



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

Marília Rodrigues Silva Passos

DIVERSIDADE GENÉTICA E VARIANTES ESTRUTURAIS
DO BLOCO ALFA DO COMPLEXO PRINCIPAL DE
HISTOCOMPATIBILIDADE HUMANO

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

Orientador: Prof. Dr. Erick da Cruz Castelli

Botucatu
2020

Marília Rodrigues Silva Passos

Diversidade genética e variantes estruturais do bloco alfa do
complexo principal de histocompatibilidade humano

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

Orientador: Prof. Dr. Erick da Cruz Castelli

Botucatu

2020

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Passos, Marília Rodrigues Silva.

Diversidade genética e variantes estruturais do bloco alfa do complexo principal de histocompatibilidade humano / Marília Rodrigues Silva Passos. - Botucatu, 2020

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu

Orientador: Erick da Cruz Castelli

Capes: 21103003

1. Genes MHC classe I. 2. Antígenos HLA. 3. Histocompatibilidade. 4. Variação genética.

Palavras-chave: Alpha block; HLA; MHC; NGS; Variantes estruturais.

DEDICATÓRIA

À minha família, meu filho Miguel Schitini Silva Passos, meu marido João Vitor Schitini Passos e nossos dois anjos que estão no céu. Meus pais, Valdicéia Alves R. Silva e Luiz Antonio da Silva Junior, meu irmão Luiz Leonardo R. Silva, minha madrinha Neusa Maria R. F. da Silva e à nossa Ana Luiza.

AGRADECIMENTOS

Agradeço primeiramente a Deus pela vida, pelas oportunidades, pelos livramentos e pelas pessoas maravilhosas que surgiram em meu caminho. Pela chance de estudar e tentar ser uma pessoa melhor no mundo.

À meus pais, Valdicéia e Luiz Antônio ou Val e Luizão, pela criação e carinho recebido em toda minha vida. A dificuldade que superaram para nos criar em uma casa simples, num bairro simples no interior só me ensinou a valorizar a família e que é preciso esforço e dedicação para crescer. Todo esforço de vocês está refletido na vida maravilhosa que levam, e todo amor que sentem pelo meu filho Miguel está refletido no sorriso do nosso pequeno.

Ao melhor amigo, companheiro e marido do mundo, João Vitor, que se dispôs a cuidar de nossa família sozinho para que eu realizasse o sonho do mestrado. Foram muitas noites sem voltar pra casa, viagens para congressos, diagnósticos preocupantes, semanas estressantes e você ali comigo, sem deixar que eu desista. Nossa casa, nosso filho, nossa família toda sendo cuidada por você nesses três anos, representando o pilar que você sempre foi para nós. Te amo.

Aos meus “amigos-professores” da graduação Lígia Maria Micai Gomide, Messias Miranda Junior, Renato Paschoal Prado, Thiago Cândido e Bruna Ventura de Campos Camargo, pelo incentivo e ajuda nos anos que passamos juntos e mais ainda depois que decidi enfrentar mais esse desafio. Vocês foram essenciais para minha inserção no mundo científico. Para todo sempre meu muito obrigada!

Ao meu orientador Dr Erick da Cruz Castelli por me aceitar em sua equipe e apresentar o maravilhoso mundo da Bioinformática e Imunogenética, sempre ensinando com muita paciência e se preocupando com todos seus alunos como se fossem filhos. Tenho certeza que sentirei muita falta durante nosso “período de pausa”. Serei sempre grata por tudo.

Aos meus colegas do Laboratório de Genética Molecular e Bioinformática por todos os momentos de alegria, tristeza, concentração e preocupação. Tudo que aprendemos juntos e ensinamos juntos só me tornou uma pessoa ainda melhor, absorvendo um pouco do que cada um traz no coração. Thálitta e Andreia, obrigada pela ótima recepção e pela parceria que sempre tivemos, vocês são excelentes educadoras e sempre possuirão um lugar na minha família. Emiliana e Heloisa, agradeço por todos os perrengues que passamos juntas, por toda preocupação, por toda superação, por me acolher sempre e sempre vibrar com minhas conquistas como se fosse de vocês. Juliana Lara, Michele, Iane e Camila, sempre serão minhas mestres e nunca me cansarei de sentir orgulho de vocês. Enfim, a todos os companheiros de vida nesse laboratório meu muito obrigada.

Ao Programa de Pós Graduação em Patologia da Faculdade de Medicina de Botucatu (FMB-UNESP) pela oportunidade de aprender e a todo o corpo docente pelos ensinamentos nas

disciplinas do programa. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Por fim, a todos os membros da banca que se disponibilizaram a estar aqui e contribuir para minha formação bem como para a melhoria desse trabalho. A todos que contribuíram direta ou indiretamente para a realização de mais essa conquista.

Muito obrigada!

“O êxito da vida não se mede pelo caminho que você conquistou, mas sim pelas dificuldades que superou no caminho”

Abraham Lincoln

RESUMO

Resumo

O MHC humano possui cerca de 224 genes, a maioria deles está relacionada de alguma forma com o sistema imunitário. O bloco alfa do MHC humano compreende genes importantes do complexo HLA, como *HLA-F*, *HLA-G* e *HLA-A*. *HLA-F* e *HLA-G* são genes HLA não clássicos que possuem baixa variabilidade, expressão restrita em tecidos e tem função imunomodulatória, enquanto *HLA-A* é um gene altamente polimórfico e importante para a apresentação de antígenos. Este segmento também apresenta algumas variantes estruturais, como elementos *Alu* e CNVs. O *HLA-G* e o *HLA-A* foram anteriormente descritos para frequência e variabilidade em pequenas amostras, além disso, a diversidade de *HLA-F* em populações mundiais é desconhecida. Aqui aplicamos um *pipeline* de bioinformática que otimiza o mapeamento de leituras curtas para genes HLA e permite a determinação de haplótipos confiáveis para pesquisar a diversidade genética do bloco alfa do MHC em 1323 brasileiros, juntamente da avaliação de três variantes estruturais, *AluyHF*, *AluyHG* e *esv3608493*, em uma amostra altamente miscigenada da maior cidade brasileira, São Paulo. O *HLA-F* é o mais conservado, com 82,65% dos cromossomos codificando a mesma proteína *HLA-F* e com apenas uma proteína adicional frequente. O gene *HLA-A* mostrou-se altamente polimórfico, com 96 sequências genômicas codificando 59 proteínas, incluindo as raras. Foi encontrado forte LD entre *AluyHF* e *HLA-F*, mas não com alelos específicos de *HLA-F*. Enquanto *AluyHG* foi encontrado em forte LD com um alelo específico de *HLA-G*, *G*01:01:01:01/UTR-1*, e *esv3608493* foi associado apenas aos alelos *HLA-A*23* e *A*24*. Também detectamos haplótipos frequentes em nossa amostra, especialmente a estreita relação entre os alelos *HLA-G* e *HLA-A*. Por esse motivo, a evidência de seleção balanceadora em ambos os genes pode não ser independente, pois a alta heterozigose em um locus aumentaria a heterozigose no outro. Esses dados reúnem um conjunto de informações valiosas sobre a variabilidade e estrutura de haplótipos estendidos na região do alfa-bloco do complexo HLA, que podem contribuir em estudos clínicos e evolutivos.

ABSTRACT

Abstract

Human MHC has about 224 genes, most of which are related in some way to the immune system. The alpha block of human MHC comprises important genes of the HLA complex, such as *HLA-F*, *HLA-G*, and *HLA-A*. *HLA-F* and *HLA-G* are non-classical HLA genes that have low variability, restricted expression in tissues and have an immunomodulatory function, while *HLA-A* is a highly polymorphic gene and important for the presentation of antigens. This segment also has some structural variants, such as Alu elements and CNVs. *HLA-G* and *HLA-A* have previously been described for frequency and variability in small samples, also, the diversity of *HLA-F* in populations worldwide is unknown. Here we apply a bioinformatics pipeline that optimizes the mapping of short readings for HLA genes and allows the determination of reliable haplotypes to research the genetic diversity of the MHC alpha block in 1323 Brazilians, together with the evaluation of three structural variants, *AluHF*, *AluHG*, and ns823470, in a highly mixed sample from the largest Brazilian city, São Paulo. *HLA-F* is the most conserved, with 82.65% of chromosomes encoding the same *HLA-F* protein and with only one additional frequent protein. The *HLA-A* gene was shown to be highly polymorphic, with 96 genomic sequences encoding 59 proteins, including rare ones. Strong LD was found between *AluHF* and *HLA-F*, but not with specific *HLA-F* alleles. While *AluHG* was found in strong LD with a specific *HLA-G* allele, G*01:01:01:01/UTR-1 and ns823470 were associated only with the *HLA-A**23 and A*24 alleles. We also detected frequent haplotypes in our sample, especially the close relationship between the *HLA-G* and *HLA-A* alleles. For this reason, the evidence of balanced selection in both genes may not be independent, since high heterozygosity in one locus would increase heterozygosity in the other. These data bring together a set of valuable information about the variability and structure of extended haplotypes in the alpha-block region of the HLA complex, which can contribute to clinical and evolutionary studies.

Lista de ilustrações

Capítulo 1. Revisão de Literatura

- Figura 1.** Mapa genômico humano da localização dos genes MHC de classe I, divididos nos blocos de segregação: Beta, Kappa e Alfa. Na imagem é possível observar os genes *HLA-A*, *-G* e *-F* (objetos do presente estudo), além dos genes *MICA*, *MICB* e os pseudogenes *HLA-H* e *-J*.....**21**
- Figura 2.** Composição gênica do bloco alfa do MHC humano e as variantes estruturais presentes nele.....**26**

Capítulo 2. Artigo

- Figure 3.** A map of the human MHC alpha block and the genes and structural variants evaluated in a large admixed Brazilian sample from São Paulo.**47**
- Figure 4.** Linkage disequilibrium across the alpha-block, considering 638 biallelic SNPs, and the structural variants *AluyHF*, *AluyHG*, and *nsv823470*.**47**

Lista de tabelas

Capítulo 2. Artigo

Table 1.	<i>HLA-F</i> CDS alleles in an admixed sample from São Paulo (city), Brazil.....	48
Table 2.	<i>HLA-F</i> coding alleles in an admixed sample from São Paulo (city), Brazil.....	48
Table 3.	<i>HLA-G</i> CDS alleles in an admixed sample from São Paulo (city), Brazil.	49
Table 4.	<i>HLA-G</i> genomic alleles in an admixed sample from São Paulo (city), Brazil.	50
Table 5.	<i>HLA-A</i> CDS alleles in an admixed sample from São Paulo (city), Brazil.....	51
Table 6.	<i>HLA-A</i> genomic alleles in an admixed sample from São Paulo (city), Brazil..	52
Table 7.	<i>HLA-F</i> genomic alleles and the AluyHF haplotypes in an admixed sample from São Paulo (city), Brazil.	53
Table 8.	<i>HLA-G</i> genomic alleles and the AluyHG haplotypes in an admixed sample from São Paulo (city), Brazil.	54

Material suplementar

Table S1.	List of variants observed in the across the <i>HLA-F</i> gene in 1323 Brazilians from São Paulo, and their frequencies.....	56
Table S2.	List of variants observed in the across the <i>HLA-G</i> gene in 1323 Brazilians from São Paulo, and their frequencies.	59
Table S3.	List of variants observed in the across the <i>HLA-A</i> gene in 1323 Brazilians from São Paulo, and their frequencies.	62
Table S4.	List of all <i>HLA-G</i> / <i>HLA-A</i> haplotypes observed in 1323 Brazilians from São Paulo and their frequencies.	74
Table S5.	List of all <i>HLA-G</i> / <i>HLA-A</i> haplotypes observed in 1323 Brazilians from São Paulo and their frequencies.....	76

Lista de abreviaturas e siglas

MHC	<i>Major Histocompatibility Complex</i> (Complexo Principal de Histocompatibilidade)
Mb	Megabases (106 bases)
HLA	<i>Human Leukocyte Antigen</i> (Antígeno Leucocitário Humano)
β	Beta
α	Alfa
Ig	Imunoglobulina
MICA	<i>Major histocompatibility complex class I chain-related gene A</i>
MICB	<i>Major histocompatibility complex class I chain-related gene B</i>
NK	<i>Natural Killer</i>
IPD	<i>Immuno Polymorphism Database</i>
IMGT	<i>ImMunoGeneTics information system</i>
3'UTR	<i>3' Untranslated region</i>
5'UTR	<i>5' Untranslated region</i>
HIV-1	Vírus da imunodeficiência humana tipo 1
Aids	Síndrome da imunodeficiência adquirida
DM-1	Diabetes melitus tipo 1
i. e.	<i>id est</i> (isto é)
SINE	<i>Short Interspersed Nuclear Element</i>
bp ou pb	<i>base pairs</i> (pares de base)
RNA	Ácido ribonucleico
LINE	<i>Long interspersed Nuclear Elements</i>
PCR	Reação em cadeia da polimerase
DNA	Ácido desoxirribonucleico
SNP	<i>Single nucleotide polymorphism</i> (Polimorfismo de base única)
Kb	Kilobases (103 bases)
CNV	<i>Copy Number Variation</i>
LD	<i>Linkage disequilibrium</i> (Desequilíbrio de ligação)
ILT2	<i>Human immunoglobulin-like transcript 2</i>
ILT4	<i>Human immunoglobulin-like transcript 4</i>
KIR3DL2	<i>Killer cell immunoglobulin-like receptor 3DL2</i>
KIR2DS4	<i>Killer cell immunoglobulin-like receptor 2DS4</i>
HCG4B	<i>HLA complex group 4B</i>
WGS	<i>Whole Genome Sequencing</i>
SABE	Saúde, Bem Estar e Envelhecimento
CEGH-CEL	Centro de Estudos do Genoma Humano e Células-Tronco
EDTA	Ethylene diaminetetracetic acid
Hg38	<i>Humam genome version 38</i>
BAM	<i>Binary Alignment Map</i>
GATK	<i>Genome Analysis Toolkit</i>

GVCF	<i>GenomicVariant Call Format</i>
RBP	<i>ReadBackedPhasing</i>
VCF	<i>Variant Call Format</i>
MAF	<i>Minimum allele frequency</i>
CDS	<i>Coding Sequence</i>
STR	<i>Short tandem repeats</i>
NGS	<i>Next generation sequencing</i>
REDOME	Registro Nacional de Doadores de Medula Óssea
Chr6	Cromossomo 6
SNPid	Identificação do SNP

SUMÁRIO

CAPÍTULO 1	18
1. Revisão de Literatura	19
1.1. O MHC humano	19
1.2. HLA-A.....	21
1.3. HLA-G.....	22
1.4. HLA-F	23
1.5. Inserções Alu	23
1.6. Forças seletivas.....	23
1.7. Referências	27
CAPÍTULO 2	30
2. Artigo	31
2.1. Introduction	32
2.2. Methods	36
2.3. Results	38
2.4. Discussion.....	41
MATERIAL SUPLEMENTAR	55
Table S1. List of variants observed in the across the <i>HLA-F</i> gene in 1323 Brazilians from São Paulo, and their frequencies.....	56
Table S2. List of variants observed in the across the <i>HLA-G</i> gene in 1323 Brazilians from São Paulo, and their frequencies.....	59
Table S3. List of variants observed in the across the <i>HLA-A</i> gene in 1323 Brazilians from São Paulo, and their frequencies.....	62
Table S4. List of all <i>HLA-G/HLA-A</i> haplotypes observed in 1323 Brazilians from São Paulo and their frequencies.....	74
Table S5. List of all <i>HLA-G/HLA-A</i> haplotypes observed in 1323 Brazilians from São Paulo and their frequencies.....	76
2.5. References	81
3. CONCLUSÃO.....	87

1. REVISÃO DE LITERATURA

1.1. O MHC humano

O complexo principal de histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*) localiza-se no braço curto do cromossomo seis (6p21.3) em um segmento de aproximadamente 3,5 Mb, com cerca de 224 genes. Esses genes e as moléculas que eles codificam, bem como sua organização cromossômica, estão presentes em todos os vertebrados que possuem mandíbula, os gnatostomados, numa organização que vem sendo mantida conservada por mais de 500 milhões de anos (KLEIN; SATO, 1998).

O MHC foi descoberto por meio de estudos de rejeição a transplantes em camundongos. Essa descoberta rendeu um prêmio Nobel em 1980 a George Davis Snell, Jean Baptiste Dausset e Baruj Benacerraf (RAJU, 1999). Esses pesquisadores perceberam que enxertos de pele entre camundongos isogênicos não eram rejeitados, enquanto transplantes entre linhagens diferentes eram rejeitados pelo receptor. Em humanos, essa descoberta foi feita a partir de três grupos de pacientes: (i) aqueles que haviam passado por múltiplas transfusões sanguíneas, (ii) pacientes que haviam passado por transplantes renais e (iii) mulheres múltiparas. Analisando o sangue desses pacientes, foi encontrado, respectivamente, anticorpos que reconheciam as células sanguíneas e renais dos doadores e as células paternas. Por esse motivo, o MHC humano foi chamado de Antígenos Leucocitários Humanos (HLA, do inglês *Human Leucocyte Antigens*) (TERASAKI, 2007; THORSBY, 2009). A maioria dos genes do MHC está relacionada de alguma forma com o sistema imunitário (GOLDBERG; RIZZO, 2015).

Alguns genes desse complexo são particularmente variáveis e seus produtos são fundamentais ao desenvolvimento da resposta imunológica. Com base em estrutura molecular e funções, esses genes foram divididos em três classes. Os genes de classe I e II também são conhecidos como genes HLA, uma vez que atuam na apresentação antigênica à linfócitos. Esses genes, principalmente aqueles cuja molécula codificada é responsável pela apresentação antigênica, estão entre os mais polimórficos do genoma humano (SHIINA et al., 2009).

Os principais genes de classe I envolvidos diretamente na apresentação de antígenos aos linfócitos T durante a resposta imune são os genes *HLA-A*, *HLA-B* e *HLA-C*, também chamados de genes clássicos de classe I (KLEIN; SATO, 2000). Estes genes são extremamente polimórficos e expressos na maioria das células nucleadas, apresentando juntos mais de 12.000 alelos conhecidos (<http://www.ebi.ac.uk/ipd/imgt/hla/stats.html>). O mais variável desses genes é o *HLA-B*, com 4.828 alelos, seguido pelo *HLA-A* (3.968 alelos) e *HLA-C* (3.579 alelos). Esses 3 genes codificam a cadeia alfa de um heterodímero de membrana, responsável pela

apresentação antigênica (a molécula HLA). As cadeias alfas ligam-se não covalentemente a uma molécula de $\beta 2$ -microglobulina, codificada por um gene localizado fora do complexo HLA, no cromossomo 15. A molécula HLA de classe I apresenta cinco domínios: dois domínios de ligação a peptídeos, $\alpha 1$ e $\alpha 2$, que juntos formam a fenda de ligação de peptídeo; um domínio de imunoglobulina (Ig), $\alpha 3$; um domínio transmembrana e a cauda citoplasmática.

A maioria das células humanas nucleadas expressam moléculas HLA de classe I clássicas na sua superfície, juntamente com um repertório de peptídeos de origem endógena associados a estas moléculas HLA. Os linfócitos T CD8⁺ precisam diferir peptídeos normais dos estranhos ao organismo (AMIGORENA, 2016; BRUTKIEWICZ, 2016). Desse modo, as moléculas HLA de classe I participam desse processo de reconhecimento do próprio e do não próprio. Para isso, os antígenos são mantidos na fenda de ligação a peptídeos por meio de interações não covalentes, entre a sequência do peptídeo e pontes de hidrogênio formadas entre aminoácidos conservados da molécula de HLA e aminoácidos do peptídeo (localizados nos finais da fenda) (JENSEN, 2007).

A variabilidade dos genes HLA clássicos está diretamente relacionada com a capacidade de apresentação de diferentes peptídeos, tanto do ponto de vista pessoal quanto populacional. Desse modo, a capacidade de apresentação antigênica das moléculas HLA possui um importante papel em infecções virais. Além disso, vários alelos de genes HLA associam-se à uma série de contextos patológicos, como doenças autoimunes, doenças infecciosas, neoplasias e o resultado de transplantes (OGAHARA et al., 1998; VANDIEDONCK et al., 2009).

Na mesma região genômica dos genes clássicos estão os genes *HLA-G*, *HLA-E* e *HLA-F*, conhecidos como genes HLA não clássicos de classe I. Ainda, na região de classe I do MHC humano temos os genes *MICA* e *MICB* (*Major histocompatibility complex class I chain-related gene A and B*). Esses genes codificam moléculas que apresentam similaridades estruturais com as moléculas HLA de classe I. No entanto, MICA e MICB não se ligam à $\beta 2$ -microglobulina ou apresentam peptídeos aos linfócitos T. A transcrição destas moléculas é induzida em condições de estresse celular, e ocorre principalmente na superfície celular de fibroblastos, epitélio intestinal e tumores. Essas moléculas interagem com o receptor NKG2D presente em células NK e em linfócitos T, modulando a atividade dessas células. MICA, por exemplo, é expresso em uma série de tecidos, principalmente tecidos endócrinos, fígado, vesícula biliar, pulmão e testículo (<http://www.proteinatlas.org/ENSG00000204520-MICA/tissue>).

O MHC é composto por blocos altamente polimórficos que foram gerados por uma série de duplicações, deleções e vários eventos de recombinação genômica durante a evolução de mamíferos e primatas. A região está dividida nos blocos Beta, Kappa e Alfa, conforme Figura

X e, este último, compreende os genes *HLA-A*, *HLA-G* e *HLA-F*, além de variantes estruturais abordadas posteriormente.

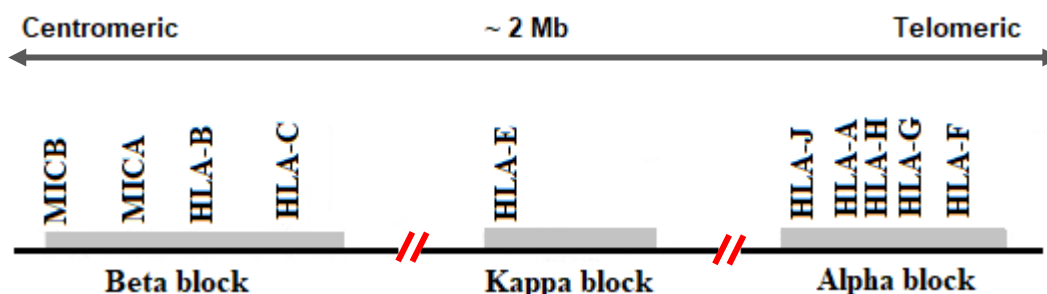


Figura 1. Mapa genômico humano da localização dos genes MHC de classe I, divididos nos blocos de segregação: Beta, Kappa e Alfa. Na imagem é possível observar os genes *HLA-A*, *-G* e *-F* (objetos do presente estudo), além dos genes *MICA*, *MICB* e os pseudogenes *HLA-H* e *-J*.

1.2. HLA-A

Considerando o banco de dados IPD-IMGT/HLA na versão 3.38.0, existem 18.691 alelos reconhecidos para os loci HLA de classe I. O *HLA-A* possui 8 éxons, sendo o oitavo também chamado de região 3' não traduzida (3'UTR) e a região 5'UTR, com 5.735 alelos que codificam 3.629 diferentes proteínas completas (IPD-IMGT/HLA). A maior parte dessa variabilidade é encontrada nos éxons 2 e 3, responsáveis pela codificação dos domínios $\alpha 1$ e $\alpha 2$, responsáveis pela formação da fenda de ligação ao peptídeo. Considerando que esse banco de dados, que rastreia e organiza todas as sequências conhecidas de genes HLA, geralmente solicita a avaliação e caracterização de apenas os éxons 2 e 3 para a anexação de uma nova sequência de alelos ao banco de dados, e não inclui o promotor completo e a região 3'UTR da maioria dos genes HLA, não está claro se os segmentos reguladores e outros éxons dos genes da classe I seguem a mesma natureza polimórfica dos éxons 2 e 3. Além disso, poucos estudos caracterizaram a variabilidade dessas regiões reguladoras, mesmo que de forma parcial.

Os dados de associação dos alelos *HLA-A* com doenças auto-imunes, infecciosas e o câncer estão bem presentes na literatura, com ênfase na relação do alelo *HLA-A*74:01* com a proteção contra a infecção pelo HIV-1 e com a lenta progressão para a aids. O grupo de alelos *HLA-A*02* possui muitas relações com patologias, como Alzheimer e tireoide de Hashimoto, além dos alelos *A*24:02* relacionado com a predisposição para diabetes mellitus tipo 1 (DM-1) e *A*11:01*, *A*32:01* ou *A*66:01* associados com a proteção contra essa doença (CARVALHO DOS SANTOS et al., 2013; LIMA et al., 2019). É pela função de reconhecer o próprio e o não próprio do organismo que os genes HLA de classe I estão relacionados com o

resultado de transplantes, dependendo da compatibilidade de doador-receptor do enxerto para o sucesso da técnica.

1.3. HLA-G

O *HLA-G* faz parte do grupo de genes não clássicos do MHC humano, caracterizados como conservados em relação aos genes clássicos do HLA. Apesar da molécula produzida por este gene apresentar a mesma estrutura dos genes clássicos, o mesmo está pouco envolvido com a apresentação antigênica e tem como função principal a imunomodulação das células T e Natural Killer (NK). Graças a essa imunomodulação das NK na interface materno-fetal, é viabilizado o crescimento de um feto semi alogênico, porém, tal expressão de *HLA-G* em células tumorais possibilita um escape imunológico da célula tumoral, inibindo a ação das células T contra esta célula alvo. Em situações patológicas é expresso *HLA-G* em câncer de cólon, câncer de mama, câncer renal, carcinoma de bexiga, entre outros, mas em situações não patológicas a expressão é restrita a alguns tecidos como córnea, timo e placenta (CAROSELLA et al., 2008; DUNKER et al., 2008; CASTELLI et al., 2008; GERAGHTY et al., 1992; DONADI et al., 2011).

Atualmente diversos estudos tem sido realizados para detalhar sua estrutura e variabilidade no Brasil e em diferentes populações e, de fato, as 19 proteínas diferentes descritas codificadas pelos seus 69 alelos (IPD-IMGT/HLA) não tem evidência de modificar sua função (CASTELLI et al., 2017; CASTELLI et al., 2010; CASTELLI et al., 2007; CASTELLI et al., 2010; CASTELLI et al., 2011; CASTELLI et al., 2014; LIMA et al., 2016; FELICIO et al., 2014; DONADI et al., 2011). Entretanto, a região promotora do gene *HLA-G* e sua região 3'UTR incluem muitos elementos reguladores transcricionais e pós-transcricionais, portanto, alguns sítios polimórficos nesses segmentos foram associados a diferentes perfis de expressão de *HLA-G*. Considerando que há baixa variabilidade na região de codificação do *HLA-G* e apenas poucas moléculas de proteína *HLA-G* de comprimento total são realmente detectadas em todo o mundo (CASTELLI et al., 2014, 2017); e que as principais funções biológicas da molécula, incluindo a dimerização e a interação com os receptores de leucócitos, são aparentemente conservadas para todas as frequentes moléculas de proteína *HLA-G* de comprimento total; a magnitude da produção de *HLA-G*, em condições normais ou patológicas, depende da constituição das regiões regulatórias (promotor e 3'UTR) e os fatores secretados no microambiente (CASTELLI et al., 2009; MANASTER et al., 2012; MOREAU; FLAJOLLET; CAROSELLA, 2009; PORTO et al., 2015; TAN et al., 2007; YIE et al., 2008; DONADI et al., 2011).

1.4. HLA-F

O *HLA-F* também é um gene não clássico do MHC humano, porém pouco explorado em termos de função e diversidade genética. Sabe-se atualmente que esse gene é pouco variável, ou seja, possui um baixo grau de polimorfismos, com apenas 44 alelos e 6 proteínas (HLA-F*01:01, HLA-F*01:02, HLA-F*01:03, HLA-F*01:04, HLA-F*01:05 e HLA-F*01:06) descritas no banco de dados IPD-IMGT/HLA versão 3.38.0. Como o gene *HLA-F* apresenta as mesmas características dos demais genes não clássicos, i.e., expressão restrita a alguns tecidos ou baixa expressão, baixo polimorfismo e interação da proteína HLA-F com receptores inibitórios ILT e KIR (GOODRIDGE et al., 2013; LEPIN et al., 2000), acredita-se que sua função também esteja relacionada com imunomodulação.

Até o momento foi detectada a expressão de HLA-F em situações fisiológicas no timo, baço e amígdala (LEPIN et al., 2000; WAINWRIGHT; BIRO; HOLMES, 2000). Em situações patológicas, a expressão de HLA-F foi descrita em vários tipos de câncer e doenças autoimunes (ISHIGAMI et al., 2015; JUCAUD et al., 2016), em geral relacionado ao prognóstico da doença. Os genes HLA não clássicos *HLA-G* e *HLA-E* são importantes durante a gestação e o transplante. No entanto, ainda existem poucos estudos que descrevam a relação do gene *HLA-F* com essas situações.

Salvo um estudo que avaliou a variabilidade do gene *HLA-F* em uma amostra da população brasileira (LIMA et al., 2016) não há estudos que tenham avaliado a diversidade desse gene em populações mundiais. Portanto, é possível que a variabilidade desses gene esteja bastante subestimada na literatura e nos bancos de dados de variantes HLA.

1.5. Variantes estruturais

As inserções *Alu* são sequências repetidas no genoma e os representantes mais fartos de uma família de retrotransposons chamada de SINE (*Short Interspersed Nuclear Element*), que possuem em sua estrutura um sítio de restrição para a enzima *AluI* (BATZER; DEININGER, 2002; HOUCK; RINEHART; SCHMID, 1979). Sua estrutura é composta por aproximadamente 300 pb, dividida em dois braços praticamente idênticos, onde o monômero direito tem cerca de 31 bases a mais que o esquerdo, resultando em torno de 150 bases, separados por uma região rica em adeninas (A) (STRACHAN; READ, 1999). O motivo rico em “A” presente no centro da estrutura geralmente possui a repetição A₅TACA₆ (BATZER; DEININGER, 2002). Além disso, esses autores comentam que as sequências *Alu* possuem uma cauda terminal oligo-(dA) de tamanho variável. As inserções *Alu* também têm repetições curtas de timinas e adeninas (T + A) flanqueando a sequência, refletindo sua inserção cromossomal

(JURKA, 1997). Como parte da sequência, ainda, há a presença de um promotor para a enzima RNA polimerase III (BATZER; DEININGER, 2002), necessária à sua transcrição. Um grande número de inserções Alu estão localizadas em diferentes cromossomos e foram utilizadas em estudos populacionais, sendo estimadas contribuições das mesmas em cerca de 1% das doenças genéticas humanas (DUNN et al., 2002). Esses marcadores são normalmente encontrados em íntrons, regiões não traduzidas dos genes e regiões genômicas intergênicas, chegando a mais de um milhão de cópias no genoma, representando 10% do total, de forma não uniforme (BATZER; DEININGER, 2002).

Esses elementos móveis são derivados, ancestralmente, do RNA 7SL, envolvido na ligação com o peptídeo sinal, que sinaliza o transporte de proteínas ao retículo endoplasmático (TIAN et al., 2008). O seu mecanismo de transposição envolve a retrotranscrição de um produto originado a partir da ação da enzima RNA polimerase III, provavelmente pela ação de uma transcriptase reversa codificada por uma sequência do tipo LINE (*Long interspersed Nuclear Elements*) (MATHIAS et al., 1991), já que os elementos *Alu* perderam, ao longo do processo evolutivo, a sequência que codifica essa enzima (BATZER; DEININGER, 2002). Esses autores comentam, ainda, a importância de regiões conservadas (TTAAA) flanqueando a sequência *Alu* necessárias ao reconhecimento por endonucleases para que ocorra o processo de integração ao genoma.

Esses elementos são classificados em subfamílias, baseado em seu padrão de mutação, idade genética e diferenças na sequência, acumuladas durante a evolução. Entre as sequências presentes na espécie humana existem aquelas mais antigas, compartilhadas com outros primatas, chamadas de *AluJ* e *AluS* (JURKA; SMITH, 1988) e algumas sequências específicas da espécie humana, notadamente as de inserções mais recentes, chamadas *AluY* (BATZER et al., 1996). Para a diferenciação das famílias de inserções *Alu* há consenso em utilizarem-se diferenças nas sequências, como algumas mutações diagnósticas (JURKA; SMITH, 1988; BATZER et al., 1996; KAPITONOV; JURKA, 1996).

Dentre as famílias *Alu* descritas acima, a *AluY* mostra-se a mais útil no estudo populacional, devido ao fato de ainda não terem alcançado sua completa fixação no genoma (KULSKI et al., 2001), sendo dimórficas entre a população humana, ou seja, ou estão presentes ou ausentes (DUNN et al., 2002), além de sua característica de terem identidade por descendência (DEININGER; BATZER, 1999). Ainda, sua determinação é um procedimento rápido e fácil através da reação em cadeia da polimerase (PCR) e posterior eletroforese (ROY-ENGEL et al., 2001). A presença ou ausência dessas inserções *Alu* resultam em marcadores de DNA que apresentam vantagens em relação aos SNPs e outros marcadores: (1) a probabilidade

é quase nula de duas retroposições idênticas acontecerem no mesmo local cromossômico; (2) a condição ancestral de cada inserção *Alu* polimórfica é conhecida, i.e., a ausência de inserção é a condição ancestral, possibilitando o estabelecimento preciso das origens dos diferentes alelos (TIAN et al., 2008), pois não há mecanismos de retirada dessas inserções *Alu* do genoma. Portanto, dois indivíduos possuindo a mesma inserção no mesmo *locus* possuem um ancestral comum com a mesma inserção.

Foram identificadas pelo menos cinco diferentes sequências de *AluY* na região de classe I do MHC humano, sendo elas: *AluYMICB*, *AluYTF*, *AluYHJ*, *AluyHF* e *AluyHG* (DUNN et al., 2002). A *AluYMICB* é encontrada no primeiro íntron do gene *MICB* (KULSKI et al., 2002). *AluYTF* está localizada entre os genes *HLA-B* e *HLA-E*, a cerca de 505 kb em relação ao último (DUNN et al., 2003). A *AluyHG* está localizada a aproximadamente 20 kb da ponta 3' do gene *HLA-G* (KULSKI et al., 2001) e foi recentemente explorada em um estudo do grupo (SANTOS et al., 2013). A *AluYHJ* está localizada a aproximadamente 190 kb em relação ao *HLA-G* (DUNN et al., 2002). A *AluyHF* se encontra a aproximadamente 130 kb telomérico em relação a *HLA-G* e próxima ao gene *HLA-F* (DUNN et al., 2002). Na Figura 2 observa-se o bloco alfa evidenciado, com cerca de 235 kb, onde estão presentes os genes *HLA-A*, *-G* e *-F* e as variantes estruturais, foco principal do presente trabalho.

Inicialmente, a diversidade genética e os padrões de associação entre *AluyHF*, *AluyHG* e *AluYHJ* com *HLA-A* foram relatados nas populações australiana e japonesa (DUNN et al., 2002). Posteriormente, estudos de associação *Alu* com *HLA-A* e *HLA-B* foram realizadas em uma população do nordeste tailandês (DUNN et al., 2005), em um estudo na Malásia (DUNN et al., 2007) e outro estudo relacionou cinco marcadores *Alu* com o MHC de classe I (KULSKI; DUNN, 2005).

Embora estudos prévios tenham sido realizados, os dados disponíveis no momento desses estudos eram de baixa resolução, i.e., apenas alguns éxons de cada gene HLA foram explorados, em geral buscando variantes já conhecidas. Esses marcadores são bastante informativos e contribuem para o entendimento da história evolutiva dos genes que os circundam. Ainda, estes marcadores podem indicar a presença de haplótipos específicos, alguns associados com fenótipos específicos e susceptibilidade a doenças. Por exemplo, em um estudo prévio do grupo (SANTOS et al., 2013), a presença do marcador *AluyHG* acompanhava um haplótipo de *HLA-G* relacionado com alta produção dessa molécula regulatória.

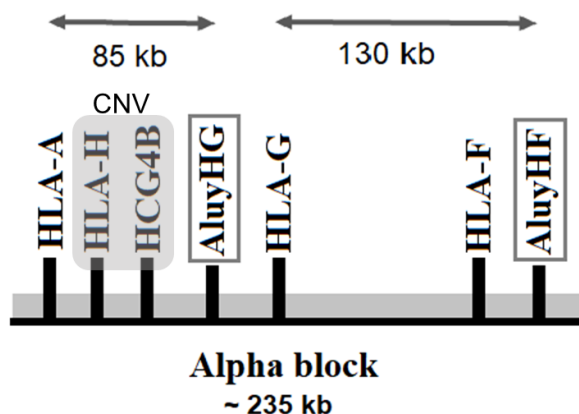


Figura 2. Composição gênica do bloco alfa do MHC humano e as variantes estruturais presentes nele.

O conhecimento do perfil de variabilidade das inserções *AluY* ao longo do MHC humano pode refinar a determinação de haplótipos estendidos deste complexo gênico, possibilitando, por exemplo, determinar quais alelos e haplótipos estão associados com algum fenótipo importante. Além disso, a correlação entre estes marcadores e polimorfismos detectados nos genes de classe 1 pode ajudar na diferenciação dos fatores intrínsecos, como recombinações, e extrínsecos, como a seleção adaptativa, na variabilidade observada no MHC humano, determinando quais alelos são mais antigos ou possuem uma origem comum. Ainda, por meio da determinação de haplótipos estendidos, podemos estimar quais alelos HLA acompanham a presença desses marcadores *Alu*. A análise de inserções *Alu* consiste em um método rápido e barato, que pode indiretamente indicar a presença de alelos HLA relevantes.

Outra variante estrutural no bloco alfa do MHC é a CNV (*copy number variation*) entre o *HLA-G* e o *HLA-A* (Figura 2), excluindo mais de 60Kb em uma região de 5Kb à montante do *HLA-A*. Essa CNV (esv3608493) inclui os genes *HLA-H* e *HCG4B*, como visto na Figura 2, que podem estar presentes ou não em alguns indivíduos. Essa CNV tem sido associada com a menor expressão de *HCG4B* e *HLA-A* e tem influência na doença pulmonar obstrutiva crônica e na asma (CHEN et al., 2017; OLIVEIRA et al., 2018b; PNG et al., 2019).

1.6. Forças seletivas

A seleção balanceadora já foi bem documentada para os genes *HLA-A* e *HLA-G*, entretanto ainda existem questões sobre a ação de seleção purificadora na região codificante de *HLA-G* e a um possível efeito carona entre esses dois genes, onde as forças evolutivas atuando em um deles poderiam influenciar as frequências alélicas do outro. Neste estudo, como parte de um projeto maior que visa caracterizar todos genes e variantes estruturais na região do MHC humano e a variabilidade encontrada em cada gene, avaliamos a diversidade genética observada

no bloco alfa do MHC humano e como esta diversidade se correlaciona com as variantes estruturais presentes nele.

1.7. Referências

AMIGORENA, S. Antigen presentation: from cell biology to physiology. **Immunological Reviews**, v. 272, n. 1, p. 5-7, Jul 2016.

BATZER, M. A.; DEININGER, P. L. Alu Repeats and Human Genomic Diversity. **Nature Reviews Genetics**, v. 3, n. 5, p. 370–379, 2002.

BATZER, M. A. et al. Standardized nomenclature for Alu repeats. **Journal of Molecular Evolution**, v. 42, n. 1, p. 3-6, Jan 1996.

BRUTKIEWICZ, R. R. Cell Signaling Pathways That Regulate Antigen Presentation. **Journal of Immunology**, v. 197, n. 8, p. 2971-2979, Oct 2016.

CAROSELLA, E. D.; MOREAU, P.; LEMAOULT, J.; ROUAS-FREISS, N. *HLA-G*: from biology to clinical benefits. **Trends in Immunology**, v. 29, n. 3, p. 125–132, Mar 2008.

CASTELLI, E. C. et al. A comprehensive study of polymorphic sites along the *HLA-G* gene: implication for gene regulation and evolution. **Molecular Biology and Evolution**, v. 28, n.11, p. 3069-3086, Nov 2011.

CASTELLI, E. C. et al. *HLA-E* coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. **Human Immunology**, v. 76, n. 12, p. 945-53, Dec 2010.

CASTELLI, E. C. et al. *HLA-G* alleles and *HLA-G* 14 bp polymorphisms in a Brazilian population. **Tissue Antigens**, v. 70, n. 1, p. 62-68, Jul 2007.

CASTELLI, E. C. et al. *HLA-G* polymorphism and transitional cell carcinoma of the bladder in a Brazilian population. **Tissue Antigens**, v. 72, n. 2, p. 149-157, Aug 2008.

CASTELLI, E. C. et al. *HLA-G* variability and haplotypes detected by massively parallel sequencing procedures in the geographically distinct population samples of Brazil and Cyprus. **Molecular Immunology** v. 83, p. 110-126, Mar 2017.

CASTELLI, E. C. et al. Insights into *HLA-G* Genetics Provided by Worldwide Haplotype Diversity. **Frontiers in Immunology**, v. 5, Oct 2014.

CASTELLI, E. C. et al. The genetic structure of 3' untranslated region of the *HLA-G* gene: polymorphisms and haplotypes. **Genes and Immunity**, v. 11, n. 2, p. 134-141, Mar 2010.

DEININGER, P. L.; BATZER, M. A. *Alu* repeats and human disease. **Molecular Genetics and Metabolism**, v. 67, n. 3, p. 183–193, Jul 1999.

DONADI, E. A., et al. Implications of the polymorphism of *HLA-G* on its function, regulation, evolution and disease association. **Cellular and Molecular Life Sciences**, v. 68, n. 3, p. 369-395, Feb 2011.

DUNKER, K.; SCHLAF G.; BUKUR J. et al. Expression and regulation of non-classical *HLA-G* in renal cell carcinoma. **Tissue Antigens**, v. 72, n. 2, p. 137-148, Aug 2008.

DUNN, D. S.; CHOY, M. K.; PHIPPS, M. E.; KULSKI, J. K. The distribution of major histocompatibility complex class I polymorphic *Alu* insertions and their associations with HLA alleles in a Chinese population from Malaysia. **Tissue Antigens**. v. 70, n. 2, p. 136–143, Aug 2007.

DUNN, D. S. et al. Dimorphic *Alu* element located between the *TFIIH* and *CDSN* genes within the major histocompatibility complex. **Electrophoresis**. v. 24, n. 16, p. 2740-2748, Aug 2003.

DUNN, D. S. et al. Polymorphic *Alu* insertions and their associations with MHC class I alleles and haplotypes in the northeastern Thais. **Annals of Human Genetics**, v. 69, pt. 4, p. 364–72, Jul 2005.

DUNN, D. S. et al. The association between *HLA-A* alleles and young *Alu* dimorphisms near the *HLA-J*, *-H*, and *-F* genes in workshop cell lines and Japanese and Australian populations. **Journal of Molecular Evolution**, v. 55, n. 6, p. 718-726, Dec 2002.

FELICIO, L. P. et al. Worldwide *HLA-E* nucleotide and haplotype variability reveals a conserved gene for coding and 3' untranslated regions. **Tissue Antigens**. v. 83, n. 2, p. 82-93, Feb 2014.

GERAGHTY, D. E. et al. The *HLA* class I gene family includes at least six genes and twelve pseudogenes and gene fragments. **Journal of Immunology**, v. 149, n. 6, p. 1934-1946, Sep 1992.

GOLDBERG, A. C.; RIZZO, L. V. MHC structure and function – antigen presentation. Part 1. **Einstein (São Paulo)**. v. 13, n. 1, p. 103-106, Mar 2010.

HOUCK, C. M.; RINEHART, F. P.; SCHMIDT, C. W. A ubiquitous family of repeated DNA sequences in the human genome. **Journal of Molecular Biology**, v. 132, n. 3, p. 289-306, Aug 1979.

JENSEN, P. E. Recent advances in antigen processing and presentation. **Nature Immunology**, v. 8, n. 10, p. 1041-8, Oct 2007.

JURKA, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 5, p. 1872-1877, Mar 1997.

JURKA, J.; SMITH, T. A fundamental division in the *Alu* family of repeated sequences. **Proceedings of the National Academy of Sciences of the United States of America**, v. 85, n. 13, p. 4775-4778, Jul 1988.

KAPITONOV, V.; JURKA, J. The age of *Alu* subfamilies. **Journal of Molecular Evolution**, v. 42, n.1, p. 59-65. Jan 1996.

KLEIN, J.; SATO, A. Birth of the major histocompatibility complex. **Scandinavian Journal of Immunology**, v. 47, n. 3, p. 199-209, Mar 1998.

KLEIN, J.; SATO, A. The HLA system. First of two parts. **New England Journal of Medicine**, v. 343, n. 10, p. 702-709, Sep 2000.

KLEIN, J.; SATO, A. The HLA system. Second of two parts. **New England Journal of Medicine**, v. 343, n. 11, p. 782-786, Sep 2000.

KULSKI, J. K. et al. The association between *HLA-A* alleles and an Alu dimorphism near *HLA-G*. **Journal of Molecular Evolution**, v. 53, n. 2, p. 114-123, Aug 2001.

KULSKI, J. K.; D. S. DUNN. Polymorphic Alu insertions within the Major Histocompatibility Complex class I genomic region: a brief review. **Cytogenetic and Genome Research**, v. 110, n. 1-4, p. 193-202, Jul 2005.

KULSKI, J. K. et al. Alu polymorphism within the MICB gene and association with HLA-B alleles. **Immunogenetics**, v. 53, n. 10-11, p. 975-979, Feb 2002. Epub 26 Jan 2002.

LIMA, T. H. et al. HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. **Human Immunology**, v. 77, n. 10, p. 841-853, Oct 2016.

MATHIAS, S. L. et al. Reverse transcriptase encoded by a human transposable element. **Science**, v. 254, n. 5039, p. 1808-1810, Dec 1991.

OGAHARA, S. et al. Effect of mismatched combinations of *HLA-A* antigens on graft survival in the transplanted kidney. **Transplantation Proceedings**, v. 30, n. 7, p. 3500-1, Nov 1998.

RAJU, T. N. The Nobel chronicles. 1980: George Davis Snell (1903-96); Jean Baptiste Dausset (b 1916); Baruj Benacerraf (b 1920). **Lancet**, v. 354, n. 9191, p. 1738, Nov 1999.

ROY-ENGEL, A. M. et al. Alu insertion polymorphisms for the study of human genomic diversity. **Genetics**, v. 109, n. 1, p. 279-290, Sep 2001.

SANTOS, K. E. et al. Insights on the *HLA-G* evolutionary history provided by a nearby Alu Insertion. **Molecular Biology and Evolution**, v. 30, n. 11, p. 2423-2434, Nov 2013.

SHIINA, T. et al. The HLA genomic loci map: expression, interaction, diversity and disease. **Journal of Human Genetics**, v. 54, n. 1, p. 10-39, Jan 2009.

STRACHAN, T.; READ, A. P. **Human molecular genetics 2 / Tom Strachan and Andrew P. Read**. 2. ed. New York : Wiley-Liss, 1999.

TERASAKI, P. I. A brief history of HLA. **Immunol Res**, v. 38, n. 1-3, p. 139-148, 2007.

THORSBY, E. A short history of HLA. **Tissue Antigens**, v. 74, n. 2, p. 101-16, Aug 2009.

TIAN, W. et al. Polymorphic insertions in 5 Alu loci within the major histocompatibility complex class I region and their linkage disequilibria with HLA alleles in four distinct populations in mainland China. **Tissue Antigens**, v. 72, n. 6, p. 559-567, Dec 2008.

VANDIEDONCK, C. et al. Association of *HLA-A* in autoimmune myasthenia gravis with thymoma. **Journal of Neuroimmunology**, v. 210, n. 1-2, p. 120-3, May 2009.

CAPÍTULO 2

2. ARTIGO

GENETIC DIVERSITY AND STRUCTURAL VARIANTS IN THE HUMAN MHC ALPHA-BLOCK

Marilia Rodrigues Silva Passos¹, Thálitta H. A. Lima², Andreia S. Souza², Nayane dos Santos B. Silva¹, Heloisa S. Andrade², Camila Ferreira B. Castro³, Michel S. Naslavsky⁴, Marília Scliar⁴, Eduardo Antônio Donadi⁵, Nicolas Vince⁶, Diogo Meyer⁷, Celso T. Mendes-Junior⁸, Erick C. Castelli¹

¹ Universidade Estadual Paulista (UNESP), Faculdade de Medicina de Botucatu, Brasil.

² Universidade Estadual Paulista (UNESP), Instituto de Biociências de Botucatu, Brasil.

³ Centro Universitário Sudoeste Paulista – UniFSP, Brasil.

⁴ Centro de Pesquisas sobre o Genoma Humano e Células Tronco, Instituto de Biociências, Universidade de São Paulo (USP), Brasil.

⁵ School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, State of São Paulo, Brazil.

⁶ Nantes Université, Centrale Nantes, CHU Nantes, Inserm, Centre de Recherche en Transplantation et Immunologie, UMR 1064, ITUN, F-44000 Nantes, France.

⁷ Faculdade de Saúde Pública, Escola de Enfermagem, Universidade de São Paulo (USP), Brasil.

⁸ Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brasil.

Contact: Erick C. Castelli

Departamento de Patologia, Faculdade de Medicina Unesp, Botucatu – SP, CEP 18618970, Brazil

erick.castelli@unesp.br, +55 14 3880 1696

Abstract

The MHC (Major Histocompatibility Complex) alpha block comprehends important HLA genes such as *HLA-F* and *HLA-G*, both non-classical HLA loci with low variability and restricted tissue expression, and *HLA-A*, a highly polymorphic gene important for antigen presentation. This segment also presents some structural variants such as *Alu* elements and CNVs. *HLA-G* and *HLA-A* have been described as targets for balancing selection in different scenarios, the coding region for *HLA-A* and the regulatory segments for *HLA-G*. Moreover, *HLA-F* diversity in worldwide populations is unknown. Here we applied a bioinformatics pipeline that optimizes short reads mapping for HLA genes and allows the determination of reliable haplotypes, to survey the MHC alpha block genetic diversity in 1323 Brazilians. We also evaluated three structural variants, *AluyHF*, *AluyHG*, and esv3608493. These individuals configure a highly admixed sample from the largest Brazilian city, São Paulo. Our data reviewed the presence of a few *HLA-F* and *HLA-G* sequences in spite of the large sample size. *HLA-F* is the most conserved one, with 82.65% of the chromosomes encoding the same *HLA-F* protein, and with only one additional frequent protein. *HLA-A* was highly polymorphic with 96 genomic sequences encoding 59 proteins, including rare ones. We also assessed linkage disequilibrium and haplotypes along the MHC alpha block. *AluyHF* was in LD only with *HLA-F* but not with any particular allele. *AluyHG* was in LD with G*01:01:01:01/UTR-1, while esv3608493 was associated only with alleles *HLA-A**23 and *A**24. We also detected frequent haplotypes in our sample, especially the close relationship between *HLA-G* and *HLA-A* alleles. Because of that, the evidence of balancing selection in both these genes might not be independent since high heterozygosis in one locus would increase the heterozygosis in the other. This data provides a valuable resource of variants, frequencies, and haplotypes in HLA genes of admixed samples for clinical and evolutive studies.

2.1. Introduction

The MHC (Major Histocompatibility Complex) alpha block comprehends three HLA genes important for antigen presentation and immune modulation (Figure 1). At the telomeric end, we find the *HLA-F* gene, a non-classical HLA gene with low variability and low expression levels in different tissues (LEPIN et al., 2000; LIMA et al., 2016). The IPD-IMGT/HLA database version 3.38.0 describes only 44 alleles for *HLA-F*, encoding 6 different proteins. *HLA-F* may interact with inhibitory receptors such as *ILT2*, *ILT4*, *KIR3DL2* and *KIR2DS4*, which may modulate NK and T cell activity (GOODRIDGE et al., 2013; LEPIN et al., 2000).

HLA-F expression has been detected in physiological situations in the thymus, spleen, and amygdala (LEPIN et al., 2000; WAINWRIGHT; BIRO; HOLMES, 2000), and pathological in various types of cancer and autoimmune diseases usually associated with a bad prognosis (ISHIGAMI et al., 2015; JUCAUD et al., 2016). *HLA-F* genetic diversity using molecular techniques able to detect any variant has been explored mainly in Brazil and Benin (LIMA et al., 2016; SONON et al., 2018) and, because of that, its genetic diversity might be underestimated.

In the center of the alpha block (Figure 1) we find the *HLA-G* locus, another non-classical HLA gene important for immune modulation. *HLA-G* was first identified on the maternal-fetal interface, especially on the placenta's cytotrophoblast, where immunotolerance contributes to the fetus maintenance. *HLA-G* is constitutively expressed on the trophoblast cells surface, cornea, pancreas, and thymus (KING et al., 2000). In endocrine tissues, *HLA-G* is regulated according to growth and inflammatory stimulus and can exhibit regulatory functions in human pancreatic islets, including autoimmune progression (CIRULLI et al., 2006). *HLA-G* can inhibit Natural Killer cells (NK) and T CD8⁺ lymphocytes cytotoxicity, by the interaction with NK and T cell surface receptors (CAROSELLA et al., 2008). *HLA-G* binding to these receptors modulates T cytotoxic and NK cells activity, inducing the expression of inhibitory receptors with tyrosine motifs (CAROSELLA et al., 2003).

HLA-G is conserved in the nucleotide and protein levels, with 69 alleles encoding 19 full-length proteins according to the IPD-IMGT/HLA database version 3.38.0. However, when studying different population samples few different *HLA-G* proteins are detected worldwide, and the majority of the *HLA-G* alleles present intronic or synonymous mutations, and they are present in all population samples with different frequencies (CASTELLI et al., 2014, 2017; SONON et al., 2018). The *HLA-G* promoter and 3' untranslated region (3'UTR) enclose many transcriptional (MOREAU; FLAJOLLET; CAROSELLA, 2009) and post-transcriptional regulatory elements (MANASTER et al., 2012; PORTO et al., 2015; TAN et al., 2007; YIE et al., 2008), respectively. Therefore, some *HLA-G* polymorphic sites in regulatory sequences have been associated with different expression profiles (HVIID, 2006; MOREAU; FLAJOLLET; CAROSELLA, 2009; YIE et al., 2008).

HLA-G deserves especial attention in pregnancy, as the molecules codified by these genes play a role in immunosuppression and homeostasis maintenance during pregnancy (HVIID, 2006). The non-expression of classical class I HLA molecules by the trophoblastic cells make then potential targets for the lysis mediated by NK cells. However, the interaction between the trophoblast *HLA-G* and NK inhibitory receptors hampers the trophoblast cell lysis

(PONTE et al., 1999). Also, *HLA-G* activates cytokines production that promotes vascular remodeling at the maternal-fetal interface (KIRWAN et al., 2002).

In the centromeric end of the MHC alpha block, we find the *HLA-A* locus. *HLA-A* is a classical class I gene important for antigen presentation. It is one of the most polymorphic genes in the human genome, presenting 5.735 different alleles according to the IPD-IMGT/HLA database version 3.38.0. The Brazilian population has been studied regarding the *HLA-A* genetic diversity, and the *HLA-A* promoter sequence is highly variable, probably related to different expression profiles for each of the main *HLA-A* allele groups (LIMA et al., 2019). The most common *HLA-A* variant is *HLA-A*02*, which accounts for more than 920 alleles already described (IPD-IMGT/HLA), with some of them related to diseases such as Alzheimer, Hashimoto thyroiditis, and type 1 diabetes mellitus (NOBLE et al., 2010; SHIINA; INOKO; KULSKI, 2004). Another frequent *HLA-A* allele group is *HLA-A*24*, with some alleles (*A*24:02*), together with alleles of other groups (*A*11:01*, *A*32:01*, and *A*66:01*) associated with type 1 diabetes mellitus (DM-1) (CARVALHO DOS SANTOS et al., 2013; LIMA et al., 2019). It is noteworthy that these alleles are relatively frequent in Brazil (CARVALHO DOS SANTOS et al., 2013; LIMA et al., 2019). There is a functional link between the *HLA-A* high nucleotide and protein diversity and the antigen presentation compatibility since different *HLA-A* molecules would present different subsets of peptides (BUHLER; NUNES; SANCHEZ-MAZAS, 2016).

As the antigen presentation is the main function of the HLA classical class I molecules such as *HLA-A*, one can consider that the natural selection has shaped the genetic variability the HLA classical genes, increasing variability at the peptide binding site, thus allowing a great diversity of antigen presentation. Likewise, since the expression of classical class I molecules should be constitutive in virtually all somatic cells, and that the expression of class I molecules occurs in a codominant fashion (FISCHER; MAYR, 2001; KLEIN; SATO, 2000a, 2000b), it is expected that the regulatory regions of these genes might be less variable, avoiding too many gene expression variations. However, variations in the regulatory regions may modify this scenario. It is believed that polymorphisms in these regions could substantially affect the expression levels, leading to an unbalanced expression among the different alleles of an individual.

Balancing selection has been well documented for both *HLA-A* and *HLA-G*, but at different fashions. While the most polymorphic segment at *HLA-A* is the coding sequence, which is associated with the presence of many frequent and divergent coding alleles in worldwide populations, the most polymorphic at *HLA-G* are the regulatory segments. In fact,

many studies have detected signatures of balancing selection at the *HLA-G* promoter and 3'UTR, while the coding segment is conserved (GINEAU et al., 2015). However, it is not clear whether balancing selection is indeed operating at the *HLA-G* regulatory segments, or these results are a hitchhiking effect caused by the selective pressures acting on *HLA-A*.

In the MHC alpha block there are some structural variants, such as two Alu insertions, *AluyHF* and *AluyHG*. Alu insertions are repeated sequences with approximately 300 bp distributed across the genome. Its structure is divided into two nearly identical arms separated by an adenine-rich region (A), a variable length oligo- (dA) tail, short thymine, and adenine (T + A) repeats flanking the sequence and a promoter for RNA polymerase III enzyme, required for its transcription. These markers are usually found in introns, untranslated gene regions, and intergenic regions (as it is in this case), reaching more than 10% of the genome, nonuniformly. Among the Alu families, *Aluy* is the most useful for population genetics studies because of its identity by descent and because it has not yet reached complete fixation in the genome, being dimorphic in the population. The presence or absence of these insertions results in DNA markers that have advantages over SNPs and other markers: (1) the probability is almost zero of two identical elements occurring at the same chromosomal site; (2) have a known ancestral condition. These allowed the exact establishment of the origins of the different alleles since there are no mechanisms for removing these Alu insertions from the genome. Therefore, two individuals with the same insertion in one place and may have the same common ancestor (BATZER; DEININGER, 2002; DUNN et al., 2002; JURKA, 1997; KULSKI et al., 2001; MATHIAS et al., 1991). Previous studies report associations between *AluyHF* and *AluyHG* with *HLA-A* in the Australian and Japanese populations (DUNN et al., 2002) and *AluyHG* following the presence of a *HLA-G* haplotype related to high HLA-G production, G*01:01:01:01/UTR-01 (SANTOS et al., 2013).

Another structural variant in the MHC alpha block is copy number variation (CNV) between the *HLA-G* and *HLA-A* (Figure 1), deleting more than 60Kb in a region 5Kb upstream *HLA-A*. This CNV (esv3608493, nsv823469, nsv823470) includes the genes *HLA-H* and *HCG4B*, which may be present or not in some individuals. This CNV loss has been associated with lower *HCG4B* and *HLA-A* expression and influences chronic obstructive pulmonary disease and asthma (CHEN et al., 2017; OLIVEIRA et al., 2018b; PNG et al., 2019).

Here we evaluate the *HLA-F*, *-G* and *-A* genetic diversity, and also the presence or absence of the *AluyHF*, *AluyHG*, and esv3608493, in 1323 individuals living in the largest Brazilian city, São Paulo, by using Whole Genome Sequencing (WGS) and a bioinformatics workflow developed to optimize short-read alignment for HLA genes. Then, we evaluate

Linkage disequilibrium along the MHC alpha-block and relationship among genes and structural variants, defining the most common haplotypes in our highly admixed population.

2.2. Methods

2.2.1. Samples and sequencing

We evaluated polymorphisms in the MHC alpha-block in 1323 DNA samples from individuals from São Paulo (city). These individuals are part of the SABE project (Saúde, Bem Estar e Envelhecimento – Health, Well-Being, and Aging), led by the *Centro de Estudos do Genoma Humano e Células-Tronco* (CEGH-CEL), University of São Paulo (IB/USP).

Peripheral blood was drawn into ethylene diaminetetracetic acid (EDTA) tubes and DNA was extracted from leukocytes using the automatic extraction robot QIAAsymphony (Qiagen, Hilden, Alemanha) automated purification system with the lysis/isopropanol precipitation protocol as recommended by the manufacturer. The quality and concentration of DNA was evaluated with NanoDrop® (Thermo Fisher Scientific, Carlsbad, Califórnia, EUA).

Sequencing was performed by the CEGH-CEL using sonication (Covaris, Inc., Woburn, Massachusetts, USA) and Illumina HiSeq (Illumina, San Diego, Califórnia, EUA). Then, reads were aligned using Isaac Aligner against the human reference hg38.

2.2.2. MHC alpha-block analysis

We used *samtools* to extract reads mapped to the MHC segment and unmapped ones and convert them to paired FASTQ. We resynchronizing the FASTQ files with local Perl script. We used *hla-mapper* version 3.05 to produce *HLA-F*, *HLA-G*, and *HLA-A* specific aligned files (BAM format) to the reference genome version hg38 (CASTELLI et al., 2018). This software was previously introduced evaluating the *HLA-A*, *HLA-G*, *HLA-E*, and *HLA-F* in different population samples (BUTTURA et al., 2019; CASTELLI et al., 2015, 2017; LIMA et al., 2016, 2019; RAMALHO et al., 2017; SONON et al., 2018).

The Genome Analysis Toolkit HaplotypeCaller (GATK, version 4.1.2) in the GVCF mode and GATK GenotypeGVCFs was used to infer genotypes with default parameters (DEPRISTO et al., 2011; MCKENNA et al., 2010; VAN DER AUWERA et al., 2013). Then, variant refinement was performed using *vcfx* (available at www.castelli-lab.net/apps/vcfx). This software introduces missing alleles in genotypes with low likelihoods (function *checkpl*) and unbalanced genotypes (function *checkad*), and also annotates each variant with some quality control parameters (function *evidence*). To ensure that only high-quality genotypes are

passed forward to the next imputation and haplotyping steps, all variants not indicated as “PASS” by vcfx evidence were manually inspected and, if necessary, removed. Since hla-mapper version 3 reports the depth of coverage in each HLA locus, we used the *HLA-H* depth of coverage to genotype esv3608493 as presenting none, one, or two copies of this segment. A local Perl script was used to genotype the *Alu**YHG* and *Alu**YHG* in all samples. This script considers the distance observed between known sequences surrounding the insertion site, and calculates whether the samples present none, one, or two copies of the Alu insertion.

We inferred haplotypes combining two computational strategies. First, GATK ReadBackedPhasing (RBP) inferred the relationship between close heterozygous variants, with the minimal PhaseQualityThreshold set to 500 (25x the default value). This procedure resulted in phasing sets, i.e., blocks of known phases but unphased among them. Multi-allelic variants, indels, and missing alleles are not considered by RBP. Then, the second step consisted of inferring the final haplotypes considering these phase sets. For that, we used an in-house software named PHASEX (available upon request). PHASEX uses Shapeit4 (DELANEAU; MARCHINI; ZAGURY, 2011) to phase all bi-allelic variants considering the phase sets, in 50 independent runs, comparing the results afterward. Samples presenting the same pair of haplotypes in at least 48 runs ($P > 0.95$) are fixed as completely phased samples and passed forward to the next round. This procedure is repeated until the number of samples with $P > 0.95$ no longer increases. Then, the haplotypes defined in the previous step are passed forward to the next, in which Beagle 4.1 is used to infer the final set of haplotypes, now including the multi-allelic variants. Here we also used 50 independent runs for each round, fixing the samples in which the same haplotypes are inferred in at least 48 independent runs ($P > 0.95$). The final set of haplotypes are the ones with $P > 0.95$ after the final run. Shapeit4 and Beagle also imputed the bi-allelic and multi-allelic missing alleles, respectively. We have removed all singletons before the haplotyping procedure. This step is necessary because singletons are ambiguous by definition. When possible, singletons were manually inserted in the final VCF file comparing the RBP results and the BAM files. The procedure described above was performed for *HLA-F*, *HLA-G*, and *HLA-A*, separately. Then, a joined VCF file was produced with the *Alu**YHF*, *HLA-F*, *HLA-G*, *Alu**YHG*, the *HLA-H* CNV, and *HLA-A* data, and another haplotyping round was performed. This procedure allowed us to identify the extended haplotypes of 1306 individuals (98.7% of the original sample).

The final VCF file (phased) was converted into two complete sequences for each locus and each individual, using the hg38 chromosome 6 as a draft and replacing the correct nucleotide in each position by using vcfx fasta. Then, a local Perl script identified the closest

known HLA allele for each sequence, considering the IPD-IMGT/HLA database (ROBINSON et al., 2015), version 3.38.0. In cases in which the sequence observed was identical to one known sequence, the same name was used. In cases in which the sequence observed was not identical to any one already reported, we classified these sequences as “unknown”. Linkage disequilibrium (LD) was assessed using Haploview 4.2 and variants with minor allele frequency (MAF) greater than 1%.

2.3. Results

***HLA-F* genetic diversity among Brazilians**

We detected 16 different *HLA-F* CDS sequences in this sample (Table 1). The list of variants in exons, introns, and regulatory segments, and their frequencies, is available at table S1. The most frequent CDS allele was F*01:01:01 (64.81%) and F*01:01:02 (17.64%), all encoding the same *HLA-F* protein F*01:01. The *HLA-F* molecule F*01:01 has a summed frequency of 82.65%, and F*01:03 has a frequency of 14.93%. Thus, in this large admixed Brazilian sample, there are only two frequent *HLA-F* molecules, indicating that *HLA-F* is indeed conserved in the DNA and protein level. There were nine *HLA-F* CDS sequences detected in our sample, with a summed frequency of 1.58%. One of these new sequences was detected in 22 individuals. When the genomic sequence is taken into account (exons + introns), we found only 19 alleles compatible with known sequences as described in Table 2, with a summed frequency of 92.78%. The new genomic sequences are mostly related to rare intronic variants and number of repeats in STRs. Although not shown in this study, the *HLA-F* promoter and 3'UTR haplotypes, and their association with coding alleles, are compatible with the ones previously described for other Brazilian and African samples (BUTTURA et al., 2019; LIMA et al., 2016; SONON et al., 2018).

***HLA-G* genetic diversity among Brazilians**

We detected 27 different *HLA-G* CDS sequences in this sample (Table 3), 20 of them compatible with known *HLA-G* sequences and with summed frequency of 99.51%. The list of variants in exons, introns, and regulatory segments, and their frequencies, is available at table S2. The most frequent CDS allele was G*01:01:01 (42.75%) and G*01:01:02 (17.18%). Converting the alleles at Table 3 in protein sequences, there were 9 variants. The most frequent was G*01:01 (68.09%), followed by G*01:04 (15.94%), G*01:03 (8.04%), G*01:06 (5.02%),

G*01:05N (truncated protein, 2.11%), and other four rare molecules (G*01:08, G*01:09, G*01:11, and G*01:21N). In terms of protein and nucleotide diversity, *HLA-G* is more polymorphic than *HLA-F*.

When the genomic sequence is taken into account (exons + introns), we found only 27 alleles compatible with known sequences as described in Table 4, with a summed frequency of 96.71%. The new genomic sequences are mostly related to rare intronic variants, but some are even frequent than the alleles reported in Table 4. The most frequent *HLA-G* genomic sequence was G*01:01:01:01 (24.62%) and G*01:01:02:01 (16.16%). This large sample confirms the existence of rare *HLA-G* alleles already reported in the IMGT/HLA database, such as G*01:01:22:01, G*01:01:19, G*01:09, and G*01:21N. Although not shown in this study, the *HLA-G* promoter and 3'UTR haplotypes, and their association with coding alleles, are compatible with the ones previously described for other Brazilian and African samples (CASTELLI et al., 2014, 2017; SONON et al., 2018).

***HLA-A* genetic diversity among Brazilians**

HLA-A is one of the most polymorphic genes in the human genome, with thousands of alleles reported in the IPD-IMGT/HLA database. In this large Brazilian sample, we found 70 different CDS sequences and one new and rare one (Table 5). There were copies of each allele group, including rare ones such as A*80. The list of variants is described in Table S3. The most frequent CDS sequence was A*02:01:01, A*24:02:01, A*03:01:01, and A*01:01:01, but at least 19 of them present a frequency higher than 1%. In terms of different *HLA-A* proteins, this Brazilian sample encodes 59 different *HLA-A* proteins.

When the genomic sequence is taken into account (exons + introns), we found only 96 alleles compatible with known sequences as described in Table 6, with a summed frequency of 98.71%. The most frequent *HLA-A* genomic sequence was A*02:01:01:01 (21.08%), A*01:01:01:01 (8.61%), and G*24:02:01:01 (8.23%), but many alleles present high frequencies. Here we confirm the presence of uncommon alleles such as A*23:17:01:01 and A*80:01:01:02, all presenting many copies in our sample, and also the existence of very rare *HLA-A* alleles such as A*24:104, A*30:29, A*68:17:01. Although not shown in this study, the *HLA-A* promoter and 3'UTR haplotypes, and their association with coding alleles, are compatible with the ones previously described for another Brazilian sample (LIMA et al., 2019).

Linkage Disequilibrium across the MHC alpha block

We observed high Linkage Disequilibrium across the MHC alpha block considering 638 biallelic SNPs and 3 structural variants, *AluyHF*, *AluyHG*, and esv3608493. There were 3 large segregation blocks (Figure 2). The first one encompasses the Alu insertion *AluyHF* and most of the *HLA-F* variants, corresponding to a segment of 13Kb. The LD among *AluyHF* and non-*HLA-F* variants is low. The second block encompasses a few 3' *HLA-F* variants, all *HLA-G* SNPs up to the *AluyHG* and the CNV esv3608493, in a segment of 104 Kb. The transition between blocks coincides with these two structural variants. The third block, of 114 Kb, includes esv3608493 and the majority of the *HLA-A* variants. Because of this LD pattern, we opted to track the association between the *AluyHF* insertion and *HLA-F* alleles, the *AluyHG* insertion with *HLA-G* alleles, and esv3608493 with *HLA-A* allele. Besides, since there are many *HLA-G/HLA-A* pairs with high LD, we will also track the association between *HLA-G* and *HLA-A* alleles.

***AluyHF* and *HLA-F* haplotypes**

The frequency of the *AluyHF* insertion was 16.08%, mostly associated with the allele F*01:01:01:08 and F*01:01:01:18 (Table 7). Nevertheless, many copies of both F*01:01:01:08 and F*01:01:01:18 are associated with the absence of the *AluyHF* insertion. Although there is a high LD between this Alu element and *HLA-F* (Figure 2), this Alu element can be used to track specific *HLA-F* variants.

***AluyHG* and *HLA-G* haplotypes**

The frequency of the *AluyHG* insertion was 32.39%, mostly associated with the allele G*01:01:01:01 and other uncommon *HLA-G* alleles that were derived from G*01:01:01:01, such as G*01:01:01:08 (Table 8).

The CNV nsv823470 and *HLA-A* haplotypes

The loss of the 60Kb related to CNV esv3608493 has a frequency of 15.77% in our Brazilian sample. All alleles A*23 and A*24 described in Table 6 (and only them) are associated with the loss of this segment.

***HLA-G* and *HLA-A* haplotypes**

There were 159 *HLA-G/HLA-A* haplotypes, 60 of them occurring just once and 13 occurring twice. The most frequent haplotype was G*01:01:01:01 / A*02:01:01:01, but at least 21 haplotypes present a frequency higher than 1%. The full table of haplotypes is available as Table S4. Among the associations in which one *HLA-G* allele is mostly associated with the

same *HLA-A* allele or allele group, with minor deviations, we may list: G*01:01:01:04 with A*29/A*74, G*01:01:01:05 with A*03, G*01:01:01:06 with A*80, G*01:01:01:08 and A*30:02, G*01:01:01:02 with A*66, G*01:01:03 with A*11, G*01:01:15 with A*03:01:01:15, G*01:01:17 with A*32:01:01:01, A*01:04:04 with A*23, the null allele G*01:05N with A*30:01:01:01, G*01:06 with A*01:01:01:01, and others.

MHC alpha block common haplotypes

Here we evaluated the common haplotypes observed when considering the genomic sequences of *HLA-F*, *HLA-G*, and *HLA-A*. Since *HLA-F* and *HLA-A* are 200 Kb apart (Figure 1), with weak LD between them (Figure 2), we found 268 extended haplotypes, 120 of them occurring just once and 66 occurring twice. The most frequent haplotype was F*01:01:01:01 / G*01:01:01:01 / A*02:01:01:01, with a frequency of 10.29%. Table S5 described all the haplotypes and their frequencies.

2.4. Discussion

Here we evaluated the genetic diversity of the HLA alpha-block in highly admixed samples from Brazil, focusing on the *HLA-F*, *HLA-G*, and *HLA-A* loci, and three structural variants, *AluyHF*, *AluyHG*, and *esv3608493*. Since the HLA region is prone to mapping and genotyping errors because of NGS short reads (BRANDT et al., 2015; CASTELLI et al., 2018), we used the *hla-mapper* software to optimize read mapping as described elsewhere (CASTELLI et al., 2018; LIMA et al., 2019).

This is the largest survey of *HLA-F* genetic variability so far. *HLA-F* diversity is poorly explored in worldwide populations, with a smaller and different Brazilian sample and a sample from Benin evaluated using a similar methodology and describing many new and frequent *HLA-F* alleles (LIMA et al., 2016; SONON et al., 2018) that were subsequently confirmed by independent analysis (PAGANINI et al., 2019). This large admixed Brazilian sample confirms that *HLA-F* is indeed conserved, with 82.65% of the chromosomes encoding the same *HLA-F* protein, and with only one additional frequent protein (Table 1 and 2). Moreover, many alleles that were recently described, such as F*01:01:02:09, is quite frequent among Brazilians, reaching near 5%. It should be mentioned that many *HLA-F* alleles differ only by the number of TG repeats in a Short Tandem Repeat (STR) located in the 3'UTR region (LIMA et al., 2016; PAGANINI et al., 2019b). Because of that, the frequencies displayed in Table 2 might be biased towards specific allele sizes predicted by the GATK HaplotypeCaller. The high number of unknown *HLA-F* genomic sequences (Table 2) might be a consequence of this issue. Only

cloning and Sanger sequencing would confirm the presence of each of those alleles and their frequencies, which is not practical and cost-effective in this sample size. Nevertheless, this issue does not affect the results in Table 1, which demonstrates the frequency of each exonic sequence found in Brazil.

Since Brazilians are highly admixed, with major European, African, and Amerindian contributions, but also with recently Asian contributions in the last century, we may expect to find many different alleles in Brazil. However, this was not the case for *HLA-F*. Considering the exonic sequences we may find only three frequent ones (Table 1), two of them carrying synonymous mutations (F*01:01:01 and F*01:01:02). Recently, we described F*01:01:02 as a high-expressing allele because it carries a unique and highly divergent promoter sequence (BUTTURA et al., 2019), and the frequency of this allele in Brazil was 17.65%. Although we did not present data for the *HLA-F* regulatory sequences, we have included promoter and 3'UTR variants in this analysis (Table S1), and this sample presents the same pattern observed previously (LIMA et al., 2016; SONON et al., 2018). We also noticed that at least 1.58% of the chromosomes present an unknown *HLA-F* CDS sequence, and some of them encoded different *HLA-F* molecules. Thus, considering the genomic and CDS sequences, and in spite of the *HLA-F* conservation, *HLA-F* diversity is greater than the one reported at the IPD-IMGT/HLA database. We will clone and sequence these new sequences shortly.

The only structural variant presenting LD with *HLA-F* was *AluyHF* (Figure 2). The insertion was associated mainly with alleles F*01:01:01:08 and F*01:01:01:18, but other copies of these alleles follow *AluyHF* absence. *AluyHF* is not marking the presence of any specific allele. Nevertheless, the STR issue might be influencing these results.

HLA-G is the most studied non-classical HLA gene. Many manuscripts are addressing *HLA-G* diversity in worldwide populations (ALVAREZ et al., 2009; ARNAIZ-VILLENA et al., 2018; CARLINI et al., 2013; CASTELLI et al., 2014, 2017; EMMERY et al., 2017; GINEAU et al., 2015; MENDES-JUNIOR et al., 2013; NILSSON et al., 2018; OLIVEIRA et al., 2018a; RIBEYRE et al., 2018; SABBAGH et al., 2014; SANTOS et al., 2013; SONON et al., 2018; SVENDSEN et al., 2018; ZIDI et al., 2016), and many others addressing association with diseases. However, here we present the largest dataset of *HLA-G* variants in a population-based study in admixed samples. This Brazilian sample presented many *HLA-G* genomic sequences, with frequencies compatible with previous studies (CASTELLI et al., 2017; CASTELLI; MENDES; DONADI, 2007). In this sample we found many rare alleles such as G*01:09 and G*01:21N (Tables 7 and 8), but the protein polymorphism is still restricted to five frequent versions: G*01:01, G*01:03, G*01:04, G*01:05N, and G*01:06, with G*01:01

encoded by 68.09% of the chromosomes. However, these nucleotide exchanges seem to be modifying the main *HLA-G* features including peptide-binding, membrane expression, and binding to ILT and KIR receptors (DONADI et al., 2011; UROSEVIC; TROJAN; DUMMER, 2002). Interestingly, there was one homozygous individual for the truncated G*01:05N allele, as already reported in other populations (CASTRO et al., 2000), demonstrating that there may be other immunomodulatory mechanisms allowing the survival of individuals not producing the membrane-bound full-length *HLA-G* molecule, or that the soluble isoforms are sufficient for *HLA-G* function (LE DISCORDE et al., 2005).

The *Alu**YG* insertion occurs about 20Kb downstream *HLA-G* and it is included in the second large segregation block observed in Figure 2. This insertion is quite frequent in worldwide populations and has been previously associated with the haplotype PROMO-G010101a/G*01:01:01:01/UTR-1 (SANTOS et al., 2013), which in turn is considered a high-expressing haplotype (CARLINI et al., 2013; DURMANOVA et al., 2019; MARTELLI-PALOMINO et al., 2013; PORAS et al., 2017). Our data confirm that 84% of the chromosomes carrying the *Alu**YG* insertion also carry G*01:01:01:01. Moreover, the previous association of G*01:01:01:01 with PROMO-G010101a and UTR-1 is maintained, together with all other associations previously described in worldwide populations (reviewed at (CASTELLI et al., 2014, 2017; DONADI et al., 2011))

The admixed nature of our sample may also explain the high number of different *HLA-A* genomic alleles and *HLA-A* proteins observed (Tables 5 and 6). We found at least one copy of each *HLA-A* allele group including some rare ones such as A*80, but except A*43. This is not a methodological issue and reflects the low frequency of A*43 around the world, with an estimated frequency in Brazil according to the Brazilian Donor Registry (REDOME) of 0.007% (www.allelefrequencies.net). Since there are thousands of *HLA-A* alleles described in the IPD-IMGT/HLA database with mutations covering most of the *HLA-A* sequence, coupled with the many studies evaluating *HLA-A* diversity because of its importance for transplantation outcome, we did not find many new *HLA-A* alleles in our sample. New genomic alleles represent only 1.28% of the sequences (Table 6), most of them carrying intronic mutations and all encoding known proteins. The number of new CDS sequences (all encoding known proteins) was quite low, only 0.42%. Considering this sample size, and the fact that this population present European, African, Amerindian, and Asian component, it is possible that the IPD-IMGT/HLA database now covers most of the *HLA-A* genetic diversity observed in worldwide populations, at least at the protein level (Table 5).

The CNV esv3608493 represents a deletion of about 60Kb that includes the *HLA-H* pseudogene and gene *HCG4B* that encodes an ncRNA, about 5 Kb upstream *HLA-A*. This CNV has been describing influencing *HLA-A* expression (CHEN et al., 2017; OLIVEIRA et al., 2018a; PNG et al., 2019). Here we tracked this CNV by evaluating read depth at the *HLA-H* gene using hla-mapper. These CNV have been explored recently as allele H*Del (CARLINI et al., 2016; PAGANINI et al., 2019), with a frequency of 17% in the 1000 Genomes sample, which is compatible with the one observed in Brazil (15.77%). It should be mentioned that the previous study does not include Brazilian samples. Because of that, and combining the sample sizes, we have more than 3,500 individuals from worldwide populations confirming that CNV loss is related to only HLA-A*23 and A*24 alleles. These *HLA-A* alleles are clustered together because they present similar coding sequences and identical regulatory sequences (LIMA et al., 2019). HLA-A*24 (and probably A*23) is a high expressing lineage because of many mechanisms including a low level of methylation (RAMSURAN et al., 2015, 2017). Moreover, a previous study demonstrated that the promoters of A*23 and A*24 present fewer CpG sites (LIMA et al., 2019), and the CNV loss may reduce even more the number of CpG sites upstream *HLA-A*, favoring *HLA-A* expression. Thus, likewise *AluYHG*, esv3608493 is marking the presence of high-expression allele. This CNV is also related to alleles G*01:04 since 86.13% of the chromosomes with the CNV loss also present a G*01:04 allele, configuring two frequent haplotypes, G*01:04:01/CNVloss/A*24 and G*01:04:04/CNVloss/A*23. Other low-frequency combinations such as G*01:04:04/A*36:01 do not present the CNV loss (Table S5).

There are previous studies addressing the relationship between *HLA-F*, *HLA-G*, and *HLA-A* alleles, but using low-resolution typing methods (CARLINI et al., 2016; PAGANINI et al., 2019). Because of that, we will not compare our results with the previous ones, although they are similar when we consider the same resolution level. Our study aimed to measure the linkage disequilibrium between genes and markers in the HLA alpha block and to evaluate the extended haplotypes. The association of these markers with nearby SNPs can detect large blocks of segregation between neighboring, as was observed for each structural variant and their neighbor genes (Figure 2). Because of the distance between *HLA-F* and *HLA-A* (Figure 1) and the presence of many recombination sites in this segment (MIRETTI et al., 2005), we detected several extended haplotypes as described in Table S5. However, there was a close relationship between *HLA-G* and *HLA-A*, which are 100Kb apart (Table S4). There are many frequent haplotypes, and some *HLA-G* alleles (as discussed for G*01:04 earlier) follow specific *HLA-A* alleles. Other examples include the association of G*01:06 with A*01:01:01:01,

G*01:01:02:02 with A*66:01:01:01, and G*01:05N with A*30:01:01:01 or A*30:95 (Table S4).

HLA-A is a notorious target for natural selection (BITARELLO; FRANCISCO; MEYER, 2016; DOS SANTOS FRANCISCO et al., 2015; MEYER; THOMSON, 2001), considering that balancing selection has been shaping the *HLA-A* coding variability (MACK et al., 2009; SOLBERG et al., 2008; TU et al., 2007), increasing heterozygosity to improving antigen presentation. Because of that, high heterozygosity is expected for *HLA-A* and was indeed observed in our sample, with positive and significant Tajima's *D* considering the entire *HLA-A* locus (*Tajima's D* = 2.9005). Balancing selection has already been reported for *HLA-G*, especially at the regulatory segments (promoter and 3'UTR) (CASTELLI et al., 2011; DONADI et al., 2011; GINEAU et al., 2015; TAN; SHON; OBER, 2005). However, there is no strong functional evidence that different *HLA-G* regulatory regions are indeed associated with different *HLA-G* expression profiles besides an SNP at position -725 (OBER et al., 2006). Our data demonstrate a close relationship between *HLA-G* and *HLA-A* (Table S4). Therefore, natural selection acting on one locus will probably influence frequencies in the other locus, as a hitchhiking effect.

Because of that, high heterozygosity in *HLA-A* may lead to high heterozygosity in *HLA-G*, and vice-versa. For *HLA-G*, this high heterozygosity would be detected in intronic segments because most of the *HLA-G* alleles differ only by intronic mutations. Using a 150-bp window approach to calculate Tajima's *D* across *HLA-G* we detected Tajima's *D* over 3.00 in just one promoter region coinciding with position -725, and in introns 1, 4, and 5, while exonic segments and most of the promoter sequence presented negative values or low positive values (data not shown). Exception made to position -725, the full *HLA-G* promoter and 3'UTR presented a Tajima's *D* of 0.4834 and 0.3953, respectively. Since most of the *HLA-G* alleles differ only in intron sequences, and different intron sequences are associated with specific *HLA-A* allele groups, high heterozygosity in *HLA-A* we would lead to a high heterozygosity in the *HLA-G* introns as a hitchhiking effect. For instance, G*01:01:01:04 and G*01:01:01:05 are 1 mutation apart, and the first is associated with *HLA-A**29 and *HLA-A**74, while the later with *HLA-A**03 (Table S4). However, we cannot rule out balancing selection at the *HLA-G* position -725 since it was the second strongest signal observed across *HLA-G* after intron 5.

In conclusion, here we report the *HLA-F*, *HLA-G*, and *HLA-A* full variability in a large admixed Brazilian sample and the relationship among the observed alleles, together with their associations with three structural variants. The *Alu* element *Alu_{YHG}* and the CNV *esv3608493* are both associated with high-expressing alleles, the first with G*01:01:01:01 and the second

with A*23 and A*24. Our data suggests a close relationship between *HLA-A* and *HLA-G*, with both loci with evidence of balancing selection (intronic segments for *HLA-G*, the entire locus for *HLA-A*). Thus, this phenomenon might not be independent, and the observations for *HLA-G* might be hitchhiking effect because of *HLA-A*. This is the largest dataset of variants and full-resolution HLA alleles (in the alpha block) for an admixed sample, providing a valuable resource of variants, frequencies, and haplotypes in HLA genes of admixed samples for clinical and evolutive studies.

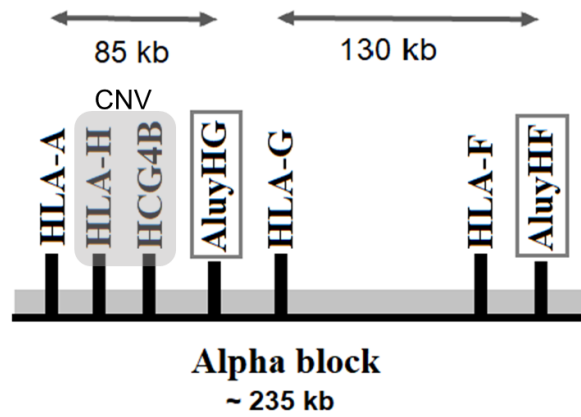


Figure 3. A map of the human MHC alpha block and the genes and structural variants evaluated in a large admixed Brazilian sample from São Paulo.

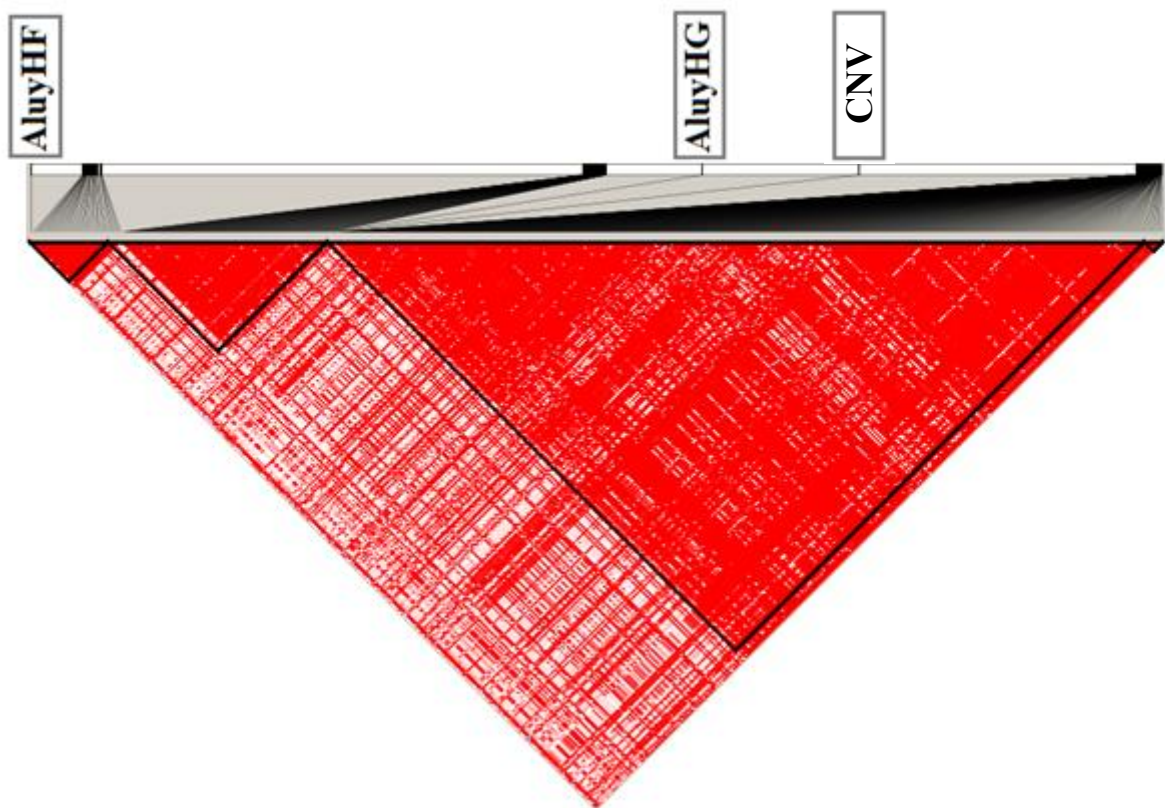


Figure 4. Linkage disequilibrium across the alpha-block, considering 638 biallelic SNPs, and the structural variants *AluHF*, *AluHG*, and *nsv823470*.

Table 1. *HLA-F* CDS alleles in an admixed sample from São Paulo (city), Brazil.

HLA-F CDS sequence ^a	Frequency (2n = 2.646)
F*01:01:01	0.6481
F*01:01:02	0.1765
F*01:03:01	0.1493
F*01:01:04	0.0011
F*01:01:05	0.0008
F*01:02	0.0026
F*01:04:01	0.0057
unknown	0.0158

^a alleles are listed in decreased order of frequency

Table 2. *HLA-F* coding alleles in an admixed sample from São Paulo (city), Brazil.

HLA-F genomic sequences (alleles) ^a	Frequency (2n= 2646)
F*01:01:01:09	0.1716
F*01:01:01:08	0.1584
F*01:01:01:01	0.1534
F*01:03:01:01	0.0835
F*01:01:01:18	0.0608
F*01:01:01:17	0.0537
F*01:03:01:03	0.0518
F*01:01:02:09	0.0480
F*01:01:02:08	0.0393
F*01:01:02:10	0.0393
F*01:01:02:07	0.0215
F*01:01:02:11	0.0185
F*01:01:01:11	0.0166
F*01:04:01:02	0.0057
F*01:01:01:13	0.0023
F*01:01:04	0.0011
F*01:03:01:04	0.0011
F*01:01:05	0.0008
F*01:01:01:19	0.0004
unknown	0.0722

^a alleles are listed in decreased order of frequency

Table 3. *HLA-G* CDS alleles in an admixed sample from São Paulo (city), Brazil.

HLA-G CDS sequence ^a	Frequency (2n = 2.646)
G*01:01:01	0.4275
G*01:01:02	0.1718
G*01:04:01	0.0921
G*01:03:01	0.0804
G*01:04:04	0.0650
G*01:01:03	0.0521
G*01:06	0.0502
G*01:05N	0.0211
G*01:01:22	0.0098
G*01:01:17	0.0049
G*01:01:15	0.0042
G*01:01:14	0.0038
G*01:01:09	0.0026
G*01:01:19	0.0023
G*01:04:05	0.0023
G*01:01:12	0.0019
G*01:08:02	0.0011
G*01:11	0.0011
G*01:09	0.0004
G*01:21N	0.0004
unknown	0.0049

^a alleles are listed in decreased order of frequency

Table 4. *HLA-G* genomic alleles in an admixed sample from São Paulo (city), Brazil.

HLA-G genomic sequences (alleles) ^a	Frequency (2n= 2646)
G*01:01:01:01	0.2462
G*01:01:02:01	0.1616
G*01:04:01:01	0.0876
G*01:01:01:05	0.0865
G*01:03:01:02	0.0789
G*01:04:04	0.0642
G*01:01:03:03	0.0517
G*01:06	0.0502
G*01:01:01:04	0.0468
G*01:01:01:08	0.0230
G*01:05N	0.0200
G*01:01:22:01	0.0098
G*01:01:02:02	0.0060
G*01:01:01:06	0.0053
G*01:01:17	0.0049
G*01:01:15	0.0042
G*01:01:14	0.0038
G*01:01:01:09	0.0034
G*01:04:01:02	0.0030
G*01:01:19	0.0023
G*01:04:05	0.0023
G*01:01:12	0.0019
G*01:08:02	0.0011
G*01:11	0.0011
G*01:01:03:04	0.0004
G*01:09	0.0004
G*01:21N	0.0004
unknown	0.0329

^a alleles are listed in decreased order of frequency

Table 5. *HLA-A* CDS alleles in an admixed sample from São Paulo (city), Brazil.

HLA-A CDS sequence ^a	Frequency (2n = 2.646)	(continue) HLA-A CDS sequence ^a	Frequency (2n = 2.646)
A*02:01:01	0.2139	A*01:02	0.0011
A*24:02:01	0.0915	A*26:02:01	0.0011
A*03:01:01	0.0907	A*26:03:01	0.0011
A*01:01:01	0.0869	A*31:15	0.0011
A*11:01:01	0.0522	A*66:02	0.0011
A*23:01:01	0.0518	A*02:09:01	0.0008
A*31:01:02	0.0378	A*02:12	0.0008
A*29:02:01	0.0374	A*11:12	0.0008
A*32:01:01	0.0344	A*24:02:13	0.0008
A*26:01:01	0.0295	A*33:03:23	0.0008
A*30:02:01	0.0295	A*69:01:01	0.0008
A*33:01:01	0.0219	A*01:03:01	0.0004
A*68:02:01	0.0212	A*02:01:04	0.0004
A*68:01:02	0.0200	A*02:01:15	0.0004
A*30:01:01	0.0193	A*02:07:01	0.0004
A*68:01:01	0.0166	A*02:14	0.0004
A*02:05:01	0.0155	A*02:17:02	0.0004
A*74:01:01	0.0113	A*02:20:01	0.0004
A*33:03:01	0.0102	A*02:221	0.0004
A*25:01:01	0.0094	A*02:334	0.0004
A*02:02:01	0.0091	A*02:52	0.0004
A*02:11:01	0.0087	A*02:66	0.0004
A*66:01:01	0.0079	A*11:02:01	0.0004
A*24:03:01	0.0076	A*23:05	0.0004
A*34:02:01	0.0068	A*24:104	0.0004
A*23:17:01	0.0064	A*24:218	0.0004
A*03:02:01	0.0042	A*29:02:07	0.0004
A*36:01	0.0042	A*30:29	0.0004
A*29:01:01	0.0038	A*31:01:13	0.0004
A*80:01:01	0.0038	A*31:04:01	0.0004
A*02:04	0.0034	A*32:02	0.0004
A*26:08:01	0.0034	A*68:01:24	0.0004
A*02:06:01	0.0030	A*68:02:02	0.0004
A*30:04:01	0.0023	A*68:17:01	0.0004
A*30:95	0.0015	unknown	0.0042

^a alleles are listed in decreased order of frequency

Table 6. *HLA-A* genomic alleles in an admixed sample from São Paulo (city), Brazil.

HLA-A genomic sequence^a	Frequency (2n = 2.646)	(continue) HLA-A genomic sequence^a	Frequency (2n = 2.646)	(continue) HLA-A genomic sequence^a	Frequency (2n = 2.646)
A*02:01:01:01	0.2109	A*36:01	0.0042	A*02:01:01:71	0.0004
A*01:01:01:01	0.0862	A*29:01:01:01	0.0038	A*02:01:04	0.0004
A*24:02:01:01	0.0824	A*80:01:01:02	0.0038	A*02:01:15	0.0004
A*03:01:01:01	0.0658	A*02:04	0.0034	A*02:02:01:04	0.0004
A*11:01:01:01	0.0518	A*26:08:01:01	0.0034	A*02:07:01:01	0.0004
A*23:01:01:01	0.0514	A*02:02:01:02	0.0030	A*02:14	0.0004
A*31:01:02:01	0.0348	A*02:06:01:01	0.0030	A*02:17:02:01	0.0004
A*32:01:01:01	0.0336	A*24:02:01:05	0.0026	A*02:20:01	0.0004
A*26:01:01:01	0.0280	A*31:01:02:04	0.0026	A*02:221	0.0004
A*29:02:01:01	0.0276	A*30:04:01:01	0.0023	A*02:334	0.0004
A*03:01:01:05	0.0215	A*33:03:01:03	0.0023	A*02:52	0.0004
A*33:01:01:01	0.0208	A*03:01:01:07	0.0019	A*02:66	0.0004
A*68:02:01:01	0.0197	A*02:01:01:05	0.0015	A*11:01:01:06	0.0004
A*30:01:01:01	0.0193	A*03:01:01:03	0.0015	A*11:02:01:01	0.0004
A*30:02:01:01	0.0185	A*30:95	0.0015	A*23:05	0.0004
A*68:01:01:02	0.0163	A*01:02	0.0011	A*24:02:01:12	0.0004
A*02:05:01:01	0.0147	A*26:02:01	0.0011	A*24:02:01:36	0.0004
A*68:01:02:01	0.0117	A*26:03:01:01	0.0011	A*24:03:01:02	0.0004
A*74:01:01:01	0.0113	A*29:02:01:03	0.0011	A*24:104	0.0004
A*25:01:01:01	0.0091	A*31:15	0.0011	A*24:218	0.0004
A*02:11:01:01	0.0087	A*66:02	0.0011	A*26:01:01:13	0.0004
A*29:02:01:02	0.0087	A*68:02:01:03	0.0011	A*29:02:07	0.0004
A*33:03:01:01	0.0079	A*02:01:01:08	0.0008	A*30:02:01:05	0.0004
A*66:01:01:01	0.0076	A*02:09:01:01	0.0008	A*30:29	0.0004
A*68:01:02:02	0.0076	A*02:12	0.0008	A*31:01:13	0.0004
A*24:03:01:01	0.0068	A*11:12	0.0008	A*31:04:01:02	0.0004
A*34:02:01:01	0.0068	A*24:02:13	0.0008	A*32:01:01:09	0.0004
A*23:17:01:01	0.0064	A*26:01:01:02	0.0008	A*32:02	0.0004
A*02:02:01:01	0.0057	A*33:03:23	0.0008	A*68:02:02	0.0004
A*30:02:01:02	0.0057	A*68:01:02:04	0.0008	A*68:17:01	0.0004
A*03:02:01	0.0042	A*69:01:01:01	0.0008	unknown	0.0128
A*24:02:01:04	0.0042	A*01:01:01:23	0.0004		
A*30:02:01:03	0.0042	A*01:03:01:02	0.0004		

^a alleles are listed in decreased order of frequency

Table 7. *HLA-F* genomic alleles and the AluyHF haplotypes in an admixed sample from São Paulo (city), Brazil.

HLA-F genomic sequences (alleles)	AluyHF	Frequency (2n= 2612)
F*01:01:01:01	Absent	0.1531
F*01:01:01:01	Present	0.0004
F*01:01:01:08	Present	0.1325
F*01:01:01:08	Absent	0.0245
F*01:01:01:09	Absent	0.1723
F*01:01:01:11	Absent	0.0168
F*01:01:01:13	Absent	0.0019
F*01:01:01:17	Absent	0.0536
F*01:01:01:17	Present	0.0004
F*01:01:01:18	Absent	0.0410
F*01:01:01:18	Present	0.0195
F*01:01:01:19	Absent	0.0004
F*01:01:01:21	Absent	0.0004
F*01:01:02:07	Absent	0.0214
F*01:01:02:08	Absent	0.0394
F*01:01:02:09	Absent	0.0482
F*01:01:02:10	Absent	0.0356
F*01:01:02:10	Present	0.0027
F*01:01:02:11	Absent	0.0188
F*01:01:04	Present	0.0011
F*01:01:05	Present	0.0008
F*01:03:01:01	Absent	0.0850
F*01:03:01:03	Absent	0.0521
F*01:03:01:04	Absent	0.0011
F*01:04:01:02	Absent	0.0057
F*01:05	Absent	0.0004
unknown	-	0.0708

Table 8. *HLA-G* genomic alleles and the AluYHG haplotypes in an admixed sample from São Paulo (city), Brazil.

HLA-G genomic sequence	AluYHG	Frequency (2n= 2612)
G*01:01:01:01	Present	0.2377
G*01:01:01:01	Absent	0.0092
G*01:01:01:04	Absent	0.0364
G*01:01:01:04	Present	0.0103
G*01:01:01:05	Absent	0.0869
G*01:01:01:05	Present	0.0004
G*01:01:01:06	Absent	0.0050
G*01:01:01:08	Present	0.0226
G*01:01:01:09	Present	0.0023
G*01:01:01:09	Absent	0.0004
G*01:01:02:01	Absent	0.1394
G*01:01:02:01	Present	0.0203
G*01:01:02:02	Absent	0.0061
G*01:01:03:03	Absent	0.0528
G*01:01:12	Present	0.0019
G*01:01:14	Absent	0.0038
G*01:01:15	Absent	0.0038
G*01:01:17	Present	0.0046
G*01:01:19	Absent	0.0019
G*01:01:22:01	Present	0.0092
G*01:01:22:01	Absent	0.0004
G*01:03:01:02	Absent	0.0789
G*01:03:01:02	Present	0.0004
G*01:04:01:01	Absent	0.0861
G*01:04:01:01	Present	0.0015
G*01:04:01:02	Absent	0.0031
G*01:04:04	Absent	0.0647
G*01:04:05	Absent	0.0023
G*01:05N	Absent	0.0199
G*01:06	Absent	0.0505
G*01:06	Present	0.0004
G*01:08:02	Present	0.0011
G*01:09	Absent	0.0004
G*01:11	Absent	0.0011
G*01:21N	Absent	0.0004
unknown	-	0.0337

Table S1. List of variants observed in the across the *HLA-F* gene in 1323 Brazilians from São Paulo, and their frequencies.

hg38 chr6 position	SNPId	Reference	Reference frequency	Alternative 1	Alternative 1 frequency	Alternative 2	Alternative 2 frequency	Alternative 3	Alternative 3 frequency	Alternative 4	Alternative 4 frequency
29723031	rs981327114	A	0.9996	AG	0.0004						
29723061	rs17875378	G	0.9773	A	0.0227						
29723079	rs1035728209	A	0.9996	C	0.0004						
29723167	.	A	0.9996	G	0.0004						
29723226	rs56044823	C	0.9509	T	0.0491						
29723241	rs41559212	G	0.9996	A	0.0004						
29723242	rs1362126	G	0.6141	A	0.3859						
29723262	rs373360195	C	0.9981	T	0.0019						
29723270	rs888682815	C	0.9996	T	0.0004						
29723313	rs1362125	T	0.5412	A	0.4588						
29723320	rs2075682	A	0.8212	T	0.1788						
29723363	rs2072896	C	0.8212	G	0.1788						
29723392	rs191310737	G	0.9996	A	0.0004						
29723420	rs771810518	A	0.9996	AT	0.0004						
29723464	rs372241614	A	0.9996	G	0.0004						
29723478	rs750462794	C	0.9996	T	0.0004						
29723501	rs17875379	C	0.9974	T	0.0026						
29723526	rs2076183	G	0.8212	A	0.1788						
29723608	rs373876297	G	0.9981	A	0.0019						
29723613	rs1632955	C	0.1493	G	0.8507						
29723614	rs1632954	T	0.1493	C	0.8507						
29723630	rs369106559	G	0.9974	T	0.0026						
29723668	rs61739958	C	0.9958	A	0.0042						
29723729	rs773661638	G	0.9996	A	0.0004						
29723805	rs17875380	C	0.9943	A	0.0057						
29723857	rs117540842	C	0.9917	G	0.0083						
29723936	rs2076182	A	0.8212	C	0.1788						
29723944	rs764490901	G	0.9996	A	0.0004						
29723967	rs2076181	C	0.8212	G	0.1788						
29723975	rs2076180	C	0.9977	T	0.0023						
29723995	rs2076179	T	0.5416	C	0.4584						
29723997	rs2076178	A	0.7816	C	0.2184						
29724023	rs17875381	C	0.9342	G	0.0658						
29724080	rs2072895	C	0.5416	G	0.4584						
29724150	rs1049628860	G	0.9996	C	0.0004						
29724252	rs747133995	G	0.9996	T	0.0004						
29724316	rs147086730	G	0.9996	C	0.0004						

29724405	rs138293991	C	0.9992	T	0.0008		
29724528	rs1632953	A	0.1493	G	0.8507		
29724557	rs9258186	A	0.8212	G	0.1788		
29724652	rs140848774	AAATTTCTGAGGG	0.8212	A	0.1788		
29724656	rs1632952	T	0.1493	C	0.6720	*	0.1788
29724685	rs142275427	TGGAATACCGATCCGC	0.9599	T	0.0401		
29724688	rs368156595	A	0.9577	C	0.0023	*	0.0401
29724693	rs13437375	C	0.8212	A	0.1387	*	0.0401
29724776	rs193250743	A	0.9996	G	0.0004		
29724785	rs1736926	A	0.1493	G	0.8507		
29724841	rs144376525	C	0.9943	T	0.0057		
29724845	rs2072899	C	0.3628	A	0.4584	G	0.1788
29724846	rs17178385	C	0.9414	G	0.0586		
29724857	rs1736925	T	0.1493	G	0.8507		
29724912	rs1012353106	G	0.9996	A	0.0004		
29724952	rs2072898	T	0.8212	G	0.1788		
29724960	rs181298082	T	0.9996	C	0.0004		
29725033	rs142311857	C	0.9989	A	0.0011		
29725053	rs376099110	C	0.9996	T	0.0004		
29725130	rs78154236	C	0.9996	T	0.0004		
29725234	rs1736924	C	0.1493	T	0.8507		
29725242	rs201844059	G	0.9989	A	0.0011		
29725336	rs2076177	C	0.8212	T	0.1788		
29725406	rs17184813	G	0.9830	A	0.0170		
29725428	rs73745582	C	0.9992	T	0.0008		
29725456	rs57237143	C	0.9996	T	0.0004		
29725671	rs17875383	C	0.9550	T	0.0450		
29725722	rs2235383	A	0.8212	G	0.1788		
29725741	rs57368914	G	0.9962	A	0.0038		
29725778	rs1024882482	G	0.9996	C	0.0004		
29725933	rs933328475	C	0.9992	G	0.0008		
29725949	rs114325231	G	0.9512	T	0.0488		
29725955	rs763552055	G	0.9996	A	0.0004		
29725984	rs180964015	C	0.9985	T	0.0015		
29726161	rs2735061	G	0.8341	A	0.1659		
29726192	rs1736922	C	0.7865	T	0.2135		
29726241	rs1032550255	C	0.9996	A	0.0004		
29726267	rs3835204	T	0.8243	TG	0.1757		
29726269	rs201715782	G	0.9966	GA	0.0030	T	0.0004
29726288	rs578160182	C	0.9989	G	0.0011		
29726344	rs913969959	T	0.9996	G	0.0004		
29726483	rs929480270	T	0.9996	C	0.0004		

29726559	rs70983993	ATGTG	0.2774	ATG	0.1149	A	0.2116	ATGTGTG	0.3896	ATGTGTGTG	0.0064
29726650	rs1059174	C	0.1493	T	0.8507						
29726666	rs62391801	A	0.9512	G	0.0488						
29726749	rs917411750	C	0.9992	T	0.0008						
29726793	rs3734813	A	0.8212	G	0.1788						
29726855	.	AG	0.9996	A	0.0004						
29726903	rs3734814	A	0.8212	C	0.1788						
29726904	rs3734815	A	0.8212	T	0.1788						

Table S2. List of variants observed in the across the *HLA-G* gene in 1323 Brazilians from São Paulo, and their frequencies.

hg38 chr6 position	SNPid	Reference	Refeence frequency	Alternative 1	Alternative 1 frequency	Alternative 2	Alternative 2 frequency
29826369	rs114252012	C	0.9898	A	0.0102		
29826435	rs978018398	C	0.9992	T	0.0008		
29826439	rs527969449	AGT	0.9940	A	0.0060		
29826468	rs113076482	T	0.9970	G	0.0030		
29826475	rs745500457	T	0.9985	C	0.0015		
29826535	rs112835540	C	0.9989	T	0.0011		
29826540	rs1736936	G	0.5185	A	0.4815		
29826602	rs994883956	G	0.9996	A	0.0004		
29826666	rs1736935	A	0.4358	G	0.5642		
29826690	rs3823321	G	0.8380	A	0.1620		
29826694	rs750652210	T	0.9989	C	0.0011		
29826705	rs1736934	A	0.6782	T	0.3218		
29826707	rs17875389	A	0.9196	G	0.0804		
29826724	rs3115630	T	0.0344	C	0.9656		
29826747	rs146374870	G	0.9936	A	0.0064		
29826772	rs192232623	G	0.9996	T	0.0004		
29826864	.	C	0.9996	T	0.0004		
29826881	rs1632947	G	0.5162	A	0.4838		
29826923	rs370338057	C	0.9981	A	0.0019		
29827035	rs182801644	C	0.9955	T	0.0045		
29827066	rs547123392	C	0.9989	T	0.0011		
29827083	rs1632946	C	0.5162	T	0.4838		
29827090	rs115528044	G	0.9996	A	0.0004		
29827120	rs1233334	G	0.0921	C	0.8274	T	0.0804
29827129	rs2249863	T	0.5162	G	0.4838		
29827156	rs2735022	A	0.5159	G	0.4841		
29827179	rs35674592	G	0.5162	T	0.4838		
29827199	rs17875391	A	0.9709	G	0.0291		
29827212	rs1632944	G	0.5162	A	0.4838		
29827293	.	G	0.9996	A	0.0004		
29827299	rs201221694	A	0.9486	AG	0.0514		
29827304	rs368205133	GA	0.9479	G	0.0521		
29827306	rs112940953	A	0.9992	G	0.0008		
29827336	rs17875393	C	0.9418	G	0.0582		
29827359	rs1736933	A	0.5155	C	0.4845		
29827362	rs149890776	A	0.9709	G	0.0291		
29827368	rs1736932	C	0.4354	G	0.5646		
29827402	rs17875394	G	0.9857	A	0.0144		
29827445	rs17875395	G	0.9207	A	0.0793		
29827446	rs866544668	G	0.9981	A	0.0019		
29827454	rs17875396	G	0.9207	A	0.0793		
29827476	rs1632943	C	0.4350	A	0.5650		
29827490	rs191630481	G	0.9955	A	0.0045		
29827548	rs781618441	G	0.9996	A	0.0004		
29827589	rs551170211	TC	0.9981	T	0.0019		
29827644	rs1233333	G	0.5162	A	0.4838		
29827660	.	G	0.9996	A	0.0004		
29827680	.	T	0.9996	C	0.0004		
29827708	.	C	0.9996	T	0.0004		
29827755	rs924682241	G	0.9992	A	0.0008		
29827789	rs17875397	C	0.9196	T	0.0804		
29827846	rs143732275	T	0.9996	C	0.0004		
29827858	.	C	0.9996	T	0.0004		
29827859	rs1630223	G	0.5162	A	0.4838		

29827880	rs1630185	G	0.5162	A	0.4838
29827943	rs56388903	A	0.1008	G	0.8992
29827970	rs6932888	G	0.7228	C	0.2772
29827974	rs6932596	C	0.7232	T	0.2768
29827975	rs547884515	G	0.9989	C	0.0011
29827991	rs1629329	T	0.4358	C	0.5642
29828017	rs539643749	G	0.9996	A	0.0004
29828032	rs1628628	C	0.5684	T	0.4316
29828045	rs535773328	A	0.9996	G	0.0004
29828136	rs41551813	A	0.9196	T	0.0804
29828137	rs72558173	C	0.9989	T	0.0011
29828141	rs80153902	G	0.9977	A	0.0023
29828150	rs72558174	G	0.9962	A	0.0038
29828168	rs9258495	G	0.9981	T	0.0019
29828210	rs78627024	G	0.9958	A	0.0042
29828216	rs1130355	G	0.5162	A	0.4838
29828326	rs1626038	T	0.4267	C	0.5733
29828329	rs17875399	G	0.9196	T	0.0804
29828337	rs535263603	G	0.9996	A	0.0004
29828338	rs1736927	A	0.4267	C	0.5733
29828342	rs201510147	G	0.9977	A	0.0023
29828349	rs3215482	A	0.5072	AC	0.4928
29828351	rs17875400	C	0.9196	A	0.0804
29828375	rs1625907	G	0.5076	C	0.4924
29828379	rs1052344302	G	0.9996	T	0.0004
29828457	rs9278333	CA	0.5162	C	0.4838
29828470	rs542016267	C	0.9996	G	0.0004
29828480	rs1625035	C	0.4362	T	0.5638
29828488	rs17875401	G	0.9196	T	0.0804
29828518	.	G	0.9996	C	0.0004
29828529	rs1624337	G	0.5162	A	0.4838
29828550	rs1130356	C	0.7323	T	0.2677
29828571	rs79303923	C	0.9940	T	0.0060
29828592	rs3873252	A	0.9479	T	0.0521
29828594	.	G	0.9996	T	0.0004
29828599	rs12722477	C	0.8380	A	0.1620
29828657	rs41557518	AC	0.9785	A	0.0215
29828679	rs777893003	G	0.9996	A	0.0004
29828715	rs17875402	G	0.9974	A	0.0026
29828746	rs55916353	T	0.9996	C	0.0004
29828860	rs17875403	C	0.9921	T	0.0079
29828863	rs1632942	T	0.4358	C	0.5642
29828898	rs17875404	G	0.9196	T	0.0804
29828908	rs1632941	T	0.5162	C	0.4838
29828923	rs148061958	C	0.9951	T	0.0049
29828938	.	G	0.9996	A	0.0004
29828948	rs370704534	C	0.9974	G	0.0026
29828972	rs62391965	C	0.9479	A	0.0521
29828991	rs1632940	T	0.2334	C	0.7666
29829061	rs1736923	A	0.5162	G	0.4838
29829079	.	C	0.9996	T	0.0004
29829157	rs114041958	G	0.9196	A	0.0804
29829158	rs1632939	G	0.5162	A	0.4838
29829188	.	G	0.9996	A	0.0004
29829205	rs967683300	A	0.9992	G	0.0008
29829209	rs1632938	G	0.5162	A	0.4838
29829222	rs910223979	C	0.9992	A	0.0008
29829226	rs766239071	C	0.9992	T	0.0008
29829248	.	A	0.9996	C	0.0004
29829255	rs919055338	A	0.9996	G	0.0004
29829266	rs145023077	C	0.9970	T	0.0030
29829275	rs116139267	C	0.9992	T	0.0008

29829296	rs188836562	G	0.9970	C	0.0030
29829378	rs17875405	G	0.9196	C	0.0804
29829396	rs1736920	A	0.4358	G	0.5642
29829434	rs41562616	C	0.9196	T	0.0804
29829525	rs45530733	C	0.9989	T	0.0011
29829533	.	G	0.9996	A	0.0004
29829546	rs144753960	C	0.9996	T	0.0004
29829575	rs140470441	T	0.9996	G	0.0004
29829643	rs12722482	C	0.9494	T	0.0506
29829644	rs76951509	G	0.9977	A	0.0023
29829671	rs17875406	G	0.9339	A	0.0661
29829725	.	T	0.9996	G	0.0004
29829732	.	A	0.9989	G	0.0011
29829776	rs1632937	G	0.5162	C	0.4838
29829862	rs1049033	C	0.7500	T	0.2500
29829919	rs1130363	A	0.4358	G	0.5642
29830005	rs1611627	T	0.5687	C	0.4313
29830117	.	T	0.9996	C	0.0004
29830122	rs1632934	T	0.5162	C	0.4838
29830156	rs1632933	C	0.5162	T	0.4838
29830174	rs1736912	A	0.5162	G	0.4838
29830256	rs17179080	G	0.9683	A	0.0317
29830262	rs1632932	G	0.5162	A	0.4838
29830306	rs114038308	C	0.9479	T	0.0521
29830342	rs913873042	C	0.9996	T	0.0004
29830363	rs915667	A	0.5340	G	0.4660
29830512	rs879242015	C	0.9992	T	0.0008
29830600	rs111781429	T	0.9966	C	0.0034
29830642	rs915670	G	0.7500	A	0.2500
29830648	rs915669	G	0.4358	T	0.5642
29830682	rs915668	C	0.4358	G	0.5642
29830707	rs190645695	G	0.9996	A	0.0004
29830786	rs551200694	T	0.9962	C	0.0038
29830804	rs371194629	G	0.5963	GATTGTTCATGCCT	0.4037
29830813	.	T	0.9996	C	0.0004
29830831	rs567747015	C	0.9962	T	0.0038
29830833	rs1707	C	0.0967	T	0.9033
29830840	rs1710	G	0.4354	C	0.5646
29830857	rs17179101	C	0.9479	A	0.0521
29830862	rs146339774	G	0.9924	C	0.0076
29830865	rs17179108	C	0.8622	T	0.1378
29830882	rs569057854	C	0.9989	T	0.0011
29830922	rs180827037	G	0.9996	T	0.0004
29830951	rs138249160	T	0.9974	C	0.0026
29830972	rs1063320	C	0.4354	G	0.5646
29831017	rs9380142	A	0.7209	G	0.2791
29831026	rs1610696	C	0.7311	G	0.2689
29831057	rs1233331	G	0.9709	A	0.0291
29831060	.	T	0.9996	C	0.0004
29831079	.	C	0.9996	A	0.0004
29831119	rs541542414	C	0.9992	T	0.0008

Table S3. List of variants observed in the across the *HLA-A* gene in 1323 Brazilians from São Paulo, and their frequencies.

hg38 chr6 position	SNPId	Reference	Refece frequency	Alternative 1	Alternative 1 frequency	Alternative 2	Alternative 2 frequency	Alternative 3	Alternative 3 frequency
29940812	rs9260064	T	0.2997	C	0.7003				
29940813	rs9260065	G	0.2997	A	0.7003				
29940816	rs113858731	G	0.9320	C	0.0680				
29940855	rs9260066	C	0.6224	T	0.3776				
29940872	rs9260067	T	0.2948	G	0.7052				
29940882	rs13219558	T	0.6723	C	0.3277				
29940890	rs1655920	A	0.1485	G	0.8515				
29940899	rs1632890	T	0.9074	A	0.0926				
29940902	rs9260068	G	0.6228	A	0.3772				
29940903	rs145284712	AATGAATTG	0.7396	A	0.2604				
29940909	rs9260069	T	0.3624	C	0.3772	*	0.2604		
29940949	rs9260070	C	0.9433	G	0.0567				
29940965	rs9260071	A	0.9119	C	0.0881				
29940976	rs72498365	G	0.9267	C	0.0733				
29941011	.	A	0.9996	G	0.0004				
29941012	rs112519287	T	0.9320	C	0.0680				
29941027	rs13203830	G	0.6716	A	0.3284				
29941060	.	C	0.9996	T	0.0004				
29941061	rs9260072	G	0.2944	C	0.7056				
29941081	rs968065011	C	0.9989	G	0.0011				
29941092	rs143741490	C	0.9581	A	0.0420				
29941114	rs9260073	C	0.9433	T	0.0567				
29941115	rs79761237	G	0.8341	A	0.1659				
29941118	rs9260074	G	0.9433	A	0.0567				
29941124	rs9260075	A	0.9433	C	0.0567				
29941151	rs67901303	GAC	0.3511	G	0.3205	AAC	0.3284		
29941160	rs78179089	C	0.6795	T	0.3205				
29941177	rs9260077	G	0.2944	A	0.7056				
29941190	rs13207054	G	0.6716	A	0.3284				
29941199	rs74218766	C	0.9267	T	0.0733				
29941233	rs976764801	T	0.9996	C	0.0004				
29941237	rs375623213	T	0.9701	C	0.0299				
29941238	rs919504401	G	0.9996	A	0.0004				
29941244	rs2571420	C	0.9074	G	0.0926				

29941256	rs9260078	A	0.2729	G	0.7271
29941267	rs28749141	G	0.6788	C	0.3212
29941293	rs9260079	T	0.6187	C	0.3813
29941294	rs9260080	G	0.8541	A	0.1459
29941334	rs191021452	C	0.9962	G	0.0038
29941376	rs9260081	G	0.2944	T	0.7056
29941404	rs1632889	A	0.1485	C	0.8515
29941420	rs1655919	G	0.0949	C	0.9051
29941428	rs1632888	C	0.8541	T	0.1459
29941440	rs1632887	T	0.8541	G	0.1459
29941442	rs1632886	T	0.8541	C	0.1459
29941461	rs9260082	T	0.9539	C	0.0461
29941472	rs9260083	A	0.8386	C	0.1614
29941474	rs112664400	T	0.2971	TA	0.7029
29941475	rs72498367	A	0.9996	G	0.0004
29941502	rs28749142	T	0.7396	A	0.2604
29941514	.	C	0.9996	A	0.0004
29941515	.	C	0.9996	T	0.0004
29941529	rs867644033	C	0.9996	T	0.0004
29941562	rs9260084	G	0.4599	A	0.5401
29941576	rs140625882	C	0.9581	G	0.0420
29941583	rs9260085	G	0.8386	A	0.1614
29941595	rs9260086	C	0.4599	G	0.5401
29941599	rs9260087	C	0.2944	A	0.7056
29941614	rs9260088	C	0.8348	A	0.1652
29941632	rs9260089	C	0.8386	T	0.1614
29941649	rs9260090	T	0.8352	C	0.1648
29941663	rs9260091	A	0.8791	G	0.1209
29941670	rs9260092	G	0.4596	C	0.5404
29941677	rs9260093	T	0.8783	C	0.1217
29941695	rs9260094	G	0.2952	C	0.7048
29941722	rs9260095	A	0.2982	T	0.7018
29941733	rs9260096	C	0.2982	T	0.7018
29941743	rs9260097	G	0.2982	A	0.7018
29941746	rs9260098	A	0.2982	C	0.7018
29941747	rs9260099	G	0.2982	T	0.7018
29941756	rs576018139	G	0.9777	C	0.0223
29941765	rs9260100	A	0.8386	G	0.1614
29941782	rs9260101	G	0.4596	A	0.5404
29941785	rs9260102	C	0.8386	A	0.1614

29941816	rs9260103	T	0.2944	G	0.7018	C	0.0038
29941820	rs114157794	T	0.9271	G	0.0729		
29941839	rs9260104	T	0.2944	C	0.7056		
29941856	rs141035764	A	0.9463	AT	0.0537		
29941890	.	G	0.9996	T	0.0004		
29941897	rs151250402	A	0.9962	G	0.0038		
29941898	rs9260105	G	0.9539	T	0.0461		
29941937	rs9260106	T	0.8791	C	0.1209		
29941968	rs9260107	A	0.2944	G	0.7056		
29941974	rs2905764	T	0.4596	C	0.5404		
29941979	rs9260109	G	0.2944	C	0.5465	T	0.1591
29941992	rs67139508	AAAC	0.7967	A	0.1995	CAAC	0.0038
29942020	rs3129022	C	0.8541	T	0.1459		
29942030	rs3094141	T	0.0949	G	0.9051		
29942034	rs35333729	G	0.9153	T	0.0847		
29942063	rs2734904	G	0.4558	A	0.5442		
29942066	.	A	0.9996	AT	0.0004		
29942075	.	C	0.9996	G	0.0004		
29942139	rs9260111	A	0.2944	T	0.7056		
29942141	rs9260112	T	0.2944	C	0.7056		
29942147	rs2734903	G	0.6187	A	0.3813		
29942168	rs112027704	G	0.9989	A	0.0011		
29942180	rs79231337	G	0.9267	C	0.0733		
29942190	.	C	0.9996	A	0.0004		
29942193	rs116395178	T	0.9388	A	0.0612		
29942211	rs9260114	C	0.2445	T	0.7555		
29942222	rs35158227	C	0.8409	T	0.1591		
29942245	rs2735115	C	0.9581	A	0.0420		
29942248	.	T	0.9996	C	0.0004		
29942250	rs9260115	C	0.2982	G	0.7018		
29942256	rs41541221	G	0.9467	A	0.0533		
29942257	rs2735114	G	0.5805	A	0.4195		
29942280	rs41272547	C	0.9267	A	0.0733		
29942306	rs3132688	A	0.0949	G	0.9051		
29942328	rs9260116	G	0.9471	A	0.0529		
29942345	rs11964254	C	0.9463	A	0.0537		
29942355	rs187514143	G	0.9962	A	0.0038		
29942358	rs115743151	T	0.9970	G	0.0030		
29942382	rs947036726	C	0.9996	A	0.0004		
29942390	rs2735113	C	0.9119	G	0.0881		

29942399	rs9260118	C	0.2948	T	0.7052		
29942407	.	C	0.9996	G	0.0004		
29942412	rs9260119	A	0.2944	T	0.7056		
29942450	rs193269784	C	0.9962	T	0.0038		
29942488	rs41545520	G	0.9467	T	0.0533		
29942499	rs41560714	T	0.4535	G	0.5465		
29942500	rs9260120	G	0.9467	C	0.0533		
29942509	rs2916801	C	0.1485	T	0.8515		
29942547	rs114945359	G	0.9467	C	0.0533		
29942553	rs112877416	G	0.9932	A	0.0068		
29942566	rs41541913	G	0.9962	C	0.0038		
29942581	rs1143146	C	0.4596	G	0.5404		
29942594	rs2230954	C	0.8386	T	0.1614		
29942601	rs41546314	C	0.7400	T	0.2600		
29942620	rs41564012	T	0.9887	A	0.0113		
29942625	rs1136656	G	0.8371	A	0.1629		
29942642	rs9260122	G	0.8390	T	0.1610		
29942643	rs9278464	GA	0.8390	G	0.1610		
29942651	rs770946958	C	0.9996	A	0.0004		
29942652	rs9260123	C	0.4562	G	0.5438		
29942660	rs9260124	C	0.6190	T	0.3810		
29942664	rs754662878	G	0.9996	T	0.0004		
29942673	rs17885299	G	0.7396	C	0.2604		
29942680	rs9260125	T	0.2944	G	0.7056		
29942683	rs9260126	T	0.7143	C	0.2857		
29942691	rs41557221	G	0.9581	A	0.0420		
29942696	rs41561116	G	0.8409	A	0.1591		
29942701	rs9260127	G	0.2944	C	0.7056		
29942705	rs9256980	G	0.2944	A	0.5442	T	0.1614
29942710	rs746723651	G	0.9996	T	0.0004		
29942715	rs9260128	G	0.8791	T	0.1209		
29942719	rs2904759	C	0.0949	G	0.9051		
29942729	rs41549214	G	0.9452	T	0.0548		
29942730	rs41554614	C	0.9592	T	0.0408		
29942731	rs9260129	A	0.2944	G	0.7056		
29942733	rs41541212	G	0.9255	A	0.0745		
29942741	rs560549696	C	0.9992	T	0.0008		
29942745	rs9260130	G	0.2944	C	0.7056		
29942750	rs17879990	C	0.5805	T	0.4195		
29942761	rs1136657	C	0.7396	T	0.2604		

29942775	rs41542421	A	0.9996	G	0.0004		
29942780	rs1136659	T	0.8847	A	0.1153		
29942781	rs2075684	T	0.4777	C	0.3288	A	0.1935
29942785	rs1059418	A	0.8065	C	0.1727	T	0.0208
29942789	rs41545116	G	0.9781	A	0.0219		
29942804	rs1059423	C	0.9452	A	0.0548		
29942806	rs1059426	C	0.9452	T	0.0548		
29942809	rs12721672	G	0.9452	A	0.0548		
29942810	rs41549014	G	0.9996	A	0.0004		
29942827	rs12721675	C	0.6810	A	0.3190		
29942846	rs1059438	A	0.9962	T	0.0038		
29942859	rs1059441	G	0.9962	A	0.0038		
29942863	rs3173429	C	0.9119	T	0.0881		
29942883	rs41559117	A	0.9739	G	0.0261		
29942886	rs707910	G	0.9074	A	0.0926		
29942902	rs199474404	G	0.9581	A	0.0420		
29942921	rs1059449	G	0.9067	A	0.0933		
29942922	rs1059451	G	0.9962	A	0.0038		
29942923	rs1059452	G	0.7396	T	0.2604		
29942926	rs41541222	G	0.9067	T	0.0933		
29942939	rs1059455	C	0.5767	G	0.4233		
29942940	rs1064588	A	0.5431	G	0.4150	T	0.0420
29942942	rs2230991	G	0.8035	A	0.1546	C	0.0420
29942944	rs199474424	G	0.8454	C	0.1546		
29942948	rs199474430	C	0.8409	G	0.1591		
29942953	rs199474436	T	0.5809	A	0.4191		
29942954	rs79361534	G	0.9074	A	0.0926		
29942964	rs199474449	A	0.9989	G	0.0011		
29942965	rs78306866	G	0.2880	C	0.7120		
29942973	rs199474457	C	0.9184	T	0.0816		
29942975	rs1136683	G	0.7430	C	0.2532	A	0.0038
29942982	rs1071742	T	0.5918	A	0.2355	C	0.1727
29942984	rs1136688	G	0.5918	A	0.4082		
29942985	rs2231004	A	0.9558	G	0.0442		
29942990	rs1136689	G	0.7967	C	0.2033		
29942993	rs755077203	ACCC	0.7967	A	0.2033		
29942997	rs778137446	TG	0.7967	T	0.2033		
29943002	rs775073538	G	0.7967	GCTCC	0.2033		
29943024	rs1136692	C	0.7914	A	0.2086		
29943039	rs41562315	G	0.9267	A	0.0733		

29943040	rs41559412	G	0.9705	C	0.0295		
29943043	rs9260142	G	0.1599	C	0.8401		
29943045	rs41541123	G	0.9985	T	0.0015		
29943053	rs281864745	GTCAGGACC	0.9989	G	0.0011		
29943057	rs63225861	G	0.1062	C	0.8927	*	0.0011
29943062	rs9260144	C	0.5888	T	0.4112		
29943072	rs377364793	C	0.9962	T	0.0038		
29943084	rs41555717	A	0.7989	G	0.2011		
29943088	rs41540614	C	0.9467	G	0.0533		
29943089	rs41543916	G	0.9267	A	0.0733		
29943110	rs560554218	G	0.9966	T	0.0034		
29943115	rs9260145	A	0.6187	G	0.3813		
29943133	rs9260146	C	0.9074	T	0.0926		
29943146	rs9260147	C	0.8537	T	0.0926	A	0.0537
29943149	.	C	0.9996	T	0.0004		
29943164	.	C	0.9992	A	0.0008		
29943170	rs41545815	T	0.8413	A	0.1587		
29943172	rs189287094	C	0.9962	G	0.0038		
29943203	rs41552627	C	0.9463	T	0.0537		
29943209	rs17882350	G	0.7392	A	0.2608		
29943213	rs539098093	T	0.9962	G	0.0038		
29943223	rs9256982	T	0.1485	C	0.3810	G	0.4705
29943226	rs9260148	G	0.8314	A	0.1686		
29943228	rs41547618	G	0.9467	A	0.0533		
29943234	rs41562213	C	0.9467	T	0.0533		
29943236	rs777171774	G	0.9992	A	0.0008		
29943242	rs9260149	T	0.6221	C	0.3779		
29943252	rs9260150	G	0.8866	T	0.1134		
29943253	rs9260151	C	0.8311	T	0.1689		
29943260	rs9260152	G	0.6259	C	0.3742		
29943279	rs1071743	A	0.5801	G	0.2343	C	0.1856
29943286	rs199474485	T	0.6670	G	0.3330		
29943287	rs1136695	A	0.3050	G	0.6950		
29943292	rs1136697	A	0.8401	T	0.1595	G	0.0004
29943300	rs199474490	G	0.9962	C	0.0038		
29943309	rs1136700	T	0.7320	C	0.2680		
29943315	rs1136702	G	0.7392	T	0.2608		
29943321	rs1059488	T	0.9539	C	0.0461		
29943326	rs41563718	C	0.9739	G	0.0261		
29943334	rs757867739	ACC	0.9460	A	0.0540		

29943337	rs3173420	G	0.3311	A	0.6149	T	0.0540		
29943338	rs12721717	G	0.5042	C	0.4418	GAA	0.0540		
29943342	rs3173419	G	0.5030	T	0.4418	C	0.0540	A	0.0011
29943343	rs41557212	A	0.9996	T	0.0004				
29943347	rs3173418	C	0.9388	T	0.0612				
29943372	rs1059506	C	0.8386	T	0.1614				
29943377	rs1059509	C	0.5208	A	0.4792				
29943401	rs41542714	G	0.9996	A	0.0004				
29943408	rs45474295	A	0.9996	G	0.0004				
29943413	rs1059514	G	0.5336	A	0.4664				
29943421	rs1059516	T	0.6791	C	0.3209				
29943426	rs1059517	A	0.6655	C	0.3345				
29943430	rs1059520	G	0.6795	A	0.3205				
29943441	rs1059526	G	0.9392	A	0.0608				
29943445	rs1059535	C	0.9074	T	0.0926				
29943448	rs1059536	A	0.7256	G	0.2744				
29943450	rs1059537	G	0.9452	T	0.0215	C	0.0333		
29943451	rs9256983	A	0.1516	T	0.6478	C	0.1459	G	0.0548
29943462	rs9260155	T	0.7411	C	0.2589				
29943463	rs9260156	T	0.6066	G	0.2309	A	0.1625		
29943469	rs9260157	C	0.9074	T	0.0926				
29943479	rs1059542	T	0.0949	G	0.9051				
29943483	rs543623321	A	0.8005	C	0.1946	G	0.0049		
29943484	rs3129017	C	0.8005	G	0.1946	A	0.0049		
29943494	rs879577815	G	0.7562	C	0.2438				
29943495	rs3098019	T	0.7562	G	0.2438				
29943496	rs199474567	G	0.9996	C	0.0004				
29943507	rs41563116	T	0.9777	C	0.0223				
29943542	rs41554520	G	0.9992	T	0.0008				
29943549	rs2517725	A	0.9581	G	0.0420				
29943558	rs199474585	G	0.9996	A	0.0004				
29943567	rs17878838	C	0.7838	A	0.1587	T	0.0574		
29943579	rs3132687	G	0.8541	A	0.1459				
29943583	rs17878860	A	0.8413	G	0.1587				
29943589	.	CG	0.9996	C	0.0004				
29943642	rs2735112	A	0.3715	G	0.6285				
29943654	rs186459750	G	0.9962	C	0.0038				
29943680	rs140855897	G	0.0949	GT	0.9051				
29943689	rs559107954	C	0.9962	T	0.0038				
29943728	rs41543412	C	0.9267	G	0.0733				

29943733	rs41553212	A	0.9705	T	0.0295		
29943745	rs562709947	T	0.9962	TA	0.0038		
29943757	rs75439900	A	0.8409	C	0.1591		
29943761	rs2735111	G	0.2944	A	0.7056		
29943768	rs76475517	G	0.8409	A	0.1591		
29943779	rs898377291	C	0.9996	G	0.0004		
29943794	rs29028728	G	0.8409	C	0.1591		
29943804	.	T	0.9996	A	0.0004		
29943806	rs139109856	G	0.7740	GACAC	0.2260		
29943810	rs9280789	C	0.6795	CA	0.3205		
29943811	rs371311437	C	0.6795	CACA	0.3205		
29943815	rs549627736	A	0.9962	G	0.0038		
29943827	rs77467746	C	0.8409	A	0.1591		
29943850	rs182296210	T	0.9962	C	0.0038		
29943860	.	C	0.9996	T	0.0004		
29943873	rs188062958	G	0.9962	A	0.0038		
29943907	rs9260164	C	0.8503	T	0.1497		
29943908	rs17879103	C	0.7649	T	0.2351		
29943936	rs190837170	A	0.9970	G	0.0030		
29943948	.	C	0.9985	CTCTTCAGGGACTAGAA	0.0015		
29943950	rs29028878	T	0.2959	C	0.7041		
29943951	rs41561516	C	0.4550	T	0.5412	G	0.0038
29943952	rs9280792	T	0.9981	TC	0.0019		
29943955	rs9260165	G	0.4573	GGGACTAGAATTTTCCAC	0.5412	C	0.0015
29943963	rs2508038	A	0.4573	G	0.5427		
29943993	rs549444296	GGT	0.9962	G	0.0038		
29944013	rs9260166	T	0.4573	C	0.5427		
29944026	rs9260167	G	0.4535	A	0.5465		
29944041	rs188231435	T	0.9962	C	0.0038		
29944047	rs193073597	G	0.9962	A	0.0038		
29944050	rs9260168	G	0.4573	A	0.5427		
29944059	rs9260169	C	0.4573	T	0.5427		
29944060	rs17879578	G	0.8409	A	0.1591		
29944067	rs9260170	T	0.4573	A	0.5427		
29944080	rs70944751	G	0.9267	T	0.0733		
29944095	rs374722929	C	0.9962	G	0.0038		
29944102	rs9260171	T	0.4573	G	0.5427		
29944103	rs17885212	T	0.6795	A	0.3205		
29944116	rs9260172	C	0.4573	T	0.5427		
29944118	rs9260173	T	0.4573	A	0.5427		

29944124	rs1136741	C	0.5306	G	0.4694		
29944131	rs41557315	A	0.9777	G	0.0223		
29944132	rs1136746	G	0.6795	A	0.3205		
29944135	rs1136749	A	0.4573	G	0.5427		
29944144	rs1136754	C	0.4573	T	0.5427		
29944147	rs45520243	C	0.9992	T	0.0008		
29944150	rs41554113	C	0.9996	T	0.0004		
29944151	rs1059563	C	0.4573	G	0.5427		
29944153	rs1059564	C	0.4573	T	0.5427		
29944154	rs9260179	A	0.4577	G	0.5423		
29944168	rs12721753	G	0.6795	A	0.3205		
29944174	rs1059570	C	0.8371	T	0.1629		
29944180	rs12721755	G	0.8409	A	0.1591		
29944193	rs9260180	G	0.4528	A	0.5472		
29944248	rs41551612	C	0.9996	T	0.0004		
29944264	rs9260181	C	0.7528	T	0.2472		
29944265	rs150028516	G	0.9996	A	0.0004		
29944281	rs41540417	C	0.9992	A	0.0008		
29944291	rs1059617	A	0.9471	G	0.0529		
29944308	rs148794540	C	0.9407	T	0.0593		
29944309	rs199474634	G	0.8405	A	0.1591	C	0.0004
29944310	rs1059632	G	0.7751	T	0.2249		
29944321	rs1059638	G	0.8371	A	0.1629		
29944331	rs2231095	G	0.4528	C	0.5435	A	0.0038
29944337	rs1059644	C	0.9962	A	0.0038		
29944370	rs1059686	C	0.6795	T	0.3205		
29944372	rs114818671	G	0.9278	C	0.0722		
29944376	rs1136835	A	0.9962	G	0.0038		
29944411	rs9260184	T	0.6187	C	0.3813		
29944427	rs2276447	C	0.4577	T	0.5423		
29944431	rs9260186	G	0.9539	C	0.0461		
29944450	rs146444598	TGGA	0.4577	T	0.5423		
29944472	rs188245242	G	0.9962	A	0.0038		
29944503	rs1136903	T	0.2948	C	0.7052		
29944504	rs1136917	G	0.8371	A	0.1629		
29944520	rs1136949	A	0.8409	G	0.1591		
29944524	rs41558618	C	0.9482	A	0.0518		
29944528	rs41555922	C	0.9962	T	0.0038		
29944549	rs1059846	G	0.8386	A	0.1614		
29944556	rs3179982	C	0.4573	T	0.5427		

29944565	rs1059896	G	0.9539	A	0.0461		
29944568	rs3180278	A	0.8386	T	0.1614		
29944571	rs1137078	A	0.7740	G	0.2260		
29944591	rs1141891	C	0.2978	T	0.7022		
29944596	rs1137110	T	0.8386	G	0.1614		
29944605	rs41550119	G	0.9958	A	0.0042		
29944609	rs1137160	G	0.8409	C	0.1591		
29944618	rs376806238	T	0.8386	TGG	0.1614		
29944621	rs9260191	A	0.4577	T	0.1614	G	0.3810
29944628	rs9260192	G	0.8511	A	0.1489		
29944633	rs41551619	G	0.9467	A	0.0533		
29944636	rs9278466	TG	0.4577	T	0.5423		
29944674	rs9260193	C	0.4569	A	0.5427	T	0.0004
29944680	rs9260194	C	0.4573	G	0.5427		
29944681	rs554587009	T	0.9996	A	0.0004		
29944686	rs9260195	A	0.4573	G	0.5427		
29944696	rs9260196	T	0.4573	C	0.5427		
29944700	rs9260197	A	0.4573	G	0.5427		
29944703	rs183484257	A	0.9962	C	0.0038		
29944716	rs1655894	G	0.0949	A	0.7593	T	0.1459
29944717	rs2975059	T	0.6757	C	0.3243		
29944725	rs9260198	C	0.4573	T	0.5427		
29944729	rs9260199	G	0.9392	A	0.0608		
29944733	rs1632885	T	0.6912	G	0.3050	C	0.0038
29944740	rs1655895	T	0.0949	C	0.9051		
29944749	rs146309471	T	0.9981	C	0.0019		
29944785	rs17880291	T	0.8371	G	0.1629		
29944787	rs17885126	A	0.8371	C	0.1629		
29944817	rs17882753	C	0.7657	T	0.2343		
29944818	rs2975058	G	0.6799	A	0.3201		
29944821	rs41551518	G	0.9679	A	0.0321		
29944825	.	A	0.9992	G	0.0008		
29944836	rs189337983	T	0.9974	A	0.0026		
29944837	rs41545117	C	0.9467	A	0.0533		
29944872	rs17887246	G	0.8409	C	0.1587	T	0.0004
29944875	rs67131684	CAGA	0.4573	C	0.5427		
29944879	rs17885779	C	0.8394	T	0.1591	*	0.0015
29944880	rs9260200	C	0.4573	T	0.5427		
29944891	rs9260201	T	0.2944	G	0.7056		
29944892	rs9260202	A	0.4535	G	0.5465		

29944903	rs17884187	A	0.8409	G	0.1591		
29944910	rs17883986	C	0.8409	T	0.1591		
29944936	rs17879496	C	0.8360	T	0.1640		
29944951	rs141912766	G	0.9962	T	0.0038		
29944961	rs150666121	T	0.9913	A	0.0087		
29944977	rs3823323	G	0.4573	A	0.5427		
29944986	rs548326797	A	0.9962	G	0.0038		
29944989	rs3823324	A	0.4573	T	0.5427		
29944990	rs3823325	G	0.9962	C	0.0038		
29945019	rs2735103	G	0.6784	A	0.3216		
29945033	rs184394934	T	0.9962	G	0.0038		
29945038	rs200725098	C	0.9974	T	0.0026		
29945053	rs2571381	T	0.1485	C	0.8515		
29945059	rs41551512	A	0.9962	T	0.0038		
29945075	rs1137296	T	0.2944	C	0.7056		
29945079	rs2231119	A	0.2944	T	0.7056		
29945108	rs139910379	C	0.9267	A	0.0733		
29945113	rs41546014	G	0.9267	C	0.0733		
29945115	rs3823329	C	0.9267	A	0.0733		
29945116	rs3823330	T	0.9267	A	0.0733		
29945119	rs41549018	G	0.9267	A	0.0733		
29945120	rs41555715	G	0.9267	A	0.0733		
29945123	rs41558113	C	0.9267	A	0.0733		
29945124	rs41561414	T	0.9267	C	0.0733		
29945137	rs541705597	T	0.9962	G	0.0038		
29945148	rs2735102	A	0.7778	G	0.2222		
29945158	rs2916802	C	0.0714	T	0.9286		
29945161	rs147947139	C	0.8409	A	0.1591		
29945167	rs2916803	T	0.0949	C	0.9051		
29945179	rs535529462	C	0.9996	T	0.0004		
29945191	rs3823339	C	0.9471	G	0.0529		
29945202	rs902382249	T	0.9996	A	0.0004		
29945224	rs2735101	T	0.7782	C	0.2218		
29945230	rs200140272	C	0.9951	T	0.0049		
29945260	rs1137631	G	0.7778	A	0.2222		
29945265	rs14288	C	0.0949	T	0.9051		
29945290	rs3823342	T	0.4535	C	0.5465		
29945297	rs9260205	A	0.1485	G	0.8515		
29945301	rs41546815	A	0.1485	G	0.8515		
29945321	rs3129015	C	0.9036	G	0.0038	T	0.0926

29945322	rs41553916	G	0.9690	A	0.0310		
29945334	rs3823343	A	0.7778	G	0.2222		
29945350	rs367651478	G	0.9992	A	0.0008		
29945358	rs28749156	G	0.6795	C	0.3205		
29945360	rs3132686	A	0.9074	G	0.0926		
29945370	.	T	0.9996	C	0.0004		
29945394	rs2975057	G	0.2944	A	0.7056		
29945398	rs114353087	T	0.9958	A	0.0042		
29945402	rs2975056	C	0.7778	T	0.2222		
29945407	rs549086393	A	0.9996	G	0.0004		
29945414	rs2975055	T	0.2944	C	0.7018	G	0.0038
29945415	rs9260210	G	0.2944	A	0.7018	T	0.0038
29945420	rs2975054	T	0.4573	C	0.5427		
29945424	rs41557912	T	0.8409	G	0.1591		
29945436	rs35830199	TA	0.8409	T	0.1591		
29945437	rs45497894	AT	0.8371	*	0.1591	A	0.0038
29945489	rs1061156	G	0.7778	T	0.2222		
29945521	rs1061235	A	0.9267	T	0.0733		
29945566	rs578027483	C	0.9962	G	0.0038		
29945567	rs13395	G	0.6187	T	0.2222	A	0.1591
29945588	rs181754482	G	0.9962	A	0.0038		
29945590	rs113042145	G	0.9962	A	0.0038		
29945597	rs1061815	C	0.9074	G	0.0926		
29945609	rs41545022	C	0.9467	G	0.0533		
29945653	rs112863520	C	0.8541	T	0.1459		
29945657	rs186532503	A	0.9962	G	0.0038		
29945672	rs66585529	C	0.8409	G	0.1591		
29945706	rs79244404	T	0.7778	C	0.2222		
29945724	rs144149732	C	0.9962	T	0.0038		
29945738	rs112525507	C	0.9354	T	0.0646		
29945762	rs577937205	C	0.9962	A	0.0038		
29945765	rs2499	T	0.0949	G	0.9051		
29945777	rs9385	A	0.9974	C	0.0026		
29945811	rs573809399	G	0.9962	C	0.0038		
29945853	rs2975053	A	0.8537	T	0.0926	C	0.0537
29945861	rs13488	G	0.1485	C	0.8515		
29945862	rs2975052	A	0.8541	G	0.1459		

Table S4. List of all *HLA-G/HLA-A* haplotypes observed in 1323 Brazilians from São Paulo and their frequencies.

HLA-G genomic allele	HLA-A genomic allele	Haplotype Frequency	(continue) HLA-G genomic allele	HLA-A genomic allele	Haplotype Frequency
G*01:01:01:01	A*02:01:01:01	0.2002	G*01:03:01:02	A*31:15	0.0011
G*01:01:01:05	A*03:01:01:01	0.0655	G*01:04:01:01	A*32:01:01:01	0.0011
G*01:04:01:01	A*24:02:01:01	0.0616	G*01:04:04	unknown	0.0011
G*01:04:04	A*23:01:01:01	0.0517	G*01:11	A*23:17:01:01	0.0011
G*01:01:03:03	A*11:01:01:01	0.0505	unknown	A*29:02:01:02	0.0011
G*01:06	A*01:01:01:01	0.0498	unknown	A*31:01:02:01	0.0011
G*01:01:02:01	A*01:01:01:01	0.0337	G*01:01:01:01	A*02:09:01:01	0.0008
G*01:01:01:04	A*29:02:01:01	0.0272	G*01:01:01:01	A*02:12	0.0008
G*01:03:01:02	A*31:01:02:01	0.0268	G*01:01:01:01	A*24:02:13	0.0008
G*01:01:02:01	A*26:01:01:01	0.0260	G*01:01:02:01	A*24:03:01:01	0.0008
G*01:03:01:02	A*33:01:01:01	0.0211	G*01:01:02:01	A*69:01:01:01	0.0008
G*01:01:02:01	A*68:02:01:01	0.0191	G*01:01:03:03	A*11:12	0.0008
G*01:01:01:08	A*30:02:01:01	0.0176	G*01:04:04	A*02:05:01:01	0.0008
G*01:05N	A*30:01:01:01	0.0176	G*01:05N	unknown	0.0008
G*01:01:02:01	A*68:01:01:02	0.0165	G*01:08:02	A*32:01:01:01	0.0008
G*01:01:01:01	A*32:01:01:01	0.0161	unknown	A*03:01:01:01	0.0008
G*01:01:01:05	A*03:01:01:05	0.0142	unknown	A*03:01:01:07	0.0008
G*01:03:01:02	A*02:05:01:01	0.0138	unknown	A*33:03:01:01	0.0008
G*01:01:02:01	A*24:02:01:01	0.0134	unknown	unknown	0.0008
G*01:01:02:01	A*68:01:02:01	0.0115	G*01:01:01:01	A*02:01:01:08	0.0004
G*01:01:01:04	A*74:01:01:01	0.0103	G*01:01:01:01	A*02:01:04	0.0004
G*01:01:02:01	A*25:01:01:01	0.0088	G*01:01:01:01	A*02:01:15	0.0004
G*01:01:01:01	A*02:11:01:01	0.0084	G*01:01:01:01	A*02:07:01:01	0.0004
G*01:01:22:01	A*32:01:01:01	0.0084	G*01:01:01:01	A*02:20:01	0.0004
G*01:01:01:04	A*29:02:01:02	0.0077	G*01:01:01:01	A*02:221	0.0004
G*01:01:02:01	A*68:01:02:02	0.0077	G*01:01:01:01	A*02:334	0.0004
G*01:04:01:01	A*24:03:01:01	0.0061	G*01:01:01:01	A*02:66	0.0004
G*01:01:02:02	A*66:01:01:01	0.0057	G*01:01:01:01	A*23:01:01:01	0.0004
G*01:03:01:02	A*02:02:01:01	0.0057	G*01:01:01:01	A*32:02	0.0004
unknown	A*30:02:01:02	0.0057	G*01:01:01:04	unknown	0.0004
G*01:04:01:01	A*33:03:01:01	0.0050	G*01:01:01:05	A*01:01:01:01	0.0004
G*01:01:02:01	A*02:01:01:01	0.0046	G*01:01:01:05	A*29:02:07	0.0004
G*01:01:17	A*32:01:01:01	0.0046	G*01:01:01:08	A*30:02:01:05	0.0004
G*01:04:04	A*23:17:01:01	0.0046	G*01:01:01:08	unknown	0.0004
G*01:01:01:05	A*03:02:01	0.0042	G*01:01:02:01	A*02:52	0.0004
G*01:01:01:08	A*30:02:01:03	0.0042	G*01:01:02:01	A*26:01:01:13	0.0004
G*01:04:01:01	A*24:02:01:04	0.0042	G*01:01:02:01	A*30:02:01:01	0.0004
G*01:01:01:06	A*80:01:01:02	0.0038	G*01:01:02:01	A*31:01:13	0.0004
G*01:01:15	A*03:01:01:05	0.0038	G*01:01:02:01	A*31:04:01:02	0.0004

G*01:04:04	A*36:01	0.0038	G*01:01:02:01	A*68:01:02:04	0.0004
unknown	A*29:01:01:01	0.0038	G*01:01:02:01	A*68:02:02	0.0004
G*01:01:01:01	A*02:04	0.0034	G*01:01:02:01	A*68:17:01	0.0004
G*01:01:02:01	A*26:08:01:01	0.0034	G*01:01:02:01	A*74:01:01:01	0.0004
G*01:03:01:02	A*34:02:01:01	0.0034	G*01:01:02:02	unknown	0.0004
unknown	A*02:01:01:01	0.0034	G*01:01:03:03	A*11:01:01:06	0.0004
unknown	A*03:01:01:05	0.0034	G*01:01:03:03	A*11:02:01:01	0.0004
G*01:01:01:01	A*02:06:01:01	0.0031	G*01:01:03:03	A*24:02:01:01	0.0004
G*01:01:01:01	A*24:02:01:01	0.0031	G*01:01:03:03	A*68:01:02:04	0.0004
G*01:01:02:01	A*31:01:02:01	0.0031	G*01:01:19	A*01:01:01:01	0.0004
G*01:03:01:02	A*02:02:01:02	0.0031	G*01:01:22:01	A*02:01:01:01	0.0004
G*01:04:01:02	A*24:02:01:01	0.0031	G*01:01:22:01	A*11:01:01:01	0.0004
unknown	A*34:02:01:01	0.0031	G*01:01:22:01	A*32:01:01:09	0.0004
G*01:01:01:09	A*02:01:01:01	0.0027	G*01:03:01:02	A*02:02:01:04	0.0004
G*01:01:14	A*31:01:02:04	0.0027	G*01:03:01:02	A*02:14	0.0004
G*01:03:01:02	unknown	0.0027	G*01:03:01:02	A*03:01:01:01	0.0004
G*01:04:01:01	A*24:02:01:05	0.0027	G*01:03:01:02	A*31:01:13	0.0004
G*01:04:01:01	A*33:03:01:03	0.0027	G*01:04:01:01	A*02:05:01:01	0.0004
G*01:01:01:01	A*31:01:02:01	0.0023	G*01:04:01:01	A*24:02:01:12	0.0004
G*01:01:02:01	unknown	0.0023	G*01:04:01:01	A*24:02:01:36	0.0004
G*01:04:01:01	unknown	0.0023	G*01:04:01:01	A*24:03:01:02	0.0004
G*01:01:12	A*32:01:01:01	0.0019	G*01:04:01:01	A*33:03:23	0.0004
G*01:04:04	A*01:01:01:01	0.0019	G*01:04:04	A*03:01:01:05	0.0004
G*01:04:05	A*33:03:01:01	0.0019	G*01:04:04	A*23:05	0.0004
unknown	A*30:01:01:01	0.0019	G*01:04:05	A*02:01:01:71	0.0004
unknown	A*30:04:01:01	0.0019	G*01:06	A*01:01:01:23	0.0004
G*01:01:01:01	A*02:01:01:05	0.0015	G*01:06	A*26:01:01:01	0.0004
G*01:01:01:01	A*11:01:01:01	0.0015	G*01:06	unknown	0.0004
G*01:01:01:05	A*03:01:01:03	0.0015	G*01:08:02	unknown	0.0004
G*01:01:19	A*66:01:01:01	0.0015	G*01:09	A*01:01:01:01	0.0004
G*01:05N	A*30:95	0.0015	G*01:21N	A*24:02:01:01	0.0004
unknown	A*24:02:01:01	0.0015	unknown	A*02:01:01:08	0.0004
G*01:01:01:01	unknown	0.0011	unknown	A*02:17:02:01	0.0004
G*01:01:01:04	A*29:02:01:03	0.0011	unknown	A*24:104	0.0004
G*01:01:01:05	A*03:01:01:07	0.0011	unknown	A*26:01:01:01	0.0004
G*01:01:01:06	A*66:02	0.0011	unknown	A*29:02:01:01	0.0004
G*01:01:02:01	A*01:02	0.0011	unknown	A*30:02:01:01	0.0004
G*01:01:02:01	A*26:02:01	0.0011	unknown	A*30:29	0.0004
G*01:01:02:01	A*26:03:01:01	0.0011	unknown	A*32:01:01:01	0.0004
G*01:01:02:01	A*68:02:01:03	0.0011	unknown	A*68:01:02:01	0.0004
G*01:01:14	A*31:01:02:01	0.0011			

Table S5. List of all *HLA-G/HLA-A* haplotypes observed in 1323 Brazilians from São Paulo and their frequencies.

HLA-F genomic allele	HLA-G genomic allele	HLA-A genomic allele	Haplotype Frequency
F*01:01:01:01	G*01:01:01:01	A*02:01:01:01	0.102986217
F*01:03:01:01	G*01:01:01:05	A*03:01:01:01	0.062787136
F*01:01:01:09	G*01:06	A*01:01:01:01	0.047856049
F*01:03:01:03	G*01:04:04	A*23:01:01:01	0.046324655
F*01:01:02:09	G*01:01:03:03	A*11:01:01:01	0.045941807
F*01:01:01:08	G*01:01:01:01	A*02:01:01:01	0.04517611
F*01:01:01:09	G*01:04:01:01	A*24:02:01:01	0.032542113
F*01:01:01:09	G*01:01:02:01	A*01:01:01:01	0.03101072
F*01:01:01:17	G*01:01:01:01	A*02:01:01:01	0.029862175
F*01:01:01:08	G*01:01:01:04	A*29:02:01:01	0.025650842
F*01:01:01:08	G*01:01:02:01	A*26:01:01:01	0.024119449
F*01:01:02:07	G*01:03:01:02	A*33:01:01:01	0.019908116
F*01:01:01:08	G*01:04:01:01	A*24:02:01:01	0.018759571
F*01:01:01:18	G*01:01:02:01	A*68:02:01:01	0.017993874
F*01:01:01:09	G*01:05N	A*30:01:01:01	0.016462481
F*01:01:01:17	G*01:01:01:08	A*30:02:01:01	0.015696784
F*01:01:01:01	G*01:01:01:01	A*32:01:01:01	0.015313936
F*01:01:01:11	G*01:03:01:02	A*31:01:02:01	0.014548239
F*01:01:02:11	G*01:03:01:02	A*02:05:01:01	0.013399694
F*01:01:01:18	G*01:01:02:01	A*24:02:01:01	0.012633997
F*01:01:01:18	G*01:01:02:01	A*68:01:01:02	0.012633997
F*01:01:01:09	G*01:03:01:02	A*31:01:02:01	0.010336907
F*01:01:02:08	G*01:01:01:04	A*74:01:01:01	0.010336907
unknown	G*01:01:01:01	A*02:01:01:01	0.010336907
F*01:01:01:08	G*01:01:02:01	A*25:01:01:01	0.008805513
F*01:01:02:08	G*01:01:02:01	A*68:01:02:01	0.008422665
F*01:01:01:01	G*01:01:01:01	A*02:11:01:01	0.008039816
F*01:01:02:10	G*01:01:22:01	A*32:01:01:01	0.008039816
F*01:01:02:10	G*01:01:01:05	A*03:01:01:05	0.007274119
F*01:03:01:01	G*01:04:01:01	A*24:02:01:01	0.006508423
F*01:01:01:09	G*01:04:01:01	A*24:03:01:01	0.006125574
F*01:01:02:08	G*01:01:02:01	A*68:01:02:02	0.005742726
F*01:01:01:08	G*01:01:02:02	A*66:01:01:01	0.004977029
F*01:04:01:02	G*01:01:01:01	A*02:01:01:01	0.004977029
F*01:01:01:09	G*01:01:02:01	A*02:01:01:01	0.004211332
F*01:01:01:09	G*01:04:01:01	A*24:02:01:04	0.004211332
F*01:01:02:10	G*01:01:17	A*32:01:01:01	0.004211332
F*01:01:02:10	G*01:04:01:01	A*33:03:01:01	0.004211332
F*01:03:01:01	G*01:01:01:05	A*03:01:01:05	0.004211332
F*01:03:01:01	G*01:01:01:05	A*03:02:01	0.004211332
unknown	G*01:04:04	A*23:17:01:01	0.004211332
unknown	unknown	A*30:02:01:02	0.004211332
F*01:01:01:17	G*01:01:01:08	A*30:02:01:03	0.003828484
F*01:01:01:18	G*01:01:15	A*03:01:01:05	0.003828484
F*01:01:02:08	unknown	A*29:01:01:01	0.003828484
F*01:03:01:03	G*01:04:04	A*36:01	0.003828484
unknown	G*01:01:01:04	A*29:02:01:02	0.003828484
unknown	G*01:01:01:06	A*80:01:01:02	0.003828484
F*01:01:01:08	G*01:01:02:01	A*26:08:01:01	0.003445636
unknown	G*01:01:01:01	A*02:04	0.003445636
unknown	G*01:03:01:02	A*34:02:01:01	0.003445636
F*01:01:01:01	G*01:01:01:01	A*24:02:01:01	0.003062787
F*01:01:02:11	G*01:03:01:02	A*02:02:01:01	0.003062787
unknown	G*01:01:02:01	A*68:01:02:01	0.003062787
unknown	G*01:03:01:02	A*02:02:01:02	0.003062787

F*01:01:01:01	unknown	A*02:01:01:01	0.002679939
F*01:01:01:08	G*01:01:14	A*31:01:02:04	0.002679939
F*01:01:01:09	G*01:04:01:01	A*24:02:01:05	0.002679939
F*01:01:02:10	G*01:04:01:02	A*24:02:01:01	0.002679939
unknown	G*01:01:03:03	A*11:01:01:01	0.002679939
F*01:01:01:01	G*01:01:01:09	A*02:01:01:01	0.00229709
F*01:01:01:08	G*01:01:02:01	A*01:01:01:01	0.00229709
F*01:01:01:08	G*01:04:04	A*23:01:01:01	0.00229709
F*01:01:01:18	G*01:01:01:01	A*02:01:01:01	0.00229709
F*01:01:01:18	G*01:01:01:04	A*29:02:01:02	0.00229709
F*01:01:02:10	G*01:04:01:01	A*33:03:01:03	0.00229709
unknown	G*01:01:02:01	A*31:01:02:01	0.00229709
unknown	unknown	A*34:02:01:01	0.00229709
F*01:01:01:01	G*01:01:01:01	A*31:01:02:01	0.001914242
F*01:01:01:08	G*01:01:02:01	A*68:01:01:02	0.001914242
F*01:01:01:09	G*01:04:01:01	unknown	0.001914242
F*01:01:01:13	G*01:01:01:01	A*02:06:01:01	0.001914242
F*01:01:01:18	G*01:01:01:05	A*03:01:01:05	0.001914242
F*01:01:02:10	G*01:01:12	A*32:01:01:01	0.001914242
F*01:01:02:10	G*01:04:05	A*33:03:01:01	0.001914242
unknown	G*01:03:01:02	A*31:01:02:01	0.001914242
unknown	unknown	A*30:04:01:01	0.001914242
F*01:01:01:01	G*01:01:01:01	A*02:01:01:05	0.001531394
F*01:01:01:01	G*01:01:01:01	A*11:01:01:01	0.001531394
F*01:01:01:01	G*01:01:01:08	A*30:02:01:01	0.001531394
F*01:01:01:08	G*01:01:19	A*66:01:01:01	0.001531394
F*01:01:01:08	unknown	A*03:01:01:05	0.001531394
F*01:01:01:09	unknown	A*30:01:01:01	0.001531394
F*01:01:02:07	G*01:03:01:02	unknown	0.001531394
F*01:01:02:08	G*01:01:01:04	A*29:02:01:02	0.001531394
F*01:03:01:01	G*01:01:01:01	A*02:01:01:01	0.001531394
unknown	G*01:01:01:05	A*03:01:01:01	0.001531394
unknown	G*01:01:01:05	A*03:01:01:03	0.001531394
unknown	G*01:01:02:01	A*68:01:01:02	0.001531394
unknown	G*01:04:01:01	A*24:02:01:01	0.001531394
unknown	unknown	A*03:01:01:05	0.001531394
F*01:01:01:01	G*01:01:02:01	A*68:01:02:02	0.001148545
F*01:01:01:01	G*01:01:03:03	A*11:01:01:01	0.001148545
F*01:01:01:01	G*01:04:01:01	A*24:02:01:01	0.001148545
F*01:01:01:08	G*01:01:02:01	A*26:02:01	0.001148545
F*01:01:01:08	G*01:01:02:01	A*26:03:01:01	0.001148545
F*01:01:01:08	G*01:01:14	A*31:01:02:01	0.001148545
F*01:01:01:09	G*01:01:01:01	A*02:01:01:01	0.001148545
F*01:01:01:09	G*01:01:02:01	A*01:02	0.001148545
F*01:01:01:09	G*01:04:01:01	A*32:01:01:01	0.001148545
F*01:01:01:09	G*01:05N	A*30:95	0.001148545
F*01:01:01:09	unknown	A*30:02:01:02	0.001148545
F*01:01:01:09	unknown	A*31:01:02:01	0.001148545
F*01:01:01:11	G*01:03:01:02	A*31:15	0.001148545
F*01:01:01:18	G*01:01:02:01	A*68:02:01:03	0.001148545
F*01:01:01:18	G*01:04:04	A*23:01:01:01	0.001148545
F*01:01:02:08	G*01:01:01:04	A*29:02:01:01	0.001148545
F*01:01:02:08	G*01:01:01:04	A*29:02:01:03	0.001148545
F*01:01:02:08	G*01:01:01:06	A*66:02	0.001148545
F*01:01:02:08	unknown	A*29:02:01:02	0.001148545
F*01:01:04	G*01:01:01:01	A*02:01:01:01	0.001148545
F*01:03:01:01	G*01:01:01:05	A*03:01:01:07	0.001148545
F*01:03:01:04	G*01:03:01:02	A*02:02:01:01	0.001148545
unknown	G*01:03:01:02	A*33:01:01:01	0.001148545
unknown	G*01:11	A*23:17:01:01	0.001148545
F*01:01:01:01	G*01:01:01:01	A*02:06:01:01	0.000765697
F*01:01:01:01	G*01:01:01:01	A*02:09:01:01	0.000765697

F*01:01:01:01	G*01:01:01:01	A*02:12	0.000765697
F*01:01:01:01	G*01:01:01:01	A*24:02:13	0.000765697
F*01:01:01:01	G*01:01:01:01	unknown	0.000765697
F*01:01:01:01	G*01:06	A*01:01:01:01	0.000765697
F*01:01:01:08	G*01:01:02:01	A*31:01:02:01	0.000765697
F*01:01:01:08	G*01:01:02:01	unknown	0.000765697
F*01:01:01:08	G*01:01:03:03	A*11:01:01:01	0.000765697
F*01:01:01:08	G*01:04:04	A*02:05:01:01	0.000765697
F*01:01:01:08	G*01:04:04	unknown	0.000765697
F*01:01:01:08	G*01:06	A*01:01:01:01	0.000765697
F*01:01:01:09	G*01:05N	unknown	0.000765697
F*01:01:01:09	unknown	A*24:02:01:01	0.000765697
F*01:01:01:17	G*01:01:02:01	A*68:02:01:01	0.000765697
F*01:01:01:18	G*01:01:02:01	A*24:03:01:01	0.000765697
F*01:01:01:18	G*01:01:02:01	A*69:01:01:01	0.000765697
F*01:01:01:18	G*01:01:02:01	unknown	0.000765697
F*01:01:02:08	G*01:03:01:02	A*02:02:01:01	0.000765697
F*01:01:02:08	G*01:04:04	A*23:01:01:01	0.000765697
F*01:01:02:08	unknown	A*34:02:01:01	0.000765697
F*01:01:02:09	G*01:01:03:03	A*11:12	0.000765697
F*01:01:02:10	G*01:04:04	A*01:01:01:01	0.000765697
F*01:01:02:10	G*01:08:02	A*32:01:01:01	0.000765697
F*01:01:02:11	G*01:03:01:02	unknown	0.000765697
F*01:03:01:01	G*01:01:02:02	A*66:01:01:01	0.000765697
F*01:03:01:01	unknown	A*03:01:01:01	0.000765697
F*01:03:01:01	unknown	A*03:01:01:07	0.000765697
F*01:03:01:03	G*01:04:04	A*01:01:01:01	0.000765697
F*01:04:01:02	G*01:04:01:01	A*24:02:01:01	0.000765697
unknown	G*01:01:01:05	A*03:01:01:05	0.000765697
unknown	G*01:01:02:01	A*26:01:01:01	0.000765697
unknown	G*01:04:04	A*23:01:01:01	0.000765697
F*01:01:01:01	G*01:01:01:01	A*02:07:01:01	0.000382848
F*01:01:01:01	G*01:01:01:01	A*02:20:01	0.000382848
F*01:01:01:01	G*01:01:01:01	A*02:221	0.000382848
F*01:01:01:01	G*01:01:01:01	A*02:334	0.000382848
F*01:01:01:01	G*01:01:01:01	A*02:66	0.000382848
F*01:01:01:01	G*01:01:01:01	A*23:01:01:01	0.000382848
F*01:01:01:01	G*01:01:01:01	A*32:02	0.000382848
F*01:01:01:01	G*01:01:02:01	A*26:01:01:01	0.000382848
F*01:01:01:01	G*01:04:01:01	A*24:03:01:02	0.000382848
F*01:01:01:01	G*01:05N	A*30:95	0.000382848
F*01:01:01:01	unknown	A*02:17:02:01	0.000382848
F*01:01:01:01	unknown	A*24:02:01:01	0.000382848
F*01:01:01:08	G*01:01:01:01	A*02:01:01:08	0.000382848
F*01:01:01:08	G*01:01:01:01	A*32:01:01:01	0.000382848
F*01:01:01:08	G*01:01:01:04	unknown	0.000382848
F*01:01:01:08	G*01:01:01:05	A*03:01:01:01	0.000382848
F*01:01:01:08	G*01:01:02:01	A*26:01:01:13	0.000382848
F*01:01:01:08	G*01:01:02:01	A*30:02:01:01	0.000382848
F*01:01:01:08	G*01:01:02:01	A*31:01:13	0.000382848
F*01:01:01:08	G*01:01:02:01	A*31:04:01:02	0.000382848
F*01:01:01:08	G*01:03:01:02	A*02:05:01:01	0.000382848
F*01:01:01:08	unknown	A*02:01:01:01	0.000382848
F*01:01:01:08	unknown	A*02:01:01:08	0.000382848
F*01:01:01:08	unknown	A*24:02:01:01	0.000382848
F*01:01:01:08	unknown	A*26:01:01:01	0.000382848
F*01:01:01:08	unknown	A*29:02:01:01	0.000382848
F*01:01:01:08	unknown	unknown	0.000382848
F*01:01:01:09	G*01:01:01:01	A*02:01:04	0.000382848
F*01:01:01:09	G*01:01:02:01	A*02:52	0.000382848
F*01:01:01:09	G*01:01:19	A*01:01:01:01	0.000382848
F*01:01:01:09	G*01:04:01:01	A*24:02:01:12	0.000382848

F*01:01:01:09	G*01:04:01:01	A*24:02:01:36	0.000382848
F*01:01:01:09	G*01:04:01:01	A*33:03:01:01	0.000382848
F*01:01:01:09	G*01:06	A*01:01:01:23	0.000382848
F*01:01:01:09	G*01:06	A*26:01:01:01	0.000382848
F*01:01:01:09	G*01:09	A*01:01:01:01	0.000382848
F*01:01:01:09	G*01:21N	A*24:02:01:01	0.000382848
F*01:01:01:09	unknown	A*02:01:01:01	0.000382848
F*01:01:01:09	unknown	A*24:104	0.000382848
F*01:01:01:09	unknown	unknown	0.000382848
F*01:01:01:11	G*01:01:01:05	A*03:01:01:01	0.000382848
F*01:01:01:11	G*01:01:02:01	A*68:01:02:02	0.000382848
F*01:01:01:11	G*01:03:01:02	A*31:01:13	0.000382848
F*01:01:01:17	G*01:01:01:01	A*02:01:15	0.000382848
F*01:01:01:17	G*01:01:01:01	unknown	0.000382848
F*01:01:01:17	G*01:01:01:04	A*29:02:01:01	0.000382848
F*01:01:01:17	G*01:01:01:05	A*03:01:01:01	0.000382848
F*01:01:01:17	G*01:01:01:08	A*30:02:01:05	0.000382848
F*01:01:01:17	G*01:01:01:08	unknown	0.000382848
F*01:01:01:17	G*01:01:02:01	A*26:01:01:01	0.000382848
F*01:01:01:17	G*01:01:17	A*32:01:01:01	0.000382848
F*01:01:01:17	G*01:05N	A*30:01:01:01	0.000382848
F*01:01:01:17	unknown	A*30:02:01:01	0.000382848
F*01:01:01:18	G*01:01:02:01	A*02:01:01:01	0.000382848
F*01:01:01:18	G*01:01:02:01	A*68:02:02	0.000382848
F*01:01:01:18	G*01:04:01:01	A*33:03:01:01	0.000382848
F*01:01:01:18	G*01:04:04	A*01:01:01:01	0.000382848
F*01:01:01:18	G*01:05N	A*30:01:01:01	0.000382848
F*01:01:01:18	unknown	A*03:01:01:05	0.000382848
F*01:01:01:19	G*01:01:02:01	A*24:02:01:01	0.000382848
F*01:01:01:21	G*01:01:01:01	A*02:01:01:01	0.000382848
F*01:01:02:08	G*01:01:01:01	A*02:01:01:01	0.000382848
F*01:01:02:08	G*01:01:02:01	A*24:02:01:01	0.000382848
F*01:01:02:08	G*01:01:02:01	A*68:01:01:02	0.000382848
F*01:01:02:08	G*01:01:02:01	A*68:01:02:04	0.000382848
F*01:01:02:08	G*01:01:02:01	A*68:17:01	0.000382848
F*01:01:02:08	G*01:01:02:01	unknown	0.000382848
F*01:01:02:08	unknown	A*68:01:02:01	0.000382848
F*01:01:02:09	G*01:01:03:03	A*11:01:01:06	0.000382848
F*01:01:02:09	G*01:01:03:03	A*11:02:01:01	0.000382848
F*01:01:02:09	G*01:01:03:03	A*24:02:01:01	0.000382848
F*01:01:02:09	G*01:04:01:01	A*24:02:01:01	0.000382848
F*01:01:02:10	G*01:01:22:01	A*02:01:01:01	0.000382848
F*01:01:02:10	G*01:01:22:01	A*11:01:01:01	0.000382848
F*01:01:02:10	G*01:01:22:01	A*32:01:01:09	0.000382848
F*01:01:02:10	G*01:04:01:01	A*02:05:01:01	0.000382848
F*01:01:02:10	G*01:04:01:01	A*33:03:23	0.000382848
F*01:01:02:10	G*01:04:01:01	unknown	0.000382848
F*01:01:02:10	G*01:04:05	A*02:01:01:71	0.000382848
F*01:01:02:10	G*01:06	unknown	0.000382848
F*01:01:02:10	G*01:08:02	unknown	0.000382848
F*01:01:02:10	unknown	A*32:01:01:01	0.000382848
F*01:01:02:10	unknown	A*33:03:01:01	0.000382848
F*01:01:02:11	G*01:03:01:02	A*02:02:01:04	0.000382848
F*01:01:02:11	G*01:03:01:02	A*02:14	0.000382848
F*01:01:02:11	G*01:03:01:02	A*03:01:01:01	0.000382848
F*01:01:02:11	unknown	A*30:02:01:02	0.000382848
F*01:01:05	G*01:01:02:01	A*01:01:01:01	0.000382848
F*01:01:05	G*01:01:02:01	A*26:01:01:01	0.000382848
F*01:03:01:01	G*01:01:01:05	A*01:01:01:01	0.000382848
F*01:03:01:01	G*01:01:01:05	A*29:02:07	0.000382848
F*01:03:01:01	G*01:01:01:09	A*02:01:01:01	0.000382848
F*01:03:01:01	G*01:01:02:02	unknown	0.000382848

F*01:03:01:01	G*01:03:01:02	A*02:02:01:01	0.000382848
F*01:03:01:01	G*01:04:04	A*23:01:01:01	0.000382848
F*01:03:01:03	G*01:04:04	A*03:01:01:05	0.000382848
F*01:03:01:03	G*01:04:04	A*23:17:01:01	0.000382848
F*01:03:01:03	G*01:04:04	unknown	0.000382848
F*01:05	G*01:01:01:01	A*02:06:01:01	0.000382848
unknown	G*01:01:01:01	A*02:11:01:01	0.000382848
unknown	G*01:01:01:01	A*31:01:02:01	0.000382848
unknown	G*01:01:01:01	A*32:01:01:01	0.000382848
unknown	G*01:01:01:08	A*30:02:01:01	0.000382848
unknown	G*01:01:01:08	A*30:02:01:03	0.000382848
unknown	G*01:01:02:01	A*68:01:02:02	0.000382848
unknown	G*01:01:02:01	A*68:02:01:01	0.000382848
unknown	G*01:01:02:01	A*74:01:01:01	0.000382848
unknown	G*01:01:02:01	unknown	0.000382848
unknown	G*01:01:03:03	A*68:01:02:04	0.000382848
unknown	G*01:01:22:01	A*32:01:01:01	0.000382848
unknown	G*01:03:01:02	A*02:02:01:01	0.000382848
unknown	G*01:03:01:02	unknown	0.000382848
unknown	G*01:04:01:01	A*33:03:01:03	0.000382848
unknown	G*01:04:01:02	A*24:02:01:01	0.000382848
unknown	G*01:04:04	A*23:05	0.000382848
unknown	G*01:05N	A*30:01:01:01	0.000382848
unknown	G*01:06	A*01:01:01:01	0.000382848
unknown	unknown	A*30:01:01:01	0.000382848
unknown	unknown	A*30:29	0.000382848
unknown	unknown	A*33:03:01:01	0.000382848

2.5. References

- ALVAREZ, M. et al. HLA-G 3'-UTR SNP and 14-bp deletion polymorphisms in Portuguese and Guinea-Bissau populations. **International Journal of Immunogenetics**, v. 36, n. 6, p. 361–366, 2009.
- ARNAIZ-VILLENA, A. et al. HLA-G, -A haplotypes in Amerindians (Ecuador): HLA-G*01:05N World distribution. **Human Immunology**, v. 79, n. 2, p. 89–90, 2018.
- BATZER, M. A.; DEININGER, P. L. Alu repeats and human genomic diversity. **Nature Reviews Genetics**, v. 3, n. 5, p. 370–379, 2002.
- BITARELLO, B. D.; FRANCISCO, R. DOS S.; MEYER, D. Heterogeneity of dN/dS Ratios at the Classical HLA Class I Genes over Divergence Time and Across the Allelic Phylogeny. **Journal of Molecular Evolution**, v. 82, n. 1, p. 38–50, 2016.
- BRANDT, D. Y. C. et al. Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 genomes project phase I data. **G3: Genes, Genomes, Genetics**, v. 5, n. 5, p. 931–941, 2015.
- BUHLER, S.; NUNES, J. M.; SANCHEZ-MAZAS, A. HLA class I molecular variation and peptide-binding properties suggest a model of joint divergent asymmetric selection. **Immunogenetics**, v. 68, n. 6–7, p. 401–416, 2016.
- BUTTURA, R. V. et al. HLA-F displays highly divergent and frequent haplotype lineages associated with different mRNA expression levels. **Human Immunology**, v. 80, n. 2, p. 112–119, fev. 2019.
- CARLINI, F. et al. HLA-G UTR haplotype conservation in the Malian population: Association with soluble HLA-G. **PLoS ONE**, v. 8, n. 12, 2013.
- CARLINI, F. et al. Association of HLA-A and non-classical HLA class I alleles. **PLoS ONE**, v. 11, n. 10, p. 1–17, 2016.
- CAROSELLA, E. D. et al. HLA-G molecules: from maternal-fetal tolerance to tissue acceptance. **Advances in immunology**, v. 81, n. 0, p. 199–252, 2003.
- CAROSELLA, E. D. et al. HLA-G: from biology to clinical benefits. **Trends in Immunology**, v. 29, n. 3, p. 125–132, 2008.
- CARVALHO DOS SANTOS, L. et al. HLA-E polymorphisms in an Afro-descendant Southern Brazilian population. **Human Immunology**, v. 74, n. 2, p. 199–202, 2013.
- CASTELLI, E. C. et al. A comprehensive study of polymorphic sites along the HLA-G gene: Implication for gene regulation and evolution. **Molecular Biology and Evolution**, v. 28, n. 11, p. 3069–3086, 2011.
- CASTELLI, E. C. et al. Insights into HLA-G genetics provided by worldwide haplotype diversity. **Frontiers in Immunology**, v. 5, n. OCT, 2014.
- CASTELLI, E. C. et al. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. **Human Immunology**,

v. 76, n. 12, p. 945–953, 2015.

CASTELLI, E. C. et al. HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographically distinct population samples of Brazil and Cyprus. **Molecular Immunology**, v. 83, p. 115–126, 2017.

CASTELLI, E. C. et al. Hla-mapper: An application to optimize the mapping of HLA sequences produced by massively parallel sequencing procedures. **Human Immunology**, v. 79, n. 9, p. 678–684, 2018.

CASTELLI, E. C.; MENDES, C. T.; DONADI, E. A. HLA-G alleles and HLA-G 14 bp polymorphisms in a Brazilian population. **Tissue Antigens**, v. 70, n. 1, p. 62–68, 2007.

CASTRO, M. J. et al. Homozygous HLA-G*0105N healthy individuals indicate that membrane-anchored HLA-G1 molecule is not necessary for survival. **Tissue Antigens**, v. 56, n. 3, p. 232–239, set. 2000.

CHEN, X. et al. Association of nsv823469 copy number loss with decreased risk of chronic obstructive pulmonary disease and pulmonary function in Chinese. **Scientific Reports**, v. 7, n. September 2016, p. 1–8, 2017.

CIRULLI, V. et al. The class I HLA repertoire of pancreatic islets comprises the nonclassical class Ib antigen HLA-G. **Diabetes**, v. 55, n. 5, p. 1214–1222, 2006.

DELANEAU, O.; MARCHINI, J.; ZAGURY, J. F. A linear complexity phasing method for thousands of genomes. **Nature Methods**, v. 9, n. 2, p. 179–181, 2011.

DEPRISTO, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. **Nature Genetics**, v. 43, n. 5, p. 491–501, 2011.

DONADI, E. A. et al. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. **Cellular and Molecular Life Sciences**, v. 68, n. 3, p. 369–395, 2011.

DOS SANTOS FRANCISCO, R. et al. HLA supertype variation across populations: new insights into the role of natural selection in the evolution of HLA-A and HLA-B polymorphisms. **Immunogenetics**, v. 67, n. 11–12, p. 651–663, 2015.

DUNN, D. S. et al. The association between HLA-A alleles and young Alu dimorphisms near the HLA-J, -H, and -F genes in workshop cell lines and Japanese and Australian populations. **Journal of Molecular Evolution**, v. 55, n. 6, p. 718–726, 2002.

DURMANOVA, V. et al. Association of HLA-G Polymorphisms in the 3'UTR Region and Soluble HLA-G with Kidney Graft Outcome. **Immunological Investigations**, v. 48, n. 6, p. 644–658, 2019.

EMMERY, J. et al. Associations between fetal HLA-G genotype and birth weight and placental weight in a large cohort of pregnant women – Possible implications for HLA diversity. **Journal of Reproductive Immunology**, v. 120, p. 8–14, 2017.

FISCHER, G. F.; MAYR, W. R. **Molecular genetics of the HLA complex** Wiener Klinische Wochenschrift, 2001.

GINEAU, L. et al. Balancing immunity and tolerance: Genetic footprint of natural selection in the transcriptional regulatory region of HLA-G. **Genes and Immunity**, v. 16, n. 1, p. 57–70, 2015.

GOODRIDGE, J. P. et al. HLA-F and MHC Class I Open Conformers Are Ligands for NK Cell Ig-like Receptors. **The Journal of Immunology**, v. 191, n. 7, p. 3553–3562, 2013.

HVIID, T. V. F. HLA-G in human reproduction: Aspects of genetics, function and pregnancy complications. **Human Reproduction Update**, v. 12, n. 3, p. 209–232, 2006.

ISHIGAMI, S. et al. Human leukocyte antigen (HLA)-E and HLA-F expression in gastric cancer. **Anticancer Research**, v. 35, n. 4, p. 2279–2285, 2015.

JUCAUD, V. et al. Serum antibodies to human leucocyte antigen (HLA)-E, HLA-F and HLA-G in patients with systemic lupus erythematosus (SLE) during disease flares: Clinical relevance of HLA-F autoantibodies. **Clinical and Experimental Immunology**, v. 183, n. 3, p. 326–340, 2016.

JURKA, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 5, p. 1872–1877, 1997.

KING, A. et al. Recognition of trophoblast HLA class I molecules by decidual NK cell receptors - A review. **Placenta**, v. 21, n. SUPPL.1, p. 81–85, 2000.

KIRWAN, J. P. et al. TNF- Is a Predictor of Insulin Resistance in Human Pregnancy. **Diabetes**, v. 51, n. 7, p. 2207–2213, 1 jul. 2002.

KLEIN, J.; SATO, A. Advances in immunology: The HLA system (Second of two parts). **New England Journal of Medicine**, v. 343, n. 11, p. 782–786, 2000a.

KLEIN, J.; SATO, A. The HLA System (First of Two Parts). **New England Journal of Medicine**, v. 343, n. 10, p. 702–709, 7 set. 2000b.

KULSKI, J. K. et al. The association between HLA-A alleles and an Alu dimorphism near HLA-G. **Journal of Molecular Evolution**, v. 53, n. 2, p. 114–123, 2001.

LE DISCORDE, M. et al. HLA-G*0105N Null Allele Encodes Functional HLA-G Isoforms1. **Biology of Reproduction**, v. 73, n. 2, p. 280–288, 2005.

LEPIN, E. J. M. et al. Functional characterization of HLA-F and binding of HLA-F tetramers to ILT2 and ILT4 receptors. **European Journal of Immunology**, v. 30, n. 12, p. 3552–3561, 2000.

LIMA, T. H. A. et al. HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. **Human Immunology**, v. 77, n. 10, p. 841–853, 2016.

LIMA, T. H. A. et al. HLA-A promoter, coding, and 3'UTR sequences in a Brazilian cohort, and their evolutionary aspects. **Hla**, v. 93, n. 2–3, p. 65–79, 2019.

MACK, S. J. et al. HLA-A, -B, -C, and -DRB1 allele and haplotype frequencies distinguish

Eastern European Americans from the general European American population. **Tissue Antigens**, v. 73, n. 1, p. 17–32, 2009.

MANASTER, I. et al. MiRNA-mediated control of HLA-G expression and function. **PLoS ONE**, v. 7, n. 3, p. 1–9, 2012.

MARTELLI-PALOMINO, G. et al. Polymorphic Sites at the 3' Untranslated Region of the HLA-G Gene Are Associated with Differential hla-g Soluble Levels in the Brazilian and French Population. **PLoS ONE**, v. 8, n. 10, p. 1–10, 2013.

MATHIAS, S. L. et al. Reverse transcriptase encoded by a human transposable element. **Science**, v. 254, n. 5039, p. 1808–1810, 1991.

MCKENNA, A. et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Research**, v. 20, n. 9, p. 1297–1303, 2010.

MENDES-JUNIOR, C. T. et al. Genetic diversity of the HLA-G coding region in Amerindian populations from the Brazilian Amazon: A possible role of natural selection. **Genes and Immunity**, v. 14, n. 8, p. 518–526, 2013.

MEYER, D.; THOMSON, G. How selection shapes variation of the human major histocompatibility complex: A review. **Annals of Human Genetics**, v. 1996, n. 1, p. 1–26, 2001.

MIRETTI, M. M. et al. A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. **American Journal of Human Genetics**, v. 76, n. 4, p. 634–646, 2005.

MOREAU, P.; FLAJOLLET, S.; CAROSELLA, E. D. Non-classical transcriptional regulation of HLA-G: An update. **Journal of Cellular and Molecular Medicine**, v. 13, n. 9 B, p. 2973–2989, 2009.

NILSSON, L. L. et al. Next-generation sequencing of HLA-G based on long-range polymerase chain reaction. **Hla**, v. 92, n. 3, p. 144–153, 2018.

NOBLE, J. A. et al. HLA class I and genetic susceptibility to type 1 diabetes: Results from the type 1 diabetes genetics consortium. **Diabetes**, v. 59, n. 11, p. 2972–2979, 2010.

OBER, C. et al. The miscarriage-associated HLA-G -725G allele influences transcription rates in JEG-3 cells. **Human Reproduction**, v. 21, n. 7, p. 1743–1748, 2006.

OLIVEIRA, M. L. G. DE et al. Extended HLA-G genetic diversity and ancestry composition in a Brazilian admixed population sample: Implications for HLA-G transcriptional control and for case-control association studies. **Human Immunology**, v. 79, n. 11, p. 790–799, 2018a.

OLIVEIRA, P. et al. Genome-wide burden and association analyses implicate copy number variations in asthma risk among children and young adults from Latin America. **Scientific Reports**, v. 8, n. 1, p. 1–10, 2018b.

PAGANINI, J. et al. Validation of new HLA-F alleles assigned by next-generation sequencing. **Hla**, v. 93, n. 2–3, p. 131–132, 2019.

PNG, G. et al. Population-wide copy number variation calling using variant call format files from 6,898 individuals. **Genetic Epidemiology**, n. May, 2019.

PONTE, M. et al. Inhibitory receptors sensing HLA-G1 molecules in pregnancy: Decidua-associated natural killer cells express LIR-1 and CD94/NKG2A and acquire p49, an HLA-G1-specific receptor. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 10, p. 5674–5679, 1999.

PORAS, I. et al. Haplotypes of the HLA-G 3' untranslated region respond to endogenous factors of HLA-G⁺ and HLA-G⁻ cell lines differentially. **PLoS ONE**, v. 12, n. 1, p. 1–18, 2017.

PORTO, I. O. P. et al. MicroRNAs targeting the immunomodulatory HLA-G gene: A new survey searching for microRNAs with potential to regulate HLA-G. **Molecular Immunology**, v. 65, n. 2, p. 230–241, 2015.

RAMALHO, J. et al. HLA-E regulatory and coding region variability and haplotypes in a Brazilian population sample. **Molecular Immunology**, v. 91, p. 173–184, nov. 2017.

RAMSURAN, V. et al. Epigenetic regulation of differential HLA-A allelic expression levels. **Human molecular genetics**, v. 24, n. 15, p. 4268–4275, 2015.

RAMSURAN, V. et al. Sequence and Phylogenetic Analysis of the Untranslated Promoter Regions for HLA Class I Genes. **The Journal of Immunology**, v. 198, n. 6, p. 2320–2329, 2017.

RIBEYRE, C. et al. HLA-G haplotypes are differentially associated with asthmatic features. **Frontiers in Immunology**, v. 9, n. FEB, 2018.

ROBINSON, J. et al. The IPD and IMGT/HLA database: Allele variant databases. **Nucleic Acids Research**, v. 43, n. D1, p. D423–D431, 2015.

SABBAGH, A. et al. Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: Balancing selection influencing genetic diversity. **Genes and Immunity**, v. 15, n. 2, p. 95–106, 2014.

SANTOS, K. E. et al. Insights on the HLA-G evolutionary history provided by a nearby alu insertion. **Molecular Biology and Evolution**, v. 30, n. 11, p. 2423–2434, 2013.

SHIINA, T.; INOKO, H.; KULSKI, J. K. An update of the HLA genomic region, locus information and disease associations: 2004. **Tissue Antigens**, v. 64, n. 6, p. 631–649, 2004.

SOLBERG, O. D. et al. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. **Human Immunology**, v. 69, n. 7, p. 443–464, 2008.

SONON, P. et al. HLA-G, -E and -F regulatory and coding region variability and haplotypes in the Beninese Toffin population sample. **Molecular Immunology**, v. 104, n. August, p. 108–127, 2018.

SVENDSEN, S. G. et al. Extended HLA-G haplotypes in patients with age-related macular degeneration. **Hla**, v. 92, n. 2, p. 83–89, 2018.

TAN, Z. et al. Allele-specific targeting of microRNAs to HLA-G and risk of asthma. **American Journal of Human Genetics**, v. 81, n. 4, p. 829–834, 2007.

TAN, Z.; SHON, A. M.; OBER, C. Evidence of balancing selection at the HLA-G promoter region. **Human Molecular Genetics**, v. 14, n. 23, p. 3619–3628, 2005.

TU, B. et al. HLA-A, -B, -C, -DRB1 allele and haplotype frequencies in an African American population. **Tissue Antigens**, v. 69, n. 1, p. 73–85, 2007.

UROSEVIC, M.; TROJAN, A.; DUMMER, R. HLA-G and its KIR ligands in cancer – another enigma yet to be solved ? v. 4, p. 252–253, 2002.

VAN DER AUWERA, G. A. et al. **From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline.** [s.l: s.n.].

WAINWRIGHT, S. D.; BIRO, P. A.; HOLMES, C. H. HLA-F Is a Predominantly Empty, Intracellular, TAP-Associated MHC Class Ib Protein with a Restricted Expression Pattern. **The Journal of Immunology**, v. 164, n. 1, p. 319–328, 2000.

YIE, S. M. et al. A single base-pair mutation in the 3'-untranslated region of HLA-G mRNA is associated with pre-eclampsia. **Molecular Human Reproduction**, v. 14, n. 11, p. 649–653, 2008.

ZIDI, I. et al. Nonclassical human leukocyte antigen (HLA-G, HLA-E, and HLA-F) in coronary artery disease. **Human Immunology**, v. 77, n. 4, p. 325–329, 2016.

3. CONCLUSÃO

O MHC humano foi chamado de HLA e a maioria dos genes do MHC estão relacionados de alguma forma com o sistema imunitário, sendo alguns genes desse complexo particularmente variáveis e seus produtos fundamentais ao desenvolvimento da resposta imunológica. Os genes HLA aqui estudados estão localizados no bloco alfa do MHC humano, assim como três variantes estruturais e a associação entre eles, bem como a variabilidade completa dos genes em uma grande população miscigenada, é capaz de fornecer ainda dados para estudos clínicos e evolutivos.

Por fim, encontramos associação entre o elemento *Alu*_{YHG} e o CNV esv3608493 com alelos de alta expressão, sendo o primeiro com G*01:01:01:01 e o segundo com A*23 e A*24. Nossos dados sugerem uma estreita relação entre o *HLA-A* e o *HLA-G*, com ambos possuindo evidências de seleção balanceadora (segmentos intrônicos para o *HLA-G* e todo o locus para o *HLA-A*). Assim, esse fenômeno pode não ser independente e o *HLA-G* pode sofrer efeito carona por causa do *HLA-A*. Este é o maior conjunto de dados de variantes e alelos HLA de resolução completa (no bloco alfa) de uma amostra misturada, fornecendo um recurso valioso de variantes, frequências e haplótipos nesses genes.