Human Immunology 80 (2019) 112–119

Contents lists available at ScienceDirect

Human Immunology

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

HLA-F displays highly divergent and frequent haplotype lineages associated with different mRNA expression levels

Renato V. Buttura[a], Jaqueline Ramalho[a], Thálitta H.A. Lima[a], Eduardo A. Donadi[b], Luciana C. Veiga-Castelli[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] Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil
[d] São Paulo State University (UNESP), Pathology Department, School of Medicine, Botucatu, State of São Paulo, Brazil

ARTICLE INFO

Keywords:
HLA-F
Variability
Polymorphisms
Haplotypes
Next Generation Sequencing (NGS)
Promoter
Brazil
Gene expression
RNA-Seq
Expression regulation

ABSTRACT

HLA-F is one of the most conserved *loci* among the HLA gene family. The exact function of *HLA-F* is still under investigation. HLA-F might present tolerogenic features, participate in the stabilization of HLA molecules in open conformation, and also participate in the recycling of HLA molecules. Here we evaluate the variability and haplotype structure of the *HLA-F* distal promoter segment (from −1893 to −943) and how this segment is correlated with the coding region. Variability at the promoter segment was surveyed in 196 Brazilian samples using second-generation sequencing. The *HLA-F* promoter region presents two major haplotype lineages. Most of the variable sites are in perfect linkage and associated with a single promoter haplotype, here named F*distal-C. This haplotype is associated with F*01:01:02 alleles, while alleles from the F*01:01:01 or F*01:03 groups present closely related promoter sequences. F*distal-C is quite frequent in Brazil and in worldwide populations, with frequencies ranging from 8.41% at the Iberian Population in Spain to 34.34% in Vietnam. F*distal-C is also present in Neanderthal and Denisovan samples. *In silico* analyses demonstrated that F*distal-C presents a different transcription factor binding profile compared with other *HLA-F* promoters. Moreover, individuals carrying this haplotype present higher *HLA-F* mRNA expression levels. Functional studies are required to define the exact mechanism underlying this higher *HLA-F* mRNA expression level associated with F*distal-C and F*01:01:02 alleles.

1. Introduction

HLA-F is a conserved non-classical HLA class I gene presenting a differentiated expression profile when compared with other *HLA* class I genes. HLA-F is expressed intracellularly in lymphocytes and monocytes, but it can be detected at the cell surface upon activation [1–3]. The tissue-specific expression has been reported on tonsil, spleen, thymus [1], trophoblasts [4,5], and fetal liver [2,6]. HLA-F expression was also reported for different cell lines, including EBV-transformed lymphoblastoid and monocytic cell lines [7], B, and HUT-78 cell lines [1], J82 transitional cell carcinoma, MOLT-3T cell leukemia, JEG choriocarcinoma, BeWo choriocarcinoma, and 3A-subE SV40 transformed [5]. Moreover, HLA-F expression was detected for gastric, non-small-cell lung and breast cancer [8–10], and also on esophageal squamous cell carcinoma [11].

Unlike other HLA class I molecules, HLA-F can be exported to the cell surface by a mechanism that is independent of loading with peptides in the endoplasmic reticulum but depends on the HLA-F cytoplasmic tail [12]. HLA-F also presents an intracellular function, associated with other HLA class I molecules at the “open conformer” mode. HLA-F may play a role in the stabilization of open conformers, or even participating in a recycling mechanism of these molecules, possibly being associated with the cross-presentation of exogenous antigens [13,14]. HLA class I molecules in an “open conformer” manner, including HLA-F, may interact with KIR receptors, thus modulating NK cell function [13,15,16]. Moreover, the HLA-F peptide-binding groove is different from those of other class I genes [17]. Because of a mutation (R62W) leading to an open-ended groove that can accommodate long peptides, HLA-F presents peptides of unconventional length [13,17].

Although its exact function is not well established, HLA-F may

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

https://doi.org/10.1016/j.humimm.2018.10.016
Received 13 April 2018; Received in revised form 10 October 2018; Accepted 26 October 2018
Available online 26 October 2018
0198-8859/

present tolerogenic and immunomodulatory properties as other non-classical HLA class I molecules since (a) HLA-F interacts with Natural Killer receptors (NKRs), such as ILT2, ILT4, KIR2DS4, KIR3DL2, KIR3DL1, and KIR3DS1 [1,13,15–18]; (b) *HLA-F* variants were associated with implantation success during pregnancy [19]; (c) HLA-F is expressed on trophoblasts [4,5], together with HLA-G, HLA-E and HLA-C [4]; (d) HLA-F molecule is quite conserved at the DNA and protein levels, as also observed for HLA-G and HLA-E [20–26]; (e) *HLA-F* is highly conserved among primates, including humans, bonobos, gorillas, orangutans, and chimpanzees, suggesting a critical function in the immune response [25,27–29].

Regarding its genetic diversity, there are 31 different *HLA-F* sequences (or alleles) encoding 6 different full-length proteins described so far at the International Immunogenetics database (IPD-IMGT/HLA, version 3.33.0). However, due to the lack of studies addressing *HLA-F* variability in worldwide populations, *HLA-F* might be even more polymorphic. So far, only one study has evaluated *HLA-F* variability by using massively parallel sequencing in a population-level approach [20], revealing the presence of many new *HLA-F* alleles. Nevertheless, most of these new alleles encode the same HLA-F molecules already reported, emphasizing that *HLA-F* is indeed the most conserved HLA class I gene. In addition, there is a lack of studies addressing the *HLA-F* promoter structure. Variability at the regulatory segments may directly influence the *HLA-F* expression quantitatively and qualitatively. Here we evaluated the variability and haplotype structure of the *HLA-F* distal promoter segment in the same Brazilian samples already characterized for the proximal promoter and the coding region [20], correlating the *HLA-F* haplotypes with mRNA expression levels.

2. Methods and samples

2.1. HLA-F distal promoter variability and haplotypes

We evaluated the *HLA-F* distal promoter variability in 204 individuals from the State of São Paulo, Brazil, associating each promoter haplotype with the coding alleles reported for the same samples in a previous study [20]. All participants gave written informed consent to participate in this study, and the UNESP School of Medicine Ethics Committee (#24157413.7.0000.5411) approved this study protocol. *HLA-F* amplification and sequencing library preparation was performed as described elsewhere [20].

Sequencing was performed using MiSeq Reagent Kit (V2, 500 cycles – Illumina, Inc.). Read mapping, genotype calling and haplotyping were performed as described elsewhere [20–22]. Briefly, we processed the paired-end FASTQ files with *hla-mapper* (www.castelli-lab.net/apps/hla-mapper) [30]. Then, we used the GATK HaplotypeCaller for variant calling and haplotypes were inferred using both GATK ReadBackedPhasing and the PHASE algorithm [31], as previously addressed [21,22]. The association between the distal promoter, proximal promoter and coding haplotypes was detected by using the PHASE algorithm. We used the Arlequin 3.5.1.2 software to calculate nucleotide diversity, gene diversity, and adherence of diplotype proportions to expectations under Hardy–Weinberg equilibrium [32]. Linkage disequilibrium (LD) was assessed as an LD plot using Haploview 4.2 [33]. We used HaploReg version 4.1 to infer the impact of known variants on the binding of transcription factors [34]. The evolutionary relationships of the *HLA-F* coding sequences and the *HLA-F* promoters were inferred using the Neighbor-Joining method, 500 replicates, using MEGA7 [35]. Branch lengths correspond to the evolutionary distances between sequences. The evolutionary distances were computed using the number of differences method and correspond to the number of base differences per sequence.

2.2. HLA-F expression levels among different haplotypes

To assess whether *HLA-F* variants are associated with the *HLA-F* mRNA expression levels, we evaluated RNA-Seq expression levels from 445 individuals' from the Geuvadis project [36] that have also been genotyped by the 1000Genomes Project Phase 3. This RNA-Seq data consists of paired-end reads of 75 bases, with 89 Utah Residents (CEPH) with Northern and Western European Ancestry (CEU sample), 92 Finnish from Finland (FIN sample), 86 British from England and Scotland (GBR sample), 91 Toscani from Italy (TSI sample), and 87 Yoruba from Ibadan, Nigeria (YRI sample). Because of the repetitive and polymorphic nature of HLA genes, this read length may bias read mapping and further downstream analyses. Thus, to evaluate *HLA-F* expression levels, we used a motif-searching approach as described below.

First, we computed all possible 20-base motifs that present the following characteristics: (a) it must be present in all known *HLA-F* sequences described at the IPD-IMGT/HLA database version 3.33; (b) it must be present in both *HLA-F* sequences from each sample available at the 1000Genomes database; (c) it cannot be found in any other HLA gene sequence already reported at the IPD-IMGT/HLA database; (d) it must not be found at the complete human genome reference sequence hg19 exception made for the *HLA-F* locus. Then, we selected three motifs for each *HLA-F* segment (exons and introns), keeping a distance of at least 75 bases from each other whenever possible. This set of 33 motifs allows an accurate evaluation of overall *HLA-F* expression of a given individual without taking into account genetic variability and the existence of eventual alternative transcripts. In addition, we also selected allele-specific motifs that would make it possible to evaluate allele-specific *HLA-F* expression levels in heterozygous individuals. For this additional purpose, since the motif must be specific for a given *HLA-F* allele, the first and second rules described above for motifs selection were set aside. The list of motifs is available in Table S1.

Second, the motif-searching algorithm calculated the number of fragments (or read pairs) presenting each of these motifs. At this point, we considered only read pairs in which one read presents a motif and the other presents at least 20 consecutive bases compatible with any known *HLA-F* sequence.

Third, we calculated the overall *HLA-F* mRNA expression level for each sample by adding the number of fragments detected for each of the 33 motifs, dividing the sum by the number of fragments at the RNA-Seq library, and multiplying this result by 10^6. Then, we compared the overall *HLA-F* expression levels among individuals carrying at least one copy of the haplotype F*Distal-C against individuals that do not present F*Distal-C, according to the 1000Genomes data. For the complementary analysis of allele-specific *HLA-F* expression in heterozygous samples presenting one copy of the F*Distal-C/F*01:01:02 haplotype, we counted the number of fragments associated with all motifs used to test for allele-specific expression.

For the analysis of overall *HLA-F* mRNA levels, we used the Mann-Whitney test for comparing two sample groups, while we used the Wilcoxon matched signed-rank test to evaluate allele-specific differences in expression levels of heterozygous individuals. These analyses were performed using GraphPad InStat 3.05 (GraphPad Software). For all instances, the significance level was set at 0.05. We also evaluated the association between polymorphisms and the *HLA-F* expression levels using linear regression. We converted the 1000Genomes VCF file into a PED/MAP format with plink 1.9. Then, we performed a linear regression indicating the *HLA-F* expression levels as the response variable.

3. Results

3.1. HLA-F promoter variability and extended haplotypes

We evaluated the *HLA-F* promoter segment between positions −1893 and −943, considering as +1 the Adenine of the first translated ATG. This comprises a promoter segment of 950 bp, henceforth considered as the *HLA-F* distal promoter. Within this segment, there were 13 variable sites (Table 1), ten of them presenting alternative

Table 1
Variable sites detected within the *HLA-F* distal promoter in a Brazilian population sample from the State of São Paulo, Brazil.

Position at chr6[a]	SNPid	Relative position IMGT/HLA[b]	Reference Allele[c]	Reference Allele Frequency ($2n$ = 408)	Alternative Allele	Alternative Allele Frequency ($2n$ = 408)
29689423	rs144001783	−1818	G	0.9975	A	0.0025
29689532	rs9258169	−1709	T	0.8284	G	0.1716
29689590	rs9258170	−1651	T	0.8284	G	0.1716
29689631	rs17875375	−1610	T	0.9510	C	0.0490
29689683	rs78739406	−1558	A	0.8284	AG	0.1716
29689717	rs9258171	−1524	G	0.8284	T	0.1716
29689973	rs9258172	−1268	C	0.8284	T	0.1716
29690002	rs9258173	−1239	G	0.8284	T	0.1716
29690026	rs9258174	−1215	T	0.8284	C	0.1716
29690037	.	−1204	G	0.9975	A	0.0025
29690056	rs17875376	−1185	C	0.8015	T	0.1985
29690125	rs9258175	−1116	G	0.8284	A	0.1716
29690228	rs9258176	−1013	C	0.8284	G	0.1716

[a] Position at chromosome 6 considering the human reference genome, version hg19.
[b] Relative position regarding the IPD-IMGT/HLA database, considering the Adenine of the first translated ATG as nucleotide +1.
[c] Reference allele at the human reference genome version hg19.

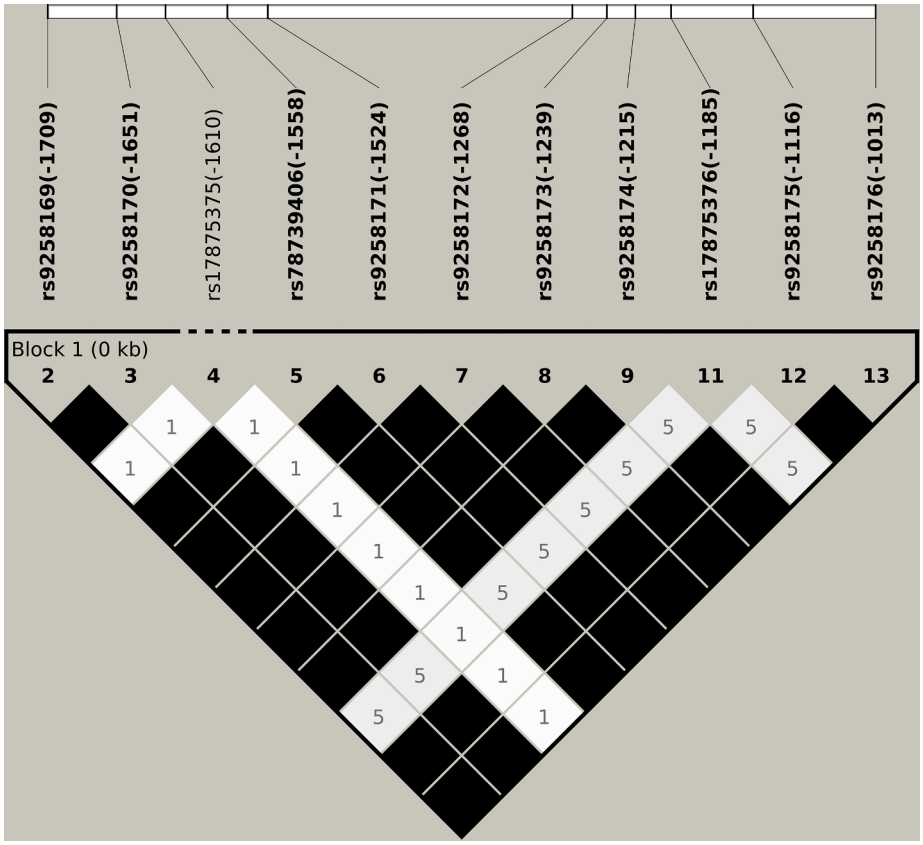

Fig. 1. Linkage disequilibrium (LD) map between pairs of variable sites in the *HLA-F* distal promoter region. This image was generated by the Haploview 4.2, using a minor allele frequency (MAF) ≥ 0.1%. The color of the spots represents the r^2 value between two variants, ranging from zero (white) to one (black). Values different from 1 are indicated as percentages.

alleles with frequencies above 15%, and all fitting Hardy-Weinberg equilibrium expectations ($P > 0.05$). Nine of these variable sites presented the same allele frequencies, indicating possible complete linkage among them, which was confirmed when evaluating the LD pattern across this segment (Fig. 1).

The distal promoter variable sites (Table 1) are arranged into six distinct haplotypes. These haplotypes were named following a decreased frequency order, as shown in Table 2. The reference haplotype, i.e., the haplotype found at the human reference genome hg19 or hg38, is the most frequent one (57.84%) and was named F*distal-A. The summed frequency of F*distal-A, -B, and -C is approximately 95%. One haplotype carries most of the alternative alleles, F*distal-C, while the remaining ones derives from F*distal-A by single point mutations. The summed frequency of F*distal-A and its derived haplotypes was 82.59%. The F*distal-C haplotype presents many alleles in absolute LD (Table 2 and Fig. 1). Therefore, there are basically two promoter profiles in this sample: haplotypes similar to F*distal-A and the highly divergent F*distal-C (Fig. S1). The frequencies of resultant diplotypes did fit Hardy-Weinberg equilibrium expectations ($P = 0.57198 \pm 0.00574$).

We evaluated the impact of the promoter variable sites (Table 1) on the binding of transcription factors (Table 3 and Table S2.). We observed a preferential binding of different subsets of transcription factors F*distal-A and its derived haplotypes when compared with F*distal-C. The positive Delta values indicate transcription factors binding preferentially to the reference allele, while negative values indicate the opposite (Table 3). Table 3 illustrates transcription factors with Delta values exceeding 5, but the complete data is available in Table S2. F*distal-A would better bind most of the transcription factors listed in

Table 2
List of *HLA-F* distal promoter haplotypes in a Brazilian population sample from the State of São Paulo, Brazil, comprehending the segment between nucleotides −1893 and −943.

Haplotypes[a]	Relative positions according to the IPD-IMGT/HLA[b]													Frequency
	-1818	-1709	-1651	-1610	-1558	-1524	-1268	-1239	-1215	-1204	-1185	-1116	-1013	(2*n*=408)
F*distal-A	G	T	T	T	A	G	C	G	T	G	C	G	C	0.5784
F*distal-B	G	T	T	T	A	G	C	G	T	G	T	G	C	0.1961
F*distal-C	G	G	G	T	AG	T	T	T	C	G	C	A	G	0.1716
F*distal-D	G	T	T	C	A	G	C	G	T	G	C	G	C	0.0490
F*distal-E	G	T	T	T	A	G	C	G	T	A	C	G	C	0.0025
F*distal-F	A	T	T	T	A	G	C	G	T	G	T	G	C	0.0025
											Nucleotide diversity			0.0031 +/- 0.0018
											Gene diversity			0.5966 +/- 0.0204

The alternative alleles are indicated in shades of gray.
[a]Haplotypes were listed in decreased frequency order.
[b]Positions were calculated considering as nucleotide +1 the adenine of the first translated ATG. Chromosome positions are listed in Table 1.

Table 3, but the opposite is observed for FEV, PU.1_known1, Foxf1, DMRT1, BDP1_disc1, and SP1_disc3, which present a stronger interaction with F*distal-C. With the exception of the SNP at position −1610, all the frequent variable sites at the *HLA-F* distal promoter were considered as eQTLs for *HLA-F* expression in many tissues according to the HaploReg database.

Table 4 presents the extended haplotypes considering the distal promoter, the proximal promoter, and the coding allele in 196 samples whose *HLA-F* coding region has been previously evaluated [20]. The *HLA-F* locus presented 39 extended haplotypes (Table 4), with F*distal-C exclusively associated with coding alleles of the F*01:01:02 group. Other distal promoter haplotypes are associated with both F*01:01:01 and F*01:03 allele groups. This association pattern between the distal promoter and coding alleles is a consequence of the elevated linkage disequilibrium along the entire *HLA-F* gene, which behaves essentially as a single segregation block (Fig. S2).

3.2. *HLA-F* mRNA expression levels according to genotypes and haplotypes

HLA-F coding sequences can be clustered into two main *HLA-F* lineages, F*01:01:01 and F*01:01:02/F*01:03 (Fig. S1). F*01:01:01 and F*01:03 are associated with closely related promoter and coding sequences, while F*01:01:02 presents a very divergent sequence considering the promoter segment and many introns. In addition, *HLA-F* presents only two frequent CDS variable sites. The first one is a synonymous mutation at position +63 (rs2076183). Allele +63A is associated with F*01:01:02 and +63G with all others. The second frequent variable site is a non-synonymous mutation at position +1771 (rs1736924). Allele +1771C is associated with F*01:03 while +1771T is associated with all others. Because of that, and since lineages F*01:01:01 and F*01:03 do present similar promoter sequences and closely related coding sequences (Table 4, Fig. S1), here we compared the *HLA-F* mRNA expression levels between F*distal-C/F*01:01:02 carriers against individuals that do not present this haplotype. Our results indicate that individuals carrying F*distal-C/F*01:01:02 (in the homozygous or heterozygous state) present a higher *HLA-F* mRNA expression level ($P < 0.0001$, Fig. 2 right panel). Moreover, we detected a larger number of fragments carrying the allele +63A, which is associated with F*01:01:02 and F*Distal-C compared with +63G ($P = < 0.0001$, Fig. 2 left panel), and also when considering two intronic alleles, +473C and +504G, both exclusively associated with the F*Distal-C/F*01:01:02 haplotype. The linear regression analysis indicated many variants associated with a higher *HLA-F* expression level, but all these variants are linked within the haplotype F*Distal-C/F*01:01:02. The strongest associations (lowest *P* values) were detected for variants that occur exclusively in the haplotype F*Distal-C/F*01:01:02 (unadjusted $P = 2.286 \times 10^{-5}$, adjusted $P = 0.0016$). Moreover, we observed that variants associated with F*Distal-C/F*01:01:02 are related to higher *HLA-F* expression levels when testing either African or European (CEU, TSI, GBR, and FIN) samples separately.

4. Discussion

Reports on *HLA-F* variability are quite scarce, mainly when the promoter segment is included. In fact, as far as we know, a single study evaluated the *HLA-F* promoter, comprehending approximately 2 kb upstream the first translated ATG, but it only explored a few cell lines [25]. This study described a pattern of variation similar to the one detected in this Brazilian population sample. Moreover, only one study evaluated the complete *HLA-F* coding segment by second-generation sequencing, but the distal promoter was not addressed [20]. The remaining studies regarding *HLA-F* variability evaluated only exons [26,37].

HLA-F is considered one of the least variable genes among the HLA gene family. Basically, two different full-length protein molecules are encoded by the sequences already reported [20,25,26,37]. These two molecules are known as F*01:01 (encoded by alleles from the F*01:01:01 and F*01:01:02 groups) and F*01:03 (encoded by alleles from the F*01:03 group). The two molecules differ by a single amino acid encoded at codon 251, a polar Threonine for F*01:01 and a hydrophobic Proline for F*01:03. In spite of this high conservation at the protein level, the *HLA-F* promoter seems to be very polymorphic when compared with other *HLA-F* segments.

Here we observed 13 variable sites detected in an *HLA-F* promoter segment of about 805 nucleotides, most of them with a minor allele frequency higher than 15%. These variable sites configure 6 different haplotypes. However, 13 variants are in perfect LD and associated with a unique and divergent haplotype, here named F*distal-C (Table 2). This same pattern has been previously detected in the aforementioned study that evaluated different cell lines [25]. Despite the presence of this divergent haplotype, *HLA-F* distal promoter can be considered a conserved segment, since its nucleotide diversity is the lowest observed among all *HLA-F* segments [20]. Gene diversity is also low, reflecting the existence of only two large promoter groups: the F*distal-A (and derivatives) and F*distal-C (Fig. S1).

F*distal-C is also present at the 1000Genomes data with a high frequency in all studied populations, from 8.41% for the Iberian Population in Spain to 34.34% in Vietnam. F*distal-C is also present in a Neanderthal sample in heterozygosis [38] and in a Denisovan sample in homozygosis [39]. This Denisovan sample seems to be homozygous for F*Distal-C/F*01:01:02, although a definitive conclusion is not possible

Table 3
In silico prediction of transcription factor binding motifs altered by the variable sites detected at the *HLA-F* distal promoter segment.

Chr 6 position (hg19)	Chr 6 position (hg38)	SNPid	Associated Haplotype	Reference allele	Alternative allele	Relative position to IMGT/HLA[a]	eQTL for *HLA-F*[b]	Regulatory motifs	Strand	HaploReg Reference[c]	HaploReg Alternative[c]	Delta (Ref − Alt)[d]
29689532	29721755	rs9258169	F*distal-C	T	G	−1709	Yes	Hoxa10	−	11.4	3.3	8.1
								Hoxd10	−	13.4	5.7	7.7
29689590	29721813	rs9258170	F*distal-C	T	G	−1651	Yes	Brachyury_2	−	8.6	−3.4	12
								Eomes	−	13.2	7	6.2
								SRF_disc1	+	5.6	−5.4	11
								SRF_known2	−	−1.8	−12.1	10.3
								TBX5_1	−	15	3.3	11.7
29689683	29721906	rs78739406	F*distal-C	A	AG	−1558	–	FEV	+	−14.3	−5	−9.3
								PU.1_disc1	+	−16.9	−22.3	5.4
								PU.1_known1	+	−13.7	−2.1	−11.6
29689717	29721940	rs9258171	F*distal-C	G	T	−1524	Yes	Pax-5_known1	+	13.3	1.9	11.4
								Zfp691	−	14.5	5.5	9
29689973	29722196	rs9258172	F*distal-C	C	T	−1268	Yes	Foxf1	+	0.9	12.9	−12
								Foxi1	+	2.3	14.3	−12
29690002	29722225	rs9258173	F*distal-C	G	T	−1239	Yes	DMRT1	+	−3.1	8.1	−11.2
29690026	29722249	rs9258174	F*distal-C	T	C	−1215	Yes	AP-1_disc2	+	13.1	1.1	12
								E2F_disc4	−	14.7	2.8	11.9
								Irf_disc1	−	13.3	1.3	12
								NF-Y_disc1	+	15.3	3.4	11.9
								NF-Y_known1	−	15.9	9.2	6.7
								NF-Y_known3	−	5.4	−6.6	12
								Pbx3_disc1	+	13.5	1.5	12
								RFX5_disc2	+	14.4	2.4	12
								SP1_disc1	−	13.2	5.8	7.4
								SP2_disc1	+	12	0	12
								TATA_disc6	+	13.7	1.8	11.9
29690056	29722279	rs17875376	F*distal-B	C	T	−1185	Yes	CEBPD	+	1.2	13.1	−11.9
								Dlx3	+	3.3	11.8	−8.5
								Pitx2	−	13.3	1.4	11.9
29690125	29722348	rs9258175	F*distal-C	G	A	−1116	Yes	SZF1-1	+	10.6	−1.4	12
29690228	29722451	rs9258176	F*distal-C	C	G	−1013	Yes	BDP1_disc1	+	−13.2	−1.3	−11.9
								SP1_disc3	+	−0.2	11.6	−11.8

[a] Position considering the Adenine of the first translated ATG as nucleotide +1.
[b] According to the HaploReg database.
[c] Position Weight Matrix ID (Kheradpour and Kellis, 2013).
[d] The difference observed between the values observed for the reference and alternative alleles. Only regulatory motifs presenting a difference (Delta) higher than 5 is presented. The complete table is available in Table S1. Positive values indicate a preferential binding to the reference allele, while negative values indicate opposite.

Table 4
HLA-F extended haplotypes in a Brazilian population sample from the State of São Paulo, Brazil, comprehending the segment between nucleotides -1893 and +4252, considering the Adenine of the first translated ATG as +1.

Distal Promoter[a]	5′ upstream[b]	*HLA F* Allele[b]	3′ downstream[b]	Frequency ($2n$ = 392)
F*distal-A	F*upstream-A	F*01:01:01:09	F*downstream-B	0.1735
F*distal-A	F*upstream-B	F*01:01:01:08	F*downstream-A	0.1684
F*distal-B	F*upstream-B	F*01:01:01:01	F*downstream-A	0.1403
F*distal-A	F*upstream-A	F*01:03:01:01	F*downstream-D	0.1020
F*distal-D	F*upstream-A	F*01:03:01:03-compatible ($[TG]_{12}$]	F*downstream-E	0.0459
F*distal-A	F*upstream-D	F*01:01:01:05 (1943G, $[TG]_{12}$)	F*downstream-A	0.0434
F*distal-A	F*upstream-A	F*01:01:01:11	F*downstream-B	0.0383
F*distal-C	F*upstream-E	F*01:01:02:01 (new-238T, 1230A, 1943G, new2486T, new3205G, $[TG]_{13}$)	F*downstream-C	0.0383
F*distal-C	F*upstream-C	F*01:01:02:08-compatible ($[TG]_{11}$)	F*downstream-C	0.0332
F*distal-B	F*upstream-B	F*01:01:01:01 ($[TG]_{12}$)	F*downstream-A	0.0332
F*distal-C	F*upstream-C	F*01:01:02:05 (($[TG]_{11}$)	F*downstream-C	0.0306
F*distal-C	F*upstream-F	F*01:01:02:03 (1943G, $[TG]_{11}$)	F*downstream-C	0.0179
F*distal-A	F*upstream-A	F*01:04 (1943G, 3189 T)	F*downstream-B	0.0128
F*distal-B	F*upstream-B	F*01:01:01:01 ($[TG]_{12}$)	F*downstream-F	0.0128
F*distal-C	F*upstream-E	F*01:01:02:08-compatible (new-238T, new2486T, new3205G, $[TG]_{13}$)	F*downstream-C	0.0128
F*distal-A	F*upstream-A	F*01:01:01:09 (new394G)	F*downstream-B	0.0077
F*distal-C	F*upstream-H	F*01:01:02:07-compatible ($[TG]_{13}$)	F*downstream-C	0.0077
F*distal-A	F*upstream-D	F*01:01:01:05 (1943G, $[TG]_{14}$)	F*downstream-A	0.0077
F*distal-C	F*upstream-C	F*01:01:02:08-compatible ($[TG]_{11}$)	F*downstream-G	0.0077
F*distal-D	F*upstream-A	F*01:03:01:03-compatible (new1378T, $[TG]_{11}$)	F*downstream-E	0.0051
F*distal-A	F*upstream-G	F*01:01:01:08 (−222G, 2698G, $[TG]_{11}$)	F*downstream-A	0.0051
F*distal-A	F*upstream-B	F*01:01:01:08 (new1779A)	F*downstream-A	0.0051
F*distal-A	F*upstream-G	F*01:01:01:08 (−222G, 2698G, $[TG]_{11}$)	F*downstream-H	0.0051
F*distal-C	F*upstream-C	F*01:01:02:08-compatible ($[TG]_{12}$)	F*downstream-C	0.0051
F*distal-B	F*upstream-B	F*01:01:01:01 (new1313G)	F*downstream-A	0.0051
F*distal-A	F*upstream-D	F*01:01:01:08 (−222G, 2698G, $[TG]_{11}$)	F*downstream-A	0.0026
F*distal-C	F*upstream-E	F*01:01:02:08-compatible (new-238 T, new2486T, new3205G, $[TG]_{10}$)	F*downstream-C	0.0026
F*distal-C	F*upstream-C	F*01:01:02:08 (new1570A, $[TG]_{11}$)	F*downstream-C	0.0026
F*distal-B	F*upstream-B	F*01:01:01:01	F*downstream-F	0.0026
F*distal-C	F*upstream-C	F*01:01:02:08-compatible ($[TG]_{11}$)	F*downstream-C	0.0026
F*distal-A	F*upstream-A	F*01:03:01:01 ($[TG]_{11}$)	F*downstream-D	0.0026
F*distal-A	F*upstream-B	F*01:01:01:08	F*downstream-J	0.0026
F*distal-C	F*upstream-C	F*01:01:02:08-compatible ($[TG]_{12}$)	F*downstream-I	0.0026
F*distal-E	F*upstream-A	F*01:01:01:11	F*downstream-B	0.0026
F*distal-A	F*upstream-A	F*01:03:01:01 (new1225C)	F*downstream-D	0.0026
F*distal-B	F*upstream-B	F*01:01:01:01 ($[TG]_{14}$)	F*downstream-A	0.0026
F*distal-A	F*upstream-A	F*01:03:01:01 ($[TG]_{13}$)	F*downstream-D	0.0026
F*distal-A	F*upstream-A	F*01:01:01:11 ($[TG]_{9}$)	F*downstream-B	0.0026
F*distal-F	F*upstream-B	F*01:01:01:01	F*downstream-A	0.0026
Nucleotide Diversity 0.0040 ± 0.0020				
Gene Diversity 0.9267 ± 0.0062				

[a] Haplotypes detected in this study and described in Table 2.
[b] Haplotypes detected in a previous study [Lima, 2016]. Some haplotype names were updated following the new IPD-IMGT/HLA database version 3.31.0. The updated names were as follows: F*01:03:01:01 (1383G) to F*01:03:01:03-compatible (($[TG]_{12}$); F*01:01:02:02 (1943G, 2208C, $[TG]_{11}$) to F*01:01:02:08-compatible ($[TG]_{11}$); F*01:01:02:03 (new1570A, 1943G, 2208C, $[TG]_{11}$) to F*01:01:02:08-compatible (new1570A, $[TG]_{11}$); F*01:01:02:02 (new-238 T, 1943G, 2208C, new2486T, new3205G) to F*01:01:02:08-compatible (new-238T, new2486T, new3205G, $[TG]_{10}$); F*01:01:02:02 (new-238 T, 1943G, 2208C, new2486T, new3205G, $[TG]_{13}$) to F*01:01:02:08-compatible (new-238T, new2486T, new3205G, $[TG]_{13}$); F*01:01:02:03 (1943G, 2208C, $[TG]_{11}$) to F*01:01:02:08-compatible ($[TG]_{11}$); F*01:01:02:03 (1943G, $[TG]_{13}$) to F*01:01:02:07-compatible ($[TG]_{13}$); F*01:01:02:04 (3189T) to F*01:01:02:05 ($[TG]_{11}$); F*01:01:02:03 (1943G, 2208C, $[TG]_{12}$) to F*01:01:02:08-compatible ($[TG]_{12}$); F*01:03:01:01 (new1378T, 1383G, $[TG]_{11}$) to F*01:03:01:03-compatible (new1378T, $[TG]_{11}$).

due to the low sequencing depth that characterizes this sample. Some variants associated with F*01:01:02 can also be detected in sample of 6 chimpanzees, including the variants -104G, +63A, +473C, +1094G, and +1882T. However, variants associated with F*Distal-C seems to be absent. This raises the speculation that this promoter associated with higher expression levels may be absent in primates, although larger and different primate samples are yet to be studied. Thus, it is possible that F*Distal-C/F*01:01:02 is present only among the genus Homo, and has emerged from an older sequence already available in primates. This haplotype is quite old since we detected it in all modern human populations here studied, as well as in ancient humans.

Previous reports showed that the *HLA-F* proximal promoter presents a similar structure to other *HLA* class I genes promoters [40,41]. For *HLA-F*, there are responsive elements to NF-kB, IRF1, IFN-γ, and CIITA [42,43]. However, differently from classical *HLA* class I genes, but similarly to *HLA-G*, *HLA-F* promoter does not have a conserved Inr ("Initiator") element, which explains the relatively limited HLA-F expression [43]. The *HLA-F* distal promoter region may comprise regulatory protein binding sites, enhancing or silencing gene expression. In 1990, when Geraghty and colleagues first described the *HLA-F* promoter comprehending 300 bp upstream the first translated ATG, they described a highly repetitive sequence and the presence of microsatellites. These repetitive sequences might play a regulatory function [40].

Although most of the promoter haplotypes are similar to F*distal-A (Fig. S1), the presence of a highly divergent and frequent haplotype (F*distal-C, associated with F*01:01:02 alleles, frequency of 17.16%) may be associated with a differential *HLA-F* transcriptional profile. The HaploReg analysis supports this hypothesis (Table 3 and Table S2), and all variable sites associated with F*distal-C are registered as eQTLs for *HLA-F* expression for many different tissues. The remaining distal promoters seem to present similar binding profiles, and they are associated with both F*01:01:01 and F*01:03. The *HLA-F* mRNA expression levels survey from 445 individuals of the 1000Genomes project also

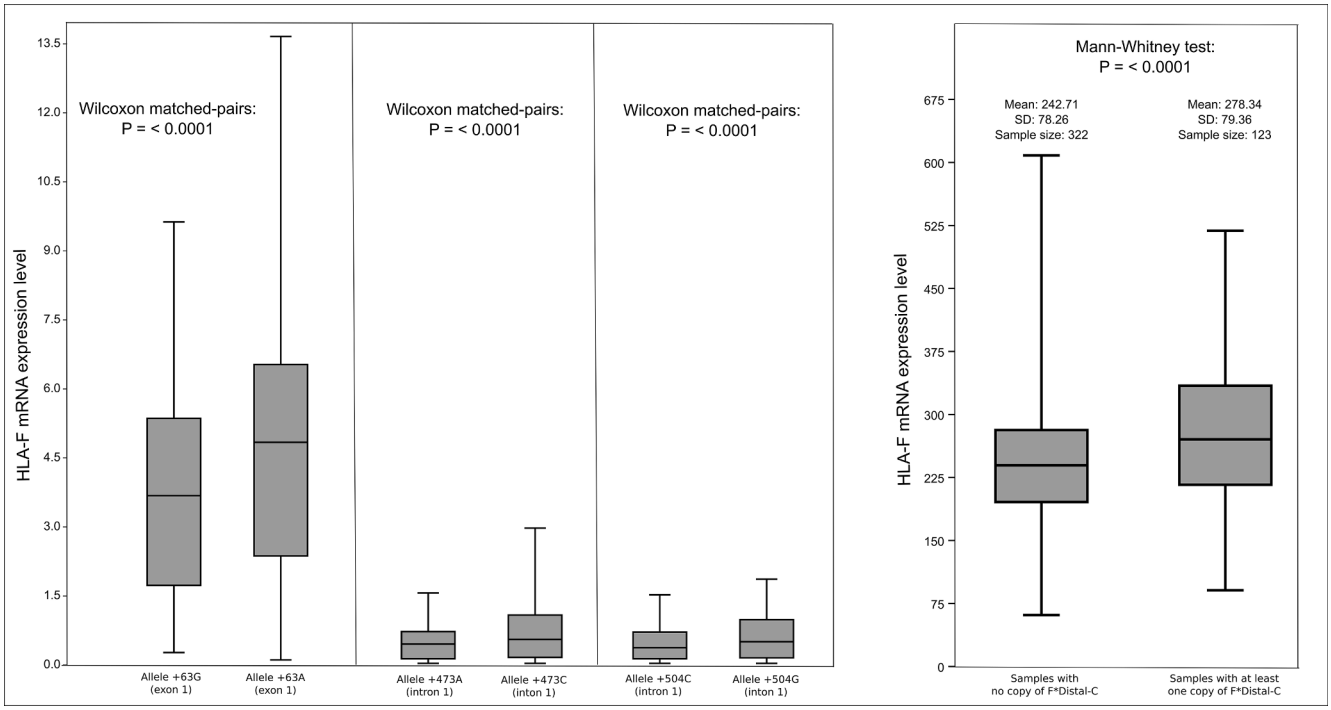

Fig. 2. *HLA-F* expression levels considering different *HLA-F* genotypes. Left panel: comparison of the number of fragments detected for each tested allele in heterozygous samples, indicating a higher expression of the chromosome carrying the *HLA-F* alleles +63A, +473C, and +504G, all exclusively associated with F*01:01:02 alleles. Right panel: comparison of the *HLA-F* mRNA expression levels in individuals that do not present F*Distal-C/F*01:01:02 alleles against individuals with F*Distal-C/F*01:01:02 alleles (in homozygous or heterozygous state), indicating a higher *HLA-F* expression level for individuals carrying F*Distal-C/F*01:01:02. The *HLA-F* expression level for each sample was calculated as follows: number of fragments detected by taking into account each one of the 33 exonic and intronic motifs, dividing the sum by the number of fragments at the RNA-Seq library, and multiplying this result by 10^6.

corroborates this hypothesis. Individuals carrying the F*distal-C/F*01:01:02 haplotype are associated with a higher *HLA-F* mRNA expression levels (Fig. 2, right panel), due to a higher presence of F*01:01:02 transcripts (Fig. 2, left panel).

According to the Human Protein Atlas (https://www.proteinatlas.org/), most of the transcription factors binding preferentially to F*distal-C presents a restrict expression pattern or are expressed ubiquitously in low levels. For instance, FOXF1 is expressed mainly on placenta, lung, gallbladder, prostate, and tissues from the digestive tract; FOXI1 is expressed on breast, kidney, prostate and salivary gland tissues; DMRT1 is only expressed on testis; and SP1 is expressed ubiquitously in low levels. Interestingly, according to the Human Protein Atlas, lung, gallbladder, testis and many tissues from the digestive tract present high levels of *HLA-F* mRNA. However, a similar rationale might be applied to transcription factors binding preferentially to the reference alleles (F*distal-A and derivatives), such as PAX5, which is mostly expressed on cells of the immune system and cells expressing high *HLA-F* mRNA levels, or RFX5, which is expressed in high levels on many tissues. Thus, it is possible that these haplotypes are associated with specific expression levels at particular tissues. Notwithstanding that, F*distal-C/F*01:01:02 is clearly related to a higher *HLA-F* expression (Fig. 2).

The exact mechanism by which F*distal-C/F*01:01:02 influences *HLA-F* expression is unknown. First, F*distal-C presents a different transcription factor binding profile, as demonstrated by the HaploReg analysis. Second, F*01:01:02 alleles present specific variants at the proximal promoter, such as Thymine at position -144 (other alleles present Adenine) and Guanine at position -101 (other alleles present Cytosine). These mutations do not modify the TATA or CAAT boxes at the *HLA-F* proximal promoter. Third, the haplotype F*distal-C/F*01:01:02 presents fewer CpG sites than others [positions -1116, -1013, and -101 destroys CpG sites, while position -116 creates one]. Thus, in this context, F*distal-C would be less methylated than others. Fourth, according to the 1000Genomes data, at least twelve other mutations are in complete LD with F*distal-C at the segment encompassing positions −3142 and −1992, and other intronic variants are also in absolute LD with F*Distal-C, which might also influence the binding of transcription factors.

From the *HLA-F* distal promoter analyses, we may conclude that this gene is indeed conserved, once its nucleotide diversity is the lowest observed among all *HLA-F* segments. *HLA-F* promoter presents two frequent and highly divergent haplotypes, with predicted different transcription factor binding capabilities. Most of the variants here detected are associated with a single haplotype here named F*distal-C, which was associated with higher *HLA-F* mRNA expression levels. F*Distal-C is associated with F*01:01:02 alleles. This high-expressing haplotype is common in worldwide populations, and is also detected at Neanderthal and Denisovan samples. Independently of the role played by HLA-F, such as the stabilization of HLA class I molecules presenting an open conformation, or the recycling of HLA class I molecules, or even as an antigen presenting molecule or ligand to KIR receptors, the conservation of this gene suggests a critical role for the cell physiology. Further functional studies are required for the understanding of the HLA-F function, and to define the exact mechanism by which F*distal-C/F*01:01:02 is related to a higher *HLA-F* mRNA expression level.

Acknowledgements

This work was supported by FAPESP/Brazil (Grant# 2013/17084-2). E.C.C. and C.T.M.J. are supported by CNPq/Brazil (Grants# 302590/2016-1 and 309572/2014-2). R.V.B, J.R., T.H.A.L., and L.C.V.C. are supported by the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES/Brazil).

Conflict of interests

There is no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2018.10.016.

References

[1] E.J. Lepin, J.M. Bastin, D.S. Allan, G. Roncador, V.M. Braud, D.Y. Mason, et al., Functional characterization of HLA-F and binding of HLA-F tetramers to ILT2 and ILT4 receptors, Eur. J. Immunol. 30 (2000) 3552.

[2] S.D. Wainwright, P.A. Biro, C.H. Holmes, HLA-F is a predominantly empty, intracellular, TAP-associated MHC class Ib protein with a restricted expression pattern, J. Immunol. 164 (2000) 319.

[3] N. Lee, A. Ishitani, D.E. Geraghty, HLA-F is a surface marker on activated lymphocytes, Eur. J. Immunol. 40 (2010) 2308.

[4] R. Hackmon, L. Pinnaduwage, J. Zhang, S.J. Lye, D.E. Geraghty, C.E. Dunk, Definitive class I human leukocyte antigen expression in gestational placentation: HLA-F, HLA-E, HLA-C, and HLA-G in extravillous trophoblast invasion on placentation, pregnancy, and parturition, Am. J. Reprod. Immunol. 77 (2017).

[5] A. Ishitani, N. Sageshima, N. Lee, N. Dorofeeva, K. Hatake, H. Marquardt, et al., Protein expression and peptide binding suggest unique and interacting functional roles for HLA-E, F, and G in maternal-placental immune recognition, J. Immunol. 171 (2003) 1376.

[6] J.M. Houlihan, P.A. Biro, A. Fergar-Payne, K.L. Simpson, C.H. Holmes, Evidence for the expression of non-HLA-A,-B,-C class I genes in the human fetal liver, J. Immunol. 149 (1992) 668.

[7] N. Lee, D.E. Geraghty, HLA-F surface expression on B cell and monocyte cell lines is partially independent from tapasin and completely independent from TAP, J. Immunol. 171 (2003) 5264.

[8] A. Harada, S. Ishigami, Y. Kijima, A. Nakajo, T. Arigami, H. Kurahara, et al., Clinical implication of human leukocyte antigen (HLA)-F expression in breast cancer, Pathol. Int. 65 (2015) 569.

[9] S. Ishigami, T. Arigami, H. Okumura, Y. Uchikado, Y. Kita, H. Kurahara, et al., Human leukocyte antigen (HLA)-E and HLA-F expression in gastric cancer, Anticancer Res. 35 (2015) 2279.

[10] A. Lin, X. Zhang, Y.Y. Ruan, Q. Wang, W.J. Zhou, W.H. Yan, HLA-F expression is a prognostic factor in patients with non-small-cell lung cancer, Lung Cancer 74 (2011) 504.

[11] X. Zhang, A. Lin, J.G. Zhang, W.G. Bao, D.P. Xu, Y.Y. Ruan, et al., Alteration of HLA-F and HLA I antigen expression in the tumor is associated with survival in patients with esophageal squamous cell carcinoma, Int. J. Cancer 132 (2013) 82.

[12] L.H. Boyle, A.K. Gillingham, S. Munro, J. Trowsdale, Selective export of HLA-F by its cytoplasmic tail, J. Immunol. 176 (2006) 6464.

[13] J.P. Goodridge, A. Burian, N. Lee, D.E. Geraghty, HLA-F and MHC class I open conformers are ligands for NK cell Ig-like receptors, J. Immunol. 191 (2013) 3553.

[14] J.P. Goodridge, N. Lee, A. Burian, C.W. Pyo, S.S. Tykodi, E.H. Warren, et al., HLA-F and MHC-I open conformers cooperate in a MHC-I antigen cross-presentation pathway, J. Immunol. 191 (2013) 1567.

[15] A. Burian, K.L. Wang, K.A. Finton, N. Lee, A. Ishitani, R.K. Strong, et al., HLA-F and MHC-I open conformers bind natural killer cell Ig-Like receptor KIR3DS1, PLoS One 11 (2016) e0163297.

[16] W.F. Garcia-Beltran, A. Holzemer, G. Martrus, A.W. Chung, Y. Pacheco, C.R. Simoneau, et al., Open conformers of HLA-F are high-affinity ligands of the activating NK-cell receptor KIR3DS1, Nat. Immunol. 17 (2016) 1067.

[17] C.L. Dulberger, C.P. McMurtrey, A. Holzemer, K.E. Neu, V. Liu, A.M. Steinbach, et al., Human leukocyte antigen F presents peptides and regulates immunity through interactions with NK cell receptors, Immunity 46 (2017) 1018.

[18] S. Song, C.J. Miranda, L. Braun, K. Meyer, A.E. Frakes, L. Ferraiuolo, et al., Major histocompatibility complex class I molecules protect motor neurons from astrocyte-induced toxicity in amyotrophic lateral sclerosis, Nat. Med. 22 (2016) 397.

[19] C.K. Burrows, G. Kosova, C. Herman, K. Patterson, K.E. Hartmann, D.R. Velez Edwards, et al., Expression quantitative trait locus mapping studies in mid-secretory phase endometrial cells identifies HLA-F and TAP2 as fecundability-associated genes, PLoS Genet. 12 (2016) e1005858.

[20] T.H. Lima, R.V. Buttura, E.A. Donadi, L.C. Veiga-Castelli, C.T. Mendes-Junior, E.C. Castelli, HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample, Hum. Immunol. 77 (2016) 841.

[21] E.C. Castelli, P. Gerasimou, M.A. Paz, J. Ramalho, I.O. Porto, T.H. Lima, et al., HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographicaly distinct population samples of Brazil and Cyprus, Mol. Immunol. 83 (2017) 115.

[22] J. Ramalho, L.C. Veiga-Castelli, E.A. Donadi, C.T. Mendes-Junior, E.C. Castelli, HLA-E regulatory and coding region variability and haplotypes in a Brazilian population sample, Mol. Immunol. 91 (2017) 173.

[23] T.V. Hviid, S. Hylenius, A.M. Hoegh, C. Kruse, O.B. Christiansen, HLA-G polymorphisms in couples with recurrent spontaneous abortions, Tissue Antigens 60 (2002) 122.

[24] C. Matte, J. Lacaille, L. Zijenah, B. Ward, M. Roger, HLA-G and HLA-E polymorphisms in an indigenous African population. The ZVITAMBO study group, Hum. Immunol. 61 (2000) 1150.

[25] C.W. Pyo, L.M. Williams, Y. Moore, H. Hyodo, S.S. Li, L.P. Zhao, et al., HLA-E, HLA-F, and HLA-G polymorphism: genomic sequence defines haplotype structure and variation spanning the nonclassical class I genes, Immunogenetics 58 (2006) 241.

[26] L.F. Manvailer, P.F. Wowk, S.B. Mattar, J.S. da Siva, Bicalho M. da Graca, V.M. Roxo, HLA-F polymorphisms in a Euro-Brazilian population from Southern Brazil, Tissue Antigens 84 (2014) 554.

[27] R. Daza-Vamenta, G. Glusman, L. Rowen, B. Guthrie, D.E. Geraghty, Genetic divergence of the rhesus macaque major histocompatibility complex, Genome Res. 14 (2004) 1501.

[28] J. Moscoso, J.I. Serrano-Vela, A. Arnaiz-Villena, MHC-F polymorphism and evolution, Tissue Antigens 69 (Suppl. 1) (2007) 136.

[29] R. Rojo, M.J. Castro, J. Martinez-Laso, J.I. Serrano-Vela, P. Morales, J. Moscoso, et al., MHC-F DNA sequences in bonobo, gorilla and orangutan, Tissue Antigens 66 (2005) 277.

[30] E.C. Castelli, M.A. Paz, A.S. Souza, J. Ramalho, C.T. Mendes-Junior, Hla-mapper: an application to optimize the mapping of HLA sequences produced by massively parallel sequencing procedures, Hum. Immunol. 79 (2018) 678.

[31] M. Stephens, N.J. Smith, P. Donnelly, A new statistical method for haplotype reconstruction from population data, Am. J. Hum. Genet. 68 (2001) 978.

[32] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows, Mol. Ecol. Resour. 10 (2010) 564.

[33] J.C. Barrett, Haploview: Visualization and analysis of SNP genotype data, Cold Spring Harb. Protoc. (2009) 2009:pdb ip71.

[34] L.D. Ward, M. Kellis, HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants, Nucleic Acids Res. 40 (2012) D930.

[35] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 1870 (2016) 33.

[36] T. Lappalainen, M. Sammeth, M.R. Friedlander, P.A. t Hoen, J. Monlong, M.A. Rivas, et al., Transcriptome and genome sequencing uncovers functional variation in humans, Nature (2013) 501–506.

[37] F.H. Pan, X.X. Liu, W. Tian, Characterization of HLA-F polymorphism in four distinct populations in Mainland China, Int. J. Immunogenet. 40 (2013) 369.

[38] K. Prufer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, et al., The complete genome sequence of a Neanderthal from the Altai Mountains, Nature 505 (2014) 43.

[39] D. Reich, R.E. Green, M. Kircher, J. Krause, N. Patterson, E.Y. Durand, et al., Genetic history of an archaic hominin group from Denisova Cave in Siberia, Nature 468 (2010) 1053.

[40] D.E. Geraghty, X.H. Wei, H.T. Orr, B.H. Koller, Human leukocyte antigen F (HLA-F). An expressed HLA gene composed of a class I coding sequence linked to a novel transcribed repetitive element, J. Exp. Med. 171 (1990) 1.

[41] X.H. Wei, H.T. Orr, Differential expression of HLA-E, HLA-F, and HLA-G transcripts in human tissue, Hum. Immunol. 29 (1990) 131.

[42] S.J. Gobin, P.J. van den Elsen, Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F, and HLA-G, Hum. Immunol. 61 (2000) 1102.

[43] T. Howcroft, D. Singer, Expression of nonclassical MHC class Ib genes: comparison of regulatory elements, Immunol. Res. 27 (2003) 1.