UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

Programa de Pós-Graduação em Ciências Biológicas (GENÉTICA)

VARIABILIDADE GENÉTICA E ESTRUTURA HAPLOTÍPICA DO GENE *KIR2DL4* AVALIADA EM UMA AMOSTRA BRASILEIRA

EMILIANA WEISS

Botucatu - SP

2019

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

Programa de Pós-Graduação em Ciências Biológicas (GENÉTICA)

VARIABILIDADE GENÉTICA E ESTRUTURA HAPLOTÍPICA DO GENE *KIR2DL4* AVALIADA EM UMA AMOSTRA BRASILEIRA

EMILIANA WEISS

ORIENTADOR: PROF. DR. ERICK DA CRUZ CASTELLI

Dissertação apresentado ao Instituto de Biociências de Botucatu, Universidade Estadual Paulista "Júlio de Mesquita Filho", para obtenção do título de Mestre no Programa de Pós-Graduação em Ciências Biológicas (Genética).

Botucatu – SP

2019

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

Weiss, Emiliana.
Variabilidade genética e estrutura haplotípica do gene *KIR2DL4* avaliada em uma amostra brasileira / Emiliana Weiss. - Botucatu, 2019

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Erick da Cruz Castelli
Capes: 21103003

1. Células Killer. 2. Antígenos HLA. 3. Haplótipos. 4. Polimorfismo (Genética). 5. Sequenciamento de nucleotídeo.

Palavras-chave: Células *Natural Killer*; HLA; Haplótipos; Polimorfismo; Sequenciamento de segunda geração.

Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution.

A. Einstein, Cosmic Religion: With Other Opinions and Aphorisms, p. 97 (1931).

AGRADECIMENTOS

Dedico essa dissertação aos meus Pais, *Angela Laura Rios Weiss* e *Emilio Carlos Weiss*. Agradeço imensamente pelo apoio e suporte, que foram e são essenciais nessa *trajetória* ∞. À minha irmã Eliana Weiss, minha eterna melhor amiga, *roommate* e conselheira, desculpe pelos momentos de ausência. À Douglas S. Mangini, por sempre me apoiar, ter sido meu porto seguro nos momentos de angústia nesses dois longos anos, e além disso revisar meus projetos e me ajudar com ferramentas do Excel. *Thank u*. À todos os meus familiares, que apesar de não entenderem minha decisões, continuam a apoiar meus sonhos. Em especial à minha *segunda mãe* Arlete Rios e minha avó Laura J. Rios. Aos avós Venceslau Rios, Hilda H. Weiss, Raulino Weiss (*in memoriam*). *Ich bin Ihnen sehr dankbar für.*

À Universidade Estadual Paulista "Júlio Mesquita Filho", campus Botucatu, onde realizei este projeto. Aos professores que participaram dessa trajetória.

Ao Programa de Pós-Graduação em Ciências Biológicas (Genética) do Instituto de Biociências de Botucatu. À equipe do Workshop de Genética (XVIII e XIX).

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – com o Processo no 2017/05042-4, pelo apoio financeiro essencial para a realização dessa pesquisa. Sou muito grata por todo o financiamento recebido, pelo investimento a minha formação científica com a bolsa de mestrado concedida.

Ao orientador Dr. Erick C. Castelli, pela oportunidade e aprendizado.

À equipe de laboratório: Andreia S. Souza, Heloísa S. Andrade, Iane O. P. Porto, Marília R. Passos, Michelle A. Paz, Thálitta H. A. Lima, Juliana R. Lara. Pela ajuda na realização dos experimentos e execução desse projeto. Em especial, à Hello, que me abrigou e é minha conselheira *top*, revisora *máster*. *Merci*. Às Amigas e colegas Juxliana R. Lara, Ticiane D. J. Farias e Bibiana S. Almeida, por me darem conselho e acreditarem no meu potencial.

À banca avaliadora, por dedicarem seu tempo e conhecimento. A contribuição de vocês é de suma importância.

RESUMO

O gene *KIR2DL4* codifica um importante receptor de células Natural Killer (NK). O único ligante de KIR2DL4 conhecido é a molécula HLA-G, expressa principalmente na placenta, modulando a ação das células NK durante a gestação. O gene *KIR2DL4* parece ser bastante variável, quando considerado os bancos de dados que armazenam suas sequências conhecidas, porém não está claro o nível de diversidade deste gene em populações reais e heterogêneas. Polimorfismos presentes no gene *KIR2DL4* poderiam influenciar a interação entre KIR2DL4 e HLA-G, modificando a ação das células NK. Neste estudo exploramos a variabilidade genética de *KIR2DL4* em 157 indivíduos oriundos do Estado de São Paulo/Brasil. Devido à alta similaridade de sequências entre os genes KIR, erros de genotipagem são esperados quando se utiliza sequenciamento de segunda geração. Por este motivo, desenvolvemos uma abordagem para classificar cada leitura com base em sequências KIR conhecidas, endereçando-as ao gene mais provável. Também utilizamos o painel SNPforID 34-plex para avaliar a ancestralidade dessas amostras. Considerando o segmento completo desse gene, indo da região 5'URR até a 3'UTR, com aproximadamente 13kb, o gene *KIR2DL4* se mostrou pouco polimórfico, com 152 pontos de variações identificados (MAF 1%). Esses pontos de variação estão organizados em 32 haplótipos estendidos que codificam 13 proteínas diferentes. Foram encontrados 11 haplótipos na região promotora, sendo que 8 possuem MAF maior que 1%. Na região codificadora 70 haplótipos foram encontrados, sendo que quatro deles correspondem a 50% das sequencias codificas (KIR2DL4*0080204, *008105, *001, *005). Na região 3'UTR, 14 haplótipos foram identificados, todos com MAF maior que 1%. A diversidade nucleotídica foi maior na região codificadora. Através do uso de AIM, percebeu-se que a variabilidade de *KIR2DL4* é influenciada pela ancestralidade. Foi também observado um forte equilíbrio de ligação ao longo do gene, similar ao seu ligante *HLA-G.* Assim, o *KIR2DL4* é bastante conservado tanto no nível de DNA quanto no de proteína.

Palavras-chave: Polimorfismo; Células *Natural Killer;* Haplótipos; HLA; Sequenciamento de segunda geração; KIR2DL4.

ABSTRACT

KIR2DL4 is the most unusual Killer Cell Immunoglobulin-Like receptor (KIR) family member in terms of structure, expression, and signaling properties. The only known KIR2DL4 ligand is HLA-G, and polymorphisms might disrupt this interaction. *KIR2DL4* variability is not well explored in admixed populations. Here we explored KIR2DL4 exon variability in 157 individuals from the State of São Paulo/Brazil. Because of sequence similarity with other KIR genes, it is expected genotyping errors when using second-generation sequencing. We developed an approach to score each read based on known KIR sequences, addressing them to the most likely locus. We evaluated the SNPforID 34-plex panel to assess ancestry. The methodology was applied to survey the variability of a very admixed population, such as Brazilian, counting with 157 samples of São Paulo State. Considering a segment of about 13-kb, *KIR2DL4* gene was conserved with few different and frequent sequences. Overall, 152 variable sites were detected, arranged in 32 haplotypes codifying 13 protein. We found 11 promoter haplotypes, 8 with a frequency greater than 1%. In the coding region we detected 70 haplotypes, four of which correspond to 50% of the coding sequences (KIR2DL4 * 0080204, * 008105, * 001, * 005). In the 3'UTR region, 14 haplotypes were identified with MAF greater than 1%. The *KIR2DL4* coding region was the most variable segment. We observed that *KIR2DL4* variability is strongly influenced by the sample ancestry background. *KIR2DL4* is highly conserved at both the DNA and protein levels.

Key-words: Polymorphisms; *Natural Killer* cells; Haplotypes; HLA; Second-generation sequencing; KIR2DL4.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

APC	*Antigen-presenting cell*
bp ou pb	pares de bases
CD56	Receptor das células natural killer
CD94/NKG2A	Receptor heterodímero das células natural killer
D0, D1 e D2	Domínios dos receptores KIR
dbSNP	Single Nucleotide Polymorphism Database
DNA	Ácido desoxirribonucleico
GM-CSF	Fator estimulador de crescimento de granulócitos
HLA	*Human Leukocyte Antigen* (Antígeno Leucocitário Humano)
IFN-γ	Interferon- γ
Ig	Imunoglobulina
IL	Interleucina
ITAM	Do inglês Immunotyrosine-based activatory motif
ITIM	Do inglês Immunotyrosine-based inhibitory motif
Kb	Kilobases ($10^3\ bases$)
KIR	Receptor de células killer semelhante à imunoglobulina, do inglês Killer-cell Imunoglobulin-like Receptor
LRC	Complexo de receptores leucocitários, do inglês Leukocyte
MAF	*Minimum allele frequency*
MHC	Major Histocompatibility Complex (Complexo Principal de Histocompatibilidade)
miRNA	micro
mRNA	RNA mensageiro
NGS	*Next generation sequencing*
NK	Células Natural Killer
PCR	Reação em Cadeia da Polimerase
RNA	Ácido ribonucleico
SNP	Single nucleotide polymorphism (Polimorfismo de base única)
Receptor Complex	

TCR	Receptor de célula T
UTR	*Unstranslated region*
VCF	*Variant Call Format*
TNF-α	Fator de necrose tumoral-α, do inglês Tumor Necrosis Factor-α

SUMÁRIO

Capitulo I

1. REVISÃO DE LITERATURA

1.1 O SISTEMA IMUNE E AS CÉLULAS *NATURAL KILLER*

O sistema imunitário é responsável pela defesa do organismo contra infecções por patógenos, incluindo bactérias, vírus e parasitas. Além disso, é capaz de reconhecer e eliminar células alteradas e células neoplásicas. O sistema imunitário apresenta um mecanismo sofisticado para a regulação da atividade das células imunocompetentes em diversos níveis, o que envolve muitas vias celulares e bioquímicas (Jager & Kuchroo, 2010).

A resposta imune adaptativa e o processo inflamatório são ativados pela primeira linha de defesa do organismo: as células do sistema imune inato, na qual atuam principalmente as células *Natural Killer* (NK), macrófagos, células dendríticas e outras células apresentadoras de antígeno (APC, do inglês *Antigen-presenting cell*), responsáveis pelo reconhecimento de moléculas não próprias ao organismo por meio de receptores de reconhecimento de padrões (Doria et al., 2012).

As células NK são cruciais para a ativação do sistema imune inato e influenciam na resposta antígeno-específica da imunidade adaptativa (Biron, Nguyen, Pien, Cousens, & Salazar-Mather, 1999). Classificadas como um subgrupo de linfócitos granulares, pela sua morfologia e expressão de marcadores linfoides, as células NK tem sua origem a partir de um progenitor de células linfoides na medula óssea (Jager & Kuchroo, 2010; Jobim et al., 2010). As células NK interagem com outros tipos celulares diretamente, por meio de receptores e ligantes presentes na sua superfície celular, ou indiretamente, por meio da produção de citocinas e quimiocinas, como IFN-γ, TNF-α e fator estimulador de crescimento de granulócitos (*GM-CSF*). Os seus receptores de membrana são responsáveis por reconhecer ligantes em células-alvo e liberar sinais ativadores ou inibidores que modulam a resposta citotóxica das NK. Dentre suas funções, pode-se dizer que elas desempenham um papel chave

no controle de infecções por vírus e parasitas intracelulares (Lee, Miyagi, & Biron, 2007; Perricone, Perricone, De Carolis, & Shoenfeld, 2008), na vigilância imunológica de tumores (Langers, Renoux, Thiry, Delvenne, & Jacobs, 2012; Purdy & Campbell, 2009), durante a gestação em humanos (Parham, Norman, Abi-Rached, Hilton, & Guethlein, 2012) e, por fim, na patogênese de diversas doenças autoimunes (Schleinitz, Vely, Harle, & Vivier, 2010; Vivier et al., 2011).

O fenótipo das células NK maduras é definido pela expressão de CD56 na superfície celular (CD56+) e pela não expressão de CD3 (CD3-). O nível de expressão de CD56 define dois subgrupos: CD56bright (maior expressão) e CD56dim (menor expressão). As células NK do tipo CD56bright localizam-se predominantemente em linfonodos e locais de inflamação e sua função reguladora ocorre através da produção de citocinas, como as IL-12 e IL-18, e quimiocinas. As células NK do tipo CD56dim circulam na corrente sanguínea. Estas células possuem grande potencial citotóxico, pois produzem perforinas e apresentam receptores KIR (do inglês *killer cell immunoglobulin-like receptors*) (Ahern & Brennan, 2011; Schleinitz et al., 2010).

As células *Natural Killers* possuem uma vasta variabilidade de receptores na sua superfície celular, que são responsáveis por interagir direta ou indiretamente com outros tipos celulares, estimulando ou inibindo sua atividade citolítica. Para promover um ataque direto às células-alvo, as NK não necessitam de uma ativação prévia, o que evidencia a participação desta célula na resposta imune inata (Perricone *et al.*, 2008; Vivier *et al.*, 2011). Após o reconhecimento da célula-alvo, as NK podem provocar a lise celular e/ou produzir citocinas e quimiocinas que ativam outras células da resposta imune adaptativa. Esses eventos dependem da natureza do estímulo entre a célula NK e a célula-alvo (Perricone et al., 2008; Vivier et al., 2011), e também evidenciam a atuação dessas células tanto na resposta imune inata quanto na adaptativa.

O reconhecimento das células-alvo por células NK se dá por uma vasta gama de receptores que não sofrem rearranjos genéticos como ocorre com os receptores das células T (TCR) (Vivier et al., 2011). Os genes que codificam esses receptores incluem lectinas da família do tipo-C (CD94/NKG2, Ly49), que ficam localizada no cromossomo 12q12.3-13.4; e os Killers Immunoglobulin (Ig)-Like Receptors (KIRs), os leukocyte Ig-like receptors (LILRs, LIRs ou Ig-like transcrips ou ILTs, e os leukocyte-associated Ig-like receptors (LAIRs), localizados em um segmento de aproximadamente 1MB na região chamada Complexo de Receptores Leucocitários, no cromossomo 19q13.4.

Os receptores KIR podem tanto estimular a atividade das células NK, pelos chamados receptores ativadores, quanto impedir a ativação destas células pelos receptores inibidores. Os receptores KIR se ligam as moléculas de HLA de classe I (HLA-A, -B, -C, -E, -F e -G) presentes em células-alvo, e podem modular a ação das NK juntamente com a ação de outros receptores(Jobim et al., 2010; Purdy & Campbell, 2009). Receptores inibitórios ligam-se a moléculas HLA de classe I alvo e impedem o ataque de células NK contra células normais. Quando um receptor ativador se liga ao seu ligante, os sinais de ativação são gerados levando à destruição da célula-alvo (Parham & Moffett, 2013).

Considerando-se que nenhum receptor age sozinho (FIGURA 1), é necessário um predomínio de sinais inibidores para que a célula NK não seja dirigida ao ataque. Da mesma forma, se faz necessário um predomínio de sinais ativadores para que ocorra ativação da célula NK a iniciar a lise da célula-alvo (Vivier et al., 2011).

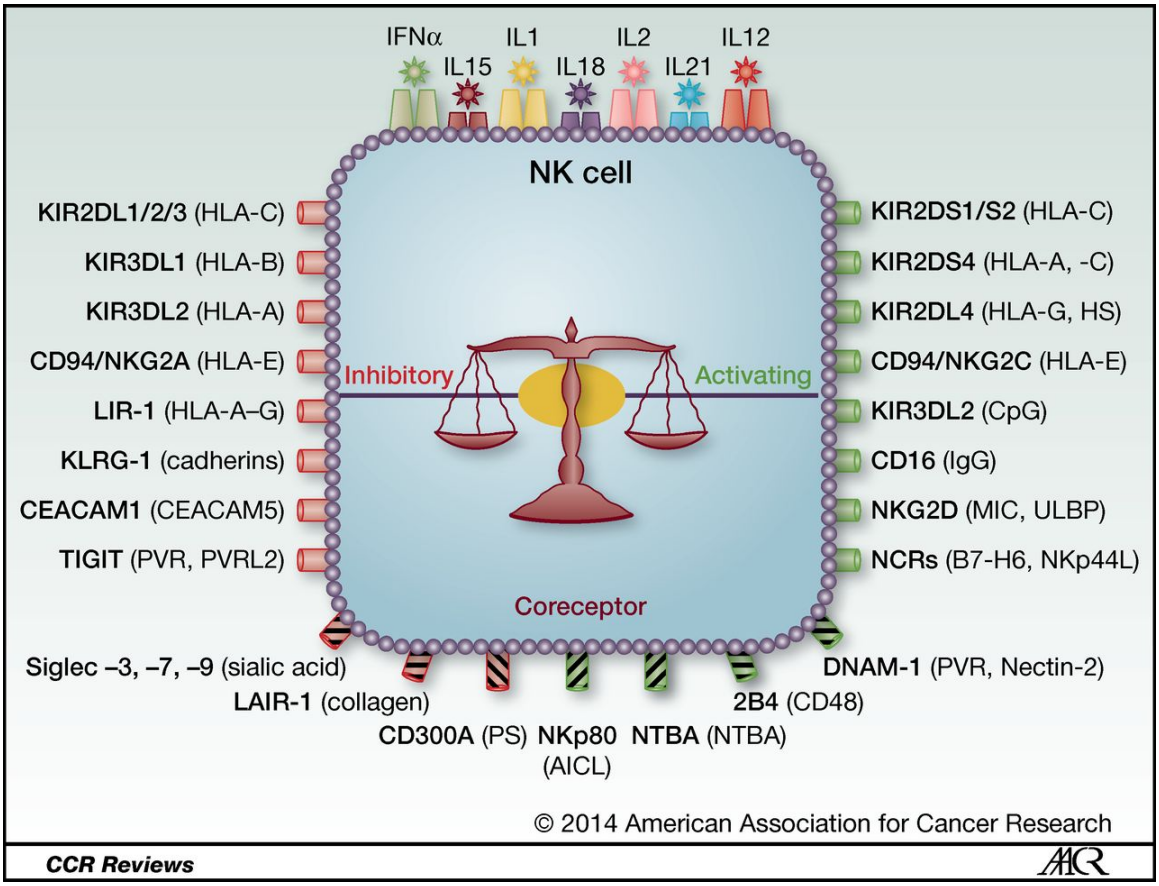

Figura 1– Receptores de membrana de células NK. Nessa figura estão representados alguns dos receptores de membrana de células NK e seus ligantes, em parênteses. Os receptores inibitórios estão representados em vermelho e os seus coreceptores (vermelho com listras pretas), os receptores ativatórios (em verde) e os seus coreceptores (em verde com listras pretas). Fonte: (Leung, 2014).

A ativação das células NK ocorre na presença do processo denominado "*missing self*", *i.e.,* perda do próprio (Karre, Ljunggren, Piontek, & Kiessling, 1986), no qual ocorre a perda da expressão de moléculas HLA de classe I. Essa situação manifesta-se quando a célula está infectada ou sofre processo neoplásico. Quando as moléculas de HLA de classe I, que são os principais ligantes dos receptores inibidores das NK, não estão presentes ou tem sua expressão diminuída na superfície celular da célula alvo, a célula-alvo será interpretada como em situação anormal, resultando em atividade citotóxica da célula NK, fato esse que pode depender também do balanço entre sinais de inibição e ativação mediados por receptores KIR (Orr & Lanier, 2010; Vivier et al., 2011). Desta forma, a ligação do receptor KIR com moléculas HLA de classe I é vital para o reconhecimento de antígenos não próprios e tolerância imunológica (Purdy & Campbell, 2009).

1.2 Os Receptores Similares a Imunoglobulinas - KIR

Os receptores KIR são responsáveis por sinais tanto de ativação quanto de inibição da resposta imune celular, dependendo da composição das caudas citoplasmáticas dos receptores em questão. Os receptores inibitórios geralmente contam com motivos de inibição baseados em tirosina (ITIMs), associados a caudas citoplasmáticas mais longas (Avril, Floyd, Lopez, Vivier, & Crocker, 2004). Os receptores ativadores contam com caudas citoplasmáticas curtas e incapazes de enviar sinais intrínsecos. Assim, eles precisam de moléculas adaptadoras que contam com motivos de ativação baseados em tirosina (ITAMs) (Humphrey, Lanier, & Nakamura, 2005).

Os receptores KIR são os receptores de células NK mais estudados, descritos pela primeira vez por Moretta et al. (1993). Os receptores KIR fazem parte de uma família gênica que se localiza em uma região genômica modificada por recentes expansões e contrações (Uhrberg, 2005). Os genes KIR são altamente polimórficos e se ligam às moléculas de HLA classe I, que por sua vez também são muito polimórficas (Vivier et al., 2011). Os genes codificadores dos receptores KIR estão localizados no cromossomo 19q13.4, na região do Complexo de Receptores Leucocitários – LCR (Figura 2). Estes genes variam entre si em estrutura, tamanho e nível de polimorfismo (Suto et al., 1996). Atualmente são reconhecidos cerca de 977 alelos (ou sequências distintas) distribuídos entre os 16 *loci* (incluindo dois pseudogenes) que compõem esse grupo (*The Immuno Polymorphism Database*, versão 2.8.0). Alguns desses genes podem estar presentes ou ausentes no genoma do indivíduo, formando conjuntos específicos de genes (haplótipos), como o representado na FIGURA 2. No entanto, deve-se destacar que a maioria dos alelos já descritos não conta com sequências de íntrons e regiões regulatórias devidamente caracterizadas.

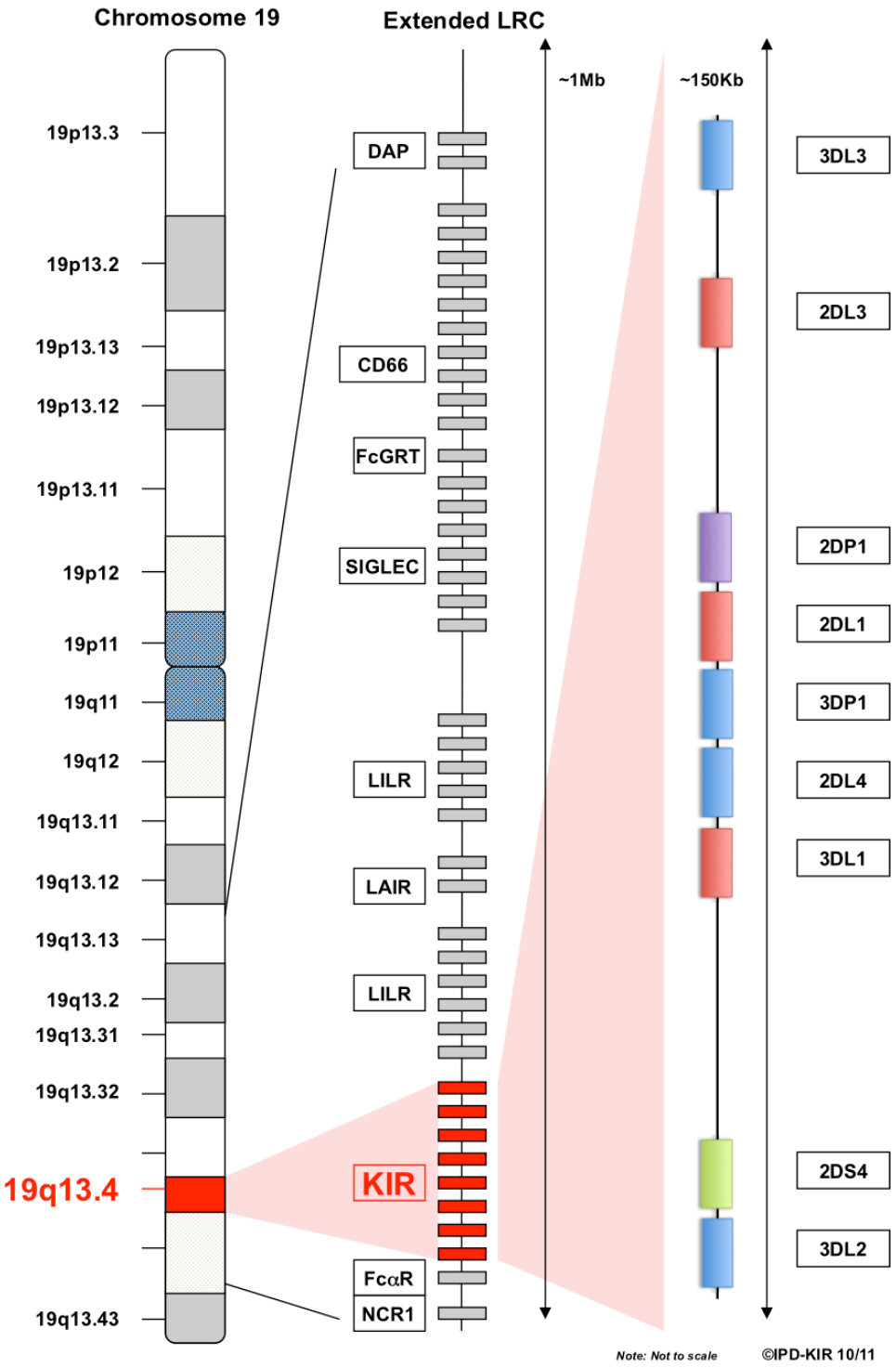

Figura 2 – Composição do Complexo de Receptores Leucocitários, com destaque para a região ocupada pelos genes KIR. Caixas azuis representam genes presentes em todos os haplótipos; caixas verdes representam genes ativadores e as vermelhas, inibidores; caixas roxas representam pseudogenes. Imagem disponível em http://www.ebi.ac.uk/ipd/kir/images/figure_01.png

Os genes *KIR* são classificados por três critérios: i) número de domínios de Ig extracelulares que eles codificam; ii) tamanho da cauda citoplasmática; e iii) similaridade com outras sequências (Vilches & Parham, 2002). Assim, de acordo com o número de domínios de Ig, eles são classificados como KIR2D (dois domínios) ou KIR3D (3 domínios), sendo que os domínios extracelulares tem por função o reconhecimento da célula-alvo. Os genes que possuem somente dois domínios de Ig apresentam um mRNA menor por edição alternativa do transcrito primário (Martin, Freitas, Witt, & Christiansen, 2000). Quanto à cauda citoplasmática, esta pode ser longa (L - *long*) ou curta (S - *small*), diferenciando os receptores inibitórios dos ativadores (FIGURA 3). O número subsequente ao L ou S, representa a ordem de descrição do gene e sua semelhança estrutural com outros genes (Martin et al., 2000; Vilches & Parham, 2002).

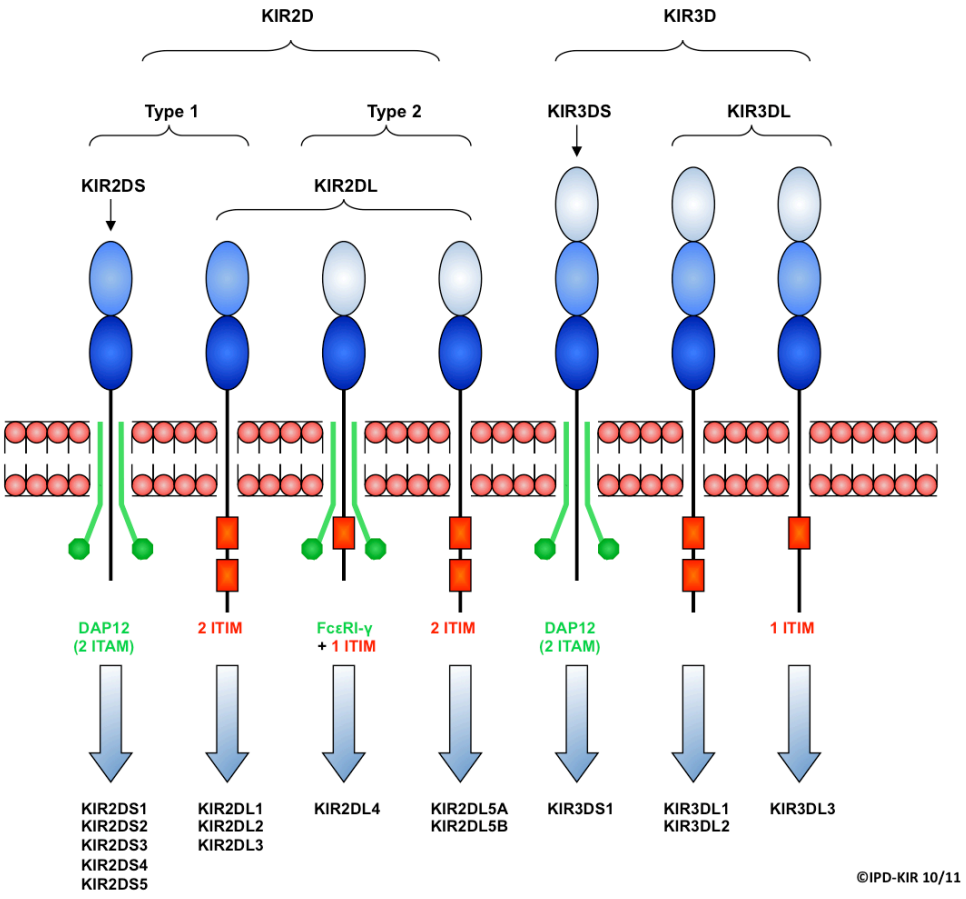

Figura 3– Representação esquemática da estrutura das moléculas KIR. Imagem disponível em http://www.ebi.ac.uk/ipd/kir/images

A nomenclatura dos alelos KIR tem denominação semelhante à antiga nomenclatura dos alelos de genes HLA (FIGURA 4), com o nome do gene sendo separado do número do alelo por um asterisco (*), os 3 primeiros dígitos referentes à variações em regiões de éxon (mudanças não sinônimas), os 2 dígitos seguintes sendo a variante sinônima dentro da sequência do éxon e os 2 últimos identificam polimorfismos encontrados nas regiões intrônicas, promotora ou não codificadora, que não causam mudanças na sequência codificadora da proteína (Marsh et al., 2003).

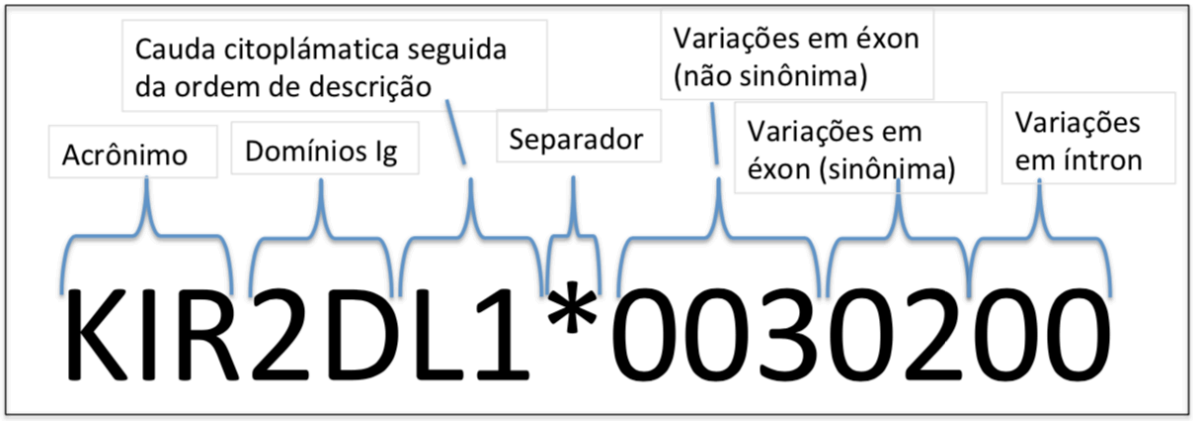

Figura 4– Representação esquemática da nomenclatura KIR. Imagem baseada na disponível em http://www.ebi.ac.uk/ipd/kir/images

Os genes KIR apresentam, em geral, nove éxons e oito íntrons, com seu tamanho variando entre 4 e 16 Kb (FIGURA 5). Os éxons 1 e 2 desses genes codificam o peptídeo líder; éxons 3, 4 e 5 são responsáveis pela codificação dos domínios de Ig, chamados respectivamente de D0, D1 e D2; o éxon 6 codifica uma haste que liga o domínio D2 ao domínio transmembrana, codificado pelo éxon 7; finalmente, os éxons 8 e 9 codificam a cauda citoplasmática (Vilches & Parham, 2002). Os diferentes tamanhos de caudas citoplasmáticas dos receptores ativadores e inibidores são causados por substituições de bases e inserções/deleções (*indels*) que dão origem a códons de parada em posições diversas (FIGURA 5) (Steffens, Vyas, Dupont, & Selvakumar, 1998).

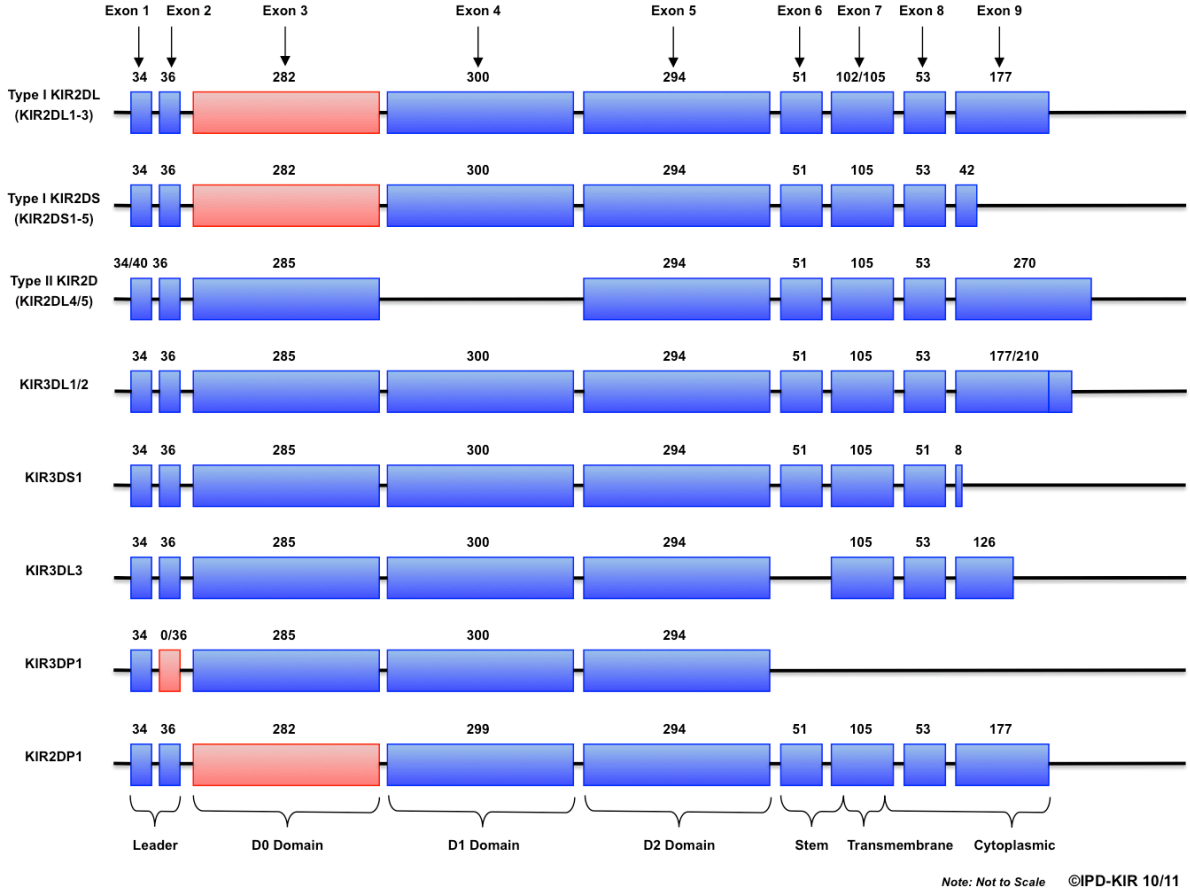

Figura 5– Organização dos genes KIR. Os genes *KIR* que possuem uma estrutura organizada de forma similar foram agrupados, enquanto que os genes *KIR* com peculiaridades estruturais são mostrados individualmente. As caixas azuis representam as regioes codificadoras (éxons) e o número acima representa a quantidade de pares de base. O pseudoexon 3 e a região do éxon 2 deletada são mostradas em vermelho. Imagem disponível em http://www.ebi.ac.uk/ipd/kir/images/figure_01.png

Os receptores KIR estão presentes na superfície de células NK, onde interagem com moléculas do HLA de classe I, em especial as clássicas como HLA-B e HLA-C, além da HLA-G que interage com KIR2DL4 (Thielens, Vivier, & Romagne, 2012). Além de ser encontrado na superfície de células NK, o receptor KIR2DL4 também pode ser encontrados em endossomos, podendo induzir uma secreção robusta de citocinas (Rajagopalan, 2010). Em tecidos normais, moléculas HLA de classe I apresentam antígenos de proteínas normais para células NK e células T citotóxicas e normalmente não estimulam respostas imunes, pois interagem com receptores inibitórios das células T e NK. No entanto, a mudança na estrutura

de peptídeos pode levar a ativação de células T efetoras pelo reconhecimento da molécula por outros receptores ativadores (Fadda et al., 2010; Norman et al., 2013).

A diversidade alélica dos genes *KIR* pode estar relacionada com os níveis de expressão dos receptores na superfície celular, além de potencialmente influenciar sua interação com os ligantes HLA de classe I. Por exemplo, a interação HLA-C/KIR2DL1 é sensível a uma substituição de um único aminoácido na posição 80 na alfa-hélice da HLA-C e na posição 44 do domínio D1 em *KIR2DL1, KIR2DL2* e *KIR2DL3*. Polimorfismos em determinadas regiões alteram a especificidade ligante-receptor, com intensidades distintas (KULKARNI et al., 2008; PARHAM, 2005).

Estudos recentes sobre a diversidade do KIR sugerem que variações no conteúdo genético e diversidade alélica podem afetar a funcionalidade das células NK, tanto no nível de especificidade quanto no nível de sensibilidade. A presença ou ausência de genes KIR em um haplótipo é sempre associada com ganho ou perda de especificidade, uma vez que eles estão relacionados a reconhecimento de genes ligantes específicos na célula-alvo. Por exemplo, células NK de indivíduos com haplótipos KIR que não possuem *KIR3DL1* não sofrem inibição pela presença da molécula HLA-Bw4(Saunders et al., 2015). Até onde se sabe, nenhum outro KIR é capaz de compensar a sua função (Uhrberg, 2005; Vilches & Parham, 2002). Além disso, é importante destacar que a expressão de genes KIR não é idêntica em todas as células. Isso porque cada célula NK é capaz de expressar um conjunto diferente de receptores, podendo estar relacionado com tecidos específicos ou a uma maior ou menor resposta imune a um determinado patógeno (Raulet, Vance, & McMahon, 2001).

Os haplótipos dos genes KIR foram classificados em dois grupos de acordo com o seu conteúdo gênico: A e B (FIRURA 6). O haplogrupo A é formado por um conjunto único de genes, sendo composto por um único tipo de haplótipo caracterizado pela presença de um gene ativador (*KIR2DS4*), cinco genes inibidores (*KIR2DL1*, *KIR2DL3*, *KIR3DL1*, *KIR3DL2*, *KIR3DL3),* um gene com função tanto ativadora quanto inibidora (*KIR2DL4)* (Hsu, Chida,

Dupont, & Geraghty, 2002; Martin et al., 2004), e dois pseudogenes (*KIR2DP1* e *KIR3DP1)* (Uhrberg et al., 1997). O haplogrupo B contém uma grande variedade de combinações entre os genes *2DS1*, *2DS2*, *2DS3, 2DS5, 3DS1, 2DS4*, além de *2DL4* (Martin et al., 2000; Vilches & Parham, 2002) (FIGURA 5). As frequências dos haplótipos A e B variam muito de uma população para outra. Já foi observado que raramente indivíduos não relacionados possuem genótipos KIR similares, e que populações com etnias diferentes possuem distintas distribuições de frequência de genótipo KIR (Parham, 2005b).

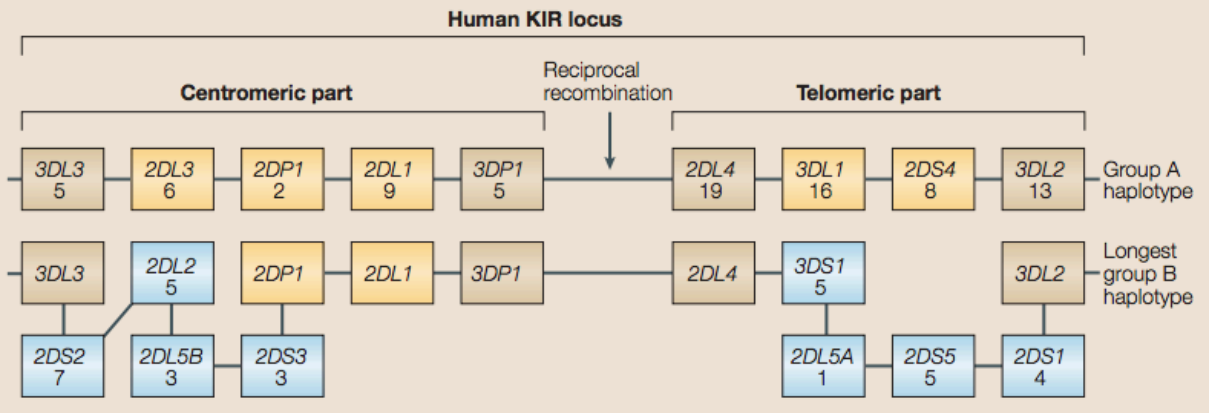

Figura 6– Organização dos haplótipos A e B. Nas caixas marrons estão os genes chamados estruturais, que estão semple presentes nos diferentes haplótipos. O haplótipo mais curto é o A, e o mais longo o haplótipo B. Genes presentes em ambos os haplótipos estão representados nas caixas amarelas. Os genes nas caixas azuis são responsáveis por formar os diferentes haplótipos do grupo B, podendo estar presentes ou ausentes. Os números logo abaixo dos genes representam os alelos específicos de cada grupo Fonte: Parham, 2005.

Considerando todos os possíveis haplótipos de KIR, apenas três genes (e um pseudogene - KIR3DP1) estão sempre presentes em um mesmo indivíduo – *KIR2DL4*, *KIR3DL2* e *KIR3DL3* – e são considerados genes estruturais, ou *frameworks* (caixas azuis da Figura 2). Todos eles contêm cauda citoplasmática longa e motivos ITIM que são responsáveis pela ancoragem de moléculas com homologia à Src-2, um domínio proteico altamente conservado relacionado ao envio de sinais intracelulares, que contêm fosfatases – mecanismo comum a outros receptores de caráter inibitório (Avril et al., 2004; Staub, Rosenthal, & Hinzmann, 2004). No entanto, o receptor KIR2DL4 tem se mostrado um tanto controverso, uma vez que ele é considerado também um ativador de respostas imunes. Isso

ocorre porque um resíduo de asparagina em seu domínio transmembranar lhe confere uma característica de atividade ativadora (Kikuchi-Maki, Yusa, Catina, & Campbell, 2003). Sua interação com a molécula HLA-G resulta no envio de sinais celulares responsáveis pela produção de citocinas. Estas serão responsáveis por modular a resposta imune, inclusive ativando outras células, porém a célula NK que é ativada por esse receptor é responsável por pouca, ou nenhuma, citotoxicidade (Vilches & Parham, 2002). Seu domínio ITIM interage com o receptor FcεRI-ɣ e estimula a produção de citocinas como o Interferon-ɣ (IFNɣ), de modo semelhante a receptores que apresentam motivos de ITAM (Kikuchi-Maki et al., 2003).

1.3 Estrutura do *KIR2DL4*

Ao contrário dos demais genes KIR, o gene *KIR2DL4* é composto por 8 éxons (Martin et al., 2000). O tradicional éxon 4 dos genes KIR, responsável por codificar o Domínio Extracelular D1, encontra-se deletado, indicando que esse gene possui uma estrutura genômica diferente quando comparado à outros genes da mesma família. Dessa forma, *KIR2DL4* apresenta apenas dois Domínios Extracelulares, D0 e D2. Os éxons 1 e 2 são responsáveis por codificar o peptídeo sinal e os éxons 3 e 5 codificam os dois Domínios semelhantes à Ig, Domínio 0 e Domínio 2, respectivamente. O Domínio de união é codificado pelo éxon 6 o transmembrana pelo éxon 7, e os domínios citoplasmáticos são codificados pelos éxons 8 e 9 (Wilson, Torkar, & Trowsdale, 1997). *KIR2DL4* possui apenas um domínio com motivo inibidor, enquanto os demais genes 2DL apresentam dois. A presença de uma Arginina carregada positivamente na região transmembranar, diferentemente dos outros 2DL, leva a acreditar que esse gene tenha tanto função ativadora quanto inibitória (Faure & Long, 2002). A região promotora de *KIR2DL4* apresenta uma vasta gama de elementos regulatórios, conferindo a esse gene uma regulação diferenciada em comparação aos demais genes KIR. Neste sentido, *KIR2DL4* apresenta apenas 69% de identidade com os outros genes da sua família (Trowsdale et al., 2001). Além disso, aparentemente, *KIR2DL4* é transcrito em todas as células NK (Valiante et al., 1997).

O gene *KIR2DL4* tem sido associado a patologias relacionadas à reprodução (Li & Durbin, 2009), o que pode ser atribuído ao papel desempenhado por seu ligante, HLA-G, durante o curso gestacional. O receptor KIR2DL4 tem uma interação importante com moléculas de HLA-G, o que influencia especialmente a gestação. Durante o primeiro trimestre, as células NK deciduais expressam grandes quantidades de KIR2DL4, e o IFNɣ resultante da ativação dessas células possui um papel importante na implantação do blastocisto e decidualização do tecido uterino (Yan et al., 2007). A relação com patologias relacionadas ao sistema imunológico fundamenta-se, provavelmente, na função de inibição e ativação da atividade das células NK.

1.4 Polimorfismos do gene *KIR2DL4*

A sequência genômica do *KIR2DL4* compreende aproximadamente 11kb. Os 70 alelos já descritos para *KIR2DL4* codificam 41 (IMGT/KIR versão 2.8.0) variantes proteicas – números semelhantes aos encontrados até o momento para seu ligante, codificado pelo gene *HLA-G*.

Variantes nas regiões codificadora e regulatórias de um gene influenciam a molécula codificada e o padrão de expressão do gene, respectivamente. O gene *KIR2DL4* possui pouca similaridade com os genes da sua família na sua região promotora, indicando que há um mecanismo diferente de regulação (Stewart, van Bergen, & Trowsdale, 2003; Witt, Martin, & Christiansen, 2000). Além disso, *KIR2DL4* possui uma deleção no éxon 6 que modifica o frame de leitura, levando a um *stop códon* prematuro na região da cauda citoplasmática, no éxon 7 e a uma proteína com cauda citoplasmática menor (Witt et al., 2002). A proteína produzida por esse mRNA pode ser expressa na membrana de células NK, mas perde a sua capacidade inibitória com motivo ITIM. Essa variante apresenta uma frequência de 50% em diversas populações e há evidências de seleção balanceadora neste segmento (Goodridge et al., 2007).

1.5 Aspectos evolutivos e ancestralidade genômica

A similaridade entre os genes KIR nas suas regiões de sequências intergênicas, íntrons e éxons, sugerem que recombinação intergênica, conversão gênica e trocas de domínio devem ter acontecido ao longo do tempo (Rajalingam et al., 2002; Vilches & Parham, 2002). Esses eventos, juntamente com mutações pontuais, duplicações gênicas e deleções, parecem ter contribuído para a rápida evolução desse gene e o perfil de diversidade observado, com grande variabilidade nas regiões de domínio extracelular, domínio de ligação e as citoplasmático (Shilling, Lienert-Weidenbach, Valiante, Uhrberg, & Parham, 1998; Vilches & Parham, 2002). Também devemos lembrar que a família genica KIR surgiu em decorrência de eventos de múltiplas duplicações genicas durante a evolução do grupo dos primatas (Vilches & Parham, 2002; Wilson et al., 1997).

Foi sugerido que a variação do conteúdo gênico e a diversidade alélica de KIR podem afetar a função e o nível de especificidade e sensibilidade das células NK. A presença ou ausência de alguns genes pode ser diretamente associada ao ganho ou perda de especificidade como, por exemplo, no caso já citado entre o gene *KIR2DL1* e seu ligante específico *HLA-C* com lisina na posição 80. A variação no número de genes também é associada à mudanças qualitativas na ação específica das NK. Desse modo, a seleção natural pode influenciar as frequências de certos genes, dependendo de fatores ambientais e endógenos, garantindo maior resistência a certos tipos de infecções e outras doenças (Goodridge, Witt, Christiansen, & Warren, 2003). Além disso, diferenças populacionais podem ser decorrentes da ação da deriva genética, efeito fundador e fluxo gênico.

Até o momento não há estudos que caracterizaram a variabilidade do gene *KIR2DL4* na profundidade aqui proposta, em especial em uma população como a brasileira, que é um reconhecido repositório de variação genética. A população brasileira apresenta contribuições Europeia, Africana e Ameríndia. O estado de São Paulo, alvo do presente estudo, é o mais

populoso do Brasil e contém uma considerável miscigenação entre os grupos citados acima. Para essa região do país, a maior contribuição é Europeia (79%), seguida pela Africana (14%) e Ameríndia (7%) (Ferreira, Mendes, Vieira, Luizon, & Simoes, 2006).

Uma das metodologias mais utilizadas atualmente para inferência do perfil de ancestralidade de um indivíduo (e controlar estratificação populacional) é a análise de Marcadores Informativos de Ancestralidade (AIMs, do inglês *ancestry informative markers*), que são *loci* (SNPs, microssatélites, inserções/ deleções, entre outros) que apresentam alelos com frequências significativamente diferentes entre populações de regiões geográficas distintas (Halder & Shriver, 2003; Phillips, 2015). Usando um conjunto de AIMs, é possível determinar as proporções de ancestralidade de cada indivíduo. Da mesma forma, as associações entre ancestralidade biogeográfica e fenótipos quantitativos ajudam a desconstruir as fontes de variação que contribuem para o risco da doença (isto é, genética *vesus* efeitos ambientais e interações gene com o ambiente) e, assim, identificar mecanismos genéticos subjacentes às doenças (Gomes et al., 2017). Um conjunto de marcadores, SNPforID 34-plex, foi utilizado para a estratificação populacional das nossas amostras.

Embora dezenas de alelos diferentes de *KIR2DL4* tenham sido descritos, não se sabe a variabilidade inerente de suas regiões regulatórias e a frequência dessas variantes em uma população real e heterogênea como a brasileira. Ainda, os poucos estudos realizados sobre o *KIR2DL4* restringiram-se à busca de variantes já conhecidas ou usaram metodologias que avaliam apenas alguns éxons e ignoram regiões regulatórias.

Ainda, o gene *HLA-G*, cuja molécula é um dos principais ligantes para KIR2DL4, sofre influência da seleção natural em pelo menos dois níveis: seleção balanceadora em suas regiões regulatórias e seleção purificadora em seu segmento codificador (Castelli et al., 2011). Como a molécula HLA-G é virtualmente invariável, espera-se que seu ligante também seja conservado ou que a variabilidade eventualmente encontrada não altere a capacidade de expressão da molécula ou sua capacidade de se ligar ao HLA-G. Assim, esse estudo é

fundamental para determinar as forças evolutivas que vem atuando para manter as respostas imunológicas finamente orquestradas.

Por fim, este estudo permitirá a criação de um banco de dados da variabilidade inerente a este gene em uma população normal do estado de São Paulo, que poderá ser usado para estudos futuros de associação com doenças, além de detectar potenciais polimorfismos funcionais no gene *KIR2DL4*.

1.6 Projeto *1000Genomes*

O projeto *1000Genomes* (*The 1000Genomes Project Consortium, 2010; The 1000Genomes Project Consortium, 2012*) foi um consórcio internacional que, utilizando diferente técnicas de sequenciamento, avaliou o genoma completo de 2.504 indivíduos oriundos de 26 populações diferentes. O projeto foi realizado em três fases, usando uma combinação de sequenciamento de genoma completo de baixa cobertura, sequenciamento de alta cobertura, exoma e genotipagem por meio de microarranjos. O objetivo do Projeto 1000 Genomas foi de genotipar e fornecer informações precisas de haplótipos sobre todas as formas de polimorfismo do DNA humano em diferentes populações. Especificamente, caracterizar mais de 95% das variantes que estão em regiões genômicas acessíveis às tecnologias de sequenciamento, e que têm frequência alélica de 1% ou mais em cada um dos cinco principais grupos populacionais com ascendência da Europa, Ásia Oriental, Sul da Ásia, África Ocidental e Americana (*The 1000 Genomes Project Consortium*).

O projeto 1000G já elucidou as propriedades e a distribuição de variações comuns e raras, forneceu *insights* sobre os processos que moldam a diversidade genética e o entendimento avançado da biologia de doenças. Esses dados são públicos e milhares de estudos relacionados à variabilidade e aspectos evolutivos de diferentes genes já foram realizados, além de estudos de associação para genes candidatos a diferentes doenças (Harrow et al., 2012).

Até o momento, a população brasileira não foi avaliada em nenhum dos projetos de sequenciamento completo do genoma humano, ou mesmo, alguma outra população tão miscigenada quanto a nossa. Devemos lembrar que a população brasileira constitui uma das populações mais heterogêneas do mundo, fruto de mais de cinco séculos de miscigenação entre as populações de diferentes continentes, resultando na população triíbrida formada por ameríndios, africanos, europeus. Por esse motivo, resolveu-se analisar os dados presentes nesse banco de dados em um primeiro momento, para depois relacionar com os dados obtidos pelo projeto aqui proposto, que utiliza amostras brasileiras.

2. JUSTIFICATIVA

Os receptores KIR são componentes fundamentais da modulação de respostas imunes, em especial aquelas mediadas por células NK – células importantes na manutenção da gestação, combate à patógenos e imunovigilância contra neoplasias. Assim, variantes alélicas do gene *KIR2DL4* e alterações em seus níveis de expressão vêm sendo constantemente associados com susceptibilidade ou proteção a diversas doenças (Goris, Dobosi, Boonen, Nagels, & Dubois, 2009; Heidenreich et al., 2012).

A coevolução entre receptores imunológicos vem atraindo a atenção dos pesquisadores nos últimos anos, de modo que avaliar a variabilidade dos genes *KIR* e em conjunto com os genes do HLA de classe I, em especial o *HLA-G* no caso do gene *KIR2DL4*, é fundamental para determinar as forças evolutivas que vem atuando para manter as respostas imunológicas finamente orquestradas. Neste contexto, em estudos anteriores, nosso grupo desenvolveu um método próprio para avaliar a variabilidade de genes *HLA* de classe I por sequenciamento de segunda geração, definindo a variabilidade completa desses genes em indivíduos do estado de São Paulo e de amostras africanas (parte dos resultados já foi publicada). A avaliação da variabilidade no gene *KIR2DL4*, nas mesmas amostras avaliadas para os genes HLA, nos trará respostas quanto aos aspectos funcionais e evolutivos da resposta imune humana. Além disso, os dados gerados constituirão um banco de dados bastante completo da variabilidade inerente desses genes em uma amostra brasileira.

3. OBJETIVOS

Este projeto tem como objetivo geral avaliar a variabilidade genética e a estrutura de haplótipos do gene *KIR2DL4* em 157 indivíduos brasileiros oriundos do estado de São Paulo, por meio de sequenciamento de segunda geração, considerando seus segmentos regulatórios, éxons e íntrons. Os objetivos específicos são:

1. Desenvolver uma metodologia para analisar a variabilidade do gene *KIR2DL4* sequenciamento de segunda geração (NGS);
2. Definir uma estratégia computacional para detecção das variantes e dos haplótipos encontrados;
3. Determinar as frequências alélica e genotípica dos pontos de variação encontrados nesse gene;
4. Classificar funcionalmente cada variante encontrada, incluindo a determinação de variantes capazes de alterar a edição do transcrito primário, mutações sinônimas e não sinônimas, entre outras;
5. Inferir a existência de Desequilíbrio de Ligação (LD) entre as variantes encontradas;
6. Definir os haplótipos encontrados em cada gene e em cada indivíduo;
7. Elaborar um banco de dados de variabilidade desses genes para uso em estudos futuros de associação com doenças e estudos evolutivos.

REFERÊNCIAS BIBLIOGRÁFICAS

AHERN, D. J.; BRENNAN, F. M. The role of Natural Killer cells in the pathogenesis of rheumatoid arthritis: Major contributors or essential homeostatic modulators? **Immunology Letters,** v. 136, n. 2, p. 115-121, May 2011. ISSN 0165-2478. Disponível em: < <Go to ISI>://WOS:000289820200001 >.

AVRIL, T. et al. The membrane-proximal immunoreceptor tyrosine-based inhibitory motif is critical for the inhibitory signaling mediated by siglecs-7 and-9, CD33-related Siglecs expressed on human monocytes and NK cells. **Journal of Immunology,** v. 173, n. 11, p. 6841-6849, Dec 2004. ISSN 0022-1767. Disponível em: < <Go to ISI>://WOS:000225307500042 >.

BIRON, C. A. et al. Natural killer cells in antiviral defense: Function and regulation by innate cytokines. **Annual Review of Immunology,** v. 17, p. 189-220, 1999. ISSN 0732-0582. Disponível em: < <Go to ISI>://WOS:000080436600007 >.

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. ISSN 0737-4038. Disponível em: < <Go to ISI>://WOS:000296336600012 >.

DORIA, A. et al. Autoinflammation and autoimmunity: Bridging the divide. **Autoimmunity Reviews,** v. 12, n. 1, p. 22-30, Nov 2012. ISSN 1568-9972. Disponível em: < <Go to ISI>://WOS:000310255700006 >.

FADDA, L. et al. Peptide antagonism as a mechanism for NK cell activation. **Proceedings of the National Academy of Sciences of the United States of America,** v. 107, n. 22, p. 10160-10165, Jun 2010. ISSN 0027-8424. Disponível em: < <Go to ISI>://WOS:000278246000047 >.

FAURE, M.; LONG, E. O. KIR2DL4 (CD158d), an NK cell-activating receptor with inhibitory potential. **Journal of Immunology,** v. 168, n. 12, p. 6208-6214, Jun 2002. ISSN 0022-1767. Disponível em: < <Go to ISI>://WOS:000176145300033 >.

FERREIRA, L. B. et al. Genomic ancestry of a sample population from the state of Sao Paulo, Brazil. **American Journal of Human Biology,** v. 18, n. 5, p. 702-705, Sep-Oct 2006. ISSN 1042-0533. Disponível em: < <Go to ISI>://WOS:000239914300012 >.

GOMES, K. F. B. et al. The influence of population stratification on genetic markers associated with type 1 diabetes. **Scientific Reports,** v. 7, Mar 2017. ISSN 2045-2322. Disponível em: < <Go to ISI>://WOS:000395428700001 >.

GOODRIDGE, J. P. et al. Three common alleles of KIR2DL4 (CD158d) encode constitutively expressed, inducible and secreted receptors in NK cells. **European Journal of Immunology,** v. 37, n. 1, p. 199-211, Jan 2007. ISSN 0014-2980. Disponível em: < <Go to ISI>://WOS:000243600400022 >.

GOODRIDGE, J. P. et al. KIR2DL4 (CD158d) genotype influences expression and function in NK cells. **Journal of Immunology,** v. 171, n. 4, p. 1768-1774, Aug 2003. ISSN 0022-1767. Disponível em: < <Go to ISI>://WOS:000184667400022 >.

GORIS, A. et al. KIR2DL4 (CD158d) polymorphisms and susceptibility to multiple sclerosis. **Journal of Neuroimmunology,** v. 210, n. 1-2, p. 113-115, May 2009. ISSN 0165-5728. Disponível em: < <Go to ISI>://WOS:000266739000016 >.

HALDER, I.; SHRIVER, M. D. Measuring and using admixture to study the genetics of complex diseases. **Hum Genomics,** v. 1, n. 1, p. 52-62, Nov 2003. ISSN 1473-9542 (Print)

1473-9542.

HARROW, J. et al. GENCODE: The reference human genome annotation for The ENCODE Project. **Genome Research,** v. 22, n. 9, p. 1760-1774, Sep 2012. ISSN 1088-9051. Disponível em: < <Go to ISI>://WOS:000308272800017 >.

HEIDENREICH, S. et al. Impact of the NK Cell Receptor LIR-1 (ILT-2/CD85j/LILRB1) on Cytotoxicity against Multiple Myeloma. **Clinical & Developmental Immunology**, 2012. ISSN 1740-2522. Disponível em: < <Go to ISI>://WOS:000307592900001 >.

HSU, K. C. et al. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. **Immunological Reviews,** v. 190, n. 1, p. 40-52, Dec 2002. ISSN 0105-2896. Disponível em: < <Go to ISI>://WOS:000179946400004 >.

HUMPHREY, M. B.; LANIER, L. L.; NAKAMURA, M. C. Role of ITAM-containing adapter proteins and their receptors in the immune system and bone. **Immunological Reviews,** v. 208, p. 50-65, Dec 2005. ISSN 0105-2896. Disponível em: < <Go to ISI>://WOS:000233446000005 >.

JAGER, A.; KUCHROO, V. K. Effector and Regulatory T-cell Subsets in Autoimmunity and Tissue Inflammation. **Scandinavian Journal of Immunology,** v. 72, n. 3, p. 173-184, Sep 2010. ISSN 0300-9475. Disponível em: < <Go to ISI>://WOS:000280638800002 >.

JOBIM, M. et al. Killer cell immunoglobulin-like receptor gene diversity in a Caucasian population of Southern Brazil. **International Journal of Immunogenetics,** v. 37, n. 2, p. 83-89, Apr 2010. ISSN 1744-3121. Disponível em: < <Go to ISI>://WOS:000275765800004 >.

KARRE, K. et al. Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. **Nature,** v. 319, n. 6055, p. 675-8, Feb 20-26 1986. ISSN 0028-0836 (Print)

0028-0836.

KIKUCHI-MAKI, A. et al. KIR2DL4 is an IL-2-regulated NK cell receptor that exhibits limited expression in humans but triggers strong IFN-gamma production. **Journal of Immunology,** v. 171, n. 7, p. 3415-3425, Oct 2003. ISSN 0022-1767. Disponível em: < <Go to ISI>://WOS:000185548000015 >.

LANGERS, I. et al. Natural killer cells: role in local tumor growth and metastasis. **Biologics,** v. 6, p. 73-82, 2012. ISSN 1177-5475 (Print)

1177-5475.

LEE, S. H.; MIYAGI, T.; BIRON, C. A. Keeping NK cells in highly regulated antiviral warfare. **Trends in Immunology,** v. 28, n. 6, p. 252-259, Jun 2007. ISSN 1471-4906. Disponível em: < <Go to ISI>://WOS:000247280700004 >.

LEUNG, W. Infusions of Allogeneic Natural Killer Cells as Cancer Therapy. **Clinical Cancer Research,** v. 20, n. 13, p. 3390-3400, Jul 2014. ISSN 1078-0432. Disponível em: < <Go to ISI>://WOS:000338920100006 >.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics,** v. 25, n. 14, p. 1754-1760, Jul 2009. ISSN 1367-4803. Disponível em: < <Go to ISI>://WOS:000267665900006 >.

MARSH, S. G. E. et al. Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. **Journal of Immunological Methods,** v. 281, n. 1-2, p. 1-8, Oct 2003. ISSN 0022-1759. Disponível em: < <Go to ISI>://WOS:000186487500001 >.

MARTIN, A. M. et al. The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. **Immunogenetics,** v. 51, n. 4-5, p. 268-280, Apr 2000. ISSN 0093-7711. Disponível em: < <Go to ISI>://WOS:000086924100003 >.

MARTIN, A. M. et al. Comparative genomic analysis, diversity and evolution of two KIR haplotypes A and B. **Gene,** v. 335, p. 121-131, Jun 2004. ISSN 0378-1119. Disponível em: < <Go to ISI>://WOS:000222434300011 >.

MORETTA, A. et al. P58 MOLECULES AS PUTATIVE RECEPTORS FOR MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) CLASS-I MOLECULES IN HUMAN NATURAL-KILLER (NK) CELLS - ANTI-P58 ANTIBODIES RECONSTITUTE LYSIS OF MHC CLASS-I-PROTECTED CELLS IN NK CLONES DISPLAYING DIFFERENT SPECIFICITIES. **Journal of Experimental Medicine,** v. 178, n. 2, p. 597-604, Aug 1993. ISSN 0022-1007. Disponível em: < <Go to ISI>://WOS:A1993LP83000024 >.

NORMAN, P. J. et al. Co-evolution of Human Leukocyte Antigen (HLA) Class I Ligands with Killer-Cell Immunoglobulin-Like Receptors (KIR) in a Genetically Diverse Population of Sub-Saharan Africans. **Plos Genetics,** v. 9, n. 10, Oct 2013. ISSN 1553-7404. Disponível em: < <Go to ISI>://WOS:000330367200087 >.

ORR, M. T.; LANIER, L. L. Natural Killer Cell Education and Tolerance. **Cell,** v. 142, n. 6, p. 847-856, Sep 2010. ISSN 0092-8674. Disponível em: < <Go to ISI>://WOS:000281855000008 >.

PARHAM, P. MHC class I molecules and KIRs in human history, health and survival. **Nature Reviews Immunology,** v. 5, n. 3, p. 201-214, Mar 2005. ISSN 1474-1733. Disponível em: < <Go to ISI>://WOS:000227302400011 >.

PARHAM, P.; MOFFETT, A. Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. **Nature Reviews Immunology,** v. 13, n. 2, p. 133-144, Feb 2013. ISSN 1474-1733. Disponível em: < <Go to ISI>://WOS:000314906600015 >.

PARHAM, P. et al. Review: Immunogenetics of human placentation. **Placenta,** v. 33, p. S71-S80, Feb 2012. ISSN 0143-4004. Disponível em: < <Go to ISI>://WOS:000301868000014 >.

PERRICONE, R. et al. NK cells in autoimmunity: A two-edg'd weapon of the immune system. **Autoimmunity Reviews,** v. 7, n. 5, p. 384-390, May 2008. ISSN 1568-9972. Disponível em: < <Go to ISI>://WOS:000256720000007 >.

PHILLIPS, C. Forensic genetic analysis of bio-geographical ancestry. **Forensic Science International-Genetics,** v. 18, p. 49-65, Sep 2015. ISSN 1872-4973. Disponível em: < <Go to ISI>://WOS:000362665800005 >.

PURDY, A. K.; CAMPBELL, K. S. Natural killer cells and cancer Regulation by the killer cell Ig-like receptors (KIR). **Cancer Biology & Therapy,** v. 8, n. 23, p. 2211-2220, Dec 2009. ISSN 1538-4047. Disponível em: < <Go to ISI>://WOS:000272796800001 >.

RAJAGOPALAN, S. Endosomal Signaling and a Novel Pathway Defined by the Natural Killer Receptor KIR2DL4 (CD158d). **Traffic,** v. 11, n. 11, p. 1381-1390, Nov 2010. ISSN 1398-9219. Disponível em: < <Go to ISI>://WOS:000282571500001 >.

RAJALINGAM, R. et al. Distinctive KIR and HLA diversity in a panel of north Indian Hindus. **Immunogenetics,** v. 53, n. 12, p. 1009-1019, Mar 2002. ISSN 0093-7711. Disponível em: < <Go to ISI>://WOS:000174860100003 >.

RAULET, D. H.; VANCE, R. E.; MCMAHON, C. W. Regulation of the natural killer cell receptor repertoire. **Annual Review of Immunology,** v. 19, p. 291-330, 2001. ISSN 0732-0582. Disponível em: < <Go to ISI>://WOS:000168234600012 >.

SAUNDERS, P. M. et al. The interaction of KIR3DL1*001 with HLA class I molecules is dependent upon molecular microarchitecture within the Bw4 epitope. **J Immunol,** v. 194, n. 2, p. 781-789, Jan 15 2015. ISSN 1550-6606 (Electronic)

0022-1767 (Linking). Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/25480565 >.

SCHLEINITZ, N. et al. Natural killer cells in human autoimmune diseases. **Immunology,** v. 131, n. 4, p. 451-458, Dec 2010. ISSN 0019-2805. Disponível em: < <Go to ISI>://WOS:000283949900001 >.

SHILLING, H. G. et al. Evidence for recombination as a mechanism for KIR diversification. **Immunogenetics,** v. 48, n. 6, p. 413-416, Nov-Dec 1998. ISSN 0093-7711. Disponível em: < <Go to ISI>://WOS:000076754500007 >.

STAUB, E.; ROSENTHAL, A.; HINZMANN, B. Systematic identification of immunoreceptor tyrosine-based inhibitory motifs in the human proteome. **Cellular Signalling,** v. 16, n. 4, p. 435-456, Apr 2004. ISSN 0898-6568. Disponível em: < <Go to ISI>://WOS:000188293600004 >.

STEFFENS, U. et al. Nucleotide and amino acid sequence alignment for human killer cell inhibitory receptors (KIR), 1998. **Tissue Antigens,** v. 51, n. 4, p. 398-413, Apr 1998. ISSN 0001-2815. Disponível em: < <Go to ISI>://WOS:000073017600012 >.

STEWART, C. A.; VAN BERGEN, J.; TROWSDALE, J. Different and divergent regulation of the KIR2DL4 and KIR3DL1 promoters. **Journal of Immunology,** v. 170, n. 12, p. 6073-6081, Jun 2003. ISSN 0022-1767. Disponível em: < <Go to ISI>://WOS:000183411100033 >.

SUTO, Y. et al. Chromosomal localization of the human natural killer cell class I receptor family genes to 19q13.4 by fluorescence in situ hybridization. **Genomics,** v. 35, n. 1, p. 270-272, Jul 1996. ISSN 0888-7543. Disponível em: < <Go to ISI>://WOS:A1996UV86900041 >.

THIELENS, A.; VIVIER, E.; ROMAGNE, F. NK cell MHC class I specific receptors (KIR): from biology to clinical intervention. **Current Opinion in Immunology,** v. 24, n. 2, p. 239-245, Apr 2012. ISSN 0952-7915. Disponível em: < <Go to ISI>://WOS:000303187600019 >.

TROWSDALE, J. et al. The genomic context of natural killer receptor extended gene families. **Immunological Reviews,** v. 181, p. 20-38, Jun 2001. ISSN 0105-2896. Disponível em: < <Go to ISI>://WOS:000170341300002 >.

UHRBERG, M. The KIR gene family: life in the fast lane of evolution. **European Journal of Immunology,** v. 35, n. 1, p. 10-15, Jan 2005. ISSN 0014-2980. Disponível em: < <Go to ISI>://WOS:000226402800002 >.

UHRBERG, M. et al. Human diversity in killer cell inhibitory receptor genes. **Immunity,** v. 7, n. 6, p. 753-763, Dec 1997. ISSN 1074-7613. Disponível em: < <Go to ISI>://WOS:000071351300003 >.

VALIANTE, N. M. et al. Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. **Immunity,** v. 7, n. 6, p. 739-751, Dec 1997. ISSN 1074-7613. Disponível em: < <Go to ISI>://WOS:000071351300002 >.

VILCHES, C.; PARHAM, P. KIR: Diverse, rapidly evolving receptors of innate and adaptive immunity. **Annual Review of Immunology,** v. 20, p. 217-251, 2002. ISSN 0732-0582. Disponível em: < <Go to ISI>://WOS:000175012600010 >.

VIVIER, E. et al. Innate or Adaptive Immunity? The Example of Natural Killer Cells. **Science,** v. 331, n. 6013, Jan 2011. ISSN 0036-8075. Disponível em: < <Go to ISI>://WOS:000285974000029 >.

WILSON, M. J.; TORKAR, M.; TROWSDALE, J. Genomic organization of a human killer cell inhibitory receptor gene. **Tissue Antigens,** v. 49, n. 6, p. 574-579, Jun 1997. ISSN 0001-2815. Disponível em: < <Go to ISI>://WOS:A1997XG75400003 >.

WITT, C. S.; MARTIN, A.; CHRISTIANSEN, F. T. Detection of KIR2DL4 alleles by sequencing and SSCP reveals a common allele with a shortened cytoplasmic tail. **Tissue Antigens,** v. 56, n. 3, p. 248-257, Sep 2000. ISSN 0001-2815. Disponível em: < <Go to ISI>://WOS:000089523200007 >.

WITT, C. S. et al. Alleles of the KIR2DL4 receptor and their lack of association with pre-eclampsia. **European Journal of Immunology,** v. 32, n. 1, p. 18-29, Jan 2002. ISSN 0014-2980. Disponível em: < <Go to ISI>://WOS:000173527400003 >.

YAN, W. H. et al. Possible roles of KIR2DL4 expression on uNK cells in human pregnancy. **American Journal of Reproductive Immunology,** v. 57, n. 4, p. 233-242, Apr 2007. ISSN 1046-7408. Disponível em: < <Go to ISI>://WOS:000244793000001 >.

Capítulo II

4. ARTIGO

KIR2DL4 variability and haplotypes in a Brazilian population sample

Emiliana Weiss[1,2], Heloisa S. Andrade[1,2], Andreia S. Souza[1,2], Thálitta H. A. Lima[1,2], Juliana R. Lara[3], Marília R. S. Passos[1,4], Eduardo A. Donadi [5], Maria Luiza G. Oliveira[6], Celso T. Mendes-Junior [6,7], Erick C. Castelli [1,2,4]

Institution: (1) Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unity, School of Medicine - UNESP, Botucatu – SP, Brazil; (2) Graduate Program on Biological Sciences (Genetics), Bioscience's Institute of Botucatu, UNESP – São Paulo State University, Brazil; (3) Graduate Program on Anesthesiology, School of Medicine, UNESP – São Paulo State University, Brazil; (4) Graduate Program on Pathology, School of Medicine of Botucatu, UNESP – São Paulo State University, Brazil; (5) School of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto – SP, Brazil; (6) Graduate Program on Genetics, School of Medicine of Ribeirão Preto, University of São Paulo (USP), Brazil; (7) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brasil.

E-mail: emiliana.weiss@unesp.br; erick.castelli@unesp.br

4.1 Introduction

Natural killer (NK) cells are among the major effectors in innate immunity, and they also influence adaptive immunity. As result, these cells integrate these two complex systems. Natural killer cells play a vital role in the immune response from virus infection (Perricone et al., 2008) to tumors (Langers et al., 2012; Purdy & Campbell, 2009) through the regulation and balancing signals transmitted by activating and inhibitory receptors. KIRs (Killer Cell Immunoglobulin-Like receptor) are expressed on the surface of NK and some T cell, and these receptors interactions with their ligands regulating the cytotoxicity activity of NK cells (Marsh et al., 2003; Miah, Hughes, & Campbell, 2008; Vilches & Parham, 2002).

KIRs are transmembrane glycoproteins with two (2D) or three (3D) Ig-like domains and cytoplasmic domains, which vary in length and size. KIRs with long (L) cytoplasmic domains (KIR2DL and KIR3DL) are mostly inhibitory receptors that may contain one or two cytoplasmic ITIMs (immunoreceptor tyrosine-based inhibitory motif) (Avril et al., 2004; Staub et al., 2004). Those KIRs with short (S) cytoplasmic domains (KIR2DS and KIR3DS) are activating receptors with ITAM (immunoreceptor tyrosine-based activate motif) associated with the adaptor DAP12 via a charged lysine residue in their transmembrane region (Altshuler et al., 2015; Martin et al., 2000).

KIR2DL4 is the most conserved and unusual (KIR) family member in terms of genetic structure, expression levels, cellular localization, and signaling properties (Miah et al., 2008;

Parham, 2005a). Also, it is unique among KIRs regarding its genomic structure. Unlike other *KIR* family members, *KIR2DL4* is found in almost all *KIR* haplotypes (Martin et al., 2000; Valiante et al., 1997), but its surface expression appears to be variable between individuals. KIR2DL4 presents a single ITIM domain in its cytoplasmic tail and a positively charged arginine in the transmembrane region. Therefore, KIR2DL4 may either activate or inhibit NK cells depending on the ligand (Faure & Long, 2002).

The *KIR2DL4* genomic sequence comprises approximately 11-kb. There are at least 70 alleles (or coding sequences) already described for this gene, encoding 41 protein variants according to the IPD-IMGT/KIR Database, *version 2.8.0.* Polymorphisms in both regulatory and coding regions may influence its expression pattern and also the strength of the receptor/ligand binding.

The only known ligand of KIR2DL4 is the HLA-G molecule, which is produced by fetal-derived trophoblast cells. Also, HLA-G is unusual among HLA molecules in its unique pattern of expression in healthy individuals. The signals that induce the activation of uterine NK cells are still unknown, but KIR2DL4 interaction with HLA-G may play an important role in this process (Rajagopalan & Long, 2012). Hence, *KIR2DL4* polymorphism may influence the interaction of decidual NK cells with HLA-G expressed on trophoblastic cells, preventing surface expression. Polymorphisms might disrupt the ligand-receptor binding, or even influencing KIR expression. However, KIR variability is not well explored, mainly in admixed populations such as Brazilian, although *KID2DL4* is related to many pathological conditions such as recurrent abortion and preeclampsia (Yan et al., 2007).

Here we present a molecular and computational methodology to evaluate the entire *KIR2DL4* gene segment, including its upstream promoter region, all exons and introns, using massively parallel sequencing, and freely available computer softwares for genotyping and haplotyping. This methodology was applied to evaluate *KIR2DL4* genetic variability in 157 samples from Southeast Brazil.

4.2 Material and Methods

4.2.1 Samples

KIR2DL4 variability was evaluated in 157 unrelated individuals (82.8% female, mean age of 31.3 years old), randomly recruited, from the State of São Paulo, Brazil. All participants

signed informed consent before blood collection. The local Human Research Ethics Committee approved the study protocol (Protocol 24157413.7.0000.5411). It is important to consider that these samples do not necessarily represent the Brazilian population, however, considering the Brazilian admixed nature it is a heterogeneous sample, a great repository of genetic diversity, and this sample was used mainly for methodological goals. The inter-ethnic admixture estimates for this population sample was 71.4% European, 19.6% African and 9% Asian/Amerindian. DNA was obtained by salting out procedures and normalized to a final concentration of 50 ng/μL after quantification with Qubit Broad Range dsDNA Assay (Thermo Fisher Scientific Inc., Waltham, MA).

4.2.2 Samples ancestry

In order to estimate the ancestry contributions in this Brazilian sample, we used 34 Ancestral Informative Markers (AIM) from the SNPforID 34-plex panel (Phillips, 2015). We used STRUCTURE v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000) for population structure analysis, defining K=3 and considering the 1000 Genomes Project data from European, African and East Asian populations as parental references. In total, 1.412 individuals from the 1000 Genomes Project (phase 3) were used to represent the parental populations in our ancestry analysis: 404 Europeans (TSI, FIN, GBR, IBS), 504 Africans (YRI, LWK, GWD, MSL, ESN), and 504 East Asians (CHB, JPT, CHS, CDX, KHV). Since the 1000 Genomes Project does not include a Native American group that could represent the Native American ancestry, we used the East Asian group as a proxy for those (Posth et al., 2018). Levels of admixture were inferred based on cluster information.

We stratified our samples within five subgroups: individuals with more than 90% of European ancestry (EUR 90), individuals with more than 70% of European ancestry (EUR70), individuals with more than 50% of African ancestry (AFR50), individuals with more than 30% of African ancestry (AFR30), and individuals with more than 30% of East Asian ancestry (EAS30).

4.2.3 *KIR2DL4* amplification, library preparation and sequencing

We designed specific primers to amplify a *KIR2DL4* segment of 13,387 bp comprising the promoter region, all exons (including the 3'UTR) and all introns. However, due to the length of the *KIR2DL4* gene (10,906 bp), we used three overlapping fragments covering the entire gene. Primers were designed in segments that are conserved (no polymorphisms) regarding

the *1000Genomes* project and the dbSNP databases.

Polymerase Chain Reaction (PCR) was carried out using primers: (a) 5'-CTATTGTATGTCTTGGTGAC-3' and 5'-CTTGATACTGGAATATTGCA-3', which amplify the position 54801519 to 54802921, chromosome 19 (*human genome draft assembly* hg38), producing an amplicon of 1,402 bp that covers the *KIR2DL4* promoter region*;* (b) 5'-GGGCAAACAGTGAGACCCAT-3' and 5'-CAGGCATTTGTCCTCCCAGT-3', which amplify from position 54802741 to 54809190 (hg38), resulting in a final product of 6,450 bp that covers first half of the *KIR2DL4* coding region; and (3) 5'-CTCTGAGATAAAACCCATTGTAA-3' and 5'- TTCAATAAACACCTGTAAATCCC-3' (hg38), which amplify from position 54807865 to 54814906 (hg38) resulting in a amplicon of 7,045 bp that includes the second half of the *KIR2DL4* coding region and the 3'UTR.

PCR reaction was carried out in a final volume of 25 µL. For the promoter region, 1.0 U of Platinum® Taq DNA polymerase and 1X buffer solution, 1.5mM of MgCl2, 0.20mM of each dNTP (Invitrogen-Carlsbad, CA, USA) and 0.30µM of each primer were used. The PCR cycling conditions were as follow: initial denaturation at 94°C by 1 minute; 32 denaturation cycles at 94°C by 30 seconds, annealing at 55°C by 30 seconds, extension at 72°C by 2 minutes; followed by final extension at 72°C for 5 minutes and a hold at 4°C. We used the same protocol for the two segments of the coding region: 1.25U of TaKaRa GXL Taq® DNA polymerase and 1X buffer solution, 0.20mM of dNTP and 0.30µM of each primer. The cycling conditions were: 30 cycles 98°C by 10 seconds denaturation, 60°C for 15 seconds annealing, 68°C for 7 minutes extension, and one hold at 4°C. For each reaction, 100ng of genomic DNA was used as template. Amplicons were evaluated on 1% agarose gel stained with GelRed® (Biotium, Inc. Hayward, CA), quantified using Qubit dsDNA High Sensitivity Assayes, normalized, pooled together and purified using ExoSAP-IT (GE Healthcare).

For library preparation, we followed recommendations for the Illumina Nextera XT Sample Preparation Kit, multiplexing 96 samples with the Nextera XT Index Kit (Illumina, Inc., San Diego, CA). However, we did not carry out the library normalization step, passing directly to library quantification with qPCR (Kapa Biosystems, Wilmington, MA). We used Agilent High Sensitivity DNA Kits (Agilent Technologies, CA, USA) to evaluate the fragmentation. Finally, libraries were normalized to the recommended concentration for MiSeq Reagent Kit version V2 (500 cycles, 2 x 250-bp).

4.2.4 Data processing

We attempted to mapping reads using the reference genome of chromosome 19 (hg38 version) and the aligner BWA-MEM (Li & Durbin, 2009). However, due to the high similarity between KIR genes and high rate of polymorphisms, the end products of these alignments were quite unsatisfactory, with large numbers of erroneously mapped or unmapped sequences. It is important to remember that KIR genes are very similar to each other due to a common paralogous origin (Vilches & Parham, 2002). For this reason, mapping bias can occur in the presence of differences that diminish the identity of a given KIR read with its corresponding haplotype sequence at the reference genome and increasing its similarity with a homologous KIR gene.

Therefore, it was necessary to adopt a strategy similar to the one used for HLA genes to obtain adequate mappings, in which a series of filters are applied to direct each sequence pair to the correct location in the genome [see the description this methodology in (Castelli, Paz, Souza, Ramalho, & Mendes-Junior, 2018). We have developed a new software capable of processing the KIR gene sequences, using as a reference all sequences of KIR genes already described in public databank. This software is available upon request.

The Genome Analysis Toolkit (GATK, version 3.6) HaplotypeCaller (McKenna et al., 2010; Van der Auwera et al., 2013) was used to infer genotypes, using the GVCF (variant call format)mode and applied for each sample separately. Next, we concatenate all genotypes in a VCF file with GATK GenotypeGVCFs. Variants were annotated based on the dbSNP database (September 2018). Inferred genotypes were treated for correction of uncertain genotypes, especially those inferred in regions with low or disproportional coverage, using the vcfx function *checkpl* program (program available at www.castelli-lab.net), which ensures that only genotypes with high quality continue to the next steps of analysis.

The haplotypes variables were inferred by combining two computational strategies. The first uses the GATK *ReadBackedPhasing* routine, considering minimum base quality set to 20, and the *phaseQualityThresh* defined as 2000, to verify the phase between the heterozygous sites of each sample. These haplotypes are obtained physically, observing the co-occurrence of variants in the same read or fragment. However, eventually, the distance between some of these polymorphic sites does not allow such inference. Moreover, *ReadBackedPhasing* does

not evaluate multi-allelic and indels polymorphisms. Thus, for these variants, the inference was performed using the PHASE algorithm (Stephens, Smith, & Donnelly, 2001), which also imputed the missing alleles observed after the vcfx treatment and considered the phase observed by ReadBackedPhasing. This strategy has been used for HLA genes (Castelli et al., 2017; Castelli et al., 2015; Lima et al., 2016; Sonon et al., 2018). In this way, it is possible to use the phase information acquired in the previous step and statistically infer what would be the most likely haplotype in these regions, based on the observations of all the samples at the same time. With the haplotypes determined in a phased VCF file (in the same format as the *1000Genomes* project), the complete sequence of each individual was determined using the vcfx fasta program (www.castelli-lab.net/apps/vcfx).

Allelic and haplotype frequencies were calculated by direct counting. Genetic diversity analyzes, including adherence to Hardy-Weinberg equilibrium (HWE), nucleotide diversity, haplotype diversity, among others, were performed using the software ARLEQUIN 3.5 (Excoffier & Lischer, 2010), converting the phased VCF into ARP files using *vcfx arlequin*. The frequency of each haplotype was computed by the direct counting method. A graphical overview of the linkage disequilibrium (LD) plot was generated using Haploview software version 4.2 (Barrett, Fry, Maller, & Daly, 2005), excluding markers with minor allele frequency (MAF) below 1%. The haplotype blocks were defined by the confidence intervals method implemented in the program.

This phased VCF file was converted into *KIR2DL4* coding sequence (CDS) sequences by using the GRhg38 reference sequence as a draft and replacing the correct nucleotide in each position, two sequences per sample, by using *vcfx fasta*. Then, we used a local BLAST server with databases containing all known *KIR2DL4* sequences downloaded from the IPD-IMGT/KIR Database version 2.8.0, and the closest known coding allele was defined for each haplotype.

4.2.5 Evaluation of *KIR2DL4* gene variability using data from the *1000Genomes* Project

In order to verify the genetic variability of *KIR2DL4* in different populations, the analysis of *KIR2DL4* gene present in the public database *1000Genomes* was carried out. The variability data were obtained from the official *1000Genomes* Project browser (http://browser.1000genomes.org/index.html). The download was performed containing the data of 2.504 individuals, using the positions of the genome assembly GRCh37 (hg37)

19:55312974 to 19:55326361, resulting in a total of 13,387 bp (overcoming the same regions amplified by the primers described here).

4.3. Results

4.3.1 Genetic variability within the *KIR2DL4* gene

We detected 152 variable sites on *KIR2DL4* and surrounding sequences, considering 157 samples from Brazil (Table S1). Many of these variable sites (11.18%) occurred just once as singletons. Nevertheless, they were included in the haplotype analysis because they were properly phased by GATK or they occurred at complete homozygous samples exception made to that singleton. The *KIR2DL4* segment can be considered highly polymorphic since at least 93 variable sites (61.18%) presented a minor allele frequency MAF > 1% and many (41.44%) presented MAF > 30% (Table S1). All the variable sites detected and their genotype frequencies are under the HWE expectations. Some of those variants described here are not included in the IPD-IMGT/KIR Database, since the aforementioned databank tracks variants only between the positions 54803385 to 54814518 (hg38). Besides, we found 14 new variants that have not yet been recorded on database online.

Additionally, we detected 21 singletons (Table S2). Given the great distance between these singletons and the nearest variable sites and, since they occurred just once in samples carrying other heterozygous sites, or some samples presented missing alleles at this particular variable site, it was not possible to properly phase these singletons to infer haplotypes. As a result, they were not included in the haplotype analysis, mainly because the PHASE algorithm does not handle singletons properly.

We will be presented the *KIR2DL4* genetic variability and haplotypes following the segments defined in Figure 1, starting from the promoter and the 5'upstream region (Table 1), followed by the coding haplotypes (Tables 2 and 3), the 3'UTR haplotypes (Table 5), and lastly the haplotypes defined for each segment will be combined as extended haplotypes (Table 6).

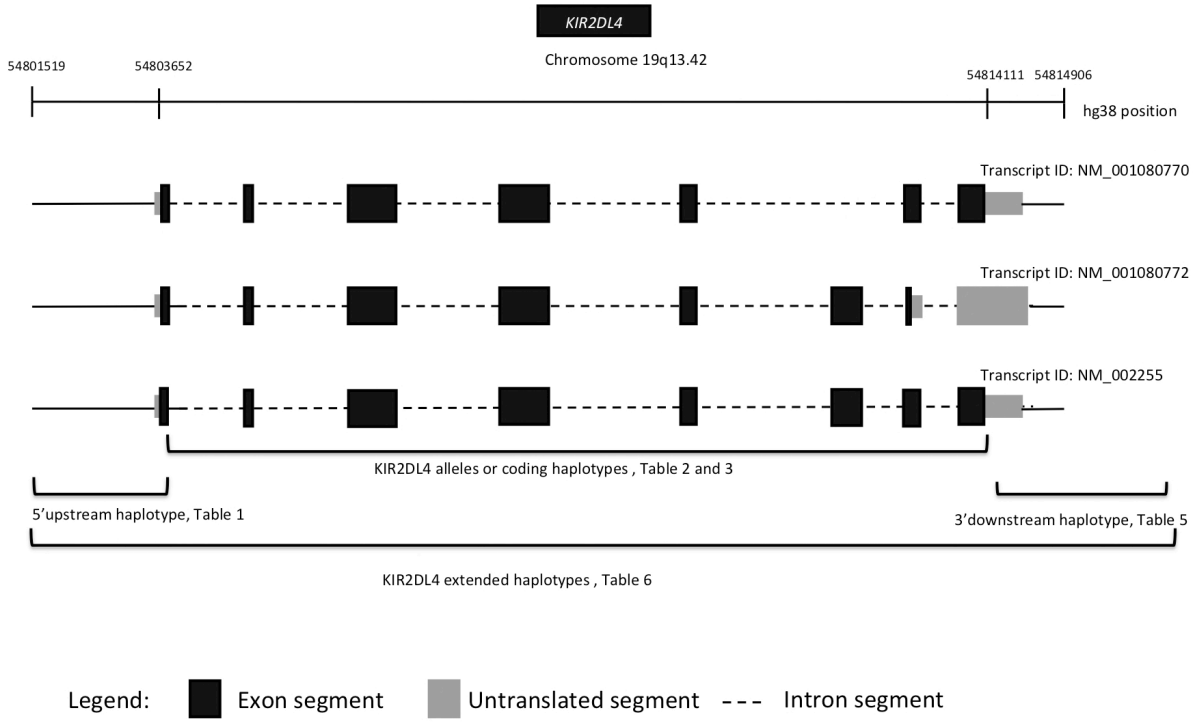

Fig 1. *KIR2DL4* gene structure and known transcripts.

We found 16 variable sites upstream from the first translated ATG, in a segment of approximately 2-kb (Table S1), most of them with MAF higher than 1% in this Brazilian sample. These 16 variants are arranged in 11 haplotypes (Table 1). The haplotype that is identical to the one described in the reference genome (with only reference alleles) was named P1. Since this region is not entirely considered by the IPD-IMGT/KIR database, and since there is no nomenclature pattern defined for it, haplotypes were named as P1 to P11, following sequence similarities and the number of alternative alleles. The most frequent haplotypes were P8 (27%), P3 (23%), P1 (20%), P5 (12.1%), and P7 (10.2%), with a summed frequency of 92.3%.

Table 1. List of the KIR2DL4 upstream and 5' untranslated region haplotypes found in a Brazilian population sample. The positions recorded here refer to the 1000Genome project (hg38).

Chr19 position	54802354	54802427	54802632	54802639	54802679	54802832	54803022	54803145	54803290	54803346	54803350	54803423	54803425	54803437	54803442	54803494						
SNPid	rs3916556	rs111803848	rs35250545	rs35160118	rs34370734	rs589158	-	rs376798518	rs55881846	rs144978425	rs36116896	rs55808487	rs56401846	rs933157437	rs35656676	rs17173106	Frequency (2n=314)	EAS 30 (2n=16)	AFR 50 (2n=14)	AFR 30 (2n=40)	EUR 90 (2n=92)	EUR 70 (2n=236)
P1	C	G	C	T	G	A	A	C	G	C	T	C	G	G	G	T	0.204	0.063	0.000	0.000	0.085	0.254
P2	C	G	C	T	A	A	A	C	G	C	T	C	G	G	G	T	0.003	0.000	0.000	0.000	0.004	0.004
P3	C	G	C	T	G	G	A	C	G	CA	G	C	G	G	G	T	0.232	0.250	0.075	0.225	0.119	0.246
P4	C	G	C	T	G	G	A	C	G	CA	G	C	G	A	G	T	0.003	0.000	0.000	0.000	0.000	0.000
P5	C	T	C	T	G	G	A	C	G	CA	G	C	G	G	G	T	0.121	0.188	0.000	0.125	0.042	0.110
P6	C	G	C	T	G	G	G	C	G	CA	G	C	G	G	G	T	0.003	0.000	0.000	0.000	0.000	0.004
P7	T	G	A	C	A	G	A	C	G	CA	G	C	G	G	G	T	0.102	0.125	0.214	0.100	0.042	0.102
P8	T	G	A	C	A	G	A	C	G	CA	G	C	G	G	C	C	0.277	0.250	0.429	0.425	0.089	0.250
P9	T	G	A	C	A	G	A	C	G	CA	G	T	A	G	C	C	0.029	0.000	0.143	0.100	0.004	0.013
P10	T	G	A	C	A	G	A	T	G	CA	G	C	A	G	C	C	0.013	0.000	0.000	0.000	0.004	0.013
P11	T	G	A	C	A	G	A	C	T	CA	G	C	G	G	C	C	0.013	0.125	0.000	0.025	0.000	0.004
																	Gene diversity	0.8040 +/- 0.0087				
																	Nucleotide diversity	0.001996 +/- 0.001098				

The alternative alleles are represented with a grey background.

Haplotypes were named considering the number of changes in bp.

We detected 127 variable sites in the CDS, from the first translated codon until the stop codon (Table S1). These variants are arranged into 70 coding haplotypes or CDS alleles, named following the haplotypes available in the IPD-IMGT/KIR Database *version 2.8.0* (Table 2). Most of the sequences were identical to the one already reported by the aforementioned database, which includes copies of the alleles: *KIR2DL4*00501* (15.9%), *KIR2D14*011* (8.6%), *KIR2DL4*0010305* (3.5%), *KIR2DL4*0010304* (0.6%), *KIR2DL4*006* (1.3%), *KIR2DL4*0080105* (0.3%), *KIR2DL4*0080204* (8%). The remaining sequences configure possible new *KIR2DL4* coding alleles. In these cases, the closest known *KIR2DL4* allele was identified, followed by all the observed mutations.

Table 2. List of KIR2DL4 coding haplotypes or coding alleles found in Brazilian population sample.

KIR2DL4 allele or coding haplotypes	Frequency (2n=314)
*KIR2DL4*011*	8.60%
*KIR2DL4*0080204*	7.96%
KIR2DL4*0010306 (A>T Position 7625)	3.50%
*KIR2DL4*0010305*	3.50%
*KIR2DL4*0080105* (new - 4 mutations)	3.18%
*KIR2DL4*0010201* (A>T Position 7625); (A Del 9620)	2.87%
*KIR2DL4*00501*	15.92%
*KIR2DL4*0080105*	14.01%
*KIR2DL4*0010201* (A>T Position 7622)	12.10%
*KIR2DL4*0010304* (new - 8 mutations)	1.59%
*KIR2DL4*0080204* (A>G Position 958)	1.59%
*KIR2DL4*0010201* (G>A Position 692); (A>T Position 7625)	1.27%
*KIR2DL4*00602*	1.27%
*KIR2DL4*0010304* (new - 7 mutations)	0.96%
*KIR2DL4*00501* (A>G Position 5077)	0.96%
*KIR2DL4*008* (new - 5 mutations)	0.96%
*KIR2DL4*0010201* (T>C Position 5988); (A>T Posição 7902)	0.64%
*KIR2DL4*0010304*	0.64%
*KIR2DL4*0010305* (T>C Position 1021)	0.64%
*KIR2DL4*00501* (A Del at position 9620)	0.64%
*KIR2DL4*00501* (A>C Position 219)	0.64%
*KIR2DL4*00501* (new - 24 mutations)	0.64%
*KIR2DL4*006* (new - 42 mutations)	0.64%
*KIR2DL4*0080105* (C>T Position 494); (G>T Position 4923)	0.64%
*KIR2DL4*001* (new - 3 mutations)	0.32%
*KIR2DL4*001* (new - 3 mutations)	0.32%
*KIR2DL4*001* (new - 3 mutations)	0.32%
*KIR2DL4*001* (new - 3 mutations)	0.32%
*KIR2DL4*0010201* (A>T Position 7625); (A>G Position 9874)	0.32%

*KIR2DL4*0010201* (G>C Position 7020); (A>T Position 7625)	0.32%
*KIR2DL4*0010201* (new - 3 mutations)	0.32%
*KIR2DL4*0010304* (A>C Position 3981)	0.32%
*KIR2DL4*0010304* (new - 7 mutations)	0.32%
*KIR2DL4*0010305* (G>C Position 1968)	0.32%
*KIR2DL4*0010305* (new - 3 mutations)	0.32%
*KIR2DL4*0010305* (T Del at position 4607)	0.32%
*KIR2DL4*0010306* (G>A Position 692); (A>T Position 7625)	0.32%
*KIR2DL4*0010306* (new - 4 mutations)	0.32%
*KIR2DL4*005* (G>A Position 10392); (C>A Position 10753)	0.32%
*KIR2DL4*005* (G>A Position 4551); (G>A Position 7832)	0.32%
*KIR2DL4*005* (new - 20 mutations)	0.32%
*KIR2DL4*005* (new - 18 mutations)	0.32%
*KIR2DL4*00501* (A insert at position 6909); (G>A Position 7743)	0.32%
*KIR2DL4*00501* (C>T Position 2028); (G>A Position 6033)	0.32%
*KIR2DL4*00501* (C>T Position 3199)	0.32%
*KIR2DL4*00501* (G>A Position -214); (C>T Position 3199)	0.32%
*KIR2DL4*00501* (G>A Position 266)	0.32%
*KIR2DL4*00501* (G>A Position 4551)	0.32%
*KIR2DL4*00501* (G>A Position 5172); (A Del at position 9620)	0.32%
*KIR2DL4*00501* (G>C Position 1389); (G>A Position 10508)	0.32%
*KIR2DL4*00501* (T Del at position 6689); (A Del 9620)	0.32%
*KIR2DL4*00501* (T inser at position 6689)	0.32%
*KIR2DL4*00501* (T>A Position 2061)	0.32%
*KIR2DL4*006* (new - 49 mutations)	0.32%
*KIR2DL4*00602* (G>A Position 266)	0.32%
*KIR2DL4*00602* (new - 40 mutations)	0.32%
*KIR2DL4*008* (new - 4 mutations)	0.32%
*KIR2DL4*008* (new - 4 mutations)	0.32%
*KIR2DL4*0080105*	0.32%
*KIR2DL4*0080105* (C>A Position 3981); (G>T Position 4923)	0.32%
*KIR2DL4*0080105* (G>C Position 2398); (G>T Position 4923)	0.32%
*KIR2DL4*0080105* (G>T Position 4923); (G>A Position 10392)	0.32%
*KIR2DL4*0080105* (new - 4 mutation)	0.32%
*KIR2DL4*0080204* (A>C Position 5156)	0.32%
*KIR2DL4*0080204* (G>A Position 10392); (C>A Position 10753)	0.32%
*KIR2DL4*0080204* (G>C Position 7103)	0.32%
*KIR2DL4*0080204* (T inser at position 6689)	0.32%
*KIR2DL4*011* (G>A Position 3974)	0.32%
*KIR2DL4*011* (G>A Position 7910)	0.32%
*KIR2DL4*011* (T>C Position 2750)	0.32%

Gene diversity 0.9212 +>- 0.0070

Nucleotide diversity 0.004887 +>- 0.002344

Haplotype names were given according to the closest known KIR2DL4 allele followed by the differences>divergences that were observed for this given haplotype according to IMGT>KIR database. The word "new" refers to variable sites that are not currently described at the IMGT>HLA database

Although 70 haplotypes have been reported, those are responsible for encoding 13 different full-length protein molecules since most of the variants lay on intronic segments (Table 3). The most frequent protein molecules were KIR2DL4*008 (31.53%), KIR2DL4*001 (29.30%), KIR2DL4*005 (20.38%), and KIR2DL4 *011 (11.46%). Table 3 also describes the frequency of each KIR2DL4 molecule in groups with different ancestry backgrounds. The frequency of some variants modifies substantially according to the ancestry background. For instance, KIR2DL4*008 seems to be a European allele since this allele was absent in the AFR50 group, and this allele was slight frequent in the AFR30 group. On the contrary, KIR2DL4*001 was considerable frequent in both groups with higher African ancestry. The amino acids exchanges that characterize each of these KIR2DL4 molecules are listed in Table 4. These two ancestry-related alleles, KIR2DL4*001 and KIR2DL4*008 are among the most divergent ones, with at least 6 amino acid exchanges between them.

Table 3. List of KIR2DL4 molecules found in Brazilian population sample separated by background ancestry.

Protein	EAS30 (2n=16)	AFR50 (2n=14)	AFR30 (2n=40)	EUR90 (2n=92)	EUR70 (2n=236)	Pop general (2n=314)
KIR2DL4*006	6.25%	0.00%	0.00%	2.17%	1.27%	1.27%
KIR2DL4*002	0.00%	0.00%	0.00%	0.00%	0.42%	0.32%
KIR2DL4*001	37.50%	57.14%	52.50%	21.74%	24.15%	29.30%
KIR2DL4*012	0.00%	0.00%	5.00%	0.00%	0.00%	0.64%
KIR2DL4*011	12.50%	28.57%	12.50%	14.13%	11.44%	11.46%
KIR2DL4*005	12.50%	14.29%	20.00%	25.00%	21.61%	20.38%
KIR2DL4*008	18.75%	0.00%	7.50%	33.70%	36.44%	31.53%
KIR2DL4*013	6.25%	0.00%	2.50%	2.17%	2.97%	3.50%
KIR2DL4*022	0.00%	0.00%	0.00%	0.00%	0.42%	0.32%
New1	6.25%	0.00%	0.00%	0.00%	0.00%	0.32%
New2	0.00%	0.00%	0.00%	0.00%	0.42%	0.32%
New3	0.00%	0.00%	0.00%	0.00%	0.42%	0.32%
New4	0.00%	0.00%	0.00%	1.09%	0.42%	0.32%

Table 4. List of KIR2DL4 proteins found in a Brazilian population sample, with the position of each aminoacid exchange. Aminoacids different from the ones encoded by reference alleles are marked in shades of grey.

Allele	Exon3			Exon4			Exon6	Exon7		Exon 8		
	Position 23	Position 53	Position 97	Position 132	Position 138	Position 209	Position 271	Position 272	Position 273	Position 274	Position 296	Position 340
KIR2DL4*001	A	Y	C	P	T	A	N	A	A	V	Q	A
KIR2DL4*006	A	Y	C	L	A	P	N	A	A	V	D	E
KIR2DL4*005	A	C	C	P	A	P	N	A	A	V	Q	A
KIR2DL4*012	A	C	C	P	A	P	N	A	A	V	E	A
KIR2DL4*011	A	C	C	P	A	P	M	L	L	V	-	A
KIR2DL4*013	A	Y	C	P	T	A	M	L	L	V	-	A
KIR2DL4*008	A	Y	C	P	A	P	M	L	L	V	-	A
KIR2DL4*002	A	Y	C	P	A	P	N	A	A	V	Q	A
New2	A	C	S	P	A	P	N	A	A	V	Q	A
KIR2DL4*022	A	Y	C	P	T	A	N	A	A	V	Q	V
New4	A	Y	C	P	A	P	N	A	A	A	Q	A
New3	T	Y	C	L	T	A	K	C	C	C*	-	A
New1	T	C	C	L	A	P	N	A	A	V	Q	A

*For this molecule, all the aminoacids between positions 274 to 289 are different when compared to other molecules.

The 3'UTR segment is important for gene regulation and mRNA stability. Overall, the 3'UTR segment presented 10 variable sites, arranged into 14 haplotypes (Table 5). Since this region is not entirely considered by the IPD-IMGT/KIR Database, and since there is no nomenclature pattern defined for it, haplotypes were named as U1 to U14 following sequence similarities. The reference haplotype (presenting only reference alleles) were named U1. The most frequent haplotypes were U2 (21,66%), followed by U7 (21,02%), U1 (19,75%), and U8 (18,79%). Together, these haplotypes present a summed frequency of 81.22% and carry at least 5 frequent variable sites.

Table 5. List of haplotypes at the KIR2DL4 3'untranslated region (exon 8) detected in a Brazilian population sample.

Relative KIR2DL4 position according to the GRhg38																
Chr19 position	**54814116**	**54814169**	**54814361**	**54814366**	**54814618**	**54814619**	**54814646**	**54814806**	**54814823**	**54814824**	**Frequency (2n=314)**	**EAS 30 (2n=16)**	**AFR 50 (2n=14)**	**AFR 30 (2n=40)**	**EUR 90 (2n=92)**	**EUR 70 (2n=236)**
SNPid	**rs370869656**	**rs376917646**	**rs34785252**	**rs374254587**	**rs676443**	**rs667205**	**rs667263**	**rs76100532**	**rs377747517**	**rs200660658**						
U1	G	G	C	C	T	G	G	T	G	G	0.197	0.063	0.000	0.025	0.217	0.246
U2	G	G	C	C	C	G	G	T	G	G	0.216	0.250	0.075	0.175	0.228	0.216
U3	G	G	A	C	T	G	G	T	G	G	0.064	0.000	0.000	0.000	0.000	0.008
U4	G	G	C	C	C	T	G	T	G	G	0.032	0.000	0.025	0.075	0.065	0.051
U5	G	G	A	C	C	T	G	T	G	G	0.032	0.000	0.000	0.025	0.000	0.000
U6	G	G	C	C	C	T	G	C	G	G	0.086	0.000	0.075	0.225	0.043	0.059
U7	G	G	C	C	C	T	C	T	G	G	0.210	0.125	0.075	0.175	0.272	0.229
U8	G	G	A	C	C	T	G	C	G	G	0.187	0.313	0.100	0.225	0.141	0.165
U9	G	G	C	G	C	T	G	C	G	G	0.012	0.125	0.000	0.025	0.000	0.004
U10	A	G	C	C	C	T	C	T	G	G	0.096	0.000	0.000	0.050	0.000	0.004
U11	G	C	C	C	C	T	G	C	G	G	0.032	0.000	0.000	0.000	0.000	0.000
U12	G	G	A	C	C	T	C	T	G	G	0.032	0.000	0.000	0.000	0.011	0.004
U13	G	G	C	C	C	T	C	C	G	A	0.016	0.125	0.000	0.000	0.022	0.013
U14	G	C	C	C	C	T	G	C	C	G	0.045	0.000	0.000	0.000	0.000	0.000

Nucleotide diversity 0.002790 +/- 0.001708

Gene diversity 0.8273 +/- 0.0071

We also evaluated KIRDL4 extended haplotypes to define the association among each of the promoter, coding and 3'UTR sequences (Table 6). We detected 32 extended haplotypes, with 11 presenting frequencies higher than 1%. Some encoded KIR2DL4 molecules might be associated with specific promoter and 3'UTR sequences. For instance, KIR2DL4*008 is associated mainly with P1 e U1 (19,43%); *001 with P8 and U8 (14,65%); *005 with P3 and U7 (18,79%); *011 with P7 and U2 (10,19%).

Table 6. KIR2DL4 extended haplotypes detected in a Brazilian population sample, considering the segment between nucleotide position chr 19:54801519 – 54814906 (hg38).

5'Upstream/5'UTR Haplotypes	***KIR2DL4* coding allele**	**3'UTR/exon8 Haplotypes**	**Frequency (2n=314)**
P10	KIR2DL4*001	U6	0.003%
P11	KIR2DL4*001	U9	0.006%
P5	KIR2DL4*001	U8	0.19%4
P8	KIR2DL4*001	U1	0.013%
P8	KIR2DL4*001	U2	0.003%
P8	KIR2DL4*001	U6	0.01%
P8	KIR2DL4*001	U8	0.003%
P9	KIR2DL4*001	U6	0.003%
P5	KIR2DL4*002	U2	0.003%
P3	KIR2DL4*005	U12	0.003%
P3	KIR2DL4*005	U10	0.003%
P3	KIR2DL4*005	U7	0.006%
P4	KIR2DL4*005	U7	0.188%
P6	KIR2DL4*005	U7	0.013%
P3	KIR2DL4*006	U13	0.013%
P1	KIR2DL4*008	U2	0.003%
P1	KIR2DL4*008	U3	0.003%
P1	KIR2DL4*008	U1	0.003%
P2	KIR2DL4*008	U1	0.003%
P5	KIR2DL4*008	U8	0.003%
P5	KIR2DL4*008	U2	0.003%
P3	KIR2DL4*011	U7	0.105%
P7	KIR2DL4*011	U2	0.003%
P5	KIR2DL4*012	U4	0.102%
P5	KIR2DL4*012	U5	0.003%
P11	KIR2DL4*013	U9	0.003%
P8	KIR2DL4*013	U8	0.003%
P8	KIR2DL4*022	U6	0.003%
P3	New1	U13	0.041%
P3	New2	U10	0.146%
P3	New3	U7	0.032%
P8	New4	U8	0.029%
			Gene diversity 0.9212 +/- 0.0070
			Nucleotide diversity 0.004887 +/- 0.002344

Although the number of different extended haplotypes is quite significant, only one block of LD was observed (Figure 2), indicating a high LD throughout the *KIR2DL4* gene segment. Considering nucleotide diversity, we may consider the coding region the most variable *KIR2DL4* segment (Table 2), followed by the 3'downstream (Table 6). The most conserved segment (in terms of nucleotide diversity) is the 5'upstream, with nucleotide diversity of only 0.001996 +/- 0.001098 (Table 1).

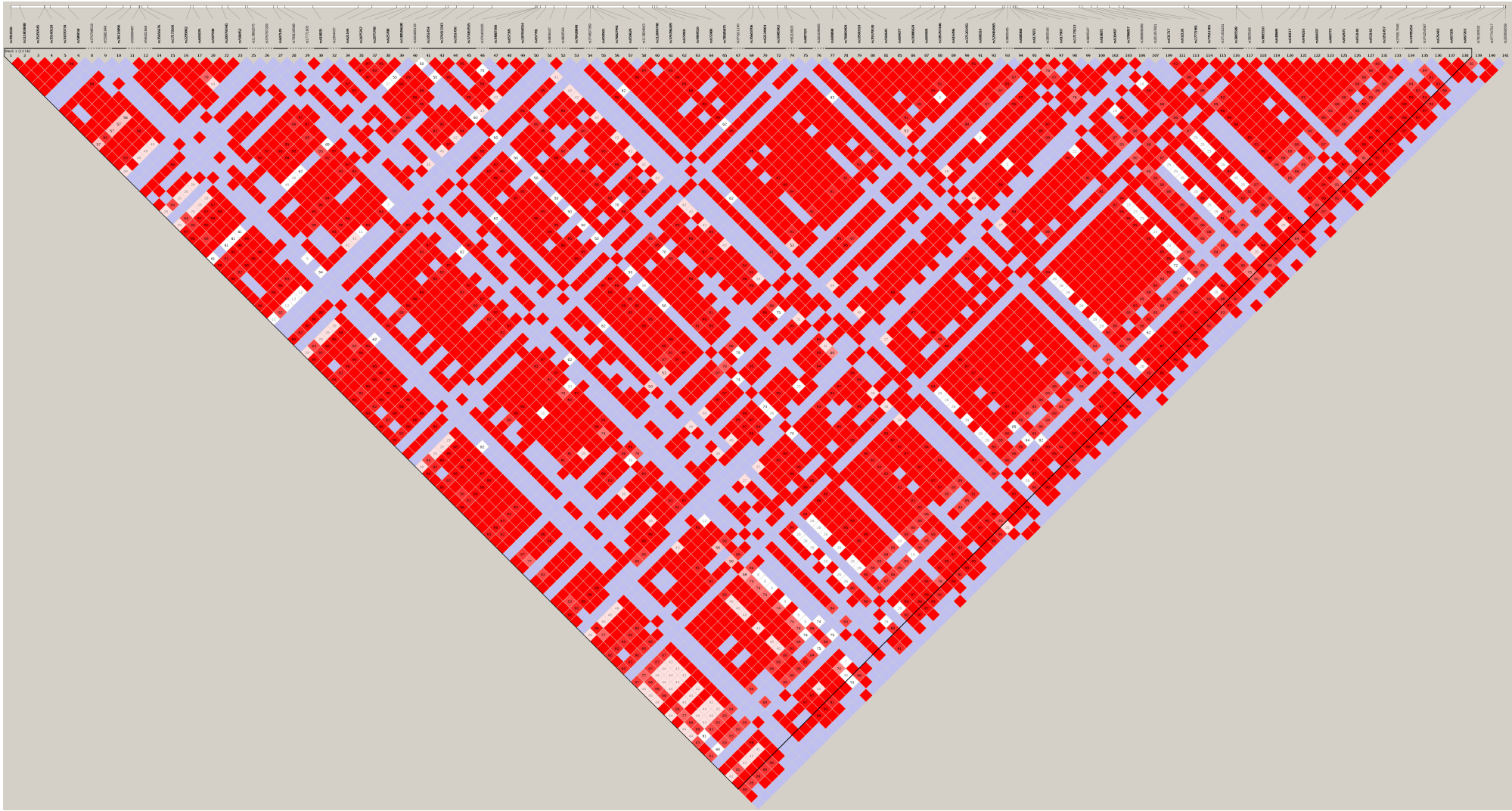

Fig 2. Linkage disequilibrium (LD) between pairs of single nucleotide polymorphisms (SNPs) at the extended *KIR2DL4* gene segment (from position 54801519 to 54814906, chromosome 19). This image was generated by Haploview 4.2 using variable sites with minimum allele frequency (MAF) > 1%. Areas in red indicate strong LD [log de odds ratio (LOD≥2, D'=1)]; areas in light red indicate moderate LD (LOD≥2, D'≤1); areas in blue or almost white indicate weak LD (LOD≤2, D'=1) and white areas indicated no LD (LOD≤2, D'≤1). D' values different from 1 are represented inside the squares as percentages. There is only one block of segregation.

To clusters the different proteins expressed by *KIR2DL4*, we calculated a phylogenetic tree using Bootstrap and *Neighbor-Joining* methodology. As a result, it was possible to observe a clear separation between the alleles that have an insertion of an Adenine at position 54.813.220 (KIR2DL4 * 006, *005, * 012, *001) from those that do not present this insertion (KIR2DL4 *011, *008, *013 and *New3).

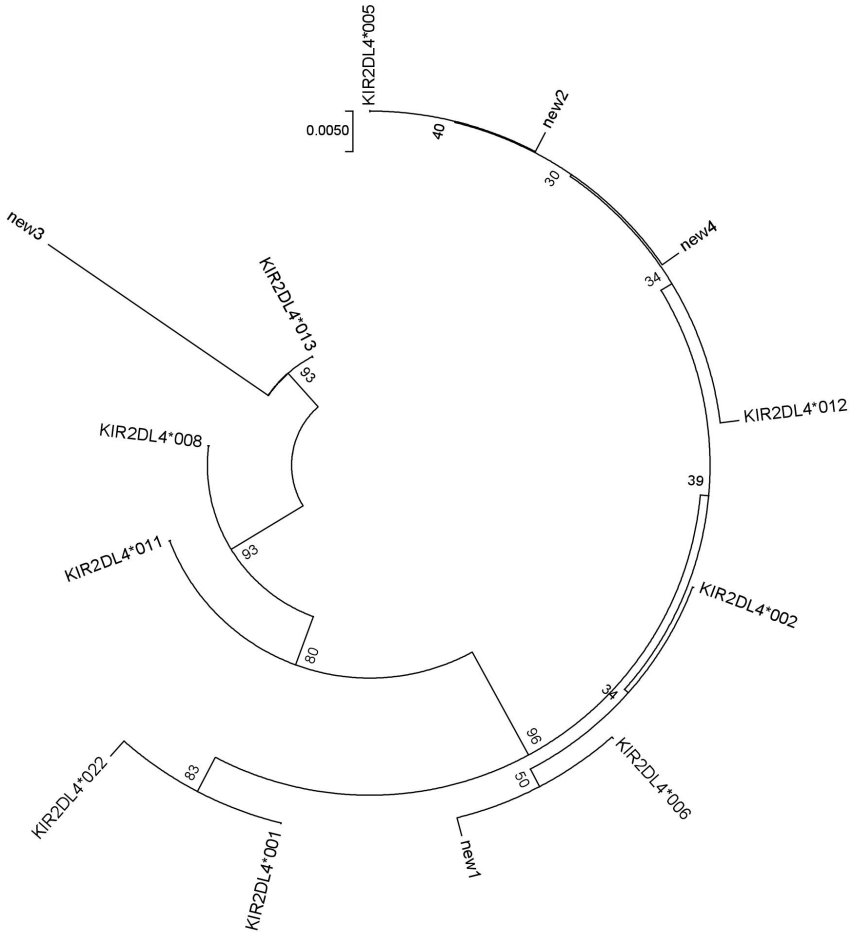

Fig 3. Phylogenetic tree of the protein encoded by the 70 haplotypes detected in a Brazilian population sample.The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.09325383 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 13 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 377 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Felsenstein, 1985; Kumar, Stecher, & Tamura, 2016; Saitou & Nei, 1987).

4.3.2 *1000Genomes* Project

We compared the variants observed in the *1000Genomes* Project and the ones observed in our Brazilian sample (Table S3). Considering 2.504 individuals from the *1000Genomes* dataset

and the segment between 55312974 and 55326339 (chr19 - hg37), we detected 261 variation sites. Only 55 of those variations were also detected in Brazil. Exon 1 (signal peptide) was invariable in Brazil and in the *1000Genomes* dataset. Exon 4 was the most variable exon in both datasets, followed by exon 3. The rs11371265-C (adenine deletion) variant is absent from the *1000Genomes* dataset. We understand that this might be a genotyping error issue, or all the 2.504 individuals would encode only truncated KIR2DL4 molecules.

4.4 Discussion

Polymorphic *Killer Immunoglobulin-like Receptors* (KIR) genes are involved in modulating immune responses through recognition and interactions with MHC class I molecules. Also, these genes are known to exhibit a quite substantial genetic diversity. *KIR2DL4* is characterized by low polymorphism and highly conserved sequence compared to the other *KIR* genes. Also, *KIR2DL4* is considered an atypical KIR family member in some aspects such as genetic structure, ligand specificity, expression levels, cellular localization, and signaling. This gene is located in the middle of the KIR complex, and it is referred to as a framework since it is virtually present in all individual. Although KIR genes display a variegated expression among individual NK cells, KIR2DL4 is expressed on all NK cells and also a subset of T cells (CD8+) (Rajagopalan & Long, 1999).

Polymorphisms present in the coding region of *KIR2DL4,* mainly non-synonyms, might disrupt the KIR2DL4/HLA-G binding, or even influencing KIR expression (Goodridge et al., 2007). Also, variable sites present in the regulatory region may affect the expression of KIR2DL4 on the surface of NK and T cells. Even though *KIR2DL4* is considered to have a low polymorphism among the highly variable KIR family members, the complete gene variability (including the 5'upstream promoter and 3'downstream) has never been assessed.

As far as we know this is the first study evaluating an admixed population from the State of São Paulo, which is considered a large repository of genetic variation. Moreover, this is the first study evaluating the entire *KIR2DL4* segment using massively parallel sequencing (NGS) procedures.

In this study, we investigated the level of KIR2DL4 nucleotide variation, from the 5'upstream promoter to 3'downstream, in 157 individual from São Paulo/Brazil. We found 152 variable sites in approximately 13-kb, with at least 14 variants been described for the first time. On the IPD-IMGT/KIR database *version 2.8.0*, a total of 70 alleles have been reported, those are

responsible for coding 40 different protein. Here we detected 32 alleles responsible for coding 13 proteins. These differences are probably related to (a) the small sample size here evaluated, and (b) genotyping errors observed in the *1000Genomes* dataset.

4.4.1. *KIR2DL4* variability in the 5′upstream regulatory region (5'URR)

The 5'upstream regulatory region (5'URR) was firstly characterized here. This segment encompasses approximately 2-kb upstream the first translated ATG and includes the distal and proximal promoters, and the 5'UTR. As far as we know, this is the first characterization of a continuous long segment upstream the *KIR2DL4* locus using a large sample size of an admixed population.

Previous studies indicated that KIR2DL4 presents the most divergent promoter sequence when compared to other *KIR* genes (Li, Wright, McCullen, & Anderson, 2016). The 5'URR was the most conserved segment in terms of nucleotide diversity (0.001996, Table 1), with 16 variable sites arranged into 11 haplotypes (Table 1). The core promoter (the 157 bases upstream the first translated ATG), important to the assembly of the transcription machinery assemblies, was invariant in all samples. In addition, the whole promoter segment is apparently highly conserved. This conservation is possibly related to a tight regulation mechanism, and, therefore, this segment might under strong purifying selection. Further investigation is needed to evaluate whether the variable sites in the distal promoter, and even the few variable sites in the proximal promoter would influence the binding of transcriptional factors.

Previous studies have suggested that the first *KIR2DL4* intron could overcome the function of the promoter region when silencing elements bond to the promoter (Trompeter et al., 2005). The first introns of *KIR* are generally located close to the promoter region, separated only by a small exon of approximately 34-bp. Most of the KIR genes present a mini-satellites in the first intron, however *KIR2DL4* consists of a 198-bp non-repetitive sequence, which is a unique feature amount of *KIR* family members (Trowsdale et al., 2001). Here we detected only 2 variants in this intron, showing that this segment is quite conserved and is probably involved in the KIR2DL4 expression regulation.

4.4.2 The *KIR2DL4* variability in the coding segment

The coding segment was the most variable one and presented with the highest nucleotide diversity compared to the 5'URR and 3'UTR. We detected 124 variable sites in this segment,

and they have been arranged in 70 haplotypes. However, since most of them are intronic variants or synonymous mutations, they encode only 13 different protein molecules (Table 3). Four of these molecules, KIR2DL4*008, KIR2DL4*001, KIR2DL4*005, and KIR2DL4 *011, present a summed frequency of 92,67% and represent very divergent ones.

Most of the variable sites described here were found in introns (111 variable sites). Although intron sites are unlike to alter gene function, modification in those sequences may cause alternative splicing and different expression profiles (Bruijnesteijn et al., 2018). Intron 5 is the most variable one, with 50 variable sites registered here, while, as discussed earlier, intron 1 is the most conserved. Exon 4 and exon 8 are the most variable ones, and most of the variants configure non-synonymous mutations.

To estimate the possible effects of the non-synonymous substitutions detected here, we need first to address the *KIR2DL4* genetic structure. First, the signal peptide sequence is encoded by the first two exons, where no variation was detected in this population sample. This may indicate that this segment is important to the correct exportation of the KIR2DL4 molecules to the cell surface. Second, two Ig-like domain (D0, D2, starting from the N-terminus), are encoded by exons 3 and 4, respectively. We detected six non-synonyms variable sites in this segment (Table S1). These modifications in an extracellular domain may play a role in the interaction between the receptor and its ligands. Charge complementarity and protein secondary structure have a very important role at the KIR/HLA recognition (Boyington & Sun, 2002). The modeling of KIR2DL4 docking to HLA-G may provide an insight into the role of allelic diversity in the binding of these two molecules. Third, there is a linker region encoded by the exon 5, where no variation sites were found. Forth, the transmembrane regions encoded by exon 6 presented two non-synonymous mutations. Finally, the cytoplasmic domain encoded by exons 7 and 8 was conserved, with only three variants in exon 8.

Considering only the exonic region, *KIR2DL4* is described as being quite conserved, with no more than 14 variable sites already described in other population studies (Buhler et al., 2009; Gedil, Steiner, & Hurley, 2005; Goodridge et al., 2003; Jones, Hiby, Moffett, Trowsdale, & Young, 2006; Yawata et al., 2006; Zhu, Jiang, Lv, He, & Yan, 2006). Most of the polymorphic sites in the coding region described in our sample are present in exons 3, 4, 6 and 8. There are few studies that have described the diversity of the *KIR2DL4* gene in different populations, however, in general, used sample sizes and less sensitive techniques. As

a result, the study of this region is still to be elucidated for most of the populations, and also in terms of genetic diversity and ancestry background.

It became clear that KIR2DL4 variability is influenced by ancestry (Table 3). Some KIR2DL4 alleles are more frequent in (or exclusively associated with) samples of specific ancestries. For instance, KIR2DL4*001 is highly frequent among samples with higher African ancestry. The KIR2DL4*012 allele (5,0% in the AFR30 group) was found in samples from Congo (6,12%) (Buhler et al., 2009). It is not clear why some alleles are associated with specific ancestry background, but we may indicate some possibilities. First, this may be a case of genetic drift. Second, it is possible that natural selection my playing a role in these frequencies, since the most important ligand, HLA-G, is also variable in individuals with different ancestry backgrounds. However, the sample size study here does not allow additional analysis in this sense. The allele KIR2DL4*006 (1,27%) was founded only in the individuals with Asian background (EAS30). Although this frequency is not quite significant, this allele seems to be more frequent in Asian population compared to European population (Buhler et al., 2009; Yawata et al., 2006; Zhu et al., 2006). This analysis indicates that polymorphism frequencies are influenced by the sample ancestry background, indicating that ancestry must be considered for any association study regarding *KIR2DL4*.

One important variant in the KIR2DL4 coding region is an insertion/deletion of single Adenine in a series of consecutive Adenines at the end of exon 6 (Goodridge et al., 2007; Goodridge et al., 2003). This exon ends with a mononucleotide repeats of 10 consecutive adenines (10 A). Some alleles have a deletion of one adenine (9 A) with no subsequent nucleotide replacement resulting in a frameshift, which leads to a premature stop codon. This stop codon results in a truncated protein that lacks the ITIM motif in the cytoplasmic tail. This deletion (9A) has a frequency of 46,81% in this Brazilian population sample. This frequency is compatible with the ones observed in other populations (Goodridge et al., 2003; Goris et al., 2009). However, if we take into consideration the African (AFR30) and Asian (EAS30) subgroups, the frequency of the 10A allele reaches 77,5% and 62,5%, respectively. Samples with higher European ancestry are associated frequencies around 50%. Thus, samples with greater African ancestry have a lower frequency of the truncated protein (about 22,5% in our sample, 28% in Congo) (BUHLER et al., 2009). Likewise, samples with greater Asian ancestry have a lower frequency of the truncated protein (about 37,5% in our sample; 18,9% in the Japanese population, and 19,09% in Chinese population) (Yawata et al., 2006; Zhu et al., 2006). The lower frequency of the truncated KIR2DL4 among Africans may be related to

selective pressures, since Africans also present a higher frequency of a truncated HLA-G allele G*01:05N (Castelli et al., 2014). Thus, the presence of more functional KIR2DL4 molecules might compensate the lower rate of functional HLA-G molecules in these populations.

It was not possible to evaluate this variant in the *1000Genome* databank, since this variant is not included in the dataset. In previous studies carried out by our group with HLA family genes, specifically the HLA-G and HLA-C analysis, we observed the absence of several polymorphisms in the genotypes published by the *1000Genomes* Project (Castelli et al., 2017) (WEISS et al. 2017). It is believed that this error is caused by limitations during genotype detection and sequence mapping. Here, it was necessary to create a specific tool to mapping KIR gene data. In this sense, considering that the *1000Genomes* project did not use specific tools to genotype gene families, such as KIR and HLA genes, publicly available data may be biased, with the presence of wrong genotypes and/or absence of well-known genotypes, emphasizing the need for proper analysis of samples for the actual knowledge of the variability of this gene in worldwide populations.

4.4.4 The *KIR2DL4* variability in the 3'downstream segment

A greater diversity is expected is in the regulatory and non-coding regions, that since these regions have little similarity when compared to other genes members of this family (Trowsdale et al., 2001). Because of that, we expect that these divergent regulatory sequences be related to differential regulation mechanisms and expression profiles. It is well know that polymorphisms present in the 3'downstream segment may alter mRNA stability, translation, localization, degradation, and influence gene expression, mainly because different mRNA 3'downstream sequences may bind differently to miRNAs (Bartel, 2009).

Here we report the 3'UTR variability and haplotypes, and the relationship among these haplotypes with promoter and coding sequences. This segment encompasses approximately 795-bp downstream the stop codon. We found 10 variable sites between the positions 54814111 and 54814906. These variable sites were arranged in 14 haplotypes. The most frequent are U2 (21%), U7 (21%) and U1 (19%). Together those haplotypes represent 61% of all sequences, and they differ from each other by three variable sites. If we take in consideration subgroups, some of those haplotypes can be found more frequently in specifics populations. For instance, U13 is more frequent in Asian (EAS30), and it is not found in any

of the African subgroup (AFR30 and AFR50). U1 is predominating found in European subgroups. This results shown that 3'UTR variability is also influenced by ancestry. Overall, we can consider that the 3'UTR of this gene is quite conserved. This can be a consequence of mechanism that involves the binding of specifics miRNAs.

4.4.5 *KIR2DL4* extended haplotype

Likewise *HLA-G* (Castelli et al., 2017), *KIR2DL4* present a strong LD along the entire gene segment (Figure 2). This reflects in a low number of extended haplotypes (32 in the present sample). The most frequent haplotypes were: P8/KIR2DL4*001/U8 (14%), P7/KIR2DL4*011/U2 (10%), P5/KIR2DL4*008/U2 (10%), P3/KIR2DL4*005/U7 (18%), P1/KIR2DL4*008/U1 (19%). Together those extended haplotypes represent 71% of all sequences. In general, each coding allele is associated with a specific promoter and 3'UTR haplotype (Table 6). Some of the promoter and 3'UTR haplotypes are quite divergent. Thus, it is possible that each of these coding *KIR2DL4* haplotypes are associated with different expression profiles.

4.5 Conclusion

As far as we know, this is the first study to fully characterize the *KIR2DL4* gene variability considering both regulatory segments and the entire coding sequence in a highly admixed Brazilian population sample. Also, this study introduces a molecular and computational methodology to evaluate the entire *KIR2DL4* gene and its regulatory segments using massively parallel sequencing and freely available software for genotyping and haplotyping. KIR2DL4 presents a high LD throughout the gene, with each coding allele usually associated with specific promoter and 3'UTR sequences. The most variable segments were exons 3 and 4 (the external domains) and the insertion/deletion polymorphism in exon 6, which is related to a truncated KIR2DL4 protein. Although the *1000Genome* Project provides a foundation for the study of human genetics and variability, its data regarding repetitive and high polymorphic genes such as *KIR2DL4* should be used with caution. Finally, *KIR2DL4* polymorphisms frequencies are influenced by sample ancestry background, indicating that ancestry must be considered for any association study regarding this important immune regulatory gene.

4.6 References

Ahern, D. J., & Brennan, F. M. (2011). The role of Natural Killer cells in the pathogenesis of rheumatoid arthritis: Major contributors or essential homeostatic modulators? *Immunology Letters, 136*(2), 115-121. doi:10.1016/j.imlet.2010.11.001

Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., . . . Genomes Project, C. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-+. doi:10.1038/nature15393

Avril, T., Floyd, H., Lopez, F., Vivier, E., & Crocker, P. R. (2004). The membrane-proximal immunoreceptor tyrosine-based inhibitory motif is critical for the inhibitory signaling mediated by siglecs-7 and-9, CD33-related Siglecs expressed on human monocytes and NK cells. *Journal of Immunology, 173*(11), 6841-6849. doi:10.4049/jimmunol.173.11.6841

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics, 21*(2), 263-265. doi:10.1093/bioinformatics/bth457

Biron, C. A., Nguyen, K. B., Pien, G. C., Cousens, L. P., & Salazar-Mather, T. P. (1999). Natural killer cells in antiviral defense: Function and regulation by innate cytokines. *Annual Review of Immunology, 17*, 189-220. doi:10.1146/annurev.immunol.17.1.189

Boyington, J. C., & Sun, P. D. (2002). A structural perspective on MHC class I recognition by killer cell immunoglobulin-like receptors. *Mol Immunol, 38*(14), 1007-1021.

Bruijnesteijn, J., van der Wiel, M. K. H., de Groot, N., Otting, N., de Vos-Rouweler, A. J. M., Lardy, N. M., . . . Bontrop, R. E. (2018). Extensive Alternative Splicing of KIR Transcripts. *Front Immunol, 9*. doi:10.3389/fimmu.2018.02846

Buhler, S., Di Cristofaro, J., Frassati, C., Basire, A., Galicher, V., Chiaroni, J., & Picard, C. (2009). High levels of molecular polymorphism at the KIR2DL4 locus in French and Congolese populations: Impact for anthropology and clinical studies. *Human Immunology, 70*(11), 953-959. doi:10.1016/j.humimm.2009.08.002

Castelli, E. C., Gerasimou, P., Paz, M. A., Ramalho, J., Porto, I. O. P., Lima, T. H. A., . . . Costeas, P. (2017). HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographicaly distinct population samples of Brazil and Cyprus. *Molecular Immunology, 83*, 115-126. doi:10.1016/j.molimm.2017.01.020

Castelli, E. C., Mendes, C. T., Sabbagh, A., Porto, I. O. P., Garcia, A., Ramalho, J., . . . Donadi, E. A. (2015). HLA-E coding and 3 ' untranslated region variability determined by next-generation sequencing in two West-African population samples. *Human Immunology, 76*(12), 945-953. doi:10.1016/j.humimm.2015.06.016

Castelli, E. C., Mendes, C. T., Veiga-Castelli, L. C., Roger, M., Moreau, P., & Donadi, E. A. (2011). A Comprehensive Study of Polymorphic Sites along the HLA-G Gene: Implication for Gene Regulation and Evolution. *Molecular Biology and Evolution, 28*(11), 3069-3086. doi:10.1093/molbev/msr138

Castelli, E. C., Paz, M. A., Souza, A. S., Ramalho, J., & Mendes-Junior, C. T. (2018). Hla-mapper: An application to optimize the mapping of HLA sequences produced by massively parallel sequencing procedures. *Hum Immunol, 79*(9), 678-684. doi:10.1016/j.humimm.2018.06.010

Castelli, E. C., Ramalho, J., Porto, I. O., Lima, T. H., Felicio, L. P., Sabbagh, A., . . . Mendes-Junior, C. T. (2014). Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity. *Front Immunol, 5*, 476. doi:10.3389/fimmu.2014.00476

Doria, A., Zen, M., Bettio, S., Gatto, M., Bassi, N., Nalotto, L., . . . Punzi, L. (2012). Autoinflammation and autoimmunity: Bridging the divide. *Autoimmunity Reviews, 12*(1), 22-30. doi:10.1016/j.autrev.2012.07.018

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x

Fadda, L., Borhis, G., Ahmed, P., Cheent, K., Pageon, S. V., Cazaly, A., . . . Khakoo, S. I. (2010). Peptide antagonism as a mechanism for NK cell activation. *Proceedings of the National Academy of Sciences of the United States of America, 107*(22), 10160-10165. doi:10.1073/pnas.0913745107

Faure, M., & Long, E. O. (2002). KIR2DL4 (CD158d), an NK cell-activating receptor with inhibitory potential. *Journal of Immunology, 168*(12), 6208-6214. doi:10.4049/jimmunol.168.12.6208

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution, 39*(4), 783-791. doi:10.1111/j.1558-5646.1985.tb00420.x

Ferreira, L. B., Mendes, C. T., Vieira, C. E., Luizon, M. R., & Simoes, A. L. (2006). Genomic ancestry of a sample population from the state of Sao Paulo, Brazil. *American Journal of Human Biology, 18*(5), 702-705. doi:10.1002/ajhb.20474

Gedil, M. A., Steiner, N. K., & Hurley, C. K. (2005). Genomic characterization of KIR2DL4 in families and unrelated individuals reveals extensive diversity in exon and intron sequences including a common frameshift variation occurring in several alleles. *Tissue Antigens, 65*(5), 402-418. doi:10.1111/j.1399-0039.2005.00380.x

Gomes, K. F. B., Santos, A. S., Semzezem, C., Correia, M. R., Brito, L. A., Ruiz, M. O., . . . da Silva, M. E. R. (2017). The influence of population stratification on genetic markers associated with type 1 diabetes. *Scientific Reports, 7*. doi:10.1038/srep43513

Goodridge, J. P., Lathbury, L. J., Steiner, N. K., Shulse, C. N., Pullikotil, P., Seidah, N. G., . . . Witt, C. S. (2007). Three common alleles of KIR2DL4 (CD158d) encode constitutively expressed, inducible and secreted receptors in NK cells. *European Journal of Immunology, 37*(1), 199-211. doi:10.1002/eji.200636316

Goodridge, J. P., Witt, C. S., Christiansen, F. T., & Warren, H. S. (2003). KIR2DL4 (CD158d) genotype influences expression and function in NK cells. *Journal of Immunology, 171*(4), 1768-1774. doi:10.4049/jimmunol.171.4.1768

Goris, A., Dobosi, R., Boonen, S., Nagels, G., & Dubois, B. (2009). KIR2DL4 (CD158d) polymorphisms and susceptibility to multiple sclerosis. *Journal of Neuroimmunology, 210*(1-2), 113-115. doi:10.1016/j.jneuroim.2009.03.001

Halder, I., & Shriver, M. D. (2003). Measuring and using admixture to study the genetics of complex diseases. *Hum Genomics, 1*(1), 52-62.

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., . . . Hubbard, T. J. (2012). GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Research, 22*(9), 1760-1774. doi:10.1101/gr.135350.111

Heidenreich, S., Eulenburg, C. Z., Hildebrandt, Y., Stubig, T., Sierich, H., Badbaran, A., . . . Kroger, N. (2012). Impact of the NK Cell Receptor LIR-1 (ILT-2/CD85j/LILRB1) on Cytotoxicity against Multiple Myeloma. *Clinical & Developmental Immunology*. doi:10.1155/2012/652130

Hsu, K. C., Chida, S., Dupont, B., & Geraghty, D. E. (2002). The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. *Immunological Reviews, 190*(1), 40-52. doi:10.1034/j.1600-065X.2002.19004.x

Humphrey, M. B., Lanier, L. L., & Nakamura, M. C. (2005). Role of ITAM-containing adapter proteins and their receptors in the immune system and bone. *Immunological Reviews, 208*, 50-65. doi:10.1111/j.0105-2896.2005.00325.x

Jager, A., & Kuchroo, V. K. (2010). Effector and Regulatory T-cell Subsets in Autoimmunity and Tissue Inflammation. *Scandinavian Journal of Immunology, 72*(3), 173-184. doi:10.1111/j.1365-3083.2010.02432.x

Jobim, M., Salim, P. H., Portela, P., Wilson, T. J., Fraportti, J., Baronio, D., . . . Schwartsmann, G. (2010). Killer cell immunoglobulin-like receptor gene diversity in a Caucasian population of Southern Brazil. *International Journal of Immunogenetics, 37*(2), 83-89. doi:10.1111/j.1744-313X.2009.00894.x

Jones, D. C., Hiby, S. E., Moffett, A., Trowsdale, J., & Young, N. T. (2006). Nature of allelic sequence polymorphism at the KIR3DL3 locus. *Immunogenetics, 58*(8), 614-627. doi:10.1007/s00251-006-0130-5

Karre, K., Ljunggren, H. G., Piontek, G., & Kiessling, R. (1986). Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. *Nature, 319*(6055), 675-678. doi:10.1038/319675a0

Kikuchi-Maki, A., Yusa, S., Catina, T. L., & Campbell, K. S. (2003). KIR2DL4 is an IL-2-regulated NK cell receptor that exhibits limited expression in humans but triggers strong IFN-gamma production. *Journal of Immunology, 171*(7), 3415-3425. doi:10.4049/jimmunol.171.7.3415

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol, 33*(7), 1870-1874. doi:10.1093/molbev/msw054

Langers, I., Renoux, V. M., Thiry, M., Delvenne, P., & Jacobs, N. (2012). Natural killer cells: role in local tumor growth and metastasis. *Biologics, 6*, 73-82. doi:10.2147/btt.s23976

Lee, S. H., Miyagi, T., & Biron, C. A. (2007). Keeping NK cells in highly regulated antiviral warfare. *Trends in Immunology, 28*(6), 252-259. doi:10.1016/j.it.2007.04.001

Leung, W. (2014). Infusions of Allogeneic Natural Killer Cells as Cancer Therapy. *Clinical Cancer Research, 20*(13), 3390-3400. doi:10.1158/1078-0432.ccr-13-1766

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, H., Wright, P. W., McCullen, M., & Anderson, S. K. (2016). Characterization of KIR intermediate promoters reveals four promoter types associated with distinct expression patterns of KIR subtypes. *Genes and Immunity, 17*(1), 66-74. doi:10.1038/gene.2015.56

Lima, T. H. A., Buttura, R. V., Donadi, E. A., Veiga-Castelli, L. C., Mendes, C. T., & Castelli, E. C. (2016). HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample. *Human Immunology, 77*(10), 841-853. doi:10.1016/j.humimm.2016.07.231

Marsh, S. G. E., Parham, P., Dupont, B., Geraghty, D. E., Trowsdale, J., Middleton, D., . . . Witt, C. (2003). Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Journal of Immunological Methods, 281*(1-2), 1-8. doi:10.1016/j.jim.2003.08.001

Martin, A. M., Freitas, E. M., Witt, C. S., & Christiansen, F. T. (2000). The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics, 51*(4-5), 268-280. doi:10.1007/s002510050620

Martin, A. M., Kulski, J. K., Gaudieri, S., Witt, C. S., Freitas, E. M., Trowsdale, J., & Christiansen, F. T. (2004). Comparative genomic analysis, diversity and evolution

of two KIR haplotypes A and B. *Gene, 335*, 121-131. doi:10.1016/j.gene.2004.03.018

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research, 20*(9), 1297-1303. doi:10.1101/gr.107524.110

Miah, S. M., Hughes, T. L., & Campbell, K. S. (2008). KIR2DL4 differentially signals downstream functions in human NK cells through distinct structural modules. *J Immunol, 180*(5), 2922-2932.

Moretta, A., Vitale, M., Bottino, C., Orengo, A. M., Morelli, L., Augugliaro, R., . . . Moretta, L. (1993). P58 MOLECULES AS PUTATIVE RECEPTORS FOR MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) CLASS-I MOLECULES IN HUMAN NATURAL-KILLER (NK) CELLS - ANTI-P58 ANTIBODIES RECONSTITUTE LYSIS OF MHC CLASS-I-PROTECTED CELLS IN NK CLONES DISPLAYING DIFFERENT SPECIFICITIES. *Journal of Experimental Medicine, 178*(2), 597-604. doi:10.1084/jem.178.2.597

Norman, P. J., Hollenbach, J. A., Nemat-Gorgani, N., Guethlein, L. A., Hilton, H. G., Pando, M. J., . . . Parham, P. (2013). Co-evolution of Human Leukocyte Antigen (HLA) Class I Ligands with Killer-Cell Immunoglobulin-Like Receptors (KIR) in a Genetically Diverse Population of Sub-Saharan Africans. *Plos Genetics, 9*(10). doi:10.1371/journal.pgen.1003938

Orr, M. T., & Lanier, L. L. (2010). Natural Killer Cell Education and Tolerance. *Cell, 142*(6), 847-856. doi:10.1016/j.cell.2010.08.031

Parham, P. (2005a). Influence of KIR diversity on human immunity. In S. Gupta, W. E. Paul, & R. Steinman (Eds.), *Mechanisms of Lymphocyte Activation and Immune Regulation X: Innate Immunity* (Vol. 560, pp. 47-50).

Parham, P. (2005b). MHC class I molecules and KIRs in human history, health and survival. *Nature Reviews Immunology, 5*(3), 201-214. doi:10.1038/nri1570

Parham, P., & Moffett, A. (2013). Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. *Nature Reviews Immunology, 13*(2), 133-144. doi:10.1038/nri3370

Parham, P., Norman, P. J., Abi-Rached, L., Hilton, H. G., & Guethlein, L. A. (2012). Review: Immunogenetics of human placentation. *Placenta, 33*, S71-S80. doi:10.1016/j.placenta.2011.11.020

Perricone, R., Perricone, C., De Carolis, C., & Shoenfeld, Y. (2008). NK cells in autoimmunity: A two-edg'd weapon of the immune system. *Autoimmunity Reviews, 7*(5), 384-390. doi:10.1016/j.autrev.2008.03.002

Phillips, C. (2015). Forensic genetic analysis of bio-geographical ancestry. *Forensic Science International-Genetics, 18*, 49-65. doi:10.1016/j.fsigen.2015.05.012

Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T. C., . . . Reich, D. (2018). Reconstructing the Deep Population History of Central and South America. *Cell, 175*(5), 1185-1197 e1122. doi:10.1016/j.cell.2018.10.027

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Purdy, A. K., & Campbell, K. S. (2009). Natural killer cells and cancer Regulation by the killer cell Ig-like receptors (KIR). *Cancer Biology & Therapy, 8*(23), 2211-2220.

Rajagopalan, S. (2010). Endosomal Signaling and a Novel Pathway Defined by the Natural Killer Receptor KIR2DL4 (CD158d). *Traffic, 11*(11), 1381-1390. doi:10.1111/j.1600-0854.2010.01112.x

Rajagopalan, S., & Long, E. O. (1999). A human histocompatibility leukocyte antigen (HLA)-G-specific receptor expressed on all natural killer cells. *J Exp Med, 189*(7), 1093-1100.

Rajagopalan, S., & Long, E. O. (2012). KIR2DL4 (CD158d): An activation receptor for HLA-G. *Front Immunol, 3*, 258. doi:10.3389/fimmu.2012.00258

Rajalingam, R., Krausa, P., Shilling, H. G., Stein, J. B., Balamurugan, A., McGinnis, M. D., . . . Parham, P. (2002). Distinctive KIR and HLA diversity in a panel of north Indian Hindus. *Immunogenetics, 53*(12), 1009-1019. doi:10.1007/s00251-001-0425-5

Raulet, D. H., Vance, R. E., & McMahon, C. W. (2001). Regulation of the natural killer cell receptor repertoire. *Annual Review of Immunology, 19*, 291-330. doi:10.1146/annurev.immunol.19.1.291

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol, 4*(4), 406-425. doi:10.1093/oxfordjournals.molbev.a040454

Saunders, P. M., Vivian, J. P., Baschuk, N., Beddoe, T., Widjaja, J., O'Connor, G. M., . . . Brooks, A. G. (2015). The interaction of KIR3DL1*001 with HLA class I molecules is dependent upon molecular microarchitecture within the Bw4 epitope. *J Immunol, 194*(2), 781-789. doi:10.4049/jimmunol.1402542

Schleinitz, N., Vely, F., Harle, J. R., & Vivier, E. (2010). Natural killer cells in human autoimmune diseases. *Immunology, 131*(4), 451-458. doi:10.1111/j.1365-2567.2010.03360.x

Shilling, H. G., Lienert-Weidenbach, K., Valiante, N. M., Uhrberg, M., & Parham, P. (1998). Evidence for recombination as a mechanism for KIR diversification. *Immunogenetics, 48*(6), 413-416. doi:10.1007/s002510050453

Sonon, P., Sadissou, I., Tokplonou, L., M'Po K, K. G., Glitho, S. S. C., Agniwo, P., . . . Donadi, E. A. (2018). HLA-G, -E and -F regulatory and coding region variability and haplotypes in the Beninese Toffin population sample. *Mol Immunol, 104*, 108-127. doi:10.1016/j.molimm.2018.08.016

Staub, E., Rosenthal, A., & Hinzmann, B. (2004). Systematic identification of immunoreceptor tyrosine-based inhibitory motifs in the human proteome. *Cellular Signalling, 16*(4), 435-456. doi:10.1016/j.cellsig.2003.08.013

Steffens, U., Vyas, Y., Dupont, B., & Selvakumar, A. (1998). Nucleotide and amino acid sequence alignment for human killer cell inhibitory receptors (KIR), 1998. *Tissue Antigens, 51*(4), 398-413.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics, 68*(4), 978-989. doi:10.1086/319501

Stewart, C. A., van Bergen, J., & Trowsdale, J. (2003). Different and divergent regulation of the KIR2DL4 and KIR3DL1 promoters. *Journal of Immunology, 170*(12), 6073-6081. doi:10.4049/jimmunol.170.12.6073

Suto, Y., Maenaka, K., Yabe, T., Hirai, M., Tokunaga, K., Tadokoro, K., & Juji, T. (1996). Chromosomal localization of the human natural killer cell class I receptor family genes to 19q13.4 by fluorescence in situ hybridization. *Genomics, 35*(1), 270-272. doi:10.1006/geno.1996.0355

Thielens, A., Vivier, E., & Romagne, F. (2012). NK cell MHC class I specific receptors (KIR): from biology to clinical intervention. *Current Opinion in Immunology, 24*(2), 239-245. doi:10.1016/j.coi.2012.01.001

Trompeter, H. I., Gomez-Lozano, N., Santourlidis, S., Eisermann, B., Wernet, P., Vilches, C., & Uhrberg, M. (2005). Three structurally and functionally divergent kinds of promoters regulate expression of clonally distributed killer cell Ig-like receptors

(KIR), of KIR2DL4, and of KIR3DL3. *Journal of Immunology, 174*(7), 4135-4143. doi:10.4049/jimmunol.174.7.4135

Trowsdale, J., Barten, R., Haude, A., Stewart, C. A., Beck, S., & Wilson, M. J. (2001). The genomic context of natural killer receptor extended gene families. *Immunological Reviews, 181*, 20-38. doi:10.1034/j.1600-065X.2001.1810102.x

Uhrberg, M. (2005). The KIR gene family: life in the fast lane of evolution. *European Journal of Immunology, 35*(1), 10-15. doi:10.1002/eji.200425743

Uhrberg, M., Valiante, N. M., Shum, B. P., Shilling, H. G., Lienert-Weidenbach, K., Corliss, B., . . . Parham, P. (1997). Human diversity in killer cell inhibitory receptor genes. *Immunity, 7*(6), 753-763. doi:10.1016/s1074-7613(00)80394-5

Valiante, N. M., Uhrberg, M., Shilling, H. G., Lienert-Weidenbach, K., Arnett, K. L., D'Andrea, A., . . . Parham, P. (1997). Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. *Immunity, 7*(6), 739-751. doi:10.1016/s1074-7613(00)80393-3

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., . . . DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics, 43*, 11.10.11-33. doi:10.1002/0471250953.bi1110s43

Vilches, C., & Parham, P. (2002). KIR: Diverse, rapidly evolving receptors of innate and adaptive immunity. *Annual Review of Immunology, 20*, 217-251. doi:10.1146/annurev.immunol.20.092501.134942

Vivier, E., Raulet, D. H., Moretta, A., Caligiuri, M. A., Zitvogel, L., Lanier, L. L., . . . Ugolini, S. (2011). Innate or Adaptive Immunity? The Example of Natural Killer Cells. *Science, 331*(6013). doi:10.1126/science.1198687

Wilson, M. J., Torkar, M., & Trowsdale, J. (1997). Genomic organization of a human killer cell inhibitory receptor gene. *Tissue Antigens, 49*(6), 574-579. doi:10.1111/j.1399-0039.1997.tb02804.x

Witt, C. S., Martin, A., & Christiansen, F. T. (2000). Detection of KIR2DL4 alleles by sequencing and SSCP reveals a common allele with a shortened cytoplasmic tail. *Tissue Antigens, 56*(3), 248-257. doi:10.1034/j.1399-0039.2000.560307.x

Witt, C. S., Whiteway, J. M., Warren, H. S., Barden, A., Rogers, M., Martin, A., . . . Christiansen, F. T. (2002). Alleles of the KIR2DL4 receptor and their lack of association with pre-eclampsia. *European Journal of Immunology, 32*(1), 18-29. doi:10.1002/1521-4141(200201)32:1<18::aid-immu18>3.3.co;2-z

Yan, W. H., Lin, A. F., Chen, B. G., Zhou, M. Y., Dai, M. Z., Chen, X. J., . . . Li, B. L. (2007). Possible roles of KIR2DL4 expression on uNK cells in human pregnancy. *American Journal of Reproductive Immunology, 57*(4), 233-242. doi:10.1111/j.1600-0897.2007.00469.x

Yawata, M., Yawata, N., Draghi, M., Little, A. M., Partheniou, F., & Parham, P. (2006). Roles for HLA and KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function. *Journal of Experimental Medicine, 203*(3), 633-645. doi:10.1084/jem.20051884

Zhu, F. M., Jiang, K., Lv, Q. F., He, J., & Yan, L. X. (2006). Investigation of killer cell immunoglobulin-like receptor KIR2DL4 diversity by sequence-based typing in Chinese population. *Tissue Antigens, 67*(3), 214-221. doi:10.1111/j.1399-0039.2006.00562.x

4.7 Supplementary material

Table S1. List of variable sites detected at the *KIR2DL4* gene segment in a Brazilian population sample, considering the segment between the positions 54801519 to 54814906 (chr 19 – hg38).

KIR2DL4 segment	Chr 19 Position (hg38)	IMGT/KIR Relative position	SNPid	Reference allele	Alternative allele	Alternative allele frequency (2n=314)	Note [a]
5'upstream	54802354		rs3916556	C	T	0.433	
5'upstream	54802427		rs111803848	G	T	0.121	
5'upstream	54802632		rs35250545	C	A	0.433	
5'upstream	54802639		rs35160118	T	C	0.433	
5'upstream	54802679		rs34370734	G	A	0.436	
5'upstream	54802832		rs589158	A	G	0.793	
5'upstream	54803022		.	A	G	0.003	
5'upstream	54803145		rs376798518	C	T	0.013	
5'upstream	54803290		rs55881846	G	T	0.013	
5'upstream	54803346		rs144978425	C	CA	0.793	
5'upstream	54803350		rs36116896	T	G	0.793	
5'upstream	54803425	-227	rs56401846	G	A	0.041	A
5'upstream	54803423	-229	rs55808487	C	T	0.029	A
5'upstream	54803437	-215	rs933157437	G	A	0.003	A
5'upstream	54803442	-210	rs35656676	G	C	0.331	B
5'upstream	54803494	-158	rs17173106	T	C	0.331	B
Intron 1	54803852	201	rs2295802	C	G	0.446	B
Intron 1	54803871	221	rs604076	A	C	0.223	B
Exon 2	54803892	241	rs56186508	G	A	0.003	A, C
Exon 2	54803917	266	rs369033322	G	A	0.006	A, C
Intron 2	54803985	334	rs597598	A	G	0.793	B
Intron 2	54804045	532	rs56133375	G	*	0.417	B
Intron 2	54804111	460	rs56136619	C	T	0.006	B
Intron 2	54804112	461	rs200791941	C	T	0.131	B
Intron 2	54804161	544	rs598452	C	T	0.462	B
Intron 2	54804200	583	rs903918210	T	A	0.006	A
Intron 2	54804309	692	rs113882675	G	A	0.016	A
Intron 2	54804329	712	rs376787209	G	C	0.013	B
Intron 2	54804542	925	rs607149	C	G	0.793	B
Intron 2	54804576	959	rs768109588	A	G	0.016	A
Intron 2	54804637	1020	rs17771905	C	T	0.041	B
Exon 3	54804874	1257	rs618835	A	G	0.331	B, D (Y>C)
Exon 3	54805006	1389	.	G	C	0.003	A, D (C/Y)
Intron 3	54805248	1631	rs2364437	G	A	0.013	B
Intron 3	54805365	1748	rs621169	T	C	0.331	B
Intron 3	54805570	1953	rs2075767	C	A	0.229	B
Intron 3	54805585	1968	.	G	C	0.003	A
Intron 3	54805645	2028	rs2075768	C	T	0.334	B

Intron 3	54805678	2061	rs625788	A	T	0.223	B
Intron 3	54805821	2204	rs148546628	G	A	0.108	A
Intron 3	54805833	2232	.	T	TGAGAGA	0.331	B
Exon 4	54805984	2367	rs200485159	C	T	0.019	B, D (P>L)
Exon 4	54806001	2384	rs1051454	G	A	0.334	B, D (R>T)
Exon 4	54806015	2398	.	G	C	0.003	A, D (V>A)
Exon 4	54806069	2452	rs374413343	G	A	0.111	B, C
Exon 4	54806214	2597	rs1051456	C	G	0.334	B, D (Q>E)
Intron 4	54806367	2750	rs373463555	T	C	0.105	B
Intron 4	54806730	3113	rs376403025	G	C	0.032	B
Intron 4	54806736	3119	rs34887390	G	A	0.102	A
Intron 4	54806745	3128	rs657395	T	C	0.790	B
Intron 4	54806770	3153	rs187656554	C	T	0.111	B
Intron 4	54806816	3199	rs657781	T	C	0.452	B
Intron 4	54806888	3271	rs4806447	G	A	0.016	B
Intron 4	54806987	3370	rs4806564	C	T	0.016	B
Intron 4	54807013	3396	rs112384686	C	CTAAA	0.338	A
Intron 4	54807170	3553	rs78428848	C	T	0.334	B
Intron 4	54807186	3573	rs374697369	C	T	0.013	B
Intron 4	54807205	3592	rs659565	A	T	0.331	B
Intron 4	54807212	3599	rs76827991	G	C	0.334	B
Intron 4	54807249	3636	rs659629	T	A	0.838	B
Intron 4	54807463	3850	rs113804843	G	T	0.045	A
Intron 4	54807587	3974	.	G	A	0.003	A
Intron 4	54807594	3981	rs112849748	C	A	0.334	B
Intron 4	54807705	4092	rs376786209	T	C	0.108	B
Intron 4	54807723	4110	rs672001	A	C	0.793	B
Intron 4	54807744	4131	rs4969516	T	C	0.334	B
Intron 4	54807818	4205	rs672486	T	C	0.793	B
Intron 4	54808145	4532	rs78585875	G	T	0.334	B
Intron 4	54808164	4551	rs750734058	G	A	0.006	A
Intron 4	54808219	4606	.	A	AT	0.331	A
Intron 4	54808515	4903	rs373211160	G	C	0.048	B
Intron 4	54808535	4923	rs377627213	T	G	0.003	B
Intron 4	54808545	4933	rs35603706	C	T	0.334	B
Intron 4	54808591	4979	rs62124064	C	T	0.102	A
Intron 4	54808689	5077	rs770466380	A	G	0.010	A
Intron 4	54808745	5133	rs16985962	T	G	0.334	B
Intron 4	54808768	5156	.	A	C	0.003	A
Intron 4	54808784	5172	rs35837163	G	A	0.003	A
Intron 4	54808812	5200	rs56218903	C	A	0.013	A
Intron 4	54808824	5214	rs687423	G	A	0.793	B
Intron 5	54809026	5416	rs200439905	C	G	0.016	B
Intron 5	54809212	5602	rs600898	A	C	0.331	B
Intron 5	54809237	5627	rs76869409	A	G	0.334	B
Intron 5	54809324	5714	rs10500318	G	A	0.102	A
Intron 5	54809414	5804	rs35579530	C	T	0.331	B
Intron 5	54809472	5862	rs592645	T	A	0.223	B

Intron 5	54809598	5988	rs1015105183	T	C	0.006	A
Intron 5	54809643	6033	rs112006653	G	A	0.003	A
Intron 5	54809822	6212	rs111735328	A	G	0.006	A
Intron 5	54809941	6331	rs604077	T	C	0.682	B
Intron 5	54810058	6448	rs16986024	G	A	0.334	B
Intron 5	54810126	6516	rs138956580	GT	G	0.108	B
Intron 5	54810142	6532	rs604999	A	G	0.838	B
Intron 5	54810151	6541	rs185767446	C	T	0.108	B
Intron 5	54810285	6675	.	A	AT	0.111	A
Intron 5	54810298	6688	.	A	AT	0.239	B
Intron 5	54810401	6791	rs616496	A	C	0.223	B
Intron 5	54810518	6908	rs544859823	G	GA	0.003	A
Intron 5	54810630	7020	rs371261951	G	C	0.111	B
Intron 5	54810646	7036	rs608316	C	G	0.338	B
Intron 5	54810647	7037	rs529464965	A	G	0.108	B
Intron 5	54810672	7062	rs3865505	G	A	0.016	B
Intron 5	54810689	7079	rs608368	A	T	0.790	B
Intron 5	54810714	7104	rs617872	C	G	0.771	B
Intron 5	54810724	7114	rs3865506	C	T	0.016	B
Intron 5	54810735	7125	rs617907	T	C	0.347	B
Intron 5	54810746	7136	rs17173113	A	C	0.108	B
Intron 5	54810921	7311	rs3865507	C	A	0.016	B
Intron 5	54810968	7358	rs618871	G	C	0.223	B
Intron 5	54811247	7637	rs111490149	A	T	0.006	A
Intron 5	54811269	7659	rs630497	G	A	0.347	B
Intron 5	54811291	7681	rs3786857	G	A	0.277	B
Intron 5	54811324	7714	rs112954606	G	C	0.006	A
Intron 5	54811353	7743	.	G	A	0.003	A
Intron 5	54811364	7754	rs368909099	C	T	0.022	B
Intron 5	54811419	7809	rs201457831	C	A	0.016	B
Intron 5	54811442	7832	.	G	A	0.003	A
Intron 5	54811520	7910	rs631717	G	A	0.561	B
Intron 5	54811574	7964	rs112264748	T	A	0.006	A
Intron 5	54811602	7992	rs632126	A	G	0.229	B
Intron 5	54811619	8009	rs56136222	G	C	0.006	B
Intron 5	54811638	8028	rs17771961	C	G	0.108	B
Intron 5	54812145	8535	rs77611365	G	C	0.350	B
Intron 5	54812181	8571	rs371450241	G	T	0.045	B
Intron 5	54812575	8965	rs3865508	T	G	0.350	B
Intron 5	54812586	8976	rs3865509	C	G	0.016	B
Intron 5	54812784	9174	rs3865510	C	A	0.357	B
Intron 5	54812863	9254	rs11375633	A	AC	0.793	B
Intron 5	54813019	9410	rs648689	T	C	0.580	B
Intron 5	54813111	9502	rs649117	C	T	0.580	B
Exon 6	54813180	9571	rs649216	T	C	0.580	B, C (F>L)
Exon 6	54813219	9620	.	C	CA	0.532	E
Intron 6	54813377	9769	rs660437	C	A	0.223	B
Intron 6	54813405	9797	rs660773	G	A	0.580	B

Intron 6	54813482	9874	.	A	G	0.003	A
Intron 6	54813668	10060	rs652671	T	C	0.229	B
Intron 7	54813785	10177	rs653140	G	C	0.580	B
Intron 7	54813786	10178	rs653142	A	T	0.580	B
Exon 8	54813863	10255	rs56184317	C	G	0.006	A, D (Q>E)
Exon 8	54813995	10387	rs35946789	G	C	0.010	A, D (A/P)
Exon 8	54813996	10388	rs374021043	C	T	0.003	A, D (A>L)
Exon 8	54814000	10392	rs1051457	G	A	0.204	B, C
Exon 8|3'UTR	54814116	10508	rs370869656	G	A	0.010	A
Exon 8|3'UTR	54814169	10561	rs376917646	G	C	0.048	B
Exon 8|3'UTR	54814361	10753	rs34785252	C	A	0.201	B
Exon 8|3'UTR	54814366	10758	rs374254587	C	G	0.013	B
3'downstream	54814618		rs676443	T	C	0.796	
3'downstream	54814619		rs667205	G	T	0.580	
3'downstream	54814646		rs667263	G	C	0.239	
3'downstream	54814806		rs76100532	T	C	0.350	
3'downstream	54814823		rs377747517	G	C	0.045	
3'downstream	54814824		rs200660658	G	A	0.016	

Table S2. List of singletons (variable sites detected only in one sample) without defined haplotype detected at the *KIR2DL4* gene on a Brazilian population sample.

***KIR2DL4* segment**	**Chr 19 Position (hg38)**	**SNPid**	**IMGT/KIR relative position**	**Reference allele**	**Alternative allele**
5'upstream	54803043	rs748539788		T	C
Intron 1	54803730	rs372376615	122	T	C
Intron 2	54804426	.	809	C	T
Intron 2	54804731	rs560341968	1114	C	A
Intron 3	54805255	rs755023073	1638	A	G
Exon 4	54805998	rs989403226	2381	C	T
Exon 4	54806126	rs577091862	2509	G	A
Intron 4	54807462	rs77055523	3845	C	G
Intron 4	54807657	rs113418417	4040	C	A
Intron 4	54808115	rs113064120	4498	T	C
Intron 4	54808482	.	4865	G	A
Intron 4	54808514	rs796547461	4897	G	T
Intron 5	54809639	.	6029	C	T
Intron 5	54810869	rs572249134	7259	C	G
Intron 5	54811872	rs938247389	8262	G	A
Intron 5	54811879	.	8269	A	T
Intron 5	54812599	.	8989	A	G
Intron 5	54812753	.	9143	C	T
Intron 5	54812771	.	9161	G	C
Intron 5	54812975	rs55736818	9365	G	A
Exon 6	54813158	rs192263392	9593	A	G
Intron 6	54813679	rs372293579	10071	G	A

The segment that is tracked by the IMGT/HLA database is indicated in gray.

Table S3. List of variable sites detected at the *KIR2DL4* gene on the *1000Genomes* Project (Phase 3) database considering the segment between the positions 55312974 to 55326339 (GRhg37) followed by the alternative allele frequency found in a Brazilian population sample.

KIR2DL4 **segment**	**Chr 19 Position (hg37)**	**SNPid**	**Reference allele**	**Alternative allele**	**Alternative allele frequency 1000G (2n=5008)**	**Alternative allele frequency Brazilian population (2n=314)**
5'upstream	55314317	rs544778719	T	A	0.0473	.
5'upstream	55314838	rs555206298	G	A	0.0002	.
5'upstream	55314880	rs375557142	G	A	0.0555	.
5'upstream	55314897	rs35656676	G	C	0.5116	0.3312
5'upstream	55314949	rs17173106	T	C	0.5118	0.3312
5'upstream	55314997	rs528053789	C	T	0.0002	.
5'upstream	55315054	rs544626886	C	G	0.0002	.
5'upstream	55315068	rs564804446	C	T	0.0002	.
5'upstream	55315103	rs530717281	C	T	0.0004	.
Intron 1	55315249	rs550507964	A	G	0.0002	.
Intron 1	55315276	rs560915366	G	A	0.0002	.
Intron 1	55315307	rs2295802	C	G	0.626	0.4459
Exon 2	55315347	rs200150084	G	A	0.0018	.
Exon 2	55315372	rs369033322	G	A	0.0002	0.0064
Intron 2	55315411	rs538996766	A	C	0.0006	.
Intron 2	55315440	rs597598	A	G	0.7987	0.7930
Intron 2	55315566	rs569528157	C	T	0.0002	.
Intron 2	55315567	rs200791941	C	T	0.1713	0.1306
Intron 2	55315590	rs371120439	A	G	0.0002	.
Intron 2	55315687	rs574920795	G	T	0.0004	.
Intron 2	55315764	rs113882675	G	A	0.0076	0.0159
Intron 2	55315784	rs376787209	G	C	0.017	0.0127
Intron 2	55315954	rs542783030	AT	A	0.0022	.
Intron 2	55315997	rs607149	C	G	0.4091	0.7930
Intron 2	55316050	rs374093219	G	A	0.013	.
Intron 2	55316059	rs564536710	C	G	0.0002	.
Intron 2	55316092	rs376449393	C	T	0.0178	.
Intron 2	55316098	rs544084384	G	A	0.0004	.
Intron 2	55316107	rs561261944	G	A	0.0004	.
Intron 2	55316170	rs367959035	G	A	0.001	.
Intron 2	55316174	rs546785513	G	A	0.0002	.
Intron 2	55316186	rs560341968	C	A	0.0004	.
Exon 3	55316279	rs532657894	G	A	0.0002	.
Exon 3	55316286	rs369994438	G	A	0.0006	.
Exon 3	55316304	rs200922243	G	A	0.0036	.
Exon 3	55316306	rs538279134	A	C	0.0002	.
Exon 3	55316319	rs372914780	C	T	0.0002	.

Exon 3	55316334	rs191351791	C	G	0.0002	.
Exon 3	55316354	rs534098694	G	A	0.0002	.
Exon 3	55316463	rs554354206	C	T	0.0004	.
Exon 3	55316477	rs577410556	G	A	0.0004	.
Intron 3	55316564	rs538336951	T	A	0.0012	.
Intron 3	55316595	rs558263071	G	A	0.0002	.
Intron 3	55316663	rs574870505	G	A	0.0004	.
Intron 3	55316703	rs367555183	G	A	0.0232	.
Intron 3	55316738	rs560988603	T	A	0.0048	.
Intron 3	55316800	rs574682397	G	A	0.0002	.
Intron 3	55316966	rs368657774	G	A	0.0002	.
Intron 3	55317025	rs2075767	C	A	0.348	0.2293
Intron 3	55317100	rs2075768	C	T	0.5074	0.3344
Intron 3	55317101	rs546305156	C	T	0.0002	.
Intron 3	55317125	rs372323051	C	T	0.0006	.
Intron 3	55317194	rs531711276	G	A	0.0022	.
Intron 3	55317262	rs548580350	C	T	0.0002	.
Intron 3	55317276	rs148546628	G	A	0.1052	0.1083
Intron 3	55317291	rs182171195	G	A	0.0036	.
Intron 3	55317344	rs547816805	G	C	0.0002	.
Intron 3	55317345	rs570695268	G	T	0.0004	.
Exon 4	55317414	rs539671377	G	A	0.0002	.
Exon 4	55317439	rs200485159	C	T	0.0541	0.0191
Exon 4	55317448	rs376428959	C	T	0.0006	.
Exon 4	55317454	rs537400112	G	A	0.0002	.
Exon 4	55317456	rs1051454	G	A	0.516	0.3344
Exon 4	55317524	rs374413343	G	A	0.1122	0.1115
Exon 4	55317528	rs540514355	G	A	0.0002	.
Exon 4	55317550	rs553817989	C	A	0.0034	.
Exon 4	55317581	rs577091862	G	A	0.0004	.
Exon 4	55317585	rs546044168	G	A	0.0004	.
Exon 4	55317633	rs562937136	G	A	0.004	.
Exon 4	55317669	rs1051456	C	G	0.5345	0.3344
Intron 4	55318200	rs145527611	T	C	0.5389	.
Intron 4	55318252	rs562104845	C	T	0.0002	.
Intron 4	55318270	rs527719717	C	A	0.0002	.
Intron 4	55318343	rs201322663	G	A	0.0575	.
Intron 4	55318625	rs78428848	C	T	0.5409	0.3344
Intron 4	55318641	rs374697369	C	T	0.0248	0.0127
Intron 4	55318667	rs76827991	G	C	0.5409	0.3344
Intron 4	55318704	rs659629	T	A	0.611	0.8376
Intron 4	55318918	rs113804843	G	T	0.0162	0.0446
Intron 4	55320199	rs370190236	A	T	0.0002	.
Intron 4	55320200	rs16985962	T	G	0.4553	0.3344
Exon 5	55320292	rs533582948	C	A	0.0004	.
Intron 5	55320401	rs554082856	C	T	0.0004	.
Intron 5	55320442	rs577226306	C	A	0.0006	.
Intron 5	55320481	rs200439905	C	G	0.0519	0.0159

Intron 5	55320512	rs556397525	G	A	0.0004	.
Intron 5	55320687	rs576344643	G	A	0.0002	.
Intron 5	55320692	rs76869409	A	G	0.5567	0.3344
Intron 5	55320779	rs10500318	G	A	0.094	0.1019
Intron 5	55320869	rs35579530	C	T	0.5122	0.3312
Intron 5	55320907	rs541080836	G	A	0.0002	.
Intron 5	55320984	rs564396483	C	G	0.0054	.
Intron 5	55321213	rs533445355	C	T	0.0002	.
Intron 5	55321277	rs111735328	A	G	0.0142	0.0064
Intron 5	55321396	rs604077	T	C	0.6615	0.6815
Intron 5	55321399	rs192835571	T	C	0.0034	.
Intron 5	55321403	rs548246951	G	T	0.0004	.
Intron 5	55321423	rs568087247	C	G	0.0002	.
Intron 5	55321513	rs16986024	G	A	0.5052	0.3344
Intron 5	55321606	rs185767446	C	T	0.0801	0.1083
Intron 5	55321740	rs199888594	A	AT	0.0903	.
Intron 5	55321741	rs570712469	T	C	0.0004	.
Intron 5	55321742	rs539742342	T	A	0.0002	.
Intron 5	55321753	rs370767457	A	T	0.0004	.
Intron 5	55321792	rs576163531	A	G	0.0002	.
Intron 5	55321863	rs535311853	T	C	0.0002	.
Intron 5	55321880	rs555796172	C	A	0.0006	.
Intron 5	55321973	rs544859823	G	GA	0.0152	0.0032
Intron 5	55321989	rs572454893	C	A	0.0002	.
Intron 5	55322004	rs540977253	C	T	0.0002	.
Intron 5	55322045	rs564135163	A	G	0.001	.
Intron 5	55322075	rs577653426	G	A	0.0002	.
Intron 5	55322085	rs371261951	G	C	0.0729	0.1115
Intron 5	55322101	rs608316	C	G	0.0897	0.3376
Intron 5	55322102	rs529464965	A	G	0.0729	0.1083
Intron 5	55322127	rs113969567	G	A	0.0749	.
Intron 5	55322134	rs561783405	C	T	0.0162	.
Intron 5	55322135	rs113577571	A	G	0.0234	.
Intron 5	55322162	rs548243607	T	G	0.0004	.
Intron 5	55322169	rs78704235	C	G	0.7338	.
Intron 5	55322179	rs199966994	C	T	0.0553	.
Intron 5	55322201	rs184787616	A	C	0.0074	.
Intron 5	55322207	rs535594746	C	T	0.0166	.
Intron 5	55322209	rs555272549	T	C	0.0166	.
Intron 5	55322324	rs572249134	C	G	0.0008	.
Intron 5	55322376	rs3865507	C	A	0.0591	0.01592
Intron 5	55322408	rs557867687	G	A	0.0002	.
Intron 5	55322423	rs618871	G	C	0.0002	0.2229
Intron 5	55322476	rs543436315	T	G	0.0004	.
Intron 5	55322672	rs557601074	G	A	0.0002	.
Intron 5	55322690	rs3786855	T	A	0.3688	.
Intron 5	55322702	rs111490149	A	T	0.0164	0.0064
Intron 5	55322746	rs3786857	G	A	0.4549	0.0064

Intron 5	55322779	rs112954606	G	C	0.0164	.
Intron 5	55322790	rs541104573	T	C	0.0002	.
Intron 5	55322797	rs377528700	G	C	0.007	.
Intron 5	55322810	rs568153700	G	A	0.0002	.
Intron 5	55322819	rs368909099	C	T	0.0455	0.0223
Intron 5	55322874	rs201457831	C	A	0.0553	0.0159
Intron 5	55322893	rs529169038	A	G	0.0002	.
Intron 5	55322942	rs549298662	C	G	0.0002	.
Intron 5	55323021	rs565719007	G	A	0.0002	.
Intron 5	55323029	rs112264748	T	A	0.0142	0.0064
Intron 5	55323093	rs17771961	C	G	0.095	0.1083
Intron 5	55324171	rs571426631	T	G	0.0559	.
Intron 5	55324351	rs537009792	G	A	0.0004	.
Intron 5	55324430	rs190228734	G	A	0.0066	.
Intron 5	55324474	rs648689	T	C	0.4179	0.5796
Exon 6	55324566	rs649117	C	T	0.4147	0.5796
Exon 6	55324596	rs371301480	G	A	0.001	.
Exon 6	55324613	rs192263392	A	G	0.0004	.
Exon 6	55324617	rs545633761	A	C	0.0004	.
Exon 6	55324635	rs649216	T	C	0.5907	0.5796
Intron 6	55324757	rs533060800	G	A	0.0012	.
Intron 6	55324860	rs660773	G	A	0.4814	0.5796
Intron 6	55324902	rs563126913	C	T	0.0006	.
Intron 6	55324920	rs528914319	T	G	0.0002	.
Intron 6	55324970	rs548986354	C	G	0.0002	.
Intron 6	55324975	rs565988654	A	G	0.0004	.
Intron 6	55325016	rs528239645	T	G	0.0004	.
Intron 6	55325044	rs551174013	C	T	0.0008	.
Intron 6	55325049	rs571361166	G	A	0.0002	.
Intron 6	55325076	rs537145225	G	A	0.0004	.
Intron 6	55325123	rs652671	T	C	0.0004	0.2293
Exon 7	55325147	rs201370666	G	A	0.0016	.
Exon 7	55325193	rs536671464	G	T	0.0002	.
Exon 7	55325197	rs376492770	G	A	0.0012	.
Intron 7	55325271	rs573588301	C	T	0.0002	.
Intron 7	55325287	rs570638678	C	G	0.0002	.
Exon 8	55325302	rs559190217	T	G	0.0002	.
Exon 8	55325450	rs143893213	G	C	0.0034	.
Exon 8	55325451	rs374021043	C	T	0.011	0.0032
Exon 8	55325455	rs1051457	G	A	0.2278	0.2038
Exon 8	55325531	rs200678651	C	T	0.0034	.
3'downstream	55325571	rs370869656	G	A	0.0004	0.0096
3'downstream	55325573	rs539322264	G	C	0.0024	.
3'downstream	55325816	rs34785252	C	A	0.2796	0.2006
3'downstream	55326067	rs551309750	T	C	0.0002	.
3'downstream	55326073	rs676443	T	C	0.4105	0.7962
3'downstream	55326199	rs530769114	C	A	0.0002	.
3'downstream	55326261	rs76100532	T	C	0.5631	0.3503

3'downstream	55326278	rs377747517	G	C	0.0272	0.0446