



Ministério da Saúde
Instituto Nacional de Câncer
Coordenação de Pós-graduação *Stricto sensu*

INSTITUTO NACIONAL DE CÂNCER

Pós-Graduação em Oncologia

ISABEL MARIA PRELLWITZ

**Determinação dos alelos de HLA-A, B e C por NGS e predição de
afinidade a epítomos do HIV-1 de pacientes de diferentes regiões do Brasil**

Orientadora: Dra. Esmeralda Augusta Jardim Machado Soares

Co-orientador: Dr. Marcelo Alves Soares

RIO DE JANEIRO

2019



Ministério da Saúde
Instituto Nacional de Câncer
Coordenação de Pós-graduação *Stricto sensu*

INSTITUTO NACIONAL DE CÂNCER

Pós-Graduação em Oncologia

ISABEL MARIA PRELLWITZ

Determinação dos alelos de HLA-A, B e C por NGS e predição de afinidade a epítomos do HIV-1 de pacientes de diferentes regiões do Brasil

Tese apresentada ao Instituto Nacional de Câncer como parte dos requisitos para obtenção do título de Doutor em Oncologia

Orientadora: Dra. Esmeralda Augusta Jardim Machado Soares

Co-orientador: Dr. Marcelo Alves Soares

RIO DE JANEIRO

2019



**Ministério da Saúde
Instituto Nacional de Câncer
Coordenação de Pós-graduação *Stricto sensu***

**INSTITUTO NACIONAL DE CÂNCER
Pós-Graduação em Oncologia**

AUTORA: Isabel Maria Prellwitz

**DETERMINAÇÃO DOS ALELOS DE HLA-A, B E C POR NGS E PREDIÇÃO
DE AFINIDADE A EPÍTOPOS DO HIV-1 DE PACIENTES DE DIFERENTES
REGIÕES DO BRASIL**

ORIENTADORA: Dra. Esmeralda Augusta Jardim Machado Soares

CO-ORIENTADOR: Dr. Marcelo Alves Soares

EXAMINADORES:

Prof. Dr.

Prof. Dr.

Prof. Dr.

Prof. Dr.

Prof. Dr.

Prof. Dr.

RIO DE JANEIRO

2019

DEDICATÓRIA

**Aos meus queridos
filhos Alice e Luís.**

AGRADECIMENTOS

Nada disso seria possível sem a ajuda de três pessoas muito importantes. Elas doaram seu tempo, se dedicaram e muitas vezes me socorreram. Só vocês conhecem todos os sacrifícios feitos ao longo do caminho. Francisco, Brunna e Juliana, a vocês eu agradeço do início ao fim desse percurso.

Francisco, meu companheiro, amigo e porto seguro. Obrigada por não me deixar desistir e ao mesmo tempo me deixar segura de que, tudo bem se eu não conseguir... Todo o meu esforço se baseou na sua confiança em mim e a sua dedicação para com a nossa família sempre foi o meu maior exemplo de perseverança.

Brunna, amiga querida. Dividimos uma grande amizade, a tese e tudo mais desde a faculdade (até a Lilice!). Sou eternamente grata por ter estado presente em todos experimentos e em todos momentos da minha vida. Não poderia desejar amiga melhor.

Juliana, quem tem a sorte de ter você em sua vida ganhou um enorme tesouro. E que sorte a minha! Com toda a sua boa vontade e enorme paciência eu fui capaz de crescer e desenvolver. Para cada momento de desespero nessa tese você me ajudou a achar a solução! Obrigada por todo o carinho e amizade.

Agradeço enormemente aos meus orientadores Esmeralda e Marcelo Soares por todos esses anos de parceria. Agradeço a confiança em mim e a enorme paciência com todas as minhas gestações ao longo desses anos. Obrigada pelo respeito à minha dedicação a família e por me dar a oportunidade de conciliar os dois trabalhos (maternidade e doutorado). Esmeralda, sua orientação e amizade foram de grande importância na minha formação profissional e pessoal. Marcelo, tenho grande admiração pelo seu incansável trabalho de pesquisador e educador. Obrigada por sempre me desafiar e me incentivar a buscar o melhor.

Agradeço ao Dr. Hector Seuáñez, chefe do Programa de Genética do INCA, e a todos os pesquisadores e funcionários por todo o apoio técnico e

insumos do laboratório que me permitiram realizar todos esses experimentos.

Meninas do grupo de Oncovirologia, em especial Mariana, Ornella, Mirela, Gislaine, Laura e Gisele, obrigada pela companhia sempre divertida e produtiva ao longo desses anos.

Agradeço à minha família e amigos que me ajudaram, seja me incentivando ou como babá das crianças. Meus pais, Lulu, Ana Clara, Martinha, Goretti, Auxília, Espe e tantos mais, muito obrigada por tudo.

Agradeço aos pacientes que generosamente participaram desse estudo e que espero que futuramente recebam os frutos da pesquisa.

Por fim, agradeço as agências Capes, CNPq, FAPERJ e INCT do Câncer pelo apoio financeiro.

Determinação dos alelos de HLA-A, B e C por NGS e predição de afinidade a epítomos do HIV-1 de pacientes de diferentes regiões do Brasil

RESUMO –TESE DE DOUTORADO

Isabel Maria Prellwitz

A erradicação do HIV ainda é um grande desafio e um grande número de pessoas infectadas pelo HIV está sob terapia antirretroviral. Frente a isso, novas estratégias estão surgindo com base na vacinação terapêutica de indivíduos infectados pelo HIV, provocando respostas imunes que podem ajudar a controlar a replicação do HIV após a interrupção do tratamento. A capacidade de controlar a infecção pelo HIV tem sido correlacionada com certos alelos do antígeno leucocitário humano (HLA). Na presente proposta, determinaremos a composição dos alelos HLA classe I com resolução ultra profunda por NGS e preveremos a afinidade dos alelos mais frequentes encontrados aos epítomos conservados do HIV-1. Quarenta e quatro adultos HIV+ foram selecionados no Rio de Janeiro e 40 no Rio Grande do Sul. Os critérios de inclusão foram idade ≥ 18 anos, estar sob HAART de primeira linha e com carga viral indetectável de HIV por pelo menos 12 meses. Uma amostra periférica de sangue total foi coletada e o DNA genômico foi extraído. Amplificações separadas dos loci HLA-A, B e C foram realizadas e reunidas para a construção das bibliotecas genômicas usando o kit Nextera XT e sequenciadas na plataforma Illumina MiSeq. Os alelos HLA foram tipados usando o algoritmo HLA-Twin. Os alelos HLA indicativos de serem novos foram analisados manualmente usando um pipeline desenvolvido internamente. A afinidade entre a fenda de ligação do HLA e os peptídeos do HIV-1 foi prevista usando o banco de dados de epítomos imunogênicos considerando apenas aqueles com $IC_{50} < 50nM$. Todos os 84 pacientes foram coletados e tiveram seus alelos HLA amplificados, sequenciados e analisados. Todos os pacientes tiveram pelo menos um alelo HLA tipado com sucesso, totalizando 482 alelos tipados. Dezesseis pacientes apresentaram homozigose em um locus. Os alelos mais frequentes no Rio de Janeiro foram A*02:01:01 (22%), B*07:02:01 (12%), C*4:01:01 (16%) e no Rio Grande do Sul foram A*01:01:01:01 (19%), B*08:01:01:01 e B*14:02:01:01 (8%) e C*03:03:01:01 (8%). Oito alelos apresentaram evidências de serem novos e precisam ser confirmados. Encontramos quatro epítomos candidatos (KARVLAEAM, EMMTACQGV, MGGIGGFI, VGSLQYLAL) localizados em regiões conservadas (Gag, Pol e Vif), com alta afinidade de ligação aos alelos de HLA-A, B e C mais frequentes na população estudada e específicos de cada paciente. A análise de variabilidade dos epítomos virais identificou 22 epítomos variantes com frequência superior a 1%, entrando apenas seis dessas variantes mostraram uma menor afinidade a apenas um dos seis alelos dos pacientes. A combinação da composição dos alelos HLA e a afinidade para os epítomos do HIV-1 que são restritos intrapacientes permitirá a prova de princípios para o desenho de vacinas terapêuticas personalizadas contra o HIV.



Ministério da Saúde

Instituto Nacional de Câncer

Coordenação de Pós-graduação *Stricto sensu*

Determination of HLA-A, B and C alleles by NGS and affinity prediction to HIV-1 epitopes of patients from different regions of Brazil

ABSTRACT - EXAME - PhD

Isabel Maria Prellwitz

Eradication of HIV is still a major challenge and large numbers of people infected with HIV are on antiretroviral therapy. In light of this, new strategies are emerging based on the therapeutic vaccination of HIV-infected individuals, inducing immune responses that may help control HIV replication after treatment discontinuation. The ability to control HIV infection has been correlated with certain human leukocyte antigen (HLA) alleles. In the present proposal, we will determine by NGS the composition of HLA class I alleles with ultra-deep resolution and predict the affinity of the most frequent alleles found with conserved HIV-1 epitopes. Forty-four HIV+ adults were selected from Rio de Janeiro and 40 from Rio Grande do Sul. The inclusion criteria were age ≥ 18 years, being under first-line HAART and with undetectable HIV viral load for at least 12 months. A peripheral whole blood sample was collected and genomic DNA was extracted. Separate amplifications of the HLA-A, B and C loci were performed and pooled to construct genomic libraries using the Nextera XT kit and sequenced on the Illumina MiSeq platform. HLA alleles were typed using the HLA-Twin algorithm. HLA alleles indicative of being novel were analyzed manually using a pipeline developed in-house. The affinity between the HLA groove and the HIV-1 peptides was predicted using the immunogenic epitope database considering only those with $IC_{50} < 50nM$. All 84 patients were collected and had their HLA alleles amplified, sequenced and analyzed. All patients had at least one successfully typed HLA allele, totaling 482 typed alleles. Sixteen patients had homozygosity at one *loci*. The most frequent alleles from Rio de Janeiro were A*02:01:01 (22%), B*07:02:01 (12%), C*04:01:01 (16%) and from Rio Grande do Sul were A*01:01:01:01 (19%), B*08:01:01:01 and B*14:02:01:01 (8%) and C*03:03:01:01 (8%). Eight alleles showed evidence of being novel and need to be confirmed. We found four candidate epitopes (KARVLAEAM, EMMTACQGV, MIGGIGGFI, VGSLQYLAL) located in conserved regions (Gag, Pol and Vif), with high affinity binding to the most frequent HLA-A, B and C alleles in the study population and to the specific alleles from each patient. Viral epitope variability analysis identified 22 variant epitopes with a frequency greater than 1%, however only six of these variants had lower affinity for only one of the six patient alleles. The combination of the composition of HLA alleles and the affinity to HIV-1 epitopes that are restricted intrapatient will provide proof of principles for the design of personalized therapeutic vaccines against HIV.

Sumário

LISTA DE TABELAS	xii
LISTA DE FIGURAS	xiii
LISTA DE ABREVIATURAS	xiv
1. INTRODUÇÃO	18
1.1. Complexo MHC e Sistema HLA	18
1.2. HLA de classe I	19
1.3. HLA de classe II e III	23
1.4. Processamento e apresentação de antígeno via HLA de classe I	25
1.5. Polimorfismos da molécula de HLA de classe I	26
1.6. Nomenclatura do Sistema HLA	28
1.7. Tipagem do HLA por sequenciamento de nova geração	29
1.8. Função do HLA de classe I e sua associação com o sistema imune	32
1.9. O papel do HLA de classe I na infecção pelo HIV-1	34
1.10. Epidemiologia da infecção pelo HIV-1 e AIDS	36
1.11. História natural da infecção pelo HIV-1	38
1.12. Partícula viral, estrutura genômica do HIV-1 e seus produtos proteicos	41
1.13. Ciclo replicativo do HIV-1 e terapia antirretroviral	45
1.14. Novas estratégias contra o HIV-1	49
a) Vacina e cura do HIV-1	50
b) Vacinação terapêutica	51
2. JUSTIFICATIVA	53
3. OBJETIVOS	54
3.1. Objetivos específicos	54
4. MATERIAL E MÉTODOS	55
4.1. Colaborações	55

4.2.	Casuística	55
4.3.	Extração do DNA genômico dos pacientes	57
4.4.	Amplificação dos genes de HLA-A, B e C	58
4.5.	Purificação dos produtos de PCR	59
4.6.	Preparação das bibliotecas genômicas	60
4.7.	Quantificação das bibliotecas	62
4.8.	Sequenciamento de nova geração na plataforma MISEQ ILLUMINA	63
4.9.	Análise de qualidade dos <i>reads</i>	65
4.10.	Tipagem do HLA pelo software <i>HLA TWIN</i>	65
4.11.	Análise de alelos novos do HLA-A, B e C	66
a)	Filtragem dos <i>reads</i>	66
b)	Reconstrução dos alelos através de montagem com referência	66
4.12.	Comparação das frequências dos HLA-A, B e C encontrados com a população de doadores de medula óssea cadastrados no REDOME	67
4.13.	Análise de predição de epítomos de células T e de diversidade viral intra-hospedeiro	67
a)	Predição de epítomos de células T	68
b)	Afinidade dos epítomos aos HLA-A, B e C específicos	69
c)	Variação dos epítomos intrapaciente	69
5.	RESULTADOS	70
5.1.	Casuística e dados epidemiológicos	70
5.2.	Amplificação e sequenciamento dos genes de HLA-A, B e C	74
5.3.	Tipagem do HLA-A, B e C e frequência dos alelos	76
5.4.	Comparação das frequências alélicas entre as localidades de estudo e com o banco de dados do redome	85
5.5.	Descoberta e confirmação de alelos novos	86
5.6.	Análise de predição de epítomos de células T e de diversidade viral intra-hospedeiro	89

6. DISCUSSÃO	95
7. CONCLUSÕES	105
8. BIBLIOGRAFIA	106

LISTA DE TABELAS

Tabela 4.1 Descrição dos iniciadores usados para a amplificação dos alelos de HLA-A, B e C.	58
Tabela 5.1 Características clínicas e epidemiológicas retiradas dos prontuários dos pacientes HIV+ incluídos no estudo.	72
Tabela 5.2. Esquema terapêutico dos antirretrovirais utilizados pelos pacientes HIV+ incluídos no estudo.	73
Tabela 5.3 Análise dos reads gerados do sequenciamento dos genes de HLA-A, B e C das amostras estudadas na plataforma MiSeq.	74
Tabela 5.4 Tipagem dos alelos de HLA-A, B e C dos pacientes do HI.	77
Tabela 5.5 Tipagem dos alelos de HLA-A, B e C dos pacientes do HUCFF-UFRJ.....	78
Tabela 5.6 Tipagem dos alelos de HLA-A, B e C dos pacientes do HUFURG.	79
Tabela 5.7 Análises estatísticas de comparação de frequências alélicas entre as casuísticas estudadas e do REDOME	86
Tabela 5.8 Descrição dos possíveis alelos novos encontrados durante a tipagem dos HLA-A, B e C dos pacientes do estudo.....	87
Tabela 5.9 Representação dos 9 pacientes do HI que apresentaram ao menos um alelo de HLA-A, B e C sem alta afinidade aos epítomos virais de Gag.	90
Tabela 5.10 Variantes dos epítomos virais de Gag (KARVLAEAM e EMMTACQGV) e resultado de predição de afinidade aos alelos de HLA específicos desses mesmos pacientes.	91
Tabela 5.11 Representação da predição de afinidade do epítopo viral MIGGIGGFI, da região da Pol do HIV, aos alelos de HLA-A, B e C.....	92
Tabela 5.12 Variantes do epítopo viral MIGGIGGFI, da região da Pol do HIV, e resultado de predição de afinidade aos alelos de HLA-A, B e C específicos desses mesmos pacientes.....	93
Tabela 5.13 Representação da predição de afinidade do epítopo viral VGSLQYLAL, da região Vif do HIV, aos alelos de HLA-A, B e C	94
Tabela 5.14 Variantes do epítopo viral VGSLQYLAL, da região de Vif, e resultado de predição de afinidade aos alelos de HLA-A, B e C específicos desses mesmos pacientes.....	95

LISTA DE FIGURAS

Figura 1.1 Mapa da região HLA, banda 6p21.3 do cromossomo 6.....	19
Figura 1.2 Estrutura da molécula de HLA classe I.....	21
Figura 1.3 Fenda de ligação a peptídeo das moléculas de HLA classe I.	22
Figura 1.4 Estrutura da molécula de HLA de classe II.....	24
Figura 1.5 Nomeclatura do Sistema HLA	29
Figura 1.6 História natural da infecção pelo HIV-1 em adultos	39
Figura 1.7 Partícula viral do HIV-1.....	42
Figura 1.8 Organização genômica do HIV-1	43
Figura 1.9 Ciclo replicativo do HIV-1 e atuação das classes de inibidores.....	46
Figura 4.1 Representação esquemática das etapas de preparação das bibliotecas de DNA.....	62
Figura 5.1 Proporção de heterozigotos e homozigotos encontrada para todos os pacientes do Rio de Janeiro	80
Figura 5.2 Proporção de heterozigotos e homozigotos encontrada para todos os pacientes do Rio Grande	81
Figura 5.3 Alelos de HLA-A encontrados nos pacientes do Rio de Janeiro com frequência maior que 1% (n = 86).	82
Figura 5.4 Alelos de HLA-B encontrados nos pacientes do Rio de Janeiro com frequência maior que 1% (n = 81).	82
Figura 5.5 Alelos de HLA-C encontrados nos pacientes do Rio de Janeiro com frequência maior que 1% (n=86).	83
Figura 5.6 Alelos de HLA-A encontrados nos pacientes do Rio Grande com frequência maior que 1% (n = 78).	83
Figura 5.7 Alelos de HLA-B encontrados nos pacientes do Rio Grande com frequência maior que 1% (n = 74).	84
Figura 5.8 Alelos de HLA-C encontrados nos pacientes do Rio Grande com frequência maior que 1% (n = 77).	84

LISTA DE ABREVIATURAS

°C – graus Celsius

3TC - lamivudina

3TC – lamivudina

AC - antagonista de CCR5

ATV – atazanavir

AZT – zidovudina

CCR5 – receptor de quimiocina tipo 5

CD4+ - do inglês *cluster of differentiation 4* (grupo de diferenciação 4)

CD8+ - do inglês *cluster of differentiation 8* (grupo de diferenciação 8)

CDC – Centro de Controle e Prevenção de Doenças dos Estados Unidos, do inglês *Center for Disease Control and Prevention*

cDNA – DNA complementar

CTL – linfócitos T CD8+ citotóxicos

CXCR4 – receptor de quimiocina tipo 4

DNA – ácido desoxirribonucleico

dNTP – trifosfato de desoxirribonucleotídeo

DRV – darunavir

DTG – dolutegravir

EFV – efavirenz

Env – envelope

ERAP1 – do inglês *endoplasmic reticulum aminopeptidase-1*

FTC - emtricitabina

GALT - do inglês *gastrointestinal associated lymphoid tissues* (tecido linfóide associado à mucosa gastrointestinal)

H2 - H para histocompatibilidade e 2 para o antígeno II.

HAART - do inglês *Highly Active AntiRetroviral Therapy* (terapia antirretroviral altamente eficaz)

HI - Hospital Federal de Ipanema

HIV-1 - vírus da imunodeficiência humana do tipo 1

HLA - do inglês *Human Leukocyte Antigen* (Antígeno Leucocitário Humano)

HSH - homens que fazem sexo com homens

HUCFF - Hospital Universitário Clementino Fraga Filho

HU-FURG - Hospital Universitário Dr. Miguel Riet Corrêa Jr da Universidade Federal do Rio Grande

ICTV - do inglês *International Committee on Taxonomy of Viruses*

IEDB - do inglês *Immune Epitope Database Analysis Resource*

IF - inibidores de fusão

IHIW – do inglês *International HLA and Immunogenetics Workshops*

II - inibidores de integrase

IL-2 – interleucina 2

IMGT – do inglês *international ImMunoGeneTics project*

IN – integrase

INCA – Instituto Nacional de Câncer José Alencar Gomes da Silva

INNTR - inibidores não-nucleosídicos da transcriptase reversa

INTR - inibidores nucleosídicos da transcriptase reversa

IP - inibidores de protease

Kb - quilobase

kDa – unidade de massa atômica

LAV - do inglês *Lymphadenopathy-Associated Virus*

LPV – lopinavir

LTNP - do inglês *long-term nonprogressors*

LTRs - repetições terminais longas

M - molar

mAmp – miliamperes

MgCl₂ – cloreto de Magnésio

MHC - do inglês *Major Histocompatibility Complex* (Complexo Maior de Histocompatibilidade)

min – minuto

ml – mililitro

mM – milimolar

NaOH – hidróxido de sódio

Nef – do inglês *Negative Regulatory Factor*

NF-κB – fator nuclear kappa B

NGS – do inglês *Next Generation Sequencing* (sequenciamento de nova geração)

NIH – do inglês *National Institute of Health*

NK – do inglês *natural killer*
nm – nanômetro
nM - nanomol
NVP – nevirapina
ng – nanograma
OH – hidroxila
OMS – Organização Mundial da Saúde
pb – pares de base
PBMC – do inglês *peripheral blood mononuclear cell* (células mononucleares do sangue periférico)
PCR - reação em cadeia da polimerase
PCR-SSO – do inglês *sequence-specific oligonucleotide probes*
PCR-SSP - *sequence-specific primers*
PLC - complexo de carregamento de peptídeo
pM - picomol
POL – DNA polimerase
PR – protease
RE - retículo endoplasmático
REDOME - Registro Nacional de Doadores Voluntários de Medula Óssea
RH - RNase H
RJ – Rio de Janeiro
RNA – ácido ribonucléico
RNAm - RNA mensageiro
rpm – rotação por minuto
RS – Rio Grande do Sul
RT – transcriptase reversa
RTV – ritonavir
SBT - do inglês *sequence-based typing*
SUS - Sistema Único de Saúde
TAP – transportador associado ao processamento de antígeno
Tat – proteína transativadora de transcrição
TCLE - Termo de Consentimento Livre e Esclarecido
TCR – receptor de célula T
TDF – tenofovir

TDF – tenofovir

TM – do inglês *Melting Temperature*

TNF – fator de necrose tumoral

Tregs - linfócitos T reguladores

UFRJ – Universidade Federal do Rio de Janeiro

UNAIDS – do inglês *Joint United Nations Programme on HIV/AIDS* (Programa das Nações Unidas para HIV/Aids)

UPS - sistema de ubiquitina-proteassoma

UTR – do inglês *untranslated region*

Vpr – proteína viral R

Vpu – proteína viral U

ng – nanograma

µg – micrograma

µL – microlitro

pmol – picomol

Ω – Ohm

1. INTRODUÇÃO

1.1. COMPLEXO MHC E SISTEMA HLA

Em 1937, Peter Gorer realizou estudos com transplantes de camundongos descobrindo que tumores poderiam ser transplantados entre camundongos que compartilhavam certos alelos, que ele chamou de Antígeno II, enquanto que camundongos que não possuíam os mesmos alelos do Antígeno II rejeitavam o transplante (GORER e SCHUTZE, 1938). Nesse mesmo período, George Snell também descobriu o mesmo sistema ao realizar experimentos com camundongos e observar que o transplante de tumores era controlado por alguns genes, chamando-o de Histocompatibilidade Maior ou locus H (SNELL, 1948). Por terem identificado o mesmo sistema, em colaboração, George Snell o chamou de sistema H-2 (H para histocompatibilidade e 2 para o antígeno II). Esses foram os primeiros relatos da descoberta do Complexo Principal de Histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*). A designação MHC só veio a ser introduzida no início da década de 1970, quando se tornou conhecido que sistemas geneticamente homólogos ao H2 existiam em muitos outros vertebrados (BJORKMAN *et al.*, 1987b).

Em 1958, Jean Dausset, por meio de um estudo com transfusões seriadas, concluiu que pessoas submetidas a transfusões sanguíneas são capazes de aglutinar leucócitos de outras pessoas, descobrindo o primeiro antígeno humano, o HLA-A*02, na época denominado MAC (DAUSSET, 1984).

O sistema de Antígeno Leucocitário Humano (HLA, do inglês *Human Leukocyte Antigen*) é o correspondente humano do MHC. Sua função se baseia na apresentação de epítopos na superfície celular para serem reconhecidos por células T do sistema imune. Portanto, as moléculas de HLA são um elemento central na ativação e resposta do sistema imune à substâncias estranhas através da diferenciação entre próprio e não-próprio. (MCDEVITT, 2000). Dessa forma, o sistema HLA influencia na rejeição de transplantes, na predisposição a um amplo espectro de doenças crônicas e

autoimunes e na suscetibilidade a doenças infecciosas como a infecção pelo vírus da imunodeficiência humana do tipo 1 (HIV-1), tema abordado nesse trabalho (SHIINA *et al.*, 2009).

O sistema HLA está localizado no braço curto do cromossomo 6 (banda 6p21.3) e se estende por aproximadamente 3.600 kilobases do DNA. É a região com maior densidade gênica do genoma humano, sendo que, dos 224 genes identificados, 128 são genes funcionais e 96 são pseudogenes. Além disso, ele pode ser subdividido em três regiões de acordo com a estrutura e função dos seus genes: HLA de classe I, de classe II e de classe III, como pode ser visto na Figura. 1.1 (APOSTOLOPOULOS *et al.*, 2008; BECK e TROWSDALE, 2000; CONSORTIUM, 1999).

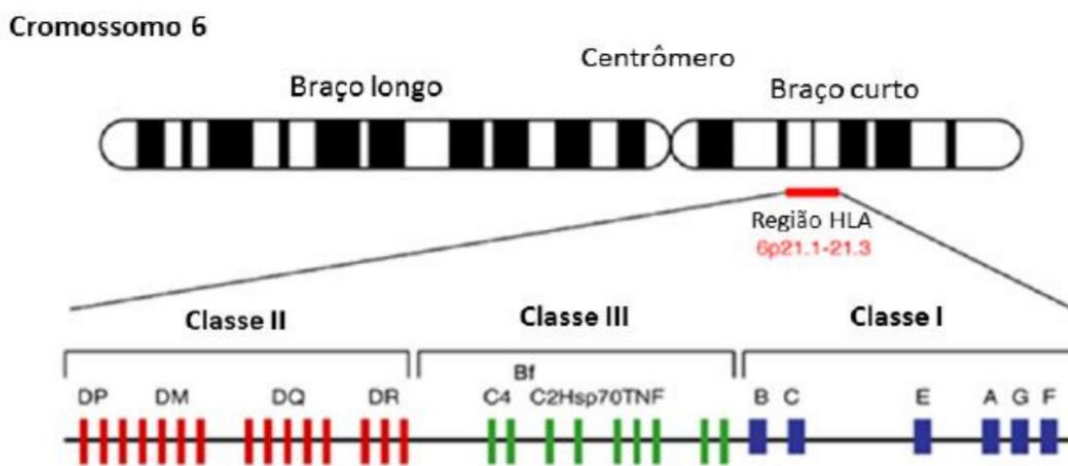


Figura 1.1 Mapa da região HLA, banda 6p21.3 do cromossomo 6 (retirada de CEREDA, 2012).

1.2. HLA DE CLASSE I

Na região mais telomérica estão localizados os genes clássicos (HLA-A, HLA-B e HLA-C) e não-clássicos (HLA-E, HLA-F e HLA-G) que correspondem à região de HLA de classe I. Esses genes expressam moléculas presentes na superfície celular da maioria das células nucleadas, embora o nível de expressão varie entre as células (CHOO, 2007). A principal função das moléculas de HLA de classe I clássicas é apresentar peptídeos derivados de processamento intracelular às células T citotóxicas CD8+. A apresentação de peptídeos de patógenos pode levar a uma resposta imune que resulta na morte

da célula infectada. Já a função das moléculas de HLA de classe I não-clássicos se baseia não na apresentação de epítomos mas sim na imunorregulação, como a inibição de células NK pelo HLA-G, por exemplo (CURIGLIANO *et al.*, 2013).

As moléculas de HLA de classe I são glicoproteínas heterodiméricas ligadas à membrana celular e constituídas por uma cadeia polimórfica alfa ou pesada (44 a 47 kDa) em associação não-covalente na porção extracelular com uma proteína não-polimórfica β 2-microglobulina (12 kDa). Em humanos, a β 2-microglobulina é invariável e seu gene está localizado no cromossomo 15 (BJORKMAN e PARHAM, 1990; CONSORTIUM, 1999).

A cadeia alfa ancora a molécula de HLA de classe I na membrana celular. Ela possui três domínios extracelulares (α 1, α 2 e α 3), uma região transmembranar e um domínio citoplasmático, que podem ser exemplificados na Figura. 1.2. Os domínios α 1 e α 2 possuem como característica serem altamente polimórficos, contendo variações na sequência de aminoácidos. Dessa forma, esses domínios determinam a especificidade antigênica das moléculas de HLA classe I. Já o domínio α 3 é altamente conservado e sítio de ligação para o receptor CD8, formando junto com a β 2-microglobulina um domínio constante (BJORKMAN *et al.*, 1987b). A região transmembranar da cadeia alfa é formada por aminoácidos hidrofóbicos e o domínio citoplasmático contém sítios para fosforilação de proteínas e ligação com o citoesqueleto celular (PARHAM, 1990).

Os domínios α 1 e α 2 da cadeia alfa formam uma estrutura única chamada fenda de ligação de peptídeo, que é o sítio de ligação para antígenos peptídicos. Essa estrutura, que pode ser observada na Figura 1.2, consiste em uma plataforma de oito folhas β -pregueadas antiparalelas, formando o assoalho da fenda, e duas α -hélices paralelas opostas no topo da plataforma constituindo um par de paredes. Na fenda se ligam peptídeos processados de 8 a 10 resíduos de aminoácidos, que são ligados através de uma série de ligações de hidrogênio e interações iônicas em cada extremidade do peptídeo (BJORKMAN *et al.*, 1987a; KLEIN e SATO, 2000a).

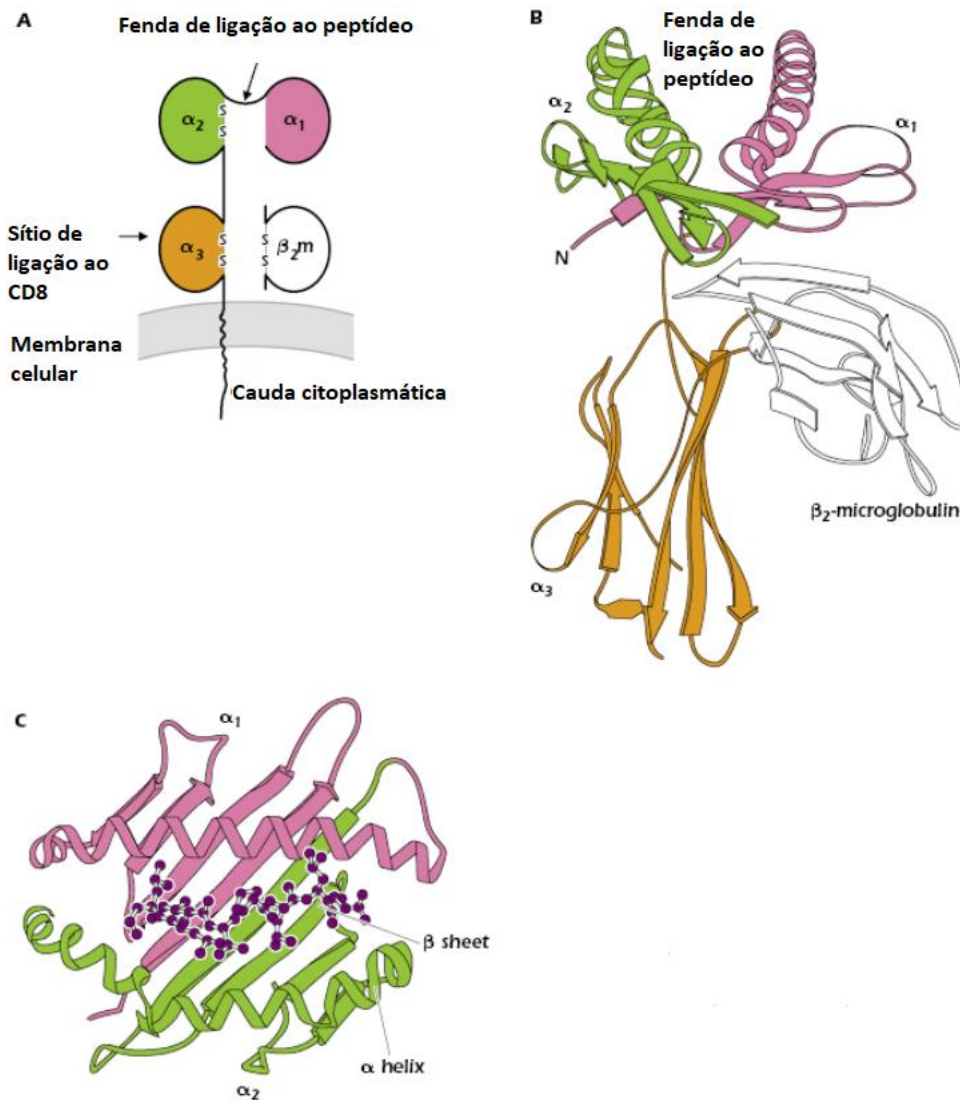


Figura 1.2 Estrutura da molécula de HLA classe I. A: representação esquemática da molécula de HLA classe I e seus domínios. B: representação da estrutura proteica da molécula de HLA classe I. C: representação da fenda de ligação da molécula de HLA classe I com um peptídeo ancorado a ela (retirada e modificada de <https://oncohemakey.com/how-t-cells-recognize-antigen-the-role-of-the-major-histocompatibility-complex/>).

A fenda de ligação está estruturada em várias subcavidades denominadas pockets (bolsos), que variam na sua constituição de aminoácidos, conferindo especificidade na ligação a diferentes peptídeos, como pode ser observado na Figura 1.3. Os aminoácidos que constituem esse sulco de ligação ao peptídeo representam a maior parte do polimorfismo do gene HLA de classe I e têm um impacto direto na natureza do peptídeo que pode se ligar a essa molécula. Dessa forma, diferentes peptídeos se ligam a diferentes moléculas de HLA, desde que certos aminoácidos específicos

estejam presentes por serem necessários para a ancoragem na fenda. Os pockets B e F são os mais importantes, pois interagem com as cadeias laterais dos resíduos P2 e P9 do peptídeo, funcionando como âncoras e determinando o tipo de peptídeo que se ligará à molécula de HLA (KLEIN e SATO, 2000a).

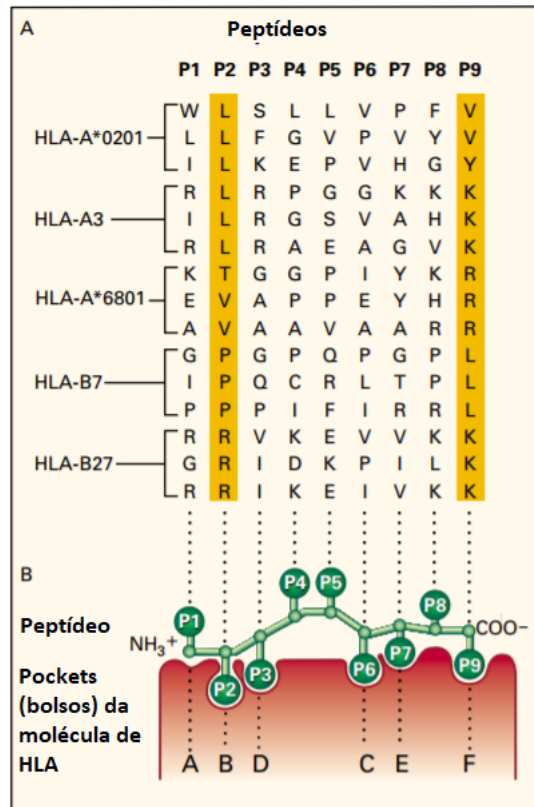


Figura 1.3 Fenda de ligação a peptídeo das moléculas de HLA classe I (Modificada de KLEIN and SATO, 2000a).

A sequência total dos genes que codificam a cadeia alfa das moléculas HLA de classe I (éxons 1 a 8) consiste em 1089-1101 nucleotídeos, e codificam para um polipeptídeo de 362-366 aminoácidos. Ela está estruturada de forma que cada éxon codifica cada um dos domínios do polipeptídeo. O éxon1 codifica o peptídeo líder; os domínios extracelulares $\alpha 1$, $\alpha 2$ e $\alpha 3$ são codificados pelos éxons 2 (343 pb), 3 (274 pb) e 4 (276 pb), respectivamente; o éxon 5 codifica o segmento transmembranar e a cauda citoplasmática pelos éxons 6 e 7. A região 3' não traduzida faz parte do éxon 8. (MARSH, 2000).

1.3. HLA DE CLASSE II E III

Os genes de HLA de classe II localizam-se na região mais centromérica do sistema HLA e são expressos nas células apresentadoras de antígenos (monócitos, macrófagos e células dendríticas), linfócitos B e linfócitos T ativados (CHOO, 2007). A principal função das moléculas HLA de classe II é apresentar peptídeo extracelular para células T CD4+. Isso leva a uma resposta imune, à ativação de células B para a produção de anticorpos e à ativação de células T citotóxicas. Dentre os genes, estão os clássicos (HLA-DR, HLA-DP e HLA-DQ) e os não-clássicos (HLA-DM e HLA-DO) (SANCHEZ-MAZAS *et al.*, 2011).

A molécula HLA de classe II é uma glicoproteína heterodimérica ligada à membrana, composta por duas cadeias polimórficas não-associadas covalentemente, α (34 kDa) e β (29 kDa). Estes receptores membranares são constituídos por uma porção extracelular, uma região transmembranar e um segmento citoplasmático como pode se observada na Figura 1.4 (BJORKMAN e PARHAM, 1990; DOHERTY e ZINKERNAGEL, 1975).

A região extracelular é formada pelos domínios $\alpha 2$ e $\beta 2$, região altamente conservada de ligação ao CD4 e pelos domínios $\alpha 1$ e $\beta 1$, região altamente polimórfica de ligação de peptídeos. Os domínios $\alpha 1$ e $\beta 1$ se dobram formando uma estrutura que consiste em um par de paredes laterais anti-paralelas que flanqueiam um piso de folhas β -pregueadas anti-paralelas. Os aminoácidos que constituem esse sulco de ligação peptídica, principalmente na cadeia β , mas também na cadeia α , representam a maioria do polimorfismo do gene HLA de classe II. A fenda de ligação de peptídeos se associa a peptídeos de 13-25 aminoácidos, sendo que alguns aminoácidos do peptídeo podem se ligar a partes externas da fenda (GATTI e PIERRE, 2003; KLEIN e SATO, 2000a).

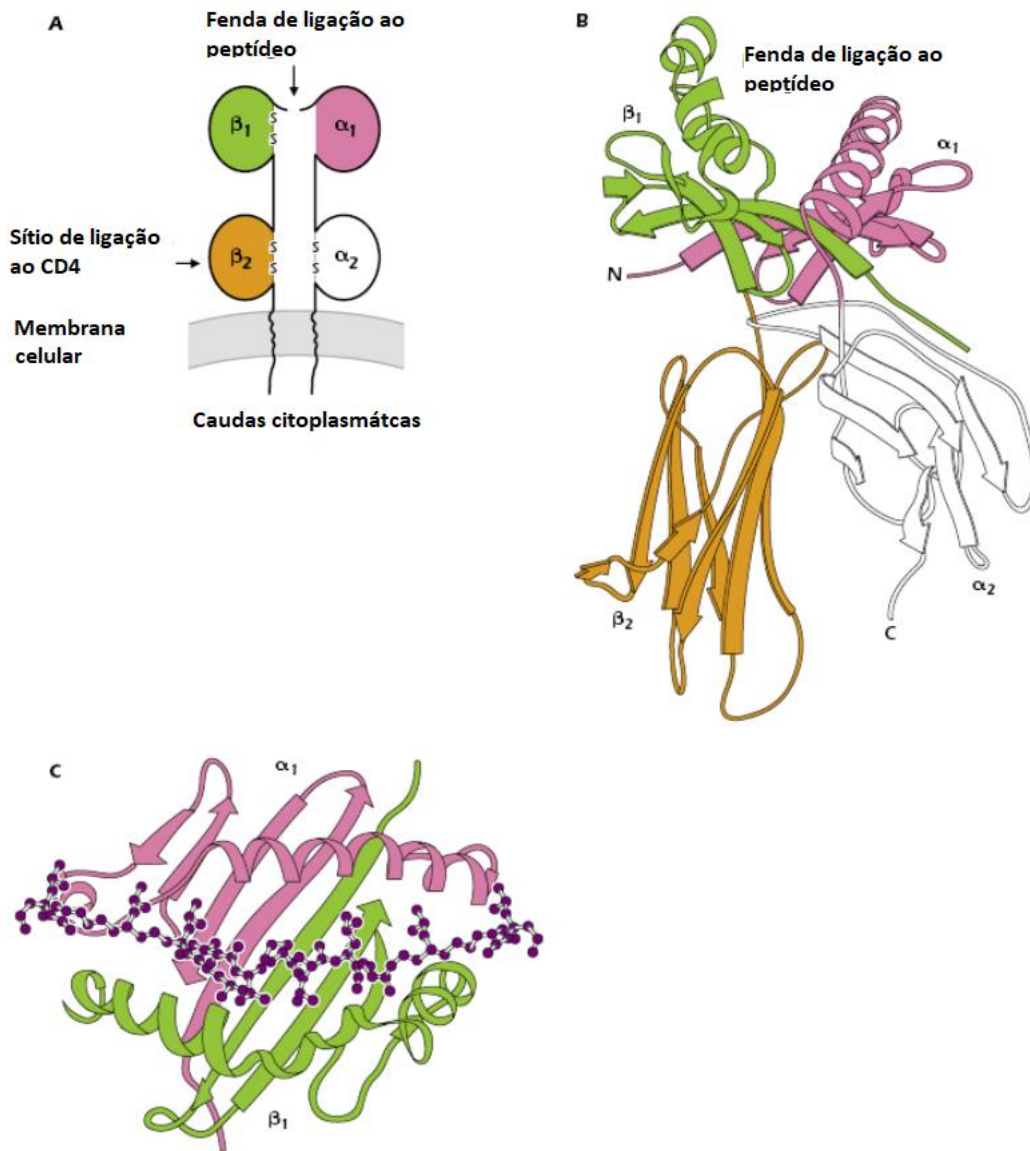


Figura 1.4 Estrutura da molécula de HLA de classe II A: representação esquemática da molécula de HLA classe II e seus domínios. B: representação da estrutura proteica da molécula de HLA classe II. C: representação da fenda de ligação da molécula de HLA classe II com um peptídeo ancorado a ela (retirada e modificada de <https://oncohemakey.com/how-t-cells-recognize-antigen-the-role-of-the-major-histocompatibility-complex/>).

Os genes de HLA de classe III constituem o segmento do genoma humano com maior densidade de genes, localizando-se entre os genes de classe I e classe II (XIE *et al.*, 2003). Nenhuma molécula de HLA é codificada por esses genes, que são genes para componentes do complemento (C2, C4, fator B), 21-hidroxilase e fator de necrose tumoral (TNF), entre outros (BECK e TROWSDALE, 2000). Os genes de classe III do sistema HLA possuem várias funções, destacando-se o seu papel na codificação de proteínas solúveis

importantes na modelação e regulação da resposta imune (XIE *et al.*, 2003).

1.4. PROCESSAMENTO E APRESENTAÇÃO DE ANTÍGENO VIA HLA DE CLASSE I

Os HLAs de classe I e II possuem funções semelhantes de apresentação de antígenos na superfície celular. Porém, o HLA de classe I é reconhecido por linfócitos T CD8+ enquanto que o HLA de classe II por linfócitos T CD4+ (PAMER e CRESSWELL, 1998). A natureza e origem dos peptídeos que vão se ligar a moléculas de classe I ou II também são diferentes (ENGELHARD, 1994). Moléculas de HLA de classe I se associam a antígenos endógenos próprios ou não-próprios, enquanto que HLAs de classe II se associam a antígenos exógenos de patógenos extracelulares (dentro de vesículas no citoplasma) (VYAS *et al.*, 2008).

A apresentação de antígenos via HLA de classe I se inicia com a clivagem de proteínas pelo proteassoma 26S em fragmentos menores (peptídeos). O proteassoma 26S é um complexo enzimático proteolítico central do sistema de ubiquitina-proteassoma (UPS) responsável pela degradação de proteínas poli-ubiquitiniladas. Ele é composto por uma subunidade 20S, que é o núcleo catalítico com atividade protease, formando uma estrutura em forma de barril, e duas subunidades reguladoras 19S, constituindo uma estrutura de tampa (SIJTS e KLOETZEL, 2011).

Uma vez gerados os peptídeos, eles deverão ser transportados do citoplasma da célula para o retículo endoplasmático (RE, onde se encontram as moléculas de HLA de classe I. Esse transporte é mediado pela proteína transportadora TAP, localizada na membrana do RE, que bombeia os peptídeos formados no citosol para o lúmen do RE. Além disso, ela também atua na estabilização da molécula de HLA, fazendo parte do complexo de carregamento de peptídeo (PLC) junto com outras moléculas como a tapasina e chaperonas (NEEFJES *et al.*, 2011).

A TAP transloca peptídeos entre 8 e 16 aminoácidos para o RE. Estes peptídeos podem requerer um corte adicional antes de se ligarem às moléculas de HLA de classe I. Uma aminopeptidase do RE, denominada ERAP1, corta os

peptídeos até um comprimento mínimo de 8 aminoácidos. Como a ERAP1 não faz parte do PLC, muitos peptídeos devem ser aparados fora do PLC, no lúmen do RE e, em seguida, se ligar a moléculas MHC de classe I. Quando esses peptídeos não conseguem se ligar a moléculas de HLA de classe I, eles são removidos pela ERAD e retornam ao citosol (NEEFJES *et al.*, 2011).

No RE, as moléculas HLA de classe I são parcialmente dobradas e estabilizadas pelas chaperonas (calreticulina e ERp57) e tapasina, que acompanham toda a síntese da molécula e seleção de peptídeo, impedindo que ela prossiga caso ocorra algum erro. As chaperonas se associam à cadeia alfa da molécula de HLA, estabilizando-a até que a cadeia β -microglobulina seja sintetizada e se ligue à molécula, e até que seja feita a correta ligação com o peptídeo. A tapasina se liga à molécula de HLA após a sua associação à cadeia β -microglobulina e medeia a interação da molécula de HLA com a TAP, otimizando a ligação do peptídeo (NEEFJES *et al.*, 2011; VYAS *et al.*, 2008).

Dessa forma, o PLC garante um carregamento eficiente de peptídeos nas moléculas de HLA de classe I. Após a ligação a um peptídeo de afinidade, o complexo HLA de classe I é liberado das chaperonas, passando pelo sistema de controle de qualidade do RE para expressão na membrana plasmática. O complexo HLA associado ao peptídeo é encaminhado para o complexo de Golgi e transportado por vesículas até a membrana celular, para ser expresso na superfície celular e reconhecido por receptores de células T CD8+ (NEEFJES *et al.*, 2011; VYAS *et al.*, 2008).

1.5. POLIMORFISMOS DA MOLÉCULA DE HLA DE CLASSE I

O sistema HLA é conhecido por ser o mais polimórfico em humanos. Os polimorfismos do HLA de classe I localizam-se essencialmente nos éxons 2 e 3, os quais codificam para as estruturas moleculares do receptor onde se localiza o sítio de ligação a peptídeos (BJORKMAN e PARHAM, 1990; BJORKMAN *et al.*, 1987a; KLEIN e SATO, 2000a; MALISSEN *et al.*, 1982). Esse número limitado de resíduos de aminoácidos é o responsável pela especificidade de ligação das moléculas de HLA. Como diferentes moléculas de HLA apresentam diferentes aminoácidos em sua sequência de ligação a

peptídeos, cada molécula de HLA pode se ligar a diferentes peptídeos através do conceito de ligação de peptídeos “degenerada” (FALK *et al.*, 1991; GARRETT *et al.*, 1989).

Os polimorfismos geram mudanças estruturais na fenda de ligação no domínio extracelular da proteína de HLA classe I. Isso afeta a função biológica das moléculas de HLA classe I de se ligar e apresentar peptídeos para o receptor antígeno-específico do linfócito T CD8+ (CTLs) (STEPHENS, 2005). Além disso, existem grupos de alelos chamados de supertipos, que compartilham preferências de ligação específica a pequenos peptídeos antigênicos com tamanho, carga e composição de aminoácidos similares (SETTE e SIDNEY, 1999; SIDNEY *et al.*, 2008).

Vários mecanismos diferentes podem contribuir para a geração e manutenção do polimorfismo. Alguns novos alelos podem se originar por mutações pontuais, porém muitos são originados por recombinação ou conversão gênica, na qual uma sequência é substituída em parte por outra de um gene homólogo. A recombinação entre variantes alélicas de um locus parece ter sido mais importante do que a conversão gênica na geração de polimorfismo (PRUGNOLLE *et al.*, 2005).

Os polimorfismos no sistema HLA parecem ter sido selecionados por pressões evolutivas e muito dessa diversidade surgiu devido a pressões seletivas de microrganismos. Diversos estudos mostraram indícios de um papel de seleção orientada por patógenos na modelagem da variação do HLA, uma vez que a pressão do patógeno influencia a variabilidade do HLA (PENN *et al.*, 2002) e o polimorfismo HLA está correlacionado com a diversidade de patógenos (PRUGNOLLE *et al.*, 2005). Além disso, os genes HLA estão associados com susceptibilidade e resistência a doenças infecciosas (CAGLIANI e SIRONI, 2013). Assim, os polimorfismos extensos no sistema HLA podem ser considerados como uma marca histórica da evolução no genoma humano, que reflete o processo de seleção natural pelos microrganismos agindo no sistema imune adaptativo, levando à sobrevivência, diversificação e expansão da espécie (STEPHENS, 2012).

Variações extensivas de alelos de HLA classe I e classe II ocorrem em diferentes grupos étnicos, fazendo com que certos haplótipos sejam encontrados mais frequentemente em uma população. Este fenômeno se dá

devido ao desequilíbrio de ligação entre os loci do HLA, possibilitando combinações estáveis de alelos ou haplótipos, que também varia em composição e frequência entre diferentes grupos étnicos (GONZALEZ-GALARZA *et al.*, 2011; STEPHENS, 2005). O desequilíbrio de ligação ocorre quando a frequência de dois alelos juntos em um mesmo haplótipo excede o previsto. Por exemplo, o haplótipo HLA-A1, B8, DR17 é o mais comum entre caucasianos, com uma frequência de 5% (CHOO, 2007).

1.6. NOMECLATURA DO SISTEMA HLA

A nomenclatura do HLA foi estabelecida e padronizada nos Workshops Internacionais de HLA e Imunogenética (*International HLA and Immunogenetics Workshops - IHIW*) realizados desde 1964, para que os pesquisadores pudessem comparar suas técnicas e resultados de tipagem de HLA e comunicar novas descobertas.

A nomenclatura para descrever os alelos de HLA utiliza um sistema de numeração hierárquico. A Figura 1.5 abaixo exemplifica a nomenclatura do HLA. Os dois primeiros dígitos são para diferenciar grupos alélicos (acurácia de dois dígitos) de alelos específicos (acurácia de quatro dígitos), o que os difere quanto ao nível proteico. Polimorfismos encontrados em regiões exônicas, mas que não estão associadas a mudanças no grupo alélico, pois são alterações nucleotídicas sinônimas, são indicados nos dígitos seguintes (seis dígitos), aumentando a resolução. Os dois dígitos finais (oito dígitos) informam as variantes com alterações em outras regiões não-codificantes como regiões intrônicas ou promotoras. A nomenclatura também permite adicionar informações como alelos de baixa expressão (L), expressão aberrante (A), expressão questionável (Q), proteína presente no citoplasma (C), alelo nulo (N) ou secretado (S) (BAUER, 2018; MARSH *et al.*, 2010).

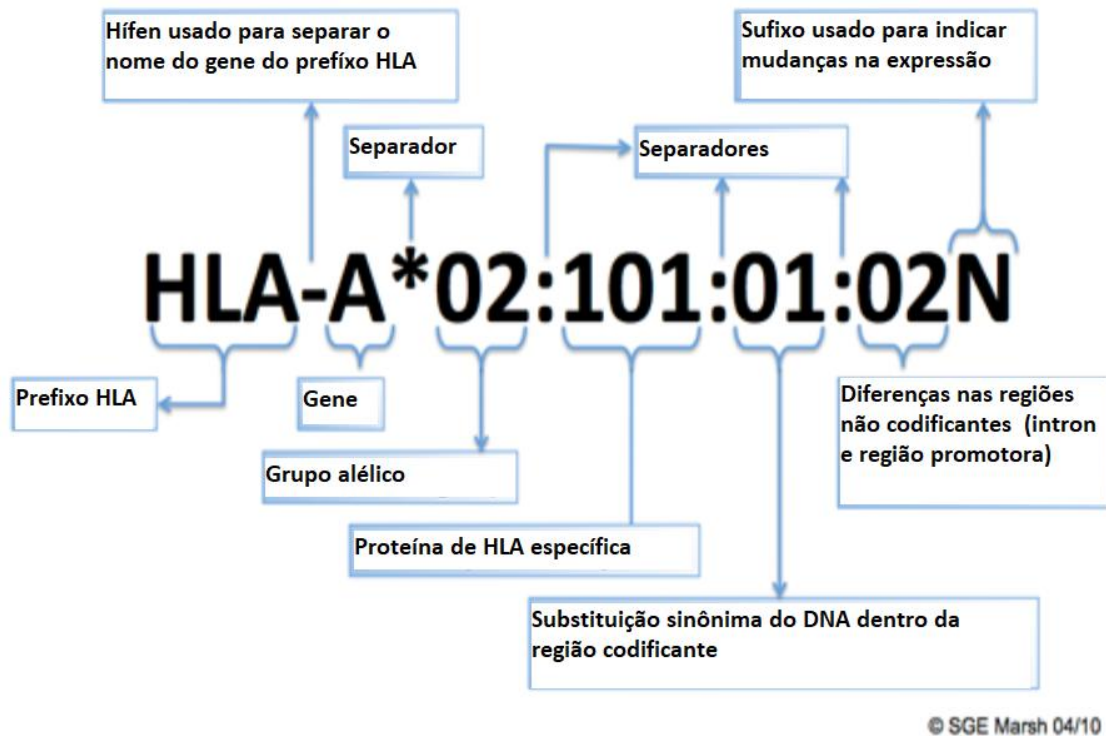


Figura 1.5 Nomeclatura do Sistema HLA (modificada de hla.alleles.org).

Atualmente já foram descritos e depositados no banco de dados IMGT/HLA sequencias referências de 16.200 alelos de HLA de classe I e 6.162 de classe II. O HLA-B é o alelo mais polimórfico entre os HLA classe I, com 6.096 alelos descritos, seguido do HLA-A com 5.018 alelos e do HLA-C com 4.852 alelos. Já entre os de classe II, o HLA-DRB é o lócus com o maior número de alelos descrito, 2.923 alelos, seguido do DQB1 com 1.560 alelos e do DPB1 com 1.360 alelos (IMGT/HLA, <http://www.ebi.ac.uk/ipd/imgt/hla/stats.html>, acessado em 24/05/2019 às 09:27).

1.7. TIPAGEM DO HLA POR SEQUENCIAMENTO DE NOVA GERAÇÃO

A tipagem do HLA tem sido amplamente usada na procura de doadores e receptores compatíveis para transplantes para redução do risco de rejeição de órgãos e doença enxerto-versus-hospedeiro. Variantes específicas de HLA estão associadas a doenças autoimunes, doenças infecciosas e reações adversas a drogas. A tipagem do HLA é muitas vezes utilizada não só complementando o diagnóstico de doenças como também na escolha de

prescrição de medicamentos para essas condições, no desenho de peptídeos imunogênicos para vacinas antivirais e tratamento de câncer, nos estudos de genética populacional, entre outros (BAUER, 2018).

A tipagem do HLA teve início com técnicas sorológica baseadas na ligação de antígenos do HLA desconhecido a anticorpos previamente conhecidos. Essa abordagem gera uma tipagem de baixa resolução com apenas dois dígitos (SCHMITZ, 2006). Com o desenvolvimento da técnica de reação em cadeia da polimerase (PCR), os genes de HLA puderam ser amplificados e identificados por sondas (PCR-SSO - *sequence-specific oligonucleotide probes*) ou sequências específicas (PCR-SSP - *sequence-specific primers*). Essas técnicas analisam apenas as regiões polimórficas dos éxons 2 e 3 dos genes de HLA de classe I, e para os HLA de classe II, o éxon 2 (DE SANTIS *et al.*, 2013). Outra metodologia de tipagem, considerada padrão-ouro, também utiliza a amplificação por PCR do HLA, o sequenciamento de Sanger e em seguida a sequência resultante é comparada com sequências-referências de HLA (SBT, do inglês *sequence-based typing*). Embora essas técnicas de genotipagem molecular permitam uma resolução maior do que a tipagem sorológica, elas são metodologias caras, e de trabalho intensivo e demorado (BAUER, 2018; ERLICH, 2012; GRUMBT *et al.*, 2013; WITTIG *et al.*, 2015).

Uma enorme desvantagem das metodologias de tipagem que se limitam ao sequenciamento parcial dos genes (apenas os éxons codificadores da fenda de ligação ao peptídeo) é a geração de tipagem ambígua. Essas ambiguidades podem ocorrer devido a padrões de polimorfismos em cis/trans de amostras heterozigotas ou devido a alguns alelos serem idênticos em regiões comumente analisadas, levando a duas ou mais combinações de alelos (BENTLEY *et al.*, 2009; HOLCOMB *et al.*, 2011; HOSOMICHI *et al.*, 2013; LIND *et al.*, 2010). A análise de éxons adicionais ou até mesmo de todo o gene e a determinação da fase dos polimorfismos seriam soluções para a tipagem. Técnicas para separação dos haplótipos são abordagens de difícil execução e caras (GRUMBT *et al.*, 2013; HOLCOMB *et al.*, 2011).

Segundo o banco de dados do IMGT/HLA, que contém as sequências dos alelos de HLA, poucos desses alelos descritos possuem a sequência completa; a grande maioria se limita apenas aos éxons mais polimórficos que

codificam a fenda de ligação a peptídeos, pois possuem grande relevância clínica em transplantes de órgãos. Porém, aproximadamente 30% dos receptores de transplante entre indivíduos com alelos “idênticos” apresentam eventos adversos de rejeição dentro de cinco anos (OTTINGER *et al.*, 2003). Não se sabe se isso se dá devido a fatores genéticos, como diferenças nucleotídicas em regiões não analisadas do HLA, ou se outros genes estariam envolvidos. Portanto, a relevância clínica dos outros éxons não habitualmente tipados permanece desconhecida.

Com o surgimento de novas metodologias de sequenciamento, conhecidas como *Next Generation Sequencing* (NGS), tornou-se possível a tipagem do HLA em alta resolução. Existem diferentes plataformas de NGS, como as da plataforma Illumina (HiSeq e MiSeq), 454/Roche e Ion torrent/Life Technologies. Apesar de diferirem em diversas etapas, as plataformas compartilham de duas etapas na preparação da biblioteca do DNA seguida do sequenciamento clonal onde é produzindo uma população de fragmentos de DNA idênticos ao DNA a ser sequenciado. Dessa forma, é gerado um grande número de dados, permitindo, assim, o sequenciamento de diversos polimorfismos, subclones, variantes e mutações (GRUMBT *et al.*, 2013; LIND *et al.*, 2010). As vantagens da metodologia de sequenciamento por NGS são: sequenciamento de grande número de bases por corrida, sequenciamento de diversos alvos (genes ou genomas) em paralelo, sequenciamento de diversas amostras ao mesmo tempo e grande volume de sequências geradas por corrida.

Essas novas tecnologias de NGS permitem a tipagem de diversas amostras e diversos genes do sistema HLA de uma única vez e também a tipagem em alta resolução. Conforme descrito por Nunes e colaboradores, a tipagem de baixa resolução se equivale à tipagem sorológica, definindo o alelo apenas a nível de família ou antígeno. A tipagem intermediária envolve a análise de polimorfismos nos éxons 2 e 3 para HLA classe I e éxons 2 para classe II, enquanto que a tipagem de alta resolução analisa polimorfismos em éxons adicionais (NUNES *et al.*, 2011a; NUNES *et al.*, 2011b).

Frente a isso, a tipagem de toda a sequência do HLA sem ambiguidades através de técnicas de NGS permite a caracterização completa em alta resolução de alelos novos e já existentes. Essas novas metodologias de

tipagem de HLA podem ter implicações relevantes para o transplante de células hematopoiéticas, assim como para diversas outras doenças associadas ao sistema HLA, como no caso da infecção pelo HIV-1. Recentemente, muitos protocolos para tipagem de HLA foram descritos em diferentes contextos de abordagem metodológica e análise dos dados, mostrando um panorama da tipagem em alta resolução do HLA por NGS e suas implicações (CAO *et al.*, 2013; CASTELLI *et al.*, 2018; DANZER *et al.*, 2013; EHRENBERG *et al.*, 2018; ERLICH, 2012; LANGE *et al.*, 2014; LANK *et al.*, 2012; LIMA *et al.*, 2019; MAJOR *et al.*, 2013; MONOS e MAIERS, 2015; NORMAN *et al.*, 2016; WANG *et al.*, 2012; WITTIG *et al.*, 2018).

1.8. FUNÇÃO DO HLA DE CLASSE I E SUA ASSOCIAÇÃO COM O SISTEMA IMUNE

A principal função do HLA de classe I é a de apresentar antígenos próprios e não-próprios na superfície celular como os produzidos no interior de células infectadas por vírus ou oriundos de desenvolvimento tumoral. Esses antígenos são reconhecidos pelos receptores TCR das células T CD8+, específicos para o reconhecimento do antígeno apresentado, desenvolvendo, então, uma resposta imune adaptativa mediada por linfócitos T citotóxicos, conhecida por ser a resposta primária contra células infectadas por vírus (CARRINGTON e O'BRIEN, 2003; ZINKERNAGEL, 1997).

Com o reconhecimento e interação do complexo HLA classe I-antígeno com o receptor TCR da célula T CD8+, são induzidos dois caminhos principais de morte da célula infectada. Os linfócitos T CD8+ induzem a via de morte celular programada (apoptose) na célula-alvo pela ação de perforinas e granzimas. A perforina é liberada formando canais na membrana da célula-alvo. Em seguida, as granzimas, serino-proteases, penetram na célula-alvo através dos canais e ativam caspases e nucleases na célula-alvo que vão induzir a fragmentação do DNA, resultando em apoptose. Além disso, as células T citotóxicas também podem levar à apoptose pela expressão do receptor Fas L (CD95), que interage com a molécula Fas nas células-alvo, também induzindo a apoptose (MIGUELES *et al.*, 2008; PARKIN e COHEN,

2001; POONIA *et al.*, 2009).

Outra interação do HLA com o sistema imune seria com as células *natural killer* (NK), que possuem uma habilidade inata de reconhecer e destruir células infectadas ou tumorais (DOHERTY e ZINKERNAGEL, 1975; PARHAM, 2005; TRINCHIERI, 1989). As células NK possuem receptores em sua superfície para o HLA de classe I e reconhecem a perda de expressão de moléculas de HLA classe I. Se na interação com uma célula, esse receptor não é reconhecido, a célula NK está programada para lisar o alvo. Esse mecanismo também se baseia na secreção de perforinas na superfície da célula que fazem furos na membrana celular possibilitando a passagem de granzimas que, por sua vez, causam indução da apoptose. Células tumorais e alguns vírus muitas vezes diminuem a expressão de HLA de classe I. Embora isso possa oferecer alguma vantagem ao patógeno, pois prejudica o reconhecimento por células T citotóxicas, eles estão sujeitos a serem alvos do mecanismo imune pelas células NK. Outras células com expressão normal de HLA de classe I ainda podem ser alvos de NK se elas proverem sinais apropriados para ativação de seus receptores. Muitos receptores de células NK foram identificados e a maioria dos seus ligantes são moléculas de HLA de classe I (CHOO, 2007).

Essa resposta imune gerada pela apresentação de antígenos pelo sistema HLA é de extrema importância para o transplante de órgãos e de células-tronco hematopoiéticas. A combinação de alelos de HLA entre doadores e receptores se faz necessária a fim de evitar a rejeição e a doença enxerto-versus-hospedeiro. Além disso, a menor expressão de moléculas de HLA em células tumorais e conseqüente menor apresentação de antígenos tumorais é uma das principais hipóteses para se explicar a inabilidade do sistema imune de reconhecer as células malignas, mostrando mais uma vez o importante papel da resposta imune mediada por HLA de classe I (APTSIAURI *et al.*, 2007a; APTSIAURI *et al.*, 2007b; DEL CAMPO *et al.*, 2012).

Os alelos de HLA têm sido relacionados à predisposição a diversas doenças e desfechos diferentes. Vários estudos estabeleceram associações de diferentes alelos de HLA-DRB1 com maior predisposição a desenvolvimento de doenças reumáticas autoimunes como artrite reumatóide e do alelo HLA-B*27 com a espondilite anquilosante. Outras doenças como esclerose múltipla, psoríase, diabetes tipo 1, narcolepsia e alergias também são associadas a

diferentes alelos de HLA. Os genes do sistema HLA também apresentam uma importância particular na progressão ou resistência a doenças infecciosas como, por exemplo, a infecção pelo HIV-1 (KLEIN e SATO, 2000b).

1.9. O PAPEL DO HLA DE CLASSE I NA INFECÇÃO PELO HIV-1

Algumas pessoas possuem eventos que indicam uma resposta natural à infecção pelo HIV-1. Em alguns casos, pessoas infectadas pelo HIV-1 conseguem manter o controle espontâneo do vírus a níveis indetectáveis ou permanecer por décadas sem manifestar evidências clínicas de progressão da doença. Há ainda os que progridem muito lentamente para a aids, que é caracterizada pelo colapso do sistema imune, aumento da carga viral, níveis de linfócitos T CD4+ <200 células por milímetro cúbico de sangue e susceptibilidade a doenças oportunistas (CARRINGTON e WALKER, 2012; DEEKS e WALKER, 2007; SABIN e LUNDGREN, 2013). Estudos mostraram a presença de algumas mutações em comum em grupos de pessoas expostas e não-infectadas e de pacientes infectados que apresentavam uma progressão lenta da doença, sugerindo que características do hospedeiro estariam impedindo ou dificultando a entrada do vírus na célula e reduzindo a infecção, e, se a infecção ocorresse, essas mesmas características eliminariam ou retardariam a progressão da doença (MARMOR *et al.*, 2006).

As respostas celulares contra infecções são dependentes da constituição imunogenética do indivíduo, e dentro do genoma humano, os genes de HLA de classe I são os mais convincentes loci de traços quantitativos para o controle da infecção pelo HIV-1 (BRUMME *et al.*, 2012; CARLSON *et al.*, 2012; CARRINGTON e O'BRIEN, 2003; DALMASSO *et al.*, 2008; FELLAY, 2009; FELLAY *et al.*, 2009; KAWASHIMA *et al.*, 2009; MOORE *et al.*, 2002; PELAK *et al.*, 2010). Diferentes aspectos da infecção pelo HIV e da aids como a transmissão viral, controle da dinâmica e progressão viral, desenvolvimento de doenças oportunistas e até a resposta à terapia, incluindo a hipersensibilidade observada a drogas antivirais, são influenciados pela diversidade genética do HLA (KAUR e MEHRA, 2009a, b; SINGH *et al.*, 2008; STEPHENS, 2005).

De fato, os peptídeos virais apresentados por algumas moléculas HLA de classe I induzem fortes respostas contra o HIV a partir das células T CD8+ específicas, responsáveis pelo controle da replicação viral, e, conseqüentemente, bloqueiam a depleção de linfócitos T CD4+ e a progressão da doença (LETVIN e WALKER, 2003; MCMICHAEL e JONES, 2010; PEREYRA et al., 2010). A resposta imune restrita por HLA de classe I possui, portanto, influência direta na diversidade, replicação e adaptabilidade evolutiva viral, através das mutações de escape imune que surgem no genoma do HIV (SHARMA et al., 2011).

Vários estudos demonstraram que alguns alelos de HLA estão associados ao controle imune da infecção e à progressão mais lenta da doença, que é o caso dos alelos HLA-B*27:05 e B*57:01, tidos como “protetores” do hospedeiro. Por outro lado, alelos como B*07, B*35 e B*53 levam a um rápido desenvolvimento para a aids (BANSAL et al., 2007; CARRINGTON e O'BRIEN, 2003; MCMICHAEL e JONES, 2010; MIGUELES et al., 2000; TANG e KASLOW, 2003). Os alelos B*57:01 e B*27:05 são os mais bem descritos na literatura quanto à associação com o controle da carga viral e progressão mais lenta para aids. Tem sido proposto que os peptídeos virais apresentados por estas moléculas induzem fortes respostas citotóxicas contra o HIV, que são responsáveis pelo controle viral observado (MCMICHAEL e JONES, 2010; PEREYRA et al., 2010). Além disso, estudos mostraram que células T CD8+ restritas pelos alelos HLA-B*27 e B*57 são muito mais resistentes à supressão pela regulação de Tregs (linfócitos T reguladores) do que células T CD8+ restritas por alelos não-protetores da infecção pelo HIV. A falta de supressão dessas células CTLs restritas por esses alelos permite que essas células continuem a proliferar e lisar os alvos infectados durante a infecção crônica, o que pode gerar uma progressão mais lenta para a doença em pessoas com esses alelos protetores (ELAHI et al., 2011).

Apesar da grande influência do HLA-B em haplótipos relacionados à proteção contra a infecção pelo HIV, alguns haplótipos protetores não podem ser explicados apenas pela ligação ao HLA-B, evidenciando um efeito interdependente entre os alelos. Apesar do HLA-B ter um grande impacto, o controle do HIV é provavelmente influenciado por efeitos aditivos de outros alelos de HLA presente (LESLIE et al., 2010; MCMICHAEL e JONES, 2010;

ZIPETO e BERETTA, 2012). A progressão para aids também foi associada com uma variante na região regulatória 5'UTR do HLA-C (KULKARNI *et al.*, 2011). Isso indica que a seleção dos genes de HLA não é apenas restrita aos domínios estruturais envolvidos na ligação do peptídeo, mas também envolve variantes em regiões regulatórias (MEYER *et al.*, 2018).

A heterozigosidade de alelos de HLA de classe I está associada a uma menor progressão para aids em pacientes infectados pelo HIV, enquanto a homozigosidade está associada à progressão mais rápida para a doença. Isso pode ser resultante da diversidade de peptídeos de HIV apresentados para células T devido à presença de uma maior diversidade alélica em heterozigotos (CARRINGTON, 1999; CARRINGTON e O'BRIEN, 2003).

Outro dado interessante são estudos que demonstraram a associação de alelos de HLA de classe I a subtipos específicos do HIV-1 (FARQUHAR *et al.*, 2004; FRAHM *et al.*, 2005; FRAHM *et al.*, 2006; KASLOW *et al.*, 1996; KIEPIELA *et al.*, 2004; LAZARYAN *et al.*, 2006; LESLIE *et al.*, 2010). Como exemplo, temos a associação do HLA-B*35:01 à progressão mais rápida no subtipo B, mas não ao subtipo C. Isso se dá, em parte, porque a sequência consenso subtipo B do epítipo NY10 (códon 253 a 262 de Gag) tem uma afinidade reduzida ao HLA-B*35:01. Dessa forma, o impacto dos alelos HLA, como o HLA-B*35:01, no desfecho da doença pelo HIV podem diferir de acordo com o subtipo infectante do vírus (MATTHEWS *et al.*, 2012).

1.10. EPIDEMIOLOGIA DA INFECÇÃO PELO HIV-1 E AIDS

Em 1983, Françoise Barré-Sinoussi, Luc Montagnier e colaboradores do Instituto Pasteur de Paris isolaram um retrovírus a partir de linfonodos de um paciente com síndrome da imunodeficiência adquirida (aids) e o denominaram LAV, do inglês *Lymphadenopathy-Associated Virus* (BARRE-SINOUSSE *et al.*, 1983). Nesse mesmo ano, nos Estados Unidos, Robert Gallo e colaboradores do *National Institute of Health* (NIH) isolaram o mesmo vírus e o denominaram HTLV-III (GALLO, 1984; GALLO *et al.*, 1983). A unificação da nomenclatura ocorreu em 1986, pelo Comitê Internacional para a Taxonomia dos Vírus (ICTV, do inglês *International Committee on Taxonomy of Viruses*),

denominando-o, então, de vírus da imunodeficiência humana (HIV) (COFFIN *et al.*, 1997).

No entanto, os primeiros relatos de manifestações da aids datam de 1981 em usuários de drogas via parenteral e homens que fazem sexo com homens (HSH), que apresentavam sintomatologia de imunodeficiência, grande perda de peso, febre prolongada e doenças infecciosas oportunistas (GOTTLIEB *et al.*, 1981). Ao longo dos anos, surgiram novos casos da doença que indicaram a via de transmissão do vírus por meio de fluidos corporais contaminados (via parenteral e via sexual) ao acometer pacientes hemofílicos, transfundidos e crianças (JETT *et al.*, 1983).

Atualmente, a aids é considerada uma das maiores pandemias da humanidade. Segundo dados recentes da UNAIDS, até o ano de 2018, 37,9 milhões de pessoas convivem com HIV no mundo (UNAIDS, 2019b). A prevalência mundial de adultos infectados é de 0,8%. Contudo, devido aos avanços científicos e esforços globais para fortalecer os programas de prevenção e tratamento do HIV, o número de novas infecções por HIV caiu 40% desde o pico em 1997 (1,7 milhão de novas infecções em 2018, em comparação com 2,9 milhões em 1997) (UNAIDS, 2019b).

Em 2018, 23,3 milhões de pessoas vivendo com HIV tinham acesso à terapia antirretroviral. Um aumento de 7,7 milhões em comparação com 2010 (UNAIDS, 2019a). Nesse mesmo ano, 62% de todas as pessoas vivendo com HIV tinham acesso ao tratamento, e dentre estas últimas, 86% estavam com a carga viral suprimida ou indetectável (UNAIDS, 2019b).

O aumento global da terapia antirretroviral (maior cobertura de tratamento e melhor adesão ao tratamento) tem sido o principal contribuidor para um declínio de mais de 55% das mortes relacionadas à aids desde o pico em 2004. Em 2017, 770 mil pessoas morreram por doenças relacionadas à aids em todo o mundo, em comparação com 1,7 milhão em 2004 e 1,2 em 2010 (UNAIDS, 2019b).

De acordo com o último Boletim Epidemiológico divulgado pelo Departamento de Vigilância, Prevenção e Controle das Infecções Sexualmente Transmissíveis, do HIV/Aids e das Hepatites, foram detectadas 982.129 mil casos de Aids no país em 2017. De 2007 até junho de 2018 foram notificados 247.795 casos de infecção pelo HIV no Brasil, a grande maioria no Sudeste,

com 117.415 (47,4%), 50.890 no Sul (20,5%), 42.215 no Nordeste (17%), 19.781 na região Norte (8,0%) e 17.494 no Centro-Oeste (7,1%) (MINISTÉRIO DA SAÚDE, 2018a). Mais da metade das pessoas vivendo com HIV estão em tratamento antirretroviral (64% [550 mil]) e com supressão viral (59% [510 mil]) (UNAIDS, 2017).

A taxa de detecção de aids vem caindo no Brasil nos últimos anos. Em um período de dez anos, a taxa de detecção apresentou queda de 9,4%: em 2007, foi de 20,2 casos por 100 mil habitantes e, em 2017, de 18,3 casos a cada 100 mil habitantes. As regiões Sudeste e Sul apresentaram tendência de queda nos últimos dez anos; em 2007, as taxas de detecção dessas regiões foram de 22,0 e 32,9, passando para 17,1 e 24,1 casos por 100 mil habitantes em 2017: quedas de 22,2% e 26,7%, respectivamente. Vale destacar que Porto Alegre apresentou taxa de 60,8 casos/100 mil habitantes, em 2017, valor superior ao dobro da taxa do Rio Grande do Sul e 3,3 vezes maior que a taxa do Brasil (MINISTÉRIO DA SAÚDE, 2018a).

No período de 2007 a 2017, verificou-se uma queda de 14,8% no coeficiente de mortalidade por aids padronizado para o Brasil, que passou de 5,6 para 4,8 óbitos por 100 mil habitantes (MINISTÉRIO DA SAÚDE, 2018a).

1.11. HISTÓRIA NATURAL DA INFECÇÃO PELO HIV-1

O vírus pode ser transmitido horizontalmente através da exposição da mucosa a fluidos contaminados (hemoderivados, sêmen ou secreção vaginal) ou por inoculação (drogas injetáveis) (SPIRA *et al.*, 1996). Já a transmissão vertical pode ocorrer durante a gestação, parto ou aleitamento materno (COWAN *et al.*, 1984). Diversos fatores como propriedades biológicas da cepa viral quanto do hospedeiro podem afetar o sucesso da transmissão e a sua evolução para o desenvolvimento da aids.

O HIV possuiu como células-alvo as que expressam receptores CD4, como linfócitos T CD4+ do sistema imune do hospedeiro, macrófagos e células dendríticas. Inicialmente, ao entrar no hospedeiro, o vírus infecta as células inflamatórias ou células mononucleares do sangue periférico, replicando-se. Em seguida, o vírus é carregado por células dendríticas até os órgãos linfóides

que abrigam grande quantidade de células-alvo para a infecção (SPIRA *et al.*, 1996).

Após o estabelecimento da infecção, a progressão é caracterizada por três fases: (i) aguda, (ii) assintomática e (iii) aids, exemplificadas na Figura 1.6. A fase aguda é caracterizada por uma intensa replicação viral, que pode ser observada por um pico de viremia, e depleção drástica dos níveis de linfócitos T CD4+, ocorrendo nas primeiras duas a quatro semanas após a infecção. Essa fase é geralmente sintomática, porém com sintomas inespecíficos, semelhantes a uma gripe, tais como febre, cefaleia, faringite, exantema, mialgia e mal estar. Em geral este período é curto, podendo durar cerca de oito semanas, até que o sistema imune consiga controlar a alta taxa de replicação viral pela ação de linfócitos T CD8+ (BANGHAM, 2009). Ocorre, também, a produção de anticorpos neutralizantes e a ação conjunta destas duas respostas é capaz de reduzir a viremia a níveis baixos, controlando a replicação viral, além de recuperar parcialmente os níveis de células T CD4+, controlando a infecção.

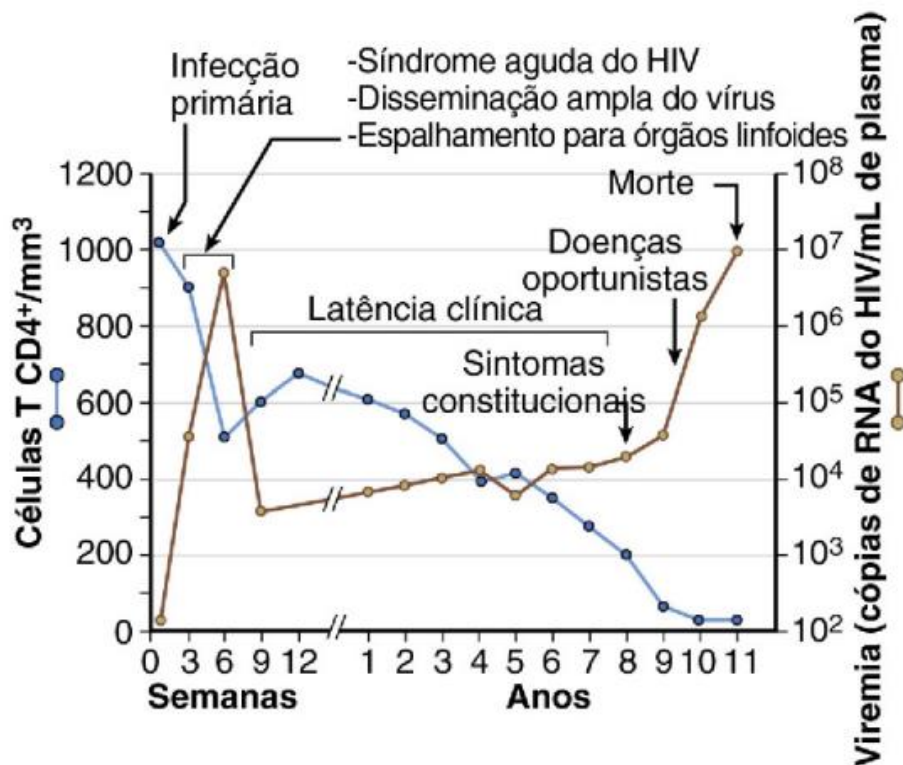


Figura 1.6 História natural da infecção pelo HIV-1 em adultos (extraído e traduzido de ABBAS, 2015).

Após esta fase inicia-se a fase crônica, na qual a resposta imune contra o vírus promove um equilíbrio dinâmico da carga viral plasmática do HIV, e, dessa forma, as taxas de produção e eliminação viral são similares (HO, 1995). Esse período pode durar até dez anos em média, sem que o hospedeiro apresente nenhum sintoma clínico, porém, a destruição progressiva dos tecidos linfóides leva a uma disfunção imune de linfócitos T CD4+, caracterizando o estágio avançado da doença.

A entrada do hospedeiro na fase aids é caracterizada pelo colapso do sistema imune por aumento da carga viral e diminuição da contagem de linfócitos T CD4+ para menos do que 200 células por milímetro cúbico de sangue. A queda acentuada dos níveis de células T CD4+ implica a susceptibilidade do indivíduo a infecções oportunistas por bactérias, fungos, vírus, infecções por protozoários e surgimento de neoplasias, sendo muitas delas mortais ao indivíduo (SABIN e LUNDGREN, 2013).

Devido à variabilidade da duração da fase assintomática, vários grupos têm sido descritos na literatura com características clínicas distintas, como os progressores rápidos, os progressores lentos ou não-progressores (LTNP, do inglês *long-term nonprogressors*) e os controladores de elite. Os indivíduos considerados progressores rápidos, na ausência de terapia antirretroviral, são caracterizados por um declínio drástico das células T CD4+ e uma evolução para a fase aids dentro de três anos em média após a infecção (LEITE *et al.*, 2017; PANTALEO e FAUCI, 1996). Os indivíduos caracterizados como LTNP desenvolvem uma lenta diminuição das células T CD4+ ao longo do tempo, podendo manter baixos níveis de viremia, por mais de 10 anos na ausência de terapia antirretroviral (CASADO *et al.*, 2010; PETRUCCI *et al.*, 1997). Os controladores de elite são caracterizados pela persistência de carga viral indetectável independente do tempo de infecção (SAKSENA *et al.*, 2007). Estes últimos são alvos de estudo para tentar esclarecer os mecanismos responsáveis pelo controle da infecção (CASADO *et al.*, 2010; KAMYA *et al.*, 2011).

1.12. PARTÍCULA VIRAL, ESTRUTURA GENÔMICA DO HIV-1 E SEUS PRODUTOS PROTEICOS

O HIV é um retrovírus da família *Retroviridae*, caracterizada por possuir a enzima transcriptase reversa, que tem por finalidade a retrotranscrição do RNA viral em uma dupla fita de cDNA para ser integrado ao genoma do hospedeiro durante a entrada do vírus na célula. Ele é membro do gênero *Lentivirus*, que apresenta progressão de sintomatologia lenta (COFFIN *et al.*, 1997).

A partícula viral do HIV-1, exemplificada na Figura 1.7, é uma estrutura esférica que mede de cerca de 100 nanômetros e é composta na sua camada mais externa por um envelope de natureza glicolipoproteica, originado da camada bilipídica da célula hospedeira infectada durante o brotamento do vírus (GOTO *et al.*, 1994). Inseridos nessa membrana externa, são encontradas glicoproteínas virais (gp120 e gp41) responsáveis pela ligação ao receptor CD4 e correceptores (CCR5 e CXCR4) presentes na superfície da célula a ser infectada e fusão entre o envelope viral e a célula hospedeira (CLAPHAM e WEISS, 1997).

Interno ao envelope encontra-se a matriz, formada pela p17, e abaixo da matriz temos o capsídeo viral de formato icosaédrico, formado pela proteína p24, que possui no seu interior o genoma do vírus (DARLIX *et al.*, 1990; FREED, 2001; MASSIAH *et al.*, 1996; WAIN-HOBSON *et al.*, 1985). Dentro do capsídeo viral se encontram proteínas virais essenciais como a protease (PR), transcriptase reversa (RT) e integrase (IN) e proteínas acessórias Vpr, Vpu e Nef (FRANKEL e YOUNG, 1998; SCARLATA e CARTER, 2003). A organização da partícula viral pode ser observada na Figura. 1.7.

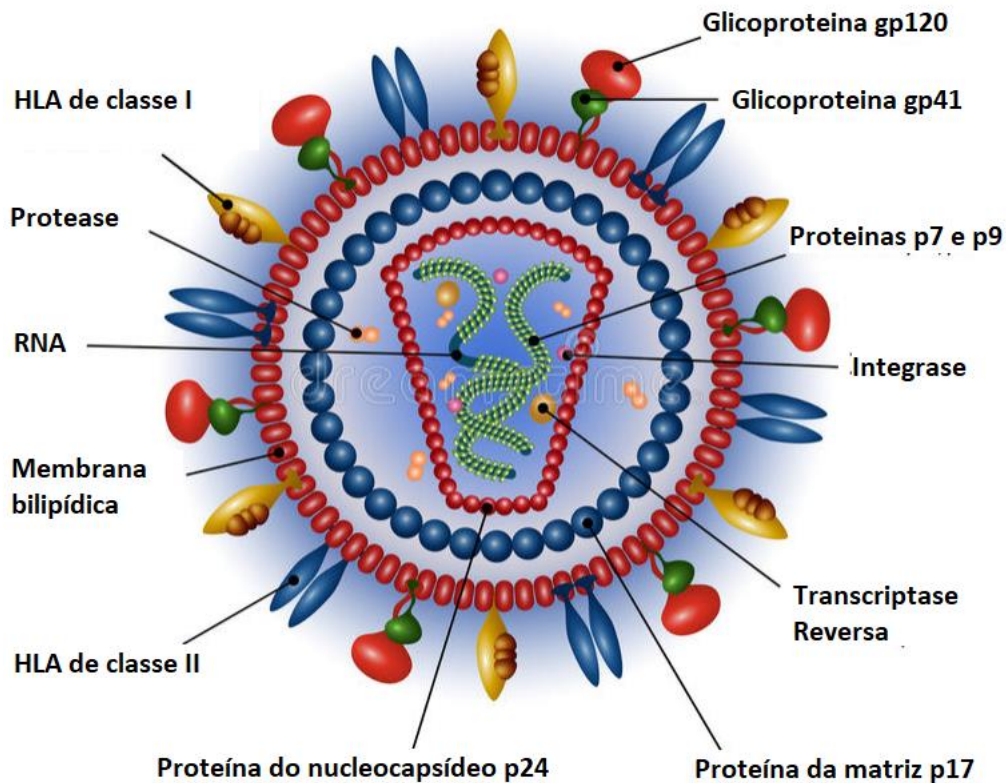


Figura 1.7 Partícula viral do HIV-1 (retirado e traduzido de <https://pt.dreamstime.com/foto-de-stock-estrutura-da-part%C3%ADcula-do-v%C3%ADrus-do-vih-image38001720>).

O genoma viral é composto por duas cópias similares de RNA de polaridade positiva de aproximadamente 9,5 kb de tamanho unidas pelas extremidades 5', formando um dímero (WAIN-HOBSON *et al.*, 1985). O genoma do HIV-1 possui três genes (gag, pol e env) que codificam proteínas essenciais, comuns a todos os retrovírus; dois genes (tat e rev) que codificam proteínas reguladoras; e quatro genes (vif, vpr, vpu e nef) que codificam proteínas acessórias (WAIN-HOBSON *et al.*, 1985). Além disso, como característica comum a todos os retrovírus, o HIV-1 também tem o genoma flanqueado por repetições terminais longas (LTRs), que constituem o local de iniciação transcricional, regulando a integração no genoma do hospedeiro, a expressão dos genes e a replicação viral (VARMUS, 1988). As particularidades do genoma do HIV podem ser observadas na Figura 1.8 abaixo.

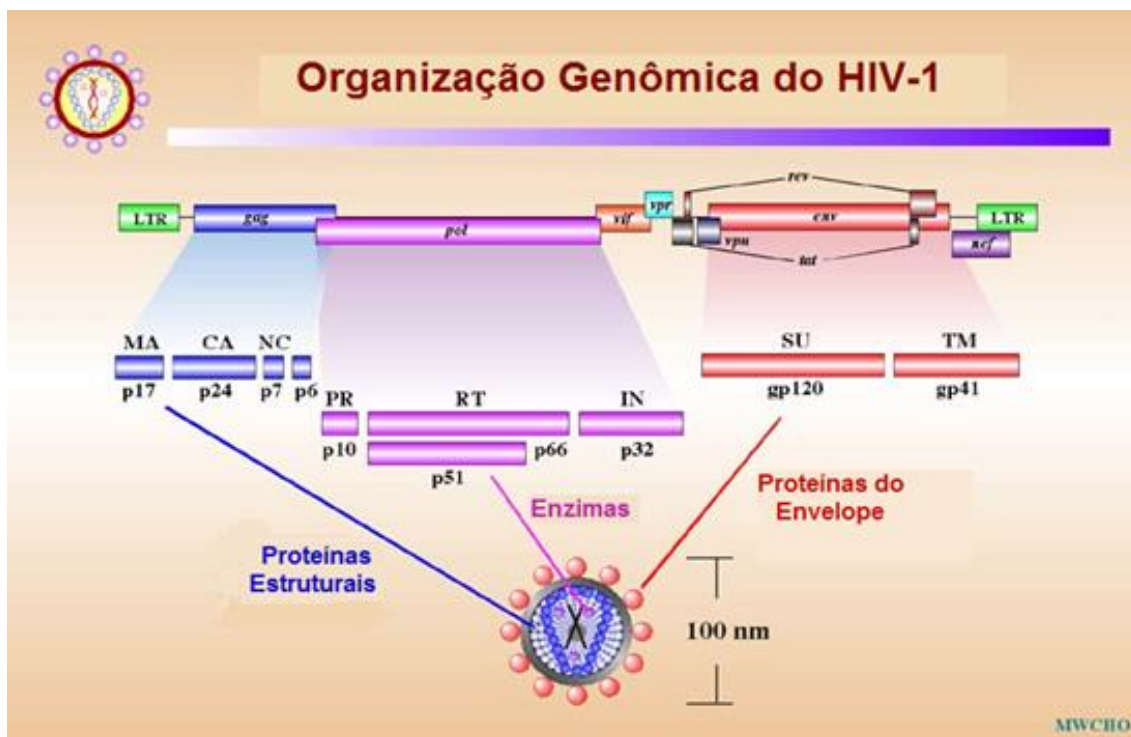


Figura 1.8 Organização genômica do HIV-1 (modificada de <http://www.liquidarea.com/2009/08/aids-decifrado-genoma-virus>).

O gene gag codifica as proteínas estruturais p17 (matriz), que auxiliam a migração do DNA viral até o núcleo; p24 (capsídeo), que conferem o formato cônico a partícula viral; p9, que é responsável por reconhecer o sinal de montagem da nova partícula viral e p6, que possibilita a incorporação do Vpr (uma das proteínas acessórias) na partícula que está sendo montada; todas traduzidas na forma de uma poliproteína precursora Gag (Pr55). Gag é clivada pela protease viral após o brotamento, dando origem às proteínas maduras (FREED, 2001; GALLAY *et al.*, 1995; GOTTLINGER *et al.*, 1989; HARRISON e LEVER, 1992; PAXTON *et al.*, 1993).

A proteína Pol, codificada pelo gene pol, é transcrita e traduzida junto com o precursor Gag, formando uma poliproteína Gag-Pol (Pr160). Após a clivagem pela protease viral, estas poliproteínas são separadas e o produto do gene pol originará proteínas fundamentais ao ciclo replicativo do HIV. São elas: a protease (PR), que é indispensável ao processo de maturação do vírus pela clivagem dos precursores virais Gag-Pol e Pol; a transcriptase reversa (RT), que realiza a transcrição do RNA viral; a RNase H (RH), que remove o molde de RNA original da primeira fita de DNA permitindo a síntese da cadeia

complementar de DNA; e a integrase (IN), que atua na inserção do DNA proviral do HIV no DNA genômico da célula infectada (BUSHMAN *et al.*, 1990). Cabe ressaltar aqui que a enzima RT não possui propriedade de correção durante o processo de replicação viral, o que resulta em uma alta taxa de erro na incorporação dos nucleotídeos (aproximadamente 10^{-4} por ciclo replicativo), e isso está diretamente relacionado à alta diversidade intra- e inter-hospedeiro do HIV (NOWAK, 1990).

O gene *env* codifica a proteína precursora gp160 (Env), uma poliproteína que, após clivada pela protease celular, originará as proteínas que compõem o envelope viral. São elas: a gp120, responsável por reconhecer a molécula de CD4 e seus correceptores (CCR5 ou CXCR4), possibilitando a ligação do vírus à superfície da célula-alvo; e gp41, que possui um peptídeo de fusão que é inserido na membrana quando a gp120 reconhece o correceptor, permitindo a entrada do vírus na célula (DITTMAR *et al.*, 1997).

O gene *tat* codifica a proteína transativadora (Tat), que atua principalmente promovendo a fase de alongamento da transcrição do HIV-1. Ela se liga a porção 5' terminal do RNA viral e induz a hiperfosforilação do domínio C-terminal da RNA polimerase II celular. Com isso, ela aumenta a afinidade da RNA pol II pelo promotor viral, aumentando a taxa de transcrição do genoma viral (PUMFERY *et al.*, 2003). O gene *rev* codifica para a proteína Rev, que regula a transição da replicação viral de uma fase precoce para uma tardia. Essa proteína é capaz de se ligar ao RNA mensageiro viral, estabilizando-o, permitindo sua exportação para o citoplasma e evitando o processamento (splicing) por proteínas celulares (MALIM *et al.*, 1989).

Por fim, os genes acessórios *vif*, *vpu*, *nef* e *vpr* codificam proteínas com funções variadas. A proteína Vif é responsável por aumentar a infecciosidade das novas partículas virais e a replicação viral, uma vez que ela é capaz de induzir degradação de fatores de restrição celular às infecções virais como as proteínas APOBEC3G e APOBEC3F (SCHUBERT *et al.*, 1996; STREBEL *et al.*, 1987; ZHENG e DANIELS, 2004). A proteína acessória Vpu, codificada pelo gene *vpu*, possui duas funções principais: atua na degradação intracelular de CD4 e auxilia o brotamento de partículas virais da célula infectada (BOUR e STREBEL, 2003). O gene *nef* codifica para o fator de regulação Nef, que atua diminuindo a expressão de proteínas na superfície de células T como as

moléculas CD4 e HLA de classe I, e expressão do fator de transcrição NF-kB e IL-2, o que provoca uma diminuição na eficiência da resposta imune de células T citotóxicas contra células infectadas, auxiliando assim na evasão do sistema imune do hospedeiro (GARCIA e MILLER, 1992; LURIA *et al.*, 1991; SCHWARTZ *et al.*, 1996). Por fim, a proteína Vpr atua nos estágios iniciais do ciclo replicativo do HIV, impedindo a progressão do ciclo celular, mantendo as células na fase G2/M e no auxílio do transporte do DNA viral para o núcleo (HEINZINGER *et al.*, 1994; JACOTOT *et al.*, 2000; JOWETT *et al.*, 1995).

1.13. CICLO REPLICATIVO DO HIV-1 E TERAPIA ANTIRRETROVIRAL

A terapia antirretroviral tem como objetivo impedir a replicação viral e restaurar a resposta imunológica do paciente. Dessa forma, os antirretrovirais diminuem o risco de transmissão do vírus, previnem a progressão da doença, complicações e mortes relacionadas à aids, contribuindo substancialmente para a melhoria na qualidade de vida dos pacientes infectados (HIRSCH *et al.*, 2008). O Brasil é tido como país referência na administração dos antirretrovirais pelo Sistema Único de Saúde (SUS), no qual os brasileiros infectados recebem o tratamento subsidiado pelo governo segundo orientações de tratamento do Ministério da Saúde. Além disso, desde outubro de 2013, a recomendação para início da terapia antirretroviral é independente da contagem de linfócitos T CD4+ para todos os indivíduos positivos para o HIV (MINISTÉRIO DA SAÚDE, 2018b).

Os diferentes medicamentos têm como alvo bloquear etapas importantes do ciclo replicativo viral, a fim de suprimir a replicação viral a níveis indetectáveis pelos métodos mais sensíveis e reduzir a mortalidade e a morbidade causadas pela aids. Existem seis classes de drogas antirretrovirais: inibidores nucleosídicos da transcriptase reversa (INTR), inibidores não-nucleosídicos da transcriptase reversa (INNTR), inibidores de protease (IP), inibidores de fusão (IF), antagonista de CCR5 (AC) e inibidores de integrase (II) (Figura 1.9).

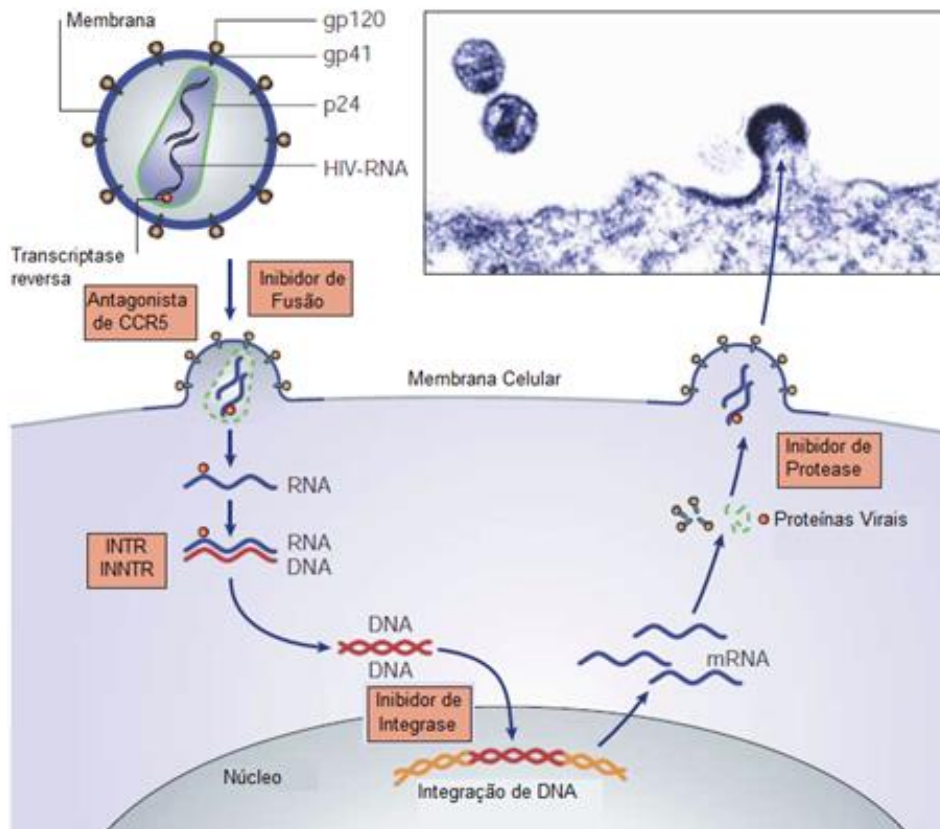


Figura 1.9 Ciclo replicativo do HIV-1 e atuação das classes de inibidores (modificada de SIMON e HO, 2003).

As células que expressam o receptor CD4 em sua superfície, como os linfócitos T auxiliares, monócitos, macrófagos, células dendríticas e da micróglia, são as células-alvo do HIV. O seu ciclo replicativo se inicia, portanto, com o reconhecimento e a ligação específica da glicoproteína viral gp120 à molécula de CD4 da célula-alvo. Esta interação promove uma alteração conformacional na gp120 permitindo sua ligação a seus correceptores, os receptores de quimiocina CCR5 ou CXCR4. Após esta interação, uma nova modificação conformacional na gp41 promove a exposição do peptídeo de fusão, que se insere na membrana celular permitindo a fusão das membranas celular e viral (BERGAMASCHI e PANCINO, 2010; DALGLEISH *et al.*, 1984; SATTENTAU e MOORE, 1993). Os IFs se ligam à gp41, inibindo a fusão das membranas viral e celular e os AC são moléculas que se ligam ao CCR5 e inibem sua interação com a gp120 (POMERANTZ e HORN, 2003; RICHMAN *et al.*, 2001).

Uma vez ocorrido o processo de fusão ocorre o desnudamento da

partícula viral, o capsídeo é liberado no citoplasma. O capsídeo é dissociado progressivamente, liberando o complexo pré-integrativo (PIC) que migra até o núcleo celular (GALLAY *et al.*, 1997; HEINZINGER *et al.*, 1994). Simultaneamente, ocorre a transcrição reversa do RNA viral, iniciado ainda no interior da partícula viral. O RNA viral é utilizado como molde para a síntese de uma fita de DNA complementar pela enzima transcriptase reversa viral (RT). À medida que o DNA viral é sintetizado, a fita-molde de RNA é degradada pela RNase H, um domínio C-terminal da RT. Em seguida, é sintetizada uma fita de DNA complementar àquela recém-sintetizada, originando uma molécula de DNA dupla-fita. Uma vez terminada a transcrição reversa, ocorre a integração do DNA viral na forma de provírus ao genoma do hospedeiro pela ação da proteína integrase (IN), onde ele servirá de molde para a transcrição dos genes virais (BOWERMAN *et al.*, 1989; FRANKEL e YOUNG, 1998).

O mecanismo de ação dos INTR se baseia em mimetizar nucleotídeos ou nucleosídeos que ao serem incorporados pela enzima RT impedem a continuidade da síntese da cadeia de cDNA, pois não possuem uma extremidade 3'OH livre. Os ITRNN atuam causando mudanças conformacionais na RT, pois inibem o sítio ativo da enzima ao se ligarem em um bolsão hidrofóbico próximo ao primeiro. Por fim, os IIs se ligam à integrase viral, bloqueando a integração do cDNA no genoma da célula hospedeira (POMERANTZ e HORN, 2003; RICHMAN *et al.*, 2001).

Após a integração, a transcrição e tradução do DNA viral será governada pela maquinaria celular com o auxílio de algumas proteínas virais (KNIPE *et al.*, 2001). Para que ocorra a produção de novos vírus é importante que as células contendo o provírus estejam ativadas. Os provírus integrados em células em repouso encontram-se em estado de latência, podendo permanecer assim por anos, gerando importantes reservatórios virais. A transcrição do DNA viral integrado origina moléculas longas de RNAs mensageiros (RNAm). Parte desses RNAm serão direcionados para os sítios de montagem de partícula viral abaixo da membrana celular para que sejam incluídos nas novas partículas virais como RNA genômico. Outra parte permanece no citoplasma onde será traduzida, dando origem às proteínas virais que também migrarão para os sítios de montagem. Os vírions brotam da célula carregando com ele uma bicamada lipídica da membrana celular da

célula hospedeira, que constituirá em seu envoltório. Após o brotamento, a protease (PR) cliva as poliproteínas virais Gag e Gag-Pol, dando origem às enzimas e proteínas estruturais, tornando o vírus maduro e infeccioso (CANN e KARN, 1989). Os IPs agem por competição, se ligam de forma irreversível ao centro ativo da PR, impedindo a maturação dos vírions (POMERANTZ e HORN, 2003; RICHMAN *et al.*, 2001).

Atualmente, utiliza-se uma estratégia de terapia antirretroviral combinada denominada HAART (terapia antirretroviral altamente eficaz - do inglês *Highly Active AntiRetroviral Therapy*). Essa estratégia é composta por três antirretrovirais pertencentes a duas classes diferentes, sendo na maioria das vezes dois INTRs acompanhado de um INNTR ou um IP. No Brasil, as recomendações do Ministério da Saúde são de que o esquema de tratamento inicial (primeira linha) deve incluir dois INTRs associados à um II (tenofovir - TDF, lamivudina - 3TC e dolutegravir - DTG) (MINISTÉRIO DASAÚDE, 2018b).

Um dos grandes desafios à terapia antirretroviral é o estado de latência do provírus integrado e o estabelecimento de reservatórios virais. A latência ocorre quando o vírus integra em uma célula que entra em seguida em repouso. Como todas as classes de antirretrovirais disponíveis até o momento atuam sobre pontos do ciclo replicativo do HIV, e vírus latentes não são atingidos por estas drogas, essas células funcionam como uma barreira (MARGOLIS, 2010). Outro grande desafio do tratamento antirretroviral seria a exposição dos pacientes a diversos efeitos colaterais decorrentes da toxicidade dos medicamentos e do estado inflamatório crônico relacionado à infecção pelo HIV (RYOM *et al.*, 2016).

Devido à alta taxa de replicação viral e à falta de atividade revisora da enzima RT, mutações podem ocorrer no genoma viral conferindo resistência ao antirretroviral utilizado pelo paciente. Algumas mutações podem conferir algumas vantagens seletivas em relação ao tratamento utilizado, tornando-se dominantes na população em relação aos vírus suprimidos pelos fármacos e podendo levar ao desenvolvimento de falha terapêutica (SANTORO e PERNO, 2013). Essas mutações de resistência aos antirretrovirais podem causar falha do tratamento, ser transmitidas para novos indivíduos e diminuir substancialmente as opções terapêuticas do paciente. Devido a esses fatores, a resistência aos antirretrovirais vem se mostrando um importante problema de

saúde pública (BENNETT *et al.*, 2009; SHAFER *et al.*, 2007).

1.14. NOVAS ESTRATÉGIAS CONTRA O HIV-1

A terapia antirretroviral, quando usada de forma ótima, consegue controlar a replicação viral, prevenir o desenvolvimento da AIDS, reduzir o risco de transmissão e prolongar a vida, diminuindo assim a mortalidade associada à infecção pelo HIV. Porém ela não é capaz de eliminar completamente o vírus presente nas células latentes (FINZI *et al.*, 1997).

Existem diversas limitações nas estratégias atuais de tratamento que acarretam desafios operacionais e logísticos no tratamento ao longo da vida. Os custos econômicos do fornecimento da terapia antirretroviral para as milhões de pessoas que atualmente vivem com o HIV vêm se tornando insustentáveis para os sistemas de saúde pública. Além disso, a adesão ao tratamento ao longo da vida é um desafio para muitos pacientes e pode levar a outros problemas como a resistência aos medicamentos antirretrovirais. Por fim, as toxicidades das drogas e a persistência da disfunção imune durante a terapia têm consequências significativas para a saúde (DEEKS *et al.*, 2016).

Dessa forma, com o tratamento atual para o HIV, os pacientes infectados, incluindo os pacientes em sucesso terapêutico, devem manter o uso dos antirretrovirais de forma crônica sem interrupções, sob o risco de rebote virológico dos reservatórios virais (DAVEY *et al.*, 1999; JOOS *et al.*, 2008).

Frente a esse contexto, novas estratégias vêm sendo estudadas pela comunidade científica, visando intensificar a eficácia da terapia antirretroviral ou evitar o uso crônico dos antirretrovirais e os problemas decorrentes a esse uso que diminuem a qualidade de vida dos pacientes infectados (ENSOLI *et al.*, 2014). Estas estratégias almejam alcançar a cura do paciente HIV+ através da eliminação do reservatório viral, ou uma cura funcional, através do controle imunológico da replicação do vírus mediante a interrupção do tratamento (DEEKS *et al.*, 2016; PERREAU *et al.*, 2017).

a) Vacina e cura do HIV-1

O grande interesse na cura do HIV reuniu diversas linhas de pesquisas sobre o vírus, dos estudos de biologia molecular e patogênese viral aos estudos de desenvolvimento de vacinas. No entanto, uma das principais barreiras para a cura do HIV é a latência, definida como a persistência do DNA viral integrado, competente para a replicação, porém transcricionalmente silencioso (DEEKS *et al.*, 2016). Após a infecção primária, o HIV-1 estabelece os reservatórios celulares, células T de memória CD4+ que se tornam latentes e os reservatórios anatômicos como o tecido linfóide associado à mucosa gastrointestinal (do inglês GALT; *gastrointestinal associated lymphoid tissues*), os sistemas nervoso e genital e os linfonodos, que não são totalmente permeáveis a alguns antirretrovirais. Os reservatórios latentes de células T CD4+ em repouso são considerados a principal barreira para curar a infecção pelo HIV (FINZI *et al.*, 1997; TUMIOTTO *et al.*, 2019).

Os ensaios de cura viral até o momento variaram de estimulação imunológica ou química de células T em repouso à vacinação antiviral, particularmente envolvendo epitopos de células T CD8+, uma vez que a importância da resposta citotóxica na diminuição da replicação viral durante a fase primária da infecção é bem conhecida (SILICIANO e SILICIANO, 2016).

Existem diversos desafios para o desenvolvimento de uma vacina contra o HIV, tais como a alta taxa mutacional do vírus que facilita a evasão do sistema imune e a ineficiência da vacina, a falta de um modelo de resposta imune natural contra o HIV e a falta de um modelo animal similar ao humano para auxiliar nos estudos. Estamos ainda longe de vislumbrar uma vacina capaz de combater o HIV, porém diversos estudos vêm sendo desenvolvidos e novas estratégias de vacinação vêm sendo estudadas (DEEKS *et al.*, 2016; SILICIANO e SILICIANO, 2016).

Como mencionado anteriormente, a resposta específica de linfócitos T CD8+ é uma das principais abordagens no desenvolvimento de vacinas eficazes contra o HIV. Isso é devido ao princípio de que as células T CD8+ são mediadoras eficientes da depuração viral. As células T CD8+ são consideradas as células efetoras que agem diretamente nas células infectadas, sendo cruciais para o controle imunológico da replicação viral (KOUP *et al.*, 2011;

MOTHE e BRANDER, 2018). O maior exemplo disso são os chamados “controladores de elite”, pacientes que apresentam um controle natural da infecção, mantendo a carga viral indetectável por muitos anos (MOTHE *et al.*, 2011; OLSON *et al.*, 2014).

b) Vacinação terapêutica

Embora ainda estejamos longe de vislumbrar a eliminação completa do HIV do paciente HIV+, muitos estudos tem se direcionado buscando a remissão prolongada da doença, uma forma de cura funcional do paciente HIV+. Nesse contexto, o paciente seria capaz de controlar o HIV através do seu próprio sistema imune, controlando a replicação e reduzindo os reservatórios virais, sem uso de terapia antirretroviral (SILICIANO e SILICIANO, 2016).

Diferentes vacinas terapêuticas já foram utilizadas em diversos ensaios clínicos testando-se diversas estratégias para alcançar o controle imunológico ideal da replicação do HIV-1 na ausência de antirretrovirais (PANTALEO e LEVY, 2016). Algumas estratégias vacinais visaram epítomos virais mais conservados, presentes em regiões conservadas do vírus; outras visaram combinações de peptídeos oriundos de diversas proteínas do HIV, junto com agentes reversores de latência para reativar o vírus dos reservatórios (STEPHENSON, 2018). Atualmente, novas vacinas baseadas na combinação de imunoterapia, modulando e mobilizando o sistema imune para atacar o HIV-1, são potentes intervenções que visam restaurar a função celular e eliminar os reservatórios virais. Apesar da ampla gama de vacinas terapêuticas já desenvolvidas, o efeito no controle da replicação viral após a interrupção do tratamento antirretroviral ainda é modesto (SEDDIKI e LEVY, 2018).

As falhas de vacinas terapêuticas testadas podem ter sido causadas por gerarem uma especificidade sub-ótima e uma resposta de célula T com amplitude limitada, devido ao design imunogênico inadequado, sua estratégia de entrega ou ambos (SILICIANO e SILICIANO, 2016). Ou seja, essas vacinas não tiveram resposta imune suficientemente abrangente para lidar com diversos vírus que emergem, não foram capazes de reduzir efetivamente os

reservatórios virais, ou não reduziram a carga viral. Outro possível motivo seria porque essas vacinas testadas acabaram gerando uma resposta modesta de células T CD8+, pouco imunogênica, que não é suficiente para controlar o rebote viral (HANCOCK *et al.*, 2017; LI *et al.*, 2011; SCHOOLEY *et al.*, 2010; STEPHENSON, 2018). Algumas vacinas testadas, apesar de reduzirem o DNA viral integrado e a quantidade de células T de memória latentes que carregam provírus, não aumentaram o tempo médio de rebote viral mais do que os controles (POLLARD *et al.*, 2014; STEPHENSON, 2018; TAPIA *et al.*, 2017).

Como uma das estratégias de cura funcional temos a vacinação terapêutica com epítomos oriundo de proteínas virais como, por exemplo, Gag, Pol e Nef, conhecidos por conter numerosos epítomos imunogênicos e potencial de reconhecimento pelas moléculas de HLA de classe I (MOTHE e BRANDER, 2018; PAPUCHON *et al.*, 2014; TUMIOTTO *et al.*, 2019). Porém, existem grandes desafios no desenho da vacina como a alta variabilidade entre as proteínas virais, assim como o alto grau de variação e diferença de distribuição dos alelos de HLA (KARLSSON *et al.*, 2012).

Um dado bastante importante acerca dessas vacinas se refere ao fato de que a maioria foi desenvolvida com base em um HIV genérico, e essas cepas circulantes podem conter variações, especialmente em relação aos epítomos que são apresentados pelo HLA de classe I. Ao interromper o tratamento, o vírus volta a se multiplicar e ele se origina principalmente dos reservatórios celulares representados pelas células T de memória que são reativadas. Ou seja, o vírus que foi arquivado é provavelmente a imagem do que será reativado (DENG *et al.*, 2015; FLEURY *et al.*, 2018; PAPUCHON *et al.*, 2014; PAPUCHON *et al.*, 2013).

Considerando todos os desafios apresentados acima, Papuchon e colaboradores questionaram que as melhores vacinas imunogênicas conteriam epítomos virais que se encontram presentes nos reservatórios provirais arquivados, e que seriam restritos pelo HLA através da apresentação ao sistema imunológico. Com base nisso, o cenário ideal para testar tal abordagem seria determinar alelos de HLA classe I e os epítomos virais arquivados do indivíduo, formando uma vacina personalizada, a fim de que se o rebote viral ocorrer, o indivíduo vacinado estará pronto para restringir especificamente os epítomos dos vírus originados a partir desse compartimento

(FLEURY *et al.*, 2018; PAPUCHON *et al.*, 2014; PAPUCHON *et al.*, 2013; TUMIOTTO *et al.*, 2019; TUMIOTTO *et al.*, 2017).

Esse novo conceito de vacina terapêutica personalizada é a base de um estudo multicêntrico ao qual esse trabalho está vinculado. Vários centros de pesquisa de diferentes regiões, não só do Brasil, como da França, Canadá, Peru e Estado Unidos, se uniram para desenvolver esse projeto. Como colaboradores nacionais, fazem parte o INCA, o Hospital Federal de Ipanema (RJ), o Hospital Universitário Clementino Fraga Filho-UFRJ (RJ) e o Hospital Universitário Dr. Miguel Riet Corrêa Jr.- FURG (RS). Como colaborador internacional temos o Dr. Hervé Fleury da Universidade de Bordeaux na França. A união de diferentes regiões do mundo implica em pacientes com diferentes constituições genéticas e também diversidade viral, dessa forma superando limitações inicialmente encontradas por Papuchon e colaboradores.

Esse projeto multicêntrico tem como objetivo a prova de conceito para o desenho de uma vacina terapêutica personalizada capaz de proporcionar a cura funcional de indivíduos HIV+ sob sucesso terapêutico há pelo menos um ano e no primeiro esquema de tratamento. Dessa forma, propõe analisar os epítomos mais invariáveis presentes nas quasispécies provirais arquivadas que seriam capazes de provocar respostas citotóxicas efetivas, ajudando a controlar a replicação viral após a descontinuação do tratamento antirretroviral. Por fim, visando uma estratégia de saúde pública de vacinação terapêutica anti-HIV, imunógenos mais universalmente ativos poderiam ser criados para grupos com alelos de HLA mais frequentes nas regiões analisadas, selecionando os epítomos do HIV mais invariáveis que sejam restritos a esses alelos.

2. JUSTIFICATIVA

Na presente proposta, seremos capazes de determinar a composição do alelo HLA de classe I (loci A, B e C) de pacientes oriundos de duas regiões do Brasil (Sudeste e Sul). A nova tecnologia de sequenciamento de nova geração aplicada à determinação do HLA irá gerar sequências de alelos com resolução ultraprofunda, possibilitando a descrição de novos alelos de HLA presentes na

população brasileira. A determinação da combinação de alelos de HLA de classe I e a identificação de epítomos conservados de proteínas do HIV-1 proviral são peças-chave para o desenvolvimento de novas estratégias como a vacinação terapêutica de pacientes HIV+. Isto, mais uma vez, irá beneficiar não só o nosso projeto, mas também muitos outros realizados no Brasil e no mundo.

3. OBJETIVOS

Determinar os alelos do HLA de classe I (A, B e C) de pacientes HIV+ em terapia antirretroviral com sucesso terapêutico das regiões Sudeste e Sul do Brasil por sequenciamento de nova geração e selecionar epítomos virais conservados com alta afinidade aos alelos encontrados para o desenho de uma vacina terapêutica personalizada.

3.1. OBJETIVOS ESPECÍFICOS

- Determinar os alelos de HLA de classe I nos loci A, B e C dos pacientes HIV-1+ de dois centros da região Sudeste (Rio de Janeiro - RJ) e de um da região Sul (Rio Grande – RS) do Brasil;
- Aplicar a nova metodologia de tipagem do HLA de alta resolução através de sequenciamento de nova geração (NGS);
- Caracterizar potenciais novos alelos de HLA de classe I presentes na população brasileira através de reconstrução completa dos genes;
- Predizer epítomos provirais do HIV-1 com alta afinidade aos alelos de HLA-A, B e C mais frequentes na população estudada;
- Analisar a afinidade dos epítomos selecionados aos alelos de HLA-A, B e C específicos dos pacientes de um dos centros estudados;
- Investigar variantes intra-hospedeiro dos epítomos virais selecionados;
- Analisar a afinidade das variantes encontradas aos alelos de HLA-A, B e C específicos do paciente.

4. MATERIAL E MÉTODOS

4.1. COLABORAÇÕES

Este estudo faz parte de um projeto maior que envolve diversas instituições colaboradoras. Dentre elas temos o Hospital Federal de Ipanema (HI), o Hospital Universitário Clementino Fraga Filho (HUCFF-UFRJ) e o Hospital Universitário Dr. Miguel Riet Corrêa Jr. (FURG-RS), que são os responsáveis pelo recrutamento dos pacientes acompanhados nesses hospitais e pela coleta das amostras. O INCA, por sua vez, é o responsável pela amplificação, sequenciamento e análise dos genomas provirais arquivados e dos alelos de HLA-A, B e C, assim como a tipagem desses alelos. Como colaborador internacional temos a Universidade de Bordeaux na França, responsável pela síntese dos peptídeos virais e testagem *in vitro* da resposta imunológica.

4.2. CASUÍSTICA

Foram selecionados para esse estudo 80 pacientes HIV-1 positivos, 40 do Rio de Janeiro, compreendendo pacientes do Hospital Federal de Ipanema (HI) e do Hospital Universitário Clementino Fraga Filho da Universidade Federal do Rio de Janeiro (HUCFF-UFRJ), e 40 pacientes acompanhados no Hospital Universitário Dr. Miguel Riet Corrêa Jr. da Universidade Federal de Rio Grande (FURG-RS), da cidade de Rio Grande, no Rio Grande do Sul. Para a cidade do Rio de Janeiro foram selecionados dois centros de coletas, HI e HUCFF-UFRJ, uma vez que no HI não foi possível completar os 40 pacientes propostos no estudo. Em vista disso, a coleta dos pacientes foi estendida também para o HUCFF-UFRJ, completando assim o total de pacientes elegíveis para esse projeto na cidade do Rio de Janeiro.

Esses pacientes possuem acompanhamento clínico regular e foram selecionados durante a visita de rotina. Os pacientes foram convidados a participar do estudo e assinaram um Termo de Consentimento Livre e Esclarecido (TCLE). O projeto foi aprovado pelos Comitês de Ética em

Pesquisa do Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), do HUCFF-UFRJ e do FURG-RS (CAAEs 52862016.9.0000.5274, 56604816.2.0000.5257 e 52862016.9.3001.5324, respectivamente).

Os critérios de inclusão para participar do estudo foram: idade maior ou igual a 18 anos, estar sob HAART de primeira linha ou na segunda linha de tratamento (neste último caso devido apenas a mudanças recomendadas pelo Ministério da Saúde), e estar em sucesso terapêutico durante pelo menos 12 meses (controle virológico, com carga viral do HIV indetectável) antes da coleta. Os critérios de exclusão foram: ter histórico de falha virológica prévia, mudança do esquema terapêutico por intolerância ou má adesão, estar em acompanhamento ou tratamento por menos de 12 meses, se recusar a participar do estudo ou não assinar o TCLE. Foi usado como critério de preferência para a seleção dos pacientes não ter sido previamente classificado em estágio clínico e/ou imunológico de aids (que tiveram doença definidora de aids ou valor de contagem de linfócitos T CD4+ menor ou igual a 200 cels/mm³, segundo critério CDC, 2014). No entanto, caso o número de pacientes não atingisse o número total definido para cada centro (40 pacientes), pacientes que não atendessem o critério de preferência também poderiam ser recrutados.

O questionário clínico e epidemiológico foi preenchido conforme os dados presentes nos prontuários médicos dos pacientes durante a visita onde foi realizada a coleta. Ele continha informações como: sexo, data de nascimento, data de diagnóstico para o HIV, histórico de carga viral plasmática do HIV e de contagens de linfócitos T CD4+, esquema terapêutico utilizado e seu tempo de uso, e provável forma de transmissão do vírus.

Durante a visita de rotina ao hospital era feita a coleta de 6 ml de sangue total periférico de cada paciente. Esse material era encaminhado para o Programa de Oncovirologia do Centro de Pesquisas do Instituto Nacional de Câncer José Alencar Gomes da Silva, onde eram separados o plasma e a parte celular. Este material era armazenado no freezer -80°C para posterior análise molecular. Para os pacientes da FURG-RS, a separação das PBMCs era realizada pelos colaboradores responsáveis locais e as mesmas eram armazenadas a -80°C até serem enviadas para nosso laboratório no INCA, RJ.

4.3. EXTRAÇÃO DO DNA GENÔMICO DOS PACIENTES

A parte celular armazenada foi processada para realizar a extração do DNA genômico com o kit de extração *Genomic DNA Extraction* (Real Genomics, BioAmericanInc, Miami, EUA), conforme o protocolo do fabricante. Primeiramente, foram misturados por inversão 900 µl do tampão de lise *RBC Lysis Buffer* e 300 µl da amostra descongelada. Essa mistura foi mantida a temperatura ambiente por cinco minutos e centrifugada a 3.000 x g por dois minutos. Em seguida, o sobrenadante foi descartado e o precipitado foi ressuspenso em 100µl do tampão de lise *RBC Lysis Buffer*. Após essa etapa, foram adicionados 200µl do tampão *GB Buffer*. Essa mistura foi vigorosamente agitada e posteriormente incubada por 10 minutos a temperatura ambiente, invertendo-se a mistura a cada três minutos. Em seguida, adicionou-se ao lisado celular 200µl de etanol absoluto (Merck, Darmstadt, Alemanha) e misturou-se por 10 segundos no Vortex®. A mistura foi transferida para a coluna de extração que possui uma membrana seletiva de sílica e centrifugada a 13.000 rpm por cinco minutos. Após essa etapa, foram realizadas duas etapas de lavagem, a primeira com 400µl do tampão *W1* seguida de centrifugação a 13.000 rpm por 30 segundos. O líquido centrifugado foi descartado e a segunda etapa de lavagem foi realizada com 600µl do tampão *Wash Buffer*, centrifugando-se a 13.000 rpm por três minutos e 30 segundos, e também descartando-se o líquido centrifugado. Ao final, o DNA genômico retido na membrana de sílica foi eluído em 100µl do tampão de eluição *Elution Buffer* aquecido a 70°C. Após incubação da mistura a temperatura ambiente por cinco minutos, foi realizada uma centrifugação a 13.000 rpm por um minuto. A coluna foi descartada e o DNA genômico eluído foi armazenado a -20 °C.

As amostras tiveram sua concentração de DNA mensurada utilizando-se um espectrofotômetro Nanodrop ND 1000 (ThermoScientific, Waltham, EUA). Todas as amostras foram medidas em duplicatas de 1 µL de amostra e considerando os valores de pureza de DNA iguais a 1,8 pela razão 260/280nm. Amostras com concentração inferior a 50 ng/µL e valores de pureza de DNA abaixo de 1,4 foram novamente extraídas até que se obtivessem amostras aceitáveis dentro dos padrões estabelecidos.

4.4. AMPLIFICAÇÃO DOS GENES DE HLA-A, B E C

A amplificação dos genes dos HLA-A, B e C foi realizada por reação em cadeia da polimerase (PCR) utilizando iniciadores e metodologia descrita por Ehrenberg e colaboradores (Tabela 4.1) (EHRENBERG *et al.*, 2014). Os genes de HLA-A, B e C foram amplificados separadamente em fragmentos de aproximadamente 5.466 pb, 4.609 pb e 4.802 pb, respectivamente, englobando as regiões promotoras e os éxons 1 ao 7.

Tabela 4.1 Descrição dos iniciadores usados para a amplificação dos alelos de HLA-A, B e C. A posição dos iniciadores é relativa às sequências-referência de HLA-A (NG_029217.2), HLA-B (NG_023187.1) e HLA-C (NG_029422.2) disponíveis na base de dados do NCBI.

Lócus	Iniciadores	Sequencia 5' - 3'	Orientação	Posição
A	AF1	AAC TCA GAG CTA AGG AAT GAT GGC AAA T	Senso	4248-4275
	AF2	AAC TCA GAG CTA TGG AAT GAT GGT AAA T	Senso	4248-4275
	AR1	ATA TAA CCA TCA TCG TGT CCC AAG GTT C	Antissenso	9684-9711
B	BF1	CCC GGT TGC AAT AGA CAG TAA CAA A	Senso	4170-4194
	BR1	GGG TCC AAT TTC ACA GAC AAA TGT	Antissenso	8755-8778
C	CF1	TGC TTA GAT GTG CAT AGT TCA CGA A	Senso	4017-4041
	CF2	TGC TTA GAT GTG CAT AGT TCC GGA A	Senso	4017-4041
	CR1	TGG ACC CAA TTT TAC AAA CAA ATA	Antissenso	8795-8818

As reações para a PCR foram realizadas utilizando os seguintes reagentes: 2,5 µL de tampão 10x, 0,75 µL de MgCl₂ a 25 mM, 0,2 µL de dNTPmix a 25 mM, 0,25 µL de cada iniciador a 25 pmol/µL, 0,2 µL de *Taq*

Platinum High Fidelity DNA polimerase a 5 U/ μ L (Invitrogen, São Paulo, Brasil), 100 ng de DNA genômico e um volume de água livre de nucleasse (Life Technologies, Carlsbad, EUA) para totalizar um volume final de 25 μ L para cada reação. Foram feitos controles negativos contendo apenas os reagentes usados para a PCR com a função de monitorar as reações para identificar possíveis contaminações.

As condições para a ciclagem no termociclador *Veriti 96 well Thermal Cycler* (Life Technologies, Carlsbad, EUA) para as reações da PCR foram: ativação inicial a 94°C por dois minutos, 35 ciclos com desnaturação a 94°C por 30 segundos, anelamento dos iniciadores a 55°C por 45 segundos e extensão a 68°C por cinco minutos e 50 segundos, seguidos de 10 minutos de extensão final dos fragmentos a 68°C. Para a etapa de anelamento, foi utilizada a temperatura correspondente à média dos TM dos iniciadores da reação.

O resultado da amplificação foi observado em gel de agarose a 1% em solução tampão de NaOH 1x (solução estoque a 20x: ácido bórico a 0,9 M e NaOH a 0,2 M; ambos da Merck, Darmstadt, Alemanha). Foram aplicados no gel 5 μ L do produto amplificado misturados a 1 μ L do corante *Gel Red*[®] (Biotium, Hayward, CA) / *SYBR Safe DNA Gel Stain* (ThermoScientific, Waltham, USA), além de 2 μ L do marcador de peso molecular 1 kb *DNA Ladder* (Amersham Biosciences, Baie-D'urfe, Canadá) e o gel foi submetido à eletroforese (100 mAmp). O gel foi analisado em um transiluminador de luz ultravioleta (Bio-Rad, Califórnia, Estados Unidos) e fotografado utilizando software *Kodak Molecular Imaging Software* (Eastman Kodak Company, Nova York, EUA). O tamanho dos fragmentos desejados foi estimado comparando-se os mesmos com o marcador de peso molecular. Os produtos da amplificação foram armazenados a 4°C até serem purificados.

4.5. PURIFICAÇÃO DOS PRODUTOS DE PCR

Os produtos das PCR que tiveram amplificações bem sucedidas foram purificados utilizando o kit *GFXTM PCR DNA and Gel Band Purification* (GE Healthcare, Chalfont St. Giles, UK), segundo protocolo do fabricante. Inicialmente, foram misturados 500 μ L do tampão de captura *Capture Buffer*

Type 3 ao volume total do produto da PCR, e então a mistura foi transferida para um tubo contendo uma membrana de sílica. Após uma centrifugação a 13.000 rpm por um minuto, o DNA ficou retido na membrana e o líquido centrifugado foi descartado. Em seguida, foram adicionados 500 µL de tampão de lavagem *Wash Buffer Type 1*, e foram realizadas duas centrifugações consecutivas a 13.000 rpm por um minuto para que a membrana ficasse bem seca. Foram adicionados 40 µL do tampão de eluição, e o material foi incubado por 10 minutos. Por fim, foi feita uma centrifugação a 13.000 rpm por dois minutos e o produto centrifugado foi armazenado a -20°C.

Os produtos purificados foram novamente analisados em gel de agarose a 1%, utilizando uma alíquota de 2 µL do purificado. Em seguida, eles foram quantificados no espectrofotômetro Nanodrop ND 1000 (ThermoScientific, Waltham, EUA) e diluídos em água ultrapura a uma concentração de 0,6 ng/µL. Por fim, 10µL de cada produto purificado referente aos HLA-A, B e C de um mesmo paciente foram misturados em um único tubo.

4.6. PREPARAÇÃO DAS BIBLIOTECAS GENÔMICAS

Os produtos purificados de cada paciente foram utilizados para preparação das bibliotecas com o kit *Nextera XT DNA Sample Preparation* (Illumina, San Diego, EUA) seguindo o protocolo do fabricante.

Na primeira etapa, ocorre simultaneamente a fragmentação do DNA e sua marcação com a inserção de sequências-etiquetas (*tags*) específicas. Para essa etapa, foram adicionados 5 µL do DNA a 0,6 ng/µL, 10 µL do tampão *TD (Tagment DNA Buffer)* e 5 µL de *ATM (Amplicon Tagment Mix)*. A reação foi incubada em um termociclador a 55 °C por cinco minutos. Nessa etapa, o DNA é fragmentado em regiões aleatórias por transposons *Nextera XT transposome* (complexo transposon-transposase) e juntamente com a quebra do DNA, são adicionados na extremidade do DNA duas sequências terminais (*tags*) de 14 pb que auxiliam na incorporação de novos identificadores utilizados nas etapas seguintes. Para interromper a ação da enzima, logo após essa etapa, adicionaram-se 5 µL do tampão *NT (Neutralize Tagment Buffer)* na reação.

A segunda etapa, de enriquecimento dos fragmentos, se constitui em

uma reação de PCR com iniciadores complementares aos que foram adicionados previamente e que possuem na sua extremidade uma sequência adaptadora. Esses adaptadores (P5 e P7) são necessários para a etapa de clusterização, uma vez que eles são complementares aos adaptadores presentes para a formação do cluster durante o sequenciamento. Além disso, esses adaptadores possuem uma sequência variável de nucleotídeos (índices; i5 e i7), necessários para separação e identificação dos fragmentos originados do sequenciamento de uma mesma amostra. Primeiramente, foram adicionados 5 µL de cada índice e 15 µL de *NPM (Nextera PCR Master Mix)* ao produto da primeira etapa. A reação foi submetida a uma PCR em termociclador, com uma ciclagem de 72°C por três minutos, 95°C por 30 segundos e 12 ciclos de 95°C por 10 segundos, 55°C por 30 segundos e 72°C por 30 segundos.

A última etapa de purificação da biblioteca foi realizada com microesferas magnéticas (*beads*), que além de purificar a biblioteca ao se ligar ao DNA, separando-o do restante da solução, também promovem uma seleção de tamanho dos fragmentos, retirando os que são muito pequenos. Nessa etapa foi utilizada uma placa de 96 poços e foram adicionados 30 µL de *AMPure XP beads* (Agencourt Bioscience Corporation, Beverly, EUA) em cada reação, misturando-se até gerar uma solução homogênea. Após cinco minutos de incubação, a placa foi colocada sobre uma placa magnética por dois minutos para a precipitação das microesferas ligadas ao DNA no fundo dos poços, e então o sobrenadante foi removido. Foram feitos dois processos de lavagem com etanol fresco a 80% (Merck, Darmstadt, Alemanha) e, após uma secagem de 15 minutos, as microesferas contendo o DNA foram ressuspensas em 32,5 µL de tampão *RSB (Resuspension Buffer)*. O *RSB* possui maior afinidade pelas microesferas, liberando o DNA na solução. Esses 30 µL de solução contendo o DNA, ou seja, a biblioteca, foram armazenados a -20°C. Um esquema da construção das bibliotecas pode ser observado na Figura 4.1 abaixo.

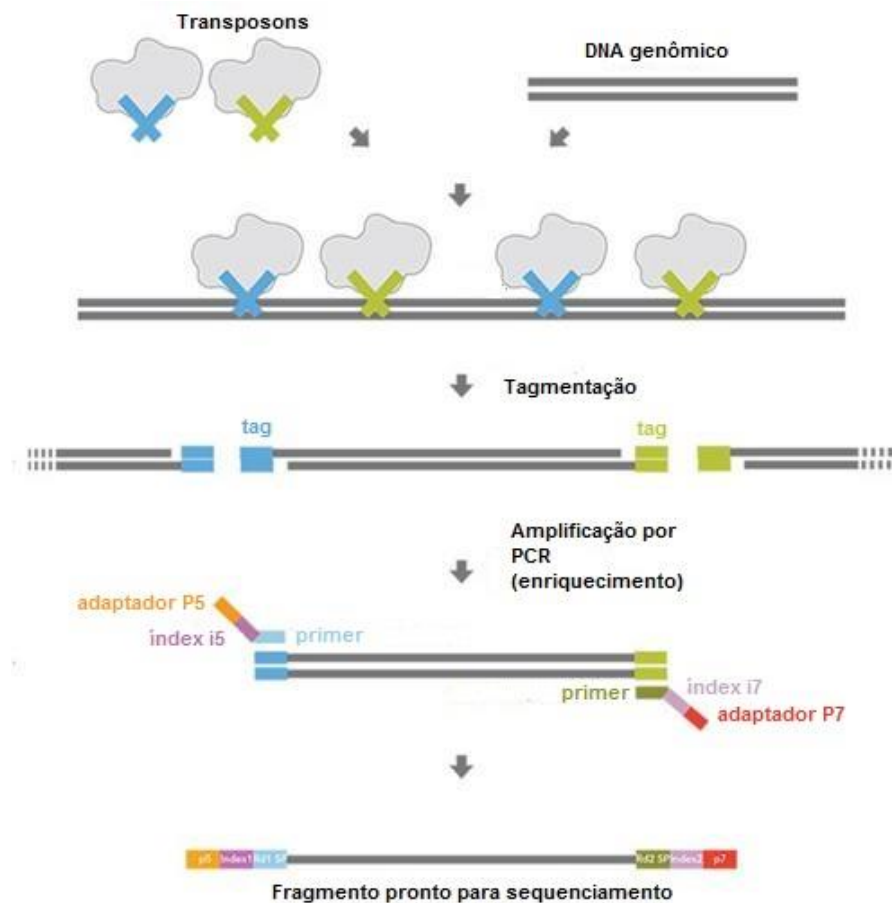


Figura 4.1 Representação esquemática das etapas de preparação das bibliotecas de DNA com o uso do kit Nextera DNA Sample Preparations para o sequenciamento na plataforma Illumina MiSeq (modificado de <http://core-genomics.blogspot.com.br/2012/07/illuminas-nextera-capture-is-this.html>).

4.7. QUANTIFICAÇÃO DAS BIBLIOTECAS

O tamanho médio dos fragmentos de DNA das bibliotecas foi visualizado por eletroforese em gel de agarose a 1% junto ao marcador de peso molecular de 1 kb como comparador.

A quantificação da biblioteca foi realizada através de quantificação absoluta por PCR em tempo real ou de quantificação por método fluorimétrico, ambos métodos recomendados pela Illumina.

A quantificação absoluta por PCR em tempo real foi realizada no sistema ECO (Illumina, San Diego, EUA) utilizando o kit *KAPA Library Quantification* (KAPA Biosystems, Wilmington, EUA) conforme protocolo

fornecido pelo fabricante. Inicialmente, as bibliotecas foram diluídas 10.000 vezes em solução contendo 0,1% de *Tween* 20 + 10 mM Tris-HCL a pH 8,5. As reações foram preparadas em microplacas fornecidas com o kit. Como padrões para a quantificação, foram utilizados seis padrões fornecidos no kit, com quantificações conhecidas e de 20 pM, 2 pM, 0,2 pM, 0,02 pM, 0,002 pM e 0,0002 pM. Quatro microlitros da diluição da biblioteca foram adicionados a 6 µL de *Master Mix*. Os controles negativos foram feitos adicionando-se 4 µL da mesma solução de *Tween* descrita acima ao invés da biblioteca. Todas as reações foram feitas em duplicatas para melhor confiabilidade da quantificação. A microplaca foi submetida a uma ciclagem de 95°C por 5 minutos, 35 ciclos de 95°C por 30 segundos e 60°C por 1 minuto e 30 segundos. A concentração de cada biblioteca foi normalizada pelo tamanho médio dos fragmentos obtidos pela eletroforese e pela taxa da diluição, chegando-se a uma concentração final.

Essa quantificação descrita acima foi realizada para as bibliotecas das amostras do HI, as primeiras a serem produzidas. Já as bibliotecas posteriores (amostras de HU-UFRJ e HU-FURG) foram quantificadas utilizando o método fluorimétrico de quantificação no sistema Qubit (Life Technologies, Carlsbad, EUA). Foi preparada uma solução de 199 µL de *master mix* e 1 µL de corante por amostra. Foram acrescentados 190 µL dessa solução em dois tubos fornecido pelo fabricante e adicionados 10 µL do padrão 0 e do padrão 10. Nos tubos de amostra, foram utilizados 198 µL da solução e 2 µL de cada biblioteca.

A concentração obtida foi ajustada pelo tamanho médio da biblioteca, chegando-se a uma concentração final de cada biblioteca. Posteriormente, todas as bibliotecas foram diluídas e reunidas de forma a obter um produto final a 4 nM.

4.8. SEQUENCIAMENTO DE NOVA GERAÇÃO NA PLATAFORMA MISEQ ILLUMINA

Com a colaboração do professor Dr. Rodolpho Albano do Departamento de Bioquímica da UERJ e com o Departamento de Genética da UFRJ, os

sequenciamentos das bibliotecas foram realizados nas plataformas MiSeq da Illumina presentes em cada laboratório.

As bibliotecas genômicas foram reunidas e diluídas a 4 nM. Em seguida, o *pool* de bibliotecas foi diluído a 12 pM, misturado a uma biblioteca controle *PhiX Control v3* (Illumina, San Diego, EUA) e o mix foi desnaturado usando NaOH fresco. O produto final foi então sequenciado usando o protocolo “*paired end 600*” (dois ciclos de 301 pb utilizando os iniciadores senso e antissenso, respectivamente) com o kit MiSeq v3 (Illumina, San Diego, EUA).

A metodologia de sequenciamento por síntese da plataforma Illumina MiSeq se inicia com a amplificação de fragmentos de DNA (bibliotecas) ligados à superfície de uma lâmina (*flowcell*) na forma de pontes através da enzima DNA polimerase, que produz múltiplas cópias da sequência original (*clusters*), se assemelhando à técnica da clonagem. Nesta etapa de clusterização, ocorre a ligação dos adaptadores presentes nas extremidades dos fragmentos das bibliotecas (P5 e P7) a sequências complementares que estão presentes na superfície da lâmina. O fragmento de DNA desnaturado irá se ligar em apenas uma extremidade, e ocorrerá um processo de cópia deste. Essa dupla-fita de fragmento de DNA (fragmento original + fragmento cópia) se ligará pela outra extremidade livre ao adaptador complementar presente na superfície da lâmina, formando uma estrutura de ponte e em seguida será desnaturada novamente. Essas etapas irão se repetir e os fragmentos serão copiados várias vezes.

Após a clusterização, ocorre o sequenciamento por síntese. Durante o sequenciamento das múltiplas cópias, são adicionados nucleotídeos fluorescentes com a terminação OH bloqueada, a fim de que seja adicionado apenas um a cada etapa. Em seguida, uma excitação por feixe de luz faz com que seja emitida uma fluorescência, que é diferente para cada um dos quatro nucleotídeos diferentes. A cada adição de uma única base, a lâmina é escaneada e são feitas várias aquisições de imagens. As imagens são interpretadas em atribuições de bases do DNA, identificando o nucleotídeo adicionado pela sua cor emitida. A lâmina é lavada e segue-se a seguinte adição de nucleotídeo. O ciclo de adição de nucleotídeo e aquisição de imagem é repetido por duas vezes, sendo sequenciados 301 nucleotídeos da extremidade do fragmento no sentido senso e 301 nucleotídeos no sentido

antissenso, caracterizando uma corrida com extremidades pareadas (*paired end*, 2x301).

As sequências oriundas do sequenciamento se denominam *reads* e estão no formato FASTQ. Esse arquivo contém a sequência de bases nucleotídicas e a qualidade correspondente a cada base. A qualidade da base é apresentada na escala de Phred, uma escala logarítmica ligada à probabilidade de erro de determinação da base (*base-calling*) usada para caracterizar a qualidade da sequência de DNA. Pode ser usada também para comparar a eficácia entre diferentes métodos de sequenciamento. Nessa escala, um valor de Q=20 significa que a probabilidade de incorporação errada da base é de 1 em 100, ou seja, a acurácia de determinação da base é de 99%.

4.9. ANÁLISE DE QUALIDADE DOS *READS*

Inicialmente, os *reads* das amostras foram analisados no programa FastQC (Babraham Bioinformatics, Cambridge, UK). Esse programa gera um relatório onde podemos observar a quantidade de *reads* gerados para cada biblioteca, a qualidade relativa a cada base e a qualidade média da sequência, o tamanho médio de cada *read*, além de outras características.

4.10. TIPAGEM DO HLA PELO SOFTWARE *HLA TWIN*

Para a tipagem dos alelos de HLA-A, B e C foi utilizado o software *HLA Twin* (Omixon, Budapeste, Hungria). Esse software realiza tipagem em alta resolução de todo o gene dos *loci* de HLA de classe I e II e vêm sendo utilizados em estudos de tipagem de HLA por NGS (PROFAIZER *et al.*, 2016; WEIMER, 2016). Esse programa realiza duas análises independentes, uma utilizando um algoritmo para mapear os *reads* contra sequências-referência publicadas no banco de dados IMGT-HLA (EMBL-EBI, UK). O IMGT-HLA fornece um banco de dados especializados em sequência do Sistema de Histocompatibilidade Humana (HLA) e inclui todas as sequências oficiais do Comitê de Nomenclatura da OMS para fatores do Sistema HLA. A segunda

metodologia de análise do software utiliza um algoritmo de alinhamento *de novo*. Após analisar a concordância do resultado da tipagem pelos dois algoritmos, o programa faz 23 verificações para cada *locus* de cada amostra onde são analisados parâmetros como desequilíbrio de alelos, comprimento das *reads*, qualidade das *reads*, ruído, tamanho do fragmento, cobertura de consenso, determinação da fase, além de outros fatores. Dessa forma, é possível obter uma maior confiabilidade na tipagem e encontrar alelos novos.

4.11. ANÁLISE DE ALELOS NOVOS DO HLA-A, B E C

Os possíveis alelos novos indicados pelo software e as tipagens ambíguas (quando mais de um alelo é selecionado como o melhor candidato) passaram por uma análise de tipagem *in-house* desenvolvida durante em minha Dissertação de Mestrado.

a) Filtragem dos *reads*

Primeiramente, os *reads* foram filtrados no programa *Sickle-Master* (disponível em <https://github.com/najoshi/sickle>), selecionando os *reads* com qualidade superior a Q=28 na escala Phred, o que significa uma acurácia maior que 99%, e com tamanho superior a 70 pb.

b) Reconstrução dos alelos através de montagem com referência

Para a montagem dos alelos de HLA foi utilizada a metodologia de montagem com referência com o programa *BWA – Burrows Wheeler Aligner* (LI e DURBIN, 2009). Os *reads* filtrados foram mapeados às sequências-referência dos alelos mais próximos determinados pelo programa *Assign 2.0* ou *HLA Twin* retiradas do banco de dados do IMGT-HLA. Além do *BWA*, foram utilizadas outras ferramentas como o *Samtools* (LI *et al.*, 2009) e *Picard* (disponível em: <http://picard.sourceforge.net>) para manipulação de referências e formatos de arquivo. A identificação das variantes (polimorfismos) com relação às

sequências-referência foi realizada utilizando a ferramenta *UnifiedGenotyper* do software *GATK* (MCKENNA *et al.*, 2010). Por fim, o alinhamento foi visualizado no programa *IGV* (do inglês, *Integrative Genomics Viewer*) (ROBINSON *et al.*, 2011).

4.12. COMPARAÇÃO DAS FREQUÊNCIAS DOS HLA-A, B E C ENCONTRADOS COM A POPULAÇÃO DE DOADORES DE MEDULA ÓSSEA CADASTRADOS NO REDOME

Os HLAs-A, B e C mais frequentes encontrados nas populações estudadas foram comparados com os dados de frequência de HLA da população doadora de medula óssea do banco de dados do Registro Nacional de Doadores Voluntários de Medula Óssea (REDOME - INCA). O REDOME é o terceiro maior banco de doadores de medula óssea do mundo com mais de 3,7 milhões de doadores cadastrados. Os dados de frequência dos grupos alélicos HLA-A, HLA-B e HLA-DRB1 de 2.847.869 doadores cadastrados no REDOME, distribuídos por UF de residência e grupo étnico do doador são disponibilizados pela *Rede Brasil Imunogenética* no site <http://imunogenetica.org/resultados/perfil-genomico-do-redome-rereme/>.

As análises estatísticas foram realizadas comparando as frequências dos alelos de HLA-A, B e C encontrados nas casuísticas do Rio de Janeiro e na do Rio Grande contra as frequências encontrada nos dados do REDOME dos estados do Rio de Janeiro (RJ) e do Rio Grande do Sul (RS). Foi utilizado o teste Exato de Fisher e os valores de $p \leq 0,05$ foram considerados estatisticamente significativos para diferenças entre as frequências em nossos pacientes e no REDOME.

4.13. ANÁLISE DE PREDIÇÃO DE EPÍTOPOS DE CÉLULAS T E DE DIVERSIDADE VIRAL INTRA-HOSPEDEIRO

Vinte e três pacientes HIV+ recrutados no HI tiveram a sequência proviral completa ou parcial sequenciada, depositada no Genbank como resultado da Tese de Doutorado da aluna Brunna L. M. Alves, recentemente publicada (ALVES *et al.*, 2017). Essas sequências-consenso de cada amostra

representam os nucleotídeos presentes em cada posição nucleotídica com frequência maior ou igual a 50% na população viral.

a) Predição de epítomos de células T

As regiões proteicas virais Gag, Pol, Nef e Vif foram submetidas individualmente à ferramenta de predição de epítomos de célula T disponível no banco de dados IEDB (do inglês, *Immune Epitope Database Analysis Resource*), disponível online no site <http://tools.iedb.org/mhci/>. Esta ferramenta prevê valores de IC_{50} para peptídeos que se ligam a moléculas de MHC específicas. Dessa forma, essa ferramenta analisa uma sequência, ou um conjunto de sequências, de aminoácidos e determina a capacidade desta(s) se ligar(em) a uma molécula MHC classe I específica. O resultado da análise é dado em IC_{50} na escala nM. Como uma diretriz aproximada, os peptídeos com valores de $IC_{50} < 50$ nM são considerados de alta afinidade, aqueles entre 50 e 500 nM de afinidade intermediária e aqueles entre 500 e 5000 nM de baixa afinidade. Portanto, números menores indicam maior afinidade.

Primeiramente as sequências proteicas a serem analisadas são submetidas à ferramenta online no formato fasta. O algoritmo escolhido para identificação dos epítomos foi o recomendado pelo IEDB (*IEDB Recommended*) que é a seleção padrão do método de previsão. Essa seleção busca usar o melhor método possível para uma dada molécula de MHC com base nos preditores disponíveis e no desempenho preditivo observado em análises prévias.

Com a finalidade de selecionar epítomos que sejam mais bem reconhecidos por uma grande parcela da população analisada, as sequências de aminoácidos foram submetidas à análise de afinidade aos dois HLAs-A, B e C encontrados com maior frequência na casuística dos pacientes do HI (HLA-A*02:01, HLA-A*30:01, HLA-B*07:02, HLA-B*44:03, HLA-C*04:01, HLA-C*07:02). Além disso, foi selecionada a opção de filtrar apenas os epítomos que apresentaram resultado de $IC_{50} < 50$ nM, de forma a selecionar apenas os epítomos que tenham alta afinidade aos alelos.

A partir do resultado da predição de afinidade dos epítomos encontrados

foram removidas as duplicatas de forma a ter cada epítopo representado uma única vez. Em seguida, foram filtrados os epítopos presentes nas sequências virais de todos os pacientes ou na maioria deles. Dessa forma, selecionamos os melhores epítopos: que possuem alta afinidade ($IC_{50} < 50$ nM) pelos HLAs-A, B e C mais frequentes da nossa população estudada e frequentemente encontrados nos consensos virais desses pacientes.

b) Afinidade dos epítopos aos HLA-A, B e C específicos

O próximo passo foi analisar a afinidade desses epítopos previamente selecionados para os alelos de HLAs-A, B e C específicos de cada um dos pacientes, de forma a observar se os epítopos apresentariam também alta afinidade a eles, além da alta afinidade apresentada aos HLAs mais frequentes nessa população. Os epítopos selecionados foram submetidos à ferramenta online de predição juntamente com os HLAs específicos dos pacientes. Foram excluídos das análises os epítopos que tivessem $IC_{50} > 50$ nM para mais de um alelo do paciente.

c) Variação dos epítopos intrapaciente

Em seguida, foram analisadas as variações intrapaciente dos epítopos escolhidos, ou seja, nas quasiespécies virais do paciente. Para essa análise foram utilizados os *reads* obtidos dos sequenciamentos dos provírus arquivados. Esses arquivos foram analisados no programa Geneious versão 9.1.3 (Biomatters Limited, Auckland, Nova Zelândia). O epítopo selecionado foi localizado no alinhamento dos *reads* ao genoma viral e foram selecionados apenas os *reads* contidos nesse intervalo. Esses *reads* tiveram sua sequência nucleotídica convertida em aminoácido e foram removidas as duplicatas de forma que ficasse representada apenas uma variante de cada epítopo. Ao fazer isso, o programa indica o número de *reads* que essa variante possui, ou seja, sua frequência dentro do número total de *reads* da população viral sequenciada. Com isso foi estimada a frequência de cada variante. Foram selecionadas as variantes que apresentavam frequência maior ou igual a

0,01% para análises posteriores. Esses epítomos virais variantes com frequência maior que 0,01% foram submetidos à mesma análise de predição de afinidade aos HLAs-A, B e C específicos do paciente. Foram descartados os epítomos que apresentaram variantes com $IC_{50} > 50$ nM a mais de um alelo de HLA do paciente.

5. RESULTADOS

5.1. CASUÍSTICA E DADOS EPIDEMIOLÓGICOS

Foram coletadas amostras de 32 pacientes do HI e 12 pacientes do HUCFF-UFRJ, totalizando 44 amostras na cidade do Rio de Janeiro (RJ). Para a região sul do Brasil, foram coletados 40 pacientes atendidos no hospital universitário da FURG-RS na cidade de Rio Grande, no Rio Grande do Sul (RS).

Foram preenchidos questionários com dados clínicos e epidemiológicos para cada paciente. Esses dados foram compilados e podem ser observados na Tabela 5.1. Com os dados obtidos dos prontuários clínicos dos pacientes do HI ($n = 32$), foi possível observar um maior número de pacientes do sexo masculino (75%) e uma mediana de idade de 38 anos à data da coleta (variando de 23 a 63 anos). Os valores são referentes à data da aplicação dos questionários clínico e epidemiológico ao paciente (fevereiro a junho de 2016). Os pacientes apresentavam, até a data da coleta, uma mediana de tempo de diagnóstico de 4,7 anos. A mediana da contagem de células T CD4+ para o momento da coleta foi de $712,5 \text{ cel/mm}^3$ e para as células T CD8+ foi de $657,5 \text{ cel/mm}^3$. Para 31 pacientes, foi observado um intervalo médio de 1,2 anos do diagnóstico da infecção pelo HIV até o início do tratamento. Um paciente não possuía essa informação em seu questionário. Até o momento da coleta, os pacientes tinham uma mediana de 3,1 anos de tratamento antirretroviral. Trinta e um pacientes apresentavam os valores da contagem de carga viral antes do início do tratamento antirretroviral em seus prontuários, com uma mediana de 19.924 cópias/ml. Para 27 pacientes, foram encontrados os valores de contagem de células T CD4+ antes do início do tratamento antirretroviral, com

uma mediana de 373 cel/mm³.

Para os pacientes do HUCFF-UFRJ (n = 12), foi possível observar a partir dos dados retirados dos prontuários médicos que a mediana de idade na data da coleta era de 43,5 anos (variando de 26 a 60 anos) e a grande maioria era do sexo masculino (67%). Os valores são referentes à data da aplicação dos questionários clínico-epidemiológicos ao paciente (dezembro de 2016 a junho de 2017). Os pacientes apresentavam, até a data da coleta, uma mediana de tempo de diagnóstico de 13,3 anos. Foi observado também um intervalo de 1,7 anos do diagnóstico da infecção pelo HIV até o início do tratamento antirretroviral, este com uma mediana de tempo de duração de 9,3 anos. O valor da contagem de células T CD4+ no momento da coleta foi observado para apenas dois pacientes (1.201 e 1.033 cel/mm³). Para o restante dos pacientes, não foi realizado o exame nessa data ou o resultado não estava presente no prontuário. E para nenhum dos doze pacientes o valor da contagem de células T CD8+ na data da coleta estava disponível. Oito pacientes apresentavam os valores da contagem de carga viral antes do início do tratamento antirretroviral em seus prontuários, com uma mediana de 48.029,5 cópias/ml. Para 10 pacientes, foram disponibilizados os valores de contagem de células T CD4+ antes do início do tratamento antirretroviral, com uma mediana de 213,5 cel/mm³.

Ao analisar os dados dos prontuários clínicos dos pacientes do HU da FURG-RS (n = 40), observamos um maior número de pacientes do sexo feminino (62%) e uma mediana de idade de 43 anos na data da coleta (variando de 23 a 62 anos). Os valores são referentes à data da aplicação dos questionários clínico-epidemiológicos aos pacientes (agosto a dezembro de 2017). Os pacientes apresentavam, até a data da coleta, uma mediana de tempo de diagnóstico de 4,8 anos. A mediana da contagem de células T CD4+ no momento da coleta foi de 780,5 cel/mm³, e para as células T CD8+ foi de 848 cel/mm³. Foi observado um intervalo de menos de um ano (0,88) do diagnóstico da infecção pelo HIV até o início do tratamento. Até o momento da coleta, os pacientes estavam com uma mediana de 3,83 anos de tratamento antirretroviral. Doze pacientes apresentavam os valores da contagem de carga viral antes do início do tratamento antirretroviral em seus prontuários, com uma mediana de 13.509 cópias/ml. Para 21 pacientes, foram disponibilizados os

valores de contagem de células T CD4+ antes do início do tratamento antirretroviral, com uma mediana de 545 cel/mm³.

Tabela 5.1 Características clínicas e epidemiológicas retiradas dos prontuários dos pacientes HIV+ incluídos no estudo referentes ao Hospital Federal de Ipanema (HI), Hospital Universitário Clementino Fraga Filho (HUCFF-UFRJ) e Hospital Universitário Dr. Miguel Riet Corrêa Jr. (HU FURG-RS).

Característica	Hospital Federal de Ipanema (n=32)	HUCFF- UFRJ (n=12)	HU FURG-RS (n=40)
Homens (%)	24 (75%)	8 (67%)	15 (38%)
Idade (anos) (mediana; IQR ₅₀)	38 (31,5-45,25)	43,5 (41-52)	43 (35-49,5)
Mediana da contagem de cels T-CD4 (cels/mm ³ ; IQR ₅₀)	712,5 (606,5-856)	1117 ¹	780,5 (680,75-938,5)
Mediana da contagem de cels T-CD8 (cels/mm ³ ; IQR ₅₀)	657,5 (529-1047,25)	ND	848 (750,5-1053)
Mediana de tempo de diagnóstico HIV+ (anos; IQR ₅₀)	4,7 (3,9-6,5)	13,3 (7,7-14,8)	4,83 (2,42-10,9)
Mediana de tempo desde o diagnóstico HIV+ até o início do tratamento (anos; IQR ₅₀)	1,2 (0,6-2,8) ²	1,7 (0,1-5,3)	0,88 (0,25-3,2)
Mediana de tempo de tratamento (anos; IQR ₅₀)	3,1 (2,4-3,9)	9,34 (4,6-11,4)	3,83 (2,2-4,77)

¹ Informação disponível para apenas dois pacientes (1.201 e 1.033 cels/mm³); ² Dados baseados em 31 pacientes; ND, informação não disponível; IQR₅₀, variação interquartil

Em relação ao tratamento utilizado pelo pacientes, 19 pacientes do HI (19/32; 60%), quatro pacientes do HUCFF-UFRJ (4/12; 33%) e 23 pacientes do HU FURG-RS (23/40; 58%) utilizavam o esquema de antirretrovirais recomendado na época pelo Ministério da Saúde que compreende os medicamentos tenofovir (TDF), lamivudina (3TC) e efavirenz (EFV). O restante dos pacientes utilizava esquemas de medicamentos diferentes daquele, compostos por inibidores de protease e outros inibidores de transcriptase reversa. Esses dados podem ser visualizados de forma mais completa na Tabela 5.2 abaixo.

Tabela 5.2. Esquema terapêutico dos antirretrovirais utilizados pelos pacientes HIV+ incluídos no estudo do Hospital Federal de Ipanema (HI; n= 32), do Hospital Universitário Clementino Fraga Filho (HUCFF-UFRJ; n= 12) e do Hospital Universitário Dr. Miguel Riet Corrêa Jr. (HU FURG-RS; n= 40).

Esquema de tratamento antirretroviral	HI	HUCFF-UFRJ	FURG-RS
3TC+TDF+EFV	19	4	23
AZT+3TC+TDF+RTV	0	0	6
AZT+3TC+EFV	3	4	3
AZT+3TC+NVP	1	1	0
AZT+3TC+RTV	0	1	2
AZT+3TC+ATV+RTV	3	0	0
AZT+3TC+ATV	1	0	0
AZT+3TC+LPV/r+RTV	2	0	1
AZT+3TC+RTV+DRV	0	0	1
AZT+EFV+EFV	0	1	0
ATV+3TC+TDF+RTV	1	1	0
ATV+3TC+TDF	0	0	1
EFV+TDF+FTC	1	0	0
3TC+LPV/ r+RTV+TDF	1	0	1
3TC+DRV+RTV+TDF	0	0	2

TDF, tenofovir; 3TC, lamivudina; EFV, efavirenz; AZT, zidovudina; ATV, atazanavir; LPV, lopinavir; RTV, ritonavir; DRV, darunavir; NVP, nevirapina; FTC, emtricitabina.

A grande maioria dos pacientes do HI declarou que a provável via de transmissão foi por a sexual (27/32, 84%), sendo que sete destes especificaram a via sexual HSH (homens que fazem sexo com homens). A grande maioria dos pacientes do HUCFF-UFRJ declarou transmissão por via heterossexual (7/8, 90%) e quatro pacientes não especificaram a forma provável de contágio. Por fim, a grande maioria dos pacientes do HU da FURG-RS declarou a provável via de transmissão como a sexual (34, 85%), sendo que desses, dois especificaram a via sexual HSH. Dois pacientes (5%) declararam serem usuários de drogas injetáveis (UDI).

5.2. AMPLIFICAÇÃO E SEQUENCIAMENTO DOS GENES DE HLA-A, B E C

Todas as 44 amostras do Rio de Janeiro e as 40 do Rio Grande tiveram os genes de HLA-A, B e C amplificados com sucesso. As bibliotecas de todas as amostras foram preparadas com uma concentração média de 11,43 nM e tamanho médio de 600 pb. Todas as 84 amostras amplificadas foram sequenciadas na plataforma MiSeq (Illumina, San Diego, EUA), totalizando 504 alelos sequenciados. Essas amostras não foram sequenciadas todas em uma mesma reação de sequenciamento, mas divididas em sequenciamentos diferentes conforme a rotina do laboratório e o andamento dos experimentos.

Na Tabela 5.3 pode ser observado o número médio de *reads*, o tamanho médio dos *reads* e a qualidade média para cada amostra. A média de pares de *reads* obtidos para cada amostra foi de 380.299,6, o tamanho médio dos *reads* foi de 250 pb e a qualidade média obtida foi de 36 na escala Phred.

Tabela 5.3 Análise dos reads gerados do sequenciamento dos genes de HLA-A, B e C das amostras estudadas na plataforma MiSeq.

Amostra	N. <i>reads</i> (R1)	Qualidade média dos <i>reads</i> (Q) **	Tamanho médio dos <i>reads</i> (pb)
1HI	352.361	36	250 pb (35-251 pb)
2HI	108.020	36	250 pb (35-251 pb)
3HI	114.261	36	250 pb (35-251 pb)
4HI	93.428	38	250 pb (35-251 pb)
5HI	329.559	38	250 pb (35-251 pb)
6HI	300.109	38	250 pb (35-251 pb)
7HI	498.710	38	250 pb (35-251 pb)
8HI	142.366	36	250 pb (35-251 pb)
9HI	141.650	36	250 pb (35-251 pb)
10HI	94.749	36	250 pb (35-251 pb)
11HI	107.718	36	250 pb (35-251 pb)
12HI	118.421	36	250 pb (35-251 pb)
13HI	71.376	36	250 pb (35-251 pb)
14HI	223.117	38	250 pb (35-251 pb)
15HI	392.669	38	250 pb (35-251 pb)
16HI	396.440	38	250 pb (35-251 pb)
17HI	307.627	36	250 pb (35-251 pb)
18HI	139.445	38	250 pb (35-251 pb)
19HI	267.916	38	250 pb (35-251 pb)
20HI	345.867	38	250 pb (35-251 pb)
21HI	384.044	38	250 pb (35-251 pb)

22HI	414.466	38	250 pb (35-251 pb)
23HI	370.916	38	250 pb (35-251 pb)
24HI	335.527	38	250 pb (35-251 pb)
25HI	506.040	38	250 pb (35-251 pb)
26HI	398.804	38	250 pb (35-251 pb)
27HI	409.530	38	250 pb (35-251 pb)
28HI	423.763	38	250 pb (35-251 pb)
29HI	138.696	38	250 pb (35-251 pb)
30HI	154.290	38	250 pb (35-251 pb)
31HI	115.380	38	250 pb (35-251 pb)
32HI	112.770	38	250 pb (35-251 pb)
2HUCFF	144.531	28	250 pb (35-251 pb)
3HUCFF	178.839	28	250 pb (35-251 pb)
4HUCFF	180.947	28	250 pb (35-251 pb)
5HUCFF	107.576	28	250 pb (35-251 pb)
6HUCFF	282.446	28	250 pb (35-251 pb)
7HUCFF	106.530	28	250 pb (35-251 pb)
9HUCFF	136.280	28	250 pb (35-251 pb)
10HUCFF	159.030	36	250 pb (35-251 pb)
12HUCFF	129.982	36	250 pb (35-251 pb)
14HUCFF	146.923	36	250 pb (35-251 pb)
16HUCFF	164.240	36	250 pb (35-251 pb)
17HUCFF	76.861	36	250 pb (35-251 pb)
1FURG	119.048	36	250 pb (35-251 pb)
2FURG	112.876	36	250 pb (35-251 pb)
3FURG	106.701	36	250 pb (35-251 pb)
4FURG	176.807	36	250 pb (35-251 pb)
5FURG	146.116	36	250 pb (35-251 pb)
6FURG	93.193	36	250 pb (35-251 pb)
7FURG	112.161	36	250 pb (35-251 pb)
8FURG	111.371	36	250 pb (35-251 pb)
9FURG	88.935	36	250 pb (35-251 pb)
10FURG	123.672	36	250 pb (35-251 pb)
11FURG	122.695	36	250 pb (35-251 pb)
12FURG	238.459	36	250 pb (35-251 pb)
13FURG	82.438	37	250 pb (35-251 pb)
14FURG	44.719	37	250 pb (35-251 pb)
15FURG	89.268	37	250 pb (35-251 pb)
16FURG	115.616	37	250 pb (35-251 pb)
17FURG	121.573	37	250 pb (35-251 pb)
18FURG	119.807	37	250 pb (35-251 pb)
19FURG	102.281	37	250 pb (35-251 pb)
20FURG	152.620	37	250 pb (35-251 pb)
21FURG	85.193	37	250 pb (35-251 pb)
22FURG	142.616	37	250 pb (35-251 pb)
23FURG	161.553	37	250 pb (35-251 pb)
24FURG	154.453	37	250 pb (35-251 pb)
25FURG	89.083	37	250 pb (35-251 pb)
26FURG	142.611	37	250 pb (35-251 pb)

27FURG	139.912	37	250 pb (35-251 pb)
28FURG	210.539	37	250 pb (35-251 pb)
29FURG	171.281	37	250 pb (35-251 pb)
30FURG	132.044	37	250 pb (35-251 pb)
31FURG	87.784	37	250 pb (35-251 pb)
32FURG	232.229	37	250 pb (35-251 pb)
33FURG	140.643	37	250 pb (35-251 pb)
34FURG	229.253	37	250 pb (35-251 pb)
35FURG	273.246	37	250 pb (35-251 pb)
36FURG	136.034	37	250 pb (35-251 pb)
37FURG	271.139	37	250 pb (35-251 pb)
38FURG	302.102	38	250 pb (35-251 pb)
39FURG	194.605	37	250 pb (35-251 pb)
40FURG	112.021	37	250 pb (35-251 pb)

* Sequenciamento realizado em datas diferentes (Fev, Maio, Jun e Jul).

** Qualidade dos *reads* na escala Phred.

5.3. TIPAGEM DO HLA-A, B E C E FREQUÊNCIA DOS ALELOS

Os *reads* obtidos do sequenciamento das amostras foram submetidos ao software *HLA Twin*. Todos os 32 pacientes do HI tiveram pelo menos um alelo de HLA tipado com o auxílio do programa. Ao todo foram, tipados 185 alelos de HLA: 62 alelos de HLA-A (31 pacientes), 59 alelos de HLA-B (31 pacientes) e 64 alelos de HLA-C (32 pacientes). Os 12 pacientes do HUCFF-UFRJ tiveram pelo menos um alelo de HLA tipado, totalizando 68 alelos tipados. Foram tipados 24 alelos para o HLA-A (12 pacientes) e 22 alelos para o HLA-B e também para o HLA-C (11 pacientes cada). Já para os pacientes do HU FURG-RS, foram tipados 229 alelos de HLA: 78 alelos de HLA-A (40 pacientes), 74 alelos de HLA-B (38 pacientes) e 77 alelos de HLA-C (40 pacientes). Os resultados das tipagens podem ser observados nas Tabelas 5.4, 5.5 e 5.6.

Tabela 5.4 Tipagem dos alelos de HLA-A, B e C dos pacientes do Hospital Federal de Ipanema (HI) com o software de tipagem HLA Twin.

Amostra HI	HLA-A		HLA-B		HLA-C	
1	A*11:01:01	A*26:01:01	B*38:01:01	B*44:03:01	C*12:03:01	C*14:03
2	A*02:01:01	A*29:02:01	x	x	C*01:02:01	C*05:01:01
3	A*02:01:01	A*68:01:01	B*35:08:01	B*51:01:01	C*04:01:01	C*14:02:01
4	A*02:02:01:01	A*33:03:01	B*81:01	x	C*04:01:01	C*18:01
5	A*24:02:01:01	A*30:01:01	x	B*50:01:01:01	C*04:01:01:06	C*04:01:01:06
6	A*25:01:01:01	A*30:02:01:01	B*18:01:01:02	B*58:01:01:03	C*03:02:02:01	C*12:03:01:01
7	A*03:01:01:01	A*30:01:01	x	B*45:01:01	C*04:01:01:02	C*16:01:01:01
8	A*02:01:01	A*24:02:01	B*07:02:01	B*52:01:02	C*04:01:01	C*07:02:21
9	x	x	B*35:01:01	B*39:01:01L	C*04:01:01	C*07:02:01
10	A*24:02:01	A*68:02:01	B*27:05:02	B*39:10:01	C*02:02:02	C*12:03:01
11	A*02:01:01	A*24:03:01	B*41:02:01	B*44:03:02	C*07:06	C*17:03
12	A*01:01:01	A*68:02:01	B*39:13:01	B*44:03:01	C*03:04:02	C*06:02:01
13	A*01:01:01	A*68:01:01	B*81:01	B*81:01	C*07:01:01	C*08:04:01
14	A*01:01:01:01	A*02:11:01	B*07:06:01	B*48:02:01	C*04:01:01:02	C*15:05:02
15	A*02:01:01:01	A*34:02:01	B*39:13:01	B*44:03:01	C*04:01:01:02	C*07:02:01:01
16	A*02:01:01:01	A*68:02:01:01	B*07:02:01:01	B*14:02:01:01	C*07:02:01:16	C*08:02:01:01
17	A*11:01:01	A*11:01:01	B*40:02:01	B*44:02:01:01	C*02:02:02	C*05:01:01
18	A*29:02:01:01	A*30:01:01	B*15:03:01:02	B*51:01:01:01	C*02:10:01:01	C*14:02:01:04
19	A*24:03:01	A*30:02:01	B*15:20	B*18:01:01:02	C*04:01:01:02	C*07:04:01:01
20	A*02:01:01:01	A*24:02:01:01	B*44:03:01:02	B*49:01:01:01	C*02:02:02:01	C*07:01:01:01
21	A*02:01:01:01	A*26:01:01:01	B*27:05:02:05	B*57:03:01:02	C*01:02:01:01	C*07:01:02
22	A*02:01:01:01	A*02:01:01:01	B*07:02:01:01	B*08:01:01:01	C*07:01:01:01	C*07:02:01:01
23	A*30:02:01	A*31:01:02:04	B*15:03:01	B*15:17:01	C*02:10:01	C*05:01:01
24	A*02:01:01:01	A*03:01:01:01	B*35:03:01:01	B*44:02:01:01	C*02:02:02:01	C*04:01:01:02
25	A*02:01:01:01	A*30:02:01:01	B*15:03:01:02	B*15:04:04	C*02:10:01:01	C*03:03:01:01
26	A*30:01:01	A*32:01:01:01	B*15:03:01:02	B*42:01:01	C*02:10:01:01	C*17:01:01:02
27	A*02:01:01:01	A*33:03:01:01	B*07:02:01:01	B*58:01:01:03	C*03:02:02:01	C*07:02:01:03
28	A*01:01:01:01	A*24:05:01	B*15:17:01:01	B*39:01:01:05	C*07:01:02	C*12:03:01:02
29	A*34:02:01	A*03:01:01:01	B*08:01:01:01	B*52:01:01:02	C*07:01:01:01	C*12:02:02:01
30	A*30:01:01	A*30:02:01:02	B*39:10:01	B*42:01:01	C*12:03:01:01	C*17:01:01:02
31	A*02:01:01:01	A*26:01:01:01	B*44:02:01:01	B*57:02:01	C*05:01:01:02	C*07:18
32	A*25:01:01:01	A*29:02:01:01	B*18:01:01:01	B*56:01:01:03	C*01:02:01:-- (alelo novo)	C*12:03:01:01

x – Alelo não tipado pelo software

(alelo novo) – Alelo sem identidade com nenhum outro já descrito no banco de dados.

Tabela 5.5 Tipagem dos alelos de HLA-A, B e C dos pacientes do Hospital Universitário Clementino Fraga Filho(HUCFF-UFRJ) com o software de tipagem HLA Twin.

Amostra HUCFF- UFRJ	HLA-A		HLA-B		HLA-C	
2	A*23:17	A*31:01:02:-- (alelo novo)	x	x	x	x
3	A*02:01:01:01	A*11:01:01:01	B*40:04	B*40:04	C*07:01:01:01	C*03:04:01:02
4	A*02:01:01:01	A*02:01:01:01	B*07:02:01:01	B*07:02:01:01	C*07:02:01:03	C*07:02:01:01
5	A*03:01:01:01	A*24:57	B*35:01:01:03	B*51:01:01	C*04:01:01	C*14:02:01
6	A*31:01:02:01	A*30:02:01:01	B*27:05:02	B*18:01:01:02	C*02:02:02:01	C*12:03:01:01
7	A*29:02:01:01	A*30:01:01	B*27:02:01:01	B*39:10:01:-- (alelo novo)	C*02:02:02:01	C*12:03:01:01
9	A*03:01:01:01	A*03:01:01:-- (alelo novo)	B*07:02:01:01	B*07:02:01:01	C*07:02:01:03	C*07:02:01:03
10	A*23:01:01:01	A*36:01	B*07:02:01:03	B*45:01:01	C*07:02:01:01	C*16:01:01
12	A*74:01:01	A*80:01:01	B*15:03:01:02	B*42:01:01	C*02:10:01:01	C*17:01:01:02
14	A*02:11:01	A*23:17	B*42:11	B*48:02:01	C*04:01:01	C*17:01:01
16	A*02:01:01:01	A*30:01:01	B*42:01:01	B*44:02:01	C*05:01:01:02	C*17:01:01
17	A*02:01:01:01	A*11:01:01:01	B*07:02:01	B*35:01:01:03	C*04:01:01:03	C*07:04:07

x – Alelo não tipado com o software.

(alelo novo) – Alelo sem identidade com nenhum outro já descrito no banco de dados.

Tabela 5.6 Tipagem dos alelos de HLA-A, B e C dos pacientes do Hospital Universitário Dr. Miguel Riet Corrêa Jr. (HUFURG) com o software de tipagem HLA Twin.

Amostra HUFURG	HLA-A		HLA-B		HLA-C	
1	A*24:02:01:01	A*24:02:40	B*14:01:01:01	B*41:01:01:01	C*08:02:01:02	C*17:01:01:-- (alelo novo)
2	A*02:01:01:22	A*23:01:01:-- (alelo novo)	B*44:100	B*49:01:01:01	x	C*08:127N
3	A*24:02:01:05	A*24:02:40	B*15:01:01:01	B*35:03:01:01	C*03:03:01:01	C*12:03:01:02
4	x	A*33:03:01:01	B*07:02:06	B*15:01:01:01	C*03:03:01:01	C*07:02:01:03
5	A*02:01:01:22	A*24:02:01:01	B*35:08:01:02	B*51:01:01:01	C*04:01:01:06	C*14:02:01:01
6	A*01:01:01:01	A*01:01:01:01	B*15:01:01:01	B*39:03:01:01	C*03:03:01:01	C*07:02:01:01
7	A*23:01:01:01	A*02:01:10	B*50:01:01:01	B*44:03:11	C*04:01:01:02	C*06:02:01:02
8	A*23:01:01:01	x	B*44:03:01:08	B*57:03:01:-- (alelo novo)	C*14:03	C*18:02
9	A*01:01:01:01	A*32:65	B*07:02:01:01	B*51:01:01:02	C*12:02:02:01	C*07:02:01:03
10	A*01:01:01:01	A*02:02:01:02	B*14:02:01:01	B*51:01:01:04	C*08:02:01:01	C*15:13:01:01
11	A*01:01:01:01	A*23:01:01:01	B*44:03:01:08	B*52:01:01	C*04:01:01:02	C*12:02:02:01
12	A*02:01:01:01	A*24:03:01:01	B*35:05:01:02	B*39:09:01:01	C*07:02:01:01	C*04:01:01:11
13	A*01:01:01:01	A*01:01:01:01	B*08:01:01:01	B*51:01:01:05	C*02:02:02:01	C*07:01:01:01
14	A*30:02:01:02	A*33:03:01:-- (alelo novo)	B*45:01:01	B*53:47	C*04:01:01:02	C*16:01:01:02
15	A*11:01:01:01	A*29:02:01:03	B*14:02:01:01	B*15:03:01:02	C*02:10:01:01	C*08:02:01:01
16	A*03:01:01:01	A*29:02:01:01	B*08:01:01:01	B*44:03:01:02	C*02:02:02:01	C*07:01:01:12
17	A*80:01:01:02	A*01:01:01:01	B*08:01:01:01	B*58:01:01:01	C*07:18	C*07:01:01:12
18	A*34:02:01	A*68:01:01:02	B*44:03:02	B*08:01:01:01	C*07:06:01:01	C*07:01:01:12
19	A*32:01:01:01	A*24:02:01:01	B*15:199	x	C*07:01:01:01	C*03:03:01:01
20	A*68:01:01:02	A*01:01:01:01	B*07:02:01:03	B*08:01:01:01	C*07:02:01:01	C*07:01:01:01
21	A*02:01:01:01	A*29:02:01:01	B*15:04:01:01	x	C*03:03:01:06	C*04:01:01:02
22	A*01:01:01:01	A*01:01:01:01	B*14:02:01:01	B*57:01:01	C*06:02:01:01	C*15:125
23	A*01:01:01:01	A*26:01:01:01	B*08:01:01:01	B*14:01:01:01	C*08:02:01:02	x
24	A*01:01:01:01	A*01:01:01:07	B*14:02:01:01	B*35:01:01:05	C*04:01:01:11	C*04:01:01:11
25	A*32:01:01:01	A*33:01:01:01	B*52:01:01:02	B*14:02:01:01	C*08:02:01:01	C*12:02:02:02
26	A*02:01:01:01	A*26:01:01:01	B*07:02:01:01	B*40:01:02:01	C*03:04:01:01	C*07:02:01:03
27	A*02:01:01:01	A*24:02:02	B*44:02:01:01	B*51:119	C*12:03:01:01	C*15:02:01:01
28	A*30:01:01	A*68:02:01:01	B*15:10:01	B*35:01:01:05	C*03:04:02	C*04:01:01:05
29	A*02:01:01:01	A*26:01:01:01	B*38:01:01	B*51:01:01:03	C*14:02:01:04	C*12:03:01:02
30	A*02:01:01:01	A*31:01:02:01	B*15:04:01:01	B*51:01:01	C*03:03:01:01	C*15:02:01:01
31	A*23:01:01:01	A*68:01:02:01	B*50:02	B*40:02:01:01	C*03:04:01:01	C*04:01:01:08
32	A*31:01:02:01	A*33:01:01:01	B*14:02:01:01	B*39:02:03	C*07:02:01:01	C*08:02:01:01
33	A*01:01:01:01	A*29:02:01:01	B*35:03:01:03	B*58:01:01:01	C*04:01:01:11	C*04:01:01:11
34	A*02:01:01:01	A*31:01:02:01	x	x	C*12:03:01:02	C*15:02:01:01
35	A*02:01:01:01	A*02:01:01:01	B*15:04:01:01	B*44:05:01	C*02:02:02:01	x
36	A*01:01:01:01	A*02:01:01:01	B*07:03	B*15:01:01	C*03:04:01:01	C*07:103
37	A*03:01:01:01	A*02:01:01:01	B*15:04:01:01	B*35:03:01:01	C*03:03:01:01	C*12:03:01:01
38	A*02:01:01:01	A*11:01:01:01	B*07:02:01:01	B*07:02:01:01	C*02:02:02:01	C*07:02:01:03
39	A*24:02:01:01	A*31:01:02:01	x	x	C*15:02:01:01	C*02:02:02:01
40	A*02:05:01:01	A*68:02:01:01	B*49:01:01	B*50:01:01:01	C*06:02:01:04	C*06:02:01:02

x – Alelo não tipado com software.

(alelo novo) – Alelo sem identidade com nenhum outro já descrito no banco de dados.

Para todas as amostras tipadas do Rio de Janeiro, a frequência de homozigotos foi de 7% (3/43 [31 HI + 12 HUCFF-UFRJ]) para o HLA-A, 10% (4/39 [285 HI + 11 HUCFF-UFRJ]) para o HLA-B e 5% (2/43 [325 HI + 11 HUCFF-UFRJ]) para o HLA-C. A frequência de homozigotos no Rio Grande para o HLA-A foi de 11% (4/38), 3% (1/36) para o HLA-B e 5% (2/37) para o HLA-C. Esses resultados podem ser visualizados nas Figuras 5.1 e 5.2 abaixo.

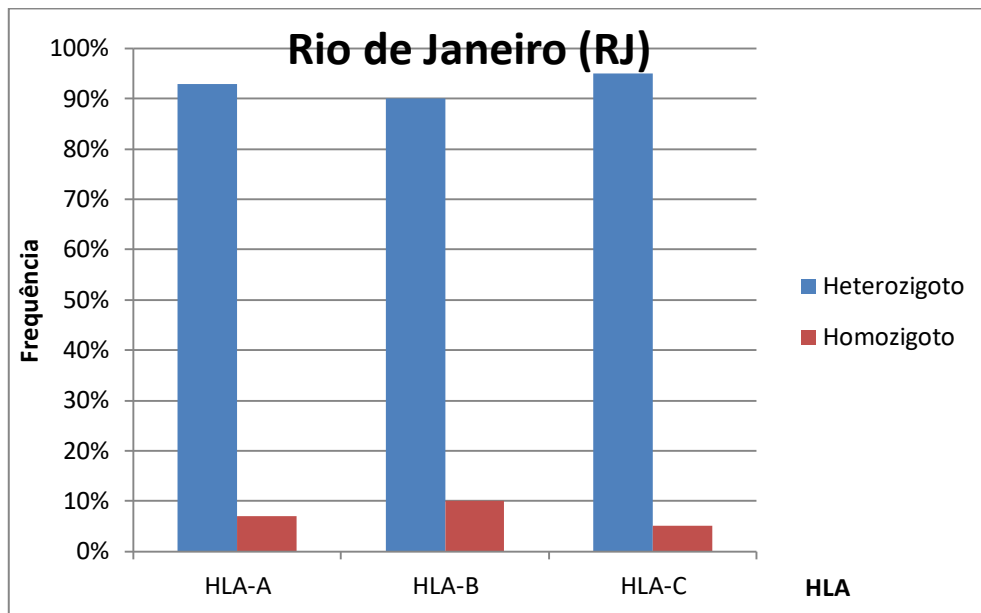


Figura 5.1 Proporção de heterozigotos e homozigotos encontrada para todos os pacientes do Rio de Janeiro (Hospital Federal de Ipanema – HI e Hospital Universitário Clementino Fraga Filho - HUCFF-UFRJ) para cada lócus de HLA analisado.

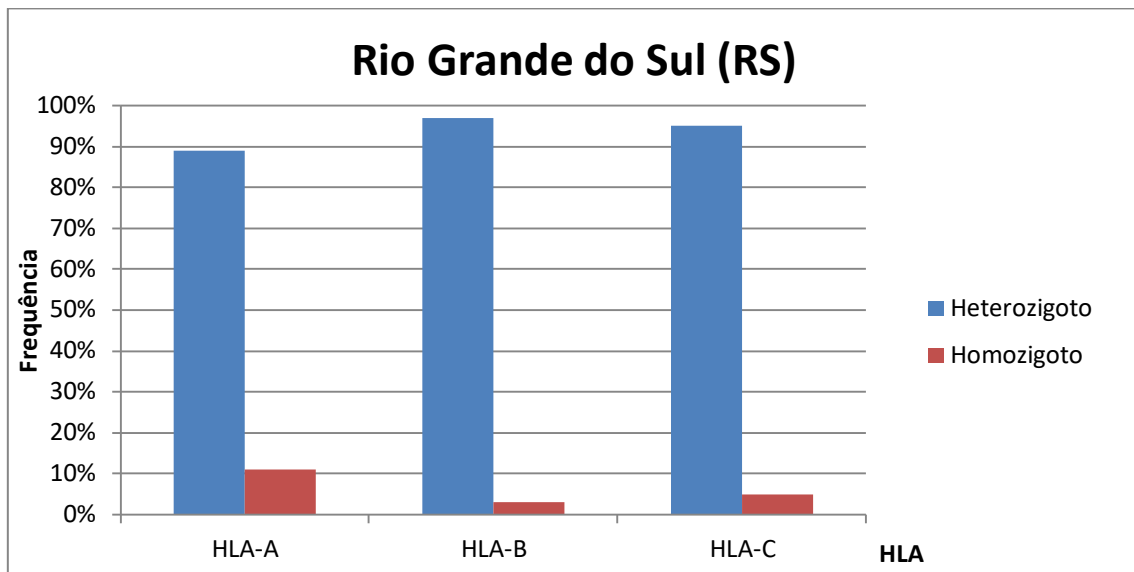


Figura 5.2 Proporção de heterozigotos e homozigotos encontrada para todos os pacientes do Rio Grande do Hospital Universitário Dr. Miguel Riet Corrêa Jr. (HUFURG-RS) para cada lócus de HLA analisado.

O alelo mais frequente para o HLA-A dentre os pacientes do Rio de Janeiro foi o A*02:01:01 (22%; 19/86 [62 HI + 24 HUCFF-UFRJ]), para o HLA-B foi o alelo B*07:02:01 (12%; 10/81 [59 HI + 22 HUCFF-UFRJ]) e para o HLA-C foi o alelo C*04:01:01 (16%; 14/86 [64 HI + 22 HUCFF-UFRJ]). Esses dados podem ser visualizados nas Figuras 5.3, 5.4 e 5.5.

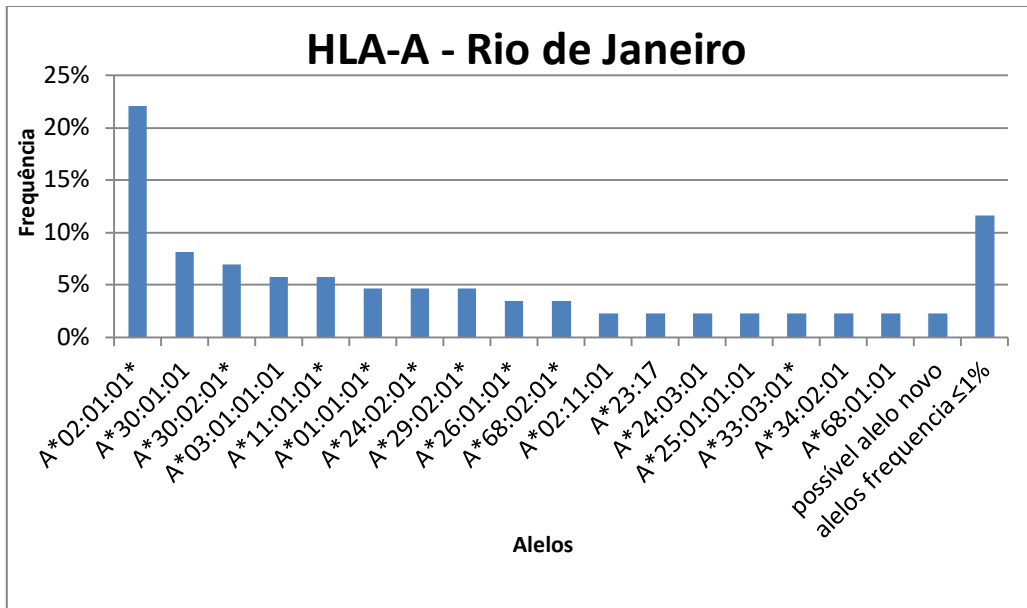


Figura 5.3 Alelos de HLA-A encontrados nos pacientes do Rio de Janeiro com frequência maior que 1% (n = 86). Marcados com * estão os alelos que sofreram redução de resolução de tipagem (seis dígitos) para melhor agrupar os alelos que não alcançaram a resolução de tipagem de 8 dígitos.

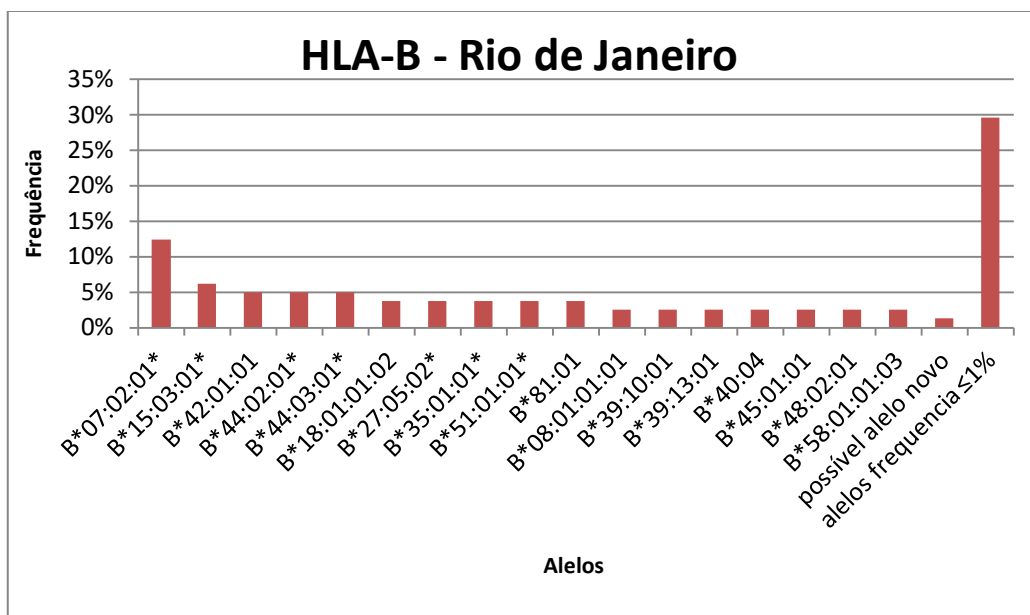


Figura 5.4 Alelos de HLA-B encontrados nos pacientes do Rio de Janeiro com frequência maior que 1% (n = 81). Marcados com * estão os alelos que sofreram redução de resolução de tipagem (seis dígitos) para melhor agrupar os alelos que não alcançaram a resolução de tipagem de 8 dígitos.

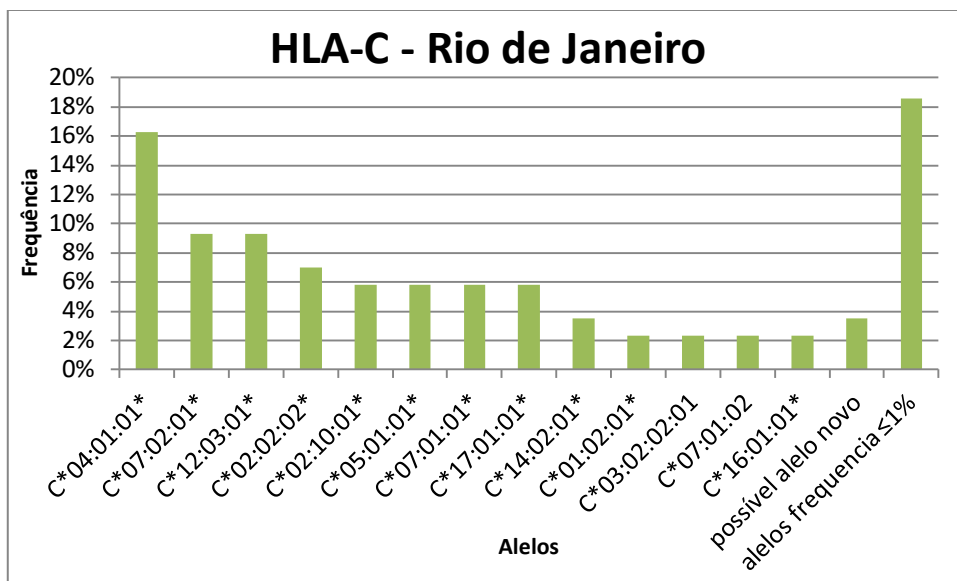


Figura 5.5 Alelos de HLA-C encontrados nos pacientes do Rio de Janeiro com frequência maior que 1% (n=86). Marcados com * estão os alelos que sofreram redução de resolução de tipagem (seis dígitos) para melhor agrupar os alelos que não alcançaram a resolução de tipagem de 8 dígitos.

Para os pacientes do Rio Grande, o alelo mais frequente para o HLA-A foi o A*01:01:01:01 (19%; 15/78), para o HLA-B foram os alelos B*08:01:01:01 e B*14:02:01:01 (8%; 6/74) e para o HLA-C foi o alelo C*03:03:01:01 (8%; 6/77). Esses dados podem ser visualizados nas Figuras 5.6, 5.7 e 5.8.

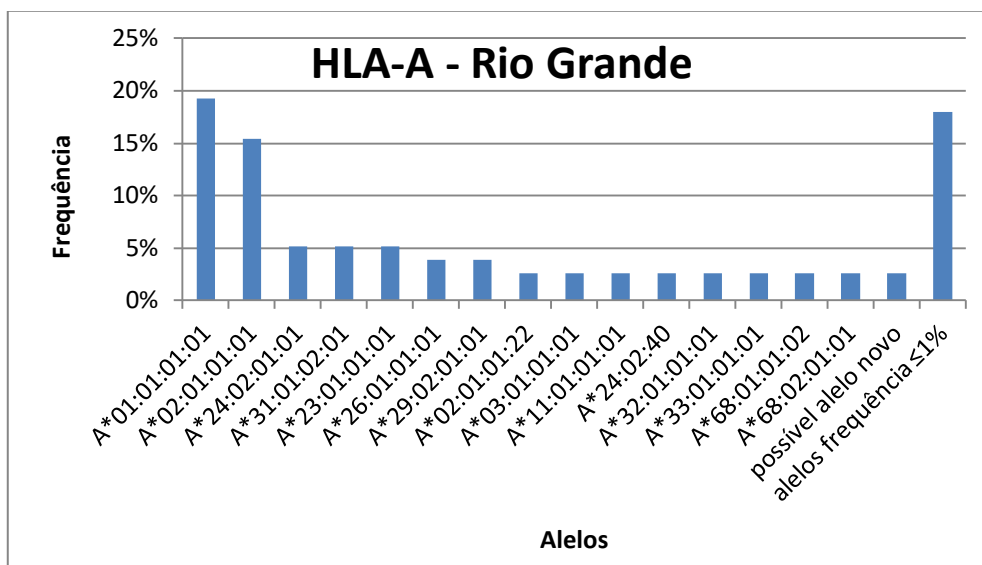


Figura 5.6 Alelos de HLA-A encontrados nos pacientes do Rio Grande com frequência maior que 1% (n = 78).

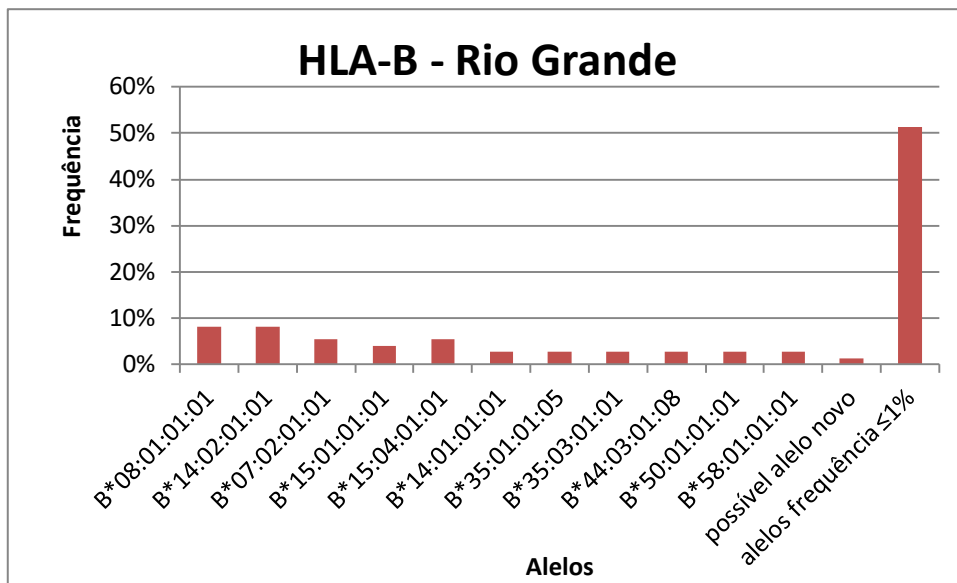


Figura 5.7 Alelos de HLA-B encontrados nos pacientes do Rio Grande com frequência maior que 1% (n = 74).

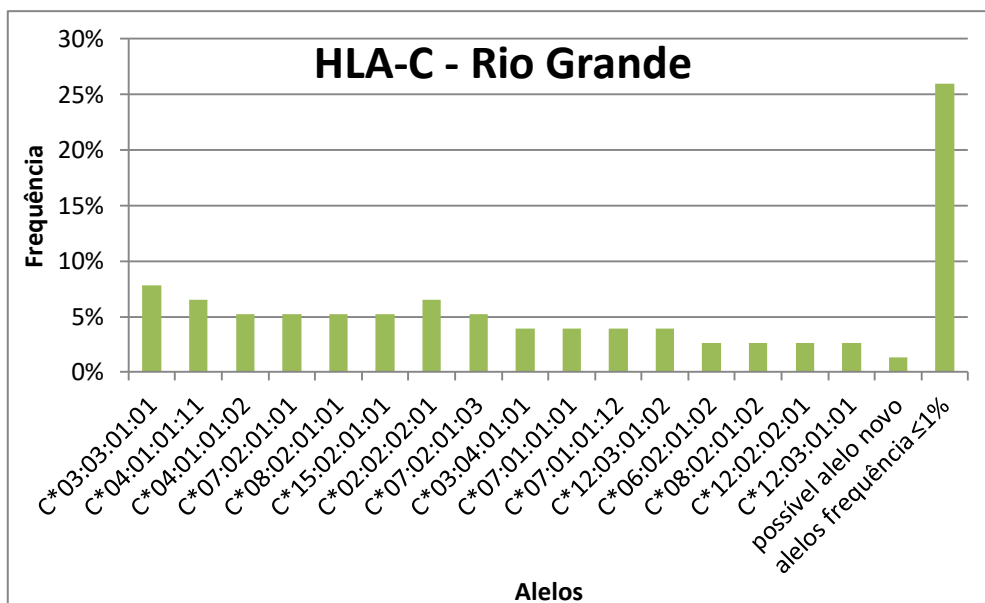


Figura 5.8 Alelos de HLA-C encontrados nos pacientes do Rio Grande com frequência maior que 1% (n = 77).

5.4. COMPARAÇÃO DAS FREQUÊNCIAS ALÉLICAS ENTRE AS LOCALIDADES DE ESTUDO E COM O BANCO DE DADOS DO REDOME

As análises estatísticas realizadas comparando as frequências dos alelos de HLA-A, B e C encontrados nas casuísticas do Rio de Janeiro e na do Rio Grande com as frequências encontrada nos dados do REDOME dos estados do Rio de Janeiro (RJ) e do Rio Grande do Sul (RS) mostrou algumas diferenças de frequência de alelos estatisticamente significativas.

O alelo A*01 apresentou uma maior frequência na população estudada do Rio Grande quando comparado com a do Rio de Janeiro ($p = 0,0014$). Essa mesma frequência também se mostrou maior do que a encontrada na população de doadores de medula óssea do REDOME do Rio Grande do Sul ($p = 0,00001$).

O alelo B*07 apresentou uma maior frequência na casuística de pacientes do HI do Rio de Janeiro quando comparada com a frequência desse alelo na população de doadores de medula óssea do REDOME-RJ ($p = 0,0001$). Já o alelo B*14 foi observado com maior frequência nos pacientes estudados do HU FURG-RS quando comparado com os pacientes do HI-RJ ($p = 0,014$) e também com a população de doadores de medula óssea do REDOME-RS ($p = 0,0068$).

O alelo C*03 foi encontrado com maior frequência também nos pacientes estudados do Rio Grande quando comparados com os do Rio de Janeiro ($p = 0,028$).

Todas as análises descritas acima podem ser observadas na Tabela 5.7 abaixo.

Tabela 5.7 Análises estatísticas de comparação de frequências alélicas entre as casuísticas estudadas e do REDOME dos estados do Rio de Janeiro e Rio Grande do Sul.

	Casuística	Presença	Ausência	p
Alelo A*01	HI - RJ	03 (4%)	71	0,0014
	HUFURG - RS	16 (20%)	58	
	HUFURG - RS	16 (20%)	58	0,00001
	REDOME - RS	26546 (11%)	214783	
Alelo B*07	HI - RJ	10 (14%)	62	0,0001
	REDOME - RJ	9752 (7%)	129570	
Alelo B*14	HI - RJ	1 (1%)	71	0,014093
	HUFURG - RS	8 (9%)	62	
	HUFURG - RS	8 (9%)	62	0,0068
	REDOME - RS	9653 (4%)	231676	
Alelo C*03	HI - RJ	03 (4%)	67	0,028042
	HUFURG - RS	11 (15%)	61	

5.5. DESCOBERTA E CONFIRMAÇÃO DE ALELOS NOVOS

Ao todo foram identificados oito possíveis alelos novos, que podem ser visualizados na Tabela 5.8.

Tabela 5.8 Descrição dos possíveis alelos novos encontrados durante a tipagem dos HLA-A, B e C dos pacientes do estudo.

Amostra	Alelo mais próximo	Troca nucleotídica	Região gênica	Posição*
32 HI	C*01:02:01:01	G>T	3'UTR	4085
7 HUCFF	B*39:10:01	T>C	3'UTR	3745
2 HUCFF	A*31:01:02:01	A>G	5'UTR	292
9HUCFF	A*03:01:01:01	G>C	5'UTR	75
1 FURG	C*17:01:01:05	G>A	INTRON 1	365
		A>G	INTRON 2	768
2 FURG	A*23:01:01:01	G>A	INTRON 5	2565
8 FURG	B*57:03:01:01	C>T	INTRON 7	3620
14 FURG	A*33:03:01:01	T>G	5'UTR	105

* Posição relativa da troca nucleotídica de acordo com as sequências-referência de HLA-A (NG_029217.2), HLA-B (NG_023187.1) e HLA-C (NG_029422.2) disponíveis na base de dados do NCBI.

Para uma das amostras do HI, a amostra 32, o software de tipagem *HLA Twin* indicou a possibilidade de que um dos alelos de HLA-C do paciente fosse um alelo novo. O programa identificou o primeiro alelo como C*12:03:01:01 e o segundo alelo como C*01:02:01:--. Foi encontrada uma troca G>T na posição 4085. Frente a isso, essa amostra foi submetida à análise de tipagem desenvolvida *in-house*. Primeiramente foi realizada a filtragem dos *reads* para qualidade e tamanho. Em seguida os *reads* filtrados foram alinhados com as sequências-referências A*25:01:01:01, A*29:02:01:01, B*18:01:01:02, B*56:01:01:03, C*12:03:01:01 e C*01:02:01:01 retiradas do banco de dados do IMGT-HLA. Ao término da análise realizada pelo pipeline já descrito, o alinhamento foi visualizado no programa IGV e as alterações foram inspecionadas individualmente onde pode-se observar a troca nucleotídica identificada pelo programa.

O HLA-A da amostra 09 do HUCFF-UFRJ apresentou uma tipagem ambígua, sem definição dos alelos que seriam os melhores candidatos, pelo programa *HLA Twin*. O mesmo procedimento de análise de possível alelo novo

relatado acima foi realizado para essa amostra. Na visualização do alinhamento no IGV foi possível observar dois nucleotídeos diferentes na posição 75 correspondente a região promotora (5'UTR). A presença das bases G e C indicam que possivelmente um dos alelos corresponderia ao A*03:01:01:01 contendo a base nucleotídica G nessa posição, enquanto que o outro, o possível alelo novo contendo a troca G>C, seria A*03:01:01: - -.

O programa HLA Twin indicou a presença de possíveis alelos novos ao tipar as amostras 02 e 07 do HUCFF-UFRJ. Com isso, essas amostras foram submetidas a análises para verificação da troca nucleotídica apontada pelo programa e confirmação da possibilidade de serem alelos novos. Os *reads* foram filtrados e alinhados com as sequências-referências A*23:17 e A*31:01:02:01 para a amostra 02 e A*29:02:01:01, A*30:01:01, B*27:02:01:01, B*39:10:01, C*02:02:02:01 e C*12:03:01:01 para a amostra 07. Os alinhamentos finais foram visualizados no programa IGV. Foi possível observar que o alinhamento da sequência-referência HLA-A*31:01:02:01 (alelo mais próximo) com a amostra 02 apresentou a troca A>G na região promotora 5'UTR (292pb), indicando ser um alelo novo. O alinhamento da sequência-referência HLA-B*39:10:01 com os *reads* da amostra 07 apresentou a troca T>C na região promotora 3'UTR (3745pb), como indicado pelo software usado para a tipagem.

Para os pacientes do sul do Brasil, quatro tiveram suspeitas de possíveis alelos novos devido a ambigüidades de tipagem pelo software HLA Twin. Esses passaram pelas análises de confirmação de tipagem pelo *pipeline in house*. O paciente 1FURG apresentou uma troca G>A na posição 365 pb (intron 1) e outra troca A>G na posição 768 (intron 2). Tanto o outro alelo C*28:02:01:02 quanto o alelo mais próximo ao alelo novo (C*17:01:01:05) possuem os nucleotídeos G e A nessas posições respectivas, atribuindo a troca a um provável alelo novo ainda não descrito.

O paciente 2FURG apresentou uma troca G>A na posição 2565 (intron 5) em relação ao alelo A*23:01:01:01. Todas as referências de HLA-A*23:01:01 possuem o nucleotídeo G nessa posição, indicando que essa troca se atribui a um possível alelo novo ainda não descrito. Já o paciente 8FURG apresentou uma troca C>T na posição 3620 no alinhamento da sequência-referência B*57:03:01:01 com os *reads* da amostra. O paciente 14FURG também

apresentou um possível alelo novo, tendo como alelo mais próximo o A*33:03:01:01, com a troca T>G na posição 105.

5.6. ANÁLISE DE PREDIÇÃO DE EPÍTOPOS DE CÉLULAS T E DE DIVERSIDADE VIRAL INTRA-HOSPEDEIRO

Vinte e três sequências provirais, completas ou parcialmente sequenciadas, de pacientes do HI foram separadas nas regiões gênicas virais correspondentes a Gag, Pol, Nef e Vif e submetidas ao algoritmo de predição de epítomos de célula T do banco de dados IEDB (<http://tools.iedb.org/mhci/>).

a) Epítomos da região genômica viral *gag*

Para a região da Gag foram analisadas sequências-consenso de 19 pacientes. O algoritmo encontrou 30.725 epítomos com $IC_{50} < 50$ nM. Após remover as duplicatas restaram 3.264 epítomos únicos. Dois epítomos estavam presentes na maioria das sequências virais analisadas. O epítopo KARVLAEAM (posição nucleotídica referente às posições 1410-1436 da sequência-referência HXB2) estava presente em 18 pacientes, e apenas a sequência-consenso da amostra 18HI não apresentou esse epítopo. Já o segundo epítopo, EMMTACQGV (posição nucleotídica 1368-1395 referente à sequência-referência HXB2), não foi encontrado apenas na sequência-consenso da amostra 3HI. Todas as sequências de Gag restantes apresentavam esses dois epítomos com IC_{50} de alta afinidade para os alelos mais frequentes entre os pacientes do HI.

Ao testar esses dois epítomos contra os HLA-A, B e C específicos dos 18 pacientes, foi observado que nove amostras tiveram um alelo sem alta afinidade por um dos dois epítomos. No entanto, os outros cinco alelos desses nove pacientes apresentavam alta afinidade aos dois epítomos. Nenhuma amostra apresentou resultado de média ou baixa afinidade a dois ou mais alelos ao mesmo tempo. Esses resultados podem ser observados na Tabela 5.9 abaixo.

Tabela 5.9 Representação dos 9 pacientes do HI que apresentaram ao menos um alelo de HLA-A, B e C sem alta afinidade aos epítomos virais de Gag KARVLAEAM e EMMTACQGV.

Paciente	HLA					
	A		B		C	
1	A*11:01:01	A*26:01:01	B*38:01:01	B*44:03:01	C*12:03:01	C*14:03
2	A*02:01:01	A*29:02:01	x	x	C*01:02:01	C*05:01:01
3	A*02:01:01	A*68:01:01	B*35:08:01	B*51:01:01	C*04:01:01	C*14:02:01
8	A*02:01:01	A*24:02:01	B*07:02:01	B*52:01:02	C*04:01:01	C*07:02:21
12	A*01:01:01	A*68:02:01	B*39:13:01	B*44:03:01	C*03:04:02	C*06:02:01
13	A*01:01:01	A*68:01:01	B*81:01	B*81:01	C*07:01:01	C*08:04:01
20	A*02:01:01	A*24:02:01	B*44:03:01	B*49:01:01	C*02:02:02	C*07:01:01
23	A*30:02:01	A*31:01:02:04	B*15:03:01	B*15:17:01	C*02:10:01	C*05:01:01
32	A*25:01:01	A*29:02:01	B*18:01:01	B*56:42	x	x

Em verde: IC50 < 50nM para os dois epítomos

Em vermelho claro: IC50 > 50 nM para um dos epítomos

Em vermelho escuro: IC50 > 50 nM para os dois epítomos.

X: alelo sem informação de tipagem

As amostras 18HI e 3HI, que não apresentavam esses epítomos, possuíam variantes (KARILAEAM e EMMSACQGV, respectivamente) que ao passarem pela mesma análise apresentaram alta afinidade (IC50 < 50 nM) aos alelos de HLA mais frequentes e aos específicos do paciente.

Ao analisarmos a conservação desses dois epítomos selecionados em Gag (KARVLAEAM e EMMTACQGV) nos *reads* virais, foram encontrados 4 variantes com frequência > 0,01 (1%) do epítopo EMMTACQGV e 5 do epítopo KARVLAEAM. Todas as variantes do epítopo EMMTACQGV apresentaram IC50 < 50 nM para os alelos de seus pacientes, e o mesmo pôde ser observado para a maioria das variantes do epítopo KARVLAEAM, com exceção de duas delas. A variante KARVLAKAM não demonstrou ter alta afinidade por um dos alelos do paciente 11HI (A*02:01) e a variante KARVLAEAV, por um dos alelos do paciente 13HI (A*68:02). No entanto, as variantes apresentaram alta afinidade aos outros HLAs desses pacientes. Essas variantes podem ser observadas na Tabela 5.10 abaixo.

Tabela 5.10 Variantes dos epítomos virais de Gag (KARVLAEAM e EMMTACQGV) encontradas nos reads virais dos pacientes do HI analisados e resultado de predição de afinidade aos alelos de HLA específicos desses mesmos pacientes.

Paciente	Variante	Frequência	HLA					
			A		B		C	
11	<u>K</u> IITACQGV	4%	A*02:01	A*24:03	B*41:02	B*44:03	C*07:06	C*17:13
	KARVLA <u>K</u> AM	5%	A*02:01	A*24:03	B*41:02	B*44:03	C*07:06	C*17:13
12	<u>E</u> ARVLAEAM	2%	A*01:01	A*68:02	B*39:13	B*44:03	C*03:04	C*06:02
13	KARVLAE <u>A</u> V	1%	A*68:01	A*01:01	B*81:01	B*81:01	C*07:01	C*08:04
19	<u>K</u> MMTACQGV	1%	A*24:03	A*30:02	B*15:20	B*18:01	C*04:01	C*07:04
23	EMMTACQ <u>K</u> V	22%	A*30:02	A*31:01	B*15:03	B*15:17	C*02:10	C*05:01
	K <u>A</u> KVLAEAM	22%	A*30:02	A*31:01	B*15:03	B*15:17	C*02:10	C*05:01
28	<u>K</u> <u>V</u> RVLAEAM	5%	A*01:01	A*24:05	B*15:17	B*39:01	C*07:01	C*12:03
32	EMMTAC <u>R</u> GV	1%	A*25:01	A*29:02	B*18:01	B*56:01	C*01:02	C*12:03

Em verde: IC50 < 50 nM

Em vermelho: IC50 > 50 nM

b) Epítomos da região genômica viral *pol*

Para a região da Pol viral, foram analisados 19 pacientes. O algoritmo encontrou 148.939 epítomos com IC50 < 50 nM. Após remover as duplicatas, restaram 7.288 epítomos únicos. Dez epítomos estavam presentes nas 19 sequências virais analisadas. Os 10 epítomos foram analisados quanto à afinidade aos alelos de HLA-A, B e C específicos dos 19 pacientes, totalizando 60 alelos testados. Três desses epítomos foram descartados, pois só apresentaram alta afinidade a menos de 50 alelos. Em seguida foi analisado se um mesmo epítopo apresentava IC50 > 50 nM a mais de um alelo de um mesmo paciente e com essa análise foram descartados seis epítomos. Apenas um epítopo (MIGGIGGFI) apresentou alta afinidade a mais de 50 alelos de HLA (57 alelos com alta afinidade) e a pelo menos 5 alelos de um mesmo paciente, podendo ser observado na Tabela 5.11.

Tabela 5.11 Representação da predição de afinidade do epítipo viral MIGGIGGFI, da região da Pol do HIV, aos alelos de HLA-A, B e C dos 19 pacientes do HI analisados.

Paciente	HLA					
	A		B		C	
1	A*11:01:01	A*26:01:01	B*38:01:01	B*44:03:01	C*12:03:01	C*14:03
2	A*02:01:01	A*29:02:01	x	x	C*01:02:01	C*05:01:01
3	A*02:01:01	A*68:01:01	B*35:08:01	B*51:01:01	C*04:01:01	C*14:02:01
8	A*02:01:01	A*24:02:01	B*07:02:01	B*52:01:02	C*04:01:01	C*07:02:21
11	A*02:01:01	A*24:03:01	B*41:02:01	B*44:03:02	C*07:06	C*17:03
12	A*01:01:01	A*68:02:01	B*39:13:01	B*44:03:01	C*03:04:02	C*06:02:01
13	A*01:01:01	A*68:01:01	B*81:01	B*81:01	C*07:01:01	C*08:04:01
14	A*01:01:01:01	A*02:11:01	B*07:06:01	B*48:02:01	C*04:01:01:02	C*15:05:02
18	A*29:02:01:01	A*30:01:01	B*15:03:01:02	B*51:01:01:01	C*02:10:01:01	C*14:02:01:04
19	A*24:03:01	A*30:02:01	B*15:20	B*18:01:01:02	C*04:01:01:02	C*07:04:01:01
20	A*02:01:01:01	A*24:02:01:01	B*44:03:01:02	B*49:01:01:01	C*02:02:02:01	C*07:01:01:01
21	A*02:01:01:01	A*26:01:01:01	B*27:05:02:05	B*57:03:01:02	C*01:02:01:01	C*07:01:02
22	A*02:01:01:01	A*02:01:01:01	B*07:02:01:01	B*08:01:01:01	C*07:01:01:01	C*07:02:01:01
23	A*30:02:01	A*31:01:02:04	B*15:03:01	B*15:17:01	C*02:10:01	C*05:01:01
27	A*02:01:01:01	A*33:03:01:01	B*07:02:01:01	B*58:01:01:03	C*03:02:02:01	C*07:02:01:03
28	A*01:01:01:01	A*24:05:01	B*15:17:01:01	B*39:01:01:05	C*07:01:02	C*12:03:01:02
29	A*34:02:01	A*03:01:01:01	B*08:01:01:01	B*52:01:01:02	C*07:01:01:01	C*12:02:02:01
31	A*02:01:01:01	A*26:01:01:01	B*44:02:01:01	B*57:02:01	C*05:01:01:02	C*07:18
32	A*25:01:01:01	A*29:02:01:01	B*18:01:01:01	B*56:01:01:03	C*01:02:01:-- (alelo novo)	C*12:03:01:01

Em verde: IC50 < 50nM

Em vermelho claro: IC50 > 50 nM

X: alelo sem informação de tipagem

O epítipo candidato MIGGIGGFI (posição nucleotídica 1934-1960 referente à sequência-referência HXB2) foi analisado quanto à presença de variantes nos *reads* virais das amostras. Foram encontradas três variantes com frequência > 0,01 (1%) nos *reads*. A variante (IIKEIEGFI) não demonstrou ter alta afinidade por um dos alelos do paciente 11HI (A*44:03) e a variante (MIRGIRGFI) não demonstrou ter alta afinidade por dois alelos do paciente 23HI (B*15:03 e C*05:01). No entanto, as variantes apresentaram alta afinidade aos outros HLAs desses pacientes. Já a variante MITGGIGGFI apresentou alta afinidade a todos os alelos do paciente. Essas variantes podem ser observadas na Tabela 5.12 abaixo.

Tabela 5.12 Variantes do epítipo viral MIGGIGGFI, da região da Pol do HIV, encontradas nos reads virais dos pacientes do HI analisados e resultado de predição de afinidade aos alelos de HLA-A, B e C específicos desses mesmos pacientes.

Paciente	Variante	Frequência	HLA					
			A		B		C	
11	I <u>I</u> K <u>E</u> I <u>E</u> GFI	2%	A*02:01	A*24:03	B*41:02	B*44:03	C*07:06	C*17:13
20	M <u>I</u> G <u>G</u> I <u>G</u> GFI	16%	A*02:01	A*24:02	B*44:03	B*49:01	C*02:02	C*07:01
23	M <u>I</u> R <u>G</u> I <u>R</u> GFI	22%	A*30:02	A*31:01	B*15:03	B*15:17	C*02:10	C*05:01

Em verde: IC50 < 50 nM

Em vermelho claro: IC50 > 50 nM

c) Epítipos da região genômica viral *nef*

Para a região de Nef foram analisadas 19 amostras. Foram encontrados 12.734 epítipos com IC50 < 50 nM. Após remover as duplicatas, restaram 1.973 epítipos únicos. Apenas um dos epítipos (LTFGWCFKL) estava presente em apenas 15 amostras, sendo então descartado como candidato.

d) Epítipos da região genômica viral *vif*

Para a região de Vif, foram analisados 22 pacientes. Foram encontrados 14.164 epítipos com IC50 < 50 nM. Após remover as duplicatas, restaram 1.992 epítipos únicos. Três epítipos estavam presentes na maioria das sequências virais analisadas. O epítipo VGSLQYLAL estava presente em 21 amostras, e apenas a sequência consenso da amostra 1HI não apresentou esse epítipo. Já os epítipos MIVWQVDRM e QVMIVWQVD só não foram encontrados na sequência-consenso da amostra 31HI. Todas as sequências de Vif restantes apresentavam esses epítipos com IC50 de alta afinidade para os alelos mais frequentes entre os pacientes do HI.

Ao testar esses três epítipos contra os HLA-A, B e C específicos dos 22 pacientes, totalizando 62 alelos testados, foi observado que os dois epítipos

MIVWQVDRM e QVMIVWQVD apresentaram média ou baixa afinidade a mais de um alelo de um mesmo paciente. O resultado pode ser observado na Tabela 5.13. Esses epítomos foram descartados e as análises seguiram apenas com o epítopo VGSLQYLAL.

Tabela 5.13 Representação da predição de afinidade do epítopo viral VGSLQYLAL, da região Vif do HIV, aos alelos de HLA-A, B e C dos 22 pacientes do HI analisados.

Paciente	HLA					
	A		B		C	
1	A*11:01:01	A*26:01:01	B*38:01:01	B*44:03:01	C*12:03:01	C*14:03
2	A*02:01:01	A*29:02:01	x	x	C*01:02:01	C*05:01:01
3	A*02:01:01	A*68:01:01	B*35:08:01	B*51:01:01	C*04:01:01	C*14:02:01
5	A*24:02:01:01	A*30:01:01	x	B*50:01:01:01	C*04:01:01:06	C*04:01:01:06
8	A*02:01:01	A*24:02:01	B*07:02:01	B*52:01:02	C*04:01:01	C*07:02:21
11	A*02:01:01	A*24:03:01	B*41:02:01	B*44:03:02	C*07:06	C*17:03
12	A*01:01:01	A*68:02:01	B*39:13:01	B*44:03:01	C*03:04:02	C*06:02:01
13	A*01:01:01	A*68:01:01	B*81:01	B*81:01	C*07:01:01	C*08:04:01
14	A*01:01:01:01	A*02:11:01	B*07:06:01	B*48:02:01	C*04:01:01:02	C*15:05:02
15	A*02:01:01:01	A*34:02:01	B*39:13:01	B*44:03:01	C*04:01:01:02	C*07:02:01:01
16	A*02:01:01:01	A*68:02:01:01	B*07:02:01:01	B*14:02:01:01	C*07:02:01:16	C*08:02:01:01
18	A*29:02:01:01	A*30:01:01	B*15:03:01:02	B*51:01:01:01	C*02:10:01:01	C*14:02:01:04
19	A*24:03:01	A*30:02:01	B*15:20	B*18:01:01:02	C*04:01:01:02	C*07:04:01:01
20	A*02:01:01:01	A*24:02:01:01	B*44:03:01:02	B*49:01:01:01	C*02:02:02:01	C*07:01:01:01
21	A*02:01:01:01	A*26:01:01:01	B*27:05:02:05	B*57:03:01:02	C*01:02:01:01	C*07:01:02
23	A*30:02:01	A*31:01:02:04	B*15:03:01	B*15:17:01	C*02:10:01	C*05:01:01
26	A*30:01:01	A*32:01:01:01	B*15:03:01:02	B*42:01:01	C*02:10:01:01	C*17:01:01:02
27	A*02:01:01:01	A*33:03:01:01	B*07:02:01:01	B*58:01:01:03	C*03:02:02:01	C*07:02:01:03
28	A*01:01:01:01	A*24:05:01	B*15:17:01:01	B*39:01:01:05	C*07:01:02	C*12:03:01:02
29	A*34:02:01	A*03:01:01:01	B*08:01:01:01	B*52:01:01:02	C*07:01:01:01	C*12:02:02:01
31	A*02:01:01:01	A*26:01:01:01	B*44:02:01:01	B*57:02:01	C*05:01:01:02	C*07:18
32	A*25:01:01:01	A*29:02:01:01	B*18:01:01:01	B*56:01:01:03	C*01:02:01:-- (alelo novo)	C*12:03:01:01

Em verde: IC50 < 50nM

Em vermelho claro: IC50 > 50 nM

X: alelo sem informação de tipagem

O epítopo candidato VGSLQYLAL (posição nucleotídica 4608-4634 referente à sequência-referência HXB2) foi analisado quanto à presença de variantes nos *reads* virais das amostras. Foram encontradas 10 variantes com frequência > 0,01 (1%) nos *reads*. Dentre as 10 variantes, sete apresentaram alta afinidade a todos os alelos de HLA dos pacientes. Duas variantes não apresentaram alta afinidade a um alelo específico do paciente, mas apresentaram alta afinidade aos outros cinco alelos do paciente. Por fim, uma variante não apresentou alta afinidade a três alelos de um mesmo paciente.

Essas variantes podem ser observadas na Tabela 5.14 abaixo.

Tabela 5.14 Variantes do epítipo viral VGSLQYLAL, da região de Vif, encontradas nos reads virais dos pacientes do HI analisados e resultado de predição de afinidade aos alelos de HLA-A, B e C específicos desses mesmos pacientes.

Paciente	Variante	Frequência	HLA					
			A		B		C	
2	IIVWQVDRI	9%	A*02:01	A*29:01	x	x	C*01:02	C*05:01
8	MIVWQVGRM	5%	A*02:01	A*24:02	B*07:02	B*52:01	C*04:01	C*07:02
13	MIAWQVDRM	2%	A*68:01	A*01:01	B*81:01	B*81:01	C*07:01	C*08:04
14	IIVWQVDRM	2%	A*01:01	A*02:11	B*07:06	B*48:02	C*04:01	C*15:05
	IIVWQVNRI	6%	A*01:01	A*02:11	B*07:06	B*48:02	C*04:01	C*15:05
20	MIVWQIDRM	2%	A*02:01	A*24:02	B*44:03	B*49:01	C*02:02	C*07:01
26	MIVWQVDGM	5%	A*30:01	A*32:01	B*15:03	B*42:01	C*02:10	C*17:01
27	MIWVWQVDRM	7%	A*02:01	A*33:03	B*07:02	B*58:01	C*03:02	C*07:02
	IIVWQVDRM	1%	A*02:01	A*33:03	B*07:02	B*58:01	C*03:02	C*07:02
28	MIVWQVDRI	2%	A*01:01	A*24:05	B*15:17	B*39:01	C*07:01	C*12:03

Em verde: IC50 < 50 nM

Em vermelho claro: IC50 > 50 nM

X: alelo sem informação de tipagem

6. DISCUSSÃO

O HLA é extremamente importante no controle imune das infecções virais, ao apresentar epítomos virais e desencadear uma resposta citotóxica contra o patógeno (CARRINGTON e O'BRIEN, 2003). Frente a esse contexto, vacinas personalizadas contendo epítomos virais conservados do HIV-1 com alta afinidade aos HLAs de classe I do paciente podem ser altamente eficazes contra o vírus, controlando a infecção após a interrupção do tratamento antirretroviral. Esse trabalho identificou os alelos de HLA de classe I de pacientes HIV-1+ sob terapia antirretroviral e controle da carga viral, e selecionou os melhores epítomos virais candidatos para a formação de uma vacina personalizada para duas diferentes regiões geográficas do Brasil onde diferentes subtipos do vírus circulam.

Os pacientes recrutados para esse estudo eram em sua maioria do sexo masculino e com aproximadamente um ano de intervalo desde o diagnóstico da

infecção pelo HIV até o início do tratamento antirretroviral. Os pacientes das casuísticas do Hospital Federal de Ipanema e do Hospital Universitário da FURG-RS tinham aproximadamente quatro anos e meio de diagnóstico HIV+ e aproximadamente três anos de tratamento. Já os pacientes do Hospital Universitário da UFRJ-RJ apresentaram um perfil de pacientes com maior tempo de diagnóstico HIV+ (aproximadamente 13 anos) e de tempo de tratamento (aproximadamente nove anos). Em outubro de 2013, o Ministério da Saúde, seguindo orientações da Organização Mundial da Saúde, atualizou as diretrizes para início do tratamento antirretroviral incentivando o acesso imediato à terapia ao paciente HIV+ recém-diagnosticado, independente da contagem de linfócitos T CD4+ (MINISTÉRIO DA SAÚDE, 2013). Essa mudança pode justificar o maior tempo entre o diagnóstico e início do tratamento visto para os pacientes do Hospital Universitário da UFRJ-RJ quando comparado aos outros dois hospitais estudados. Como os pacientes do Hospital Universitário da UFRJ-RJ foram diagnosticados há mais tempo, seguiam as recomendações de início do tratamento vigentes na época (contagem de linfócitos CD4+ < 350 céls/mm³ ou sintomatologia definidora de aids) (MINISTÉRIO DA SAÚDE, 2008).

O tempo de seleção dos pacientes pode ter grande impacto nos resultados de estudos de vacinas terapêuticas. Isso porque pacientes diagnosticados e tratados no início da infecção respondem diferentemente à vacinação comparados aos que foram tratados apenas após a fase crônica da infecção (MOTHE e BRANDER, 2018). Os reservatórios virais são formados durante os primeiros dias após a infecção. Desta forma, com o início precoce da terapia antirretroviral, é possível diminuir a formação dos reservatórios virais, possibilitando um maior controle da infecção por parte do hospedeiro (DEEKS *et al.*, 2016; GARCIA *et al.*, 2018). O intervalo de tempo médio de um ano para o início do tratamento antirretroviral dos nossos pacientes pode ter grande impacto na formação dos reservatórios virais, diversidade e escape imune viral.

Os experimentos realizados de extração de DNA genômico das amostras, amplificação e sequenciamento do HLA-A, B e C foram bem sucedidos, sem exclusão de amostras. O sequenciamento de nova geração é uma tecnologia que está se tornando rapidamente o novo padrão-ouro para a

tipagem de HLA de alta resolução. Uma característica importante é a capacidade de gerar dados de genotipagem de alta resolução mais completos e de melhor qualidade a um preço comparável ou mesmo menor do que os métodos tradicionais (PROFAIZER e KUMANOVICS, 2018).

Métodos de genotipagem baseados apenas nos éxons 2 e 3 (e às vezes 4) para HLAs de classe I e somente no éxon 2 para HLAs de classe II negligenciam variantes fora da região sequenciada, resultando em alelos ambíguos. Isso é particularmente grave no transplante de células-tronco hematopoiéticas, se uma variante fora dessas regiões comumente analisadas levar à formação de um alelo nulo. Além disso, limitar o sequenciamento à região da fenda de ligação negligencia qualquer polimorfismo intrônico ou outros polimorfismos não-codificantes que podem levar a alterações no splicing ou podem afetar a expressão da proteína (PROFAIZER e KUMANOVICS, 2018). Dessa forma, o sequenciamento de toda a região gênica do HLA vem sendo cada vez mais desenvolvido em protocolos bem estabelecidos, promovendo a análise dos polimorfismos encontrados fora da região que codifica a fenda de ligação de peptídeos. Tal interesse visa estudar o papel desses polimorfismos que podem não só se comportar como polimorfismos neutros, como também apresentar um papel funcional (SANCHEZ-MAZAS e MEYER, 2014).

No Brasil, poucos estudos realizaram a tipagem de alta resolução de alelos de HLA através de sequenciamento de nova geração. Além do nosso trabalho aqui apresentado, mais dois grupos também utilizaram a plataforma MiSeq para sequenciamento e tipagem em alta resolução dos alelos de HLA-A, B ou C. Lima e colaboradores analisaram o HLA-A de 408 amostras de São Paulo e encontraram alta frequência do alelo A*02:01:01:01 (19,5%) quando comparado com os outros alelos de HLA-A encontrados. O segundo alelo mais frequente foi o A*24:02:01:01, com 7,9% de frequência (LIMA *et al.*, 2019). Nosso estudo também encontrou uma alta frequência (22%) do alelo (A*02:01:01:01) na nossa casuística do Rio de Janeiro. Cabe destacar que as duas cidades pertencem à região sudeste do Brasil. Nosso segundo alelo mais frequente (A*30:01:01) também apresentou uma frequência de 8%, destacando a representatividade do alelo A*02:01:01:01. Já o alelo A*24:02:01:01 apresentou uma frequência de 4% na nossa população estudada.

Castro e colaboradores analisaram 108 indivíduos de Curitiba, Paraná, na região Sul do Brasil, e tiparam 12 loci do sistema HLA. De forma interessante, eles também encontraram o alelo A*02:01:01 com alta frequência (22,6%), enquanto para a nossa casuística de Rio Grande esse alelo foi o segundo mais frequente (15%). Já o alelo A*01:01:01, que encontramos com maior frequência (19%) dentre nossos pacientes, não foi encontrado com frequência tão elevada (7,5%) na casuística de Castro e colaboradores (CASTRO *et al.*, 2019).

Tais variações de frequência podem ser justificadas não só pela diferença de composição da casuística como também pela diferença regional. O Brasil apresenta uma grande diversidade étnica devido à colonização por diferentes países e migrações que ocorreram ao longo dos anos (CASTRO *et al.*, 2019; KEHDY *et al.*, 2015). A herança étnica difere entre as regiões do Brasil e assim também os alelos de HLA. Além disso, nossa casuística pode estar sofrendo algum viés devido à seleção de apenas pacientes HIV+, diferente das casuísticas dos outros estudos, em que os participantes pertenciam à população geral. Outro fator importante é a tipagem em alta resolução, que aumenta a especificidade do alelo. Dessa forma, encontramos diferentes frequências para alelos que seriam agrupados com uma frequência única se fossem tipados com baixa resolução (de apenas 2 dígitos).

Ao reduzirmos nossa resolução de tipagem, comparamos as frequências alélicas encontradas com diferentes estudos realizados no Rio Grande do Sul e Rio de Janeiro. Também comparamos com as frequências encontradas nos dados do REDOME de cada estado, e para alguns alelos foram encontradas diferenças estatisticamente significantes.

Bouquett e colaboradores estudaram uma população de 97.292 doadores de medula do estado do Rio Grande do Sul. Para a cidade de Rio Grande, eles encontraram os alelos A*02, A*01, B*35 e B*44 com maior frequência (26%, 12%, 13% e 11% respectivamente) (BOQUETT *et al.*, 2018). Os dados do REDOME-RS também mostraram frequência similar para o alelo A*01 (11%). No entanto, ao compararmos a frequência encontrada na nossa casuística do Rio Grande para o alelo A*01 (20%) com as demais, esta se mostrou significativamente mais alta ($p = 0,00001$).

Já o alelo B*14, que foi um dos mais frequentes na nossa casuística do

Rio Grande (9%), se mostrou menos frequente quando comparado com os dados do REDOME-RS (4%) e essa diferença foi significativa ($p = 0,0068$). De forma interessante, Bouquett e colaboradores encontraram frequência próxima à nossa para o alelo B*14 (7%) (BOQUETT *et al.*, 2018) e outro grupo que estudou 5.000 doadores de medula óssea de Porto Alegre e municípios próximos encontrou uma frequência de 5% (BORTOLOTTO *et al.*, 2012), demonstrando assim uma variação de frequência do alelo B*14 nessa região.

Para a casuística do Rio de Janeiro, encontramos uma maior frequência do alelo B*07 (14%) quando comparado com os dados de frequência do REDOME-RJ (7%), e essa diferença apresentou relevância estatística ($p = 0,0001$). Além disso, Teixeira e colaboradores também encontraram frequência similar (7%) ao do REDOME-RJ para esse alelo em questão ao estudar 96 usuários de drogas injetáveis HIV+ do Rio de Janeiro (TEIXEIRA *et al.*, 2009). Essas diferenças de frequência encontradas podem ser explicadas, como mencionado anteriormente, devido a diferenças de seleção da casuística, metodologia de tipagem, entre outros fatores.

Durante a tipagem dos alelos fomos capazes de identificar possíveis alelos novos ainda não descritos no banco de dados IMGT-HLA. O software utilizado para tipagem apontou alelos classificados como alelos novos ou apresentou ambiguidades que não tornaram possível concluir a tipagem. Esses casos foram submetidos ao nosso pipeline de tipagem desenvolvido in-house, e com ele inspecionamos e definimos as trocas nucleotídicas, excluindo os casos de erros de tipagem do software. Após essa análise, foi possível observar que todas as trocas encontradas foram polimorfismos de apenas uma base nucleotídica em regiões não-codificantes. Esses alelos novos serão submetidos a outras metodologias de tipagem, como o sequenciamento de Sanger seguido da clonagem, para a confirmação e validação dos mesmos. Posteriormente, serão depositados no IMGT-HLA.

Vários fatores como a expressão de proteínas virais, processamento de antígeno, imunodominância e a presença de um repertório de células T são importantes para a previsão de epítopos de células T, além da predição de peptídeos que se ligam às moléculas de HLA. No entanto, como a ligação do peptídeo à molécula de HLA classe I é sempre necessária para a apresentação de antígenos, a predição eficiente dessa associação deve reduzir

consideravelmente o número de peptídeos a serem testados (MOUTAFTSI *et al.*, 2006). Dessa forma, foram adotados alguns critérios durante a análise para seleção dos melhores epítomos presentes em nossa casuística, tais como: concordância dos possíveis epítomos vacinais ao genoma proviral arquivado, seleção de epítomos de regiões conservadas do HIV, ligação de alta afinidade ao alelo de HLA de classe I e identificação de variantes dos epítomos presentes nas quasispécies virais.

Papuchon e colaboradores argumentam que, uma vez que o vírus emergente da interrupção do tratamento se origina do DNA proviral, é crucial verificar se os epítomos da vacina condizem com os epítomos arquivados que serão apresentados por alelos de HLA classe I. No paciente infectado sem uso de antirretrovirais, devido a erros na transcriptase reversa e desenvolvimento de quasispécies, o número de epítomos mutados aumenta com o tempo desde a fase de infecção primária. Ao iniciar o tratamento antirretroviral essa atividade cessa, porém ao interromper o tratamento o vírus que irá se replicar tem origem dos reservatórios celulares reativados. No entanto, esse vírus pode ser diferente do vírus circulante antes do tratamento (FLEURY *et al.*, 2018; PAPUCHON *et al.*, 2014; PAPUCHON *et al.*, 2013).

Dessa forma, potenciais epítomos identificados com base na sequência-referência HXB2 ou com base em sequências-consenso não são adequados, uma vez que os epítomos muito provavelmente serão diferentes do provírus arquivado e uma reação cruzada não será normalmente esperada. Além disso, epítomos genéricos não são sistematicamente adequados para apresentação devido à diversidade de antígenos e alelos HLA de classe I correspondentes (FLEURY *et al.*, 2018; PAPUCHON *et al.*, 2014; PAPUCHON *et al.*, 2013). Com base nessas informações, optamos por utilizar em nossas análises a sequência do próprio vírus majoritário que infecta o paciente, oriunda do sequenciamento do provírus arquivado, para buscar os epítomos que se ligam ao HLA de classe I próprio do paciente.

Nem toda região do genoma do HIV-1 desencadeiam respostas CTL eficazes contra o vírus. Vários estudos mostraram que regiões que correspondem às proteínas Gag e Pol são mais eficazes do que regiões do envelope viral, por exemplo, e devem ser consideradas como potenciais alvos para o desenho de imunógenos (KIEPIELA *et al.*, 2007; MOTHE e BRANDER,

2018; ZUNIGA, 2006). Além disso, essas regiões são mais conservadas do que o restante do genoma viral e possuem indícios de induzir respostas de célula T com alta avidéz funcional, uma característica de caráter decisivo na resposta efetiva contra patógenos hipervariáveis (MOTHE *et al.*, 2012; YERLY *et al.*, 2008). Para a predição dos epítomos virais, selecionamos as regiões gênicas Gag, Pol, Nef e Vif do HIV-1. Essas regiões foram previamente analisadas por Mothe e colaboradores, que destacaram as regiões Gag e Pol como regiões promissoras, e propuseram a inclusão de fragmentos de Nef e Vif (MOTHE *et al.*, 2011).

Já em 2012, um estudo de Mothe e colaboradores mostrou que os indivíduos controladores do HIV-1 montam resposta imune de avidéz funcional significativamente mais alta do que os não-controladores do HIV-1. As células T de alta avidéz funcional apresentam maior capacidade de reconhecer células infectadas por vírus e também de implantar funções efetoras mesmo quando há baixa densidade de antígeno, quando comparadas às populações de células T de baixa avidéz. Uma vez que as respostas de alta avidéz também se mostraram mais aptas a reagir a diferentes variantes do epítomo, sua indução em uma futura vacina contra o HIV-1 poderia ser crucial para impedir o escape imune (MOTHE *et al.*, 2012; TUMIOTTO *et al.*, 2019). Com base nesses experimentos, optamos por selecionar apenas antígenos que apresentaram alta afinidade de ligação ($IC_{50} < 50nM$) à molécula restritiva de HLA, priorizando respostas de alta avidéz.

Estudos recentes vêm demonstrando que respostas de células T à epítomos não-dominantes, ou seja, que visam regiões conservadas de proteínas do HIV-1, mas imunologicamente subdominantes, seriam mais protetoras e efetivas do que as aquelas que reconhecem epítomos canônicos imunodominantes. Isso ocorre porque os epítomos imunodominantes são hipervariáveis e mais passíveis de escape imune viral. Com isso, alguns grupos estão selecionando imunógenos mais promissores e conservados, ao invés de proteínas de HIV-1 completas que induziriam resposta de CTL contra epítomos imunodominantes. Dessa forma, seria possível combater a diversidade viral e o escape imune do HIV-1 e contornar o efeito da genética do hospedeiro na resposta natural ao HIV-1 (DEEKS *et al.*, 2016; MOTHE *et al.*, 2019).

Após as análises, encontramos 4 epítomos candidatos (KARVLAEM,

EMMTACQGV, MIGGIGGFI, VGSLQYLAL) que atendem aos critérios estabelecidos localizados nas regiões proteicas de Gag (2 epítomos), Pol (1 epítomo) e Vif (1 epítomo). Esses epítomos estavam presentes em mais de 95% das sequências virais analisadas dos pacientes do Rio de Janeiro. Nenhum epítomo apresentou baixa afinidade a mais de um alelo de HLA de um mesmo paciente, ou seja, todos os pacientes possuíam pelo menos cinco alelos que reconheciam os epítomos com alta afinidade. Apenas nove dos 18 pacientes analisados apresentaram apenas um alelo de HLA sem alta afinidade para os epítomos KARVLAEAM e EMMTACQGV da região Gag. Seis pacientes, dos 19 analisados apresentaram apenas um alelo de HLA sem alta afinidade para o epítomo MIGGIGGFI encontrado na região Pol. Já para o epítomo VGSLQYLAL da região Vif, apenas um paciente apresentou um alelo sem alta afinidade. Esses epítomos encontrados no nosso estudo já foram também identificados em outros estudos ou se encontram descritos na lista de epítomos de CTL do banco de dados de Los Alamos Immunology Database (KARLSSON *et al.*, 2007; LUO *et al.*, 2012; MUELLER *et al.*, 2007).

Após selecionarmos os epítomos candidatos presentes na maioria dos consensos virais dos pacientes analisados e que fosse apresentado não só pelos HLAs mais frequentes como também pelos HLAs específicos de cada paciente, também observamos a variabilidade desses epítomos entre as quasispecies virais com a finalidade de evitar escape imune viral por parte de variantes presentes. As variantes dos epítomos com alta frequência entre as quasispecies e que apresentam baixa afinidade de ligação aos alelos de HLA mais frequentes poderia comprometer a eficácia da vacina. Portanto, é crucial identificar epítomos de CTL que não sejam mutados ou, caso mutados, que mantenham sua capacidade de se ligar à molécula HLA classe I presente (TUMIOTTO *et al.*, 2017).

Nosso estudo encontrou um total de 22 epítomos variantes, com frequência entre 1% e 22% entre os reads virais, para os quatro epítomos candidatos. Foram cinco variantes para o epítomo KARVLAEAM, 4 variantes para o epítomo EMMTACQGV, 3 variantes para o epítomo MIGGIGGFI e 10 para o VGSLQYLAL. No entanto, apenas seis dessas variantes mostraram uma menor afinidade a apenas um alelo do paciente. Dessa forma o paciente ainda teria mais cinco alelos de HLA de classe I com alta afinidade à variante, caso

ela venha a ressurgir após a interrupção do tratamento. Uma variante do epítopo encontrado em Vif não apresentou alta afinidade a três alelos de um paciente. No entanto, como essa variante se encontra com uma frequência de apenas 2% nos reads virais, optamos por manter esse epítopo como candidato, uma vez que esse paciente que possui essa variante ainda teria três alelos de HLA com alta afinidade à variante, caso venha a ressurgir ao se interromper o tratamento.

Como mencionado anteriormente, esse estudo faz parte de um projeto maior (“Provir / Latitude 45”) que objetiva identificar epítomos de CTL provirais de HIV conservados apresentados pelos alelos HLA classe I mais prevalentes em pacientes tratados com cART de diferentes países, com o objetivo de projetar uma sequência imunogênica mais universalmente utilizável (TUMIOTTO *et al.*, 2019). Assim como nós, o estudo concentrou-se nas partes genômicas do HIV-1 consideradas cruciais para os epítomos de CTL, Gag, Pol e Nef, onde já foram descritos alguns epítomos potencialmente benéficos.

Em seu artigo recentemente publicado, Tumiotto e colaboradores apresentaram 15 epítomos conservados nas regiões da Pol, Gag e Nef. De todos os epítomos selecionado por eles, apenas um deles (VLAEAMSQV), da região da Gag, se sobrepõe a um dos nossos epítomos em alguns aminoácidos (KARVLAEAM). Essa diferença de epítomos encontrada pode ser justificada pela diferença de casuística entre os estudos, uma vez que além das amostras incluídas no nosso estudo, Tumiotto também analisou amostras de outras regiões como Canadá, França, Estados Unidos e Peru. Essa variação geográfica influencia não só na diversidade viral como nos alelos de HLA. Além disso, assim como nós, alguns epítomos selecionados por Tumiotto e colaboradores já estavam descritos na lista de epítomos CTL descrita por Los Alamos (https://www.hiv.lanl.gov/content/immunology/tables/ctl_summary.html). Por fim, Tumiotto e colaboradores também encontraram casos raros de epítomos variantes com baixa afinidade pelo alelo de HLA, sendo todos com baixa frequência (ao redor de 1%) (TUMIOTTO *et al.*, 2019).

Algumas questões importantes devem ser tratadas frente às limitações de nosso estudo. Primeiramente, embora seja bem aceito que o reservatório viral se estabeleça precocemente na infecção aguda, a evolução viral e mudanças ao longo do tempo nos alvos imunes podem complicar a definição

do melhor tempo de coleta das amostras e da análise das sequências virais de pacientes sob tratamento antirretroviral a longo prazo. Alguns estudos indicam que os reservatórios virais se mantêm estáveis e com poucas alterações genéticas, porém ainda seria necessário verificar por NGS se o mesmo seria aplicável aos epítomos CTL do vírus arquivado em diferentes compartimentos (TUMIOTTO *et al.*, 2017).

Além disso, a incerteza de que de qual compartimento e tipo celular o vírus irá emergir após a interrupção do tratamento torna desejável a inclusão de sequências virais de diversos locais anatômicos e origens celulares. No entanto, alguns sítios e subconjuntos celulares são de difícil acesso, não só por limitações técnicas como éticas (DE SCHEERDER *et al.*, 2019). Frente a isso, as implicações terapêuticas de identificar um reservatório específico é de grande importância para aumentar a eficiência dos esforços no controle do HIV em pacientes sem terapia antirretroviral.

Do ponto de vista econômico e industrial, vacinas personalizadas podem apresentar grandes obstáculos, uma vez que elas podem exigir a síntese de imunógenos individuais e personalizados. No entanto, na tentativa de reduzir essas barreiras, a formação de grupos regionais dentro de um país pode reduzir custos ao agruparmos indivíduos que compartilhem os alelos de HLA mais frequentes da região e sequências virais conservadas. Dessa forma, esse estudo se destaca ao realizar a análise dos alelos de HLA em alta resolução e dos epítomos provirais de diferentes regiões do genoma do HIV-1, considerando os alelos mais frequentes e os alelos de cada paciente de diferentes regiões do Brasil.

Cabe destacar também que esse é o primeiro estudo de tipagem de alta resolução dos alelos de HLA baseado em sequenciamento de nova geração de indivíduos HIV+ das cidades do Rio de Janeiro, RJ e Rio Grande, RS. Considerando as informações aqui obtidas, foi possível definir quatro epítomos provirais presentes em mais de 95% da população estudada e com alta afinidade para desencadear uma resposta imune CTL predita eficaz no combate à replicação viral em caso de interrupção do tratamento.

De maneira geral, as informações desse estudo contribuem para avanços nos estudos de vacina terapêutica personalizada contra o HIV-1 e destacam a relevância de novos estudos baseados em sequenciamento de

nova geração para melhor correlacionar a tipagem de alta resolução dos alelos de HLA e a definição de epítomos virais eficazes, considerando-se a presença de variantes de epítomos intra-paciente. A partir dessas informações será possível investigar uma nova estratégia de tratamento contra o HIV, capaz de reduzir em longo prazo os custos associados ao tratamento crônico com uso de antirretrovirais, além de contribuir na diminuição de sua transmissão e morbidade.

7. CONCLUSÕES

- Com o sequenciamento de nova geração na plataforma Illumina MiSeq e o pipeline de tipagem desenvolvido por nosso grupo, foi possível a tipagem em alta resolução do HLA de classe I nos loci A, B e C de 84 pacientes HIV-1+ provenientes de duas regiões diferentes do Brasil, Hospital Federal de Ipanema (RJ) e Hospital Universitário Clementino Fraga Filho-UFRJ (RJ) na região Sudeste, e Hospital Universitário Dr. Miguel Riet Corrêa Jr.- FURG (RS), na região Sul do Brasil. Essa nova metodologia de tipagem, que possibilita a descrição completa do alelo, vem se tornando um novo padrão, e é amplamente utilizada no mundo, no entanto ainda é pouco aplicada no Brasil, onde predomina a tipagem de baixa e média resolução.
- Com essa nova metodologia aplicada, também foi possível a reconstrução completa dos genes, resolução de ambiguidades, além da identificação e caracterização de potenciais novos alelos de HLA de classe I presentes na população brasileira. Ao todo, foram identificados oito possíveis alelos novos, requerendo confirmação com metodologias adicionais para seu estabelecimento final.
- Com as análises de predição de epítomos, encontramos quatro epítomos candidatos (KARVLAEAM, EMMTACQGV, MIGGIGGFI, VGSLQYLAL) localizados em regiões conservadas do HIV-1 (Gag, Pol e Vif), com alta afinidade de ligação à molécula restritiva de HLA classe I ($IC_{50} < 50$ nM).
- Todos os epítomos foram avaliados quanto à afinidade aos alelos de HLA-A,B e C mais frequentes na população estudada e específicos de cada paciente, de forma a favorecer a ligação do epítomo ao HLA e seu reconhecimento por receptores TCR, desencadeando uma resposta CTL contra

o vírus.

- A análise de variabilidade dos epítomos virais identificou 22 epítomos variantes com frequência superior a 1% para os quatro epítomos candidatos. Entretanto, apenas seis dessas variantes mostraram uma menor afinidade a apenas um dos seis alelos dos pacientes. Isso indica que, apesar da variabilidade dos epítomos decorrente de variações nas quasisespécies virais, os epítomos candidatos possuem alta afinidade para os alelos de HLA testados, o que auxilia na prevenção de um possível escape imune viral.
- A partir dessas informações será possível realizar uma nova etapa deste estudo de prova de conceito para o desenvolvimento de uma vacina terapêutica contra o HIV, na qual os epítomos aqui identificados serão submetidos à análise de ELISpot para validar seu potencial uso em uma vacina.

8. BIBLIOGRAFIA

ALVES, B. M. *et al.* **Characterization of HIV-1 Near Full-Length Proviral Genome Quasispecies from Patients with Undetectable Viral Load Undergoing First-Line HAART Therapy.** 2017. *Viruses*. 9 (12):

APOSTOLOPOULOS, V. *et al.* **MHC and MHC-like molecules: structural perspectives on the design of molecular vaccines.** 2008. *Adv Exp Med Biol*. 640 252-267.

APTSIAURI, N. *et al.* **MHC class I antigens and immune surveillance in transformed cells.** 2007a. *Int Rev Cytol*. 256 139-189.

APTSIAURI, N. *et al.* **Role of altered expression of HLA class I molecules in cancer progression.** 2007b. *Adv Exp Med Biol*. 601 123-131.

BANGHAM, C. R. **CTL quality and the control of human retroviral infections.** 2009. *Eur J Immunol*. 39 (7):1700-1712.

BANSAL, A. *et al.* **Immunological control of chronic HIV-1 infection: HLA-mediated immune function and viral evolution in adolescents.** 2007. *AIDS*. 21 (18):2387-2397.

BARRE-SINOUSI, F. *et al.* **Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).** 1983. *Science*. 220 (4599):868-871.

BAUER, D. C. Z., A; WILSON, L. O. W.; MELBOURNE GENOMICS HEALTH

- ALLIANCE; THORNE, N. P. **Evaluation of computational programs to predict HLA genotypes from genomic sequencing data.** 2018. *Briefings in bioinformatics.* 19 (2):179–187.
- BECK, S. e TROWSDALE, J. **The human major histocompatibility complex: lessons from the DNA sequence.** 2000. *Annu Rev Genomics Hum Genet.* 1 117-137.
- BENNETT, D. E. *et al.* **Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update.** 2009. *PLoS One.* 4 (3):e4724.
- BENTLEY, G. *et al.* **High-resolution, high-throughput HLA genotyping by next-generation sequencing.** 2009. *Tissue Antigens.* 74 (5):393-403.
- BERGAMASCHI, A. e PANCINO, G. **Host hindrance to HIV-1 replication in monocytes and macrophages.** 2010. *Retrovirology.* 7 31.
- BJORKMAN, P. J. e PARHAM, P. **Structure, function, and diversity of class I major histocompatibility complex molecules.** 1990. *Annu Rev Biochem.* 59 253-288.
- BJORKMAN, P. J. *et al.* **The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens.** 1987a. *Nature.* 329 (6139):512-518.
- BJORKMAN, P. J. *et al.* **Structure of the human class I histocompatibility antigen, HLA-A2.** 1987b. *Nature.* 329 (6139):506-512.
- BOQUETT, J. A. *et al.* **Spatial analyzes of HLA data in Rio Grande do Sul, south Brazil: genetic structure and possible correlation with autoimmune diseases.** 2018. *International journal of health geographics.* 17 (1):34.
- BORTOLOTTO, A. S. *et al.* **HLA-A, -B, and -DRB1 allelic and haplotypic diversity in a sample of bone marrow volunteer donors from Rio Grande do Sul State, Brazil.** 2012. *Hum Immunol.* 73 (2):180-185.
- BOUR, S. e STREBEL, K. **The HIV-1 Vpu protein: a multifunctional enhancer of viral particle release.** 2003. *Microbes Infect.* 5 (11):1029-1039.
- BOWERMAN, B. *et al.* **A nucleoprotein complex mediates the integration of retroviral DNA.** 1989. *Genes Dev.* 3 (4):469-478.
- BRUMME, Z. L. *et al.* **Modulation of HIV reservoirs by host HLA: bridging the gap between vaccine and cure.** 2012. *Curr Opin Virol.* 2 (5):599-605.
- BUSHMAN, F. D. *et al.* **Retroviral DNA integration directed by HIV integration protein in vitro.** 1990. *Science.* 249 (4976):1555-1558.
- CAGLIANI, R. e SIRONI, M. **Pathogen-driven selection in the human**

- genome**. 2013. International journal of evolutionary biology. 2013 204240.
- CANN, A. J. e KARN, J. **Molecular biology of HIV: new insights into the virus life-cycle**. 1989. AIDS. 3 Suppl 1 S19-34.
- CAO, H. *et al.* **An integrated tool to study MHC region: accurate SNV detection and HLA genes typing in human MHC region using targeted high-throughput sequencing**. 2013. PLoS One. 8 (7):e69388.
- CARLSON, J. M. *et al.* **Widespread impact of HLA restriction on immune control and escape pathways of HIV-1**. 2012. J Virol. 86 (9):5230-5243.
- CARRINGTON, M. **Recombination within the human MHC**. 1999. Immunol Rev. 167 245-256.
- CARRINGTON, M. e O'BRIEN, S. J. **The influence of HLA genotype on AIDS**. 2003. Annu Rev Med. 54 535-551.
- CARRINGTON, M. e WALKER, B. D. **Immunogenetics of spontaneous control of HIV**. 2012. Annu Rev Med. 63 131-145.
- CASADO, C. *et al.* **Host and viral genetic correlates of clinical definitions of HIV-1 disease progression**. 2010. PLoS One. 5 (6):e11079.
- CASTELLI, E. C. *et al.* **Hla-mapper: An application to optimize the mapping of HLA sequences produced by massively parallel sequencing procedures**. 2018. Hum Immunol. 79 (9):678-684.
- CASTRO, M. S. *et al.* **High-resolution characterization of 12 classical and non-classical HLA loci in Southern Brazilians**. 2019. Hla. 93 (2-3):80-88.
- CEREDA, C. G., S.; COVA, E.; DIAMANTI, L; CERONI, M. **The Role of TNFAlpha in ALS: New Hypotheses for Future Therapeutic Approaches**. 2012.
- CHOO, S. Y. **The HLA system: genetics, immunology, clinical testing, and clinical implications**. 2007. Yonsei Med J. 48 (1):11-23.
- CLAPHAM, P. R. e WEISS, R. A. **Immunodeficiency viruses. Spoilt for choice of co-receptors**. 1997. Nature. 388 (6639):230-231.
- COFFIN, J. M. *et al.* **The Interactions of Retroviruses and their Hosts**. 1997.
- CONSORTIUM, T. M. S. **Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium**. 1999. Nature. 401 (6756):921-923.
- COWAN, M. J. *et al.* **Maternal transmission of acquired immune deficiency syndrome**. 1984. Pediatrics. 73 (3):382-386.

CURIGLIANO, G. *et al.* **Molecular pathways: human leukocyte antigen G (HLA-G).** 2013. Clin Cancer Res. 19 (20):5564-5571.

DALGLEISH, A. G. *et al.* **The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus.** 1984. Nature. 312 (5996):763-767.

DALMASSO, C. *et al.* **Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study.** 2008. PLoS One. 3 (12):e3907.

DANZER, M. *et al.* **Rapid, scalable and highly automated HLA genotyping using next-generation sequencing: a transition from research to diagnostics.** 2013. BMC Genomics. 14 221.

DARLIX, J. L. *et al.* **Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1.** 1990. J Mol Biol. 216 (3):689-699.

DAUSSET, J. **The birth of MAC.** 1984. Vox Sang. 46 (4):235-237.

DAVEY, R. T., JR. *et al.* **HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression.** 1999. Proc Natl Acad Sci U S A. 96 (26):15109-15114.

DE SANTIS, D. *et al.* **16(th) IHIW : review of HLA typing by NGS.** 2013. Int J Immunogenet. 40 (1):72-76.

DE SCHEERDER, M. A. *et al.* **HIV Rebound Is Predominantly Fueled by Genetically Identical Viral Expansions from Diverse Reservoirs.** 2019. Cell Host Microbe. 26 (3):347-358 e347.

DEEKS, S. G. *et al.* **International AIDS Society global scientific strategy: towards an HIV cure 2016.** 2016. Nat Med. 22 (8):839-850.

DEEKS, S. G. e WALKER, B. D. **Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy.** 2007. Immunity. 27 (3):406-416.

DEL CAMPO, A. B. *et al.* **Targeting HLA class I expression to increase tumor immunogenicity.** 2012. Tissue Antigens. 79 (3):147-154.

DENG, K. *et al.* **Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations.** 2015. Nature. 517 (7534):381-385.

DITTMAR, M. T. *et al.* **HIV-1 tropism and co-receptor use.** 1997. Nature. 385 (6616):495-496.

DOHERTY, P. C. e ZINKERNAGEL, R. M. **A biological role for the major histocompatibility antigens.** 1975. Lancet. 1 (7922):1406-1409.

EHRENBERG, P. K. *et al.* **High-throughput multiplex HLA genotyping by next-generation sequencing using multi-locus individual tagging.** 2014. BMC Genomics. 15 864.

EHRENBERG, P. K. *et al.* **High-Throughput Contiguous Full-Length Next-Generation Sequencing of HLA Class I and II Genes from 96 Donors in a Single MiSeq Run.** 2018. Methods Mol Biol. 1802 89-100.

ELAHI, S. *et al.* **Protective HIV-specific CD8+ T cells evade Treg cell suppression.** 2011. Nat Med. 17 (8):989-995.

ENGELHARD, V. H. **Structure of peptides associated with MHC class I molecules.** 1994. Curr Opin Immunol. 6 (1):13-23.

ENSOLI, B. *et al.* **Challenges in HIV Vaccine Research for Treatment and Prevention.** 2014. Front Immunol. 5 417.

ERLICH, H. **HLA DNA typing: past, present, and future.** 2012. Tissue Antigens. 80 (1):1-11.

FALK, K. *et al.* **Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules.** 1991. Nature. 351 (6324):290-296.

FARQUHAR, C. *et al.* **Human leukocyte antigen (HLA) B*18 and protection against mother-to-child HIV type 1 transmission.** 2004. AIDS Res Hum Retroviruses. 20 (7):692-697.

FELLAY, J. **Host genetics influences on HIV type-1 disease.** 2009. Antivir Ther. 14 (6):731-738.

FELLAY, J. *et al.* **Common genetic variation and the control of HIV-1 in humans.** 2009. PLoS Genet. 5 (12):e1000791.

FINZI, D. *et al.* **Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy.** 1997. Science. 278 (5341):1295-1300.

FLEURY, H. *et al.* **Therapeutic Vaccine Against HIV, Viral Variability, Cytotoxic T Lymphocyte Epitopes, and Genetics of Patients.** 2018. AIDS Res Hum Retroviruses. 34 (1):27-30.

FRAHM, N. *et al.* **HLA-B63 presents HLA-B57/B58-restricted cytotoxic T-lymphocyte epitopes and is associated with low human immunodeficiency virus load.** 2005. J Virol. 79 (16):10218-10225.

FRAHM, N. *et al.* **Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes.** 2006. Nat Immunol. 7 (2):173-178.

FRANKEL, A. D. e YOUNG, J. A. **HIV-1: fifteen proteins and an RNA.** 1998.

Annu Rev Biochem. 67 1-25.

FREED, E. O. **HIV-1 replication**. 2001. Somat Cell Mol Genet. 26 (1-6):13-33.

GALLAY, P. *et al.* **HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway**. 1997. Proc Natl Acad Sci U S A. 94 (18):9825-9830.

GALLAY, P. *et al.* **HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase**. 1995. Cell. 83 (4):569-576.

GALLO, R. C. **Kaplan memorial lecture. The family of human lymphotropic retroviruses called HTLV: HTLV-I in adult T-cell leukemia (ATL), HTLV-II in hairy cell leukemias, and HTLV-III in AIDS**. 1984. Princess Takamatsu symposia. 15 13-38.

GALLO, R. C. *et al.* **Human T-cell leukemia-lymphoma virus (HTLV): a progress report**. 1983. Haematology and blood transfusion. 28 311-319.

GARCIA, J. V. e MILLER, A. D. **Downregulation of cell surface CD4 by nef**. 1992. Research in virology. 143 (1):52-55.

GARCIA, M. *et al.* **Peering into the HIV reservoir**. 2018. Reviews in medical virology. 28 (4):e1981.

GARRETT, T. P. *et al.* **Specificity pockets for the side chains of peptide antigens in HLA-Aw68**. 1989. Nature. 342 (6250):692-696.

GATTI, E. e PIERRE, P. **Understanding the cell biology of antigen presentation: the dendritic cell contribution**. 2003. Curr Opin Cell Biol. 15 (4):468-473.

GONZALEZ-GALARZA, F. F. *et al.* **Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations**. 2011. Nucleic Acids Res. 39 (Database issue):D913-919.

GORER, P. A. e SCHUTZE, H. **Genetical studies on immunity in mice: II. Correlation between antibody formation and resistance**. 1938. J Hyg (Lond). 38 (6):647-662.

GOTO, T. *et al.* **Projection structures of human immunodeficiency virus type 1 (HIV-1) observed with high resolution electron cryo-microscopy**. 1994. J Electron Microsc (Tokyo). 43 (1):16-19.

GOTTLIEB, M. S. *et al.* **Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency**. 1981. N Engl J Med. 305 (24):1425-1431.

GOTTLINGER, H. G. *et al.* **Role of capsid precursor processing and**

myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. 1989. Proc Natl Acad Sci U S A. 86 (15):5781-5785.

GRUMBT, B. *et al.* **Diagnostic applications of next generation sequencing in immunogenetics and molecular oncology.** 2013. Transfus Med Hemother. 40 (3):196-206.

HANCOCK, G. *et al.* **Evaluation of the immunogenicity and impact on the latent HIV-1 reservoir of a conserved region vaccine, MVA.HIVconsv, in antiretroviral therapy-treated subjects.** 2017. Journal of the International AIDS Society. 20 (1):21171.

HARRISON, G. P. e LEVER, A. M. **The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure.** 1992. J Virol. 66 (7):4144-4153.

HEINZINGER, N. K. *et al.* **The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells.** 1994. Proc Natl Acad Sci U S A. 91 (15):7311-7315.

HIRSCH, M. S. *et al.* **Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel.** 2008. Top HIV Med. 16 (3):266-285.

HO, D. D. **HIV-1 dynamics in vivo.** 1995. J Biol Regul Homeost Agents. 9 (3):76-77.

HOLCOMB, C. L. *et al.* **A multi-site study using high-resolution HLA genotyping by next generation sequencing.** 2011. Tissue Antigens. 77 (3):206-217.

HOSOMICHI, K. *et al.* **Phase-defined complete sequencing of the HLA genes by next-generation sequencing.** 2013. BMC Genomics. 14 355.

JACOTOT, E. *et al.* **The HIV-1 viral protein R induces apoptosis via a direct effect on the mitochondrial permeability transition pore.** 2000. J Exp Med. 191 (1):33-46.

JETT, J. R. *et al.* **Acquired immunodeficiency syndrome associated with blood-product transfusions.** 1983. Ann Intern Med. 99 (5):621-624.

JOOS, B. *et al.* **HIV rebounds from latently infected cells, rather than from continuing low-level replication.** 2008. Proc Natl Acad Sci U S A. 105 (43):16725-16730.

JOWETT, J. B. *et al.* **The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle.** 1995. J Virol. 69 (10):6304-6313.

KAMYA, P. *et al.* **T cell Activation does not drive CD4 decline in longitudinally followed HIV-infected Elite Controllers.** 2011. *AIDS Res Ther.* 8 (1):20.

KARLSSON, A. C. *et al.* **Sequential broadening of CTL responses in early HIV-1 infection is associated with viral escape.** 2007. *PLoS One.* 2 (2):e225.

KARLSSON, I. *et al.* **Identification of conserved subdominant HIV Type 1 CD8(+) T Cell epitopes restricted within common HLA Supertypes for therapeutic HIV Type 1 vaccines.** 2012. *AIDS Res Hum Retroviruses.* 28 (11):1434-1443.

KASLOW, R. A. *et al.* **Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection.** 1996. *Nat Med.* 2 (4):405-411.

KAUR, G. e MEHRA, N. **Genetic determinants of HIV-1 infection and progression to AIDS: immune response genes.** 2009a. *Tissue Antigens.* 74 (5):373-385.

KAUR, G. e MEHRA, N. **Genetic determinants of HIV-1 infection and progression to AIDS: susceptibility to HIV infection.** 2009b. *Tissue Antigens.* 73 (4):289-301.

KAWASHIMA, Y. *et al.* **Adaptation of HIV-1 to human leukocyte antigen class I.** 2009. *Nature.* 458 (7238):641-645.

KEHDY, F. S. *et al.* **Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations.** 2015. *Proc Natl Acad Sci U S A.* 112 (28):8696-8701.

KIEPIELA, P. *et al.* **Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.** 2004. *Nature.* 432 (7018):769-775.

KIEPIELA, P. *et al.* **CD8+ T-cell responses to different HIV proteins have discordant associations with viral load.** 2007. *Nat Med.* 13 (1):46-53.

KLEIN, J. e SATO, A. **The HLA system. First of two parts.** 2000a. *N Engl J Med.* 343 (10):702-709.

KLEIN, J. e SATO, A. **The HLA system. Second of two parts.** 2000b. *N Engl J Med.* 343 (11):782-786.

KOUP, R. A. *et al.* **The quest for a T cell-based immune correlate of protection against HIV: a story of trials and errors.** 2011. *Nat Rev Immunol.* 11 (1):65-70.

KULKARNI, S. *et al.* **Differential microRNA regulation of HLA-C expression and its association with HIV control.** 2011. *Nature.* 472 (7344):495-498.

- LANGE, V. *et al.* **Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing.** 2014. BMC Genomics. 15 63.
- LANK, S. M. *et al.* **Ultra-high resolution HLA genotyping and allele discovery by highly multiplexed cDNA amplicon pyrosequencing.** 2012. BMC Genomics. 13 378.
- LAZARYAN, A. *et al.* **Human leukocyte antigen B58 supertype and human immunodeficiency virus type 1 infection in native Africans.** 2006. J Virol. 80 (12):6056-6060.
- LEITE, T. C. *et al.* **Impact of HIV-1 Subtypes on AIDS Progression in a Brazilian Cohort.** 2017. AIDS Res Hum Retroviruses. 33 (1):41-48.
- LESLIE, A. *et al.* **Additive contribution of HLA class I alleles in the immune control of HIV-1 infection.** 2010. J Virol. 84 (19):9879-9888.
- LETVIN, N. L. e WALKER, B. D. **Immunopathogenesis and immunotherapy in AIDS virus infections.** 2003. Nat Med. 9 (7):861-866.
- LI, H. e DURBIN, R. **Fast and accurate short read alignment with Burrows-Wheeler transform.** 2009. Bioinformatics. 25 (14):1754-1760.
- LI, H. *et al.* **The Sequence Alignment/Map format and SAMtools.** 2009. Bioinformatics. 25 (16):2078-2079.
- LI, J. Z. *et al.* **Factors associated with viral rebound in HIV-1-infected individuals enrolled in a therapeutic HIV-1 gag vaccine trial.** 2011. J Infect Dis. 203 (7):976-983.
- LIMA, T. H. A. *et al.* **HLA-A promoter, coding, and 3'UTR sequences in a Brazilian cohort, and their evolutionary aspects.** 2019. Hla. 93 (2-3):65-79.
- LIND, C. *et al.* **Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing.** 2010. Hum Immunol. 71 (10):1033-1042.
- LUO, M. *et al.* **For protection from HIV-1 infection, more might not be better: a systematic analysis of HIV Gag epitopes of two alleles associated with different outcomes of HIV-1 infection.** 2012. J Virol. 86 (2):1166-1180.
- LURIA, S. *et al.* **Expression of the type 1 human immunodeficiency virus Nef protein in T cells prevents antigen receptor-mediated induction of interleukin 2 mRNA.** 1991. Proc Natl Acad Sci U S A. 88 (12):5326-5330.
- MAJOR, E. *et al.* **HLA typing from 1000 genomes whole genome and whole exome illumina data.** 2013. PLoS One. 8 (11):e78410.
- MALIM, M. H. *et al.* **The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA.** 1989.

Nature. 338 (6212):254-257.

MALISSEN, M. *et al.* **Exon/intron organization and complete nucleotide sequence of an HLA gene.** 1982. Proc Natl Acad Sci U S A. 79 (3):893-897.

MARGOLIS, D. M. **Mechanisms of HIV latency: an emerging picture of complexity.** 2010. Curr HIV/AIDS Rep. 7 (1):37-43.

MARMOR, M. *et al.* **Resistance to HIV infection.** 2006. J Urban Health. 83 (1):5-17.

MARSH, S. G. **Nomenclature for factors of the HLA system, update September 2000. WHO Nomenclature Committee for Factors of the HLA System.** 2000. Tissue Antigens. 56 (6):565-566.

MARSH, S. G. *et al.* **Nomenclature for factors of the HLA system, 2010.** 2010. Tissue Antigens. 75 (4):291-455.

MASSIAH, M. A. *et al.* **Comparison of the NMR and X-ray structures of the HIV-1 matrix protein: evidence for conformational changes during viral assembly.** 1996. Protein Sci. 5 (12):2391-2398.

MATTHEWS, P. C. *et al.* **Differential clade-specific HLA-B*3501 association with HIV-1 disease outcome is linked to immunogenicity of a single Gag epitope.** 2012. J Virol. 86 (23):12643-12654.

MCDEVITT, H. O. **Discovering the role of the major histocompatibility complex in the immune response.** 2000. Annu Rev Immunol. 18 1-17.

MCKENNA, A. *et al.* **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.** 2010. Genome Res. 20 (9):1297-1303.

MCMICHAEL, A. J. e JONES, E. Y. **Genetics. First-class control of HIV-1.** 2010. Science. 330 (6010):1488-1490.

MEYER, D. *et al.* **A genomic perspective on HLA evolution.** 2018. Immunogenetics. 70 (1):5-27.

MIGUELES, S. A. *et al.* **Lytic granule loading of CD8+ T cells is required for HIV-infected cell elimination associated with immune control.** 2008. Immunity. 29 (6):1009-1021.

MIGUELES, S. A. *et al.* **HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors.** 2000. Proc Natl Acad Sci U S A. 97 (6):2709-2714.

MONOS, D. e MAIERS, M. J. **Progressing towards the complete and thorough characterization of the HLA genes by NGS (or single-molecule DNA sequencing): Consequences, opportunities and challenges.** 2015.

Hum Immunol. 76 (12):883-886.

MOORE, C. B. *et al.* **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level.** 2002. Science. 296 (5572):1439-1443.

MOTHE, B. e BRANDER, C. **HIV T-Cell Vaccines.** 2018. Adv Exp Med Biol. 1075 31-51.

MOTHE, B. *et al.* **Definition of the viral targets of protective HIV-1-specific T cell responses.** 2011. J Transl Med. 9 208.

MOTHE, B. *et al.* **CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control.** 2012. PLoS One. 7 (1):e29717.

MOTHE, B. *et al.* **Therapeutic Vaccination Refocuses T-cell Responses Towards Conserved Regions of HIV-1 in Early Treated Individuals (BCN 01 study).** 2019. EClinicalMedicine. 11 65-80.

MOUTAFTSI, M. *et al.* **A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus.** 2006. Nat Biotechnol. 24 (7):817-819.

MUELLER, S. M. *et al.* **Dual selection pressure by drugs and HLA class I-restricted immune responses on human immunodeficiency virus type 1 protease.** 2007. J Virol. 81 (6):2887-2898.

NEEFJES, J. *et al.* **Towards a systems understanding of MHC class I and MHC class II antigen presentation.** 2011. Nat Rev Immunol. 11 (12):823-836.

NORMAN, P. J. *et al.* **Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing.** 2016. Am J Hum Genet. 99 (2):375-391.

NOWAK, M. **HIV mutation rate.** 1990. Nature. 347 (6293):522.

NUNES, E. *et al.* **Definitions of histocompatibility typing terms.** 2011a. Blood. 118 (23):e180-183.

NUNES, J. M. *et al.* **Allele frequency estimation from ambiguous data: using resampling schema in validating frequency estimates and in selective neutrality testing.** 2011b. Hum Biol. 83 (3):437-447.

OLSON, A. D. *et al.* **An evaluation of HIV elite controller definitions within a large seroconverter cohort collaboration.** 2014. PLoS One. 9 (1):e86719.

OTTINGER, H. D. *et al.* **Hematopoietic stem cell transplantation: contrasting the outcome of transplantations from HLA-identical siblings, partially HLA-mismatched related donors, and HLA-matched unrelated**

- donors.** 2003. *Blood*. 102 (3):1131-1137.
- PAMER, E. e CRESSWELL, P. **Mechanisms of MHC class I--restricted antigen processing.** 1998. *Annu Rev Immunol*. 16 323-358.
- PANTALEO, G. e FAUCI, A. S. **Immunopathogenesis of HIV infection.** 1996. *Annual review of microbiology*. 50 825-854.
- PANTALEO, G. e LEVY, Y. **Therapeutic vaccines and immunological intervention in HIV infection: a paradigm change.** 2016. *Curr Opin HIV AIDS*. 11 (6):576-584.
- PAPUCHON, J. *et al.* **Kinetics of HIV-1 CTL epitopes recognized by HLA I alleles in HIV-infected individuals at times near primary infection: the Provir/Latitude45 study.** 2014. *PLoS One*. 9 (6):e100452.
- PAPUCHON, J. *et al.* **Resistance mutations and CTL epitopes in archived HIV-1 DNA of patients on antiviral treatment: toward a new concept of vaccine.** 2013. *PLoS One*. 8 (7):e69029.
- PARHAM, P. **Structure of class-I MHC molecules: HLA-B27 and disease.** 1990. *Scand J Rheumatol Suppl*. 87 11-20.
- PARHAM, P. **Immunogenetics of killer cell immunoglobulin-like receptors.** 2005. *Mol Immunol*. 42 (4):459-462.
- PARKIN, J. e COHEN, B. **An overview of the immune system.** 2001. *Lancet*. 357 (9270):1777-1789.
- PAXTON, W. *et al.* **Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis.** 1993. *J Virol*. 67 (12):7229-7237.
- PELAK, K. *et al.* **Host determinants of HIV-1 control in African Americans.** 2010. *J Infect Dis*. 201 (8):1141-1149.
- PENN, D. J. *et al.* **MHC heterozygosity confers a selective advantage against multiple-strain infections.** 2002. *Proc Natl Acad Sci U S A*. 99 (17):11260-11264.
- PEREYRA, F. *et al.* **The major genetic determinants of HIV-1 control affect HLA class I peptide presentation.** 2010. *Science*. 330 (6010):1551-1557.
- PERREAU, M. *et al.* **Targeted Immune Interventions for an HIV-1 Cure.** 2017. *Trends Mol Med*. 23 (10):945-961.
- PETRUCCI, A. *et al.* **How many HIV-infected individuals may be defined as long-term nonprogressors? A report from the Italian Seroconversion Study. Italian Seroconversion Study Group (ISS).** 1997. *J Acquir Immune Defic Syndr Hum Retrovirol*. 14 (3):243-248.

- POLLARD, R. B. *et al.* **Safety and efficacy of the peptide-based therapeutic vaccine for HIV-1, Vacc-4x: a phase 2 randomised, double-blind, placebo-controlled trial.** 2014. *The Lancet. Infectious diseases.* 14 (4):291-300.
- POMERANTZ, R. J. e HORN, D. L. **Twenty years of therapy for HIV-1 infection.** 2003. *Nat Med.* 9 (7):867-873.
- POONIA, B. *et al.* **Role of the Fas/FasL pathway in HIV or SIV disease.** 2009. *Retrovirology.* 6 91.
- PROFAIZER, T. e KUMANOVICS, A. **Human Leukocyte Antigen Typing by Next-Generation Sequencing.** 2018. *Clinics in laboratory medicine.* 38 (4):565-578.
- PROFAIZER, T. *et al.* **HLA genotyping in the clinical laboratory: comparison of next-generation sequencing methods.** 2016. *Hla.* 88 (1-2):14-24.
- PRUGNOLLE, F. *et al.* **Pathogen-driven selection and worldwide HLA class I diversity.** 2005. *Curr Biol.* 15 (11):1022-1027.
- PUMFERY, A. *et al.* **Chromatin remodeling and modification during HIV-1 Tat-activated transcription.** 2003. *Curr HIV Res.* 1 (3):343-362.
- RICHMAN, D. D. *et al.* **In vitro evaluation of experimental agents for anti-HIV activity.** 2001. *Current protocols in immunology.* Chapter 12 Unit 12 19.
- ROBINSON, J. T. *et al.* **Integrative genomics viewer.** 2011. *Nat Biotechnol.* 29 (1):24-26.
- RYOM, L. *et al.* **Use of antiretroviral therapy and risk of end-stage liver disease and hepatocellular carcinoma in HIV-positive persons.** 2016. *AIDS.* 30 (11):1731-1743.
- SABIN, C. A. e LUNDGREN, J. D. **The natural history of HIV infection.** 2013. *Curr Opin HIV AIDS.* 8 (4):311-317.
- SAKSENA, N. K. *et al.* **Elite HIV controllers: myth or reality?** 2007. *AIDS Rev.* 9 (4):195-207.
- SANCHEZ-MAZAS, A. *et al.* **Immunogenetics as a tool in anthropological studies.** 2011. *Immunology.* 133 (2):143-164.
- SANCHEZ-MAZAS, A. e MEYER, D. **The relevance of HLA sequencing in population genetics studies.** 2014. *Journal of immunology research.* 2014 971818.
- SANTORO, M. M. e PERNO, C. F. **HIV-1 Genetic Variability and Clinical Implications.** 2013. *ISRN microbiology.* 2013 481314.

SATTENTAU, Q. J. e MOORE, J. P. **The role of CD4 in HIV binding and entry.** 1993. *Philos Trans R Soc Lond B Biol Sci.* 342 (1299):59-66.

SAÚDE, M. D. **Recomendações para terapia antirretroviral em adultos infectados pelo HIV.** 2008.

SAÚDE, M. D. **Boletim Epidemiológico - Departamento de DST/Aids.** 2013.

SAÚDE, M. D. **Boletim Epidemiológico - HIV Aids.** 2018a.

SAÚDE, M. D. **Protocolo Clínico e Diretrizes Terapêuticas para Manejo da Infecção pelo HIV em Adultos.** 2018b.

SCARLATA, S. e CARTER, C. **Role of HIV-1 Gag domains in viral assembly.** 2003. *Biochim Biophys Acta.* 1614 (1):62-72.

SCHMITZ, J. L. **HLA Typing Using Molecular Methods.** 2006.

SCHOOLEY, R. T. *et al.* **AIDS clinical trials group 5197: a placebo-controlled trial of immunization of HIV-1-infected persons with a replication-deficient adenovirus type 5 vaccine expressing the HIV-1 core protein.** 2010. *J Infect Dis.* 202 (5):705-716.

SCHUBERT, U. *et al.* **The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains.** 1996. *J Virol.* 70 (2):809-819.

SCHWARTZ, O. *et al.* **Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein.** 1996. *Nat Med.* 2 (3):338-342.

SEDDIKI, N. e LEVY, Y. **Therapeutic HIV-1 vaccine: time for immunomodulation and combinatorial strategies.** 2018. *Curr Opin HIV AIDS.* 13 (2):119-127.

SETTE, A. e SIDNEY, J. **Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism.** 1999. *Immunogenetics.* 50 (3-4):201-212.

SHAFER, R. W. *et al.* **HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance.** 2007. *AIDS.* 21 (2):215-223.

SHIINA, T. *et al.* **The HLA genomic loci map: expression, interaction, diversity and disease.** 2009. *J Hum Genet.* 54 (1):15-39.

SIDNEY, J. *et al.* **HLA class I supertypes: a revised and updated classification.** 2008. *BMC Immunol.* 9 1.

SIJTS, E. J. e KLOETZEL, P. M. **The role of the proteasome in the**

generation of MHC class I ligands and immune responses. 2011. *Cell Mol Life Sci.* 68 (9):1491-1502.

SILICIANO, J. D. e SILICIANO, R. F. **Recent developments in the effort to cure HIV infection: going beyond N = 1.** 2016. *J Clin Invest.* 126 (2):409-414.

SIMON, V. e HO, D. D. **HIV-1 dynamics in vivo: implications for therapy.** 2003. *Nat Rev Microbiol.* 1 (3):181-190.

SINGH, P. *et al.* **Immunogenetic basis of HIV-1 infection, transmission and disease progression.** 2008. *Vaccine.* 26 (24):2966-2980.

SNELL, G. D. **Methods for the study of histocompatibility genes.** 1948. *Journal of genetics.* 49 (2):87-108.

SPIRA, A. I. *et al.* **Cellular targets of infection and route of viral dissemination after an intravaginal inoculation of simian immunodeficiency virus into rhesus macaques.** 1996. *J Exp Med.* 183 (1):215-225.

STEPHENS, H. A. **HIV-1 diversity versus HLA class I polymorphism.** 2005. *Trends Immunol.* 26 (1):41-47.

STEPHENS, H. A. **Immunogenetic surveillance of HIV/AIDS.** 2012. *Infect Genet Evol.* 12 (7):1481-1491.

STEPHENSON, K. E. **Therapeutic vaccination for HIV: hopes and challenges.** 2018. *Curr Opin HIV AIDS.* 13 (5):408-415.

STREBEL, K. *et al.* **The HIV 'A' (sor) gene product is essential for virus infectivity.** 1987. *Nature.* 328 (6132):728-730.

TANG, J. e KASLOW, R. A. **The impact of host genetics on HIV infection and disease progression in the era of highly active antiretroviral therapy.** 2003. *AIDS.* 17 Suppl 4 S51-60.

TAPIA, G. *et al.* **Sequential Vacc-4x and romidepsin during combination antiretroviral therapy (cART): Immune responses to Vacc-4x regions on p24 and changes in HIV reservoirs.** 2017. *J Infect.* 75 (6):555-571.

TEIXEIRA, S. L. *et al.* **Distribution of CCR5 genotypes and HLA Class I B alleles in HIV-1 infected and uninfected injecting drug users from Rio de Janeiro, Brazil.** 2009. *Infect Genet Evol.* 9 (4):638-642.

TRINCHIERI, G. **Biology of natural killer cells.** 1989. *Adv Immunol.* 47 187-376.

TUMIOTTO, C. *et al.* **Provir/Latitude 45 study: A step towards a multi-epitopic CTL vaccine designed on archived HIV-1 DNA and according to dominant HLA I alleles.** 2019. *PLoS One.* 14 (2):e0212347.

TUMIOTTO, C. *et al.* **Sanger and Next-Generation Sequencing data for characterization of CTL epitopes in archived HIV-1 proviral DNA.** 2017. PLoS One. 12 (9):e0185211.

UNAIDS **Country factsheets Brazil.** 2017.

UNAIDS **Global Aids Update - COMMUNITIES AT THE CENTRE.** 2019a.

UNAIDS **RELATÓRIO INFORMATIVO – ATUALIZAÇÃO GLOBAL DA AIDS.** 2019b.

VARMUS, H. **Regulation of HIV and HTLV gene expression.** 1988. Genes Dev. 2 (9):1055-1062.

VYAS, J. M. *et al.* **The known unknowns of antigen processing and presentation.** 2008. Nat Rev Immunol. 8 (8):607-618.

WAIN-HOBSON, S. *et al.* **Nucleotide sequence of the AIDS virus, LAV.** 1985. Cell. 40 (1):9-17.

WANG, C. *et al.* **High-throughput, high-fidelity HLA genotyping with deep sequencing.** 2012. Proc Natl Acad Sci U S A. 109 (22):8676-8681.

WEIMER, E. T. **Clinical validation of NGS technology for HLA: An early adopter's perspective.** 2016. Hum Immunol. 77 (10):820-823.

WITTIG, M. *et al.* **Development of a high-resolution NGS-based HLA-typing and analysis pipeline.** 2015. Nucleic Acids Res. 43 (11):e70.

WITTIG, M. *et al.* **High-Resolution HLA-Typing by Next-Generation Sequencing of Randomly Fragmented Target DNA.** 2018. Methods Mol Biol. 1802 63-88.

XIE, J. *et al.* **Beta 2-microglobulin as a negative regulator of the immune system: high concentrations of the protein inhibit in vitro generation of functional dendritic cells.** 2003. Blood. 101 (10):4005-4012.

YERLY, D. *et al.* **Increased cytotoxic T-lymphocyte epitope variant cross-recognition and functional avidity are associated with hepatitis C virus clearance.** 2008. J Virol. 82 (6):3147-3153.

ZHENG, N. N. e DANIELS, R. S. **Substitution of hiv type-1 non-B env genes in C2, a subtype B cassette system, results in functional chimeric viruses.** 2004. AIDS Res Hum Retroviruses. 20 (9):964-971.

ZINKERNAGEL, R. M. **Cellular immune recognition and the biological role of major transplantation antigens.** 1997. Biosci Rep. 17 (2):91-111.

ZIPETO, D. e BERETTA, A. **HLA-C and HIV-1: friends or foes?** 2012. Retrovirology. 9 39.

ZUNIGA, J. M. **State of HIV treatment: Results of the International Association of Physicians in AIDS Care surveys of HIV-positive patients and HIV-treating physicians in the United States.** 2006. *J Int Assoc Physicians AIDS Care (Chic)*. 5 (2):51-56.