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
Available online 23 September 2017
0161-5890/

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-*

E locus, excluding the reverse primer segment, according to the annotations provided by the human genome draft (versions hg19 or hg38). Polymerase Chain Reaction (PCR) was carried out using primers HEPR·F_1: 5′- GCTTCGCAGTGAATGTGGCA – 3′ and HE_3UTR·R_1: 5′-GGACTCCCTGGGCTTTCTCACCG – 3′, in a final volume of 50 μL, containing 1.5U of DNA polymerase and 0.8X of buffer solution (Long PCR Enzyme Mix, Thermo Fisher Scientific Inc., Waltham, MA, USA), 1.5 mM of $MgCl_2$, 0.20 mM of each dNTP (Invitrogen − Carlsbad, CA, USA) and 0.30 μM of each primer. PCR cycling conditions were: 94 °C for 5 min, 35 cycles at 94 °C for 30 s, 62 °C for 45 s and 68 °C for 8 min, finalizing with 68 °C for 4 min and hold at 4 °C. Amplicons were evaluated by electrophoresis on 1% agarose gel stained with GelRed® (Biotium, Inc. Hayward, CA).

For the 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 × 250-bp).

2.3. Data processing and analysis

All sequences (reads) produced were firstly trimmed for adapters and primer sequences. Sequences were also processed to remove low quality 5′ and 3′ ends using *seqtk trimfq* 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) (Castelli et al., 2015; Lima et al., 2016) to generate the *HLA-E*-specific aligned files (BAM format) to the human reference genome version hg19 (www.castelli-lab.net/apps/hla-mapper). This strategy was introduced when the *HLA-E* variability was evaluated in two African population samples (Castelli et al., 2015) and then performed both in a recent study of *HLA-F* variability in Brazilian samples (Lima et al., 2016) and *HLA-G* in Brazilian and Cypriot samples (Castelli et al., 2017).

After mapping, genotypes were inferred using the Genome Analysis Toolkit (GATK, version 3.6) (Van der Auwera et al., 2013; McKenna et al., 2010; DePristo et al., 2011): 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 for nucleotide positions ranging from 30,454,700 to 30,462,100 of chromosome 6, 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* version 0.10.1 with default parameters (www.castelli-lab.net/apps/vcfx), as described elsewhere (Castelli et al., 2015; Lima et al., 2016; Castelli et al., 2017), a strategy that guarantees that only high quality genotypes are passed to a further imputation step. The vcfx-treated VCF file was manually inspected in two segments (one at the promoter and one at intron 5) that presented evidence of genotyping errors due to low sequence complexity and to 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 after inspection of BAM files. 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), and almost all variable sites detected within this region by GATK (exception made to a variable site at position −682, which did not present any unbalance) were removed from the analysis due to low quality genotypes. It is not clear whether these removed variable sites really exist or are artifacts because of the low complexity of this segment.

The methodology for haplotype inference has been previously described elsewhere (Castelli et al., 2017). In brief, 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 variable sites are sometimes quite distant from each other in a same sample, not all variable sites were straightforwardly phased. This is particularly evident for the low variable non classical *HLA* genes such as *HLA-E*. 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 (Stephens et al., 2001), which also imputed the 0.087% 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. The scripts to perform such analysis are available online (https://github.com/erickcastelli/phase-readbackedphasing).

Because of the *HLA-E* 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 into a final phased VCF file. The variable sites with just one 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 directly counted. Complete sequences for all individuals (one per chromosome) were generated by using *vcfx*. Nucleotide and haplotype diversities, as well as the adherence of the genotype frequencies to the Hardy-Weinberg expectations, were evaluated using Arlequin 3.5 (Excoffier and Lischer, 2010). The linkage disequilibrium, considering only variable sites with a frequency higher than 1%, was inferred using the software Haploview 4.2 (Barrett et al., 2005).

3. Results

Considering the *HLA-E* segment here evaluated (please refer to the methods section for segment description), we detected 63 variable sites (Table 1). Many of these (23.81%) variable sites occurred just once as singletons and were tagged with a “D” at Table 1. At least 22 variable sites reached polymorphic proportions in this population sample, presenting a minor allele frequency (MAF) higher than 1% (Table 1). Some of these sites were detected with a MAF higher than or close to 30%. Among them, we could 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 here detected and their genotype frequencies are under the Hardy-Weinberg equilibrium expectations. All the variable sites described at Table 1 were included in the haplotype analysis.

Besides the variable sites described at Table 1, 25 additional singletons were detected (Table 2), but not included in the haplotype analysis. They were not included because the ReadBackedPhasing/PHASE approach, described earlier, could not properly infer the correct phase of these variable sites because they occurred just once, impairing

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]	IPD-IMGT/HLA relative position[d]	Reference allele[e]	Alternative allele	Alternative allele frequency (2*n*=840)
30455166	rs2078675	E	5' upstream	-2143	T	C	0.2536
30455167	rs188844729	E	5' upstream	-2142	G	A	0.0036
30455186	rs139422860	E	5' upstream	-2123	G	A	0.0131
30455203	rs17875359	E	5' upstream	-2106	G	A	0.0690
30455240	rs28780108	E	5' upstream	-2069	G	C	0.0488
30455294	rs12207974	E	5' upstream	-2015	C	G	0.2750
30455321	rs17875360	E	5' upstream	-1988	T	C	0.1345
30455364	rs139555215	E	5' upstream	-1945	G	T	0.0071
30455429	rs147590779	E	5' upstream	-1880	A	G	0.0036
30455436	.	D	5' upstream	-1873	G	A	0.0012
30455532	rs3757335	E	5' upstream	-1777	G	T	0.0024
30455541	.	D	5' upstream	-1768	A	C	0.0012
30455846	rs188359497	D,E	5' upstream	-1463	C	T	0.0012
30455886	rs762324	E	5' upstream	-1423	G	A	0.0881
30455903	.	D	5' upstream	-1406	C	T	0.0012
30455920	rs1264459	E	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	E	5' upstream	-1229	C	G	0.0048
30456093	rs17875363	.	5' upstream	-1216	A	T	0.0024
30456142	rs116253207	E	5' upstream	-1167	A	G	0.0119
30456150	rs17875364	E	5' upstream	-1159	A	G	0.0750
30456151	rs17875365	.	5' upstream	-1158	T	C	0.0024
30456192	rs574018332	E	5' upstream	-1117	C	G	0.0071
30456230	rs146647219	E	5' upstream	-1079	G	T	0.0071
30456316	rs17875366	.	5' upstream	-993	G	A	0.0024
30456627	rs138267201	E	5' upstream	-682	AG	A	0.0083
30456931	rs186882745	E	5' upstream	-378	G	C	0.0024
30457196	rs141224659	C,E	5'UTR	-113	T	C	0.0060
30457204	rs9468784	C,E	5'UTR	-105	A	G	0.0036
30457205	rs61356961	C,E	5'UTR	-104	A	G	0.0238
30457283	rs76971248	E	5'UTR	-26	G	T	0.0167
30457732	rs1059510	A,E	Exon 2	+424	T	C	0.7167
30457766	rs150949676	B,E	Exon 2	+458	G	A	0.0024
30458064	rs1264457	B,E	Exon 3	+756	G	A	0.5738
30458279	rs145034129	A,C,D,E	Exon 3	+971	G	A	0.0012
30458298	rs370425404	C,D,E	Intron 3	+990	A	C	0.0012
30458322	rs41265828	E	Intron 3	+1014	T	A	0.0060
30458502	.	C,D	Intron 3	+1194	A	G	0.0012
30458586	rs182627071	C,E	Intron 3	+1278	C	T	0.0024
30458591	rs114425530	E	Intron 3	+1283	G	A	0.0071
30458630	rs116563630	C,E	Intron 3	+1322	G	A	0.0024
30458933	rs11548296	A,E	Exon 4	+1625	G	C	0.0238
30458999	rs17875370	A,D,E	Exon 4	+1691	G	A	0.0012
30459165	rs62621992	B,E	Exon 4	+1857	C	T	0.0060
30459296	rs200270359	E	Intron 4	+1988	G	A	0.0024
30459577	rs183165297	C,E	Intron 5	+2269	T	C	0.0024
30459935	rs41543014	E	Intron 5	+2627	G	A	0.0024
30460232	rs17875371	E	Intron 6	+2924	C	T	0.0393
30460245	rs147326588	C,D,E	Intron 6	+2937	G	A	0.0012
30460253	rs374002055	C,D,E	Intron 6	+2945	T	C	0.0012
30460755	rs17195369	C,E	Exon 8/3' UTR	+3447	C	T	0.0048
30460776	rs74295295	C,E	Exon 8/3' UTR	+3468	A	C	0.0333
30460798	rs765254483	C	Exon 8/3' UTR	+3490	C	A	0.0036
30460942	rs17195376	E	Exon 8/3' UTR	+3634	G	A	0.0369
30460999	.	D	Exon 8/3' UTR	+3691	A	C	0.0012
30461085	rs1059655	E	Exon 8/3' UTR	+3777	G	A	0.7429
30461086	rs115051198	E	Exon 8/3' UTR	+3778	A	G	0.0226
30461574	.	D	Exon 8/3' UTR	+4266	G	A	0.0012
30461605	rs9283	E	Exon 8/3' UTR	+4297	G	A	0.0357
30461728	rs566930407	D,E	Exon 8/3' UTR	+4420	C	T	0.0012
30461844	.	D	Exon 8/3' UTR	+4536	T	C	0.0012
30462084	rs1264456	E	3' downstream	+4776	A	G	0.7429

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

[a] Chromosome 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 IPD-IMGT/HLA database (release 3.29.0)
D) Singleton with a defined haplotype
E) Variable site reported at the 1000 Genomes database, Phase 3

[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] IPD-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 PHASE haplotype estimation. Also, there was a great distance between these singletons and the next heterozygous site, impairing the ReadBackedPhasing approach. Considering these singletons (Table 2) and the segment between nucleotides −300 and +3522, which is the *HLA-E* region tracked by the IPD-IMGT/HLA database, we noticed 16 new variable sites not described at the aforementioned database, including two non-synonymous mutations at positions +1644 (exon 4) and +2009 (exon 5) (tagged with a "B" at Table 2), 11 at intron segments and one at the 3′UTR. In addition, a known uncommon non-synonymous mutation at exon 4 was also detected at position +1822, which characterizes the known allele E*01:09 (Table 2).

The *HLA-E* 5′ upstream region (Fig. 1) presented 32 variable sites in approximately 2.5 kb evaluated (Table 3), 12 of which presenting frequencies higher than 1% in this Brazilian population sample (Table 1). These 32 variable sites were arranged into 37 haplotypes (Table 3). The *HLA-E* proximal promoter and the 5′UTR segment are quite conserved, since only four low-frequency variants were detected between nucleotides −300 and −1. Since this region is not entirely considered by the IPD-IMGT/HLA database and since there is no nomenclature pattern defined for it, the haplotypes here detected were organized in a descendent frequency order and named as E-Promo-01 to −37 (Table 3). The two most frequent haplotypes were E-Promo-01 (40.48%) and E-Promo-02 (14.05%) and they differ at only one position (Table 3).

In order to name the coding haplotypes according to the IPD-IMGT/HLA guidelines and also to consider the current IMGT/HLA release (3.29.0), the variability between nucleotides −300 and +3522 was converted in complete sequences. We detected 27 haplotypes, which were named following the haplotypes available at the aforementioned database. Most of the sequences were identical to the ones already described, including 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*01:01:02, 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. The summed frequency of these well-documented alleles was 95.48%. The remaining sequences configure new *HLA-E* coding alleles. In these cases, the closest known *HLA-E* allele was indicated, followed by the mutations eventually observed. These 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). In addition, in the coding segment, we detected 26 variable sites (Table 1), 4 of which 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 IPD-IMGT/HLA database version 3.29.0 (tagged with a "C" at Tables 1 and 2), and they occur mostly in introns or untranslated sequences.

Although 27 haplotypes have been detected, they encode only four different HLA-E molecules (Table 4), with molecule E*01:01 occurring at a frequency of 57.38% and E*01:03 at a frequency of 41.79%. However, it should be noted that, for this population sample, *HLA-E* presents additional encoded molecules beyond the ones described at Table 4. These molecules would have been generated if the singletons described at Table 2 had been included in the haplotype analysis.

The *HLA-E* 3′UTR is formed by a small segment at exon 7 and the entire exon 8 (Fig. 1). No variable sites were detected at exon 7 after the

Table 2
List of variable sites with a single occurrence (singletons) and 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]	IPD-IMGT/HLA relative position[d]	Reference allele[e]	Alternative allele
30455328	rs555000642	D,E	5' upstream	-1981	G	A
30455405	.	D,E	5' upstream	-1904	A	G
30455628	.	.	5' upstream	-1681	T	C
30456121	rs192094585	E	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, E	Exon 4	+1644	G	A
30459130	rs141487266	B (E*01:09), E	Exon 4	+1822	A	C
30459232	rs199731038	C, E	Intron 4	+1924	C	T
30459317	rs148162840	B (Pro→His), C, E	Exon 5	+2009	C	A
30459384	rs146142563	A, C	Exon 5	+2076	C	T
30459513	.	C	Intron 5	+2205	G	A
30459540	rs193246407	C, E	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), E	Exon 8/3'UTR	+4326	G	A

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

[a] Chromosome 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 IPD-IMGT/HLA database (release 3.29.0)
D) Variable site detected certainly at once but with missing alleles
E) Variable site reported at the 1000 Genomes database, Phase 3
Others: The non-synonymous mutation at exon (B) are accompanied with notes about the associated allele at IPD-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] IPD-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).

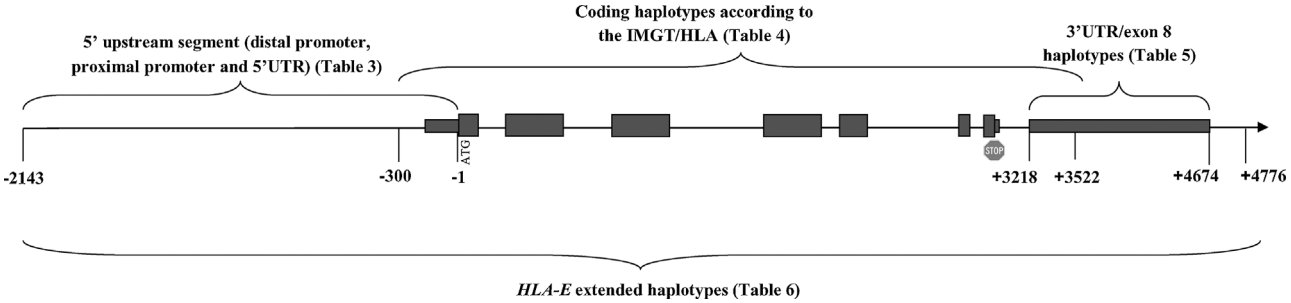

Fig. 1. *HLA-E* segments for haplotype analysis. The 5′ upstream segment (distal promoter, proximal promoter and 5′UTR) variability and haplotypes are represented at Table 3, comprising nucleotides from −2143 to −1 (considering the Adenine of the first ATG translated as nucleotide +1). The coding segment is covered by the IPD-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 the promoter, coding and 3′UTR segments and also the 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 IPD-IMGT/HLA[b,c]																														Frequency (2*n*=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	
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-21	T	G	G	G	G	C	T	G	G	G	G	A	C	G	C	G	T	C	C	A	A	A	T	C	G	G	AG	G	T	A	A	G	0.0036
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-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
E-Promo-36	C	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
E-Promo-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
																															Haplotype diversity	0.7925 +/- 0.0115

[a] Haplotypes were named considering the frequency in descending order.
[b] The position was calculated following the pattern used by the IPD-IMGT/HLA database, i.e, nucleotide +1 is the Adenine of the first translated ATG.
[c] The alternative alleles are represented with a grey background.

stop codon. Twelve haplotypes were detected at exon 8 (Table 5), and 5 of them had previously been described in African samples (3UTR-1, 3UTR-2, 3UTR-3, 3UTR-4 and 3UTR-5)(Castelli et al., 2015). Thus, the new haplotypes were arranged in descendent frequency order and named as 3UTR-8 to 3UTR-14 (Table 5). The most frequent haplotypes, known as 3UTR-1 and 3UTR-2 present a summed frequency of 89.41% and differ by a single variable site, at +3777 (Table 5).

The haplotypes detected for each segment (promoter, coding and 3′UTR) were then combined as extended haplotypes (Fig. 1, Table 6). Table 6 also includes the association of each extended haplotype and a downstream variable site at position +4776. We detected 75 extended haplotypes in this Brazilian sample, 13 of which presenting a frequency higher than 1% (Table 6). Two sets of general associations can be noticed: (a) the association between E-Promo-01 or similar haplotypes (E-Promo-05, −06, −09, −12, −14, −16, −18, −19, −22, −25, −26, −33 and −37), the coding allele group E*01:01:01 and the 3UTR-1 haplotype; and (b) E-Promo-02 or similar haplotypes (E-Promo-15, −17, −21 and −24), the coding allele E*01:03 and 3UTR-2. However, these associations are not straightforward.

Linkage disequilibrium (LD) was also assessed along the entire *HLA-E* segment, and a low LD pattern was observed (Fig. 2B), with two segregation blocks at the promoter segment, one at the coding region

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

HLA-E allele (coding haplotype)[a,c]	Encoded molecule[b]	Frequency ($2n$ = 840)
E*01:01:01:01	E*01:01	0.4750
E*01:01:01:01(+1194G)	E*01:01	0.0012
E*01:01:01:01(+1278T)	E*01:01	0.0024
E*01:01:01:01(+2937A, +2945C)	E*01:01	0.0012
E*01:01:01:01(+3447T)	E*01:01	0.0048
E*01:01:01:01(+3468C)	E*01:01	0.0155
E*01:01:01:01(+3490A)	E*01:01	0.0036
E*01:01:02 compatible	E*01:01	0.0012
E*01:01:01:01(−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(+1322A)	E*01:03	0.0012
E*01:03:01:01(+3468C)	E*01:03	0.0012
E*01:03:01:01(+971A, +1322A)	E*01:03	0.0012
E*01:03:01:01(+990C)	E*01:03	0.0012
E*01:03:01:01(−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(+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
Haplotype diversity		0.6878 +/− 0.0121

[a] Haplotype names were given according to the closest official *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 "compatible" indicates that the sequence found in Brazil is compatible with the sequence defined by the IPD-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.

[c] The variable sites at positions −113, −105, +971, +990, +1194, +1278, +1322, +2269, +2937, +2945, +3447, 3468 and +3490 are not currently described at the IPD-IMGT/HLA database, version 3.29.0.

formed by only two variable sites (positions +424 and +756) and another at the 3′UTR segment. Nucleotide diversity was also assessed allowing us to conclude that the 5′ upstream region is the most variable *HLA-E* segment (Table 4), while the coding region is the most conserved one (Table 3).

4. Discussion

The HLA system is considered the most polymorphic segment of the human genome, and this polymorphism is functionally related to the antigen presentation function of many HLA genes. Although the HLA-E molecule does present peptides, and its encoded gene is located between two of the most variable human genes (*HLA-A* and *HLA-C*) (Beck et al., 1999; Shiina et al., 2009), the *HLA-E* locus presented low genetic variability and limited encoded protein diversity in worldwide populations, including Brazil. This low polymorphism have already been reported in different populations (Felício et al., 2014; Pyo et al., 2006; Castelli et al., 2015; Olieslagers et al., 2017; Pratheek et al., 2014), but the complete gene variability (including the 5′ upstream promoter) has never been assessed in an admixed population such as Brazilians, reinforcing its low protein diversity.

Here we sequenced approximately 7 kb encompassing the *HLA-E* locus and surrounding sequences, in 420 Brazilians from the São Paulo state, Southeast Brazil. We have found 30 new variable sites along the coding segment (tagged with a letter "C" at Tables 1 and 2) tracked by the official HLA database (the IPD-IMGT/HLA database). Many of these new variable sites were detected as singletons (occurred only in one individual in a heterozygous state) and most of them were not included in the haplotype analysis for reasons previously described in the results section (Table 2). However, this pronounced number of singletons disagrees with what has previously been observed for three other HLA genes, *HLA-A* (unpublished data), *HLA-G* (Castelli et al., 2017) and *HLA-F* (Lima et al., 2016), in this same population sample.

One may consider that this pronounced number of singletons is related to technical and/or genotype errors. However, reads were mapped against the reference genome by using the hla-mapper software, which takes into account known *HLA-E* sequences to optimize read mapping (Lima et al., 2016; Castelli et al., 2017). All these new variable sites were manually inspected and the proportion of reads indicating each allele was indeed around 50%. Sequencing was performed targeting a minimum coverage of at least 500. Therefore, in view of this scenario, it is rational to downplay mapping bias and genotyping errors, and also to consider these singletons as true *HLA-E* variable sites. In addition, most of these singletons do characterize synonymous mutations in exons or variable sites in introns, which is in

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

Haplotypes[a]	Relative *HLA-E* positions according to the IPD-IMGT/HLA[b,c]											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
										Nucletide diversity		0.0005 +/- 0.0004
										Haplotype diversity		0.5250 +/- 0.0155

[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 IPD-IMGT/HLA database, i.e, nucleotide +1 is the Adenine of the first translated ATG.

[c] The alternative alleles are represented with a grey background.

Table 6
HLA-E extended haplotypes detected in 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)
E-Promo-01	E*01:01:01:01	3UTR-1	G	0.3500
E-Promo-02	E*01:03:02:01	3UTR-2	A	0.1179
E-Promo-03	E*01:03:02:01	3UTR-2	A	0.0917
E-Promo-04	E*01:03:01:01	3UTR-1	G	0.0726
E-Promo-05	E*01:01:01:01	3UTR-1	G	0.0417
E-Promo-06	E*01:01:01:01	3UTR-1	G	0.0381
E-Promo-07	E*01:03:02:01	3UTR-4	G	0.0357
E-Promo-01	E*01:01:01:03	3UTR-2	A	0.0179
E-Promo-01	E*01:01:01:01(+3468C)	3UTR-3	G	0.0155
E-Promo-08	E*01:03:05 compatible	3UTR-1	G	0.0155
E-Promo-09	E*01:01:01:06	3UTR-3	G	0.0131
E-Promo-10	E*01:01:01:01	3UTR-5	G	0.0119
E-Promo-02	E*01:03:02:01	3UTR-1	G	0.0107
E-Promo-05	E*01:01:01:03	3UTR-1	G	0.0095
E-Promo-13	E*01:01:01:01	3UTR-1	G	0.0071
E-Promo-14	E*01:01:01:01	3UTR-1	G	0.0071
E-Promo-05	E*01:01:01:04	3UTR-1	G	0.0071
E-Promo-11	E*01:03:01:01	3UTR-5	G	0.0071
E-Promo-12	E*01:03:01:01	3UTR-1	G	0.0071
E-Promo-16	E*01:01:01:03	3UTR-1	G	0.0060
E-Promo-03	E*01:03:02:02	3UTR-2	A	0.0060
E-Promo-04	E*01:06	3UTR-1	G	0.0060
E-Promo-02	E*01:01:01:01	3UTR-1	G	0.0048
E-Promo-01	E*01:01:01:03	3UTR-1	G	0.0048
E-Promo-17	E*01:03:02:01	3UTR-2	A	0.0048
E-Promo-06	E*01:01:01:01	3UTR-2	A	0.0036
E-Promo-19	E*01:01:01:01(−105G)	3UTR-1	G	0.0036
E-Promo-01	E*01:01:01:01(+3490A)	3UTR-9	G	0.0036
E-Promo-22	E*01:01:01:06	3UTR-3	G	0.0036
E-Promo-15	E*01:03:01:01(−113C)	3UTR-1	G	0.0036
E-Promo-21	E*01:03:02:01	3UTR-2	A	0.0036
E-Promo-18	E*01:03:05 compatible	3UTR-1	G	0.0036
E-Promo-20	E*01:03:05 compatible	3UTR-1	G	0.0036
E-Promo-25	E*01:01:01:01	3UTR-1	G	0.0024
E-Promo-01	E*01:01:01:01(+1278T)	3UTR-1	G	0.0024
E-Promo-05	E*01:01:01:01(+1988A)	3UTR-1	G	0.0024
E-Promo-01	E*01:01:01:01(+3447T)	3UTR-8	G	0.0024
E-Promo-26	E*01:01:01:01(+3447T)	3UTR-8	G	0.0024
E-Promo-24	E*01:03:01:01	3UTR-1	G	0.0024
E-Promo-06	E*01:03:01:01	3UTR-1	G	0.0024
E-Promo-27	E*01:03:01:02	3UTR-5	G	0.0024
E-Promo-02	E*01:03:02:01(+2269C)	3UTR-2	A	0.0024
E-Promo-15	E*01:05 compatible	3UTR-1	G	0.0024
E-Promo-02	E*01:01:01:01	3UTR-2	A	0.0012
E-Promo-01	E*01:01:01:01	3UTR-14	G	0.0012
E-Promo-34	E*01:01:01:01	3UTR-1	G	0.0012
E-Promo-05	E*01:01:01:01	3UTR-2	A	0.0012
E-Promo-03	E*01:01:01:01	3UTR-1	G	0.0012
E-Promo-28	E*01:01:01:01	3UTR-1	G	0.0012
E-Promo-01	E*01:01:01:01	3UTR-12	G	0.0012
E-Promo-01	E*01:01:01:01(+1194G)	3UTR-1	G	0.0012
E-Promo-01	E*01:01:01:01(+2937A,new+2945C)	3UTR-1	G	0.0012
E-Promo-01	E*01:01:01:01(+424T)	3UTR-1	G	0.0012
E-Promo-06	E*01:01:01:03	3UTR-1	G	0.0012
E-Promo-29	E*01:03:01:01	3UTR-5	G	0.0012
E-Promo-33	E*01:03:01:01	3UTR-1	G	0.0012
E-Promo-01	E*01:03:01:01	3UTR-1	G	0.0012
E-Promo-30	E*01:03:01:01	3UTR-1	G	0.0012
E-Promo-35	E*01:03:01:01	3UTR-1	G	0.0012
E-Promo-04	E*01:03:01:01	3UTR-2	A	0.0012
E-Promo-37	E*01:03:01:01	3UTR-1	G	0.0012
E-Promo-32	E*01:03:01:01	3UTR-1	G	0.0012
E-Promo-23	E*01:03:01:01(+1322A)	3UTR-1	G	0.0012
E-Promo-04	E*01:03:01:01(+3468C)	3UTR-3	G	0.0012
E-Promo-23	E*01:03:01:01(+971A,new+1322A)	3UTR-1	G	0.0012
E-Promo-02	E*01:03:01:01(+990C)	3UTR-1	G	0.0012
E-Promo-02	E*01:03:02:01	3UTR-10	A	0.0012
E-Promo-31	E*01:03:02:01	3UTR-2	A	0.0012
E-Promo-07	E*01:03:02:01	3UTR-1	G	0.0012
E-Promo-36	E*01:03:02:01	3UTR-2	A	0.0012
E-Promo-07	E*01:03:02:01	3UTR-11	G	0.0012
E-Promo-01	E*01:03:02:01	3UTR-2	A	0.0012
E-Promo-03	E*01:03:02:01	3UTR-1	G	0.0012
E-Promo-02	E*01:03:04	3UTR-2	A	0.0012

(*continued on next page*)

Table 6 (*continued*)

5′ Upstream/5′UTR Haplotypes[a]	*HLA-E* coding allele[b]	3′UTR/exon8 haplotypes[c]	+4776[d]	Frequency (2*n* = 840)
E-Promo-08	E*01:03:05 [compatible]	3UTR-13 Nucleotide diversity Haplotype diversity	G	0.0012 0.0006 +/− 0.0003 0.8445 +/− 0.0101

[a] Haplotypes displayed at Table 3.
[b] *HLA-E* coding alleles according to the IPD-IMGT/HLA database and displayed at Table 4.
[c] Haplotypes displayed at Table 5.
[d] Variable site at 3′ downstream region.

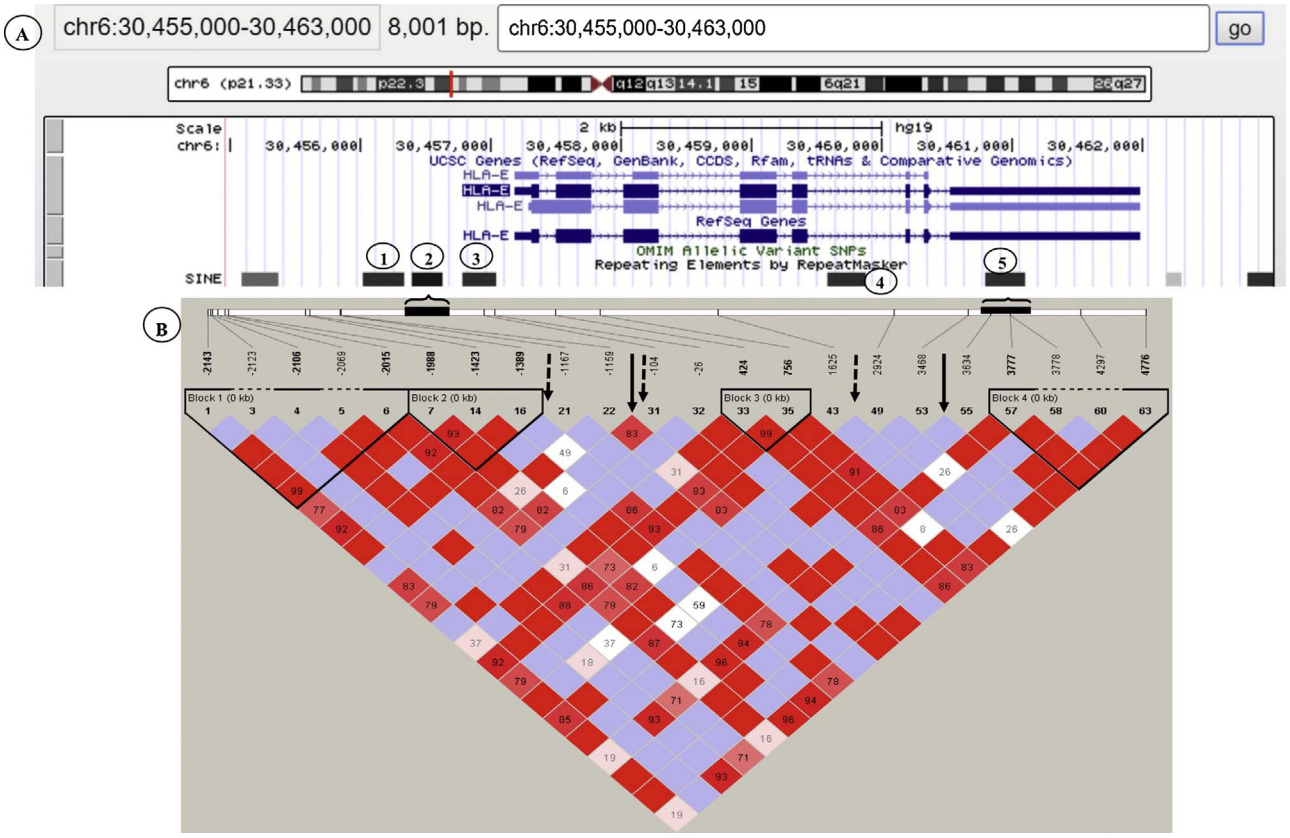

Fig. 2. *Alu* elements within the HLA-E segment and the linkage disequilibrium (LD) between pairs of single nucleotide polymorphisms (SNPs) in the extended HLA-E gene segment (from −2143 to +4776 positions).

Footnotes.

(A) Image obtained from UCSC Genome Browser comprising the extended HLA-E segment with Alu elements represented as SINE family elements (black and dark gray blocks). There are three Alu elements in the 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. Other Alu elements include one at intron 5, between nucleotides +2273 and +2568, at negative strand, and one within the HLA-E 3′UTR, from +3480 to +3792, encompassing variable sites +3634, +3777 and +3778. The Alu elements present at negative strand are represented by dashed arrow, besides Alu elements present at positive strand, the same of HLA-E gene, are represented by black boxes at the Linkage disequilibrium (LD) plot map and full arrows.

(B) Linkage disequilibrium (LD) between pairs of single nucleotide polymorphisms (SNPs) in 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) ≥ 1%. Areas in red or dark gray indicate strong LD [log de odds ratio (LOD ≥ 2, D' = 1)]; areas in light red or light gray indicate moderate LD (LOD ≥ 2, D' ≤ 1); areas in blue or almost white indicate weak LD (LOD ≤ 2, D' = 1) and white areas indicated no LD (LOD ≤ 2, D' ≤ 1). D′ values different from 1 are represented inside the squares as percentages.

agreement with the evidences of purifying selection already reported for *HLA-E* (Felício et al., 2014; Veiga-Castelli et al., 2012a; Castelli et al., 2015). In fact, only two singletons at Table 2 characterize non-synonymous mutations, thus encoding new HLA-E molecules. Finally, many of these singletons have already been reported at the 1000 genomes database, phase 3.

4.1. The HLA-E variability in the 5′upstream segment

The 5′ upstream region was firstly characterized here and its variability pattern is discussed below. This segment encompasses 2 kb upstream the first translated ATG and includes the distal and proximal promoter, and the 5′UTR. As far as we know, this is the first characterization of a continuous long segment upstream the *HLA-E* locus using a large sample size of an admixed population.

The 5′ upstream segment was quite variable when compared with the coding and 3′UTR regions. In fact, it presented the highest nucleotide diversity. Considering a previous study by Pyo and colleagues, who evaluated a similar extended *HLA-E* segment in 33 DNA samples (Pyo et al., 2006), and previous studies from our group, in which African samples were evaluated regarding the *HLA-E* coding and 3′UTR variability by next generation sequencing (Castelli et al., 2015) and Brazilian samples were evaluated regarding the proximal promoter variability by conventional sequencing (Veiga-Castelli et al., 2015), the majority of the haplotypes detected for the *HLA-E* upstream segment (Table 3) is here described for the first time.

Although the *HLA-E* 5′ upstream segment presented the highest nucleotide diversity, there are many low frequency variants within this segment, and few variable sites are truly polymorphic (MAF greater than 1%). The large number of low-frequency variants explains the large number of low-frequency haplotypes for this segment. Interestingly, a single haplotype (E-Promo-01) presents a frequency of about 40%. Moreover, only twelve variable sites are observed within the haplotypes that reached frequencies higher than 1%. The frequent variants are concentrated upstream position −1389, and mostly low frequency variants were detected in the proximal *HLA-E* promoter (exception made for position −104). In fact, the number of variable sites increases with the distance from the translation start point. In addition, most of the promoter variable sites detected in Brazil were also described by the 1000 Genomes database, phase 3 (Auton et al., 2015).

Therefore, there is a highly conserved sequence at a segment encompassing the 200 nucleotides upstream the first translated ATG, which matches with the core and proximal promoters, where the transcription machinery should bind. This conservation may be related to a tight regulation mechanism characteristic of the *HLA-E* gene, and, therefore, this proximal promoter may be under strong purifying selection. The only exception here is position −104, but there are no functional studies evaluating the impact of a mutation at position −104 on the *HLA-E* expression pattern.

This low variability at the proximal promoter has been previously observed (Veiga-Castelli et al., 2015) in a different Brazilian sample using Sanger sequencing. However, the aforementioned study could not evaluate the extended promoter since the sequencing reaction was impaired upstream position −440 because of the presence of a double-stranded branch formation. With the methodology here described, we evaluated more than 1 kb upstream this position, revealing only one variable site from position −440 to −900. However, many low quality and greatly unbalanced variable sites were detected within 3 *Alu* elements that occur between positions −1389 and −104 (Fig. 2). These variable sites were then manually inspected and further removed because it was unlikely that they represent true variable sites, but it is possible that the low nucleotide complexity of this segment and the double-stranded branch formation previously reported may also impair NGS procedures, underestimating variability in this segment.

Although the proximal and core *HLA-E* promoters are quite conserved, further investigation is needed to evaluate whether the variable sites in the distal promoter, and even the few variable sites in the proximal promoter, would influence the binding of transcriptional factors. In fact, considering the ENCODE project data (Dunham et al., 2012), while the region encompassing the variable site −104 (which is the only frequent one here) seems to be quite active, there are no regulatory site within the *HLA-E* segment between nucleotides −2143 and −378. Nevertheless, further studies are needed to better understand the impact of the promoter variability on the *HLA-E* expression profile and function.

4.2. The HLA-E variability in the coding segment

The lowest nucleotide diversity was found in the *HLA-E* coding region. Although 27 different haplotypes were detected, they encode mainly two protein molecules, known as E*01:01 and E*01:03; similar phenomenon has been observed in different studies, with different methods and population samples (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; Grimsley and Ober, 1997). This molecule conservation may be strongly associated with the HLA-E function in the immune system. HLA-E binds and presents a restricted peptide repertoire, mostly peptides derived from HLA class I leader sequences (Pietra et al., 2009; Borrego et al., 1998; Llano et al., 1998; Lauterbach et al., 2015a; Lauterbach et al., 2015b), which are, in turn, also quite conserved. In addition, some viral and tumor peptides that are similar to the HLA class I signal sequences can also bind to HLA-E (Morandi and Pistoia, 2014; Halenius et al., 2015; Wolpert et al., 2012; Bossard et al., 2012; Gong et al., 2012; Li et al., 2013; Jørgensen et al., 2012).

Besides this protein conservation, we demonstrated here that the *HLA-E* locus does present a number of coding alleles far higher from our current knowledge when the IPD-IMGT/HLA database is taken into account (Table 4). Among the sequences detected here but not described in the aforementioned database, nine have been previously described in at least one of the following studies addressing *HLA-E* variability: African samples using NGS (Castelli et al., 2015), different Brazilian samples using conventional sequencing (Felício et al., 2014; Veiga-Castelli et al., 2012b) and the 1000 Genomes database (1kgen), phase 3 (Auton et al., 2015). Among these sequences, we may found (a) E*01:01:01:01$^{(+1278T)}$, with two copies in the 1000 Genomes (1kgen) database; (b) E*01:01:01:01$^{(+3447T)}$, with a copy in Brazil and one in the 1kgen database; (c) E*01:01:01:01$^{(+3468C)}$, with 37 copies in the 1kgen database and a frequency of 3% in Brazil; (d) E*01:01:01:01$^{(-105G)}$, with 2 copies in the 1kgen database; (e) E*01:03:01:01$^{(+1322A)}$, with 31 copies in the 1kgen database and a frequency of 2% in Susu from Guinea-Conakry; (f) E*01:03:01:01$^{(+971A,+1322A)}$, detected 4 times in the 1kgen database; (g) E*01:03:01:01$^{(+990C)}$, with a frequency of 3.45% in Lobi from Burkina Faso; (h) E*01:03:01:01$^{(-113C),}$ detected 12 times in the 1kgen database; and (i) E*01:03:02:01$^{(+2269C)}$, with 13 copies in the 1kgen database and a frequency of 4% in Susu from Guinea-Conakry. Thus, although we have not cloned and sequenced them in a conventional manner, we may consider that at least 9 out of the 13 new alleles here detected are true *HLA-E* alleles. In addition, the 5 new alleles that have not been previously detected (the majority of which presenting singletons) were manually inspected. The variable sites that characterize these new alleles were detected as balanced genotypes (about 50% of reads for each allele) in segments presenting a minimum coverage of 100 reads. Besides these new *HLA-E* alleles, they encode the same E*01:01 or E*01:03 molecules (Table 4), since they harbor mutations in introns or untranslated segments.

Notwithstanding that, despite the larger number of different *HLA-E* alleles here reported, *HLA-E* is still the most conserved HLA gene in the class I region, since the E*01:01 and E*01:03 molecule types are encoded by about 99% of all *HLA-E* sequences.

4.3. The HLA-E variability in the 3′untranslated region

Variable sites in the 3′UTR segment may influence gene expression, mainly because different mRNA 3′UTR sequences may bind differently to microRNAs (Bartel, 2009), for instance. Here we report the *HLA-E* 3′UTR variability and haplotypes, and the relationship among these haplotypes with promoter and coding sequences, in a very admixed sample. Some haplotypes, such as 3UTR-8 to 3UTR-14 (Table 5), have not been described so far, while other, such as 3UTR-1 to 3UTR-5, have been previously described in African samples (Castelli et al., 2015). These shared haplotypes are the most frequent ones in our sample and do represent 98.57% of all sequences.

This study corroborates the previously reported weak association between specific *HLA-E* coding alleles and 3′UTR haplotypes (Felício et al., 2014; Castelli et al., 2015). For instance, haplotypes 3UTR-1 and 3UTR-2 (which are the most frequent ones) are associated with alleles encoding either E*01:01 or E*01:03 (Table 6). However, specific associations may also occur, such as 3UTR-4 and E*01:03:02:01, but this allele may present other 3′UTR haplotypes.

Although twelve 3′UTR haplotypes have been detected here, 3UTR-1 and 3UTR-2 represent almost 90% of all sequences, and they differ from each other by only one variable site. Thus, similarly to what has been observed for the *HLA-E* coding region, the 3′UTR is quite conserved. In fact, it presents the lowest haplotype diversity observed for *HLA-E*. The lack of variability in the 3′UTR segment might provide a

straightforward link between functional knowledge and population variation, providing evidence for the importance of this segment to a proper *HLA-E* expression profile. This probably occurs due to a mechanism that involves the binding of specific microRNAs that should not be influenced by variable sites, or the opposite (not binding microRNAs at all), since HLA-E is expressed in a variety of tissues expressing different microRNAs. It is interesting to note that the only frequent variable site within the 3′UTR segment occurs at position +3777, which coincides with the poly Adenine tail of a known *Alu* element (Fig. 2A) and, theoretically, should not bind any microRNA. Nevertheless, when comparing the *HLA-E* 3′UTR data with the 3′UTR data of another non classical HLA gene, *HLA-G,* from a similar sample (Castelli et al., 2017), we noticed that the nucleotide diversity in *HLA-E* is 60 times lower than *HLA-G*, mainly because *HLA-E* presents one of the largest HLA 3′UTR segment, but only one frequent variable site is detected there. It is possible that this segment is under purifying selection, which would be in agreement with the negative *Tajima's* D detected for this segment (−1.16444). Likewise, a negative and significant Tajima's D was observed for the *HLA-E* 3′UTR for another Brazilian sample evaluated by conventional sequencing (Felício et al., 2014). Nevertheless, functional studies are necessary to evaluate if and which microRNAs would bind to the HLA-E mRNA 3′UTR.

4.4. The HLA-E variability in the extended HLA-E segment

Unlike what has been observed for other non classical *HLA* class I genes such as *HLA-G* and *HLA-F* (Lima et al., 2016; Castelli et al., 2017; Castelli et al., 2014; Castelli et al., 2011), *HLA-E* does not present a strong LD along the entire gene (Fig. 2B). It is possible that this weak LD might be a consequence of the presence of several *Alu* elements in the *HLA-E* region (Fig. 2A). Because of that, and differently from what has been observed for other non-classical HLA genes, there is no clear association among the *HLA-E* promoter, coding and 3′UTR haplotypes.

According to the UCSC Genome Browser (Kent et al., 2002) there are at least five *Alu* elements within the *HLA-E* segment considered here: three within the distal and proximal promoter, coinciding with the segment in which we did not detected polymorphisms, one within the coding region at intron 5 and one in the 3′UTR segment (Fig. 2A). Two of these *Alu* elements are in the forward *HLA-E* strand: 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 to the human genome draft version hg19. Comparing the LD plot and the location of these *Alu* elements (Fig. 2), we may notice that they coincide with the breaks between the inferred segregation blocks (Fig. 2B). Therefore, it is possible that the presence of these *Alu* elements is increasing the *HLA-E* recombination rate (Deininger, 2011), which would explain the presence of so many extended haplotypes (Table 6) and the fact that the same promoter haplotypes are associated with different coding alleles.

4.5. Conclusion

Overall, this study introduces a new method to evaluate the entire *HLA-E* gene and its regulatory segments using massively parallel sequencing and freely available softwares, and present the full variability of the *HLA-E* gene in a highly 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, possibly because of the presence of several *Alu* elements along the *HLA-E* segment. Thus, the same promoter and 3′UTR haplotypes were sometimes associated with different *HLA-E* coding haplotypes. The *HLA-E* 3′UTR is quite conserved, with only one frequent variable site.

Conflict of interests

None.

Acknowledgements

This work was supported by grants obtained from the São Paulo Research Foundation (FAPESP/Brazil − grants # 2013/17084-2 and # 2014/18730-8), and from CNPq/Brazil (grants # 302590/2016-1, 304931/2014-4 and 309572/2014-2).

References

Allan, D.S.J., Lepin, E.J.M., Braud, V.M., O'Callaghan, C.A., McMichael, A.J., 2002. Tetrameric complexes of HLA-E, HLA-F, and HLA-G. J. Immunol. Methods 268, 43–50.
Auton, A., et al., 2015. A global reference for human genetic variation. Nature 526, 68–74.
Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265.
Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233.
Beck, S., Geraghty, D., Inoko, H., Rowen, I., 1999. Complete sequence and gene map of a human major histocompatibility complex. Nature 401, 921–923.
Borrego, F., Ulbrecht, M., Weiss, E.H., Coligan, J.E., Brooks, A.G., 1998. 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. 187, 813–818.
Bossard, C., et al., 2012. HLA-E/β2 microglobulin overexpression in colorectal cancer is associated with recruitment of inhibitory immune cells and tumor progression. Int. J. Cancer 131, 855–863.
Braud, V., Jones, E.Y., McMichael, A., 1997. 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. 27, 1164–1169.
Braud, V.M., et al., 1998. HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. Nature 391, 795–799.
Braud, V.M., Allan, D.S., McMichael, A.J., 1999a. Functions of nonclassical MHC and non-MHC-encoded class I molecules. Curr. Opin. Immunol. 11, 100–108.
Braud, V.M., Allan, D.S., McMichael, A.J., 1999b. Functions of nonclassical MHC and non-MHC-encoded class I molecules. Curr. Opin. Immunol. 11, 100–108.
Carretero, M., et al., 1998. 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. 28, 1280–1291.
Carvalho dos Santos, L., et al., 2013. HLA-E polymorphisms in an afro-descendant southern brazilian population. Hum. Immunol. 74, 199–202.
Castelli, E.C., et al., 2011. A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. Mol. Biol. Evol. 28, 3069–3086.
Castelli, E.C., Veiga-Castelli, L.C., Yaghi, L., Moreau, P., Donadi, E.A., 2014. Transcriptional and posttranscriptional regulations of the HLA-G gene. J. Immuno. Res. 2014.
Castelli, E.C., et al., 2015. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. Hum. Immunol. 76, 945–953.
Castelli, E.C., et al., 2017. HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographicaly distinct population samples of Brazil and Cyprus. Mol. Immunol. 83, 115–126.
DePristo, M.A., et al., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.
Deininger, P., 2011. Alu elements: know the SINEs. Genome Biol. 12, 236.
Djurisic, S., Hviid, T.V.F., 2014. HLA class Ib molecules and immune cells in pregnancy and preeclampsia. Front. Immunol. 5, 1–17.
Dunham, I., et al., 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.
Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567.
Felício, L.P., et al., 2014. Worldwide HLA-E nucleotide and haplotype variability reveals a conserved gene for coding and 3' untranslated regions. Tissue Antigens 83, 82–93.
García, P., et al., 2002. Human T cell receptor-mediated recognition of HLA-E. Eur. J. Immunol. 32, 936–944.
Gobin, S.J.P., Van Den Elsen, P.J., 2000. Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F and HLA-G. Hum. Immunol. 61, 1102–1107.
Gong, F., et al., 2012. Human leukocyte antigen E in human cytomegalovirus infection: friend or foe? Acta Biochim. Biophys. Sin. (Shanghai) 44, 551–554.
Grimsley, C., Ober, C., 1997. Population genetic studies of HLA-E: evidence for selection. Hum. Immunol. 52, 33–40.
Guethlein, L.A., Norman, P.J., Hilton, H.H.G., Parham, P., 2015. Co-evolution of MHC class I and variable NK cell receptors in placental mammals. Immunol. Rev. 267, 259–282.
Gunturi, A., Berg, R.E., Forman, J., 2004. The role of CD94/NKG2 in innate and adaptive immunity. Immunol. Res. 30, 29–34.

Halenius, A., Gerke, C., Hengel, H., 2015. Classical and non-classical MHC I molecule manipulation by human cytomegalovirus: so many targets—but how many arrows in the quiver? Cell. Mol. Immunol. 12, 139–153.
Howcroft, T.K., Singer, D.S., 2003. Expression of nonclassical MHC class Ib genes: comparison of regulatory elements. Immunol. Res. 27, 1–30.
Ishitani, A., Sageshima, N., Hatake, K., 2006. The involvement of HLA-E and – F in pregnancy. J. Reprod. Immunol. 69, 101–113.
Iwaszko, M., Bogunia-Kubik, K., 2011. Clinical significance of the HLA-E and CD94/NKG2 interaction. Arch. Immunol. Ther. Exp. (Warsz) 59, 353–367.
Jørgensen, P.B., Livbjerg, A.H., Hansen, H.J., Petersen, T., Höllsberg, P., 2012. Epstein-Barr virus peptide presented by HLA-E is predominantly recognized by CD8bright cells in multiple sclerosis patients. PLoS One 7.
Kent, W.J., et al., 2002. The human genome browser at UCSC. Genome Res. 12, 996–1006.
Kochan, G., Escors, D., Breckpot, K., Guerrero-Setas, D., 2013. Role of non-classical MHC class I molecules in cancer immunosuppression. Oncoimmunology 2, e26491–8.
Lauterbach, N., Wieten, L., Popeijus, H.E., Voorter, C.E.M., Tilanus, M.G.J., 2015a. HLA-E regulates NKG2C+ natural killer cell function through presentation of a restricted peptide repertoire. Hum. Immunol. 76, 578–586.
Lauterbach, N., et al., 2015b. Peptide-induced HLA-E expression in human PBMCs is dependent on peptide sequence and the HLA-E genotype. Tissue Antigens 85, 242–251.
Lee, N., et al., 1998a. HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A. Proc. Natl. Acad. Sci. U. S. A. 95, 5199–5204.
Lee, N., Goodlett, D.R., Ishitani, A., Marquardt, H., Geraghty, D.E., 1998b. HLA-E surface expression depends on binding of TAP-dependent peptides derived from certain HLA class I signal sequences. J. Immunol. 160, 4951–4960.
Li, F., et al., 2013. Blocking the natural killer cell inhibitory receptor NKG2A increases activity of human natural killer cells and clears hepatitis B virus infection in mice. Gastroenterology 144, 392–401.
Lima, T.H.A., et al., 2016. HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. Hum. Immunol. 77, 841–853.
Llano, M., et al., 1998. HLA-E-bound peptides influence recognition by inhibitory and triggering CD94/NKG2 receptors: preferential response to an HLA-G-derived nonamer. Eur. J. Immunol. 28, 2854–2863.
McKenna, A., et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
Meuleman, T., et al., 2015. HLA associations and HLA sharing in recurrent miscarriage: a systematic review and meta-analysis. Hum. Immunol. 76, 362–373.
Moffett, A., Hiby, S.E., Sharkey, A.M., 2015. The role of the maternal immune system in the regulation of human birthweight. Philos. Trans. R. Soc. London B Biol. Sci. 370.
Morandi, F., Pistoia, V., 2014. Interactions between HLA-G and HLA-E in physiological and pathological conditions. Front. Immunol. 5, 394.
O'Callaghan, C.A., et al., 1998. Structural features impose tight peptide binding specificity in the nonclassical MHC molecule HLA-E. Mol. Cell 1, 531–541.
Olieslagers, T.I., et al., 2017. New insights in HLA-E polymorphism by refined analysis of the full-length gene. HLA 89, 143–149.
Pietra, G., Romagnani, C., Moretta, L., Mingari, M.C., 2009. HLA-E and HLA-E-bound peptides: recognition by subsets of NK and T cells. Curr. Pharm. Des. 15, 3336–3344.
Pietra, G., Romagnani, C., Manzini, C., Moretta, L., Mingari, M.C., 2010. The emerging role of HLA-E-restricted CD8+ T lymphocytes in the adaptive immune response to pathogens and tumors. J. Biomed. Biotechnol. 2010, 907092.
Pratheek, B.M., et al., 2014. Mammalian non-classical major histocompatibility complex I and its receptors: importante contexts of gene, evolution, and immunity. Indian J Hum Genet. 20, 129–141.
Pyo, C.W., et al., 2006. HLA-E, HLA-F, and HLA-G polymorphism: genomic sequence defines haplotype structure and variation spanning the nonclassical class I genes. Immunogenetics 58, 241–251.
Robinson, J., et al., 2015. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 43, D423–31.
Shiina, T., Hosomichi, K., Inoko, H., Kulski, J.K., 2009. The HLA genomic loci map: expression, interaction, diversity and disease. J. Hum. Genet 54, 15–39.
Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989.
Sullivan, L.C., Clements, C.S., Rossjohn, J., Brooks, A.G., 2008. The major histocompatibility complex class Ib molecule HLA-E at the interface between innate and adaptive immunity. Tissue Antigens 72, 415–424.
Van der Auwera, G.A., et al., 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinf. 43 (11), 1–33 (10).
Veiga-Castelli, L.C., et al., 2012a. Non-classical HLA-E gene variability in Brazilians: a nearly invariable locus surrounded by the most variable genes in the human genome. Tissue Antigens 79, 15–24.
Veiga-Castelli, L.C., et al., 2012b. Non-classical HLA-E gene variability in Brazilians: a nearly invariable locus surrounded by the most variable genes in the human genome. Tissue Antigens 79, 15–24.
Veiga-Castelli, L.C., da Silveira, Bulcão, Bertuol, J.M., Castelli, E.C., Donadi, E.A., 2015. Low variability at the HLA-E promoter region in the Brazilian population. Hum. Immunol. 77, 172–175.
Wada, H., Matsumoto, N., Maenaka, K., Suzuki, K., Yamamoto, K., 2004. 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. 34, 81–90.
Wei, X., Orr, H.T., 1990. Differential expression of HLA-E, HLA-F, and HLA-G transcripts in human tissue. Hum. Immunol. 29, 131–142.
Wolpert, F., et al., 2012. HLA-E contributes to an immune-inhibitory phenotype of glioblastoma stem-like cells. J. Neuroimmunol. 250, 27–34.
Zheng, H., et al., 2015. Human leukocyte antigen-E alleles and expression in patients with serous ovarian cancer. Cancer Sci. 106, 522–528.