



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

Gabriela Sato Paes

**DIVERSIDADE DE NÚMERO DE CÓPIAS E VARIANTES
ALÉLICAS DO GENE *KIR2DS4* EM POPULAÇÕES
MUNDIAIS**

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

Orientador: Prof. Dr. Erick da Cruz Castelli
Co-orientadora: Profa. Dra. Camila Ferreira Bannwart Castro

**Botucatu
2024**

Gabriela Sato Paes

DIVERSIDADE DE NÚMERO DE CÓPIAS E
VARIANTES ALÉLICAS DO GENE *KIR2DS4* EM
POPULAÇÕES MUNDIAIS

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

Orientador (a): Prof.Dr. Erick da Cruz Castelli
Co-orientadora: Profa.Dra. Camila Ferreira Bannwart Castro

Botucatu
2024

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: MARIA CAROLINA A. CRUZ E SANTOS-CRB 8/10188

Paes, Gabriela Sato.

Diversidade de número de cópias e variantes alélicas do gene *KIR2DS4* em populações mundiais / Gabriela Sato Paes. - Botucatu, 2024

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu
Orientador: Erick da Cruz Castelli
Coorientador: Camila Ferreira Bannwart Castro
Capes: 40105008

1. Receptores KIR. 2. Sequenciamento de Nucleotídeos em Larga Escala. 3. Cromossomos - Polimorfismo.

Palavras-chave: KIR2DS4; KIR2DS4-del; NGS; Polimorfismo.

Dedicatória

Dedico este trabalho à minha família Sato e Paes, obrigada por me apoiarem, pelo suporte e o incentivo nas minhas escolhas.

Agradecimientos

Agradeço ao meu orientador, Prof. Dr. Erick da Cruz Castelli, pela oportunidade em me aceitar como aluna de mestrado. Obrigada pelo suporte, orientação, apoio, ensinamentos principalmente na área da bioinformática e pela paciência durante este período de mestrado.

A minha co-orientadora, Profa. Dra. Camila Ferreira Bannwart Castro, pela confiança em apresentar para o laboratório GeMBio desde 2020, pelo apoio, suporte, por me mostrar uma visão diferente em todos os aspectos, inclusive a respeito de mim mesma. Obrigada por mostrar para mim que a imunologia é linda, e me incentivar a continuar a estudar essa área.

Agradeço à minha família Sato, Paes e Kanashiro por sempre me apoiar, me incentivar e jamais largaram a minha mão nos momentos mais difíceis. Em especial, minha avó Irene, meu avô Shideo, minha mãe Emiliana, minha tia Daiane que me apoiaram, me incentivaram, me ajudaram em momentos que foram difíceis, vocês são minha inspiração e a minha motivação.

Ao meu pai Gerson, padrasto Milton, madrasta Simone, meus irmãos Mioko, Sthe, José e Guilherme, primos Benedicto, Miguel e Luísa, tio Rogério, tio Anderson, tia Silvana e tia Estela, muito obrigada pelo apoio.

A todos os meus amigos, em especial as minhas amigas Gabi Rando, Rosane, Thaís, Marcela, Maria Júlia Sako, Juliana Rizatto, Karen Mariotto, Thaís Zucchi e aos meus amigos Guilherme Augusto, Higor Oliveira, Hugo Akira, José, Daniel Nascimento, José Zucchi por estarem comigo. Obrigada por todo o apoio e companheirismo.

As minhas amigas e “*roommates*” Viviane Aparecida de Oliveira Ciriaco e Raphaela Neto Pereira, agradeço muito por me receberem em Botucatu, me acolherem quando mais precisei, por aguentar todos os dias, pelos conselhos e nossas sessões de terapia em conjunto. Estarei sempre grata por terem vocês em minha vida, amo vocês!

Agradeço ao nosso grupo KIR por estarmos juntos nessa batalha para analisar esses genes.

As minhas parceiras e parceiros de laboratório, Nayane Silva (minha francesa preferida), Heloísa Andrade com todo apoio e suporte, Isabelle Silva (minha parceira de I.C), Lívia Ramos (minha querida perita), Marcel Ferreira (grande Marshal), Ícaro Scalisse (guerreiro do grupo KIR), João Toloí, Ingrid Ferraz, Luise Tini, Nathaly Ogata, Luciana Borali (ombro amigo). Obrigada por me ajudarem, pelas conversas e desabafos.

Ao programa de Pós-Graduação em Patologia da Faculdade de Medicina de Botucatu (FMB-UNESP), pela oportunidade e por confiarem em mim.

Aos membros da banca avaliadora, por aceitarem o convite e contribuírem com a minha formação.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), que possibilitou dedicação exclusiva para a elaboração do trabalho através da bolsa de mestrado.

A todos, que contribuíram com os meus estudos e pesquisa, durante todos esses anos.

Por fim, gostaria de agradecer a minha companheira de quatro patas, Nina, por sempre estar comigo em todos os momentos, me acompanhar durante os momentos de escrita, não sei o que seria de mim sem você, Ninoca! Te amo muito e espero viver mais oportunidades com você por mais anos.

Epígrafe

“Como a neve que acabou de se acomodar
Vamos respirar, como na primeira vez”

(Zero O’Clock - BTS)

SUMÁRIO

Lista de Figuras	12
Lista de Figuras - Capítulo I.....	12
Lista de Figuras - Capítulo II (artigo).....	12
Lista de Figuras do Capítulo II (artigo) - Material Suplementar.....	12
Lista de Tabelas	14
Lista de Tabelas - Capítulo I.....	14
Lista de Tabelas - Capítulo II (artigo).....	14
Lista de Tabelas do Capítulo II (artigo) - Material Suplementar.....	14
Lista de Abreviaturas e Siglas	14
RESUMO	18
ABSTRACT	19
CAPÍTULO I - REVISÃO DE LITERATURA	20
INTRODUÇÃO	21
1. As células Natural Killer e os receptores KIR.....	21
2. Receptores KIR e associação com doenças.....	29
3. Gene KIR2DS4.....	31
Frequência de KIR2DS4-full e KIR2DS4-del em diferentes populações.....	34
4. Dificuldades metodológicas para análise dos genes KIR.....	35
Caracterização dos bancos de dados genômicos públicos: Projeto 1000 Genomas utilizado no estudo.....	38
JUSTIFICATIVA	39
OBJETIVOS	40
Objetivo Geral.....	40
Objetivos Específicos.....	40
REFERÊNCIAS	41
CAPÍTULO II - ARTIGO	47
Abstract	48
Keywords: KIR2DS4, copy number variation, NGS, KIR2DS4-del	48
Introduction	49
Methods	51
The kir-mapper workflow and read alignment: the map algorithm.....	51
Determining copy numbers: the ncopy algorithm.....	53
Calling SNPs and KIR alleles: the genotype algorithm.....	54
Testing kir-mapper with samples from all biogeographic regions.....	55
Comparing Kir-mapper and PING.....	56
Results	56
Presence/absence of KIR2DS4 in different biogeographic regions.....	56
The number of KIR2DS4 copies among individuals.....	57
KIR2DS4 allele distribution.....	58
The frequency of a functional KIR2DS4.....	59
Comparing Kir-mapper and PING.....	60
Discussion	62

References.....	65
Supplementary.....	68
CONCLUSÃO.....	77

Lista de Figuras

Lista de Figuras - Capítulo I

Figura 1. Mecanismos de regulação da atividade de células NK, de acordo com a interação de seus receptores com os ligantes da célula-alvo ou ausência desses ligantes.....	23
Figura 2. Localização dos genes <i>KIR</i> no cromossomo 19.....	25
Figura 3. Haplótipos A e B dos genes <i>KIR</i>	26
Figura 4. Estrutura dos receptores KIR.....	28
Figura 5. Nomenclatura utilizada para os genes <i>KIR</i>	29

Lista de Figuras - Capítulo II (artigo)

Figure 1: The kir-mapper workflow to produce BAM files for each sample, call copy numbers, SNPs, and alleles.....	51
Figure 2: The differences observed for read alignment when using BWA (A) and kir-mapper (B) considering gene <i>KIR2DS4</i>	52
Figure 3: The <i>KIR2DS4</i> copy number determination using kir-mapper::ncopy for CLM population.....	54
Figure 4: The proportion of individuals carrying at least one copy of <i>KIR2DS4</i> in different biogeographic regions.....	57
Figure 5: The number of copies of <i>KIR2DS4</i> in different populations.....	58

Lista de Figuras do Capítulo II (artigo) - Material Suplementar

Figure S1: Kir-mapper and PING genotypes compare (<i>KIR2DS4</i>) for Colombia (CLM).....	68
Figure S2: Kir-mapper and PING genotypes compare (<i>KIR2DS4</i>) for British (GBR).....	69
Figure S3: Kir-mapper and PING genotypes compare (<i>KIR2DS4</i>) for Yoruba (ITU).....	69
Figure S4: Kir-mapper and PING genotypes compare (<i>KIR2DS4</i>) for Japanese (JPT).....	70
Figure S5: Kir-mapper and PING genotypes compare (<i>KIR2DS4</i>) for Yoruba (YRI).....	71

Figure S6: The KIR2DS4 copy number determination using kir-mapper::ncopy for GBR.....	72
Figure S7: The KIR2DS4 copy number determination using kir-mapper::ncopy for JPT.....	73
Figure S8: The KIR2DS4 copy number determination using kir-mapper::ncopy for ITU.....	74
Figure S9: The KIR2DS4 copy number determination using kir-mapper::ncopy for YRI.....	75
Figure S10: Number of copies of KIR2DS4 in all the samples. The graphic represents frequency (%) and the quantity of the samples of each population of 1000 Genomes Project (17), CLM, GBR, JPT, ITU and YRI.....	75

Lista de Tabelas

Lista de Tabelas - Capítulo I

Tabela 1: A distribuição dos alelos de acordo com os genótipos KIR2DS4-full e KIR2DS4-del.....	34
--	----

Lista de Tabelas - Capítulo II (artigo)

Table 1: The frequency of <i>KIR2DS4</i> alleles in different populations.....	59
Table 2: Frequency (%) of samples carrying different KIR2DS4 genotypes in five populations from the 1000Genomes Project.....	60

Lista de Tabelas do Capítulo II (artigo) - Material Suplementar

Table S1: The frequency of different genotypes regarding the number of copies of KIR2DS4 in different populations.....	68
--	----

Lista de Abreviaturas e Siglas

1000 Genomas	Projeto 1000 Genomas
aGVHD	Doença aguda do enxerto contra o hospedeiro
bp	Pares de base
BWA	Alinhador Burrows-Wheeler
CLM	Colombianos de Medellin
CCL3	Quimiocina de CC-3
CCL4	Quimiocina de CC-4
CDS	Sequências codificadoras
CNV	Varição no número de cópias
DNA	Ácido desoxirribonucleico
CD16	Cluster de diferenciação 16
CD56^{bright}	Subpopulação de células NK com alta densidade de CD56
CD56^{dim}	Subpopulação de células NK com baixa densidade de CD56

CEACAM	Molécula de adesão celular relacionada com antígeno de carcinoembrionário
COVID-19	Doença por Coronavírus 2019
DAP10	Proteína ativadora de DNX-10
DAP12	Proteína ativadora de DNX-12
DC	Célula dendrítica
DM	Diabetes mellitus
GBR	Britânicos na Inglaterra e Escócia
GPVI	Glicoproteína VI
GWAS	Estudos de associação genômica ampla
HCV	Hepatite C
hg38	Homo sapiens (human) genome assembly GRCh38
HIV	Vírus da Imunodeficiência Humana
HLA	Antígeno Leucocitário Humano
HLA-Cw4	Variação de alelo do sistema de antígenos leucocitários humanos (HLA)-C
HLA-Cw6	Variação de alelo do sistema de antígenos leucocitários humanos (HLA)-C
HLA-I	Antígeno Leucocitário Humano de Classe I
HSCT	Transplantes de células-tronco hematopóéticas
IFN-γ	Interferon-gamma
IL-10	Interleucina 10
IL-2	Interleucina 2
InDel	Inserção/Deleção
IPD-KIR	The Immuno Polymorphism Database - Killer Cell Immunoglobulin-Like Receptor
ITAM	Domínio citoplasmático do imunoreceptor ativador baseado em tirosina
ITIM	Imunoreceptor inibidor baseado em tirosina
ITU	Indiano Telugu

JPT	Japoneses em Tóquio
KIR	Receptores Killer Cell Immunoglobulin-Like Receptor
KIR2DS4	Killer cell immunoglobulin like receptor, two Ig domains and short cytoplasmic tail 4
KIR2DS4*unresolved	Alelo de KIR2DS4 novo ou não-resolvido
KIR2DS4-del	KIR2DS4 com deleção de 22 pares de base no éxon 5
KIR2DS4-full	Alelo de KIR2DS4 funcional
KIR2DS4-null	Ausência de KIR2DS4
KLR	Receptores semelhantes à lectina em células assassinas
LAIR	Leucócitos associados com domínio com receptor-I
LAMP-1	Proteína 1 de membrana associada ao lisossomo
LILR	Receptores leucocitários semelhantes à imunoglobulina
LRC	Complexo de Receptores Leucocitários
MHC	Complexo principal de histocompatibilidade
MIP-1β	Proteína inflamatória de macrófagos-1 beta
NGS	Sequenciamento de Nova Geração
NK	Célula Natural Killer
NKG2	Natural Killer Group 2
OSCAR	Receptor associado com osteoclasto
PCR	Reação em Cadeia da Polimerase
PCR-SSP	Reação em Cadeia da Polimerase com primer de sequência específica
PSG	Glicoproteína gravidez específica
qRT-PCR	PCR quantitativa em tempo real
RA	Artrite Reumatoide
SABE	Saúde, Bem-Estar e Envelhecimento
SARS-CoV-2	Coronavirus 2 da Síndrome Respiratória Aguda Grave
SHP-1/2	Proteína Tirosina Fosfatase Src Homology 2

Siglecs	Sigleucinas
SNP	Polimorfismo de base única
T CD4+	Células T com marcador de superfície CD4
T CD8+	Células T com marcador de superfície CD8
Tel-A	Região telomérica A
TGS	Sequenciamento de terceira geração
TNF-α	Fator de necrose tumoral-alfa
VCF	Variant call format
WGS	Sequenciamento completo
YRI	Yoruba em Ibadã

Título da Dissertação - Diversidade de Número de Cópias e Variantes Alélicas do gene *KIR2DS4* em Populações Mundiais

Impacto científico esperado: Desenvolvimento de uma nova metodologia para análise dos genes *KIR*, em especial *KIR2DS4*; compreender a distribuição dos polimorfismos do gene *KIR2DS4* e suas frequências em diversas populações ao redor do mundo, fornecendo uma base sólida para investigações futuras envolvendo análise funcional.

Impacto social esperado: a possibilidade de avaliação dos genes *KIR* em genomas completos e exomas permitirá a criação de modelos estatísticos que detectam esses alelos em plataformas de genotipagem. Com isso, poderemos avaliar se esses genes estão relacionados com diversas doenças, em estudos de associação que já foram realizados, como, por exemplo, Covid-19 e Dengue, definindo variantes que podem influenciar o prognóstico e o tratamento de diversas doenças.

Impacto econômico: independência do Brasil sobre programas de análise desses genes, visto que os kits atuais de sequenciamento e as estratégias de análise são comerciais; possibilidade de incluir os genes *KIR* na gama de genes considerados para decisões terapêuticas e seleção de doadores.

RESUMO

Os receptores KIR e NKG2 estão presentes na superfície das células Natural Killer (NK), desempenhando um papel crucial principalmente na resposta imune inata. Esses receptores interagem com ligantes presentes nas células-alvo e podem levar à ativação das células NK e liberação de grânulos citotóxicos e/ou citocinas, com destruição da célula alvo, ou podem desencadear mecanismos de inibição da citotoxicidade. O gene *KIR2DS4* é responsável pela codificação de um receptor ativador dentre genes *KIR* presentes no cromossomo 19. O receptor KIR2DS4 se liga principalmente aos antígenos leucocitários de histocompatibilidade HLA-C, presentes na superfície de células somáticas em baixa quantidade. Existem duas versões do receptor KIR2DS4: funcional (proteína completa, KIR2DS4-full) e não funcional (proteína incompleta ou truncada, KIR2DS4-del) devido a uma deleção de 22 pares de base no éxon 5 que resulta em um códon de parada prematuro. Estudos indicam possíveis associações entre este receptor e suas variantes em transplantes, gravidez e infecções virais. Para esta dissertação, foi desenvolvido uma metodologia computacional para minimizar erros de alinhamento e mapeamento devido à alta similaridade e à natureza polimórfica dos genes *KIR*. Com essa nova tecnologia, detectamos SNPs e *indels* ao longo do gene *KIR2DS4*, inferimos o número de cópias deste genes e quais alelos estão presentes em cada indivíduo estudado. Aplicamos esse método em 490 amostras oriundas da Europa, África, América do Sul e Ásia, todas do projeto 1000 Genomas. As maiores frequências de ausência de KIR2DS4 foram observadas no sul da Ásia e na América (Colômbia), com 10.8% e 8.6%, respectivamente. Identificamos 12 alelos diferentes, sendo os mais mais frequentes: KIR2DS4*001 (33.1%); KIR2DS4*003 (15.9%) e KIR2DS4*null (ausência de KIR2DS4, 22.3%). Os indivíduos do sul da Ásia geralmente apresentam KIR2DS4 não funcional, seja pela ausência do gene ou pela presença de um alelo truncado. A descrição da diversidade do gene *KIR2DS4* em todas essas populações contribuirá para um melhor entendimento do papel desse receptor na gestação ou em diversas condições patológicas, como doenças autoimunes ou infecciosas e tumores.

Palavras-chave: *KIR2DS4*, polimorfismo, KIR2DS4-del, NGS.

ABSTRACT

The KIR and NKG2 receptors in the Natural Killer (NK) cell surface play a crucial role mainly in the innate immune response. These receptors interact with ligands present on target cells, modulating the NK cell cytotoxic function. The *KIR2DS4* gene is responsible for encoding an NK activating receptor. The KIR2DS4 receptor binds mainly to the human leukocyte antigen HLA-C antigens, which are present in low quantities on the surface of somatic cells. This receptor might be present or absent in an individual, and, when present, might present two versions, the full-length one and a truncated and non-functional one. Studies indicate possible associations between this receptor and its variants in transplants, pregnancy, and viral infections. For this dissertation, we developed a computational methodology to genotype KIR genes from whole-genome sequencing data. It allowed for a more accurate detection of SNPs and indels along the *KIR2DS4* gene, determination of copy numbers of different alleles. This method was applied to 490 samples from Europe, Africa, South America, and Asia, all from the 1000 Genomes Project. The highest frequencies of KIR2DS4 absence were observed in South Asia and America (Colombia), with 10.8% and 8.6%, respectively. We also detected 12 different alleles, and the most frequent ones are: KIR2DS4*001 (33.1%); KIR2DS4*003 (15.9%), and KIR2DS4*null (absence of KIR2DS4, 22.3%). Most individuals in South Asia present a non-functional KIR2DS4 (comprising KIR2DS4 absence or the KIR2DS4* absence). Describing the diversity of the KIR2DS4 gene in all these populations will contribute to a better understanding of the role of this receptor in gestation or in various pathological conditions, such as autoimmune diseases, infectious diseases, and tumors.

Key-words: *KIR2DS4*, polymorphism, KIR2DS4-del, NGS.

CAPÍTULO I - REVISÃO DE LITERATURA

INTRODUÇÃO

1. As células *Natural Killer* e os receptores KIR

O sistema imune tem sua origem durante o desenvolvimento fetal, mas sua maturação completa ocorre após o nascimento, continuando ao longo da infância e adolescência (1). A pressão exercida por patógenos tem sido a principal força na evolução do sistema imunológico, que é caracterizado por um conjunto de tecidos, órgãos e células que desempenham a função de proteção contra patógenos e promovem o equilíbrio homeostático de saúde e o bem-estar (2–4). Essas respostas são didaticamente divididas em duas formas, imunidade inata e adaptativa, com componentes específicos de cada uma delas e componentes comuns a ambas.

A resposta inata é representada por barreiras físicas, fisiológicas e químicas, que atuam como defesas primárias do organismo. Além disso, um grupo de proteínas e células atuam no reconhecimento de patógenos e na defesa inicial. Entre essas células, podemos citar os neutrófilos, as células dendríticas (DC), os monocitos/macrófagos, as células *natural killer* (NK), e outros componentes do sistema imune que atuam juntamente com os fatores do sistema complemento (2). A resposta adaptativa é organizada em torno de duas classes de linfócitos especializados: células T, especificamente os linfócitos T CD4+ e T CD8+, e células B (2,5), que, através da diversidade de receptores, são capazes de direcionar a eliminação específica de patógenos e desenvolver memória imunológica de longa duração contra reinfecção (5). As células B desempenham um papel importante na resposta imune humoral, enquanto as células T estão mais envolvidas na resposta imune mediada por células (6).

Na imunidade inata, as células NK são capazes de induzir a morte celular por apoptose de células que estejam sob estresse oxidativo, infectadas por vírus, ou neoplásicas (7,8). Elas atuam na resposta imunológica por meio de citotoxicidade celular dependente ou não de anticorpos (7,8), liberando perforinas e granzimas de grânulos intracelulares após se ligarem com as células-alvo (9) e produzem citocinas que regulam a atividade dos linfócitos T e células dendríticas (DCs) (7). As NKs secretam grandes quantidades de citocinas pró-inflamatórias, como interferon-gama (IFN- γ) e fator de necrose tumoral-alfa (TNF- α) (7,8) e

imunossupressoras, como a interleucina (IL)-10, além de quimiocinas como CCL3 e CCL4 (7,10).

As células NK podem apresentar diferenças em função, fenótipo e localização tecidual (11,12). As NKs representam 5-15% dos linfócitos circulantes nos humanos e podem ser categorizados em dois subconjuntos fenotípicos e funcionais distintos com base na densidade de CD56 em sua superfície, as células CD56^{bright} apresentam maior expressão dessa proteína. Este subgrupo é caracterizado pela produção de altos níveis de citocinas em resposta à IL-2, enquanto as células CD56^{dim}, que compõem cerca de 90% das NK que são encontradas no sangue periférico, têm uma capacidade robusta de liberar perforinas e granzimas, quando está com CD16+ apresenta um importante papel na regulação do sistema imune via produção de citocinas, mas reduzida a atividade citolítica (9,13). Além de estar envolvida na morte celular, CD56^{dim} também atua na indução de respostas inflamatórias envolvendo recrutamento de outras células de defesa e promovendo a resistência celular contra infecção juntamente com a formação inicial das respostas imunes adaptativas (14).

As células NK apresentam um repertório de receptores em sua superfície, que influenciam a atividade dessas células. Entre eles, podemos citar os receptores KIR (do inglês, *killer cell immunoglobulin-like receptors*), LILRs (do inglês, *leukocyte immunoglobulin-like receptors*) e KLR (do inglês, *the killer cell lectin-like receptors*) (15). Esses receptores têm a capacidade de ativar ou inibir as células NK, por meio da interação com ligantes das células-alvo, em especial as moléculas do sistema de Antígeno Leucocitário Humano (HLA, do inglês, *human leukocyte antigen*), de classe I clássicas (HLA-A, -B, e -C) e não clássicas (HLA-E, -F e -G). As moléculas de HLA-I (HLA de classe I, clássicas) tem como função apresentar peptídeos na superfície celular de células nucleadas, seja peptídeos próprios ou alterados devido a transformações de células neoplásicas ou infectadas por patógenos intracelulares, como vírus (2) aos linfócitos T, um processo que inicia a resposta imune adaptativa. A interação receptor/ligante (KIR/HLA, por exemplo) e das citocinas liberadas, contribuem para a funcionalidade da célula NK (15–23). Por exemplo, os receptores inibitórios KIR2DL1 e KIR2DL3 reconhecem moléculas HLA-C (24); KIR2DL4 interage com HLA-G; KIR3DL1 reconhece o epítipo HLA-Bw4 no resíduo 77-83 presente em algumas moléculas de HLA-A e HLA-B, e KIR3DL2 reconhece as variantes HLA-A*3 e HLA-A*11 (25). O receptor ativador KIR2DS4, alvo deste

estudo, reconhece HLA-A e HLA-C (26). Na literatura, há descrita a interação entre KIR2DS4 e HLA-F somente pelos autores Goodridge et al., 2009 (27), demonstrando que é necessário mais estudos para comprovar essa interação entre as duas moléculas.

Por parte das células NK, há um mecanismo central de reconhecimento denominado como hipótese “missing self”, em que uma das funções destas células é reconhecer e eliminar células que apresentam ausência ou poucas quantidades de HLA-I na superfície (28), processo comum que ocorre principalmente em células neoplásicas e durante infecções virais (29). Além disso, as células alteradas podem superexpressar ligantes induzidos por estresse (“induced self-recognition”), aumentando a interação por receptores ativadores (30) presentes na NK. Portanto, os principais receptores expressos na superfície celular da NK contribuem para a sinalização negativa (inibitória) ou positiva (ativadora) durante a resposta imune mediada por célula NK (31). (Figura 1)

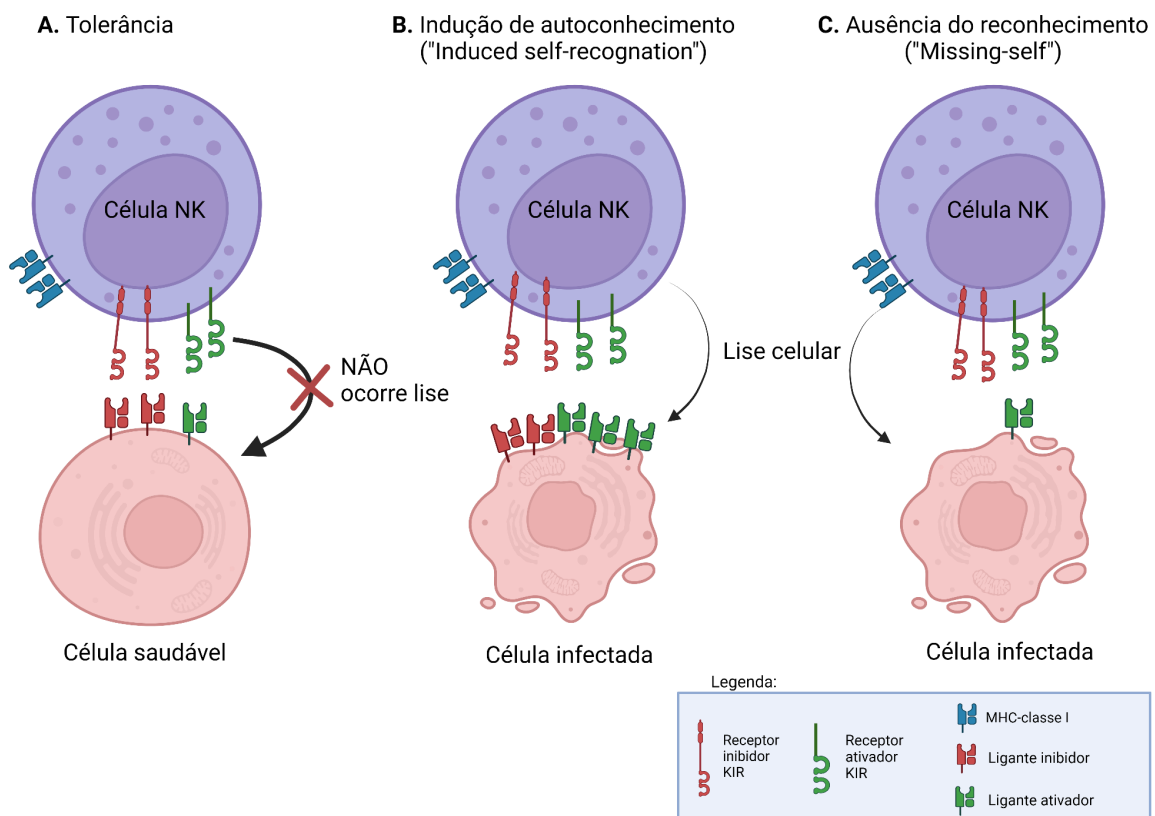


Figura 1: Mecanismos de regulação da atividade de células NK, de acordo com a interação de seus receptores com os ligantes da célula-alvo ou ausência desses ligantes. A) Maiores quantidades de ligantes inibitórios (em vermelho na célula-alvo) com receptores inibitórios de NK (em vermelho na célula NK) em relação aos ligantes com receptores ativadores(em verde na célula NK), resulta na tolerância da célula NK. B) Maior quantidade de ligantes ativadores(em verde), interagindo com

receptores ativadores (em verde na célula NK), comumente visto em células infectadas por alguns vírus e células tumorais, resulta em ativação da citotoxicidade da NK. C) Redução ou ausência de ligantes na superfície celular resulta na citotoxicidade celular. Criado no BioRender. Adaptado de Quatrini et al. 2021 (30).

Nos humanos, a família de genes *KIR* é composta por 15 genes responsáveis pela codificação dos receptores KIR, com a função de ativação (KIR2DS1-S5, KIR3DS1), inibição (KIR2DL1-L5, 3DL1-L3), ou dupla função, como o KIR2DL4 (32), expressos na superfície das células NK e em um subconjunto de linfócitos T ativados ou de memória, e 2 pseudogenes (*KIR2DP1* e *KIR3DP1*). Estes genes estão presentes no cromossomos 19 (19q13.4), no Complexo de Receptores Leucocitários (LRC, do inglês, *leukocyte receptor complex*), caracterizado por ser altamente diversificado e em rápida evolução (32), que compreende um grande cluster de receptores de superfície, juntamente com receptores LILR, leucócitos associados com domínio com receptor-I (LAIR, do inglês, *leukocyte-associated Ig-like receptor-1*), receptor associado com osteoclasto (OSCAR, do inglês, *osteoclast-associated receptor*), NKp46, glicoproteína VI (GPVI, do inglês, *glycoprotein IV*) e FcR α RI (receptor Fc de alta afinidade I das imunoglobulinas G, IgG), cujo ligantes são tão diversos quanto imunoglobulinas, colágenos e hemaglutininas virais. Na região estendida do LRC, há um conjunto adicional de genes que surgiram por múltiplas e antigas ondas de duplicação, Siglecs, glicoproteína gravidez específica (PSG, do inglês, *pregnancy-specific glycoprotein*) e molécula de adesão celular relacionada com antígeno de carcinoembrionário (CEACAM, do inglês, *carcinoembryonic antigen-related cell adhesion molecule*), além de estarem dentro da região estendida também o DAP10 e DAP12 (moléculas adaptadoras), Axl e CD79A. (33–41) (Figura 2). As famílias gênicas codificadas na região do LRC são organizadas em haplótipos. Em cada haplótipo, pode haver uma combinação única de genes que codificam receptores ativadores e inibidores (41).

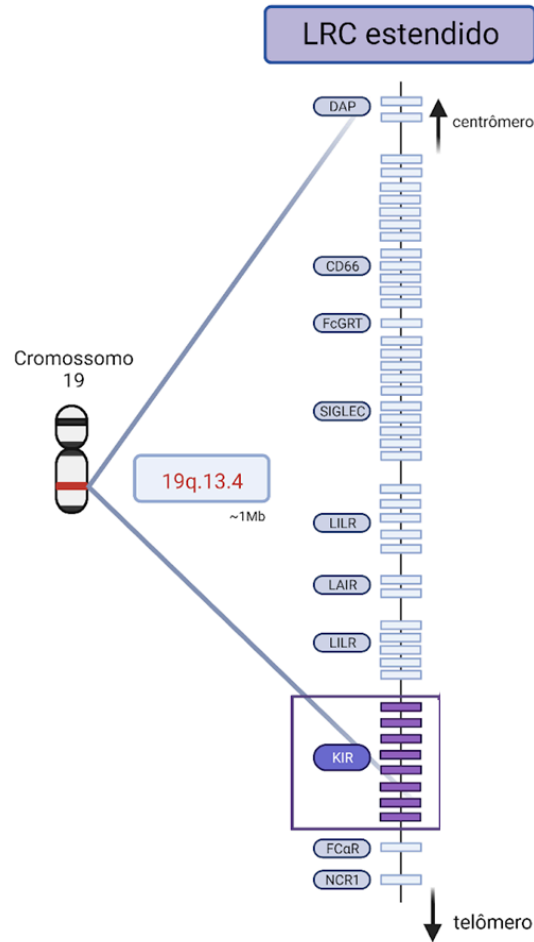


Figura 2: Localização dos genes *KIR* no cromossomo 19. No cromossomo 19, especificamente em 19q.13.4 (em roxo escuro), estão localizados diversos genes dentro do Complexo de Receptores Leucocitários (LRC), incluindo-se os genes *KIR*. Criado pelo BioRender. Adaptado de: Dębska-Zielkowska et al., 2021 (33) e IPD-KIR (42).

A ordem dos genes *KIR* dentro do complexo KIR de um cromossomo 19 define um haplótipo de *KIR* (Figura 3). Os haplótipos podem variar tanto do ponto de vista da ordem dos genes, como também da presença ou ausência de alguns dos genes.

Esses haplótipos são didaticamente divididos em dois grupos principais, A e B, diferenciando-se pela quantidade de genes que os compõem e pelos alelos presentes em cada um dos genes *KIR* presentes (43). O haplótipo A apresenta uma maior quantidade de genes que codificam KIR inibidores (44) e apenas um gene para receptor ativador, o *KIR2DS4* (43). O haplótipo B contém maior variedade de combinações, com um ou mais genes ativadores, como *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS4*, *KIR3DS1* e *KIR2DL4* (43,45,46).

Os haplótipos também são designados (ou subdivididos) em haplótipos centroméricos (cen) e/ou teloméricos (tel). A região centromérica do haplótipo A (Cen-A) é caracterizada pela presença exclusiva de *KIR2DL1* e *KIR2DL3*, sem nenhum outro gene associado ao haplótipo A. Enquanto a região centromérica B (Cen-B) foi delimitada pela presença de qualquer gene do haplótipo centromérica B (*KIR2DL2* e *KIR2DS2*, e/ou *KIR2DL5B* e *KIR2DS3/KIR2DS5*). Na região telomérica A (Tel-A), inclui-se o gene *KIR2DS4*, potencialmente codificando o único receptor ativador do haplótipo A, e *KIR3DL1* (47). E a região telomérica B (Tel-B) é definida pela presença de qualquer gene do haplótipo B centrômero (*KIR3DS1* e/ou *KIR2DS1*, por exemplo) (47).

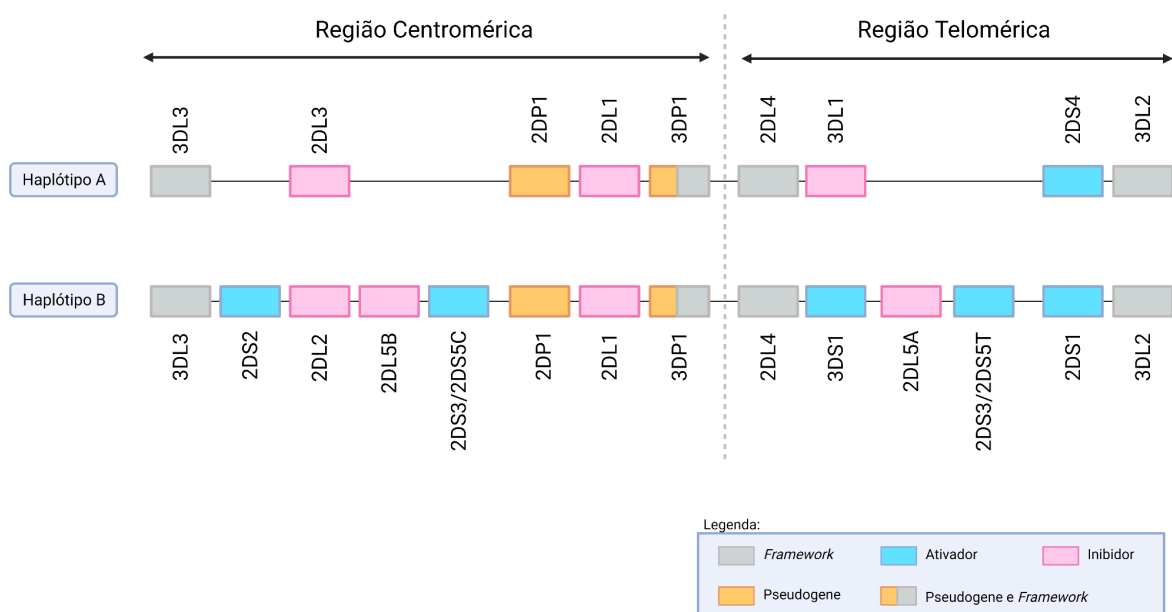


Figura 3: Haplótipos A e B dos genes *KIR*. Genes “framework” (em cinza); inibidores (em rosa); ativadores (em azul) e pseudogenes (em laranja), apenas *KIR3DP1* está presente como pseudogene e “framework” (em laranja e cinza). Criado pelo BioRender. Adaptado de: TUKWASIBWE et al., 2021 (47).

Os genes *KIR* são bastante polimórficos, tanto do ponto de vista estrutural (variação no número de cópias), como de variação alélica de cada gene (48). Os genes *KIR2DL4*, *KIR3DL3*, *KIR3DL2* e o pseudogene *KIR3DP1* geralmente estão presentes em todos os haplótipos, considerados como genes “framework” por não apresentarem variação no número de cópias. Atualmente, apenas o *KIR3DL3* parece não apresentar polimorfismo de número de cópias, e todos os demais são variáveis neste sentido. A complexidade genética dos *KIR* aumentam os riscos de

erros de genotipagem e má interpretações de associações genéticas, particularmente em estudos com pouca quantidade amostral, mostrando que é necessário desenvolver métodos analíticos aprimorados e dedicados para estes genes (48). Há ao menos 1617 alelos descritos para os genes *KIR* no banco de dados IPD-KIR, que é o banco de dados oficial de sequências conhecidas de genes *KIR* (42). Os receptores KIR têm uma alta similaridade de sequência, particularmente nos domínios extracelulares. Porém, há sinais opostos em seus domínios citoplasmáticos (41).

Cada receptor KIR é classificado de acordo com: (i) a quantidade de domínios extracelulares que codificam (KIR2D, para dois domínios, ou KIR3D, para 3 domínios); (ii) o tamanho da cauda citoplasmática intracelular, que pode ser curta (S, de *short*) e está relacionada com receptores ativadores, ou longa (L, de *long*), que está relacionada com receptores inibidores; (iii) presença ou ausência de domínio citoplasmático do imunoreceptor ativador baseado em tirosina (ITAM, do inglês *immunoreceptor tyrosine-based activation motif*) (49) que interagem com proteínas adaptadoras (como DAP12) (50) ou do imunoreceptor inibidor baseado em tirosina (ITIM, do inglês *immunoreceptor tyrosine-based inhibitory motifs*), definindo a função ativadora ou inibitória do receptor KIR, respectivamente (38,51) (Figura 4). A única exceção para este esquema de classificação é para o receptor KIR2DL4, que, apesar de possuir uma cauda citoplasmática longa (L), pode atuar tanto na inibição quanto na ativação da NK (52), o que por sua vez depende do ligante que está interagindo com KIR2DL4.

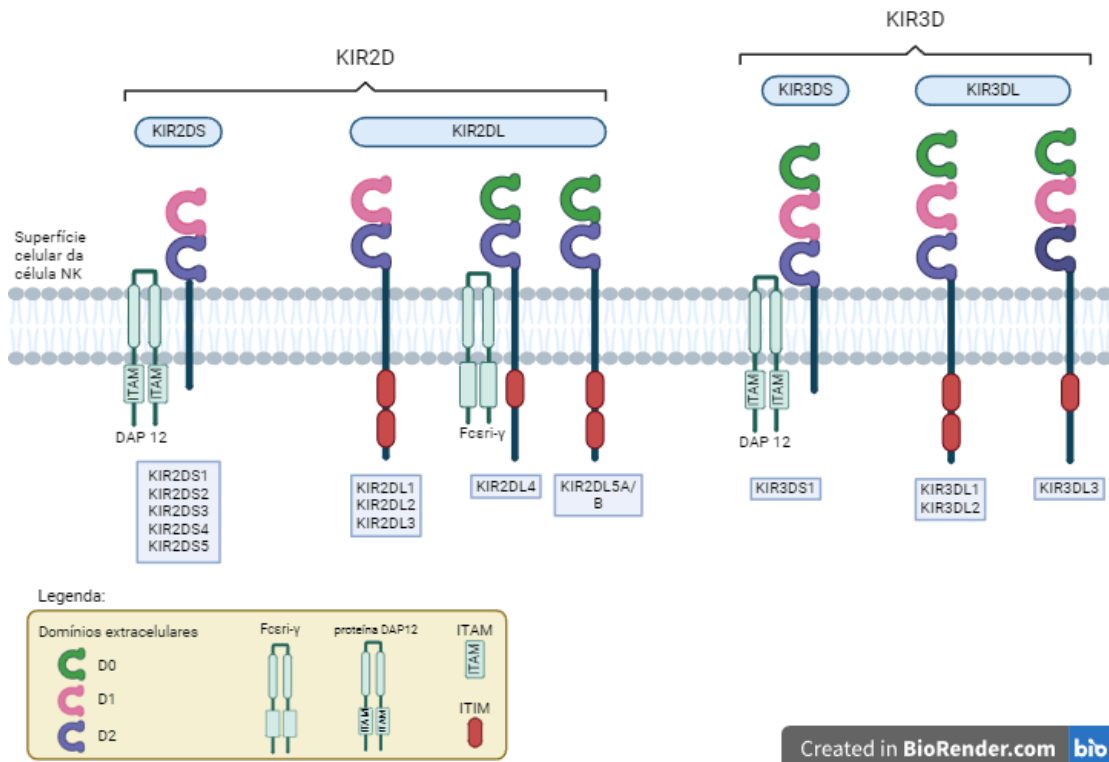


Figura 4: Estrutura dos receptores KIR. Características estruturais de 1, 2 e/ou 3 domínios, em roxo, rosa e verde, respectivamente. Adaptadores DAP12 associados com os receptores ativadores (em verde claro), ITIM associados com os inibidores (em vermelho). Criado pelo BioRender.com. Adaptado de: IPD-KIR (42).

Os receptores KIR inibidores que interagem com as moléculas HLA classe I apresentam em sua cauda citoplasmática motivos inibidores imunorreceptores baseados em tirosina, que permitem recrutar e ativar fosfatases SHP-1 (do inglês, *SH-containing tyrosine phosphatase*) e SHP-2 (15,33). Durante a gestação, por exemplo, o receptor KIR2DL1 confere um importante sinal de inibição para as células NK ao interagir com o HLA-C2 presente no trofoblasto fetal na decídua (52,53).

Devido ao sucesso da nomenclatura utilizada para os alelos de *HLA*, os pesquisadores implementaram o mesmo padrão de nomenclatura para os genes *KIR*. Inicia-se com o nome do gene, depois um asterisco (*) como um separador entre o gene e a designação numérica (38). Os três primeiros dígitos numéricos indicam diferentes versões da proteína por conta da presença de variações não-sinônimas em éxons; os dois dígitos seguintes são usados para distinguir alelos que apresentam sequências de nucleotídeos diferentes em éxons, mas que geram a mesma proteína representada pelos 3 primeiros dígitos; os dois últimos dígitos são

utilizados para indicar alelos com sequências diferentes de nucleotídeos em íntrons, região promotora ou outra região não-traduzida (38). (Figura 5)

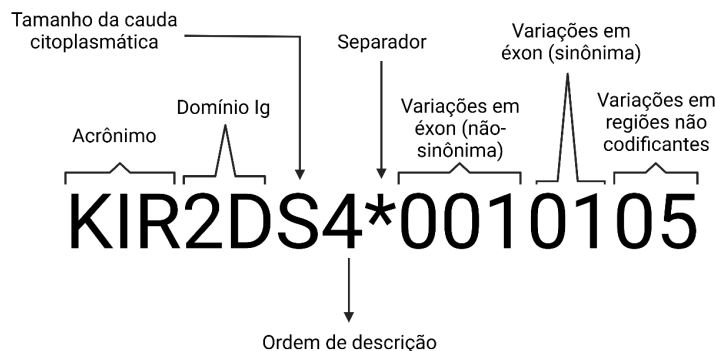


Figura 5: Nomenclatura utilizada para os genes *KIR*. Adaptado de: IPD-KIR (42).

2. Receptores KIR e associação com doenças

O papel principal das células do sistema imunológico é proteger o nosso organismo contra tumores, infecções e alérgenos (33). As células NK apresentam variações em suas propriedades funcionais dependendo da presença ou ausência de expressão dos seus receptores, principalmente em termos de degranulação (54), no balanço de sinais para regulação da maturação e ativação (19).

A família de receptores KIR, reconhecida como a mais polimórfica entre todos os receptores, exibe variações em sua sequência gênica que podem influenciar sua expressão, afinidade com os ligantes e funcionalidade. No entanto, a pesquisa sobre essas variações ainda é bastante limitada, assim como quantidade amostral e diferentes estudos com resultados conflitantes (55), o que compromete nossa capacidade de compreender muitas contribuições clinicamente relevantes, incluindo a interação entre KIR/HLA na saúde humana (19) e a diversidade genética entre populações.

Os receptores KIR e também a interação entre KIR/HLA contribuem para um papel fundamental na susceptibilidade ou proteção a doenças autoimunes, patógenos (56), e durante a gestação, apresentando uma variação entre indivíduos e populações (8). Essa relação KIR/HLA é muitas vezes negligenciada em estudos de associação genômica ampla (GWAS, do inglês, *genome-wide association study*) com doenças e infecções, em parte devido às dificuldades metodológicas de se avaliar tanto os genes *KIR* quanto os genes *HLA*, pela complexidade dos loci gênicos, pela alta variabilidade e similaridade entre as sequências. Em plataformas

de GWAS, há uma baixíssima representação de variantes dos genes *KIR* e para avaliar a variabilidade dos genes *KIR* e *HLA* em doenças, é necessário estratégias de sequenciamento direcionadas para esses genes *KIR* com uso de algoritmos específicos, ou extrair os dados de *KIR* e *HLA* em estudos de exomas ou genomas completos (57) com alta resolução de genotipagem (58), utilizando ferramentas de análise direcionadas para esses genes. Durante as duas últimas décadas, os métodos predominantes se concentravam em apenas a presença/ausência destes genes *KIR* para cada indivíduo, assim como nas análises de leucemia mielóide aguda e transplantes por exemplo (58).

Os genes *KIR* podem impactar o sucesso de transplantes. Nos casos de pós-transplantes, seja de órgão sólido ou de célula-tronco hematopoiética, há uma imunossupressão com uso de drogas para melhorar a taxa de sobrevivência do enxerto. Porém, o paciente fica suscetível a complicações por infecções (59), principalmente virais. Há uma associação significativa entre a presença de *KIR2DS4* funcional (*KIR2DS4-full*) nos doadores com haplótipo AA ou de aumento de *KIR2DL1/3DL1* com doença aguda do enxerto versus o hospedeiro após transplante alogênico de células hematopoiéticas (33,60). Geralmente, genótipos que possuem mais receptores *KIR* ativadores do que inibidores são benéficos em contextos de infecção por vírus, especialmente por imunodeficiência adquirida (HIV) e hepatite C (HCV) (61), pois os vírus interferem com a expressão de moléculas *HLA* como uma forma de driblar a resposta citotóxica por célula T (33).

Ao analisar casos de COVID-19 na Espanha, observou-se uma menor frequência de *KIR3DL1* em pacientes com *HLA-B*015* quando comparado a um controle populacional (66.7% vs 100%) (62). Além disso, *KIR3DS1/HLA-B*15:01* estava associado com casos leves/moderados da COVID-19, ressaltando o papel decisivo das células NKs na imunopatogênese da doença (62). Os genes *KIR2DL5* e *KIR2DS3* foram significativamente associados à proteção contra o vírus SARS-CoV-2 (63) em pacientes da cidade de Meca e Riad, Arábia Saudita.

Estudo envolvendo pacientes japoneses com colite ulcerosa mostrou que a presença de *KIR2DS3* aumenta o risco de desenvolver essa doença (23% nos pacientes vs 14% no grupo controle) e *KIR2DS4* diminui o risco (88% vs 95%, respectivamente) (57). Na dermatite atópica, há menor probabilidade de remissão da doença quando há o haplótipo telomérico *KIR2DL4*00102 ~ KIR2DS4*00101 ~ KIR3DL2*00201*, e mais significativo ainda quando esses alelos de *KIR2DS4* e

KIR2DL4 estavam acompanhados de HLA-C*04:01 (64). A frequência do KIR2DS4 funcional em pacientes com arterite de Takayasu é menor em relação ao grupo controle saudável (38% vs 82%), demonstrando uma significativa associação com a proteção para a doença ($p < 0.001$) (65).

Em pacientes de Kayseri (Turquia) com doença celíaca (CD) e pacientes com CD e diabetes mellitus (DM) coexistentes, foram observados dois receptores teloméricos ativadores como fatores de risco, KIR2DS5 e KIR3DS1 (66). Em uma meta-análise de estudos já descritos com populações do México, Brasil, Irã, Norte da Índia, Japão, Polônia, Irlanda do Norte, Taiwan e Europa Ocidental sobre artrite reumatóide (RA), os genes *KIR2DL3*, *KIR2DL5* e *KIR3DL3* demonstraram associação com baixo risco da doença (67).

Estas pesquisas demonstraram a importância dos receptores KIR na proteção e susceptibilidade em diversos processos patológicos, como infecções virais, doenças autoimunes e crônicas. Além disso, esses estudos podem contribuir para um melhor entendimento do mecanismo das células NK relacionadas com essa diversidade (68).

3. Gene *KIR2DS4*

O gene *KIR2DS4* codifica um dos receptores ativadores de células NK. Esse gene está localizado na região telomérica do complexo KIR e está presente no haplótipo A (*Tel-A*). Atualmente, há 41 alelos descritos, que codificam 22 proteínas distintas (diferentes sequências de aminoácidos), de acordo com o banco de dados IPD-KIR (42).

Um polimorfismo importante do gene *KIR2DS4* é uma deleção de 22 pares de bases no éxon 5 (39,42,56). Esta deleção faz com que haja um deslocamento no quadro de leitura, gerando um novo trecho com 64 aminoácidos diferentes (42,69) e um códon de parada prematuro localizado no final do éxon 7 (42,46,69), alterando a maior parte da sequência proteica do domínio D2 (63 de 92 resíduos) e toda a sequência da haste (70). Essa versão truncada é bem menor que a versão completa e funcional (42). Se a molécula truncada for expressa, provavelmente ela não se ancora na membrana celular e seria secretada, formando potencialmente uma molécula de KIR2DS4 solúvel (52,59,61). Não foram observados, até então, *splicing* alternativos que pudessem corrigir a deleção das 22 bases e re-estabelecer o quadro de leitura original (59).

Portanto, KIR2DS4 apresenta duas versões distintas, uma completa funcional (KIR2DS4-full) e uma truncada não funcional (KIR2DS4-del) (71). Os indivíduos homocigotos para o haplótipo A e com *KIR2DS4* não funcional nas duas cópias do gene apresentam somente genes *KIR* inibidores (60).

Além do polimorfismo de 22pb, mutações não sinônimas modificam alguns aminoácidos na cadeia peptídica, gerando diferentes sequências, seja na forma *full* ou *del*. Entre essas diferentes versões, podemos citar, para a KIR2DS4-full, as versões denominadas 2DS4*001, 2DS4*011, 2DS4*014, 2DS4*015, 2DS4*016, 2DS4*017, 2DS4*019, 2DS4*020, 2DS4*021, 2DS4*022, 2DS4*023 e 2DS4*024. Para a KIR2DS4-del, podemos citar 2DS4*003, 2DS4*004, 2DS4*006, 2DS4*007, 2DS4*008, 2DS4*009, 2DS4*010, 2DS4*012, 2DS4*013 e 2DS4*018 (42).

A função da proteína solúvel advinda da liberação da versão truncada (KIR2DS4-del) ainda não está completamente elucidada. Há estudos que buscam potenciais associações com doenças, além de especulações de que essa versão atue como um receptor para outro ligante, ainda não identificado, ou sendo uma armadilha para eliminar as proteínas solúveis do MHC, que podem interferir na função das células NK (71,72). Uma análise de PCR em tempo real (qRT-PCR) com confirmação por dot blot de RT-PCR demonstrou que amostras que tinham os alelos com deleção das 22 bases (KIR2DS4*003, *004 e *006) mostraram uma expressão mais baixa do gene *KIR2DS4* que os alelos sem a deleção (73).

Foram identificados alguns transcritos alternativos para KIR2DS4-full através da análise de 15 indivíduos por sequenciamento do tipo PacBio (74). Dentre os eventos de *splicing*, um possui uma deleção de 294b em parte do éxon 4 e 5, resultando na exclusão do final do domínio D1 e no começo do D2. Outro apresenta um éxon a mais, que normalmente é removido do RNAm porém mantido neste transcrito, resultando na inclusão de 54pb após o éxon 5. Como consequência, um *stop-códon* prematuro é gerado. Foi observada uma deleção de 51pb no éxon 6 que ocasiona no pulo de éxon, não sendo traduzida a região da haste da molécula (74). Os *splicing* alternativos contribuem para a complexidade dos receptores *KIR*, resultando em possíveis receptores com modificações estruturais, função e/ou perfil de expressão distintos (74).

O principal ligante de KIR2DS4 até então descrito é a molécula HLA-C, tanto o subconjunto C1 quanto C2. Portanto, KIR2DS4 se difere dos outros receptores KIRs, que são restritos a apenas um desses grupos (26). A ligação também

acontece com o HLA-A (A*11:01 e A*11:02) e HLA-F de conformação aberta (26,27). Como a funcionalidade das formas truncadas do KIR2DS4 ainda permanecem incertas, fizeram uma fusão do KIR2DS4*003 (um dos alelos de KIR2DS4-del) com beads revestidas com MHC de classe I, mas não obtiveram a ligação receptor/ligante, demonstrando que os alelos de KIR2DS4-del não são receptores de MHC de classe I (26).

Tanto o gene humano *KIR2DS4* quanto o ortólogo *Pt-2DS4* encontrado em chimpanzés, codificam moléculas que similarmente reconhecem HLA-A*11 e HLA-C*16:01, indicando que essas características evoluíram antes da separação dos ancestrais humanos e chimpanzés, resultando na conservação dessas interações em ambas as linhagens. No entanto, uma vez que apenas Pt-2DS4 se liga ao MHC-C*08:01, as especificidades entre KIR2DS4 e HLA-C são divergentes entre humanos e chimpanzés (26).

É descrito na literatura que o resíduo W na posição 14 do HLA-Cw4 (sorotipo de antígeno leucocitário humano) é importante para a ligação HLA-Cw4/KIR2DS4, enquanto que o KIR2DS4 não faz ligação com HLA-Cw6 (69). A ligação entre KIR2DS4 com HLA-Cw4 é fraca quando comparado ao KIR2DL1, receptor inibidor (69).

Segundo Augusto, 2006 (43), a perda da função de *KIR2DS4* devido à deleção no éxon 5 pode fazer com que alguns indivíduos possam ser mais suscetíveis à determinadas doenças, principalmente quando não há nenhum outro gene *KIR* ativador em seu repertório, como os indivíduos homocigotos para o haplótipo A (A/A). Todavia, existem outros receptores ativadores nas células NK que também estão envolvidos na resposta citotóxica, semelhantes à lectina (43).

Em transplantes, pacientes diagnosticados com diferentes tipos de doenças linfóides com compatibilidade total para os genes do MHC de classe I com o doador (HLA 10/10), apresentaram baixa incidência de mortalidade sem recidiva, taxas mais elevadas de sobrevida e sobrevida livre da doença, quando receberam o transplante de célula-tronco hematopoiética de um doador homocigoto para haplótipo A com KIR2DS4-del (AA/KIR2DS4-del) em comparação com aqueles que recebem um enxerto de um doador homocigoto para KIR2DS4-full (AA/KIR2DS4-full) (75). Foi visto também um aumento no risco de desenvolver doença aguda do enxerto contra o hospedeiro (aGVHD) quando o doador é AA/KIR2DS4-full (60).

Estudos realizados com população espanhola mostraram uma frequência maior de KIR2DS4-full em pacientes sintomáticos com COVID-19 quando comparado ao controle (61.7% vs 36.2%, respectivamente) e também com doença mais grave e necessidade de ventilação mecânica (62). A versão funcional do KIR2DS4 com KIR2DS1 demonstrou uma proteção adicional contra distúrbios na gravidez que podem impactar na placentação, a provável interação biológica entre o KIR2DS1 e o KIR2DS4, baseada em suas funções semelhantes e na sobreposição de seus ligantes, é provavelmente responsável pela epistasia detectada em nível estatístico (76).

Frequência de KIR2DS4-full e KIR2DS4-del em diferentes populações

A análise de alelos KIR utilizando técnicas de sequenciamento de segunda geração se faz necessária para entender a evolução desses receptores e estabelecer uma base para estudos funcionais para uma melhor percepção dos mecanismos das doenças (77). O método mais frequentemente utilizado para explorar a diversidade genética dos *KIR* ainda é a PCR-SSP (do inglês, *polymerase-chain reaction with sequence-specific primers*) (78), que limita o estudo à busca de polimorfismos já conhecidos. Descrito no IPD-KIR (42), os alelos do KIR2DS4 podem ser distribuídos de acordo com o genótipo encontrado, separando nas duas versões encontradas na tabela 1.

Tabela 1: A distribuição dos alelos de acordo com os genótipos KIR2DS4-full e KIR2DS4-del. Baseado nas sequências do IPD-KIR (42).

Genótipo	Alelos
KIR2DS4-full	KIR2DS4*00101; *00102; *00103; *00104; *00105; *00106; *01101; *01102; *014; *015; *016; *017; *018; *019; *020; *021; *022; *023 e *024
KIR2DS4-del	KIR2DS4*00301; *00401; *00601; *007; *008; *009; *010; *012; *013 e *018

Em uma pesquisa com amostras da Irlanda do Norte (n = 100), foi constatado que o alelo mais frequente encontrado foi KIR2DS4*003 (64%), seguido pelo KIR2DS4*006 (33%), KIR2DS4*00101 (36%) e KIR2DS4*005 (7%) (79). Em relação à associação com COVID-19, o KIR2DS4*00101 foi associado com maior

susceptibilidade à internação, registrado 20.7% em hospitalizados e 14.9% em não-hospitalizados, independente da ancestralidade genética. Por outro lado, KIR2DS4*00601 demonstrou relacionado ao baixo risco de internação (18.4% vs 15%) (80). Em análise de uma combinação de três populações iranianas através de sequenciamento de segunda geração, revelaram diferentes frequências alélicas para KIR2DS4: *003 (40%), *null (ausência do gene, cerca de 20%), *001 (aproximadamente 15%), *006 (10%), *010 (em torno de 7%) e *004 (com 5%) (81). Ao focar em residentes dos Estados Unidos com ascendência europeia, os alelos encontrados foram KIR2DS4*00101 (27%), KIR2DS4*null (22%), KIR2DS4*00301 (14,6%), KIR2DS4*010 (14,3%), KIR2DS4*00601 (12%), KIR2DS4*00401 (3,5%), *unresolved/new (2,8%), KIR2DS4*00104 (1,9%), KIR2DS4*014 e *00102 (frequência menor que 1%) (77).

Sobre o genótipo de KIR2DS4, a variante KIR2DS4-del é menos frequente nos indígenas, tendo sua presença em 3,4% em Kaingang (n = 37) e 11,3% em Guarani (n = 32). Em análise de pacientes com ovários policísticos, houve uma maior frequência do genótipo KIR2DS4-del/del ($P_c = 0.0002$, 54%) nas pacientes quando comparado com o controle (26%), e uma baixa frequência de KIR2DS4-full (43% nos pacientes e 70% no controle) (82). E irlandeses do norte (n = 100) apresentaram a predominância de KIR2DS4-del com uma frequência de 82%, e -full com 36% (79). A princípio, as análises de *KIR2DS4* foram realizadas, em sua maioria, em populações isoladas. O intuito deste projeto é obter informações sobre a diversidade deste gene utilizando banco de dados públicos.

4. Dificuldades metodológicas para análise dos genes *KIR*

Os genes *KIR* apresentam um elevado grau de similaridade de sequência, compartilhando cerca de 85-98% de similaridade. Há, inclusive semelhanças entre os alelos de genes distintos. Essa similaridade de sequência dificulta o alinhamento das sequências curtas produzidas em sequenciamentos de segunda geração, com vários *reads* de um gene alinhado em outro. Por consequência, temos um elevado número de variantes falso-positivas e falso-negativas (25,83).

Além disso, os genes *KIR* são altamente polimórficos, com dezenas de diferentes alelos (sequências distintas) para cada gene, e variação no número de cópias em nível individual (58,84). Os erros de alinhamento apresentados

anteriormente prejudicam a identificação desses alelos e prejudicam a análise do número de cópias, que, em geral, baseia-se no número de *reads* alinhados em cada gene.

Portanto, um dos desafios na análise bioinformática de sequenciamento dos genes *KIR* está justamente na obtenção de alinhamentos menos enviesados, que permitam captar com confiabilidade todos as variantes (SNVs e indels) ao longo de cada gene, além do número de cópias e variantes estruturais (85).

Recentemente, os avanços no sequenciamento de nova geração (NGS, do inglês, *next-generation sequence*) e de terceira geração (TGS, do inglês, *third-generation sequence*) permitiram uma análise mais abrangente dos genes *KIR*, englobando toda a extensão dos genes e haplótipos entre genes (58). O sequenciamento com sequências curtas (segunda geração) proporcionou uma melhor compreensão da estrutura de cada gene e do grau de variação observado. Além disso, embora ainda de forma inicial as plataformas de sequenciamento de sequências longas (terceira geração), como Oxford Nanopore, fornecem dados importantes acerca da estrutura de haplótipos e presença de variantes estruturais. Embora mais utilizada, a desvantagem de usar a tecnologia de sequenciamento de segunda geração reside principalmente nos erros de mapeamento/alinhamento (58). Um exemplo está no gene *KIR3DP1*, que geralmente recebe *reads* originadas de *KIR3DL3*, *KIR3DL2*, *KIR2DP1* e *KIR2DL4* (58).

Nos últimos três décadas, os métodos de genotipagem dos genes *KIR* têm se concentrado principalmente em identificar se os genes estão presentes ou ausentes, o que, em certa medida, auxilia na parcial determinação dos haplótipos individuais. No entanto, essas técnicas fornecem informações bastante limitadas sobre a extensão e a importância da variação nos níveis dos alelos (58). Obter uma genotipagem de alta resolução é importante para os inúmeros estudos que estudam a variação alélica, a história evolutiva dos genes *KIR* e o impacto da diversidade desses genes nas respostas imunológicas, como ocorre em transplante de órgãos sólidos e em transplantes de células-tronco hematopoéticas (HSCT, do inglês, *hematopoietic stem cell transplantation*) (58).

Já foram desenvolvidos métodos computacionais para análise dos genes *KIR* em dados de sequenciamento de segunda geração, como PING (86), T1K (87), KIR-IMP (88), e KPI (89,90). O T1K é compatível com dados de sequenciamento de

DNA e RNA, incluindo sequenciamento completo do genoma (WGS, do inglês, *whole genome sequencing*) e quando se avalia somente os éxons (exoma). O KIR*IMP imputa o número de cópias dos genes a partir de SNPs selecionados, e o KPI prediz os haplótipos usando sondas virtuais, mas nenhum dos dois são capazes de genotipar os alelos (91). A última versão do PING foi desenvolvida para analisar o sequenciamento de genoma completo (WGS), mas não é compatível para a análise do exoma (92). Além disso, PING exige computadores de alta performance quando se está avaliando um grande número de amostras e não produz dados de variantes (SNPs e indels) no contexto do genoma de referência humano. PING é, de longe, o método mais utilizado para avaliar os genes *KIR* em dados de sequenciamento de segunda geração.

O Kir-mapper, método apresentado nesse estudo, foi concebido para processar WGS, exomas e dados de sequenciamento obtidos por enriquecimento de genes *KIR* por sondas. Além disso, o kir-mapper foi desenvolvido para produzir alinhamentos mais confiáveis utilizando várias sequências conhecidas como referência para os alinhamentos e produzir genotipagens de SNVs e indels com alta confiabilidade, no contexto do genoma de referência hg38. Isso permite obter o número de cópias e os alelos existentes em cada gene, convertendo a informação dos SNPs em sequências e haplótipos.

PING e kir-mapper utilizam metodologias semelhantes para detectar o número de cópias de cada gene. Normalmente, o KIR3DL3 é utilizado como referência para determinar a quantidade de cópias em cada gene. Esse é o padrão utilizado pelo PING. No entanto, em casos raros em que há uma variação no número de cópias em KIR3DL3 (detectamos ao menos 3 casos como esse em amostras estudadas no laboratório), o kir-mapper permite a seleção de outros genes de referência, como HLA-E e HLA-G. Embora aqui propomos um novo método para análise dos genes *KIR*, é importante destacar a dificuldade metodológica dessas análises. Os diferentes métodos propostos apresentam pró e contras, que devem ser pesados na hora da escolha do método e considerados quando os dados são comparados. Por exemplo, a definição do número de cópias para alguns genes é, no momento, mais eficiente no PING. Por outro lado, kir-mapper é mais eficiente na detecção dos alelos, pois possui em seu *core* um sistema de faseamento entre os

SNPs e produz resultados de SNPs de mais fácil interpretação, pois estão no contexto do genoma de referência hg38.

Caracterização dos bancos de dados genômicos públicos: Projeto 1000

Genomas utilizado no estudo

O projeto 1000 Genomas criou um catálogo de variações comuns em humanos. Esse projeto conta com 3.202 amostras de indivíduos autodeclarados como saudáveis de 26 populações em 5 regiões biogeográficas. Desses, 2.504 são de indivíduos não-relacionados. Estas amostras possuem sequenciamento do genoma completo (WGS) de alta cobertura (30x), utilizando a técnica Illumina Nova Seq 6000, leitura de 150 pares de base, e a preparação da biblioteca foi por PCR. O objetivo principal deste projeto foi encontrar variantes genéticas comuns com uma frequência de pelo menos 1% na população estudada, variantes raras, variantes estruturais e fornecer informações sobre formação haplotípica nas populações de estudo. Caracterizaram cerca de 88 milhões de variantes (SNPs, InDels e variantes estruturais), faseadas em alta qualidade (93). Os dados brutos de sequenciamento foram alinhados ao genoma de referência humano GRCh38 (hg38) utilizando a ferramenta BWA (BWA-MEM 0.7.15). Todos os dados gerados a partir do projeto estão disponíveis publicamente, incluindo arquivos BAM (os sequenciamentos alinhados ao genoma de referência), VCFs (as variantes, SNPs e indels, chamados pelo HaplotypeCaller do GATK) e anotações funcionais (94) para que possam ser analisadas e estudadas por outros projetos de pesquisa. Todas as amostras estão disponíveis também em linhagens celulares de linfoblastóides (LCLs) e LCL derivados do DNA do repositório Coriell Cell (<https://catalog.coriell.org/>).

Nesta dissertação, foram selecionadas 5 populações, 1 por região biogeográfica do projeto 1000 Genomas: Indiano Telugu (ITU) (ancestralidade do Sul Asiático); Inglaterra e Escócia (GBR) (ancestralidade Européia); Japoneses (JPT) (ancestralidade do Leste Asiático); Colombianos de Medellín (CLM) (ancestralidade Americana); Yoruba (YRI) (ancestralidade Africana), totalizando em 490 amostras.

JUSTIFICATIVA

A diversidade do gene *KIR2DS4* é provavelmente subestimada, visto que há uma ausência de estudos em populações miscigenadas, como a brasileira, e outras populações não europeias, o que gera incerteza sobre a amplitude da diversidade desse gene. O estudo da diversidade de *KIR2DS4* em um conjunto amplo de amostras, de diferentes partes do mundo, irá contribuir para um melhor entendimento da funcionalidade e da história evolutiva desse receptor. A determinação das frequências alélicas, genotípicas e distribuição de *KIR2DS4*-full, -del e -null, e a identificação dos haplótipos associados, são passos fundamentais para um entendimento da função desse gene. Além disso, os métodos atuais para análise de genes *KIR* em dados de sequenciamento de segunda geração, embora úteis para a caracterização desses genes em pequenos tamanhos amostrais, dificultam a análise de grandes tamanhos amostrais, especialmente na ausência de servidores bastante potentes.

Este estudo contempla uma parte do desenvolvimento de um método computacional que se adequa à realidade brasileira, com capacidade para analisar grandes números de amostras em computadores de pequeno a médio porte. Esse método foi criado no âmbito do laboratório GeMBio pelo orientador do projeto e curado (utilizado repetidas vezes, com avaliação manual dos resultados) pelos alunos de pós-graduação. Neste estudo, mostraremos esse método com foco no gene *KIR2DS4* e exploramos a diversidade observada em amostras provenientes do projeto 1000 Genomas. Neste primeiro momento, exploraremos o número de cópias e a frequência dos alelos observados que caracterizam uma proteína diferente (primeiros três dígitos) .

OBJETIVOS

Objetivo Geral

O objetivo deste trabalho é avaliar a diversidade genética de *KIR2DS4* observada em amostras de diferentes regiões biogeográficas, considerando o número de cópias do gene e os diferentes alelos observados, e testar uma nova metodologia de análise em dados de sequenciamento de segunda geração.

Objetivos Específicos

- Consolidar uma metodologia para avaliação da variabilidade do gene *KIR2DS4* a partir de dados de sequenciamento de nova geração (NGS);
- Demonstrar como a metodologia desenvolvida no laboratório pode ser complementar a outras ferramentas já utilizadas para análise dos genes *KIR*;
- Detectar os sítios de variação encontrados para o gene *KIR2DS4* em amostras de diferentes populações;
- Detectar variação do número de cópias (presença/ausência e quantidade);
- Estudar a diversidade do gene *KIR2DS4* em diferentes populações,
- Analisar a distribuição das diferentes versões de *KIR2DS4*, com foco entre a versão funcional, não-funcional e a nula (quando o gene não está presente).

REFERÊNCIAS

1. Morimoto K, Nakajima K. Role of the Immune System in the Development of the Central Nervous System. *Front Neurosci.* 2019;13:916.
2. ABBAS AK, LICHTMAN AH, PILLAI S. *Imunologia celular e molecular.* 8º ed. Rio de Janeiro: Elsevier; 2015.
3. Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. *Proc R Soc B Biol Sci.* 22 de dezembro de 2015;282(1821):20143085.
4. Silva E dos S. Estudo de marcadores imunológicos em células Natural Killer – NK em homens e mulheres saudáveis [Internet] [TCC]. [João Pessoa]: Universidade Federal da Paraíba; 2020. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/17280>
5. Dunkelberger JR, Song WC. Complement and its role in innate and adaptive immune responses. *Cell Res.* janeiro de 2010;20(1):34–50.
6. Mandal A, Viswanathan C. Natural killer cells: In health and disease. *Hematol Oncol Stem Cell Ther.* junho de 2015;8(2):47–55.
7. Walzer T, Dalod M, Robbins SH, Zitvogel L, Vivier E. Natural-killer cells and dendritic cells: “l’union fait la force”. *Blood.* 1º de outubro de 2005;106(7):2252–8.
8. Dizaji Asl K, Velaei K, Rafat A, Tayefi Nasrabadi H, Movassaghpour AA, Mahdavi M, et al. The role of KIR positive NK cells in diseases and its importance in clinical intervention. *Int Immunopharmacol.* março de 2021;92:107361.
9. Jacobs R, Hintzen G, Kemper A, Beul K, Kempf S, Behrens G, et al. CD56bright cells differ in their KIR repertoire and cytotoxic features from CD56dim NK cells. *Eur J Immunol.* 2001 de 10DC;31(10):3121–6.
10. Vivier E, Raulet DH, Moretta A, Caligiuri MA, Zitvogel L, Lanier LL, et al. Innate or Adaptive Immunity? The Example of Natural Killer Cells. *Science.* 7 de janeiro de 2011;331(6013):44–9.
11. Romagnani C, Juelke K, Falco M, Morandi B, D’Agostino A, Costa R, et al. CD56brightCD16- killer Ig-like receptor- NK cells display longer telomeres and acquire features of CD56dim NK cells upon activation. *J Immunol Baltim Md* 1950. 15 de abril de 2007;178(8):4947–55.
12. Subedi N, Verhagen LP, Bosman EM, Van Roessel I, Tel J. Understanding natural killer cell biology from a single cell perspective. *Cell Immunol.* março de 2022;373:104497.
13. Johnson JK, Miller JS. Current strategies exploiting NK-cell therapy to treat haematologic malignancies. *Int J Immunogenet.* outubro de 2018;45(5):237–46.
14. Moretta L. Dissecting CD56dim human NK cells. *Blood.* 11 de novembro de 2010;116(19):3689–91.
15. Lanier LL. Up on the tightrope: natural killer cell activation and inhibition. *Nat Immunol.* maio de 2008;9(5):495–502.
16. Piersma SJ, Brizić I. Natural killer cell effector functions in antiviral defense. *FEBS J.* julho de 2022;289(14):3982–99.
17. Andreescu M, Frîncu F, Plotogea M, Mehedințu C. Recurrent Abortion and the Involvement of Killer-Cell Immunoglobulin-like Receptor (KIR) Genes, Activated T Cells, NK Abnormalities, and Cytokine Profiles. *J Clin Med.* 8 de fevereiro de 2023;12(4):1355.
18. Older Aguilar AM, Guethlein LA, Adams EJ, Abi-Rached L, Moesta AK, Parham P. Coevolution of Killer Cell Ig-Like Receptors with HLA-C To Become the Major Variable Regulators of Human NK Cells. *J Immunol.* 1º de outubro de

- 2010;185(7):4238–51.
19. Campbell KS, Purdy AK. Structure/function of human killer cell immunoglobulin-like receptors: lessons from polymorphisms, evolution, crystal structures and mutations: Structure/function of human KIR. *Immunology*. março de 2011;132(3):315–25.
 20. Walter L, Petersen B. Diversification of both KIR and NKG2 natural killer cell receptor genes in macaques - implications for highly complex MHC-dependent regulation of natural killer cells. *Immunology*. fevereiro de 2017;150(2):139–45.
 21. de Groot NG, Blokhuis JH, Otting N, Doxiadis GGM, Bontrop RE. Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunol Rev*. setembro de 2015;267(1):228–45.
 22. Goldberg AC, Rizzo LV. MHC structure and function – antigen presentation. Part 1. *Einstein São Paulo*. 24 de março de 2015;13(1):153–6.
 23. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med*. 7 de setembro de 2000;343(10):702–9.
 24. Pende D, Falco M, Vitale M, Cantoni C, Vitale C, Munari E, et al. Killer Ig-Like Receptors (KIRs): Their Role in NK Cell Modulation and Developments Leading to Their Clinical Exploitation. *Front Immunol*. 28 de maio de 2019;10:1179.
 25. Middleton D, Gonzelez F. The extensive polymorphism of KIR genes. *Immunology*. janeiro de 2010;129(1):8–19.
 26. Graef T, Moesta AK, Norman PJ, Abi-Rached L, Vago L, Older Aguilar AM, et al. KIR2DS4 is a product of gene conversion with KIR3DL2 that introduced specificity for HLA-A*11 while diminishing avidity for HLA-C. *J Exp Med*. 26 de outubro de 2009;206(11):2557–72.
 27. Goodridge JP, Burian A, Lee N, Geraghty DE. HLA-F and MHC Class I Open Conformers Are Ligands for NK Cell Ig-like Receptors. *J Immunol*. 1º de outubro de 2013;191(7):3553–62.
 28. Ljunggren HG, Kärre K. In search of the “missing self”: MHC molecules and NK cell recognition. *Immunol Today*. julho de 1990;11(7):237–44.
 29. Cruz-Muñoz ME, Valenzuela-Vázquez L, Sánchez-Herrera J, Santa-Olalla Tapia J. From the “missing self” hypothesis to adaptive NK cells: Insights of NK cell-mediated effector functions in immune surveillance. *J Leukoc Biol*. 19 de abril de 2019;105(5):955–71.
 30. Quatrini L, Della Chiesa M, Sivori S, Mingari MC, Pende D, Moretta L. Human NK cells, their receptors and function. *Eur J Immunol*. julho de 2021;51(7):1566–79.
 31. Moretta L, Moretta A. Unravelling natural killer cell function: triggering and inhibitory human NK receptors. *EMBO J*. 28 de janeiro de 2004;23(2):255–9.
 32. Guselnikov SV, Taranin AV. Unraveling the LRC Evolution in Mammals: IGSF1 and A1BG Provide the Keys. Zufall R, organizador. *Genome Biol Evol*. 1º de junho de 2019;11(6):1586–601.
 33. Dębska-Zielkowska J, Moszkowska G, Zieliński M, Zielińska H, Dukat-Mazurek A, Trzonkowski P, et al. KIR Receptors as Key Regulators of NK Cells Activity in Health and Disease. *Cells*. 14 de julho de 2021;10(7):1777.
 34. Gardiner CM. Killer cell immunoglobulin-like receptors on NK cells: the how, where and why: NK cell receptor expression. *Int J Immunogenet*. 18 de dezembro de 2007;35(1):1–8.
 35. Li YM, Li YX, Hu XZ, Li DY, An L, Yuan ZY, et al. Exploration of KIR genes and hematological-related diseases in Chinese Han population. *Sci Rep*. 16 de junho de 2023;13(1):9773.

36. Theeranawakam A, Vejbaesya S, Khuhapinant A, Sae-Tam P. Killer cell immunoglobulin-like receptors in Thai patients with multiple myeloma. *Asian Pac J Allergy Immunol.* 2 de janeiro de 2021;
37. Andrade HS. Diversidade do gene KIR3DL3 em populações mundiais e aspectos evolutivos [Internet] [Dissertação]. [Botucatu]: Universidade Estadual Paulista “Júlio de Mesquita Filho”; 2020. Disponível em: <https://repositorio.unesp.br/handle/11449/193008>
38. Marsh SGE, Parham P, Dupont B, Geraghty DE, Trowsdale J, Middleton D, et al. Killer-cell Immunoglobulin-like Receptor (KIR) Nomenclature Report, 2002.
39. Cardozo DM, Marangon AV, Guimarães F, Marques S, Lieber S, Delamain M, et al. Association of KIR Genes Polymorphism and its HLA Ligands in Diffuse Large B-cell Lymphoma. *Clin Lymphoma Myeloma Leuk.* junho de 2023;23(6):438–45.
40. Béziat V, Hilton HG, Norman PJ, Traherne JA. Deciphering the killer-cell immunoglobulin-like receptor system at super-resolution for natural killer and T-cell biology. *Immunology.* março de 2017;150(3):248–64.
41. Barrow AD, Trowsdale J. The extended human leukocyte receptor complex: diverse ways of modulating immune responses. *Immunol Rev.* agosto de 2008;224(1):98–123.
42. IPD-KIR [Internet]. 2022. IPD-KIR. Disponível em: <https://www.ebi.ac.uk/ipd/kir/about/>
43. Augusto DG. GENES KIR NO BRASIL: CARACTERIZAÇÃO DE TRÊS POPULAÇÕES. 2006;
44. Roe D, Kuang R. Accurate and Efficient KIR Gene and Haplotype Inference From Genome Sequencing Reads With Novel K-mer Signatures. *Front Immunol.* 26 de novembro de 2020;11:583013.
45. Oliveira LM, Portela P, Merzoni J, Beppler J, Dias FS, Graebin P, et al. KIR gene haplotype A is associated with hospital mortality in patients with sepsis. *Clin Immunol.* março de 2019;200:37–8.
46. Maxwell LD, Wallace A, Middleton D, Curran MD. A common KIR2DS4 deletion variant in the human that predicts a soluble KIR molecule analogous to the KIR1D molecule observed in the rhesus monkey: Maxwell et al : A predicted soluble KIR2DS4 deletion variant. *Tissue Antigens.* setembro de 2002;60(3):254–8.
47. Tukwasibwe S, Traherne JA, Chazara O, Jayaraman J, Trowsdale J, Moffett A, et al. Diversity of KIR genes and their HLA-C ligands in Ugandan populations with historically varied malaria transmission intensity. *Malar J.* 25 de fevereiro de 2021;20(1):111.
48. Gourraud PA, Meenagh A, Cambon-Thomsen A, Middleton D. Linkage disequilibrium organization of the human KIR superlocus: implications for KIR data analyses. *Immunogenetics.* dezembro de 2010;62(11–12):729–40.
49. Kuśnierczyk P. Are Killer Cell Immunoglobulin-Like Receptor Genes Important for the Prediction of Kidney Graft Rejection? *Arch Immunol Ther Exp (Warsz).* agosto de 2013;61(4):321–5.
50. Jiang W, Johnson C, Jayaraman J, Simecek N, Noble J, Moffatt MF, et al. Copy number variation leads to considerable diversity for B but not A haplotypes of the human KIR genes encoding NK cell receptors. *Genome Res.* outubro de 2012;22(10):1845–54.
51. Piovezan BZ. Diversidade KIR em duas populações ameríndias do Mato Grosso do Sul [TCC]. [Curitiba]: Universidade Federal do Paraná; 2007.
52. Rajagopalan S, Long EO. Understanding how combinations of HLA and KIR

- genes influence disease. *J Exp Med*. 4 de abril de 2005;201(7):1025–9.
53. Hiby SE, Walker JJ, O’Shaughnessy KM, Redman CWG, Carrington M, Trowsdale J, et al. Combinations of Maternal KIR and Fetal HLA-C Genes Influence the Risk of Preeclampsia and Reproductive Success. *J Exp Med*. 18 de outubro de 2004;200(8):957–65.
 54. Merino AM, Dugast AS, Wilson CM, Goepfert PA, Alter G, Kaslow RA, et al. KIR2DS4 Promotes HIV-1 Pathogenesis: New Evidence from Analyses of Immunogenetic Data and Natural Killer Cell Function. Unutmaz D, organizador. *PLoS ONE*. 5 de junho de 2014;9(6):e99353.
 55. González-Galarza FF, Takeshita LYC, Santos EJM, Kempson F, Maia MHT, Silva ALS da, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res*. 28 de janeiro de 2015;43(D1):D784–8.
 56. Hsu KC, Liu XR, Selvakumar A, Mickelson E, O’Reilly RJ, Dupont B. Killer Ig-like receptor haplotype analysis by gene content: evidence for genomic diversity with a minimum of six basic framework haplotypes, each with multiple subsets. *J Immunol Baltim Md 1950*. 1º de novembro de 2002;169(9):5118–29.
 57. Saito H, Hirayama A, Umemura T, Joshita S, Mukawa K, Suga T, et al. Association between KIR-HLA combination and ulcerative colitis and Crohn’s disease in a Japanese population. Matsunami H, organizador. *PLOS ONE*. 12 de abril de 2018;13(4):e0195778.
 58. Downing J, D’Orsogna L. High-resolution human KIR genotyping. *Immunogenetics*. agosto de 2022;74(4):369–79.
 59. Duncan MD. Transplant-related Immunosuppression: A Review of Immunosuppression and Pulmonary Infections. *Proc Am Thorac Soc*. 1º de dezembro de 2005;2(5):449–55.
 60. Bao XJ, Hou LH, Sun AN, Qiu QC, Yuan XN, Chen MH, et al. The impact of KIR2DS4 alleles and the expression of KIR in the development of acute GVHD after unrelated allogeneic hematopoietic SCT. *Bone Marrow Transplant*. setembro de 2010;45(9):1435–41.
 61. Kulkarni S, Martin MP, Carrington M. The Yin and Yang of HLA and KIR in human disease. *Semin Immunol*. dezembro de 2008;20(6):343–52.
 62. Bernal E, Gimeno L, Alcaraz MJ, Quadeer AA, Moreno M, Martínez-Sánchez MV, et al. Activating Killer-Cell Immunoglobulin-Like Receptors Are Associated With the Severity of Coronavirus Disease 2019. *J Infect Dis*. 15 de julho de 2021;224(2):229–40.
 63. Hajeer A, Jawdat D, Massadeh S, Aljawini N, Abedalthagafi MS, Arabi YM, et al. Association of KIR gene polymorphisms with COVID-19 disease. *Clin Immunol*. janeiro de 2022;234:108911.
 64. Margolis DJ, Mitra N, Hoffstad OJ, Chopra A, Phillips EJ. KIR Allelic Variation and the Remission of Atopic Dermatitis Over Time. *ImmunoHorizons*. 1º de janeiro de 2023;7(1):30–40.
 65. Beigmohammadi F, Aslani S, Kavosi H, Javinani A, Mostafaei S, Pournazari M, et al. Association of Killer Cell Immunoglobulin-like Receptor (KIR) Genes and their HLA Ligands with Susceptibility to Takayasu Arteritis in the Iranian Population. *Iran J Allergy Asthma Immunol [Internet]*. 25 de fevereiro de 2023 [citado 25 de novembro de 2023]; Disponível em: <https://publish.kne-publishing.com/index.php/IJAAI/article/view/12003>
 66. Akar HH, Patiroglu T, Sevinc E, Aslan D, Okdemir D, Kurtoglu S. Contribution of KIR (killer immunoglobulin-like receptor) genes, HLA class I ligands, and

- KIR/HLA class I ligand combinations on the genetic predisposition to celiac disease and coexisting celiac disease and type 1 diabetes mellitus. *Rev Esp Enferm Dig.* 2015;107.
67. Aghaei H, Mostafaei S, Aslani S, Jamshidi A, Mahmoudi M. Association study between KIR polymorphisms and rheumatoid arthritis disease: an updated meta-analysis. *BMC Med Genet.* dezembro de 2019;20(1):24.
 68. Denis L, Sivula J, Gourraud PA, Kerdudou N, Chout R, Ricard C, et al. Genetic diversity of KIR natural killer cell markers in populations from France, Guadeloupe, Finland, Senegal and Réunion. *Tissue Antigens.* outubro de 2005;66(4):267–76.
 69. Katz G, Markel G, Mizrahi S, Arnon TI, Mandelboim O. Recognition of HLA-Cw4 but not HLA-Cw6 by the NK cell receptor killer cell Ig-like receptor two-domain short tail number 4. *J Immunol Baltim Md* 1950. 15 de junho de 2001;166(12):7260–7.
 70. Middleton D, Williams F, Halfpenny IA. KIR genes. *Transpl Immunol.* agosto de 2005;14(3–4):135–42.
 71. Middleton D, Gonzalez A, Gilmore PM. Studies on the expression of the deleted KIR2DS4*003 gene product and distribution of KIR2DS4 deleted and nondeleted versions in different populations. *Hum Immunol.* fevereiro de 2007;68(2):128–34.
 72. Marangon AV, Visentainer JEL, Guelsin GAS, Clementino SL, Rudnick CCC, de Melo FC, et al. Investigation of deletion of 22pb in KIR2DS4 gene in a population of southern Brazil. *J Clin Lab Anal.* novembro de 2014;28(6):440–5.
 73. McErlean C, Gonzalez AA, Cunningham R, Meenagh A, Shovlin T, Middleton D. Differential RNA expression of KIR alleles. *Immunogenetics.* julho de 2010;62(7):431–40.
 74. Bruijnesteijn J, Van Der Wiel MKH, De Groot N, Otting N, De Vos-Rouweler AJM, Lardy NM, et al. Extensive Alternative Splicing of KIR Transcripts. *Front Immunol.* 4 de dezembro de 2018;9:2846.
 75. Gowdavally S, Tsamadou C, Platzbecker U, Sala E, Valerius T, Klein S, et al. KIR2DS4 and Its Variant KIR1D in KIR-AA Genotype Donors Showed Differential Survival Impact in Patients with Lymphoid Disease after HLA-Matched Unrelated Hematopoietic Stem Cell Transplantation. *Transplant Cell Ther.* julho de 2023;29(7):457.e1-457.e10.
 76. Kennedy PR, Chazara O, Gardner L, Ivarsson MA, Farrell LE, Xiong S, et al. Activating KIR2DS4 Is Expressed by Uterine NK Cells and Contributes to Successful Pregnancy. *J Immunol Author Choice.* 1º de dezembro de 2016;197(11):4292–300.
 77. Amorim LM, Augusto DG, Nemat-Gorgani N, Montero-Martin G, Marin WM, Shams H, et al. High-Resolution Characterization of KIR Genes in a Large North American Cohort Reveals Novel Details of Structural and Sequence Diversity. *Front Immunol.* 7 de maio de 2021;12:674778.
 78. Ordóñez D, Moraru M, Gómez-Lozano N, Cisneros E, Vilches C. KIR Typing by Non-Sequencing Methods: Polymerase-Chain Reaction with Sequence-Specific Primers. Em: Christiansen FT, Tait BD, organizadores. *Immunogenetics* [Internet]. Totowa, NJ: Humana Press; 2012 [citado 25 de novembro de 2023]. p. 415–30. (Methods in Molecular Biology™; vol. 882). Disponível em: http://link.springer.com/10.1007/978-1-61779-842-9_24
 79. Maxwell LD, Williams F, Gilmore P, Meenagh A, Middleton D. Investigation of killer cell immunoglobulin-like receptor gene diversity: II. KIR2DS4. *Hum Immunol.* junho de 2004;65(6):613–21.

80. Farias TDJ, Brugiapaglia S, Croci S, Magistroni P, Curcio C, Zguro K, et al. HLA-DPB1*13:01 associates with enhanced, and KIR2DS4*001 with diminished protection from developing severe COVID-19. *HLA*. 18 de outubro de 2023;
81. Alicata C, Ashouri E, Nemat-Gorgani N, Guethlein LA, Marin WM, Tao S, et al. KIR Variation in Iranians Combines High Haplotype and Allotype Diversity With an Abundance of Functional Inhibitory Receptors. *Front Immunol*. 2 de abril de 2020;11:556.
82. Sala Elpidio LN, De Alencar JB, Tsuneto PY, Alves HV, Trento Toretta M, It Taura SK, et al. Killer-cell immunoglobulin-like receptors associated with polycystic ovary syndrome. *J Reprod Immunol*. novembro de 2018;130:1–6.
83. Roe D, Vierra-Green C, Pyo CW, Eng K, Hall R, Kuang R, et al. Revealing complete complex KIR haplotypes phased by long-read sequencing technology. *Genes Immun*. setembro de 2017;18(3):127–34.
84. Traherne JA, Jiang W, Valdes AM, Hollenbach JA, Jayaraman J, Lane JA, et al. KIR haplotypes are associated with late-onset type 1 diabetes in European–American families. *Genes Immun*. janeiro de 2016;17(1):8–12.
85. Sakaue S, Hosomichi K, Hirata J, Nakaoka H, Yamazaki K, Yawata M, et al. Decoding the diversity of killer immunoglobulin-like receptors by deep sequencing and a high-resolution imputation method. *Cell Genomics*. março de 2022;2(3):100101.
86. Norman PJ, Hollenbach JA, Nemat-Gorgani N, Marin WM, Norberg SJ, Ashouri E, et al. Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing. *Am J Hum Genet*. agosto de 2016;99(2):375–91.
87. Song L, Bai G, Liu XS, Li B, Li H. Efficient and accurate KIR and HLA genotyping with massively parallel sequencing data. *Genome Res*. junho de 2023;33(6):923–31.
88. Vukcevic D, Traherne JA, Næss S, Ellinghaus E, Kamatani Y, Dilthey A, et al. Imputation of KIR Types from SNP Variation Data. *Am J Hum Genet*. outubro de 2015;97(4):593–607.
89. Gerada C, Campbell TM, Kennedy JJ, McSharry BP, Steain M, Slobedman B, et al. Manipulation of the Innate Immune Response by Varicella Zoster Virus. *Front Immunol*. 24 de janeiro de 2020;11:1.
90. Chen J, Madireddi S, Nagarkar D, Migdal M, Vander Heiden J, Chang D, et al. *In silico* tools for accurate HLA and KIR inference from clinical sequencing data empower immunogenetics on individual-patient and population scales. *Brief Bioinform*. 20 de maio de 2021;22(3):bbaa223.
91. Marin WM, Dandekar R, Augusto DG, Yusufali T, Heyn B, Hofmann J, et al. High-throughput Interpretation of Killer-cell Immunoglobulin-like Receptor Short-read Sequencing Data with PING. Ay F, organizador. *PLOS Comput Biol*. 2 de agosto de 2021;17(8):e1008904.
92. Marin WM, Hollenbach JA. Software update: Interpreting killer-cell immunoglobulin-like receptors from whole genome sequence data with PING. *HLA*. maio de 2023;101(5):441–8.
93. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, Abecasis GR, Steering committee, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 1º de outubro de 2015;526(7571):68–74.
94. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Regier AA, Corvelo A, et al. High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios.

CAPÍTULO II - ARTIGO

A new pipeline to evaluate *KIR2DS4* copy number and genotypes distribution in worldwide populations

Authors

Gabriela Sato Paes, Raphaela Neto Pereira, Heloisa de Souza Andrade, Camila Ferreira Bannwart Castro, Erick da Cruz Castelli

Abstract

Natural Killer (NK) cells are essential in innate immune response. NK surface receptors such as KIR modulate NK cell cytotoxicity. The KIR genes have similar sequences of DNA and a degree of polymorphism. *KIR2DS4* encodes an activator receptor and might be absent from an individual, or present in two possible isoforms, the functional full-length molecule (*KIR2DS4-full*) or the non-functional truncated one. The presence of these isoforms may influence transplantation and infection outcomes. Here, we present a new computation method to survey *KIR2DS4* copy numbers, variants, and alleles from short reads, by using the kir-mapper workflow. We have applied this method to survey *KIR2DS4* diversity in 490 samples from 5 populations: GBR (British in Europe), JPT (Japanese in East Asia), YRI (Yoruba in Africa), ITU (Telugu in South Asia), and CLM (Colombia from America) from the 1000 Genomes Project. *KIR2DS4* was detected in all populations in most individuals (more than 90%). The complete absence of *KIR2DS4* is uncommon, but more frequent in South Asia (about 10%). We detected individuals with 0, 1, 2, or 3 copies of *KIR2DS4*. The most frequent genotype was two copies, except for South Asia with one copy (49%), and rare cases of three copies, only detected in JPT (1%) and YRI (2.9%). *KIR2DS4* allele frequencies vary among populations. While some alleles are highly frequent everywhere (*KIR2DS4*00101*, **00301*, **00401*, **010*), others are only absent in specific populations (*KIR2DS4*00601* absent in Japan), and others are detected in specific populations (**00103* in Colombia and **016* in Japan). Although most individuals present at least one copy of *KIR2DS4*, the frequency of individuals carrying at least one copy of an allele encoding the full/functional version of *KIR2DS4* varied from 30.4% in South Asia, 42% in Europe, to 78.4% in Japan. These frequencies are the expected ones based on previous studies in similar populations. In conclusion, this new computational method can be an alternative or additional tool to characterize *KIR2DS4* in NGS data, and can be used in functional and disease association studies. The presence of a functional *KIR2DS4* gene is less frequent in South Asia, Europe, and admixed populations with major European and South Asian backgrounds.

Keywords: *KIR2DS4*, copy number variation, NGS, *KIR2DS4-del*.

Introduction

KIR2DS4 is the oldest and most divergent activator KIR (killer immunoglobulin-like receptor) receptor (1) and, in several cases, the only activating KIR in some individuals. Several studies detected association of *KIR2DS4* with successful pregnancy (2), protection from hepatitis C virus (3), and promoted the pathogenesis of HIV-1 (4). This receptor binds with HLA-C serotypes C1 and C2, HLA-A*11 which is a ligand of KIR3DL2 too (1), and possibly HLA-F (5–8). Analysis between the interaction of KIR/HLA in general has been increasing to investigate the impact of them on the outcome and progression of viral infection (9). For *KIR2DS4*, there are 44 alleles encoding 22 different protein sequences (10).

It is not clear whether SNPs across *KIR2DS4* influence protein/receptor function or expression patterns. However, a deletion of 22 nucleotides in exon 5 of the gene has a functional impact, because it leads to a different reading frame and a premature stop codon in the final part of exon 7 (11). Consequently, this deletion impairs the translation of the part of the D2 domain, the transmembrane domain, the stem, and the cytoplasmic domain, producing a truncated version of KIR2DS4 (11). This truncated version is known as *KIR2DS4-del* or *KIR1D*, and the alleles described until now carrying this deletion are KIR2DS4* 003, *004, *006, *007, *008, *009, *010, *012, *013, and *018 (10). It is theorized that if this truncated version is expressed, it would be secreted as a soluble isoform with loss of function (5,12,13), failing to anchor in the membrane surface of NK cells (14). They might also be unable to bind with HLA class I molecules (7,12,13). Consequently, KIR2DS4 presented two isoforms: functional, known as KIR2DS4-full and non-functional, represented by KIR2DS4-del or KIR1D. Therefore, if an individual has only KIR2DS4 as an activating receptor within the KIR receptors, and only the KIR2DS4-del in both chromosome 19, thus none of KIR receptors would be capable of activation of NK cells. The presence of this truncated KIR2DS4 protein was associated with syphilis (patients 16.85% vs control 6.59%) in the Chinese Han population (15), and with higher graft-versus-host disease-free ($p = .002$) and relapse-free survival rates ($p = .002$) (16) in hematopoietic stem cell transplantation.

The *KIR* genes are considered the most polymorphic receptors from NK cells, mainly because of the combination of allelic diversity, copy number polymorphism (17) and similarity between the sequence of DNA. The functional diversity within KIR

variations has been a subject of several studies, yielding consistent and replicable information. Nevertheless, conflicting results emerged due to distinct ancestral cohorts. Furthermore, there's a critical need to refine our understanding of copy number variations and allelic diversity. For instance, a specifically characterized allele remained undetectable using solely gene-content methods. This underscores the necessity for more nuanced approaches to fully capture the diversity and functionality of KIRs. Next-generation (NGS) and third-generation sequencing methods using short reads can generate a massive amount of sequencing data in a high throughput and cost-effective manner. The super high-resolution sequencing will provide a detailed view of the *KIR* gene and haplotype structures. This knowledge can then be carried into studies of how allelic variation might affect NK cell function, outcomes in HSC and solid organ transplantation, and the progression of infectious disease.

There is a lack of tools to sequence and analyze *KIR* genes, especially from NSG methodology. Although many pipelines and programs have been created to survey *KIR* genes and polymorphisms from whole-genomes and targeted sequencing data, such as PING and T1K (18), their results are sometimes conflicting. The most used pipeline to analyze *KIR* genes is PING (19), updated last year to be compatible with whole genomes (20). However, it depends on the availability of high-performance computers when evaluating a large number of samples, and it has not been designed for Exome sequencing (which is an increasing demand).

Here, we propose a different method to survey *KIR* genes from whole-genome sequencing data (WGS), focusing on detecting *KIR2DS4* copy numbers and alleles. Then, we applied this pipeline to survey *KIR2DS4* in 490 samples from the 1000 Genomes Project (21), divided into 5 populations and distributed by ancestry respectively: British in England and Scotland (GBR, Europe), Yoruba in Ibadan (YRI, Africa), Japanese in Tokyo (JPT, East Asia), Colombians from Medellin (CLM, South America) and Indian Telugu (ITU, South Asia).

Methods

The kir-mapper workflow and read alignment: the map algorithm

The kir-mapper is a toolkit to analyze *KIR* genes from NGS short-read sequencing data. The toolkit presents 3 main functions, one for read alignment to the reference genome (function map), one to calculate coverage and copy numbers (function ncopy), and one for genotyping SNPs *KIR* alleles (function genotype), Figure 1.

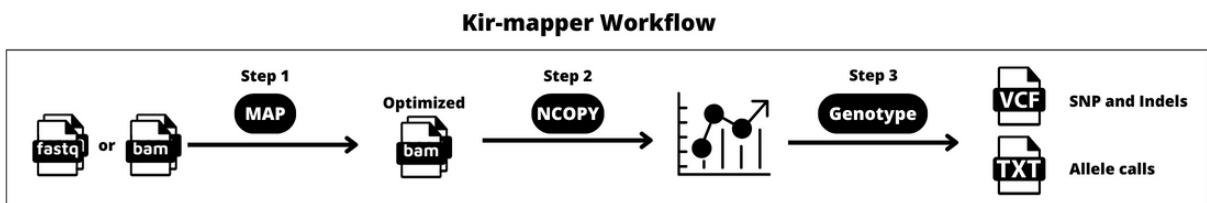


Figure 1: The kir-mapper workflow to produce BAM files for each sample, call copy numbers, SNPs, and alleles.

The map function applies an algorithm to calculate the distance between each read and known sequences for all *KIR* genes, comparing these distances to assign each read to its proper locus. Some reads present the same distance for more than one *KIR* gene. In these cases, they are assigned to multiple genes but marked as secondary alignments. Kir-mapper::map also applies an algorithm that detects two sequences for each gene that best fits the observed reads and, when possible, recovers some of these reads marked as secondary, passing them to primary alignments.

This step is important to get unbiased alignments. Because of the repetitive nature of all *KIR* genes, they present a high sequence similarity. Therefore, conventional alignment tools, such as BWA or Bowtie2, which uses a single reference genome to guide the alignments, are prone to many alignment errors, jeopardizing genotyping procedures. Kir-mapper::map employs a multi-referenced alignment because it considers many references for each gene to decide where each read should be aligned, but, in the end, forces read alignment to a single reference as other aligners do.

The effect of kir-mapper::map can be observed in Figure 2. Here, we generated a virtual NGS sequencing simulation for 50 samples, with two known

sequences for every *KIR* gene, a read length of 2x150bp (paired-end) and a mean fragment size of 450. We have aligned all samples with BWA MEM (22) and with `kir-mapper::map`, tracking the origin of each read and where they were aligned. When using BWA, the distribution of reads across *KIR2DS4* is represented in Figure 2A. We can observe that many reads from other genes are aligned to *KIR2DS4*, and the coverage across the gene is much higher than the expected 30x. These misaligned reads would jeopardize any genotyping procedure and the estimation of copy numbers based on comparing coverage between a target and a reference. When using `kir-mapper::map`, most of these misalignments are solved, and we observe a homogeneous read coverage across *KIR2DS4* around the expected value of 30X, with few exceptions (Figure 2B).

The end product of `kir-mapper::map` is a BAM file with reads aligned to the hg38 reference genome, and this BAM file can be inspected on other programs, such as Integrative Genome Viewer (IGV) (23). This step has some dependencies, including BWA MEM (22), samtools (24), freebayes (25), and picard-tools (26). The references for each *KIR* gene used by the `kir-mapper` came from IPD-KIR (10) databases. The inputs for this step can be either the raw fastq, or a BAM file with reads aligned to the hg38 reference genome. The `kir-mapper::map` step must be applied to all samples. For this analysis, we used an input fastq file to compare with another tool.

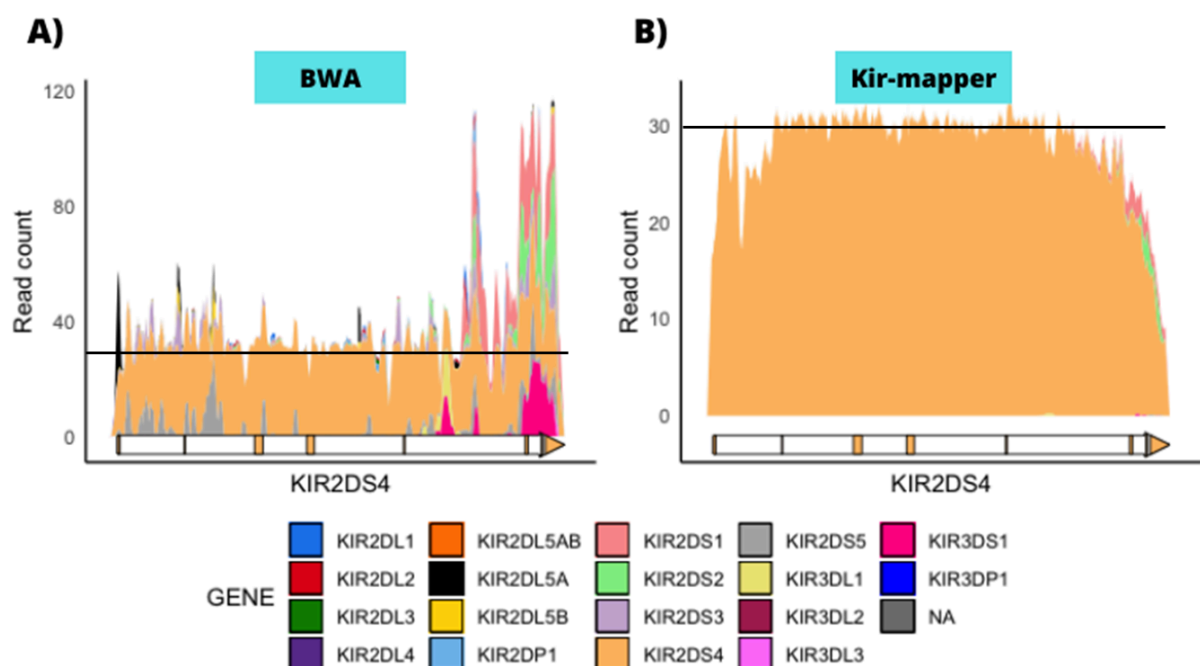


Figure 2: The differences observed for read alignment when using BWA (A) and kir-mapper (B) considering gene *KIR2DS4*. The y-axis represents the proportion of reads (coverage), different colors represent different origins for the reads, and the *KIR2DS4* structure is represented at the bottom, with exons as orange squares and introns as white squares. Specifically, *KIR2DS4* reads are colored in orange, while the presence of other colors corresponds to different *KIR* genes. The horizontal black line indicates the expected coverage, 30X.

Determining copy numbers: the ncopy algorithm

The second kir-mapper main function is ncopy, which uses the optimized BAM files produced in the previous step to calculate read depth across each *KIR* gene and some references. In this step, it can assess the presence/absence of each gene and determine their copy number, from 0 to 4 or more copies. Copy number determination is crucial to get accurate genotypes.

Kir-mapper::ncopy uses samtools depth (24) to determine the coverage of selected gene regions, calculating the ratio between the observed coverage in the target and some references. The *default* reference is usually *KIR3DL3*, which is expected to be present in two copies in all individuals (it is so far the only true framework gene). However, the user can set other references to test whether *KIR3DL3* is truly a framework, such as genes *HLA-G*, and *HLA-E*, which are HLA class I non-classical and least polymorphic of HLA region (27,28), and a region upstream of the LRC that we called 5UPKIR (some nucleotides upstream gene *KIR3DL3* and the *KIR* complex, between *ILT2* and *KIR3DL3*).

The end product of this step is plots (in PNG and HTML formats) of the ratios between the targets's coverage and the coverage of the reference. We expect a ratio around zero for samples with no copy of the target gene, 0.5 for samples with just 1 copy, 1.0 with samples with 2 copies, 1.5 with samples with 3 copies, and 2.0 for samples with 4 copies. By manually inspecting the plots, the user can determine the best thresholds (the transition between groups of samples with the same pattern), and run the algorithm again with these adjusted thresholds to determine the copy numbers of all samples. An example of the coverage plot of the Colombian population for *KIR2DS4* in Figure 3, the results of other populations are in the supplementary section (Figure S6-9). Red dots represent samples with no heterozygous SNP. Blue and green dots represent samples with heterozygous SNPs. *KIR2DS1* and *KIR2DS4* are located in the same position in the locus, so when an individual has two copies of *KIR2DS4*, it's expected to have zero copies of *KIR2DS1*.

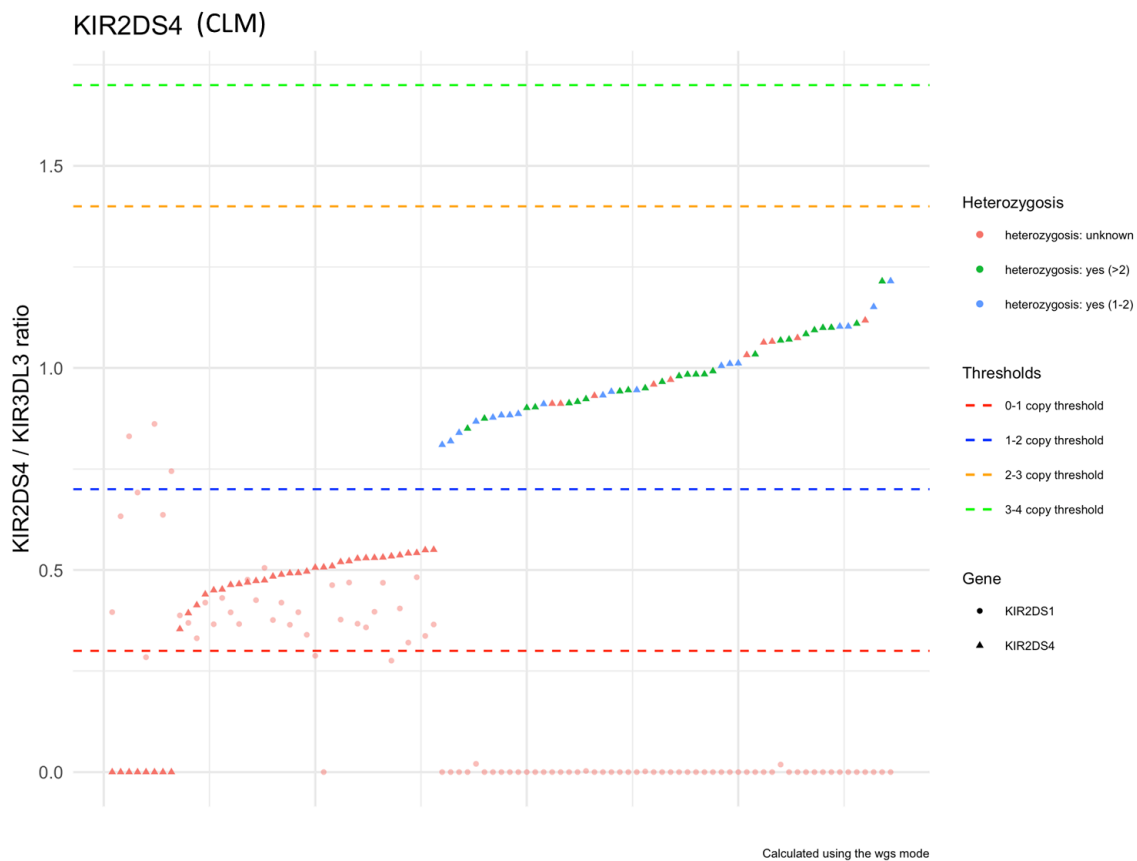


Figure 3: The KIR2DS4 copy number determination using `kir-mapper::ncopy` for CLM population. Each triangle represents a sample. The y-axis represents the ration between the coverage observed in KIR2DS4 and KIR3DL3. Because KIR2DS1 and KIR2DS4 are usually in the same position in the locus, we also plotted the ratios for KIR2DS1 (light dots).

Calling SNPs and KIR alleles: the genotype algorithm

The last algorithm is `kir-mapper::genotype`. It considers the copy numbers determined by `kir-mapper::ncopy` and the optimized bam files produced by `kir-mapper::map`, to call SNPs and Indels across each gene by using `freebayes` (25). It also applies a method to transform uncertain alleles into missing alleles, producing an unphased VCF with variants across each gene. This step also includes every allele described at the IPD-KIR (10) as a virtual sample homozygous for those alleles. Therefore, this step produces a VCF file with all tested samples and all references for a specific gene. After, it applies `WhatsHap` (29) to phase heterozygous sites that are close to each other and with reads passing by them. The comparison between the variants observed at the references (known alleles) and the samples determines the best allele combination for each sample, with reports of these allele combinations, how many variants were tested, and the proportion of matches

between the sample and the tested alleles, the proportion of missing variants, and the list of eventual mismatches that were detected. This algorithm can be applied in two formats, for whole-genome, evaluating all exons and introns of every *KIR* gene (as presented here), or for exomes, evaluating only exons. For whole-genomes, the best practice is the comparison of the results from both versions.

Sometimes, kir-mapper reports a series of possible allele combinations for the same sample (ambiguities). This might occur because there might be more than one allele combination with the same SNPs and genotypes but different phases among them. Although kir-mapper tries to phase all heterozygous sites, this phasing fails when two heterozygous sites are far apart, with no short read passing through them.

To solve this problem, kir-mapper has a built-in method to resolve ambiguities, that is only available when genotyping a minimum of 150 samples simultaneously and many *KIR* genes at the same time. We use statistical phasing to fill in the gaps among the phasing sets defined by whatshap (29), to define complete haplotypes, by using shapeit4 (30). With all variants phased, there is no ambiguity and the best combination of alleles is defined.

Testing kir-mapper with samples from all biogeographic regions

To test the kir-mapper for genotyping *KIR2DS4*, and also explore *KIR2DS4* diversity in different populations, we used high-coverage sequencing data (30x) from 490 samples from different locations. This dataset comprised related and unrelated samples from 1000 Genomes Project (21) dataset, from 5 populations (n = 490): Colombian (CLM); British (GBR); Yoruba (YRI); Japanese (JPT); and Telugu (ITU). The samples were also classified according to their biogeographic region ancestry as Europeans (EUR), Africans (AFR), East Asians (EAS), South Asians (SAS), and Americans (AMR).

For all these samples, we downloaded raw sequencing files from high-coverage (30x) from whole-genome sequencing (WGS) (fastq) or BAM files with short reads from Illumina Novaseq 6000 platform, using 2 X 150bp cycles, aligned to the human reference genome hg38 (GRCh38) using BWA-MEM (version 0.7.15) tool. (21)

Comparing Kir-mapper and PING

PING (19) is the most used tool to call *KIR* alleles from NGS data. It was created for and validated with targeted sequencing for *KIR*. To target *KIR* genes for next-generation sequencing, they used probes to the *KIR* region (19). Later, the tool was updated for whole-genome sequencing (20). PING can determine copy numbers, genotypes (allele calls), and novel sequences (19). We compared the outputs from PING (the WGS version) and kir-mapper to evaluate whether they produced similar results. The rationale here is that, since there is a lack of tools to deal with *KIR* genes and NGS data, and both methods apply an entirely different algorithm for aligning, copy number determination, and genotyping, the overlap between these two methods would greatly support an accurate *KIR* genotyping.

The graphics of comparison between the tools (kir-mapper and PING) and the results of this study were built using “ggplot” (version 3.3.6) (31) package in R (version 4.1.2) language (32) in RStudio (version 2023.06.2) (33) program.

Results

Presence/absence of *KIR2DS4* in different biogeographic regions

The *KIR* complex has been described as a polymorphic region, and this included that not all *KIR* genes are uniformly present in all individuals. The presence/absence polymorphism generates a variety of *KIR* combinations, which vary among individuals and populations as well (34). Most individuals present at least one copy of *KIR2DS4* (Figure 4). The frequency of individuals with no copy of *KIR2DS4* is higher in South Asia and Colombia. We observed that the ITU population differs from other populations ($p < 0.05$) regarding this matter. This suggests significant variation in *KIR2DS4* frequencies.

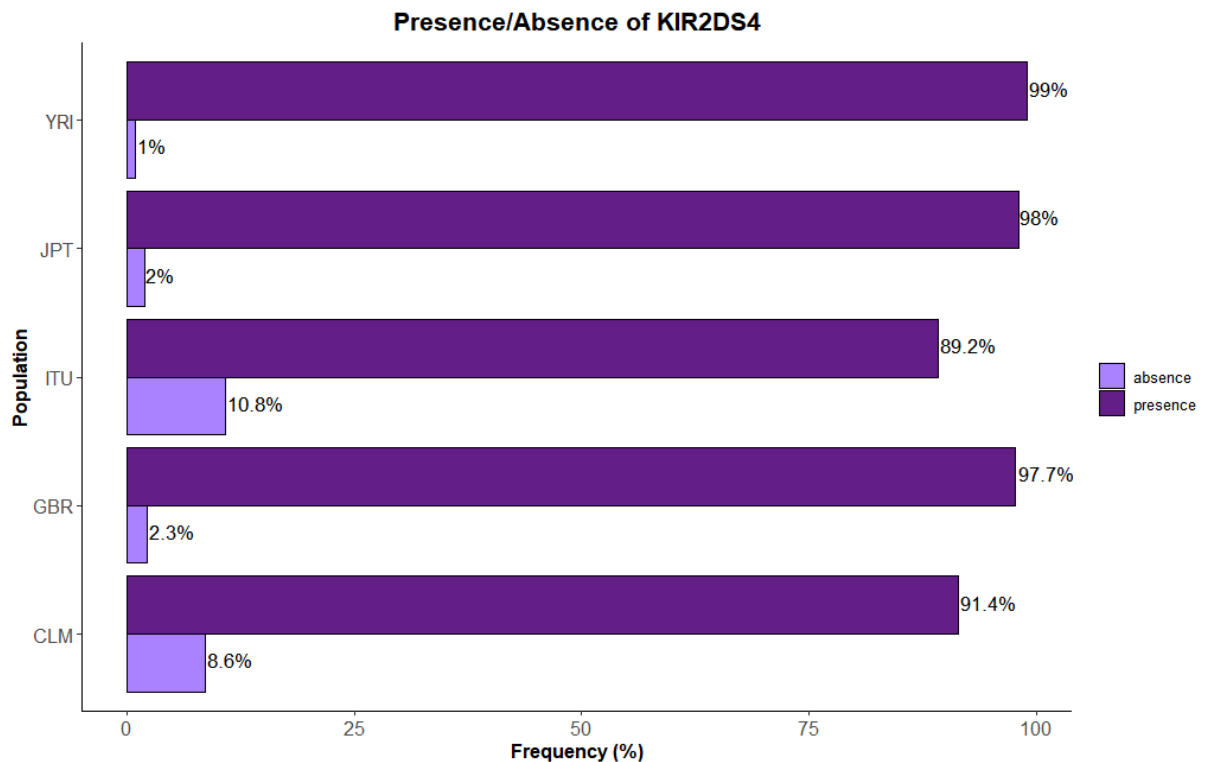


Figure 4: The proportion of individuals carrying at least one copy of *KIR2DS4* in different biogeographic regions. Population in Y axis: Colombian in America (CLM); British in Europe (GBR); Telugu in South Asia (ITU); Japanese in East Asia (JPT) and Yoruba in Africa (YRI).

The number of *KIR2DS4* copies among individuals

In addition to the presence or absence of *KIR2DS4*, we evaluated the number of copies each individual carries, by using `kir-mapper::ncopy` (Figure 3 and S6-9). We can observe the formation of at least three distinct groups of samples. The first group, with a *KIR2DS4*/*KIR3DL3* ratio around zero, represents samples with no *KIR2DS4* (*KIR2DS4* absence). The second group, with a ratio around 0.5, represents samples with one copy of *KIR2DS4*. The third group, with a ratio around 1.0 and the presence of many heterozygous sites (the blue and green triangles), represents samples with two copies of *KIR2DS4*. Sometimes, a fourth and fifth group can be noticed, indicating samples with more than 3 copies of *KIR2DS4*. The frequency of samples carrying a specific number of copies was calculated and plotted in Figure 5 and S10, and summarized per population at Table S1.

There are some rare cases of 3 copies of *KIR2DS4* in Africa (YRI) and Japan (JPT). Most individuals present two copies of *KIR2DS4* except among the South Asians (ITU).

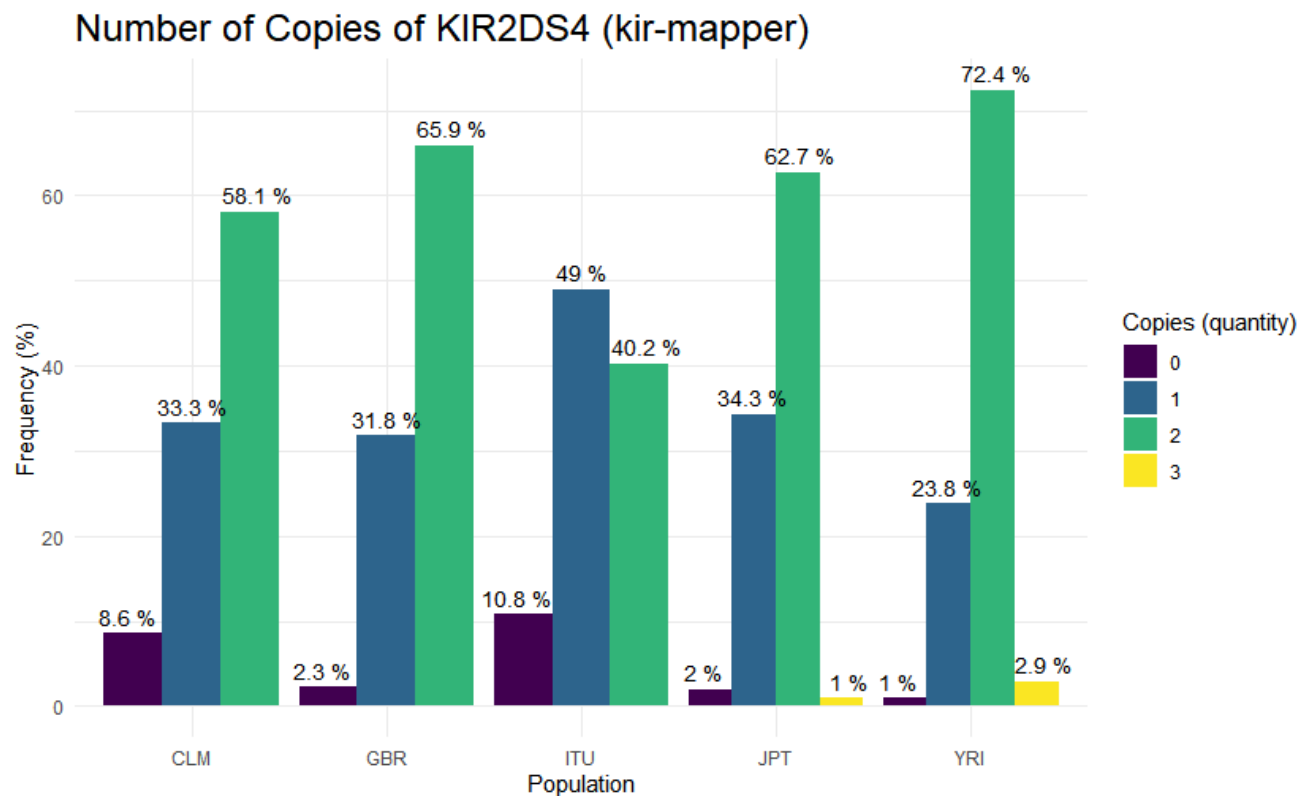


Figure 5: The number of copies of KIR2DS4 in different populations. Each color represents a quantity of *KIR2DS4*. Population in X axis: Colombian in America (CLM); British in Europe (GBR); Telugu in South Asia (ITU); Japanese in East Asia (JPT) and Yoruba in Africa (YRI).

***KIR2DS4* allele distribution**

The frequency of each allele we have detected is described in Table 1. Samples with 3 or more copies of *KIR2DS4* were deleted from this analysis to avoid a bias in allele frequencies, but this represents only 5 samples in total. We have encoded the absence of *KIR2DS4* as the *KIR2DS4*null* allele. Therefore, individuals with no *KIR2DS4* have two copies of *KIR2DS4*null*, and samples with one copy of *KIR2DS4* have one specific allele and one *KIR2DS4*null*. *KIR2DS4*unresolved* refers to a new allele or an allele that could not be resolved. In this study, the alleles encoding the full length and functional *KIR2DS4* protein (*KIR2DS4*-full) were *KIR2DS4*00101*, **00103*, **00104*, and **016*, while all the others present the 22bp deletion at exon 5 (*KIR2DS4*-del).

The alleles *KIR2DS4*00101*, **00301*, and **null* (the absence of *KIR2DS4*) are usually the most frequent ones globally, with a frequency of 33.1%, 15.9%, and 22.3%, respectively (Table 3). However, allele frequencies vary among populations. For instance, while **00101* (*KIR2DS4*-full) corresponds to about 50% of all alleles in

Japan and Africa, it drops down to 16.2% in East Asia. *KIR2DS4*00601* is frequent in most regions but absent from Japan.

Some samples presented unresolved alleles (Table 3), and the frequency of unresolved alleles is higher among Africans and East Asians. This might be a consequence of the underrepresentation of non-European samples in the database describing KIR sequences.

Table 1: The frequency of *KIR2DS4* alleles in different populations.

ALLELE	CLM (n = 93)	GBR (n = 88)	ITU (n = 102)	JPT (n = 102)	YRI (n = 105)	GLOBAL (n = 486)
KIR2DS4*00101	25.3%	21.6%	16.2%	51%	49.5%	33.1%
KIR2DS4*00103	0.5%	-	-	-	-	0.1%
KIR2DS4*00104	2.2%	-	-	-	3.9%	1.2%
KIR2DS4*00301	17.2%	25%	27.5%	5.4%	5.9%	15.9%
KIR2DS4*00401	3.2%	2.8%	6.4%	11.4%	7.4%	6.4%
KIR2DS4*00601	12.9%	18.2%	5.4%	-	8.8%	8.7%
KIR2DS4*009	-	-	-	-	2.9%	0.6%
KIR2DS4*010	12.4%	13.6%	8.8%	10.4%	2.9%	9.5%
KIR2DS4*013	-	-	-	-	2%	0.4%
KIR2DS4*016	-	-	-	1%	-	0.2%
KIR2DS4*null	25.3%	18.2%	35.3%	19.3%	13.2%	22.3%
KIR2DS4*unresolved	1.1%	0.6%	0.5%	1.5%	3.4%	1.4%

Population: Colombian (CLM); British (GBR); Telugu (ITU), Japanese (JPT), and Yoruba (YRI).

The frequency of a functional *KIR2DS4*

As described before, *KIR2DS4* presents a functional polymorphism, with two versions of the receptor: the full-length one, *KIR2DS4*-full, at the cell surface anchored to the membrane, and a truncated one, *KIR2DS4*-del, which lacks the part of D2, the transmembrane, and cytoplasmic domains, and it is possibly expressed as

a soluble molecule. We evaluated the 22bp deletion at exon 5 extracting this data from the VCF produced by kir-mapper::genotype, to survey the frequency of different genotypes among populations (Table 2).

As observed in Table 2, the frequencies of all possible genotypes are usually high in all populations. The presence of at least one functional copy of *KIR2DS4* is low in Europe (6.82%) and high in Japan (49%) and Africa (47.62%). Therefore, in Japan and Africa, *KIR2DS4* is the most frequent KIR activator gene modulating immune responses. In general, one copy of a functional *KIR2DS4* was detected in all populations, from 14.71% to 35.23%. Conversely, in South Asia, other KIR activators might be in place, because there is a high proportion of individuals with no copy of *KIR2DS4* or with non-functional copies (69.6%). Therefore, it would be necessary for future studies to analyze the mechanism of NK cells without activator KIR receptors in individuals with only *KIR2DS4*-del genotype.

Table 2: Frequency (%) of samples carrying different *KIR2DS4* genotypes in five populations from the 1000 Genomes Project.

KIR2DS4 Genotype	Copy Number	CLM (n = 93)	GBR (n = 88)	ITU (n = 102)	JPT (n = 102)	YRI (n = 105)	GLOBAL* (n = 490)
Only Full (KIR2DS4-full)	1 or 2	22.58%	6.82%	15.69%	49%	47.62%	29.19%
KIR2DS4-del KIR2DS4-full	2	23.66%	35.23%	14.71%	29.41%	29.52%	26.33%
No functional KIR2DS4 (Only Del or only *null)	-	53.76%	57.95%	69.6%	20.58%	20.95%	43.88%

Population: Colombian (CLM, America); British (GBR, Europe); Telugu (ITU, South Asia); Japanese (JPT, East Asia) and Yoruba (YRI, Africa).

Comparing Kir-mapper and PING

Figure 6 illustrates the comparative analyses of the called genotypes from kir-mapper and PING-WGS, for 490 samples from the 1000 Genomes (21). Figures S1-5 illustrated each population separately. The alleles that were compared from kir-mapper were only one combination of the alleles and PING had one combination or a list of possible alleles which were called as ambiguity.

Sometimes PING and kir-mapper report “unresolved” alleles when there are many possible allele combinations for one sample, or no allele combination could be detected. When the called alleles are the same or the kir-mapper call is among a list of possible genotypes produced by PING, they are marked as “Overlapped calls”. This represents 88.2% of the samples. There are a proportion of samples in which PING could not resolve the alleles (in blue represented by one allele unresolved, 1.4% and green with 3.9% for both alleles). Both methods disagree in about 11.8% of the samples.

The kir-mapper solved the genotypes of samples that PING failed, but we cannot address how accurate these genotypes are because there is nothing to compare with. In addition, the proportion of different calls is highly dependent on the population. The proportion of different calls was higher than 3% for African samples (YRI) (Figure S5).

Kir-mapper and PING genotypes compare (KIR2DS4) in all population

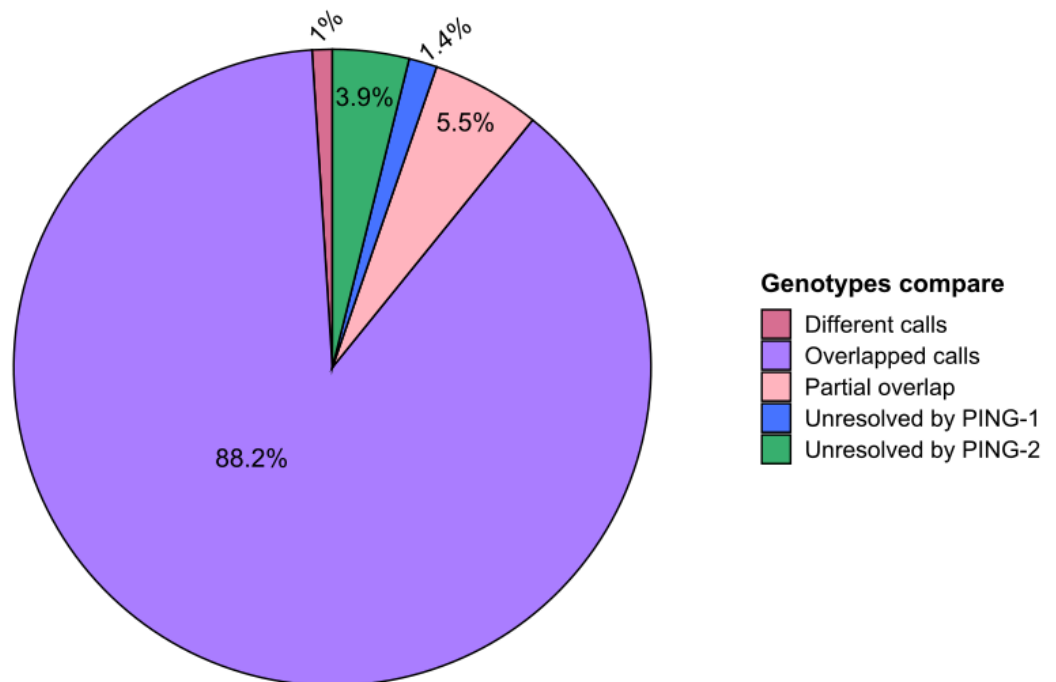


Figure 6: Comparison between the KIR2DS4 allele calls from kir-mapper and PING-WGS.

Discussion

Here, we describe and test a new computational method using a kir-mapper to evaluate *KIR2DS4* from short-read Illumina sequencing. Kir-mapper is under construction by the GeMBio group to support *KIR* analysis of a large number of samples and low computation power, and it has yet to be made available to the scientific community. All the analyses described here from Kir-mapper were conducted in a workstation with 64 GB of RAM and ten cores. Kir-mapper is multi-threaded, taking advantage of multiple cores when they are available. The time necessary for all the analyses depends highly on the sample coverage. It took 2 days to map all the samples with 30X whole-genome sequencing, 1 hour for ncopy, and 2 hours for the genotype step (with all samples).

As mentioned earlier, there is a lack of tools for *KIR* genes and sequencing data due to the diversity of the *KIR* locus, mainly due to the variability of these genes in different haplotypes, allelic polymorphism, and the combination of maternal and paternal haplotypes (35). The most used pipeline to analyze *KIR* genes is PING (20), updated last year to be compatible with whole-genomes (20). However, PING depends on a high computation power, and we could not run it on the same workstation where we performed our kir-mapper analysis. Instead, we had to run PING with just a few samples (around 100 samples) per time, which was highly time-consuming. Nevertheless, since there is a high proportion of samples with different calls between PING and kir-mapper, as described at Figure 4, particularly for African samples (FigureS3), and there are samples that failed with both methods, we support the use of different tools as T1K when dealing with *KIR* genes to compare results and get accurate genotypes. In addition, it is crucial to have another tool to be used together with PING, increasing the reliability of the called genotypes. In these cases, if different calls or partial combinations were included in a list, the best practice would be manually inspecting the outputs to decide which one is correct or use a third pipeline.

We have applied the kir-mapper workflow to characterize *KIR2DS4* in samples from 5 populations of the 1000 Genomes Project dataset, all with high coverage (30x) whole-genome sequencing data. We have selected one population per biogeographic region: Colombian (CLM) in America; British (GBR) in Europe; Yoruba (YRI) in Africa; Japanese (JPT) in East Asia; and Telugu (ITU) in South Asia. The

rationale here is to evaluate kir-mapper behavior in populations with different ancestry backgrounds, and that might present different alleles or highly divergent allele frequencies. Indeed, as observed at Figure 6 and table 3, the frequency of alleles is highly different among populations, and some of the alleles were exclusively detected in just one region. It should be noted that the genotype procedure (detection of SNPs and allele combinations) was conducted with all samples at the same time.

The complete absence of KIR2DS4 is much more frequent in South Asia (ITU, 10.8%) and America (CLM, 8.6%) (Figure 5). These frequencies are in agreement with what was observed in other populations of the same biogeographic region. For instance, the absence of KIR2DS4 occurs in about 28% of the individuals in Pakistan, 11% in Chile and 20% in Guarani Native Americans (allelefrequencies.net). This absence is uncommon in Africa (YRI, 1%) and East Asia (JPT, 2%), which was also supported by the aforementioned database. The presence and absence frequencies were statistically different ($p < 0.05$) for each population.

The copy number variation is a significant KIR polymorphism, impacting both the occurrence of a particular receptor and its expression levels. Determining copy numbers depends on a suitable reference in the genome, i.e., a gene or genomic region always presenting a fixed number of copies. Most KIR pipelines use KIR3DL3 as a reference because this gene is usually in two copies in all individuals. However, this is not always the case (data not shown), and, because of that, kir-mapper can use more than one reference to assign copy numbers. We have tested KIR3DL3 by comparing it with other genes in the genome, and, in this specific case (for these five population samples), there was no variability in KIR3DL3 copy numbers. Therefore, we used KIR3DL3 as a reference.

Copy number determination indicated that most individuals present two copies of KIR2DS4, except in South Asia (ITU) (Figure 6). KIR2DS4 in two copies is highly frequent in Africa and Europe, but not all copies correspond to a functional KIR2DS4. While in Africa 99% of the individuals present at least one copy of KIR2DS4 and most of these copies are functional (77.14%, Table 4), in Europe, 97.7% of the individuals also have at least one copy of *KIR2DS4*, but only 42.05% are functional. In fact, the presence of a KIR2DS4 copy encoding the truncated version of KIR2DS4 is highly frequent in Europe (90.9%), ranging from that to 48% in Japan. Because of

this high frequency of the KIR2DS4 truncated version, we wonder if this is truly a non-functional allele, or if this version has another function in the immune system. A higher frequency of KIR2DS4*null and KIR2DS*del might also follow a higher frequency of other KIR activators, such as KIR2DS1, but, for now, only KIR2DS4 was explored in this sample.

The alleles and genotypes observed here are close to the ones observed in other studies. Most of the individuals in Northern Ireland present KIR2DS4*00101 (35.1%); *003 (62.3%); *006 (37%); and *004 (6.5%) (36), which is similar to what was observed in Great Britain (GBR). The allelefrequencies.net database also supports the frequency observed in this study. For instance, in Ghana, the most frequent allele is *00101, followed by *006, *003, and *004, with rare copies of *009, *010, and *012. Similar results were observed for Yoruba in Africa (Table 3). Except for the functional polymorphism of 22bp del at exon 5 (the functional or truncated KIR2DS4), it is not clear whether the different alleles influence KIR2DS4 function. Nevertheless, allele association studies with diseases have revealed that KIR2DS4*00101 was associated with the severity of COVID-19 (37). This effect was particularly noticeable in patients homozygous for this allele. Conversely, KIR2DS4*00601 was associated with lower risk (37). The KIR2DS4*001 was also associated with high viral level and rapid transmission of HIV-1, with significance value ($p < 0.004$) (4).

The method presented here was able to process about 500 samples in a few days using low to medium power computers, detecting the number of copies for KIR2DS4, all the SNVs and indels across the exons, and the allele calls (the combination of SNVs and indels into haplotypes). We detected a high number of unresolved alleles in African samples, probably due the low representation of non-European samples in the IPD-KIR database. This method is straightforward and packed into a single program with multiple functionalities, called kir-mapper. The comparison with other methods, such as PING, demonstrates how important is the use of multiple tools to evaluate *KIR* genes, since there are many samples with different calls (particularly in non-European samples). For future studies, we will apply this method to evaluate *KIR2DS4* in all populations of the 1000 Genomes dataset, HGDP, and Brazilians from the SABE cohort, and integrate these findings on *KIR2DS4* studies exploring disease association, natural selection, selective pressure, and evolutionary patterns.

References

1. Falco M, Moretta L, Moretta A, Bottino C. KIR and KIR ligand polymorphism: a new area for clinical applications? *Tissue Antigens*. 2013;82(6):363–73.
2. Kennedy PR, Chazara O, Gardner L, Ivarsson MA, Farrell LE, Xiong S, et al. Activating KIR2DS4 Is Expressed by Uterine NK Cells and Contributes to Successful Pregnancy. *J Immunol Author Choice*. 1º de dezembro de 2016;197(11):4292–300.
3. Zúñiga J, Romero V, Azocar J, Terreros D, Vargas-Rojas MI, Torres-García D, et al. Protective KIR–HLA interactions for HCV infection in intravenous drug users. *Mol Immunol*. agosto de 2009;46(13):2723–7.
4. Merino AM, Dugast AS, Wilson CM, Goepfert PA, Alter G, Kaslow RA, et al. KIR2DS4 Promotes HIV-1 Pathogenesis: New Evidence from Analyses of Immunogenetic Data and Natural Killer Cell Function. Unutmaz D, organizador. *PLoS ONE*. 5 de junho de 2014;9(6):e99353.
5. Graef T, Moesta AK, Norman PJ, Abi-Rached L, Vago L, Older Aguilar AM, et al. KIR2DS4 is a product of gene conversion with KIR3DL2 that introduced specificity for HLA-A*11 while diminishing avidity for HLA-C. *J Exp Med*. 26 de outubro de 2009;206(11):2557–72.
6. Goodridge JP, Burian A, Lee N, Geraghty DE. HLA-F and MHC Class I Open Conformers Are Ligands for NK Cell Ig-like Receptors. *J Immunol*. 1º de outubro de 2013;191(7):3553–62.
7. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med*. 7 de setembro de 2000;343(10):702–9.
8. Moraes MM de. Genotipagem do repertório KIR e análise da expressão dos genes KIR (KIR2DS4) e sua interação com a variação genética do Gene HLA-C em pacientes transplantados renais com e sem episódios de rejeição aguda pós-transplante [Internet] [Tese]. [Curitiba]: Universidade Federal do Paraná, Setor de Ciências Biológicas, Programa de Pós-Graduação em Genética; 2011 [citado 4 de dezembro de 2023]. Disponível em: <https://acervodigital.ufpr.br/handle/1884/27057>
9. Vollmers S, Lobermeyer A, Körner C. The New Kid on the Block: HLA-C, a Key Regulator of Natural Killer Cells in Viral Immunity. *Cells*. 10 de novembro de 2021;10(11):3108.
10. IPD-KIR [Internet]. 2022. IPD-KIR. Disponível em: <https://www.ebi.ac.uk/ipd/kir/about/>
11. Maxwell LD, Wallace A, Middleton D, Curran MD. A common KIR2DS4 deletion variant in the human that predicts a soluble KIR molecule analogous to the KIR1D molecule observed in the rhesus monkey: Maxwell et al : A predicted soluble KIR2DS4 deletion variant. *Tissue Antigens*. setembro de 2002;60(3):254–8.
12. Middleton D, Williams F, Halfpenny IA. KIR genes. *Transpl Immunol*. agosto de 2005;14(3–4):135–42.
13. Behairy OG, El-Shimi OS, Shalan NH, Baghdady SM. Association of the killer immunoglobulin-like receptor genes KIR2DS1 and KIR2DS4 with pediatric autoimmune hepatitis type I in Egypt. *Egypt Liver J*. dezembro de 2021;11(1):75.
14. Middleton D, Gonzalez A, Gilmore PM. Studies on the expression of the deleted KIR2DS4*003 gene product and distribution of KIR2DS4 deleted and nondeleted versions in different populations. *Hum Immunol*. fevereiro de 2007;68(2):128–34.
15. Zhuang YL, Zhu CF, Zhang Y, Song YH, Wang DJ, Nie XM, et al. Association of

- KIR2DS4 and its variant KIR1D with syphilis in a Chinese Han population: Association of KIR2DS4 and syphilis. *Int J Immunogenet.* abril de 2012;39(2):114–8.
16. Gowdavally S, Tsamadou C, Platzbecker U, Sala E, Valerius T, Klein S, et al. KIR2DS4 and Its Variant KIR1D in KIR-AA Genotype Donors Showed Differential Survival Impact in Patients with Lymphoid Disease after HLA-Matched Unrelated Hematopoietic Stem Cell Transplantation. *Transplant Cell Ther.* julho de 2023;29(7):457.e1-457.e10.
 17. Dębska-Zielkowska J, Moszkowska G, Zieliński M, Zielińska H, Dukat-Mazurek A, Trzonkowski P, et al. KIR Receptors as Key Regulators of NK Cells Activity in Health and Disease. *Cells.* 14 de julho de 2021;10(7):1777.
 18. Song L, Bai G, Liu XS, Li B, Li H. Efficient and accurate KIR and HLA genotyping with massively parallel sequencing data. *Genome Res.* junho de 2023;33(6):923–31.
 19. Norman PJ, Hollenbach JA, Nemat-Gorgani N, Marin WM, Norberg SJ, Ashouri E, et al. Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing. *Am J Hum Genet.* agosto de 2016;99(2):375–91.
 20. Marin WM, Hollenbach JA. Software update: Interpreting killer-cell immunoglobulin-like receptors from whole genome sequence data with PING. *HLA.* maio de 2023;101(5):441–8.
 21. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Regier AA, Corvelo A, et al. High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios.
 22. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 15 de julho de 2009;25(14):1754–60.
 23. Robinson JT. Integrative genomics viewer. *C O Rresp O N N Ce.* 2011;29(1).
 24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 15 de agosto de 2009;25(16):2078–9.
 25. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing [Internet]. arXiv; 2012 [citado 13 de setembro de 2023]. Disponível em: <http://arxiv.org/abs/1207.3907>
 26. Picard [Internet]. Picard Toolkit. Disponível em: <https://broadinstitute.github.io/picard/>
 27. Moreau P, Contu L, Alba F, Lai S, Simoes R, Orrù S, et al. HLA-G Gene Polymorphism in Human Placentas: Possible Association of G*0106 Allele with Preeclampsia and Miscarriage. *Biol Reprod.* 1º de setembro de 2008;79(3):459–67.
 28. Olieslagers TI, Voorter CEM, Groeneweg M, Xu Y, Wieten L, Tilanus MGJ. New insights in HLA -E polymorphism by refined analysis of the full-length gene. *HLA.* março de 2017;89(3):143–9.
 29. WhatsHap [Internet]. WhatsHap. Disponível em: <https://whatshap.readthedocs.io/en/latest/guide.html>
 30. Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. *Nat Commun.* 28 de novembro de 2019;10(1):5436.
 31. ggplot2 [Internet]. Disponível em: <https://ggplot2.tidyverse.org/>
 32. The R Project for Statistical Computing [Internet]. R Software. Disponível em: <https://www.r-project.org/>
 33. Posit [Internet]. RStudio Desktop. Disponível em:

- <https://posit.co/download/rstudio-desktop/>
34. Augusto DG. The Impact of KIR Polymorphism on the Risk of Developing Cancer: Not as Strong as Imagined? *Front Genet* [Internet]. 28 de junho de 2016 [citado 7 de março de 2024];7. Disponível em:
<http://journal.frontiersin.org/Article/10.3389/fgene.2016.00121/abstract>
 35. Augusto DG. GENES KIR NO BRASIL: CARACTERIZAÇÃO DE TRÊS POPULAÇÕES. 2006;
 36. Middleton D, Meenagh A, Gourraud PA. KIR haplotype content at the allele level in 77 Northern Irish families. *Immunogenetics*. 16 de janeiro de 2007;59(2):145–58.
 37. Farias TDJ, Brugiapaglia S, Croci S, Magistrini P, Curcio C, Zguro K, et al. HLA-DPB1*13:01 associates with enhanced, and KIR2DS4*001 with diminished protection from developing severe COVID-19. *HLA*. 18 de outubro de 2023;

Supplementary

Table S1: The frequency of different genotypes regarding the number of copies of KIR2DS4 in different populations.

Number of Copies	CLM (n = 93)	GBR (n = 88)	ITU (n = 102)	JPT (n = 102)	YRI (n = 105)
0	8 (8.6%)	2 (2.27%)	11 (10.78%)	2 (1.96%)	1 (0.95%)
1	31 (33.3%)	28 (31.82%)	50 (49.02%)	35 (34.32%)	25 (23.80%)
2	54 (58.06%)	58 (65.90%)	41 (40.19%)	64 (62.75%)	76 (72.38%)
3	-	-	-	1 (0.98%)	3 (2.86%)

Divided by population: Colombian (CLM); British (GBR); Telugu (ITU); Japanese (JPT); Yoruba (YRI).

Kir-mapper and PING genotypes compare (KIR2DS4) - CLM

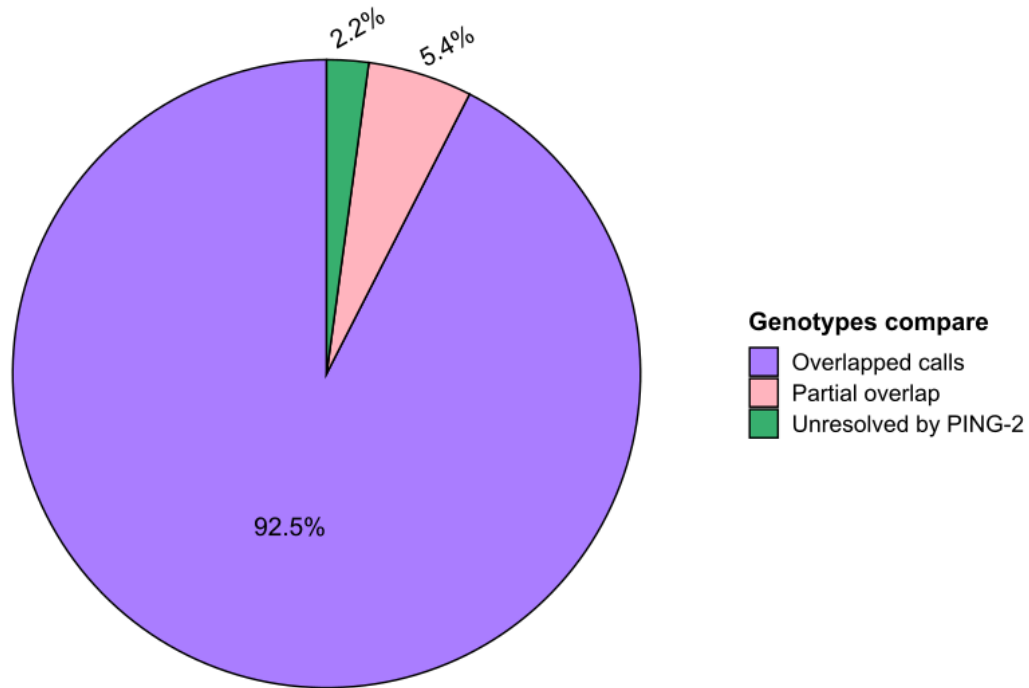


Figure S1: Kir-mapper and PING genotypes compare (KIR2DS4) for Colombia (CLM). The graphic represents in frequency (%) the proportion of similarity and differences between the two tools, PING and Kir-mapper, in *KIR2DS4* genotype in the CLM population of 1000Genomes Project (17). Each color represents a specific comparison. The green color represents that both alleles were labeled as unresolved by PING. The purple showed that the kir-mapper's combination was included in a list of combinations called by PING or the alleles were the same in both tools. The pink represented that one of the alleles was the same in kir-mapper and PING but the other wasn't. More than 90% (Overlapped calls) show that kir-mapper can be an additional tool for the analysis.

Kir-mapper and PING genotypes compare (KIR2DS4) - GBR

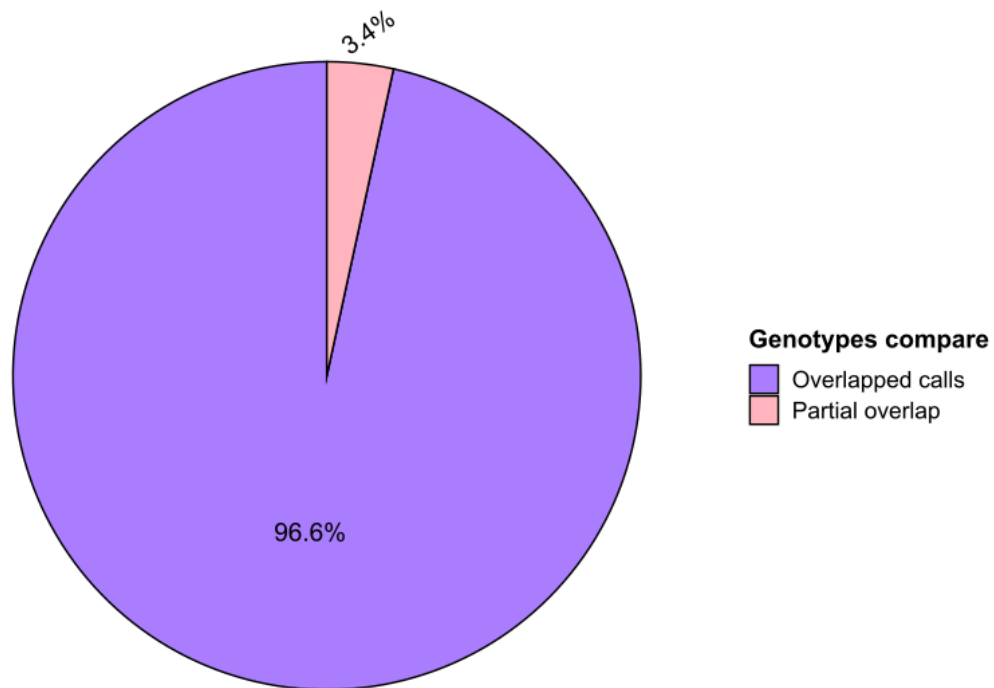


Figure S2: Kir-mapper and PING genotypes compare (KIR2DS4) for British (GBR). The graphic represents in frequency (%) the proportion of similarity and differences between the two tools, PING and Kir-mapper, in *KIR2DS4* genotype in GBR population of 1000Genomes Project (17). Each color represents a specific comparison. The purple showed that the kir-mapper's combination was included in a list of combinations called by PING or the alleles were the same in both tools. The pink represented that one of the alleles was the same in kir-mapper and PING but the other wasn't.

Kir-mapper and PING genotypes compare (KIR2DS4) - ITU

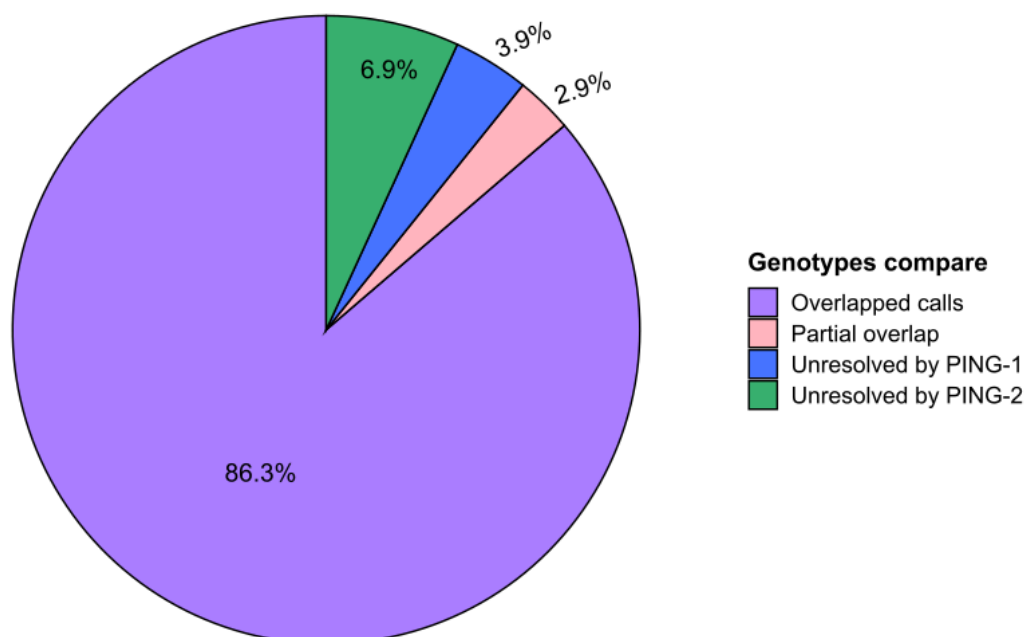


Figure S3: Kir-mapper and PING genotypes compare (KIR2DS4) for Yoruba (ITU). The graphic represents in frequency (%) the proportion of similarity and differences between the two tools, PING and Kir-mapper, in the *KIR2DS4* genotype in the ITU population of 1000Genomes Project (17). Each color represents a specific comparison. The purple shows that the kir-mapper's combination was included in a list of combinations called by PING or the alleles were the same in both tools. The pink represented that one of the alleles was the same in kir-mapper and PING but the other wasn't. The green color represents that both alleles were labeled as unresolved by PING. The blue represents that one allele by PING was identify by "unresolved".

Kir-mapper and PING genotypes compare (KIR2DS4) - JPT

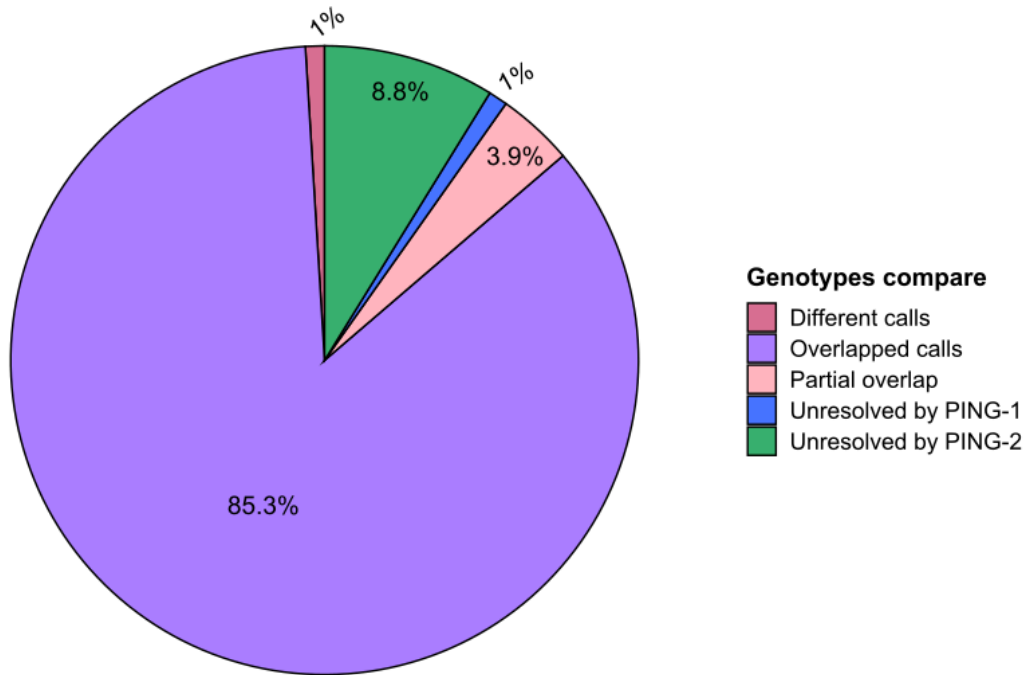


Figure S4: Kir-mapper and PING genotypes compare (KIR2DS4) for Japanese (JPT). The graphic represents in frequency (%) the proportion of similarity and differences between the two tools, PING and Kir-mapper, in the *KIR2DS4* genotype in the JPT population of 1000Genomes Project (17). Each color represents a specific comparison. The purple showed that the kir-mapper's combination was included in a list of combinations called by PING or the alleles were the same in both tools. The pink light represented that one of the alleles was the same in kir-mapper and PING but the other wasn't. The dark pink represented different alleles. The green color represents that both alleles were labeled as unresolved by PING. The blue represents that one allele by PING was identify by "unresolved".

Kir-mapper and PING genotypes compare (KIR2DS4) - YRI

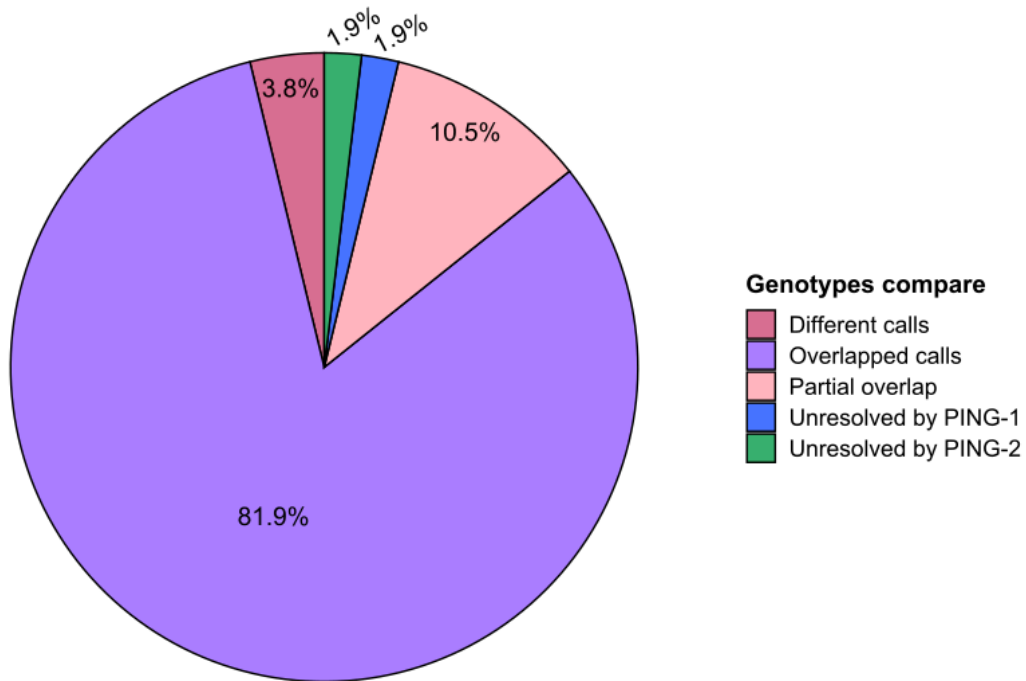


Figure S5: Kir-mapper and PING genotypes compare (KIR2DS4) for Yoruba (YRI). The graphic represents in frequency (%) the proportion of similarity and differences between the two tools, PING and Kir-mapper, in the *KIR2DS4* genotype in the YRI population of 1000Genomes Project (17). Each color represents a specific comparison. The purple showed that the kir-mapper's combination was included in a list of combinations called by PING or the alleles were the same in both tools. The pink light represented that one of the alleles was the same in kir-mapper and PING but the other wasn't. The dark pink represented different alleles. The green color represents that both alleles were labeled as unresolved by PING, while kir-mapper showed, for the same sample, an allele defined and another as unresolved. The blue color means that kir-mapper showed a combination of alleles defined, and PING showed unresolved both. The purple showed that the kir-mapper's combination was included in a list of combinations called by PING. More than 90% (identical calls + kir-mapper calls included in a list of PING) show that kir-mapper can be an additional tool for the analysis. The green color represents that both alleles were labeled as unresolved by PING. The blue represents that one allele by PING was identify by "unresolved".

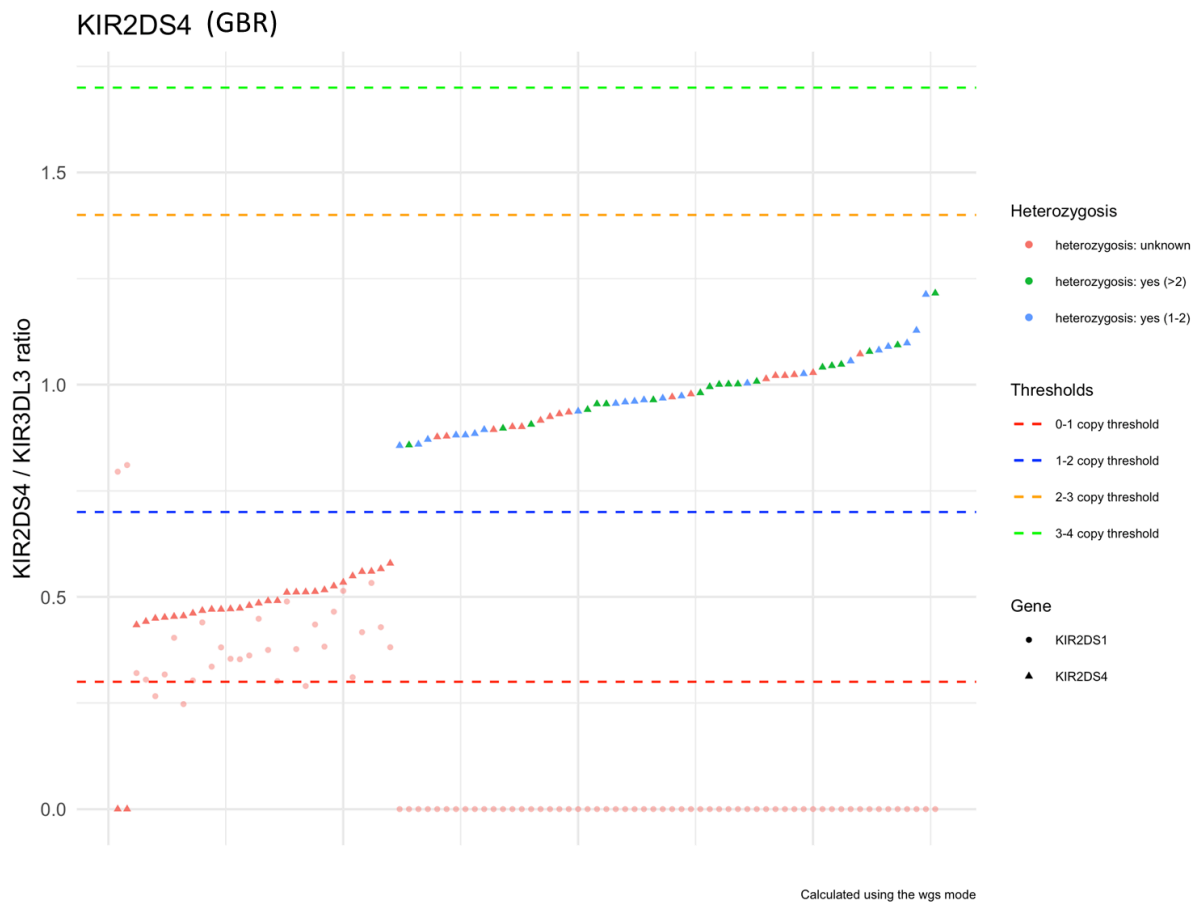


Figure S6: The *KIR2DS4* copy number determination using `kir-mapper::ncopy` for GBR. Each triangle represents a sample. The y-axis represents the ration between the coverage observed in *KIR2DS4* and *KIR3DL3*. Because *KIR2DS1* and *KIR2DS4* are usually in the same position in the locus, we also plotted the ratios for *KIR2DS1* (light dots). The lines red, blue, yellow and green defined the thresholds to separate the number of copies.

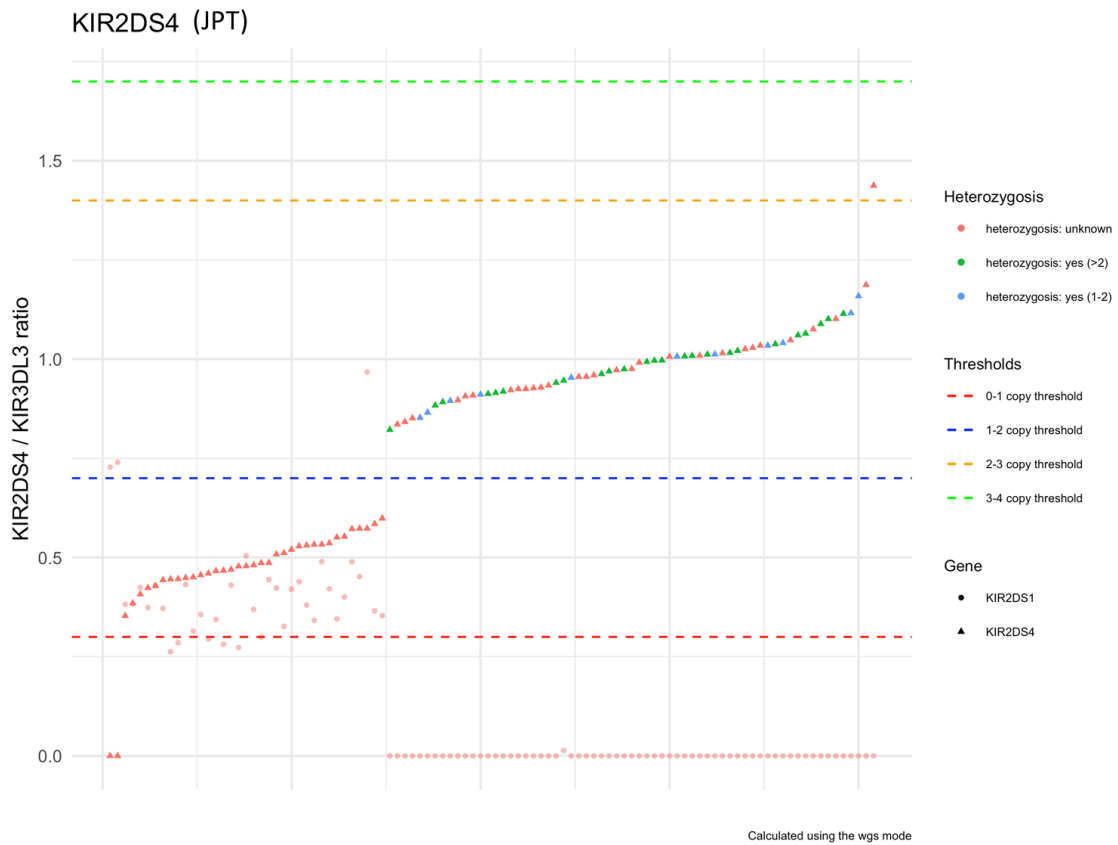


Figure S7: The *KIR2DS4* copy number determination using `kir-mapper::ncopy` for JPT. Each triangle represents a sample. The y-axis represents the ratio between the coverage observed in *KIR2DS4* and *KIR3DL3*. Because *KIR2DS1* and *KIR2DS4* are usually in the same position in the locus, we also plotted the ratios for *KIR2DS1* (light dots). The lines red, blue, yellow and green defined the thresholds to separate the number of copies.

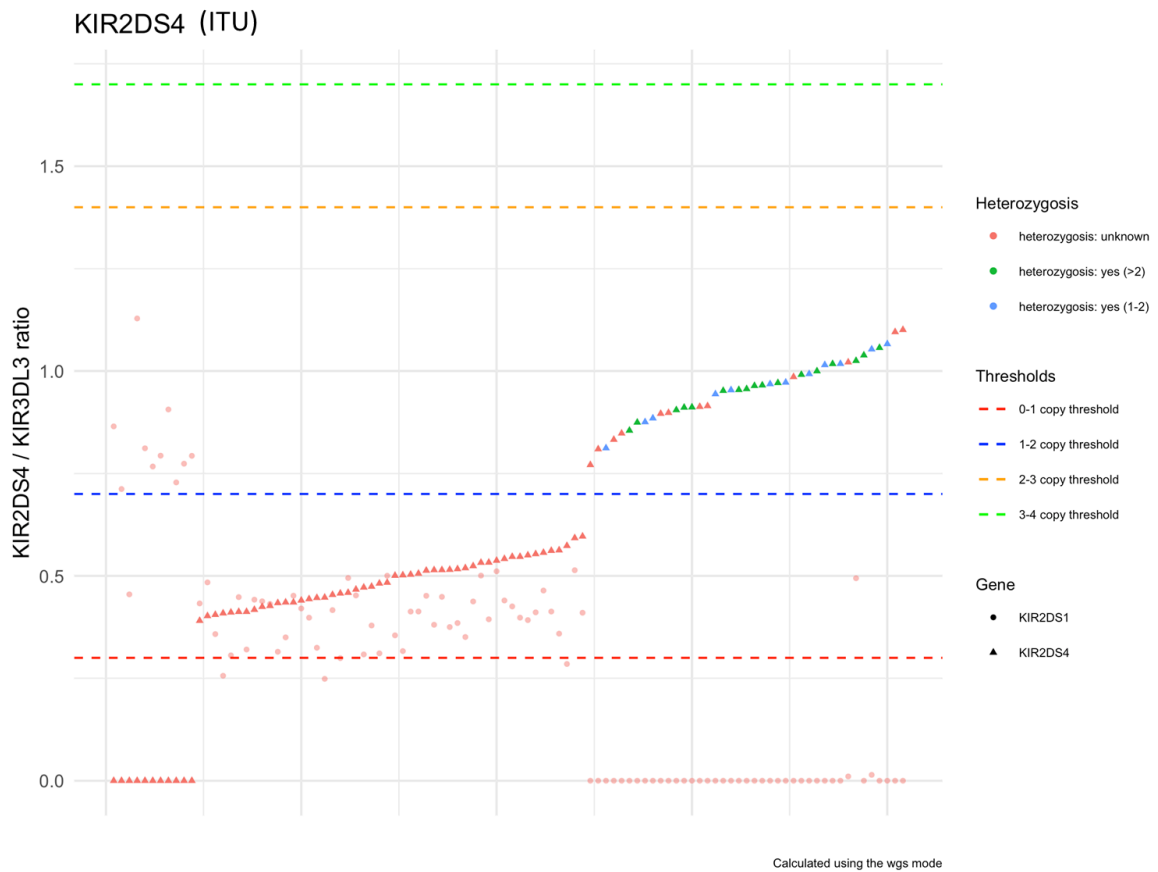


Figure S8: The *KIR2DS4* copy number determination using `kir-mapper::ncopy` for ITU. Each triangle represents a sample. The y-axis represents the ration between the coverage observed in *KIR2DS4* and *KIR3DL3*. Because *KIR2DS1* and *KIR2DS4* are usually in the same position in the locus, we also plotted the ratios for *KIR2DS1* (light dots). The lines red, blue, yellow and green defined the thresholds to separate the number of copies.

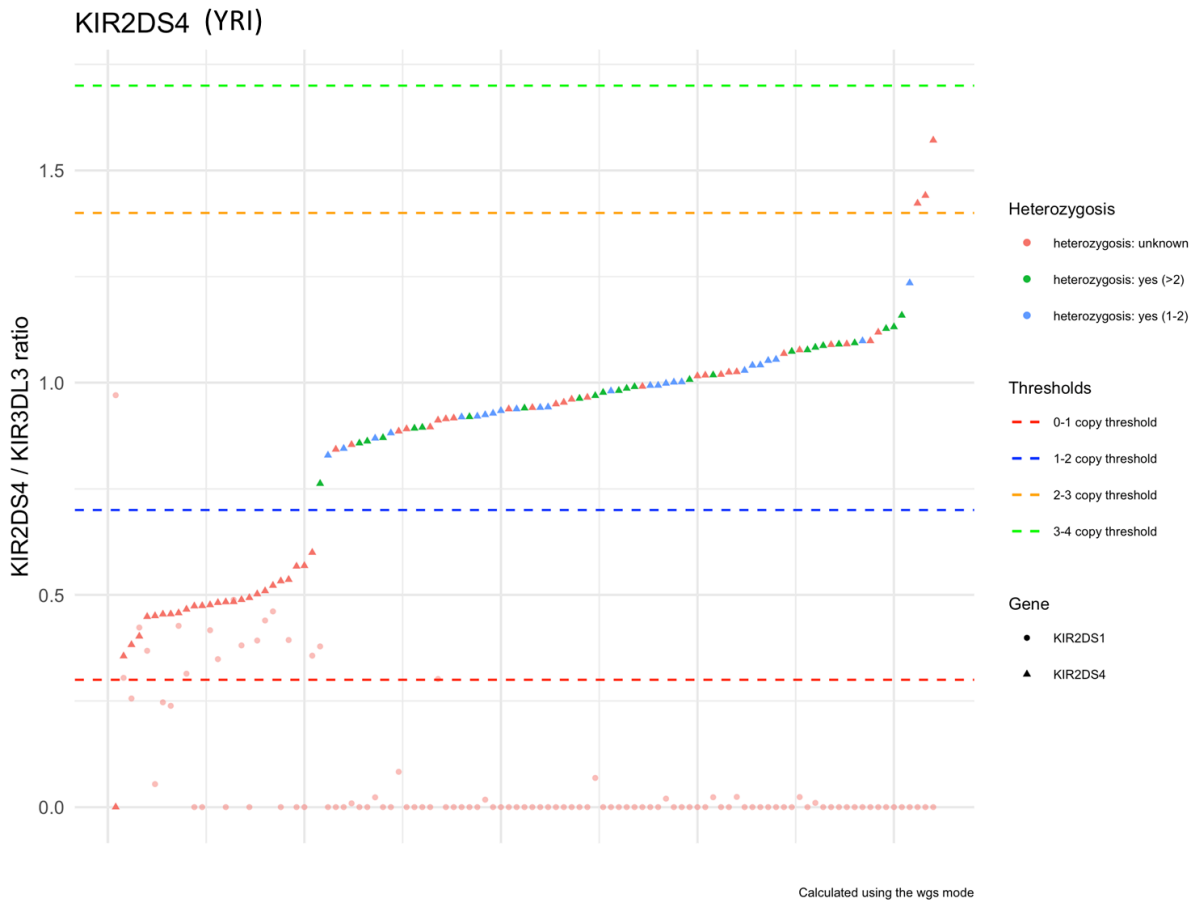


Figure S9: The *KIR2DS4* copy number determination using `kir-mapper::ncopy` for YRI. Each triangle represents a sample. The y-axis represents the ration between the coverage observed in *KIR2DS4* and *KIR3DL3*. Because *KIR2DS1* and *KIR2DS4* are usually in the same position in the locus, we also plotted the ratios for *KIR2DS1* (light dots). The lines red, blue, yellow and green defined the thresholds to separate the number of copies.

Number of Copies of *KIR2DS4* (kir-mapper)

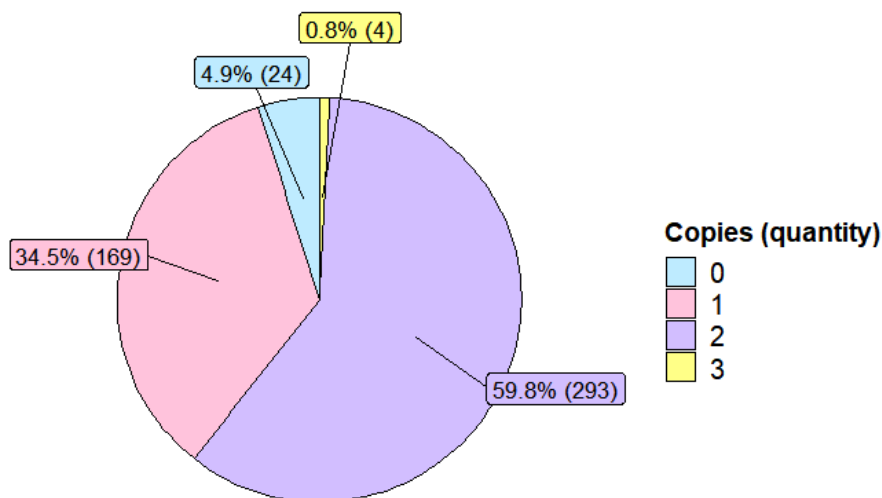


Figure S10: Number of copies of *KIR2DS4* in all the samples. The graphic represents frequency (%) and the quantity of the samples of each population of 1000Genomes (17), CLM, GBR, JPT, ITU and

YRI. Each color represents a specific quantity. The blue represents the complete absence of *KIR2DS4*. The pink one showed that one copy. The purple presents the two copies of *KIR2DS4* and the yellow, 3 copies. Population: Colombian (CLM); British (GBR); Telugu (ITU); Japanese (JPT); Yoruba (YRI).

CONCLUSÃO

Em conclusão, esta dissertação investigou a diversidade genética do gene *KIR2DS4* em um subgrupo amostral do projeto 1000 Genomas (17). Observamos que *KIR2DS4* está presente em todas as populações estudadas, porém destacou-se maiores frequências da ausência do gene no Sul da Ásia (10.8%) e na América (8.6%). Duplicação de *KIR2DS4* em um mesmo cromossomo é um evento raro, com apenas 4 exemplos (3 na África e 1 no Japão). *KIR2DS4* se mostrou bastante polimórfico, com 12 diferentes alelos. No entanto, cerca de 43.88% das amostras não apresentaram a versão funcional do gene, sugerindo que há possivelmente um mecanismo compensatório em indivíduos que não apresentam *KIR2DS4* como receptor ativador dentre os KIR.

Os resultados obtidos podem ser utilizados em futuras pesquisas, incluindo análise de associação com doenças, estudos funcionais para avaliar o mecanismo das células NK, gestação e implementação de SNPs para painéis de referência de imputação. Além disso, a metodologia aplicada neste estudo, complementar a outras utilizadas como o PING para análise dos genes *KIR*, demonstrou fornecer informações de alta confiabilidade acerca da diversidade dos genes.