presented the higher nucleotide diversity, just 4 variable sites within it were detected with MAF greater than 10%. Therefore, there are many low frequency variants in this segment, but only few variable sites are truly polymorphic (MAF greater than 1%). It may explain the large number of inferred haplotypes for this segment. Thereby, a single haplotype (Epromo-01) presents a frequency of 40% and, considering only the haplotype with a frequency greater than 1%, we may find only twelve variable sites within 2000 nucleotides.

In addition, most of the frequent variable sites were detected upstream position -1389, and only rare variants were detected in the proximal *HLA-E* promoter (exception made to position -104). It is documented that in the 5' upstream segment there are the core, proximal and distal promoter regions, which contain most of the binding sites for transcription factors (10,11). Interestingly, we observed that the number of variables sites increases with the distance from the translation start point, mainly in the distal promoter region. Thus, near the first translated ATG, located approximately 200bp from the beginning of the *HLA-E* gene, there is a highly-conserved sequence which matches with the core and proximal promoters where the transcription machinery binds. This conservation may be related to tight regulation mechanism characteristic of the *HLA-E* gene, and, therefore, this proximal promoter may be under strong selection against new mutations. The only exception here is position -104, but there are no functional studies evaluating the impact of the mutation at position -104 on the *HLA-E* expression pattern.

In fact, the low variability at the proximal promoter was previously observed (44) in a different Brazilian sample using Sanger sequencing. However, the aforementioned study could not evaluate the extended promoter by Sanger sequencing, since it was blocked upstream position -440 because of the present of a double-stranded branch formation and subsequently sequencing impairment. Using next generation sequencing, we could evaluate more than 1kb far from position -440. Besides the fact that the low variability on the core and proximal promoters was a feature, only one variable site was detected upstream position -440 up to position -900, in the segment associated with the double-stranded branch formation

discusses earlier. In addition, many low-quality variable sites were detected here and manually removed. Thus, it is possible that the variability in this segment is greater than the one reported here, but the low nucleotide complexity of this segment is impairing the NGS procedures.

Although the proximal and core *HLA-E* promoters are quite conserved, further investigation is needed to evaluate whether the variable sites at the distal promoter, and even the few variable sites in the proximal promoter, would influence the binding of transcriptional factors. Considering the ENCODE project data (54), there are no regulatory sites in the *HLA-E* segment between nucleotides -2143 and -378, but the region encompassing the variable site -104 (which is the only frequent one here) is quite active.

4.4.2 The *HLA-E* variability at the coding segment

The lowest molecular diversity was found for the *HLA-E* coding region. The different haplotypes detected encode mainly two protein molecules, E*01:01 and E*01:03. This phenomenon has been observed in different studies and population samples (34–38,55). This conservation may be strongly associated with the *HLA-E* molecule function in the immune system. *HLA-E* molecules bind and present a restricted peptide repertoire, mostly represented by the *HLA* class I leaders sequence-derived peptides (*HLA-A*, *HLA-B*, *HLA-C* and mainly *HLA-G*) (7,14,17,56,57), which in turn are also quite conserved. In addition, some viral and tumor peptides that are similar to the *HLA* class I signal sequence can also bind to the *HLA-E* groove (26–29,31–33).

Besides this protein conservation, here we demonstrated that the *HLA-E* locus do present a number of coding alleles that is far higher than our current knowledge when the IMGT/*HLA* database is considered (Table 4). Among these new haplotypes, five were previously detected in African samples (38), including (a) E*01:03:05^(new -104G/compatible), associated here with promoters carrying -104G (Table 3); (b) E*01:01:01:01^(new -26T), here named as E*01:01:01:06, associated with promoter haplotypes carrying -26T and 3'UTR haplotypes carrying +3468C (Table 4, 5 and 6); (c) E*01:03:01:01^(new +990C); (d) E*01:03:01:01^(new +1322A); and (e) E*01:03:02:01^(new +2269C). Besides these common alleles between the present study and the previously ones reported on African samples (38), 11 new alleles were detected, but all encoding the E*01:01 or E*01:03 molecules (Table 4) because of mutations in intron or untranslated segments.

Despite the number of different *HLA-E* alleles is growing, *HLA-E* is still the most conserved HLA gene in the class I region considering that E*01:01 and E*01:03 molecules types are encoded by about 99% of all *HLA-E* sequences.

4.4.3 The *HLA-E* variability at the 3'untranslated region

Besides the presence of non-synonymous mutations and variable sites at regulatory elements, variable sites at the 3'UTR segment may also influence gene function. In this sense, variability in the 3'UTR may modify the binding of microRNAs, and post transcriptionally influence gene expression (58). Here we report the *HLA-E* 3'UTR variability and haplotypes in a very admixed sample and the relationship among these haplotypes with promoter and coding ones. Some haplotypes, such as 3UTR-8 to 3UTR-14 (Table 5), were firstly described and named here, while haplotypes 3UTR-1 to 3UTR-5 were previously described in African samples (38). These last haplotypes (3UTR-1 to 3UTR-5) are the most frequent ones and do represent 98.57% of all sequences. The most frequent variable site within this segment occurs at position +3777, which coincides with the poly Adenine tail of a known *Alu* element (Figure 3).

Considering previously studies (35,38), some haplotypes and/or variable sites already described were found associated with specific coding alleles. These previous associations were in general also detected here. For instance, the haplotypes 3UTR-1 and 3UTR-2, which are the most frequent ones, were found associated with both E*01:01 and E*01:03 alleles previously (35,38) and in the present series (Table 6). Haplotype 3UTR-3 was found associated with E*01:01:01 by Felício and colleagues (35) and with E*01:01:01 and E*01:03:05 in African samples (38), but with E*01:03:01:01, E*01:01:01:01 and E*01:01:01:06 alleles in the present series.

Although twelve 3'UTR haplotypes were detected, 3UTR-1 and 3UTR-2 represents almost 90% of all sequences, which differ by only one variable site, and the remaining haplotypes are probably derived from one of them. Thus, similar to the *HLA-E* coding segment discussed previously, the *HLA-E* 3'UTR segment is also quite conserved. The 3'UTR lack of variability might indicate that this segment is a key feature to a proper *HLA-E* expression regulation, probably by interacting with specific microRNAs or other non-coding RNAs.

4.4.4 The *HLA-E* variability at extended *HLA-E* segment

Unlike what was observed for other *HLA* class I genes, *HLA-G*, *HLA-F* and *HLA-A*, in which the promoter, coding and 3'UTR haplotypes are in close association (46,59,60) and a high LD is detected along the entire gene, *HLA-E* does not present a strong LD encompassing its entire segment, and only few haplotype blocks are actually present (Figure 2). This phenomenon justifies the presence of many extended haplotypes (Table 6), and the fact that the same promoter haplotype or its derivatives are associated with different coding alleles. It is possible that this weak LD along the entire gene might be a consequence of the presence of several *Alu* elements in the *HLA-E* segment (Figure 3).

According UCSC Genome Browser (61) there are at least 5 *Alu* elements within the *HLA-E* segment considered here: three within distal and proximal promoter, mainly in the segment lacking variability possibly due to the low complexity of the segment and the presence of several repetitive elements, one in intron 5 inside the coding region and one in the 3'UTR segment (Figure 3). Two of these *Alu* elements are in the forward *HLA-E* strand and are located one in the promoter segment, between nucleotides 30,456,401 and 30,456,627 (positions -908 to -682) and the other in the 3'UTR segment, between 30,460,789 and 30,461,101 (positions +3480 to +3792), according human genome draft version hg19. Comparing the linkage disequilibrium and the location of these *Alu* elements at *HLA-E* (Figure 3), it coincides with breaks of segregation blocks, thus, it is possible that the presence of these *Alu* elements increases the *HLA-E* recombination rate (62), which would explain the presence of so many extended haplotypes.

4.4.5 Concluding remarks

Overall, this study shows the full variability of the *HLA-E* gene, including its extended promoter (more than 2 kb upstream the first translated ATG) and 100 nucleotides downstream, in a very admixed Brazilian population sample. The distal promoter was far the most variable segment, and the *HLA-E* protein conservation was reinforced by our data. *HLA-E* does not present the same pattern of LD observed for other non classical genes, mostly because of the presence of several *Alu* elements along the *HLA-E* segment. Does, the same promoter and 3'UTR haplotypes might be associated with different *HLA-E* coding haplotypes (Table 6).

Figure 3. *Alu* elements present at *HLA-E* segment. (A) Image obtained from UCSC Genome Browser comprising extended *HLA-E* segment with *Alu* elements represented as SINE family element (black and dark gray blocks) (61). There are three *Alu* elements at distal upstream segment: (1) between nucleotides -1285 and -974, at negative strand (2) -908 to -682, at positive strand and (3) -524 to -260, at negative strand; (4) one at intron 5, between nucleotides +2273 and +2568, at negative strand, and (5) one within the *HLA-E* 3'UTR, from +3480 to +3792, encompassing variable sites +3634, +3777 and +3778. The *Alus* elements that are present at negative strand are represented by dashed arrow, besides *Alus* elements present at positive strand, the same of *HLA-E* gene, are represented by black boxes at LD map (B) and full arrows.

4.5 References

1. 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.
2. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Genet*. 2009;54(1):15–39.
3. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res [Internet]*. janeiro de 2015 [citado 17 de maio de 2016];43(Database issue):D423–31. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25414341>
4. Sullivan LC, Clements CS, Rossjohn J, Brooks AG. The major histocompatibility complex class Ib molecule HLA-E at the interface between innate and adaptive immunity. Vol. 72, *Tissue Antigens*. 2008. p. 415–24.
5. Allan DSJ, Lepin EJM, Braud VM, O’Callaghan CA, McMichael AJ. Tetrameric complexes of HLA-E, HLA-F, and HLA-G. Vol. 268, *Journal of Immunological Methods*. 2002. p. 43–50.
6. Braud VM, Allan DS, McMichael AJ. Functions of nonclassical MHC and non-MHC-encoded class I molecules [Internet]. Vol. 11, *Current Opinion in Immunology*. 1999 [citado 15 de março de 2016]. p. 100–8. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/10047540>
7. Pietra G, Romagnani C, Moretta L, Mingari MC. HLA-E and HLA-E-bound peptides: recognition by subsets of NK and T cells. *Curr Pharm Des [Internet]*. 2009;15(28):3336–44. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/19860683>
8. Pietra G, Romagnani C, Manzini C, Moretta L, Mingari MC. The emerging role of HLA-E-restricted CD8+ T lymphocytes in the adaptive immune response to pathogens and tumors. *J Biomed Biotechnol [Internet]*. 2010 [citado 16 de março de 2016];2010:907092. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/20634877>
9. Iwaszko M, Bogunia-Kubik K. Clinical significance of the HLA-E and CD94/NKG2 interaction. *Arch Immunol Ther Exp (Warsz)*. 2011;59(5):353–67.
10. Gobin SJP, Van Den Elsen PJ. Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F and HLA-G. In: *Human Immunology*. 2000. p. 1102–7.
11. Howcroft TK, Singer DS. Expression of nonclassical MHC class Ib genes: comparison of regulatory elements. *Immunol Res*. 2003;27(1):1–30.
12. Wei X, Orr HT. Differential expression of HLA-E, HLA-F, and HLA-G transcripts in human tissue. *Hum Immunol*. 1990;29(2):131–42.
13. Kochan G, Escors D, Breckpot K, Guerrero-Setas D. Role of non-classical MHC class I molecules in cancer immunosuppression. *Oncoimmunology [Internet]*. 2013;2(11):e26491–8. Recuperado de: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3894240&tool=pmcentrez&rendertype=abstract>
14. Borrego F, Ulbrecht M, Weiss EH, Coligan JE, Brooks AG. Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J Exp Med [Internet]*. 2 de março de 1998 [citado 15 de março de 2016];187(5):813–8. Recuperado de:

- <http://www.ncbi.nlm.nih.gov/pubmed/9480992>
15. Carretero M, Palmieri G, Llano M, Tullio V, Santoni A, Geraghty DE, et al. Specific engagement of the CD94/NKG2-A killer inhibitory receptor by the HLA-E class Ib molecule induces SHP-1 phosphatase recruitment to tyrosine-phosphorylated NKG2-A: evidence for receptor function in heterologous transfectants. *Eur J Immunol* [Internet]. abril de 1998 [citado 16 de março de 2016];28(4):1280–91. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9565368>
 16. Lee N, Llano M, Carretero M, Ishitani A, Navarro F, López-Botet M, et al. HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A. *Proc Natl Acad Sci U S A* [Internet]. 28 de abril de 1998 [citado 16 de março de 2016];95(9):5199–204. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9560253>
 17. Llano M, Lee N, Navarro F, García P, Albar JP, Geraghty DE, et al. HLA-E-bound peptides influence recognition by inhibitory and triggering CD94/NKG2 receptors: preferential response to an HLA-G-derived nonamer. *Eur J Immunol* [Internet]. setembro de 1998 [citado 16 de março de 2016];28(9):2854–63. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9754572>
 18. Wada H, Matsumoto N, Maenaka K, Suzuki K, Yamamoto K. The inhibitory NK cell receptor CD94/NKG2A and the activating receptor CD94/NKG2C bind the top of HLA-E through mostly shared but partly distinct sets of HLA-E residues. *Eur J Immunol*. 2004;34(1):81–90.
 19. Gunturi A, Berg RE, Forman J. The role of CD94/NKG2 in innate and adaptive immunity. *Immunol Res* [Internet]. 2004 [citado 16 de março de 2016];30(1):29–34. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/15258309>
 20. Braud VM, Allan DS, O’Callaghan CA, Söderström K, D’Andrea A, Ogg GS, et al. HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* [Internet]. 19 de fevereiro de 1998 [citado 14 de outubro de 2016];391(6669):795–9. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9486650>
 21. García P, Llano M, de Heredia AB, Willberg CB, Caparrós E, Aparicio P, et al. Human T cell receptor-mediated recognition of HLA-E. *Eur J Immunol* [Internet]. abril de 2002 [citado 16 de março de 2016];32(4):936–44. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/11920559>
 22. Braud V, Jones EY, McMichael A. The human major histocompatibility complex class Ib molecule HLA-E binds signal sequence-derived peptides with primary anchor residues at positions 2 and 9. *Eur J Immunol* [Internet]. maio de 1997 [citado 13 de outubro de 2016];27(5):1164–9. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9174606>
 23. O’Callaghan CA, Tormo J, Willcox BE, Braud VM, Jakobsen BK, Stuart DI, et al. Structural features impose tight peptide binding specificity in the nonclassical MHC molecule HLA-E. *Mol Cell* [Internet]. março de 1998 [citado 16 de março de 2016];1(4):531–41. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9660937>
 24. Lee N, Goodlett DR, Ishitani A, Marquardt H, Geraghty DE. HLA-E surface expression depends on binding of TAP-dependent peptides derived from certain HLA class I signal sequences. *J Immunol* [Internet]. 15 de maio de 1998 [citado 13 de outubro de 2016];160(10):4951–60. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/9590243>
 25. Braud VM, Allan DS, McMichael AJ. Functions of nonclassical MHC and non-MHC-encoded class I

- molecules. *Curr Opin Immunol* [Internet]. fevereiro de 1999 [citado 14 de outubro de 2016];11(1):100–8. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/10047540>
26. Halenius A, Gerke C, Hengel H. Classical and non-classical MHC I molecule manipulation by human cytomegalovirus: so many targets—but how many arrows in the quiver? *Cell Mol Immunol* [Internet]. março de 2015 [citado 17 de março de 2016];12(2):139–53. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25418469>
 27. Wolpert F, Roth P, Lamszus K, Tabatabai G, Weller M, Eisele G. HLA-E contributes to an immune-inhibitory phenotype of glioblastoma stem-like cells. *J Neuroimmunol* [Internet]. Elsevier B.V.; 2012;250(1–2):27–34. Recuperado de: <http://dx.doi.org/10.1016/j.jneuroim.2012.05.010>
 28. Bossard C, Béziau S, Matysiak-Budnik T, Volteau C, Laboisie CL, Jotereau F, et al. HLA-E/ β 2 microglobulin overexpression in colorectal cancer is associated with recruitment of inhibitory immune cells and tumor progression. *Int J cancer* [Internet]. 15 de agosto de 2012 [citado 17 de março de 2016];131(4):855–63. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/21953582>
 29. Gong F, Song S, Lv G, Pan Y, Zhang D, Jiang H. Human leukocyte antigen E in human cytomegalovirus infection: friend or foe? *Acta Biochim Biophys Sin (Shanghai)* [Internet]. julho de 2012 [citado 17 de março de 2016];44(7):551–4. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/22576308>
 30. Zheng H, Lu R, Xie S, Wen X, Wang H, Gao X, et al. Human leukocyte antigen-E alleles and expression in patients with serous ovarian cancer. *Cancer Sci* [Internet]. maio de 2015 [citado 17 de março de 2016];106(5):522–8. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25711417>
 31. Morandi F, Pistoia V. Interactions between HLA-G and HLA-E in Physiological and Pathological Conditions. *Front Immunol* [Internet]. 2014;5(August):394. Recuperado de: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4141331&tool=pmcentrez&rendertype=abstract>
 32. Li F, Wei H, Wei H, Gao Y, Xu L, Yin W, et al. Blocking the natural killer cell inhibitory receptor NKG2A increases activity of human natural killer cells and clears hepatitis B virus infection in mice. *Gastroenterology* [Internet]. 2013;144(2):392–401. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/23103614>
 33. Jørgensen PB, Livbjerg AH, Hansen HJ, Petersen T, Höllsberg P. Epstein-Barr virus Peptide Presented by HLA-E is Predominantly Recognized by CD8bright Cells in multiple Sclerosis Patients. *PLoS One*. 2012;7(9).
 34. Carvalho dos Santos L, Tureck LV, Wowk PF, Mattar SB, Gelmini GF, Magalhães JCM, et al. HLA-E polymorphisms in an Afro-descendant Southern Brazilian population. *Hum Immunol* [Internet]. American Society for Histocompatibility and Immunogenetics; 2013;74(2):199–202. Recuperado de: <http://dx.doi.org/10.1016/j.humimm.2012.10.020>
 35. Felício LP, Porto IOP, Mendes-Junior CT, Veiga-Castelli LC, Santos KE, Vianello-Brondani RP, et al. Worldwide HLA-E nucleotide and haplotype variability reveals a conserved gene for coding and 3' untranslated regions. *Tissue Antigens*. 2014;83(2):82–93.
 36. Veiga-Castelli LC, Castelli EC, Mendes CT, Da Silva WA, Faucher MC, Beauchemin K, et al. Non-classical HLA-E gene variability in Brazilians: A nearly invariable locus surrounded by the most variable genes in the human genome. *Tissue Antigens*. 2012;79(1):15–24.

37. Pyo CW, 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*. 2006;58(4):241–51.
38. 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 16 de março de 2016];76(12):945–53. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/26187162>
39. Djuricic S, Hviid TVF. HLA class Ib molecules and immune cells in pregnancy and preeclampsia. *Front Immunol*. 2014;5(DEC):1–17.
40. Ishitani A, Sageshima N, Hatake K. The involvement of HLA-E and -F in pregnancy. Vol. 69, *Journal of Reproductive Immunology*. 2006. p. 101–13.
41. Moffett A, Hiby SE, Sharkey AM. The role of the maternal immune system in the regulation of human birthweight. *Philos Trans R Soc London B Biol Sci* [Internet]. 2015;370(1663). Recuperado de: <http://rspb.royalsocietypublishing.org/content/370/1663/20140071.abstract>
42. Guethlein LA, Norman PJ, Hilton HHG, Parham P. Co-evolution of MHC class I and variable NK cell receptors in placental mammals. *Immunol Rev* [Internet]. setembro de 2015 [citado 16 de março de 2016];267(1):259–82. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/26284483>
43. Meuleman T, Lashley LEO, Dekkers OM, van Lith JMM, Claas FHJ, Bloemenkamp KWM. HLA associations and HLA sharing in recurrent miscarriage: A systematic review and meta-analysis. *Hum Immunol* [Internet]. maio de 2015 [citado 16 de março de 2016];76(5):362–73. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25700963>
44. Veiga-Castelli LC, da Silveira Bulcão Bertuol JM, Castelli EC, Donadi EA. Low variability at the HLA-E promoter region in the Brazilian population. *Hum Immunol* [Internet]. American Society for Histocompatibility and Immunogenetics; 2015;77(2):172–5. Recuperado de: <http://dx.doi.org/10.1016/j.humimm.2015.11.014>
45. Seqtk. Toolkit for processing sequences in FASTA/Q formats [Internet]. Recuperado de: <https://github.com/lh3/seqtk>
46. Lima THA, Buttura RV, Donadi EA, Veiga-Castelli LC, Mendes-Junior CT, Castelli EC. HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. *Hum Immunol* [Internet]. outubro de 2016 [citado 22 de setembro de 2016];77(10):841–53. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/27448841>
47. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* [Internet]. 2013 [citado 17 de maio de 2016];43:11.10.1-33. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25431634>
48. 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]. setembro de 2010 [citado 17 de maio de 2016];20(9):1297–303. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/20644199>
49. 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]. maio de 2011 [citado 17 de maio de 2016];43(5):491–8. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/21478889>
50. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* [Internet]. abril de 2001 [citado 17 de março de 2016];68(4):978–89. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/11254454>
 51. 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]. maio de 2010 [citado 17 de março de 2016];10(3):564–7. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/21565059>
 52. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* [Internet]. 15 de janeiro de 2005 [citado 17 de março de 2016];21(2):263–5. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/15297300>
 53. 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*. 2014;20(2):129–41.
 54. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* [Internet]. *Nature Research*; 5 de setembro de 2012 [citado 10 de janeiro de 2017];489(7414):57–74. Recuperado de: <http://www.nature.com/doi/10.1038/nature11247>
 55. Grimsley C, Ober C. Population genetic studies of HLA-E: evidence for selection. *Hum Immunol* [Internet]. 1997;52(1):33–40. Recuperado de: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9021407
 56. Lauterbach N, Wieten L, Popeijus HE, Voorter CEM, Tilanus MGJ. HLA-E regulates NKG2C+ natural killer cell function through presentation of a restricted peptide repertoire. *Hum Immunol* [Internet]. agosto de 2015 [citado 17 de março de 2016];76(8):578–86. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/26382247>
 57. Lauterbach N, Wieten L, Popeijus HE, Vanderlocht J, van Zon PMH, Voorter CEM, et al. Peptide-induced HLA-E expression in human PBMCs is dependent on peptide sequence and the HLA-E genotype. *Tissue Antigens* [Internet]. abril de 2015 [citado 17 de março de 2016];85(4):242–51. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/25735891>
 58. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. Vol. 136, *Cell*. 2009. p. 215–33.
 59. Castelli EC, Veiga-Castelli LC, Yaghi L, Moreau P, Donadi EA. Transcriptional and posttranscriptional regulations of the HLA-G gene. Vol. 2014, *Journal of Immunology Research*. Hindawi Publishing Corporation; 2014.
 60. Castelli EC, Mendes-Junior CT, Veiga-Castelli LC, Roger M, Moreau P, Donadi EA. A comprehensive study of polymorphic sites along the HLA-G gene: Implication for gene regulation and evolution. *Mol Biol Evol*. 2011;28(11):3069–86.
 61. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. *Genome Res* [Internet]. 16 de maio de 2002 [citado 17 de março de 2016];12(6):996–

1006. Recuperado de: <http://www.genome.org/cgi/doi/10.1101/gr.229102>
62. Deininger P. Alu elements: know the SINEs. *Genome Biol* [Internet]. 2011 [citado 13 de outubro de 2016];12(12):236. Recuperado de: <http://www.ncbi.nlm.nih.gov/pubmed/22204421>

Conclusão

5 CONCLUSÃO

O presente trabalho demonstrou uma nova metodologia de análise de variabilidade por meio de sequenciamento de nova geração do segmento estendido do gene *HLA-E*. Considerando mais de 400 amostras, mesmo em uma população heterogênea como a brasileira, o gene *HLA-E* mostrou-se bastante conservado, principalmente em sua região codificadora. A maioria dos pontos de variação encontrados ou caracterizam mutações sinônimas ou estão presentes em regiões de íntrons, enquanto as mutações não sinônimas caracterizam alelos raros, com exceção de +756A que caracteriza os alelos E*01:03. O perfil fraco de desequilíbrio de ligação ao longo de todo segmento analisado é uma particularidade de *HLA-E*, em parte relacionado com a presença de elementos *Alu* na região promotora, íntron 5 e 3'NT que podem aumentar a taxa de recombinação dentro do gene. Este baixo desequilíbrio de ligação justifica o grande número de haplótipos estendidos, em que a mesma sequência regulatória está associada a diferentes alelos codificadores. O promotor proximal do *HLA-E* é bastante conservado, salvo pela presença de uma única mutação frequente. O promotor distal parece ser a região mais polimórfica do *HLA-E*, porém, os dados do projeto ENCODE apontam que este segmento não apresenta elementos regulatórios. Estudos futuros sobre a influência da variabilidade no promotor distal, proximal e 3'NT contribuirão para o entendimento sobre a dinâmica de ligação de fatores de transcrição e regulação por RNAs não codificadores, além de poder indicar sobre os aspectos evolutivos relacionados ao *HLA-E*.