

RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 20/02/2018.



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 TÉCN. 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 de 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, synomymous 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* present 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 positions).....	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).

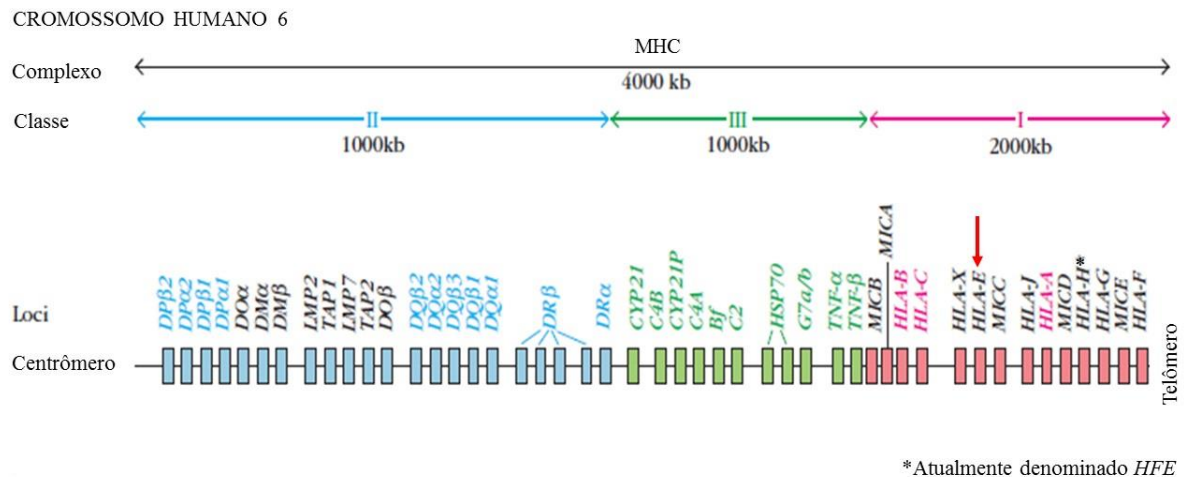


Figura 1. Esquema genômico detalhado da região do Complexo Principal de Histocompatibilidade (MHC) humano, localizado no braço curto do cromossomo 6 (10). Em azul está representada a região dos genes de classe II, na porção mais centromérica do MHC. Os genes escritos em azul representam os genes *HLA* 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 efetoras 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

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*.