Molecular Immunology 91 (2017) 173–184

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

Research paper

HLA-E regulatory and coding region variability and haplotypes in a Brazilian population sample

Jaqueline Ramalho[a], Luciana C. Veiga-Castelli[b], Eduardo A. Donadi[b], Celso T. Mendes-Junior[c], Erick C. Castelli[a,d,*]

[a] São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPEX), School of Medicine, Botucatu, State of São Paulo, Brazil
[b] School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, State of São Paulo, Brazil
[c] Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil
[d] São Paulo State University (UNESP), Department of Pathology, School of Medicine, Botucatu, State of São Paulo, Brazil

ARTICLE INFO

Keywords:
HLA-E, 3′ untranslated region
Promoter
Variability
Next generation sequencing
Polymorphism

ABSTRACT

The *HLA-E* gene is characterized by low but wide expression on different tissues. *HLA-E* is considered a conserved gene, being one of the least polymorphic class I *HLA* genes. The HLA-E molecule interacts with Natural Killer cell receptors and T lymphocytes receptors, and might activate or inhibit immune responses depending on the peptide associated with HLA-E and with which receptors HLA-E interacts to. Variable sites within the *HLA-E* regulatory and coding segments may influence the 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 gene structure, haplotype pattern and the complete *HLA-E* variability, including regulatory (promoter and 3′UTR) and coding segments (with introns), by using massively parallel sequencing. We investigated the variability of 420 samples from a very admixed population such as Brazilians by using this approach. Considering a segment of about 7 kb, 63 variable sites were detected, arranged into 75 extended haplotypes. We detected 37 different promoter sequences (but few frequent ones), 27 different coding sequences (15 representing new *HLA-E* alleles) and 12 haplotypes at the 3′UTR segment, two of them presenting a summed frequency of 90%. Despite the number of coding alleles, they encode mainly two different full-length molecules, known as E*01:01 and E*01:03, which corresponds to about 90% of all. In addition, differently from what has been previously observed for other non classical *HLA* genes, the relationship among the *HLA-E* promoter, coding and 3′UTR haplotypes is not straightforward because the same promoter and 3′UTR haplotypes were many times associated with different *HLA-E* coding haplotypes. This data reinforces the presence of only two main full-length HLA-E molecules encoded by the many *HLA-E* alleles detected in our population sample. In addition, this data does indicate that the distal *HLA-E* promoter is by far the most variable segment. Further analyses involving the binding of transcription factors and non-coding RNAs, as well as the HLA-E expression in different tissues, are necessary to evaluate whether these variable sites at regulatory segments (or even at the coding sequence) may influence the gene expression profile.

1. Introduction

The *HLA-E* gene is a non classical class I Human Leukocyte Antigen (HLA) locus 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, so far, it has been considered conserved in terms of nucleotide variability and encoded protein molecules compared to other *HLA* genes (Beck et al., 1999; Shiina et al., 2009). According to the International ImmunoGenetics (IPD-IMGT/HLA) database version 3.29.0, *HLA-E*

Abbreviations: ABC MRP7, ATP-binding cassette transporter multidrug resistance associated protein; bp, Base pairs; DNA, Deoxyribonucleic acid; EBV, Epstein-Barr virus; HCMV, Human cytomegalovirus; HCV, Hepatitis C virus; HIV, Human immunodeficiency virus; HLA, Human Leukocyte Antigen; HPV, Herpes virus; Hsp60, Heat shock protein 60; IMGT, ImMunoGeneTics information system; InflM, Influenza matrix protein; Kb, Kilobases (10^3 bases); MAF, Minor allele frequency; MHC, Major Histocompatibility Complex; mRNA, Messenger RNA; NGS, Next generation sequencing; NK, Natural Killer cell; PCR, Polymerase Chain Reaction; Prdx5, Peroxiredoxin 5; SINE, Short interspersed element; SNP, Single Nucleotide Polymorphism; TCR, T Cell Receptor; UTR, Untranslated region; VCF, Variant Call Format

* Corresponding author at: Departamento de Patologia, Faculdade de Medicina, UNESP, Botucatu, SP, CEP, 18618970, Brazil.
E-mail address: castelli@fmb.unesp.br (E.C. Castelli).

http://dx.doi.org/10.1016/j.molimm.2017.09.007
Received 17 May 2017; Received in revised form 8 September 2017; Accepted 13 September 2017

presents 26 different coding alleles (or coding sequences) encoding eight different full-length protein molecules (Robinson et al., 2015).

The HLA-E molecule may influence both the innate and adaptive immunity (Sullivan et al., 2008; Allan et al., 2002; Braud et al., 1999a; Pietra et al., 2009; Pietra et al., 2010; Iwaszko and Bogunia-Kubik, 2011). HLA-E is broadly expressed in a variety of tissues (Gobin and Van Den Elsen, 2000; Howcroft and Singer, 2003; Wei and Orr, 1990; Kochan et al., 2013), including the skin, lung, breast, urinary bladder, spleen, bone marrow, kidney, endometrium, placenta, tonsil and others, and it interacts mainly with Natural Killer (NK) cells through the CD94/NKG2A inhibitory receptors (Borrego et al., 1998; Carretero et al., 1998; Lee et al., 1998a; Llano et al., 1998; Wada et al., 2004; Gunturi et al., 2004; Braud et al., 1998). In addition, HLA-E may also interact with and activate T $CD8^+$ cells through the T cell receptor (TCR) (Sullivan et al., 2008; Pietra et al., 2009; Pietra et al., 2010; García et al., 2002). The HLA-E molecule binds to short peptides (usually 9 amino acids) derived from HLA class I signal sequences, thus, it presents self-antigens in a transporter associated with antigen processing (TAP) dependent way (Llano et al., 1998; Braud et al., 1997; O'Callaghan et al., 1998; Lee et al., 1998b). This peptide binding is related to an immune surveillance mechanism, indicating that the expression machinery and assembly of HLA molecules is operating properly, avoiding NK cytotoxicity (Borrego et al., 1998; Lee et al., 1998b; Braud et al., 1999b).

During pregnancy, for instance, there is an important interaction between HLA-E and the signal peptides derived from HLA-G, which is the main ligand for HLA-E in this scenario. The complex HLA-E/HLA-G signal peptide strongly interacts with NK cell receptors, modulating their activity, which is necessary for adequate placentation (Llano et al., 1998; Wada et al., 2004; Morandi and Pistoia, 2014). In pathological conditions, such as tumors and/or infections, in which the HLA class I expression profile may be altered, with no HLA class I signal peptides as ligands and/or with no HLA-E expressing at the cell surface, NK cells might act against the affected cells, influencing the outcome of this immune surveillance mechanism via HLA-E (Iwaszko and Bogunia-Kubik, 2011).

Also in pathological situations, besides the HLA class I signal peptides, non-self peptides may stabilize the HLA-E molecule and stimulate its expression on the cell surface. Therefore, 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) (Pietra et al., 2009; Pietra et al., 2010; Iwaszko and Bogunia-Kubik, 2011; Morandi and Pistoia, 2014; Halenius et al., 2015; Wolpert et al., 2012; Bossard et al., 2012; Gong et al., 2012; Zheng et al., 2015; Li et al., 2013). Among them we may find peptides derived from viral proteins such as gpUL40 (Human Cytomegalovirus, HCMV), BZLF-1 (Epstein-Barr virus, EBV), InflM (influenza matrix protein), core protein from Hepatitis C virus (HCV) and protein gag (Human Immunodeficiency virus, HIV). In addition, bacterial peptides from *Salmonella enterica* and *Mycobacterium tuberulosis* 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 Vβ2 TCR Vβ1, may also bind to the HLA-E molecule (Sullivan et al., 2008; Pietra et al., 2009; Pietra et al., 2010; Iwaszko and Bogunia-Kubik, 2011).

Moreover, mainly during chronic infections and cellular stress contexts, HLA-E may also interact and present different peptides to T $CD8^+$ lymphocytes restricted to HLA-E, activating them against altered cells (Jørgensen et al., 2012). Thus, HLA-E operates either modulating the immune response by interacting with CD94/NKG2 receptors of NK cells, or presenting antigens and triggering cytotoxic T cells by TCR (Sullivan et al., 2008; Pietra et al., 2009; Pietra et al., 2010; Iwaszko and Bogunia-Kubik, 2011; García et al., 2002), which is also important to consider for transplantation. Notwithstanding that, the HLA-E binding peptide repertoire is quite small compared to other HLA molecules.

This low peptide repertoire is consistent with the low variability observed so far for the *HLA-E* locus, mainly in the segment corresponding to the peptide binding site (Carvalho dos Santos et al., 2013; Felício et al., 2014; Veiga-Castelli et al., 2012a; Pyo et al., 2006; Castelli et al., 2015). Although eight different encoded protein molecules have already been described for HLA-E, only two are actually frequent worldwide. These frequent ones, known as E*01:01 and E*01:03, differ from 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 as an immune response modulator and also during pregnancy, since the fetal HLA-E interacts with the maternal NK cells (Djurisic and Hviid, 2014; Ishitani et al., 2006; Moffett et al., 2015; Guethlein et al., 2015; Meuleman et al., 2015).

Although *HLA-E* seems to be conserved in the coding region, these studies are usually conducted using molecular techniques for the evaluation of a subset of known variants, or small segments of the gene. In fact, many studies have reported the variability of exons 2 and 3, while only a few of them evaluated a continuous segment for the coding region or regulatory segments (Carvalho dos Santos et al., 2013; Felício et al., 2014; Veiga-Castelli et al., 2012a; Pyo et al., 2006; Castelli et al., 2015; Olieslagers et al., 2017). In addition, the regulatory segments, specially the promoter region, have not been properly explored. In fact, only three studies addressed variability in the regulatory segments, two of them concerning the 3′untranslated region (UTR) segment (Felício et al., 2014; Castelli et al., 2015) and the other one exploring the proximal promoter (Veiga-Castelli et al., 2015). Variable sites at regulatory segments may influence gene expression levels by different mechanisms such as differential binding of transcriptional factors, chromatin remodeling and the binding of microRNAs.

In this study, we present a methodology to evaluate the entire *HLA-E* segment, including the extended promoter, the complete coding sequence with introns and the complete 3′UTR, by using massive parallel sequencing, in order to characterize extended haplotypes for the *HLA-E* locus.

2. Material and methods

2.1. Samples

A total of 420 unrelated individuals (82.85% female, mean age of 31.3 years old) from the State of São Paulo, Brazil, accepted to participate in this study and peripheral blood samples were collected. Only 40.47% reported their ethnicity. Considering those, and according to self-reported ethnicity, individuals were classified as Euro-Brazilians (76.47%), Mulattoes (14.11%), Afro-Brazilians (4.11%), Asians (3.53%) and Amerindians (0.59%). The remaining self-declared as "unknown". These proportions are expected for a Brazilian sample from the São Paulo State. These samples may not necessarily represent the Brazilian population, however, considering the Brazilian 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.

2.2. HLA-E amplification, library preparation and sequencing

The *HLA-E* locus was amplified as a unique amplicon of approximately 7694 bp, encompassing its extended promoter segment (2775 nucleotides upstream the first translated ATG, excluding the forward primer segment), the complete coding sequence including all introns, the complete 3′UTR segment and 206 nucleotides downstream the *HLA-*