Liã Bárbara Arruda

**Caracterização molecular da gp120 do HIV-1 e suas implicações sobre
o tropismo pelos correceptores CCR5 e CXCR4**

Tese apresentada ao Instituto de Medicina Tropical de
São Paulo da Universidade de São Paulo para
obtenção do título de Doutora em Ciências.

Área de concentração: Doenças Tropicais e Saúde
Internacional

Orientador: Prof. Dr. Jorge Simão do Rosário Casseb

Versão corrigida

São Paulo
2014

Ficha catalográfica
Preparada pela Biblioteca do Instituto de Medicina Tropical de São Paulo da
Universidade de São Paulo
© Reprodução autorizada pelo autor

Arruda, Liã Bárbara

Caracterização molecular da gp120 do HIV-1 e suas implicações sobre o tropismo pelos correceptores CCR5 e CXCR4 / Liã Bárbara Arruda. – São Paulo, 2014.

Tese (Doutorado) – Instituto de Medicina Tropical de São Paulo da Universidade de São Paulo para obtenção do título de Doutor em Ciências.

Área de concentração: Doenças Tropicais e Saúde Internacional
Orientador: Jorge Simão do Rosário Casseb

Descritores: 1. HIV. 2. TROPISMO. 3. BIOINFORMÁTICA.

USP/IMTSP/BIB-02/2014.

A todo e qualquer herdeiro deste legado pio.

AGRADECIMENTOS

À minha família, meu agradecimento primordial. Minha mãe Margrit, meu pai Arruda e minha *Oma* Hildegard (*in memoriam*) são os maiores responsáveis pelas minhas conquistas longe da pequena Rio do Sul (SC), graças ao seu apoio, incentivo e confiança incondicionais.

Ao Programa de Pós-graduação em Medicina Tropical do Instituto de Medicina Tropical de São Paulo da Universidade de São Paulo pela oportunidade de realizar o Doutorado.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por prover a bolsa de doutorado (2010/00222-5) e o financiamento à pesquisa (2011/17881-4).

Aos pacientes que concordaram em contribuir com esse estudo voluntariamente.

Ao Prof. Dr. Alberto Duarte responsável pela idealização do que hoje é o Laboratório de Investigação Médica em Dermatologia e Imunodeficiências – LIM56, que foi minha segunda casa durante o Mestrado e o Doutorado.

Ao Prof. Dr. Jorge Casseb pela orientação.

Ao Dr. Sabri Sanabani e à Dra. Shirley Komninakis pelas críticas e sugestões durante o exame de qualificação que contribuíram para enriquecer meu trabalho.

À Dra. Camila Romano, minha coorientadora honorária, que transcendeu sua contribuição para fora do exame de qualificação. Sou imensamente grata por seu incansável auxílio, por sua inestimável paciência e principalmente pela amizade que consolidamos.

À Dra. Kristel Van Laethem por me receber tão solícitamente no Laboratório de Virologia Clínica e Evolutiva do Instituto Rega da Universidade Católica de Leuven, Bélgica para a realização do treinamento em *Bayesian Network Learning*. Agradeço profundamente a Andrea Piñeda, Gertjan Beheydt e Ricardo Khouri pelos ensinamentos, apoio e companheirismo constantes durante meus quatro meses belgas.

Aos meus filhos científicos Priscila Benevides, Fabrícia Madia e Bruno Aguiar que me proporcionaram tanto aprendizado quanto orgulho. Obrigada pela confiança nesta aspirante a orientadora.

Ao Rafael Martins que incitou a minha curiosidade por bioinformática, ao André Luiz de Oliveira que sempre me socorre solucionando minhas dúvidas nesta área e ao Marco Túlio Pires por me guiar no misterioso mundo da programação.

Ao Roberto Ribeiro pela enorme ajuda no sequenciamento pela plataforma *Ion Torrent*. Ao Thiago de Araújo e à Michele Gouvêa pelas corridas no sequenciador ABI3500 do LIM07.

A Noemia Orie e Rosângela Araújo pela realização dos exames de contagem de células TCD4+ e TCD8+. Ao Eduardo Martins e Dona Ideones pela realização dos exames de carga viral plasmática do HIV.

À Edna Reis pela inestimável ajuda com a burocracia da FAPESP e toda ajuda extra, essencial e decisiva.

Aos pesquisadores Dra. Maria Sato, Dr. Gil Benard, Dr. Dewton Vasconcelos e Dra. Alessandra Pontillo pela confiança e incentivo.

À Ana Carolina Soares por todo o apoio moral, pela revisão deste trabalho e amizade impagável.

Às minhas eternas amigas, Claudia Wyrepkowski e Paula Pastório, que mesmo a 4000 km de distância sempre estiveram presentes e compartilhando a árdua jornada da pós-graduação.

A Léia Rodrigues, Paula Rigato, Vanessa Batista, Camila Cácere, Adriana Ladeira e Ju Ruiz pela dedicação e confiança em nossa amizade nesses anos de LIM56.

Aos amigos de *brainstorming* regado à cerveja pelas constantes trocas de ideias e experiências na alegria e na tristeza, com bolsa ou sem bolsa: Luiz Nali, Felipe Scassi, Valdir Blasios Jr., Alexandre Bisson Filho, Luciane Schons, Érika Stein e outros amigos dos amigos dos amigos.

A Thiago Bezerra, Graziele Santana, Susana Lessa e demais colegas e colaboradores do LIM56 por toda a contribuição direta e indireta.

“Toda descoberta da ciência pura é potencialmente subversiva: até a própria ciência deve, às vezes, ser tratada como um inimigo possível. Sim, a própria ciência.”

Aldous Huxley *in* Admirável Mundo Novo

RESUMO

Arruda LB. Caracterização molecular da gp120 do HIV-1 e suas implicações sobre o tropismo pelos correceptores CCR5 e CXCR4 (tese). São Paulo: Instituto de Medicina Tropical de São Paulo da Universidade de São Paulo; 2014.

A descoberta do tropismo do Vírus da Imunodeficiência Humana tipo 1 (HIV-1) pelos correceptores CCR5 e CXCR4 propiciou o desenvolvimento de novas estratégias terapêuticas com alvo nestes correceptores. Os antagonistas de CCR5 são fármacos que bloqueiam o correceptor CCR5, impedindo a entrada do (HIV) na célula, levando a uma diminuição significativa da viremia. Porém, a administração deste medicamento depende da determinação do tropismo dos vírus do indivíduo infectado antes e ao longo do uso desta estratégia terapêutica. A alça V3 da gp120 do envelope do HIV-1 é considerada o principal determinante do tropismo, mas outros fatores como diferenças intrínsecas entre as variantes do HIV-1, o número de sítios de N-glicosilação, variações no comprimento de determinadas regiões e a carga elétrica global da gp120 também podem afetar o tropismo viral. Este estudo apresenta a caracterização molecular da gp120 e a relação dos fatores moleculares com o tropismo pelos correceptores CCR5 e CXCR4. Foram incluídos no estudo 283 indivíduos infectados pelo HIV-1, dos quais foi possível obter 265 sequências da alça V3, resultados de tropismo fenotípico para 78 amostras e 20 sequências da gp120 completa. O tropismo fenotípico demonstrou a prevalência de vírus R5 (70,5%) e a presença de apenas uma amostra contendo exclusivamente vírus X4. O tropismo genotípico classificou 71,7% das sequências como vírus R5 e 28,2% como capazes de utilizar o correceptor CXCR4. A concordância entre os resultados de tropismo fenotípico e genotípico foi de 74,5%. Houve a prevalência do subtipo B (82%), sendo que a variante B'(GWGR) esteve presente em 34,5% dos casos. Em média, as sequências da gp120 apresentaram 22 sítios de N-glicosilação, concentrados nas regiões C2 e C4. A carga elétrica líquida apresentou médias semelhantes entre vírus R5 e X4 tanto para a alça V3 quanto para a gp120. As características moleculares descritas neste estudo apresentaram distribuição semelhante entre vírus R5 e X4, não sendo possível estabelecer relações entre estes fatores a determinação do tropismo.

Descritores: HIV-1. Tropismo. Bioinformática.

ABSTRACT

Arruda LB. Molecular characterization of HIV-1-gp120 and its implications on coreceptor tropism (thesis). São Paulo: Instituto de Medicina Tropical de São Paulo da Universidade de São Paulo; 2014.

The tropism of Human Immunodeficiency Virus type 1 (HIV - 1) for CCR5 and CXCR4 coreceptors has provided the development of new therapeutic strategies targeting these coreceptors. CCR5 antagonists are able to prevent the HIV cell entry by blocking the CCR5 coreceptor, improving the viremia decreasing. However, this drug administration depends on viral tropism determination before and during the use of this therapeutic strategy. The V3 loop of the HIV-1 gp120 envelope is the major tropism determinant, but other factors such HIV-1 genetic variability, number of potential N-linked glycosylation sites, length variations in some protein domains and the gp120 global net charge may also affect the viral tropism. This study presents the molecular characterization of gp120 and the relationship between molecular factors and the tropism for CCR5 and CXCR4 coreceptors. A total of 283 HIV-1 infected subjects were included in this study. It was possible to obtain 265 V3 loop sequences, 78 results of phenotypic tropism and 20 sequences of the whole gp120. Phenotypic tropism showed the prevalence of R5 viruses (70.5%) while exclusively X4 viruses were found in only one sample. Genotypic tropism classified 71.7% of the sequences as R5 and 28.2% as virus able to use the CXCR4 coreceptor. The agreement between phenotypic and genotypic tropism results was 74.5%. There was a prevalence of subtype B (82%), and the B' (GWGR) variant was present in 34.5% of the cases. gp120 showed a mean of 22 N-linked glycosylation sites, which were more frequent in the C2 and C4 regions. The net charge showed similar means between R5 and X4 viruses for both V3 loop and gp120. The molecular characteristics described in this study showed similar distribution between R5 and X4 viruses and it was not possible to establish relationships between these factors the tropism determination.

Descriptors: HIV-1. Tropism. Bioinformatics.

LISTA DE FIGURAS

Figura 1 – Diagrama da estrutura da gp120 do envelope do HIV-1	24
Figura 2 – Estrutura esquemática da alça V3 da gp120.	26
Figura 3 – Visão geral da entrada do HIV na célula hospedeira..	28
Figura 4 – Representação esquemática dos estágios da infecção pelo HIV-1.....	29
Figura 5 – Visão geral do preparo de biblioteca por fragmentação da plataforma <i>Ion Torrent</i>	49
Figura 6 – Estrutura esquemática do chip semicondutor da plataforma <i>Ion Torrent</i>	51
Figura 7 – Esquema ilustrativo da reação de sequenciamento da plataforma <i>Ion Torrent</i>	52
Figura 8 – Exemplo hipotético de uma rede Bayesiana (BN).....	58
Figura 9 – Representação gráfica dos subconjuntos de amostras utilizados neste estudo..	61
Figura 10 – Distribuição do tropismo fenotípico (<i>Trofile</i>).....	62
Figura 11 – Distribuição do tropismo genotípico (<i>Geno2pheno</i>)	63
Figura 12 – Resultados de tropismo fenotípico (<i>Trofile</i>) e genotípico (<i>Geno2pheno</i>) das amostras do Conjunto B	64
Figura 13 - Resultados de tropismo fenotípico (<i>Trofile</i>) e genotípico (<i>Geno2pheno</i>) das amostras do Conjunto C	65
Figura 14 – Distribuição dos subtipos de HIV-1 na região V3 do Conjunto A	66
Figura 15 – Árvore filogenética das sequências da alça V3 obtidas por sequenciamento clássico e massivo paralelo	68
Figura 16 – Distribuição dos subtipos de HIV-1 na região da gp120 do Conjunto C	69
Figura 17 – Número de sítios potenciais de N-glicosilação preditos para sequências provenientes de vírus R5 e R5X4/X4.....	70
Figura 18 – Carga elétrica líquida para sequências provenientes de vírus R5 e R5X4/X4.....	70

LISTA DE TABELAS

Tabela 1 – Sequências dos <i>primers</i> utilizados para amplificar e sequenciar a região V3 da gp120	43
Tabela 2 – Amostras selecionadas para amplificação da gp120 do gene <i>env</i>	46
Tabela 3 – Sequências dos <i>primers</i> utilizados para amplificar e sequenciar a região da gp120 do gene <i>env</i>	47
Tabela 4 – Dados clínicos e epidemiológicos dos indivíduos incluídos neste estudo	59

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

ADEE 3002	Ambulatório de Imunodeficiências Secundárias do Departamento de Dermatologia – HCFMUSP
ART	Terapia antirretroviral (<i>Antiretroviral Therapy</i>)
BN	Redes Bayesianas (<i>Bayesian Network</i>)
C1	Região conservada 1 da gp120
C2	Região conservada 2 da gp120
C3	Região conservada 3 da gp120
C4	Região conservada 4 da gp120
C5	Região conservada 5 da gp120
CAPPesq	Comissão de Ética para Análise de Projetos de Pesquisa
CCL3	C-C ligante 3, também conhecido como MIP-1 α
CCL4	C-C ligante 4, também conhecido como MIP-1 β
CCL5	C-C ligante 5, também conhecido como RANTES
CCR5	C-C receptor de quimicina 5
CCR5 Δ 32	Deleção de 32 pares de base do CCR5
CD4	Agrupamento de diferenciação 4 (<i>Cluster of Differentiation</i>)
CEP	Comitê de Ética em Pesquisa
CRFs	Formas recombinantes circulantes (<i>Circulating Recombinant Forms</i>)
CXCR4	C-X-C receptor de quimiocina 4
DNA	Ácido desoxirribonucleico (<i>Deoxyribonucleic Acid</i>)
dNTPs	Desoxirribonucleosídeos 5'-trifosfatados (dATP, dTTP, dCTP, dGTP) (<i>Deoxynucleoside Triphosphates</i>)

EDTA	Ácido etilenodiaminotetracético (<i>Ethylenediaminetetraacetic Acid</i>)
EI	Código identificador das amostras deste estudo
ESTA	Ensaio <i>Trofile</i> com sensibilidade melhorada (<i>Enhanced Sensivity Trofile Assay</i>)
EUA	Estados Unidos da América
FPR	Razão de falso positivo (<i>False Positive Rate</i>)
G2P	Geno2pheno _[coreceptor]
gp	Glicoproteína
GTT	Teste genotípico de tropismo (<i>Genotypic Tropism Test</i>)
HCFMUSP	Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo
HIV-1	Vírus da Imunodeficiência Humana tipo 1 (<i>Human Immunodeficiency Virus type 1</i>)
HXB2	Variante clonal de referência para sequências de HIV-1
IPS	partículas esféricas do Ion Torrent (<i>Ion Sphere Particle</i>)
Kb	Quilobases (10^3 pares de base)
MIP-1 α	Proteína inflamatória de macrófago tipo 1 α (<i>Macrophage Inflammatory Protein 1α</i>)
MIP-1 β	Proteína inflamatória de macrófago tipo 1 β (<i>Macrophage Inflammatory Protein 1β</i>)
MgCl ₂	Cloreto de magnésio
μ L	Microlitro (10^{-6} litro)
mL	Mililitro (10^{-3} litro)
mM	Milimolar (10^{-3} molar)
μ M	Micromolar (10^{-6} molar)
min	Minuto

MVQ	Maraviroque
NaOH	Hidróxido de sódio
ng	Nanograma (10^{-9} grama)
NN	Redes neurais (<i>Neural Network</i>)
OBT	Terapia otimizada de base (<i>Optimized Background Therapy</i>)
OTA	Ensaio <i>Trofile</i> original (<i>Original Trofile Assay</i>)
pb	Pares de bases
PBMCs	Células mononucleares do sangue periférico (<i>Peripheral Blood Mononuclear Cells</i>)
PCR	Reação em cadeia da polimerase (<i>Polymerase Chain Reaction</i>)
PNGS	Sítios potenciais de N-glicosilação (<i>Potential N-linked Glycosilation Sites</i>)
PTT	Teste fenotípico de tropismo (<i>Phenotypic Tropism Test</i>)
PSSM	Matrizes de pontuação de posições específicas (<i>Position-Specific Score Matrices</i>)
R5	Vírus com tropismo pelo correceptor CCR5
R5X4	Vírus com tropismo pelos correceptores CCR5 e CXCR4
RANTES	Regulado sobre a ativação de célula T expressa e secretada (<i>Regulated upon Activation T cell Expressed and Secreted</i>)
RNA	Ácido ribonucleico (<i>Ribonucleic Acid</i>)
RT-PCR	Reação de transcrição reversa seguida de reação em cadeia da polimerase (<i>Reverse Transcriptase - Polymerase Chain Reaction</i>)
RVA	Ensaio de vírus recombinantes (<i>Recombinant Virus Assay</i>)
s	Segundo

SIV	Vírus da Imunodeficiência Símia (<i>Simian Immunodeficiency Virus</i>)
SMV	Máquina de suporte de vetores (<i>Support Vector Machine</i>)
SU	Glicoproteína de superfície
TCLE	Termo de Consentimento Livre e Esclarecido
TM	Glicoproteína transmembrana
U	Unidade
UDIV	Usuário de drogas intravenosas
URF	Formas recombinantes únicas (<i>Unique Recombinant Forms</i>)
UV	Ultravioleta
V1	Região variável 1 da gp120
V2	Região variável 2 da gp120
V3	Região variável 3 da gp120
V4	Região variável 4 da gp120
V5	Região variável 5 da gp120
X4	Vírus com tropismo pelo correceptor CXCR4
xg	Gravidade
°C	Graus Celsius

Aminoácidos

A (Ala)	Alanina
C (Cys)	Cisteína
D (Asp)	Aspartato ou ácido aspártico
E (Glu)	Glutamato ou ácido glutâmico
F (Phe)	Fenilalanina
G (Gly)	Glicina
H (His)	Histidina

I (Ile)	Isoleucina
K (Lys)	Lisina
L (Leu)	Leucina
M (Met)	Metionina
N (Asn)	Asparagina
P (Pro)	Prolina
Q (Gln)	Glutamina
R (Arg)	Arginina
S (Ser)	Serina
T (Thr)	Treonina
V (Val)	Valina
W (Trp)	Triptofano
Y (Tyr)	Tirosina

Nucleotídeos

A	Adenina
C	Citosina
G	Guanina
T	Timina

SUMÁRIO

1 INTRODUÇÃO.....	20
1.1 O envelope do HIV-1	23
1.1.1 Estrutura da gp120	23
1.2 O tropismo do HIV-1	26
1.2.1 Fatores virais associados ao tropismo do HIV-1	30
1.3 Antagonistas de correceptores	33
1.4 Testes para determinação do tropismo	35
2 OBJETIVOS	39
2.1 Objetivo geral.....	39
2.2 Objetivos específicos	39
3 MATERIAL E MÉTODOS	40
3.1 Casuística	40
3.1.1 Critérios de inclusão	40
3.1.2 Critérios de exclusão	40
3.1.3 Aspectos éticos.....	40
3.2 Coleta das amostras	41
3.3 Separação das células mononucleares do sangue periférico.....	41
3.4 Extração de DNA	42
3.5 Amplificação e sequenciamento da região V3 da gp120.....	43
3.5.1 Amplificação da região V3	43
3.5.2 Reação de sequenciamento da região V3	44
3.5.3 Análise do sequenciamento	45
3.6 Amplificação e sequenciamento da região da gp120 do gene <i>env</i>	45
3.6.1 Amplificação da região da gp120	46

3.6.2 Reação de sequenciamento da região da gp120.....	48
3.6.2.1 Preparo da biblioteca.....	48
3.6.2.2 Preparo de <i>template</i>	50
3.6.2.3 Sequenciamento.....	51
3.6.3 Análise do sequenciamento.....	52
3.7 Testes de tropismo	54
3.7.1 Teste fenotípico para determinação de tropismo	54
3.7.2 Teste genotípico para predição de tropismo	55
3.8 Determinação dos subtipos	55
3.9 Predição de sítios potenciais de N-glicosilação	56
3.10 Determinação da carga elétrica líquida.....	56
3.11 <i>Bayesian Network</i>	57
3.12 Análise estatística.....	58
4 RESULTADOS.....	59
4.1 Dados demográficos e laboratoriais	59
4.2 Conjuntos de amostras.....	60
4.3 Avaliação dos testes de tropismo	62
4.4 Caracterização molecular da gp120	65
4.4.1 Avaliação das sequências da região V3	65
4.4.2 Avaliação das sequências da região da gp120.....	66
4.5 <i>Bayesian Network</i>	71
5 DISCUSSÃO	73
5.1 Avaliação dos testes de tropismo	73
5.2 Distribuição dos subtipos de HIV-1	76
5.3 Avaliação do sequenciamento da região da gp120	77
5.4 <i>Bayesian Network</i>	79

5.5 Considerações finais.....	80
6 CONCLUSÕES.....	83
REFERÊNCIAS	84
APÊNDICES	97
Apêndice A – Termo de Consentimento Livre e Esclarecido.....	97
Apêndice B – <i>Script</i> de programação em <i>Python</i> para cálculo da carga elétrica líquida de uma sequência de aminoácidos	100
Apêndice C – Artigo científico aceito para publicação na Revista do Instituto de Medicina Tropical de São Paulo	100
Apêndice D – Resumo do trabalho apresentado oralmente na <i>4ta. Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio)</i>	105
ANEXOS	108
Anexo A – Certificado emitido pela Comissão de Ética.....	108
Anexo B – Aprovação da alteração de título	109
Anexo C – Carta de aceite para publicação de artigo na Revista do Instituto de Medicina Tropical de São Paulo	110
Anexo D – Certificado de apresentação oral na <i>4ta. Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio)</i>	111
Anexo E – Relatórios das corridas de sequenciamento da plataforma <i>Ion Torrent</i> (Life Technologies)	112
Anexo F - Predição de sítios de N-glicosilação das sequências geradas por sequenciamento massivo paralelo	116

1 INTRODUÇÃO

Uma das principais características do Vírus da Imunodeficiência Humana tipo 1 (HIV-1 – *Human Immunodeficiency Virus type 1*) é sua alta taxa de variabilidade genética, demonstrada pelo grande número de variantes que circulam em um mesmo indivíduo (Peeters, 2000; Smyth et al., 2012). Esta marcante diversidade genética é resultante não só de fatores virais, mas também, da pressão exercida pelo sistema imune do hospedeiro (Peeters, 2000; Simon-Loriere et al., 2009). A variabilidade do genoma do HIV-1 envolve um lento acúmulo de mutações pontuais (inserções e deleções) produzidas pela ausência de atividade de exonuclease da transcriptase reversa e da alta capacidade recombinogênica desta polimerase. A recombinação é um importante mecanismo evolutivo que contribui para a diversificação genética permitindo grandes alterações genômicas (Simon-Loriere et al., 2009).

Em virtude desta flexibilidade genética, o HIV-1 pode ser dividido em grupos, subtipos e sub-subtipos além de formas recombinantes circulantes (CRFs – *circulating recombinant forms*). Os grupos M (*Main*), O (*Outlier*) e N (*Non-M/Non-O*) resultam de transmissões interespecíficas independentes do SIVcpzPtt (vírus da imunodeficiência símia – *Simian Immunodeficiency Virus*) dos chimpanzés *Pan troglodytes troglodytes* para os humanos, enquanto o recente grupo P teve origem a partir da transmissão de SIVgor da espécie *Gorilla gorilla gorilla* para humanos (Keele et al., 2006; Plantier et al., 2009). A maior parte dos isolados pandêmicos de HIV-1 pertence ao grupo M que é classificado em nove subtipos: A, B, C, D, F, G, H, J e K. Existem ainda sub-subtipos (A1, A2; F1, F2, etc.) e mais de 60 CRFs descritas atualmente (Brun-Vezinet e Charpentier, 2013; Robertson et al., 2000).

A diversidade genética viral também é capaz de influenciar o seu tropismo. A afinidade de um vírus, ou tropismo, é definida pelos receptores da superfície celular de uma determinada espécie, cujas células são capazes de suportar uma infecção produtiva por aquele vírus. O tropismo do HIV-1 é bastante específico não somente em número de espécies, mas também em número de células, em uma espécie particular, que são permissivas ao vírus (Cullen, 2001). O receptor primário CD4 é essencial para o reconhecimento da célula hospedeira, enquanto os receptores de quimiocinas CCR5 e CXCR4 representam os principais correceptores *in vivo* para a entrada do HIV-1 sendo que a alça V3 da gp120 do envelope é considerada o principal determinante para a especificidade pelo correceptor (Hughes e Nelson, 2009; Sander et al., 2007).

Seguindo a história natural da infecção pelo HIV-1, durante a fase primária há prevalência de variantes CCR5-trópicas (R5) em cerca de 50% dos pacientes. Os vírus CXCR4-trópicos (X4) surgem ao longo da infecção caracterizando a progressão da doença, e há ainda a presença de variantes duplo-trópicas (R5X4) que surgem durante a transição de R5 para X4 (Jensen et al., 2003; Mild et al., 2009; Sander et al., 2007). É possível que a população viral seja modelada por fatores do hospedeiro e do sistema imune no início da infecção resultando em populações de vírus R5 que sejam mais propensos a manter o uso exclusivo de CCR5 ou desenvolvam o uso de CXCR4 ao longo do curso da infecção. Isto pode explicar porque em apenas alguns indivíduos a população viral desenvolve o tropismo por CXCR4 (Mild et al., 2009).

A descoberta dos receptores de quimiocinas CCR5 e CXCR4 e seu papel como correceptores essenciais para a entrada viral provocaram uma década de pesquisas sobre os mecanismos de entrada e possíveis alvos para tratamento e prevenção do HIV (Hughes e Nelson, 2009). O Maraviroque (MVQ) foi o primeiro antagonista do correceptor CCR5, aprovado pra uso clínico em 2007. No Brasil, o

MVQ é recomendado para situações em que, mesmo com a utilização de inibidores de protease ou inibidores de integrase, a chance de obtenção da supressão viral seja considerada baixa. Além disso, só poderá ser indicado se houver teste de tropismo viral evidenciando presença exclusiva de vírus R5, realizado no máximo seis meses antes do início do novo esquema (Brasil, 2013).

A determinação do tropismo do HIV-1 geralmente é realizada através de testes fenotípicos de tropismo (PTT – *Phenotypic Tropism Test*), ensaios que utilizam vírus recombinantes que infectam culturas celulares específicas (Low et al., 2007; Whitcomb et al., 2007). Os testes genotípicos de tropismo (GTT – *Genotypic Tropism Test*) se baseiam na sequência da alça V3 para realizar a predição do uso de correceptor utilizando de ferramentas de bioinformática e representam uma alternativa mais prática e eficaz em termos de custos (Low et al., 2007; Trabaud et al., 2012).

Apesar da determinação do tropismo do HIV-1 estar centrada na alça V3, na gp120 podem ser encontrados outros fatores associados ao tropismo, além destes 35 aminoácidos. A variabilidade genética do HIV-1 leva a diferenças intrínsecas entre os subtipos e a algumas alterações em posições específicas da gp120 que podem afetar o tropismo do HIV-1 pelo correceptor, como substituições que envolvem os sítios potenciais de N-glicosilação (PNGS – *Potential N-linked Glycosylation Sites*), inserções e deleções que resultam na variação do comprimento das regiões variáveis 1, 2 e 4 (V1/V2 e V4) e substituições capazes de alterar a carga elétrica líquida da gp120 (Hughes e Nelson, 2009; Mild et al., 2009). Assim, a avaliação de toda a extensão da gp120 pode trazer informações mais específicas sobre a prevalência e a alteração do tropismo do HIV-1 no curso da infecção.

1.1 O envelope do HIV-1

As glicoproteínas do envelope (Env) são sintetizadas como uma poliproteína precursora (gp160) no retículo endoplasmático da célula hospedeira que é codificada pelo RNAm do gene *env* viral. O Env é modificado pós-traducionalmente no retículo endoplasmático e complexo de Golgi e é clivado por proteases celulares para produzir uma glicoproteína de superfície (SU) gp120 e uma glicoproteína transmembrana (TM) gp41 (Rizzuto et al., 1998; Turner e Summers, 1999). Dentro do retículo, estas glicoproteínas são oligomerizadas em estruturas triméricas e altamente glicosiladas, gerando um complexo gp120/gp41 associado não covalentemente. Em seguida, este trímero de glicoproteínas é transportado para a membrana e, durante a montagem da partícula viral, o complexo gp120/gp41 é incorporado como espículas heterotriméricas altamente glicosiladas na bicamada lipídica dos vírions nascentes. (Checkley et al., 2011; Rizzuto et al., 1998; Turner e Summers, 1999).

1.1.1 Estrutura da gp120

A comparação do gene *env* de diversos isolados de HIV-1 identificou um alto grau de variabilidade em segmentos descontínuos da subunidade de superfície. A gp120 apresenta cinco domínios hipervariáveis (V1 – V5) inter espaçados por cinco domínios relativamente conservados (C1 – C5) (Figura 1) (Checkley et al., 2011; Willey et al., 1986; Zolla-Pazner, 2004). Tipicamente, a gp120 contém 18 resíduos de cisteína, altamente conservados, que formam nove pontes dissulfeto, essenciais para a conformação da estrutura terciária do Env (Checkley et al., 2011; Leonard et

al., 1990). A V1 é separada da V2 por duas pontes dissulfeto, assim como as alças V3 e V4 também são delimitadas por resíduos de cisteína (Leonard et al., 1990).

Figura 1 – Diagrama da estrutura da gp120 do envelope do HIV-1. As caixas demonstram as posições das regiões hipervariáveis V1 a V5. Os sítios de N-glicosilação estão representados por estruturas de ramos. Adaptado de Zolla-Pazner (2004).

Aproximadamente metade da massa molecular da gp120 é composta por N-glicosilações com uma pequena contribuição adicional de O-glicosilações (Checkley et al., 2011; Ho et al., 2008). Existem de 20 a 35 sítios de N-glicosilações na gp120 que mascaram o reconhecimento do Env pelo sistema imunológico do hospedeiro, contribuindo com a ligação do vírus à célula alvo (Wei et al., 2003; Zhang et al., 2004; Zolla-Pazner, 2004).

A variabilidade genética dos domínios variáveis é resultante de recombinações, mutações pontuais, inserções e deleções. O domínio V1/V2 é a alça com maior variação de comprimento e número de sítios de N-glicosilação (Palmer et al., 1996; Sagar et al., 2006; Shioda et al., 1997). A alça V1/V2 varia e 50 a 90 aminoácidos de comprimento, enquanto a variação de comprimento das alças V4 e V5 são de 19 a 44 e 14 a 36 aminoácidos, respectivamente. A alça V3 e os domínios C2, C3 e C4 apresentam pouca variação no comprimento. Adicionalmente, existe uma correlação entre o aumento do comprimento e o número de sítios de glicosilação na V1/V2 e a progressão da doença, sugerindo que esta modificação de comprimento pode promover o escape da resposta imunológica (Checkley et al., 2011; Curlin et al., 2010). Os domínios conservados da gp120, particularmente C1, C3 e C4 são os principais determinantes para a ligação com o CD4 (Kowalski et al., 1987; Olshevsky et al., 1990). Vários estudos demonstraram que a alça V3 é importante para a fusão das membranas e especificidade do correceptor e que contém epítomos para anticorpos neutralizantes (Checkley et al., 2011; Freed et al., 1991; Hwang et al., 1991).

Com relação à sequência genética, a V3 é a mais conservada em comparação com as demais regiões variáveis da gp120. A estrutura da V3 pode ser dividida em três regiões: base, corpo e ponta. A base conservada está acomodada no núcleo da gp120 com uma ponte dissulfeto, enquanto a haste se estende para fora e é bastante flexível. A ponta consiste em aproximadamente 13 resíduos no centro da sequência da V3, formando uma conformação β no ápice distal (Figura 2) (Huang et al., 2005; Jiang et al., 2010; Sander et al., 2007). O comprimento da V3 sofre pouca variação, apresentando na maioria dos casos 35 aminoácidos, e exhibe na ponta da alça um motivo glicina-prolina-glicina-arginina/glutamina (GPGR/Q), altamente conservado, conhecido como coroa da alça V3 (Jiang et al., 2010; LaRosa et al., 1990).

Figura 2 – Estrutura esquemática da alça V3 da gp120. A base da V3 está acomodada no núcleo da gp120, delimitada pelas pontes dissulfeto (Cys1-Cys35). A haste flexível se estende para fora da gp120 e termina na ponta que apresenta uma estrutura conservada GPG, denominada coroa. Adaptado de Sander et al. (2007).

1.2 O tropismo do HIV-1

Desde a descrição do primeiro caso da Síndrome da Imunodeficiência Adquirida (aids) em 1981, observou-se que os pacientes acometidos apresentavam uma profunda diminuição na contagem de células T CD4⁺ (Gottlieb et al., 1981). Em 1986, foi possível confirmar que a molécula de superfície celular CD4 é o receptor primário para a entrada do HIV-1 na célula hospedeira, porém ainda era necessário algum outro fator para que o vírus fosse capaz de infectar a célula (Iwamoto et al., 2010; Maddon et al., 1986). O correceptor necessário permaneceu como um enigma por 10 anos até que, em 1996, uma proteína com sete domínios transmembrana acoplados à proteína G, originalmente chamada de fusina e atualmente conhecida como o receptor de quimiocinas CXCR4, foi descrita como um correceptor do HIV-1 em linhagens de células T (Feng et al., 1996; Iwamoto et al., 2010). Mais tarde, observou-se que isolados primários apresentavam infecções

produtivas tanto em linhagens de células T CD4⁺ (vírus T-trópicos) quanto em monócitos derivados de macrófagos (vírus M-trópicos) (Collman et al., 1989). A descoberta de que as β -quimiocinas RANTES, MIP-1 α e MIP-1 β atuavam como agentes supressores da infecção por HIV-1 levou a identificação do receptor de quimiocina CCR5 como o correceptor para os vírus M-trópicos (Deng et al., 1996). Assim, foi proposta uma nova classificação para HIV-1, denominando vírus R5 para os isolados que podem utilizar o CCR5 como correceptor, mas não o CXCR4, vírus X4 para aqueles que utilizam somente o CXCR4 e vírus R5X4 ou duplo-trópicos para aqueles capazes de utilizar ambos os correceptores (Berger et al., 1998; Iwamoto et al., 2010).

O sítio de alta afinidade ao correceptor permanece interiorizado na gp120 e isolado do sistema imune do hospedeiro (Rizzuto et al., 1998). Após a ligação CD4-vírus e o reconhecimento do correceptor, ocorre uma mudança conformacional na gp120 expondo a alça V3, cuja estrutura de β -*hairpin* atua como um gancho no correceptor (Figura 3) (Iwamoto et al., 2010; Wilen et al., 2012). A interação entre a alça V3 e o correceptor implica no melhor ancoramento entre o CD4 celular e a gp120. Esse processo leva à exposição da gp41, até então inacessível, promovendo a fusão das membranas. Acredita-se que a gp41 assuma uma conformação de seis hélices aproximando as membranas o suficiente para formar o poro de fusão, no intuito de internalizar o vírus (Este e Telenti, 2007; Hughes e Nelson, 2009; Wilen et al., 2012).

Figura 3 – Visão geral da entrada do HIV na célula hospedeira. O Env, composto por subunidades de gp120 e gp41 (1), primeiramente se liga ao CD4 celular (2). Esta ligação leva a uma alteração conformacional do Env, permitindo a ligação com o correceptor, mediado pela alça V3 (3). O processo de fusão das membranas se inicia quando o peptídeo de fusão da gp41 se insere a membrana, e assume uma conformação de seis hélices aproximando as membranas o suficiente para formar o poro de fusão para a internalização do vírus (4). Adaptado de Wilen et al. (2012).

Os vírus R5 são encontrados na grande maioria das infecções primárias, enquanto variantes capazes de utilizar CXCR4 emergem somente ao longo do curso da infecção. A evidência de que o CCR5 desempenha um importante papel na transmissão e início da infecção é decorrente da presença de alelos CCR5 com uma deleção de 32 pares de bases (pb) (CCR5-Δ32) em cerca de 10% dos caucasianos (Dean et al., 1996). A homozigose deste alelo é rara (1% dos caucasianos), mas está associada com a ausência de moléculas CCR5 na superfície celular e a quase completa resistência à infecção por HIV-1 (Liu et al., 1996). A heterozigose está associada a concentrações reduzidas de CCR5 e foi correlacionada com a progressão lenta da doença (Zimmerman et al., 1997).

Vários estudos têm demonstrado que vírus X4 evoluem a partir de vírus R5 pré-existentes, seguindo uma via de mutações para a transição do correceptor e que mesmo após o aparecimento de vírus X4, as variantes R5 permanecem na maioria dos pacientes (Bunnik et al., 2011; Schuitemaker et al., 1992; Verhofstede et al., 2012). Outros estudos descrevem a presença de sequências de vírus X4 no DNA de células infectadas mesmo na ausência destas variantes no plasma dos

pacientes (Saracino et al., 2009; Shankarappa et al., 1999; Verhofstede et al., 2009). A alteração de populações R5 para X4 geralmente leva ao aumento acelerado da carga viral do paciente, diminuição da contagem de células T CD4+ e uma progressão mais rápida levando à aids e morte (Figura 4) (Goetz et al., 2009; Soulie et al., 2010; Verhofstede et al., 2012).

Figura 4 – Representação esquemática dos estágios da infecção pelo HIV-1. Painel esquerdo: pacientes infectados com vírus R5 com carga viral constante e diminuição de células T CD4+. Painel direito: pacientes infectados com vírus R5 e a emergência de vírus X4 associado com aumento acelerado da carga viral e diminuição de células T CD4+. Adaptado de Verhofstede et al. (2012).

Os mecanismos envolvidos na troca do tropismo estão bem estabelecidos e ainda não está claro se a evolução para populações X4 é a causa ou a consequência da progressão da doença (Verhofstede et al., 2012). Alguns estudos demonstraram que a terapia com antagonistas de CCR5 induziu uma rápida troca de vírus R5 para X4, sem afetar a carga viral. Após a interrupção da terapia com antagonistas de CCR5 os vírus X4 desapareceram da população viral plasmática e permaneceram como populações minoritárias, indicando a inibição da replicação de

vírus X4 por algum mecanismo ainda desconhecido (Tsibris et al., 2009; Verhofstede et al., 2012). É possível que variantes X4 surjam como consequência da progressão da doença devido ao número reduzido de células que expressam CCR5 ou pela superexpressão de CXCR4 (Briue et al., 2011). Outra hipótese, é que os vírus X4 sejam mais eficientemente reconhecidos pelo sistema imunológico do hospedeiro e consigam manter a replicação somente após a significativa deterioração deste sistema (Schreiber et al., 1996; Verhofstede et al., 2012).

1.2.1 Fatores virais associados ao tropismo do HIV-1

Os esforços para identificar o gene do HIV-1 que controla os fenótipos de tropismo revelaram que o determinante primário do tropismo viral está alocado no terceiro domínio variável (V3) da subunidade gp120 do envelope, mas que regiões do primeiro (V1) e segundo domínios (V2) também podem modular o tropismo em um menor grau (Fouchier et al., 1992; Hughes e Nelson, 2009; Mild et al., 2010).

Observou-se que isolados R5, usualmente, apresentam carga elétrica mais baixa na região V3, quando comparados a isolados X4, e inversamente, o correceptor CCR5 apresenta carga elétrica mais alta que o CXCR4 (Pollakis et al., 2004; Sierra et al., 2007). A ligação entre a V3 e o correceptor é carregada pela presença de aminoácidos básicos (K ou R), aminoácidos ácidos (D ou E) ou modificações pós-traducionais (geralmente N ou O-glicosilações ou sulfatação da tirosina), consequentemente, mutações que afetam estes fatores estão intrinsecamente correlacionadas com a seletividade do correceptor (Rosen et al., 2006; Sierra et al., 2007). Numerosas investigações confirmam que a presença de um resíduo carregado positivamente na posição 25 é capaz de converter o tropismo de um vírus R5 para X4, porém, ainda não está claro se o aparecimento de aminoácidos básicos nas posições 11 e 25 é suficiente ou mesmo necessário para

a emergência de vírus X4, mas é provável que a troca do tropismo seja um processo gradual de acúmulo de mutações (Jensen et al., 2003; Poon et al., 2007; Rosen et al., 2006). O aumento da carga positiva líquida da gp120 propicia a habilidade do vírus em aderir inespecificamente à superfície da célula devido à redução da repulsão eletrostática entre as membranas celular e viral. As substituições que resultam no aumento da carga líquida positiva ocorrem principalmente nas regiões V2 e V4, na superfície externa da gp120 (Repits et al., 2008).

Os sítios associados a N-glicosilações, ou *sequons* indicam a posição onde uma asparagina poderá ser glicosilada e para isso é requerida a presença do motivo de aminoácidos asparagina-X-serina/treonina (N-X-S/T), onde X pode ser qualquer aminoácido. A alteração de um sítio de glicosilação pode trazer impacto sobre a conformação da proteína e afetar porções distantes através de mudanças conformacionais. Padrões de glicosilação em proteínas virais podem mascarar epítomos, facilitando o escape imunológico, e influenciar na ligação com o receptor e propriedades fenotípicas dos vírus (Zhang et al., 2004). O número de glicosilações nos domínios variáveis da gp120 não está relacionado a um padrão específico de tropismo pelo correceptor, entretanto, um *sequon* particular NNT, localizado no sexto aminoácido da alça V3 é altamente conservado em vírus R5 e não em vírus X4 ou R5X4. Também foi demonstrado que o aumento do número de glicosilações na região V1/V2 pode indicar o uso potencial de CXCR4 (Clevestig et al., 2006; Ogert et al., 2001; Zhang et al., 2004).

A região V2 é um importante alvo para anticorpos neutralizantes e é conhecida por apresentar uma grande variação de comprimento, resultante de inserções ou deleções. Esta variação pode levar à inclusão de possíveis sítios de N-glicosilação, alteração de carga elétrica e mudanças conformacionais em outras regiões da gp120. Substituições de aminoácidos nesta região podem afetar a

função do envelope viral, mas não afetam a ligação gp120-CD4 (Jansson et al., 2001; Wang et al., 1995). O aumento do comprimento da V2 ocorre ao longo da infecção e existem controvérsias sobre a associação de padrões da sequência V2 e o fenótipo viral. Apesar do comprimento não estar relacionado à evolução da troca do tropismo alguns estudos relatam que vírus que trocam o tropismo para CXCR4 apresentam uma V2 mais longa que vírus que se mantém R5 ao longo da infecção, porém outros estudos não conseguiram comprovar esta relação (Jansson et al., 2001; Mild et al., 2009; Shioda et al., 1997; Wang et al., 1995).

Os principais estudos acerca do tropismo são realizados em coortes com prevalência do subtipo B. Alguns estudos indicam que a ocorrência de vírus R5X4 e X4 são raras no subtipo C (Dash et al., 2008; Tscherning et al., 1998) e que o subtipo E tem uma prevalência de vírus R5, X4 e R5X4 semelhante ao subtipo B (Subbarao et al., 2000). Alguns estudos demonstraram uma grande proporção de vírus X4 em variantes do subtipo D (Mueller e Bogner, 2007; Tscherning et al., 1998), enquanto estudos do subtipo A demonstraram uma baixa prevalência de vírus R5X4 e X4 e a presença de isolados que utilizam outros correceptores como BOB e CXCR6 (Hoffmann, 2007; Hughes e Nelson, 2009). Os poucos estudos que relatam o tropismo no subtipo F indicam a ausência de vírus X4 e R5X4 (Ferraro et al., 2001; Tscherning et al., 1998).

No Brasil, cocirculam duas variantes do subtipo B geneticamente distintas quanto ao motivo da coroa da alça V3. Uma das variantes apresenta motivo GPGR, que é prevalente em países da Europa e América do Norte enquanto a variante B' apresenta o motivo GWGR, responsável por aproximadamente um terço das infecções por subtipo B em nosso país (Couto-Fernandez et al., 1994; da Costa et al., 1995; de Brito et al., 2006). Apesar dos vírus GW terem sido observados esporadicamente em países como China, França, República Tcheca, Filipinas e Cuba, altas frequências de isolados GW têm sido observadas apenas no Brasil. Isto

sugere que a presença do triptofano (W) na ponta da alça V3, apesar de rara em outros países, não é deletéria ao HIV-1. Estudos relacionados à progressão da infecção sugerem que indivíduos infectados com variantes GW apresentam melhor prognóstico e uma evolução mais lenta da doença (de Brito et al., 2006; Leal et al., 2008). A comparação de modelos estruturais da alça V3 entre variantes GW e GP indicou diferenças principalmente na região de ligação com o correceptor, sendo vírus GW usavam exclusivamente o correceptor CCR5. Este achado pode explicar o fato de infecções causadas por variantes GWGR progredirem mais lentamente para aids que infecções causadas por variantes GPGR com vírus capazes de utilizar ambos os correceptores (Leal et al., 2008).

1.3 Antagonistas de correceptores

Os inibidores de entrada são um diverso grupo de fármacos, que têm como alvo múltiplos estágios do processo de entrada do HIV. Estes agentes apresentam um considerável potencial no tratamento de indivíduos infectados, particularmente aqueles com resistência a inibidores da transcriptase reversa, protease ou integrase. Atualmente os dois únicos inibidores de entrada aprovados para uso clínico são o inibidor de fusão Efuvirtida (T20), que atua sobre a gp41, e o antagonista de CCR5 Maraviroque (MVQ) (Haqqani e Tilton, 2013).

Logo após a identificação dos correceptores do HIV-1 para entrada na célula alvo, observou-se que pacientes que apresentavam homozigose para a CCR5 Δ 32 eram altamente resistentes à infecção sem apresentar sequelas imunológicas significativas (Liu et al., 1996). Ainda, verificou-se que os ligantes naturais do CCR5, as quimiocinas CCL3 (MIP-1 α), CCL4 (MIP1- β) e CCL5 (RANTES), eram capazes de bloquear a infecção pelo HIV-1 (Cocchi et al., 1995). Estes achados

sugeriam que o bloqueio da interação gp120-CCR5 através de moléculas inibidoras poderia ser uma estratégia efetiva e bem tolerada para a inibição da entrada do vírus (Haqqani e Tilton, 2013).

Assim, foram desenvolvidas pequenas moléculas capazes de bloquear os correceptores de entrada do HIV. Alguns fármacos bloqueadores do CXCR4 têm sido desenvolvidos, porém, nenhum foi aprovado para uso clínico até o momento (Haqqani e Tilton, 2013). As moléculas inibidoras de CCR5 se ligam à região hidrofóbica dos domínios transmembrana do CCR5 e alteram a conformação das alças extracelulares que interagem com o HIV durante a entrada a célula. Estas moléculas são nomeadas com o sufixo “viroc” derivado de ocupação do receptor viral (*viral receptor occupancy*) e vêm apresentando grande eficiência contra a replicação do HIV *in vitro* e algumas têm sido testadas em estudos clínicos (Haqqani e Tilton, 2013; Iwamoto et al., 2010).

O Aplaviroc (GW873140) chegou à fase 2b de estudos clínicos antes de ser suspenso, em 2005, devido a casos de hepatotoxicidade idiosincrática (Nichols et al., 2008). O fármaco Vicriviroc (SCH-417690, SCH-D) alcançou a fase 3 dos testes clínicos, mas seu desenvolvimento foi interrompido por não demonstrar superioridade sobre a terapia otimizada de base (OBT – *Optimized Background Therapy*) em dois ensaios duplamente cegos (Caseiro et al., 2012). O fármaco Cenicriviroc (TBR-652) é um antagonista de CCR5 e CCR2, que apresenta propriedades antivirais e anti-inflamatórias e continua em avaliação em estudos clínicos (Klibanov et al., 2010; Marier et al., 2011). O Maraviroc (UK-427857) é efetivo mesmo em baixos níveis nanomolares contra isolados de HIV que utilizam o correceptor CCR5 e foi o primeiro antagonista de CCR5 aprovado para uso clínico a partir de 2007 (Dorr et al., 2005; Emmelkamp e Rockstroh, 2007). No Brasil, é conhecido pelo nome Maraviroque (MVQ) e comercializado como Celsentri (ViiV Healthcare, Reino Unido).

1.4 Testes para determinação do tropismo

A indicação clínica do uso de antagonistas de correceptor implica na determinação do tropismo antes do início do tratamento, sendo que a administração somente poderá ser realizada após a confirmação da presença exclusiva de vírus R5 no paciente infectado. Além disso, ao longo do tratamento, a resistência ao medicamento poderá ocorrer devido à emergência de vírus X4 pré-existentes ou de vírus R5 capazes de entrarem na célula alvo mesmo na presença dos antagonistas de CCR5 e, portanto, a determinação do tropismo da população viral deve ser realizada periodicamente (Haqqani e Tilton, 2013; Obermeier et al., 2012; Wilkin e Gulick, 2012).

Atualmente, existem diferentes métodos para a determinação do tropismo do HIV-1, mas não está definido qual o mais apropriado para a rotina da prática clínica. Embora os estudos clínicos dos antagonistas de correceptores tenham preconizado o uso de ensaios fenotípicos, estes métodos são trabalhosos, demorados e caros. Uma alternativa mais rápida e barata inclui métodos genotípicos que predizem o uso do correceptor baseados em sequências da alça V3. Cada um destes métodos apresenta vantagens e limitações, e ambos dividem em comum o desafio de detectar baixos níveis de vírus X4 presentes como variantes minoritárias dentro da *quasispecie* viral (Lin e Kuritzkes, 2009).

Os primeiros ensaios para a avaliação do fenótipo viral foram desenvolvidos na década de 80, com objetivo de classificar os vírus em indutores ou não indutores de sincício. A habilidade de replicação do HIV em linhagens celulares específicas, associada à fusão celular (sincício) indica a presença de vírus X4 (Braun e Wiesmann, 2007). Após a descoberta dos correceptores, várias linhagens celulares foram desenvolvidas para expressar diversas combinações de receptores de HIV-1

em sua superfície. A maioria destes ensaios utilizam linhagens celulares de glioma humano (U87, U372 ou NP-2) que expressam CD4 e CCR5 ou CXCR4. A linhagem celular GHOST também é frequentemente utilizada como linhagem celular indicadora e carrega um gene *repórter* GFP regulado pela expressão da Tat do HIV-1 (Lin e Kuritzkes, 2009).

Os ensaios fenotípicos recombinantes (RVA – *Recombinant Viral Assays*) utilizam vírus recombinantes ou pseudovírus construídos com sequências do envelope amplificadas por RT-PCR, provenientes do plasma do paciente. Estes fragmentos de *env* amplificados preservam a diversidade da população viral presente no plasma, prevenindo qualquer seleção que possa ocorrer durante a passagem *in vitro*. A representação confiável da diversidade viral resulta principalmente da sensibilidade e precisão das reações de RT-PCR, que depende de níveis de carga viral plasmática acima de 1000 cópias/mL (Lin e Kuritzkes, 2009; Whitcomb et al., 2007). Atualmente existem quatro ensaios comerciais: *Trofile* (Monogram Biosciences, EUA), *Phenoscript* (VIRalliance, França), *Xtrack[®]/PhenX-R* (inPheno AG, Suíça) e *Virco* (Virco BVBA, Bélgica) (Braun e Wiesmann, 2007; Trouplin et al., 2001; Van Baelen et al., 2007; Whitcomb et al., 2007). O ensaio *Trofile* é conhecido como o mais amplamente utilizado em estudos clínicos e atualmente apresenta uma versão melhorada que aumentou a sensibilidade para a detecção de variantes X4 minoritárias. O ensaio atual detecta vírus X4 presentes em até 0,3% da população com 100% de sensibilidade (Lin e Kuritzkes, 2009).

Os métodos genotípicos para a determinação do tropismo viral são baseados no princípio de que a alça V3 é o principal determinante para o uso do correceptor (Jensen e van 't Wout, 2003). O algoritmo mais simples é a regra 11/25 que considera apenas a carga elétrica dos aminoácidos das posições 11 e 25 da alça V3, porém esta técnica não apresentou resultados confiáveis em amostras clínicas (Low et al., 2007). Várias abordagens de bioinformática para a

interpretação de sequências de aminoácidos da alça V3 têm sido desenvolvidas para a predição do uso do correceptor do HIV-1, incluindo árvores de decisão, máquinas de vetores de suporte (SVM – *Support Vector Machines*), redes neurais (NN – *Neural Networks*) e matrizes de pontuação de posições específicas (PSSM – *Position-Specific Scoring Matrices*). As SVMs geram uma pontuação categórica que diferencia vírus R5 e X4 (Pillai et al., 2003). Os algoritmos de NN usam um grande conjunto de dados de sequências de amostras com tropismo fenotípico conhecido para traçar correlações entre os padrões complexos da variabilidade das sequências com o fenótipo correspondente (Resch et al., 2001). A PSSM é um método de predição que utiliza um sistema de pontuação que compara a sequência do paciente com sequências X4 atribuindo uma pontuação para as semelhanças encontradas (Jensen e van 't Wout, 2003).

O sistema *Geno2pheno*_[coreceptor] é um sistema de interpretação avançado para predição de tropismo. Ele é capaz de avaliar sequências contendo misturas de nucleotídeos e interpretar todos os possíveis resultados. Ainda é possível incluir dados adicionais como carga viral, contagem de linfócitos e CCR5Δ32 e escolher valores de razão de falso positivo com diferentes níveis de restrição. Este sistema têm sido amplamente avaliado e recomendado à prática clínica (Lengauer et al., 2007; Obermeier et al., 2012; Vandekerckhove et al., 2011).

O sequenciamento de amostras de DNA ou RNA viral pode ser realizado através do método de terminação em cadeia (Sanger) ou por uma das plataformas de sequenciamento massivo paralelo (NGS – *Next Sequencing Generation*) (Chueca et al., 2009; de Mendoza et al., 2008; Swenson et al., 2011). Em estudos clínicos e em várias coortes de pacientes a amplificação e o sequenciamento da região V3 têm sido realizados em triplicatas, para melhorar o desempenho da predição (Chueca et al., 2009; Poveda et al., 2009; Raymond et al., 2011; Vandekerckhove et al., 2011). As variantes minoritárias, presentes em frequências

inferiores a 10-20%, geralmente permanecem indetectáveis em métodos convencionais de sequenciamento (Sanger), enquanto por NGS é possível sequenciar milhares de sequências da V3 em uma única sequência, permitindo uma análise mais sensível da variabilidade genética em cada paciente (Recordon-Pinson et al., 2010; Swenson et al., 2011; Vandenbroucke et al., 2008). Entretanto, as plataformas de NGS ainda estão restritas a serviços acadêmicos ou comerciais (Vandekerckhove et al., 2011).

2 OBJETIVOS

2.1 Objetivo geral

Caracterizar a glicoproteína (gp) 120 do envelope do HIV-1 de indivíduos infectados por este vírus e verificar as implicações de fatores moleculares sobre o tropismo pelos correceptores CCR5 e CXCR4.

2.2 Objetivos específicos

- Determinar o tropismo pelos correceptores CCR5 e CXCR4 utilizando ensaios fenotípicos e sistemas preditores de bioinformática;
- Determinar a prevalência do tropismo do HIV-1 pelos correceptores CCR5 e CXCR4 na população estudada;
- Avaliar as implicações dos fatores moleculares da gp120 sobre o tropismo do HIV-1 e verificar sua relação com os subtipos circulantes na população estudada.

3 MATERIAL E MÉTODOS

3.1 Casuística

Foram incluídos no estudo 283 pacientes infectados pelo HIV, que fazem acompanhamento clínico no Ambulatório de Imunodeficiências Secundárias – Departamento de Dermatologia (ADEE 3002) HCFMUSP.

3.1.1 Critérios de inclusão

Somente foram incluídos no estudo indivíduos com idade superior a 18 anos, homens ou mulheres, com sorologia positiva para HIV-1.

3.1.2 Critérios de exclusão

Foram excluídos os pacientes com dados clínicos e laboratoriais incompletos.

3.1.3 Aspectos éticos

Os indivíduos que aceitaram participar deste estudo assinaram um termo de consentimento livre e esclarecido (TCLE) (Apêndice A). Todos os dados individuais são confidenciais e em nenhum momento serão divulgados. O processo foi aprovado pelo CEP/HCFMUSP, CAPPesq nº 0108/08 (Anexos A e B).

3.2 Coleta das amostras

Foram coletadas amostras de sangue periférico através de punção venosa dos indivíduos participantes da pesquisa em três tubos de 5 mL contendo EDTA, que foram centrifugados a 450 xg em temperatura ambiente por 5 min para promover a separação do plasma.

O plasma de cada paciente foi separado em duas alíquotas de 1500 µL, enquanto o plasma excedente foi acondicionado em alíquotas de 500 µL, sendo todas armazenadas a -70°C até sua utilização.

A quantificação de carga viral plasmática do RNA do HIV-1 foi realizada utilizando o *kit* comercial *Versant HIV RNA 3.0 Assay (bDNA)* (Siemens Healthcare Diagnostics, EUA) e a contagem de células T CD4+/T CD8+ foi determinada por citometria de fluxo. A coleta de amostras para estes ensaios foi realizada no mesmo momento da coleta das amostras para este estudo. Estes ensaios fazem parte dos testes de rotina de indivíduos infectados pelo HIV-1 em acompanhamento clínico e os resultados demonstrados neste estudo foram obtidos nos registros dos prontuários.

3.3 Separação das células mononucleares do sangue periférico

A técnica de separação de células mononucleares do sangue periférico (PBMCs – *Peripheral Blood Mononuclear Cells*) é baseada na separação por gradiente de densidade utilizando *Ficoll Paque Plus* (GE Healthcare, Suécia). Após a coleta do plasma, as amostras de sangue foram diluídas em solução salina na proporção 1:1 e transferidas para um tubo cônico de 50 mL contendo *Ficoll* na proporção 1:3 para posterior centrifugação a 860 xg, por 20 min, a 16°C.

As PBMCs separadas pelo gradiente de densidade foram removidas e transferidas para um tubo cônico de 15 mL e lavadas com solução salina sob centrifugação a 450 xg, por 10 min, a 16°C. Uma segunda lavagem com solução salina foi realizada nas mesmas condições da lavagem anterior.

O sobrenadante foi removido e o *pellet* de células foi ressuspensão em 1 mL de salina para contagem de leucócitos no contador hematológico *Cell-Dyn 1400* (Abbott, EUA). Foram separadas duas alíquotas contendo $2 \cdot 10^6$ células/mL que foram centrifugadas a 6000 xg, por 3 min, a temperatura ambiente. Após a remoção do sobrenadante, uma alíquota foi armazenada a -70°C e outra a -20°C até a sua utilização.

A separação das PBMCs foi realizada para que todas as amostras destinadas à extração de DNA partissem do mesmo número de células. Esse procedimento possibilitará a realização de futuros estudos moleculares quantitativos utilizando este material.

3.4 Extração de DNA

O DNA genômico foi extraído a partir das PBMCs utilizando o *Illustra Blood GenomicPrep Mini Spin Kit* (GE Healthcare, Reino Unido), seguindo as orientações fornecidas pelo fabricante. As amostras de DNA extraído foram avaliadas quanto à concentração, qualidade e pureza utilizando o espectrofotômetro *ND-1000 Spectrophotometer v.3.5* (NanoDrop Technologies, EUA).

3.5 Amplificação e sequenciamento da região V3 da gp120

3.5.1 Amplificação da região V3

O DNA extraído foi submetido a amplificação por reação em cadeia da polimerase (PCR – *Polymarese Chain Reaction*) *nested* (aninhada), utilizando os *primers* (iniciadores) *outer* (externos) senso ED5 e antissenso ED12 e os *primers* *inner* (internos) senso ED31 e antissenso ED33 (Delwart et al., 1995) (Tabela 1).

Tabela 1 – Sequências dos *primers* utilizados para amplificar e sequenciar a região V3 da gp120

Nome	Sequência (5' – 3')	Posição ¹
ED5 <i>outer</i> senso	ATGGGATCAAAGCCTAAAGCCATGTG	6556 – 6581
ED12 <i>outer</i> antissenso	AGTGCTTCCTGCTGCTCCCAAGAACCCAAG	7822 – 7792
ED31 <i>inner</i> senso	CCTCAGCCATTACACAGGCCTGTCCAAAG	6816 – 6844
ED33 <i>inner</i> antissenso	TTACAGTAGAAAAATTCCCCTC	7359 – 7380

Nota: ¹ Posição sobre a sequência referência HXB2, acesso no GenBank nº K03455

As amplificações foram realizadas em triplicata utilizando termociclador convencional, seguindo as etapas para a PCR *outer*: uma desnaturação inicial de 94°C durante 1 min; seguida de 35 ciclos de 94°C por 45 s, 55°C por 45 s e 72°C por 2 min; e uma extensão final a 72°C por 10 min. Para a PCR *inner* foram seguidas as seguintes condições: uma desnaturação de 94°C durante 1 min; seguida de 35 ciclos de 94°C por 1 min, 55°C por 1 min e 72°C por 1:30 min; com uma extensão final de 72°C por 10 min. As condições para uma reação final de PCR de 50 µL foram de 1x tampão de PCR (10x), 2 mM de MgCl₂, 0,2 mM de cada dNTPs, 0,2 µM de cada *primer* 1 U de *Platinum Taq DNA polimerase High Fidelity*

(Invitrogen, EUA) e 5 µL de DNA extraído (20 – 50 ng) para a PCR *outer* e 5 µL de produto de PCR para a PCR *inner*.

Os produtos da PCR *inner* foram submetidos à eletroforese em gel de agarose 1,5% a 100 V por 60 min. A visualização foi realizada pela adição de brometo de etídeo (0,5 µg/mL) ao gel, sob incidência de luz UV. A quantificação e a determinação do tamanho dos fragmentos foi realizada utilizando os padrões de peso molecular de 100 pb como parâmetro de tamanho de fragmento e *Low DNA Mass* como parâmetro de concentração. Foram considerados produtos de PCR adequados os fragmentos com aproximadamente 564 pb apresentando concentração entre 20 e 100 ng. Estes produtos foram purificados com o *QIAquick Purification Kit* (Qiagen, Alemanha), conforme o protocolo fornecido pelo fabricante.

3.5.2 Reação de sequenciamento da região V3

As sequências da região V3 foram obtidas pelo método de terminação de cadeia (Sanger) e eletroforese capilar. Para a reação de sequenciamento foi utilizado o kit *ABI PRISM Big Dye Terminator v. 3.1 Cycle Sequencing Ready Reaction* (Applied Biosystems, EUA), adicionando 20 a 100 ng dos produtos amplificados purificados e 0,25 µM dos respectivos *inner primers*. As reações de sequenciamento foram realizadas em termociclador convencional seguindo 25 ciclos de 96°C por 10 s, 55°C por 5 s e 60°C por 4 min. Os produtos foram purificados com uma lavagem utilizando solução de etanol absoluto e EDTA (125mM), seguido de uma lavagem com etanol 70%. Ao *pellet* seco foram adicionados 10 µL de formamida, seguido de uma etapa de desnaturação a 95°C por 3 min e choque térmico em gelo por 15 min. A corrida foi executada em sequenciador automático *3500 Genetic Analyser* (Applied Biosystems, EUA).

3.5.3 Análise do sequenciamento

As sequências senso e antissenso foram sobrepostas, avaliadas e editadas através do programa *SeqScape* v. 2.7.0 (Applied Biosystems, EUA). A sobreposição de picos no eletroferograma foi considerada como mistura de nucleotídeos em uma mesma posição quando a diferença entre os picos foi $\leq 25\%$ e a mistura esteve presente em pelo menos duas leituras entre as seis produzidas pelo sequenciamento das triplicatas. As regiões com ruído e/ou qualidade comprometida foram removidas da análise.

3.6 Amplificação e sequenciamento da região da gp120 do gene *env*

A partir do conjunto de dados gerado através do sequenciamento da região V3 e da determinação de tropismo por ensaio *Trofile*, foram selecionadas 20 amostras para o sequenciamento da região da gp120 do gene *env*. Foram selecionadas apenas amostras do subtipo B para a região V3. Os demais critérios de seleção foram resultados de tropismo fenotípico, composição da tétrade de aminoácidos da região V3 (posições 15 a 18), uso de terapia antirretroviral (ART) e sexo com o objetivo de ter uma representatividade de metade dos indivíduos dentro de cada um destes critérios (Tabela 2).

Tabela 2 – Amostras selecionadas para amplificação da gp120 do gene *env*

Amostra	Identificação	Sexo	Uso de ART ¹	Trofile	Tétrede da V3 ²
1	EI006B	Feminino	Sim	R5	AWGR
2	EI061B	Masculino	Sim	R5	GFGR
3	EI077B	Feminino	Não	R5	GPGQ
4	EI110	Masculino	Não	R5	GWGR
5	EI113	Masculino	Não	R5X4	GWGR
6	EI115	Masculino	Sim	R5	GWGR
7	EI127	Feminino	Não	R5X4	GPGR
8	EI129	Feminino	Não	R5	GWGR
9	EI130	Feminino	Sim	R5X4	GWGR
10	EI135	Masculino	Sim	X4	EWRR
11	EI138	Feminino	Sim	R5X4	GPGR
12	EI145	Masculino	Sim	R5	GPGR
13	EI149	Masculino	Sim	R5X4	GWGR
14	EI212	Masculino	Sim	R5X4	GPGR
15	EI219	Feminino	Sim	R5X4	GQGR
16	EI240	Masculino	Sim	R5	GWGR
17	EI243	Feminino	Sim	R5	GPGR
18	EI251	Masculino	Não	R5	GPGR
19	EI256	Feminino	Não	R5	GPGQ
20	EI271	Masculino	Sim	R5X4	GWGR

Nota: ¹ terapia antirretroviral; ² Aminoácidos localizados na ponta da região V3.

3.6.1 Amplificação da região da gp120

O protocolo para amplificação da gp120 foi gentilmente cedido pela Dra. Kristel Van Laethem do Laboratório de Virologia Clínica e Evolutiva do Insituto Rega da Universidade Católica de Leuven, Bélgica, durante a realização de treinamento em bioinformática neste laboratório no período de setembro de 2011 a janeiro de 2012. O DNA extraído foi submetido a amplificação por PCR *nested*, utilizando os *primers outer* senso EnvA e antissenso KVL009 e os *primers inner* senso EnvB e antissenso EnvM (Van Laethem et al., 2005) (Tabela 3).

Tabela 3 – Sequências dos *primers* utilizados para amplificar e sequenciar a região da gp120 do gene *env*

Nome	Sequência (5' – 3')	Posição ¹
EnvA		5953 – 5981
outer senso	GGCTTAGGCATCTCCTATGGCAGGAAGAA	
KVL009		9135/60 –
outer antissenso	GCC AATCAGGGAAGWAGCCTTGTGT	9159/84
EnvB		6198 – 6224
inner senso	AGAAAGAGCAGAAGACAGTGGCAATGA	
EnvM		9058 – 9086
inner antissenso	TAGCCCTTCCAGTCCCCCTTTTCTTTTA	

Nota: ¹ Posição sobre a sequência referência NL4.3, acesso no GenBank nº AF324493

Para as PCR *outer* e *inner* foram preparadas reações com volume final de 50 µL contendo 1x tampão de PCR (10x), 2 mM de MgCl₂, 0,2 mM de cada dNTPs, 0,2 µM de cada *primer*, 1 U de *Platinum Taq DNA polimerase High Fidelity* (Invitrogen, EUA) e 5 µL de DNA extraído (20 – 50 ng) para a PCR *outer* e 1 µL de produto de PCR para a PCR *inner*. Em termociclador convencional as etapas da PCR *outer* foram compostas por uma desnaturação inicial a 94°C por 1:30 min, seguida de 40 ciclos a 94°C por 15 s, 55°C por 30 s e 68°C por 3:30 min, seguidos de uma extensão final de 10 min. A ciclagem da PCR *inner* foi composta por uma desnaturação inicial a 94°C por 1:30 min, seguido por 5 ciclos a 94°C por 15 s, 55°C por 30 s e 68°C por 2:45 s, seguidos por 30 ciclos a 94°C por 15 s, 55°C por 30 s e 68°C por 3 min iniciais adicionando 1 s por ciclo, seguidos de uma extensão final a 68°C por 7 min.

Os produtos da PCR *inner* foram submetidos à eletroforese em gel de agarose, da mesma forma que os produtos da PCR para região V3, sendo considerados produtos satisfatórios aqueles com tamanho de aproximadamente 2888 pb. A purificação destes fragmentos foi realizada com o *QIAquick Purification Kit* (Qiagen, Alemanha), seguindo instruções do fabricante. A quantificação

fluorométrica dos produtos de PCR purificados foi realizada em fluorômetro Qubit 2.0 (Invitrogen, EUA), seguindo as instruções do fabricantes, sendo consideradas satisfatórias as amostras que apresentaram concentração superior a 50 ng.

3.6.2 Reação de sequenciamento da região da gp120

Os produtos de amplificação da gp120 foram sequenciados utilizando a plataforma *Ion Torrent* (Life Technologies, EUA). Esta plataforma de sequenciamento consiste em três etapas básicas: 1) preparo da biblioteca, 2) preparo do molde (*template*) e 3) sequenciamento propriamente dito.

3.6.2.1 Preparo da biblioteca

Os produtos de PCR purificados foram submetidos à fragmentação enzimática utilizando o *Ion Xpress™ Plus Fragment Library Kit* (Life Technologies, EUA) com o objetivo de gerar fragmentos de aproximadamente 200 - 300 pb. Após o preparo seguindo as instruções do fabricante, as amostras foram incubadas a 37°C por 5 minutos, para promover a atividade enzimática.

A purificação das amostras fragmentadas foi realizada por *AMPure beads* (Invitrogen, EUA) sobre uma estante magnética. Ao material fragmentado foi adicionada uma solução contendo as pérolas (*beads*) magnéticas, se ligando ao DNA. Quando submetidos ao campo magnético, as *beads* ligadas ao DNA ficaram retidas nas paredes dos tubos, separando o DNA das impurezas. Lavagens com etanol 70% removeram o material excedente e em seguida foi realizada uma eluição para que as *beads* se desligassem do DNA.

A ligação dos adaptadores e reparo das extremidades (*nick repair*) consiste em uma reação enzimática que insere os adaptadores A e P1 bem como os

códigos de barras (*barcodes*) nas extremidades dos fragmentos de DNA. Os adaptadores A e P1 são oligonucleotídeos que são ligados às extremidades dos fragmentos de DNA que os permitem prosseguir nas etapas de amplificação clonal, enriquecimento e ligação dos *primers* de sequenciamento. Os *barcodes* são oligonucleotídeos que permitem a multiplexação da reação de sequenciamento, identificando individualmente cada amostra. A Figura 5 apresenta um resumo geral dos procedimentos descritos acima.

Figura 5 – Visão geral do preparo de biblioteca por fragmentação da plataforma *Ion Torrent* (Life Technologies, EUA). Fonte: <http://ioncommunity.lifetechnologies.com>.

Após a purificação dos fragmentos utilizando as *beads* magnéticas, foi realizada uma eletroforese utilizando *E-Gel SizeSelect 2% agarose gel* (Invitrogen,

EUA), de onde foram recolhidas bibliotecas com aproximadamente 350 pb, que correspondem ao fragmento esperado de 200-300 pb mais aproximadamente 50 pb de adaptadores e *barcodes*.

Em seguida, as bibliotecas foram quantificadas por PCR tempo real utilizando ensaio *TaqMan* específico para quantificação de bibliotecas da plataforma *Ion Torrent*. As 20 bibliotecas foram separadas em quatro grupos contendo cinco bibliotecas cada e diluídas a concentrações equimolares para proceder a reação multiplexada.

3.6.2.2 Preparo de *template*

A PCR em emulsão tem por objetivo isolar individualmente as moléculas de DNA para que possam ser copiadas. Cada micela formada pela emulsão se comporta como um microrreator onde acontecerá a PCR. A concentração dos reagentes é crucial para garantir que apenas uma molécula de DNA se ligue sobre a superfície de cada partícula esférica (IPS – *Ion Sphere Particle*) através dos adaptadores P1.

Seguindo as instruções do manual do *Ion PGM Template OneTouch 2 200 Kit* (Life Technologies) preparou-se uma solução contendo as bibliotecas, os reagentes necessários e as IPS, que foi transferida para um filtro e em seguida foi adicionada a solução oleosa. A reação de PCR em emulsão é automatizada pelo instrumento *Ion OneTouch 2* (Life Technologies, EUA) e ao final do ciclo da PCR foram recuperadas duas alíquotas de solução contendo IPS livres e IPS cobertas pelos fragmentos de DNA originados pela amplificação clonal, descartando, portanto, qualquer DNA não ligado às IPS.

O enriquecimento das IPS é automatizado pelo instrumento *Ion OneTouch ES* (Life Technologies, EUA). Para promover a separação e descarte de IPS livres,

adicionou-se ao *template* uma solução de biotina, se ligando às fitas duplas de DNA. Uma solução de estreptavidina magnética foi adicionada, ligando-se à biotina que sob a ação de uma placa magnética imobilizou o conjunto estreptavidina-biotina-DNA à parede do tubo. Ao final das lavagens com etanol 70% para remoção de impurezas, o DNA de fita dupla ligado às IPS foi eluído e submetido à desnaturação por uma solução de NaOH.

3.6.2.3 Sequenciamento

A plataforma *Ion Torrent* se baseia na tecnologia de chips semicondutores com milhões de poços sobre a sua superfície. Abaixo de cada poço existe uma camada sensível que captura sinais químicos da reação de sequenciamento e os converte em informação digital de chamada de bases (*basecalls*) (Figura 6).

Figura 6 – Estrutura esquemática do chip semicondutor da plataforma *Ion Torrent* (Life Technologies, EUA). Fonte: <http://ioncommunity.lifetechnologies.com>.

As IPS cobertas de DNA clonado de cada uma das quatro bibliotecas multiplexadas (cinco bibliotecas por multiplex) foram carregadas em chips 314, que

tem capacidade para gerar até 100 Megabases por reação, de forma que cada IPS ocupou um único micropoço dentre os 1,1 milhões distribuídos sobre a superfície do chip. Após o carregamento, o chip foi inserido no instrumento *Ion PGM Sequencer* (Life Technologies, EUA) onde foi submetido a ciclos de fluxos de cada um dos nucleotídeos intercalados por lavagens. A cada vez que uma base nitrogenada é incorporada em uma fita simples de DNA, um íon H^+ é liberado alterando o pH da solução do poço (Figura 7). Esta alteração química é detectada pela camada sensível abaixo do poço e convertida em voltagem, cuja diferença de potencial é transmitida para o sistema como *basecall*.

Figura 7 – Esquema ilustrativo da reação de sequenciamento da plataforma *Ion Torrent* (Life Technologies). Íons H^+ são liberados no meio a cada base incorporada. O chip semicondutor funciona como um micropeagâmetro detectando as alterações químicas e convertendo-as em informação digital. Fonte: <http://ioncommunity.lifetechnologies.com>.

3.6.3 Análise do sequenciamento

Os dados brutos (*raw data*) gerados pelo sequenciamento foram pré-processados em *basecalls* pelo programa *Torrent Suite* (Life Technologies, EUA).

As *basecalls* de cada fragmento sequenciado formam leituras (*reads*), que variam de dúzias a centenas de pares de bases, sendo exportadas no formato de um arquivo FASTQ.

A montagem das *reads* foi realizada por três estratégias diferentes: montagem *de novo* pelo módulo de extensão (*plugin*) *Assembler* do pacote de programas *Ion Reporter* (Life Technologies, EUA); mapeamento contra a referência HIV HXB2 (GenBank nº de acesso K03455) pelo o programa *CLC Genomics Workbench v. 6.5.1* (CLC bio, EUA) e caracterização de *reads* virais pelo programa *Segminator II v.0.1.1* (Archer et al., 2012; Archer et al., 2010). Montagens *de novo* são estratégias que não utilizam uma sequência de referência para o mapeamento das *reads*, em casos em que a referência é desconhecida, ou ainda em casos em que as sequências apresentam diversidade genética muito alta para que o desempenho da montagem alcance níveis ótimos. O *plugin Assembler* é uma extensão do montador (*assembler*) MIRA (Chevreux et al., 2004) disponibilizado no pacote *Ion Reporter* e otimizado para montagens *de novo* de sequências provenientes da plataforma *Ion Torrent*. O mapeamento contra uma sequência de referência é a estratégia de montagem mais comum, e mais recomendada para que o maior número de *reads* seja reconhecido e componha a análise. O programa *Segminator II* foi desenvolvido especificamente para genomas pequenos altamente variáveis, aplicando algoritmos específicos que minimizam a perda de dados durante a montagem (Archer et al., 2010).

As sequências da região V3 foram extraídas dos consensos obtidos através de sequenciamento clássico (Sanger) e sequenciamento massivo paralelo e comparadas entre si através de análise filogenética empregando o método de Máxima Verossimilhança com *bootstrap* de 1000 réplicas, utilizando o programa MEGA 5 (Tamura et al., 2011).

3.7 Testes de tropismo

3.7.1 Teste fenotípico para determinação de tropismo

O laboratório *Monogram Biosciences* (São Francisco, EUA) estabeleceu uma colaboração se comprometendo a realizar o ensaio *Trofile* para 100 indivíduos que apresentassem carga viral plasmática > 1000 cópias/mL, dos quais 50 deveriam estar sob terapia antirretroviral (ART) e 50 não deveriam ter sido expostos ao tratamento (*naïves*). Ao final de um ano de recrutamento foi possível enviar para o laboratório nos Estados Unidos, 99 amostras contendo 3 mL (2 alíquotas de 1500 µL) de plasma de 37 indivíduos sob ART e 62 *naïves*.

O *Trofile* (Monogram Biosciences, EUA) é o ensaio comercial mais amplamente utilizado para determinação de tropismo em estudos clínicos (Braun e Wiesmann, 2007). Este ensaio utiliza o gene *env* completo proveniente de amostras do paciente, com tamanho aproximado de 2,5 Kb, que é amplificado e inserido em um vetor de expressão do envelope. A cotransfecção do vetor de expressão do envelope do vetor de HIV genômico contendo o gene *reporter* luciferase é realizada em células de linhagem permissiva para gerar partículas de pseudovírus contendo os genes *env* e luciferase. Os pseudovírus são, então, utilizados para infectar linhagens celulares que expressam CD4/CXCR4 ou CD4/CCR5 na superfície celular. O resultado é inferido pela quantificação da luz emitida pela expressão do gene *reporter* luciferase nas células infectadas (Braun e Wiesmann, 2007; Whitcomb et al., 2007). Neste estudo foi utilizado o ensaio *Trofile* original (OTA – *Original Trofile Assay*), porém, atualmente já está disponível uma versão mais sensível (ESTA – *Enhanced Sensitivity Trofile Assay*) (Su et al., 2009).

3.7.2 Teste genotípico para predição de tropismo

As sequências da região V3 obtidas foram submetidas à predição de tropismo utilizando o programa *Geno2pheno*_[coreceptor] v. 2.5 (G2P), com razão de falso positivo (FPR – *False Positive Rate*) igual a 10%, conforme as Orientações Europeias Sobre o Manejo Clínico de Testes de Tropismo do HIV-1 (Lengauer et al., 2007; Vandekerckhove et al., 2011).

O G2P é um preditor de tropismo baseado na sequência da região V3, desenvolvido, mantido e disponibilizado na *internet* pelo Instituto de Informática Max Planck, Alemanha. A predição é realizada através de um método de aprendizagem estatística baseado em uma máquina de vetores de suporte (SVM – *Support Vector Machine*) que é um método de classificação dicotômica. Embora o tropismo do HIV-1 apresente três classificações (R5, R5X4 e X4), este modelo foi desenvolvido para determinar se o vírus pode ou não utilizar o correceptor CXCR4, portanto, este modelo dispõe vírus X4 e R5X4 na classificação positiva (Lengauer et al., 2007; Sing et al., 2007). O resultado deste teste indica “apenas o correceptor CCR5 pode ser usado” para sequências classificadas como provenientes de vírus R5 e “o correceptor CXCR4 pode ser usado” para sequências de vírus R5X4 ou X4.

3.8 Determinação dos subtipos

Os subtipos das sequências obtidas foram determinados através de métodos automáticos que compõem ferramentas de bioinformática disponíveis na *internet*: *REGA HIV Subtyping Tool* v.3 (<http://caipirinha.med.kuleuven.be/genotype/hiv/typingtool/>) e *RIP – Recombinant Identification Program* (<http://www.hiv.lanl.gov>

/content/sequence/RIP/RIP.html). A subtipagem foi confirmada através de análise filogenética empregando o método de Máxima Verossimilhança com *bootstrap* de 1000 réplicas, utilizando o programa MEGA 5 (Tamura et al., 2011). Também foram avaliados perfis de recombinação utilizando *SimPlot* v.3.5.1 (Lole et al., 1999).

3.9 Predição de sítios potenciais de N-glicosilação

Os sítios de N-glicosilação foram preditos utilizando o programa *N-GlycoSite* disponibilizado pela base de dados de HIV do Laboratório Nacional de Los Alamos, EUA (disponível em: <http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html>). Esta ferramenta reconhece e destaca sítios relacionados a N-glicosilações que são padrões NX[S ou T], chamados de *sequons*, onde X pode ser qualquer aminoácido, seguido por uma serina (S) ou treonina (T). Um *sequon* não será glicosilado se contiver ou for seguido por uma prolina (P) (Zhang et al., 2004).

3.10 Determinação da carga elétrica líquida

A carga elétrica líquida das sequências obtidas por sequenciamento massivo paralelo foi calculada utilizando um programa em *Python* escrito especificamente para este estudo. Os resíduos básicos arginina (R), histidina (H) e lisina (K) conferem à proteína, ou neste caso, à sequência de aminoácidos, a carga elétrica positiva (+1 para cada resíduo), enquanto os resíduos ácidos aspartato (D) e glutamato (E) atribuem carga negativa (-1 para cada resíduo). O *script* em *Python* para o cálculo da função da soma de cargas positivas e negativas dos aminoácidos está apresentado no Apêndice B.

3.11 *Bayesian Network*

Redes Bayesianas (BN – *Bayesian Network*) foram utilizadas para verificar relações entre as mutações e o tropismo das sequências obtidas. Para avaliar os resultados obtidos por BN, foi incluído um conjunto de dados contendo 1208 sequências com tropismo conhecido selecionado a partir do banco de dados Los Alamos (www.hiv-web.lanl.gov). Neste conjunto, foram incluídas apenas sequências apresentando 35 aminoácidos na região V3. Todas as sequências originárias do Brasil foram incluídas e dentre as demais regiões geográficas, foram selecionadas pelo menos 5% de sequências representantes dos subtipos B, C e F1 para cada um dos três resultados de tropismo possíveis. Estes conjuntos de sequências foram avaliados utilizando o programa *B-Course v.2.0* (Myllymäki et al., 2002) para modelagem de classificação e o programa *B-Right v.0.1*, que consiste em uma modificação do programa *B-Course* desenvolvida pelo grupo de pesquisa do Laboratório de Virologia Clínica e Evolutiva do Instituto Rega, KUL, Bélgica.

Uma BN é um modelo probabilístico que descreve independências estatísticas entre variáveis múltiplas. Dependências são representadas em um gráfico acíclico dirigido que forma o componente qualitativo de uma BN. Cada nó em um gráfico corresponde a uma variável, e um arco direto (seta) entre nós representa uma dependência incondicional, que corresponde a uma influência direta. Uma independência condicional é representada pela ausência de um arco relacionado diretamente em um gráfico completamente conectado (Deforche et al., 2006).

A Figura 8 ilustra um exemplo hipotético de cinco variáveis de uma amostra de um paciente. A ausência de uma linha direta entre geografia e carga viral indica que não há influência direta suficiente da geografia sobre a carga viral. A presença

de um arco indica que uma variável é dependente de outra e a influência de uma sobre a outra poderá ser tanto positiva quanto negativa. Quanto mais clara a cor do arco (**preto** > **roxo** > **azul escuro** > **azul claro**), mais fraca é a interação entre as variáveis. Isso significa que se determinada variável apresenta uma relação “fraca” (azul claro) poderia ser removida da análise probabilística sem causar grandes distúrbios na composição das demais interações na rede. Por outro lado, quando há uma forte interação (preto) significa que há uma relação não apenas importante entre as duas variáveis envolvidas, mas também sobre toda a estrutura da rede.

Figura 8 – Exemplo hipotético de uma rede Bayesiana (BN). Arcos diretos demonstram a influência direta de uma variável sobre a outra. Quanto mais clara a cor do arco, mais fraca é a interação, quanto mais escuro, maior a relação de dependência. A ausência de arcos entre duas variáveis denota sua independência probabilística. Adaptado de Deforche et al. (2006).

3.12 Análise estatística

Os testes Exato de Fisher, Chi Quadrado e Mann-Whitney foram aplicados para avaliar os resultados obtidos utilizando o programa *GraphPad Prism v. 5.00* (GraphPad Software, EUA).

4 RESULTADOS

4.1 Dados demográficos e laboratoriais

A amostragem deste estudo incluiu 283 indivíduos infectados por HIV-1 em acompanhamento clínico. A média de idade dos indivíduos foi de 42 anos, sendo a maioria constituída por homens (70,3%). O principal modo de transmissão relatado pelos indivíduos foi por via sexual em relações heterossexuais (55,1%). A terapia antirretroviral (ART) incluía inibidores de transcriptase reversa e protease, sendo que nenhum dos indivíduos estava fazendo uso de antagonistas de CCR5 no momento da coleta. Em geral, os indivíduos estavam sob terapia antirretroviral (72,1%), apresentando contagem de células T CD4+ ≥ 350 células/ μ L (68,6%) e carga viral plasmática indetectável (53%) (Tabela 4).

Tabela 4 – Dados clínicos e epidemiológicos dos indivíduos incluídos neste estudo (continua)

		Total n=283 (%)
Média de idade (anos)		41,51
Sexo	Feminino	84 (29,68)
	Masculino	199 (70,32)
Modo de transmissão	Contato sanguíneo	11 (3,89)
	Heterossexual	156 (55,12)
	Homossexual	99 (34,98)
	UDIV ¹	5 (1,77)
	Vertical	1 (0,35)
	Desconhecido	11 (3,89)
Tratamento	Naive ²	79 (27,92)
	Uso de ART ³	204 (72,08)

Tabela 4 – Dados clínicos e epidemiológicos dos indivíduos incluídos neste estudo (continuação)

		Total n=283 (%)
Contagem de células T CD4+	< 350 células/ μ L	79 (27,92)
	$\geq$ 350 células/ μ L	194 (68,55)
	Desconhecido	10 (3,53)
Carga viral plasmática	Indetectável (< 50 cópias/mL)	150 (53,0)
	log=2	11 (3,89)
	log=3	46 (16,25)
	log=4	55 (19,43)
	log=5	11 (3,89)
	Desconhecido	10 (3,53)

Nota: ¹UDIV: usuário de drogas intravenosas. ²Naive: nunca exposto a terapia antirretroviral; ³ART: terapia antirretroviral

4.2 Conjuntos de amostras

Foram obtidas amostras de DNA de 283 indivíduos, sendo possível obter sequências da região V3 (C2-C3) para 265 (93,6%). Cinco amostras não apresentaram amplificação para a região V3 e as demais 13 não apresentaram sequências com qualidade adequada para análise. Foram enviadas para o laboratório colaborador *Monogram Biosciences* amostras de plasma de 99 pacientes (35%) para a realização do teste de tropismo fenotípico (*Trofile*), das quais 21 amostras não apresentaram resultado. Dentro deste grupo de 78 indivíduos com resultado de *Trofile*, não foi possível obter a sequência da região V3 para três amostras. Dentre estas 75 amostras (26,5%) com resultado de *Trofile* e sequenciamento da região V3, foram selecionadas 20 (7,1%) para a realização de sequenciamento massivo paralelo da gp120 do gene *env*.

Portanto, a amostragem deste estudo pode ser dividida em três subconjuntos, que para fins elucidativos, serão doravante denominadas conforme a seguir (Figura 9):

- **Conjunto A:** 265 amostras com sequenciamento da região V3;
- **Conjunto B:** 75 amostras com resultado de tropismo fenotípico e sequenciamento da região V3;
- **Conjunto C:** 20 amostras com sequenciamento da região gp120 do gene *env* e resultado de tropismo fenotípico.

Figura 9 – Representação gráfica dos subconjuntos de amostras utilizados neste estudo. Foram incluídos 283 indivíduos infectados por HIV-1. O **Conjunto A** corresponde às 265 sequências da região V3 (C2-C3) obtidas. O **Conjunto B** representa as 75 amostras com resultado de tropismo fenotípico e genotípico. O **Conjunto C** inclui 20 sequências da gp120 obtidas por sequenciamento massivo paralelo que apresentam resultado de tropismo fenotípico e genotípico.

4.3 Avaliação dos testes de tropismo

Dentre as 78 amostras das quais foi possível obter o resultado de tropismo fenotípico (*Trofile*), 55 (70,5%) apresentaram vírus com tropismo para CCR5, 22 (28,2%) apresentaram tropismo duplo e apenas uma amostra (1,3%) apresentou exclusivamente vírus X4 (Figura 10). Segundo informações fornecidas pela *Monogram Biosciences*, das 21 amostras que não apresentaram resultado para o ensaio *Trofile*, 16 (76,2%) falharam na fase de amplificação inicial do ensaio, duas falharam na fase de transformação, duas na fase de ensaio celular e uma amostra falhou devido a um sítio de restrição interno da sequência do envelope.

Figura 10 – Distribuição do tropismo fenotípico (*Trofile*). Foram obtidos 78 resultados do ensaio *Trofile*: 55 amostras foram classificadas como vírus R5, 22 R5X4 e uma X4.

As sequências do **Conjunto A** (n=265) foram submetidas à predição de tropismo utilizando o sistema *Geno2pheno*_[coreceptor] (G2P), com valor de falso positivo (FPR) igual a 10%, que classificou 190 (71,7%) como provenientes de vírus R5 e 75 (28,3%) como vírus capazes de utilizar o correceptor CXCR4 para a entrada na célula hospedeira (Figura 11).

Figura 11 – Distribuição do tropismo genotípico (*Geno2pheno*). Foram submetidas à predição de tropismo 265 sequências da região V3, das quais 190 foram classificadas como vírus R5 e 75 X4.

A avaliação comparativa entre os testes fenotípico e genotípico incluiu apenas amostras que apresentavam resultados para ambos os testes, perfazendo assim o **Conjunto B** (n=75). A Figura 12 demonstra que houve concordância entre os resultados da predição por G2P e do tropismo por *Trofile* em 74,7% (n=56) dos casos. A sensibilidade e especificidade do teste G2P foram de 80,7% (IC95% = 0,68 – 0,89) e 55,6% (IC95% = 0,30 – 0,78), respectivamente, com valor preditivo positivo de 0,85 (IC95% = 0,73 – 0,93) e valor preditivo negativo de 0,48 (IC95% = 0,26 – 0,70). Um artigo apresentando resultados parciais destas análises foi aceito para publicação na Revista do Instituto de Medicina Tropical de São Paulo (Apêndice C e Anexo C).

Figura 12 – Resultados de tropismo fenotípico (*Trofile*) e genotípico (*Geno2pheno*) das amostras do **Conjunto B**.

As sequências consenso do **Conjunto C** foram submetidas ao preditor G2P (FPR=10%). Os resultados da predição foram bastante divergentes dos resultados de tropismo fenotípico. Todas as sequências de vírus determinadas fenotipicamente como R5X4/X4 foram classificadas como vírus R5 pelo preditor, enquanto as que foram preditas como X4 eram todas provenientes de amostras contendo vírus R5 (Figura 13).

Figura 13 - Resultados de tropismo fenotípico (*Trofile*) e genotípico (*Geno2pheno*) das amostras do **Conjunto C**.

4.4 Caracterização molecular da gp120

4.4.1 Avaliação das sequências da região V3

A região da alça V3 é composta por 35 aminoácidos e flanqueada pelas regiões constantes C2 e C3. As sequências do **Conjunto A** (n=265) apresentaram comprimento médio de 307 pb, cobrindo toda a alça V3 e parte das regiões C2 e C3.. A posição 16 foi a mais variável, apresentando 16 resíduos diferentes, com prevalência da prolina (P) em 60,3% (n=160) dos casos, seguida de triptofano (W) em 29% (n=77) das variantes. Posições importantes para determinação do tropismo, como 11, 24 e 25 também apresentaram alta variabilidade com 10, 12 e 14 resíduos diferentes nas respectivas posições. As posições 1, 3, 28, 31 e 33 apresentaram até quatro variantes diferentes posições, demonstrando-se como as

posições mais conservadas da alça V3 e favorecendo as pontes dissulfeto e sua conformação em β hairpin.

A frequência do subtipo B foi predominante entre as sequências do **Conjunto A**, representando mais de 82% (n=219) de prevalência, seguida pelo subtipo F (10,9%, n=29), subtipo C (4,2%, n=5), algumas recombinações BF (1,9%, n=5) e apenas um representante do subtipo A (0,4%) (Figura 14). Dentre os representantes do subtipo B, o triptofano na posição 16, o que caracteriza o subtipo B'(GWGR), esteve presente em 34,7% (n=76) das sequências.

Figura 14 – Distribuição dos subtipos de HIV-1 na região V3 do **Conjunto A**.

4.4.2 Avaliação das sequências da região da gp120

O sequenciamento massivo paralelo da gp120 foi realizado utilizando a plataforma de sequenciamento *Ion Torrent* (Life Technologies, EUA). As 20 bibliotecas geradas por esta tecnologia foram distribuídas em quatro chips 314 com o objetivo de garantir uma cobertura mínima de 1000 x do produto de aproximadamente 3 Kb gerado por PCR para a região da gp120. O carregamento

médio dos chips foi de 62,5%, gerando a média de 385.424 *reads* com tamanho médio de 168 pb e policlonalidade de 33,75% (Anexo E).

Para a análise das 20 sequências da gp120, foram aplicadas diferentes estratégias para a montagem das *reads* obtidas: montagem *de novo*, mapeamento sobre uma sequência de referência e mapeamento sobre uma referência levando em consideração a diversidade genética das amostras utilizando o programa *Segminator II*.

A avaliação inicial partiu do alinhamento das 60 sequências geradas e a sequência de referência HXB2 através da ferramenta de alinhamento MUSCLE utilizando o programa MEGA 5 (Tamura et al., 2011). As sequências consenso por montagem *de novo* não apresentaram regiões de similaridade com a sequência referência HXB2, enquanto as sequências consenso geradas por mapeamento sobre a referência e caracterização da diversidade genética apresentaram grande similaridade e regiões com alta concentração de inserções/deleções (*indels*) nas posições correspondentes as regiões V2 (nucleotídeos 468 a 587) e V4 (nucleotídeos 1152 a 1253) da gp120. A cobertura das regiões V2 e V4 foi suficientemente baixa, a ponto de não ser possível definir a composição de nucleotídeos destas regiões, em sete sequências geradas por mapeamento contra referência. Assim, optou-se por utilizar as sequências consenso geradas pelo programa *Segminator II* para a realização das análises.

As sequências da alça V3 obtidas por ambas as metodologias de sequenciamento foram avaliadas filogeneticamente (Figura 15). As amostras provenientes de vírus R5X4 apresentaram agrupadas com vírus R5 e X4, não apresentando um agrupamento individualizado, como observado em algumas sequências de vírus R5. As duplas de sequências se agruparam a partir do mesmo nó ou em ramos bastante próximos.

Figura 15 – Árvore filogenética das sequências da alça V3 obtidas por sequenciamento clássico (x) e massivo paralelo (o). As sequências de vírus R5, R5X4 e X4 estão representadas em azul, verde e vermelho, respectivamente. Os círculos representam os valores de *bootstrap* colorido de acordo com a escala indicada à esquerda.

A Figura 16 ilustra a distribuição dos subtipos encontrados entre as sequências do **Conjunto C**, havendo representantes dos subtipos F1 (n=3) e B (n=17), das quais dez (58,8%) representam o subtipo B'(GWGR).

Figura 16 – Distribuição dos subtipos de HIV-1 na região da gp120 do **Conjunto C**.

As sequências apresentaram em média 22 sítios potenciais de N-glicosilação. As regiões C2 e C3 apresentaram maior concentração de sítios, com posições conservadas entre as amostras avaliadas. As regiões V1/V2 e V4 apresentaram distribuição de sítios em diferentes posições (Anexo F). O número de sítios de N-glicosilação não apresentou diferença estatística significativa entre sequências provenientes de vírus R5 e R5X4/X4 (Figura 17).

Figura 17 – Número de sítios potenciais de N-glicosilação preditos para sequências provenientes de vírus R5 e R5X4/X4.

A carga elétrica líquida foi calculada para a gp120 completa e também para a região V3 extraída a partir destas sequências. A média da carga elétrica líquida das sequências da gp120 foi de 13,8, não apresentando diferença estatística significativa entre sequências de vírus R5 e R5X4/X4 (Figura 18 A e B). A carga elétrica líquida dos resíduos da região V3 teve média de 5,1 e também não foi possível observar diferença estatística significativa entre os grupos classificados de acordo com o tropismo fenotípico.

Figura 18 – Carga elétrica líquida para sequências provenientes de vírus R5 e R5X4/X4. A) Carga elétrica líquida da gp120 B) Carga elétrica líquida região V3 da gp120.

4.5 Bayesian Network

As possíveis interações entre as variantes genéticas virais e o tropismo pelo correceptor, foram avaliadas por modelos probabilísticos de *Bayesian Network* (BN). As sequências do **Conjunto B** foram submetidas ao modelo de classificação do *B-Course v.2.0* (Myllymäki et al., 2002) para selecionar as variáveis mais relevantes para realizar a análise Bayesiana utilizando *B-Right*. Estas avaliações demonstraram que dados de sexo, modo de transmissão, uso de terapia antirretroviral, contagem de células T CD4+, carga viral e subtipo não apresentam relações diretas com tropismo e, portanto, foram excluídas da análise para diminuir o volume do processamento de dados. As variantes da região V3 que não apresentaram relação direta com o tropismo, também foram excluídas. Um conjunto de 1208 sequências provenientes do banco de dados de Los Alamos foi incluído na avaliação sendo 82,6% (n=997) de vírus R5, 11,4% (n=138) vírus R5X4 e 6% (n=73) de vírus X4.

Entre as sequências do **Conjunto B**, a avaliação por *Bayesian Network* (BN) não demonstrou dependência entre nenhuma variante da região V3 e o tropismo fenotípico (*Trofile*). Quando foi avaliado o tropismo genotípico (G2P), observou-se uma forte dependência entre vírus R5 e as posições e variantes 9S, 11R, 19V e 25K. Entre as sequências provenientes do banco de dados de Los Alamos, as posições e variantes 9R, 18Q, 11R, 24G e 25D também demonstraram uma forte dependência com vírus R5. Dependência entre mutações e vírus R5X4 ou X4 não foram observadas em nenhuma das abordagens testadas. Resultados parciais desta análise foram apresentados oralmente na 4^{ta}. *Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio)* (Apêndice D e Anexo D).

As sequências do **Conjunto C** também foram avaliadas por BN. Foram submetidos para avaliação dados de sexo, modo de transmissão, uso de terapia antirretroviral, contagem de células T CD4+, carga viral, subtipo, tropismo fenotípico e genotípico, número de sítios de N-glicosilação e carga elétrica líquida. Entretanto, o grande número de variáveis ao longo dos 511 resíduos gerou modelos probabilísticos impraticáveis. Devido ao grande número de possíveis interações, os programas *B-Course* e *B-Right* não foram capazes de estabelecer modelos de dependências entre as variáveis submetidas nesta análise. Na tentativa de verificar interações consistentes, as variáveis foram restringidas a região V3 destas sequências, porém, também não foi possível observar relações diretas entre as variáveis e os resultados de tropismo fenotípico e genotípico.

5 DISCUSSÃO

5.1 Avaliação dos testes de tropismo

Foram reunidas 99 amostras com carga viral plasmática > 1000 cópias/mL para a realização do ensaio de tropismo fenotípico (*Trofile*). O ensaio não apresentou resultados para 21,2% (n=22) das amostras provavelmente devido a características genéticas virais. A maior parte destas amostras (76,2%) falhou na amplificação inicial do ensaio, possivelmente devido a polimorfismos presentes nas regiões de hibridização dos *primers*, uma vez que as condições de carga viral plasmática e armazenamento das amostras seguiram as instruções da *Monogram Biosciences*, executora do teste.

O ensaio *Trofile* demonstrou prevalência de vírus R5 (70,5%, n=55), 28,2% (n=22) de vírus R5X4 e a presença exclusiva de vírus X4 foi encontrada em apenas uma amostra (1,3%). As mesmas frequências foram observadas no teste genotípico para predição de tropismo (*Geno2pheno*), porém, a comparação entre os resultados dos dois testes demonstrou discordância em 25,3% entre as amostras do **Conjunto B** (n=75). A prevalência de vírus R5 é esperada e vem sendo demonstrado que vírus R5 são responsáveis pelo efeito fundador no início da infecção e a presença de vírus X4 é encontrada em cerca de 50% dos pacientes que desenvolvem aids (Mild et al., 2013; Regoes e Bonhoeffer, 2005; Scarlatti et al., 1997). Foi demonstrado que a prevalência de vírus R5X4/X4 no início da infecção é em média de 12%, durante a fase crônica estes vírus foram encontrados em cerca de 17% dos casos, enquanto a infecção tardia apresentou frequências de 18% a 63% e média de 40%, indicando a prevalência de vírus R5 no início e a emergência de vírus X4 ao longo da infecção (Verhofstede et al., 2012). No presente estudo,

não foi possível estabelecer o tempo de infecção de todos os indivíduos devido à ausência de informações precisas nos prontuários, porém a frequência de 29,5% de vírus R5X4/X4, foi semelhante aos 26,2% encontrados em um estudo que também não considerou o estágio de infecção dos indivíduos infectados (Poveda et al., 2007). Estudos em coortes brasileiras demonstraram a presença de vírus X4 em 17% de indivíduos recentemente infectados, 63% em crianças e adolescentes e de 30% a 46% em populações de indivíduos adultos independente da fase de infecção (Almeida et al., 2013; Coelho et al., 2014; Sanabani et al., 2011; Sucupira et al., 2012).

Estudos comparativos entre testes fenotípicos e genotípicos de tropismo têm por objetivo estabelecer modelos alternativos mais baratos e eficientes para determinação do tropismo. A discordância entre os resultados é frequentemente relatada e deve-se principalmente a fatores relacionados com o desenho do estudo, características dos pacientes, subtipo viral e o ensaio utilizado para a determinação do tropismo (Verhofstede et al., 2012). A avaliação das amostras do **Conjunto B** (n=75) mostrou concordância entre os resultados em 75% dos casos, semelhante a outros estudos comparativos que demonstraram uma variação de concordância entre resultados fenotípicos e genotípicos entre 70% e 90% (Coelho et al., 2014; Kagan et al., 2014; Poveda et al., 2009; Raymond et al., 2008). Todos os métodos de determinação de tropismo, fenotípicos ou genotípicos apresentam limitações, que serão refletidas no resultado final. Apesar das diferenças entre os resultados, os testes genotípicos têm sido utilizados frequentemente na prática clínica por serem estratégias mais rápidas e baratas, propiciando a adequação do tratamento do paciente de uma maneira mais adequada (Obermeier et al., 2012; Vandekerckhove et al., 2011).

Um aspecto importante a ser considerado em relação a diferença entre os resultados do ensaio *Trofile* e da predição por G2P, é o fato de que para a

realização do ensaio *Trofile* foram utilizadas sequências obtidas a partir de vírus circulantes no plasma, enquanto as sequências submetidas para predição do tropismo (*Geno2pheno*) foram provenientes de DNA proviral, uma vez que 53% do indivíduos incluídos neste estudo apresentavam carga viral plasmática indetectável. Apesar de ambos os espécimes serem provenientes de uma mesma amostra de sangue total, esta a discordância dos resultados pode ser um reflexo tanto da compartimentalização das subpopulações virais, quanto da seleção de populações durante as etapas de amplificação dos ensaios fenotípico e genotípico, que por sua vez, utilizam sequências diferentes de *primers* em cada uma das respectivas reações. A avaliação filogenética demonstrou que algumas sequências provenientes da mesma amostra foram agrupadas em ramos diferentes, porém próximos, o que pode ser devido à seleção de variantes durante a amplificação por PCR, embora discreta.

Vários estudos foram realizados para avaliar o uso de DNA a partir de células mononucleares do sangue periférico (PMBCs) para testes de tropismo genotípico. A concordância entre resultados de RNA de HIV-1 plasmático e DNA proviral de PMBCs variou de 85% a 100%, demonstrando a contemporaneidade entre as populações virais nestes espécimes, no contexto de tropismo viral (Meini et al., 2014; Paar et al., 2011; Prosperi et al., 2010; Saracino et al., 2007). Verificou-se, ainda, que a produção de vírus R5 e X4 e as subpopulações virais encontradas no sangue periférico são produzidas por células com duração de vida semelhantes e não são geneticamente isoladas de um tipo particular de células (Ince et al., 2009). Em outro estudo observou-se que aproximadamente 20% das sequências provenientes de RNA plasmático foram preditas como vírus X4, enquanto mais de 50% das sequências provenientes de PBMCs foram preditas como X4, demonstrando o DNA proviral como uma fonte mais confiável de material genético viral para ensaios de tropismo tanto genotípicos, quanto fenotípicos (Verhofstede et

al., 2009). O uso do DNA proviral pode facilitar a execução de ensaios genotípicos de tropismo na rotina laboratorial, reduzindo a complexidade, o custo e o tempo do procedimento sem depender de cargas virais acima de 1000 cópias/mL..

5.2 Distribuição dos subtipos de HIV-1

Entre as sequências do **Conjunto A** (n=265) observou-se a prevalência do subtipo B (82%), seguido dos subtipos F (11%), C (4%), alguns recombinantes BF (2%) e um representante do subtipo A. Estas frequências são corroboradas por outros estudos brasileiros de epidemiologia molecular do HIV-1 que têm demonstrado a prevalência do subtipo B (70-95%), seguido do subtipo F1 (2-24%), as formas recombinantes circulantes (CRF – *Circulating Recombinant Forms*) CRF28_BF, CRF29_BF, CRF39_BF, CRF40_BF e CRF46_BF, além de um incontável número de formas recombinantes únicas (URF – *Unique Recombinant Forms*) envolvendo os subtipos B e F em quase todo o Brasil, exceto nos estados da região Sul, onde o subtipo C e a CRF31_BC predominam (Almeida et al., 2012; Bongertz et al., 2000; Sanabani et al., 2010; Tarosso et al., 2014; Velasco-de-Castro et al., 2014). Foram incluídas no **Conjunto C** (n=20) apenas sequências classificadas como subtipo B na região V3, porém, a avaliação da gp120 completa revelou a presença de três sequências do subtipo F1. A extensão média das sequências da região V3 foi de 307 nucleotídeos, apresentando baixo sinal filogenético para uma subtipagem precisa, o que pode ter levado a classificações incorretas em alguns casos.

A conformação β *hairpin* da ponta da alça V3 é formada por cerca de 13 aminoácidos, apresentando na região central (posições 15-18) um motivo altamente conservado GPG (glicina-prolina-glicina), conhecido como coroa (LaRosa et al.,

1990). Apesar da suprema prevalência do motivo GPG em sequências do subtipo B provenientes da Europa e EUA, alguns estudos têm relatado uma alta frequência (20-60%) da assinatura GWG (glicina-triptofano-glicina) em amostras brasileiras do subtipo B, denominada variante B' (Franca et al., 2011; Morgado et al., 1994; Pimentel et al., 2013). Foi observado que esta substituição de aminoácidos altera a estrutura secundária da ponta da alça V3 e reflete sobre propriedades biológicas e bioquímicas específicas (Morgado et al., 1996; Pimentel et al., 2013). Esta substituição também tem sido associada com a progressão da doença, indicando que indivíduos que apresentam a variante B' poderiam ter uma progressão mais lenta para aids (Casseb et al., 2002; de Brito et al., 2006; Pimentel et al., 2013; Santoro-Lopes et al., 2000). No presente estudo, a variante B' foi observada em aproximadamente 35% das sequências do subtipo B, sendo que distribuição entre vírus R5 e R5X4 foi semelhante e não foi observada nenhuma associação com dados clínicos ou de progressão da infecção. Apesar dos indicativos de que vírus B' (GWGR) apresentem a tendência de apresentar tropismo pelo correceptor CCR5 (Leal et al., 2008), curiosamente, a única amostra X4 deste estudo apresentou o motivo GWGR na ponta da alça V3. Este achado foi confirmado pelo sequenciamento massivo paralelo, mas por se tratar de apenas um representante, não foi possível definir se é apenas um caso isolado e se o motivo GW apresenta ou não relação com a escolha do tropismo.

5.3 Avaliação do sequenciamento da região da gp120

O sequenciamento da gp120 completa teve por objetivo buscar características moleculares fora da alça V3 que estivessem relacionadas com o tropismo. Foram sequenciadas 20 amostras utilizando a plataforma de

sequenciamento massivo paralelo *Ion Torrent* (Life Technologies, EUA) e foi possível observar uma esperada região de baixa cobertura, em todas as sequências, sobre as regiões V1/V2 e V4, conhecidas por apresentarem grande variação em sua extensão, como resultado de duplicações e/ou deleções em suas sequências, apresentando alta variação polimórfica quanto à composição de aminoácidos (Belair et al., 2009; Jansson et al., 2001; Li et al., 2014).

Alguns estudos demonstraram que apesar da considerável variabilidade genética, foram encontrados em média 25 sítios de N-glicosilação sobre a gp120, , sugerindo uma forte pressão seletiva para a manutenção deste número (Ho et al., 2008; Wei et al., 2003; Zhang et al., 2004). A predição de sítios potenciais de N-glicosilação, a partir das sequências consenso obtidas para cada amostra, revelou uma média de 22 sítios ao longo da sequência da gp120, semelhante ao resultado descrito em outros estudos. Houve um maior número de posições conservadas sobre as regiões C2 e C4 e apresentando distribuição semelhante em vírus R5 e R5X4/X4. Estes sacarídeos estão envolvidos em várias atividades como metabolismo, transporte, conservação da estrutura de proteínas, enovelamento proteico, reconhecimento de tipos celulares e adesão a outras células (Ho et al., 2008). O grau de conservação dos sítios de N-glicosilação é importante para a manutenção da sua função nas proteínas do envelope viral, formando um “escudo sacarídico” que representa um mecanismo para bloqueio ou diminuição da ação de anticorpos neutralizantes, promovendo a persistência viral e escape imune (Wei et al., 2003).

O aumento da carga elétrica líquida da alça V3 facilita o reconhecimento do correceptor CXCR4 da gp120 do HIV-1 através de uma interação eletrostática entre a alça V3 e o CXCR4 negativamente carregado promovendo a alteração do tropismo de vírus R5 para X4 (Brelot et al., 1999). Alguns estudos têm demonstrado que aproximadamente 90% dos vírus R5 apresentam carga elétrica menor que 4 na

alça V3, enquanto mais de 90% dos vírus X4 apresentam valores superiores a 4 (Jensen et al., 2006; Xiao et al., 1998; Xu et al., 2007). Esta delimitação não foi observada entre as sequências deste estudo, que apresentaram distribuições de carga elétrica semelhantes entre vírus R5 e R5X4/X4 tanto na alça V3 quanto em toda a gp120. Este fato pode ser devido às características das populações R5X4, que podem apresentar comportamento tanto de populações R5 quanto X4, dificultado a determinação de características que sejam exclusivas de cada população.

5.4 *Bayesian Network*

Uma rede Bayesiana ou *Bayesian Network* (BN) é um modelo probabilístico que descreve independências estatísticas entre variáveis múltiplas. O melhor modelo é aquele capaz de explicar um máximo de correlações observadas em um conjunto de dados, utilizando um número mínimo de influências diretas. As dependências são visualizadas em um gráfico acíclico direto a partir de variáveis qualitativas. Neste gráfico, cada nó corresponde a uma variável e cada arco direto entre nós representa uma influência direta (Deforche et al., 2007). Em estudos sobre o HIV-1, análises de BN têm sido aplicadas para definir vias de mutações de resistência contra antirretrovirais, compartimentalização e distância genética (Deforche et al., 2008; Eshleman et al., 2011; Ho et al., 2008)

Neste estudo, dados clínicos e sequências do **Conjunto B** (n=75) foram avaliados utilizando BN para buscar interações entre as variantes de resíduos e o tropismo viral. A avaliação utilizando dados de tropismo fenotípico (*Trofile*) não estabeleceu relações de dependência direta com nenhum resíduo ou posição das sequências de aminoácidos e, portanto, não foi possível definir características

genéticas que pudessem estar relacionadas com o tropismo. Entretanto, quando foram avaliados os resultados genotípicos (G2P) observou-se a presença de arcos diretos entre vírus R5 e variantes das posições 11, 24 e 25, conhecidas como definidoras de tropismo. Estes achados foram corroborados pela avaliação das sequências da região V3 provenientes do banco de dados de Los Alamos (www.hiv-web.lanl.gov). Uma hipótese possível seria que a população viral deste estudo apresentaria características genéticas distintas daquelas consideradas como definidoras de tropismo pelas ferramentas preditivas, indicando um resultado de tropismo tendencioso baseado apenas nas características de populações virais de outras regiões geográficas. Isso implica na subestimação das características genéticas da população estudada pelo preditor de tropismo, que durante a inferência da predição atribui pontuações mais relevantes para variantes prevalentes em outras populações virais, mas que não são encontradas na população viral amostrada. Um exemplo é a prolina na posição 16 (16P) apresentou frequência de 60% na população estudada, enquanto dentre as 1208 sequências de Los Alamos a frequência da 16P foi de 92%.

5.5 Considerações finais

A principal hipótese abordada neste estudo foi de que características moleculares além da alça V3 poderiam estar envolvidas com a determinação do tropismo do HIV-1. A inclusão do Maraviroque nas estratégias de tratamento antirretroviral promoveu uma série de estudos abordando o tropismo do HIV-1, com o objetivo de garantir a eficiência terapêutica máxima deste medicamento. Neste estudo foram utilizadas diferentes estratégias para avaliar a hipótese apresentada

porém, não foi possível identificar novas características definidoras de tropismo, além dos resíduos relacionados com a clássica regra 11/25.

Alguns fatores intrínsecos aos métodos utilizados podem ter influenciado diretamente os resultados deste estudo. Idealmente, as sequências avaliadas neste estudo deveriam ser obtidas a partir das mesmas amostras de plasma enviadas para a realização do ensaio *Trofile*, porém como um grande número de indivíduos apresentava carga viral plasmática indetectável e vários estudos demonstraram alta similaridade entre as populações virais em plasma e PBMCs, optou-se por utilizar DNA como material genético. Ainda assim, é possível que haja diferenças nos resultados de tropismo fenotípico e genotípico, uma vez que ocorre uma seleção de populações virais inerente aos ensaios utilizados.

Outro ponto que poderia contribuir no enriquecimento deste estudo, seria o sequenciamento massivo paralelo de um maior número de amostras, bem como uma melhor avaliação das sequências obtidas. Uma vez que a gp120 apresenta uma alta variabilidade genética, mapeamentos sobre uma sequência de referência podem implicar no descarte de leituras (*reads*) que não apresentam alta similaridade com a referência. As *reads* geradas neste estudo apresentaram em média 160 pb, o que, aliado a alta variabilidade presente, implica na dificuldade da montagem das *reads* de maneira eficiente e precisa. Quando foi realizada a montagem *de novo*, ou seja, sem uma sequência de referência, a maioria os *contigs* obtidos não foram suficientemente longos para cobrir toda a região sequenciada e as sequências consenso originadas a partir dos *contigs* mais longos não apresentaram similaridade com região da gp120. Atualmente, estão disponíveis *kits* da plataforma *Ion Torrent* capazes de gerar *reads* de até 400 pb o que contribui com a eficiência e confiabilidade da montagem das sequências.

Desta forma, foi necessário realizar o mapeamento sobre a sequência referência HXB2, que apresentou montagens mais fidedignas, porém, diminui a

amplitude da diversidade genética viral. O programa *Segminator II* foi desenvolvido visando genomas pequenos, principalmente genomas de vírus RNA (Archer et al., 2012). A estratégia de montagem é baseada em mapeamento sobre uma sequência referência, porém os algoritmos internos são mais flexíveis em relação à baixa frequência e baixa similaridade das *reads* em relação à sequência de referência. Este programa apresentou resultados mais adequados principalmente das regiões com intensa quantidade de inserções/deleções, como as regiões V2 e V4 da gp120. A análise de dados de sequenciamento massivo paralelo ainda é um desafio. Desde o advento das novas plataformas de sequenciamento (NGS) em meados dos anos 2000, diversas metodologias foram disponibilizadas e outras tantas estão sendo desenvolvidas, revolucionando o campo do sequenciamento genético (Vrancken et al., 2010). A montagem de genomas virais, em especial, vírus de RNA, é particularmente complexa devido à extrema variabilidade intra-hospedeiro e ao tamanho das *reads* geradas. Dependendo da montagem e dos parâmetros utilizados, pode-se tanto perder a precisão na montagem dos consensos, como subestimar a real variabilidade presente na amostra, principalmente variantes minoritárias.

As análises de dados de sequenciamento massivo paralelo podem contribuir para o entendimento das divergências encontradas entre testes fenotípicos e genotípicos e melhorar a sensibilidade de detecção de vírus X4 mais precocemente que os testes tradicionais. Apesar das avaliações realizadas neste estudo não terem abrangido todo o potencial que esta estratégia de sequenciamento propicia, os dados obtidos permitiram a avaliação de características genéticas da gp120 relacionadas ao tropismo do HIV-1.

6 CONCLUSÕES

- A avaliação do tropismo na população estudada demonstrou a prevalência de vírus R5 (71%) e uma concordância de 75% entre resultados de testes fenotípicos e genotípicos para a determinação do tropismo, demonstrando que ambas as metodologias são adequadas para aplicação na prática clínica, apesar de suas limitações implícitas.
- O sequenciamento massivo paralelo da gp120 demonstrou-se uma alternativa conveniente para a detecção da ampla variabilidade genética do HIV-1, embora não tenha sido possível explorar todo o potencial deste tipo método devido à complexidade das análises dos dados gerados.
- As estratégias de análise utilizadas neste estudo não permitiram a observação de relações significativas entre características moleculares da gp120 e o tropismo do HIV-1 pelos correceptores CCR5 e CXCR4.

REFERÊNCIAS

Almeida FJ, Zaparoli MS, Moreira DH, de Souza Cavalcanti J, Rodrigues R, Berezin EN, et al. Association of X4 tropism with disease progression in antiretroviral-treated children and adolescents living with HIV/AIDS in Sao Paulo, Brazil. *Braz J Infect Dis*. 2013 Nov 22.

Almeida SE, de Medeiros RM, Junqueira DM, Graf T, Passaes CP, Bello G, et al. Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil. *Viol J*. 2012;9:306.

Archer J, Baillie G, Watson SJ, Kellam P, Rambaut A, Robertson DL. Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. *BMC Bioinformatics*. 2012;13:47.

Archer J, Rambaut A, Taillon BE, Harrigan PR, Lewis M, Robertson DL. The evolutionary analysis of emerging low frequency HIV-1 CXCR4 using variants through time--an ultra-deep approach. *PLoS Comput Biol*. 2010;6(12):e1001022.

Belair M, Dovat M, Foley B, Mayerat C, Pantaleo G, Graziosi C. The polymorphic nature of HIV type 1 env V4 affects the patterns of potential N-glycosylation sites in proviral DNA at the intrahost level. *AIDS Res Hum Retroviruses*. 2009 Feb;25(2):199-206.

Berger EA, Doms RW, Fenyo EM, Korber BT, Littman DR, Moore JP, et al. A new classification for HIV-1. *Nature*. 1998 Jan 15;391(6664):240.

Bongertz V, Bou-Habib DC, Brigido LF, Caseiro M, Chequer PJ, Couto-Fernandez JC, et al. HIV-1 diversity in Brazil: genetic, biologic, and immunologic characterization of HIV-1 strains in three potential HIV vaccine evaluation sites. Brazilian Network for HIV Isolation and Characterization. *J Acquir Immune Defic Syndr*. 2000 Feb 1;23(2):184-93.

Brasil. Ministério da Saúde. Protocolo clínico e diretrizes terapêuticas para manejo da infecção pelo HIV em adultos. Brasília: Ministério da Saúde; 2013. 216p.

Braun P, Wiesmann F. Phenotypic assays for the determination of coreceptor tropism in HIV-1 infected individuals. *Eur J Med Res*. 2007 Oct 15;12(9):463-72.

Brelot A, Heveker N, Adema K, Hosie MJ, Willett B, Alizon M. Effect of mutations in the second extracellular loop of CXCR4 on its utilization by human and feline immunodeficiency viruses. *J Virol*. 1999 Apr;73(4):2576-86.

Brieu N, Portales P, Carles MJ, Corbeau P. Interleukin-7 induces HIV type 1 R5-to-X4 switch. *Blood*. 2011 Feb 10;117(6):2073-4.

Brun-Vezinet F, Charpentier C. Update on the human immunodeficiency virus. *Med Mal Infect*. 2013 May;43(5):177-84.

Bunnik EM, Swenson LC, Edo-Matas D, Huang W, Dong W, Frantzell A, et al. Detection of inferred CCR5- and CXCR4-using HIV-1 variants and evolutionary intermediates using ultra-deep pyrosequencing. *PLoS Pathog.* 2011 Jun;7(6):e1002106.

Caseiro MM, Nelson M, Diaz RS, Gathe J, de Andrade Neto JL, Slim J, et al. Vicriviroc plus optimized background therapy for treatment-experienced subjects with CCR5 HIV-1 infection: final results of two randomized phase III trials. *J Infect.* 2012 Oct;65(4):326-35.

Casseb J, Komninakis S, Abdalla L, Brigido LF, Rodrigues R, Araujo F, et al. HIV disease progression: is the Brazilian variant subtype B' (GWGR motif) less pathogenic than US/European subtype B (GPGR)? *Int J Infect Dis.* 2002 Sep;6(3):164-9.

Checkley MA, Luttge BG, Freed EO. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J Mol Biol.* 2011 Jul 22;410(4):582-608.

Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, Wetter T, et al. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* 2004 Jun;14(6):1147-59.

Chueca N, Garrido C, Alvarez M, Poveda E, de Dios Luna J, Zahonero N, et al. Improvement in the determination of HIV-1 tropism using the V3 gene sequence and a combination of bioinformatic tools. *J Med Virol.* 2009 May;81(5):763-7.

Clevestig P, Pramanik L, Leitner T, Ehrnst A. CCR5 use by human immunodeficiency virus type 1 is associated closely with the gp120 V3 loop N-linked glycosylation site. *J Gen Virol.* 2006 Mar;87(Pt 3):607-12.

Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science.* 1995 Dec 15;270(5243):1811-5.

Coelho LP, Ferreira JL, Cabral GB, Guimaraes PM, Brigido LF. Genotypic Tropism Prediction from Paired Cell and Plasma Using Single and Replicates Sequences. *AIDS Res Hum Retroviruses.* 2014 Mar 28.

Collman R, Hassan NF, Walker R, Godfrey B, Cutilli J, Hastings JC, et al. Infection of monocyte-derived macrophages with human immunodeficiency virus type 1 (HIV-1). Monocyte-tropic and lymphocyte-tropic strains of HIV-1 show distinctive patterns of replication in a panel of cell types. *J Exp Med.* 1989 Oct 1;170(4):1149-63.

Couto-Fernandez JC, Janssens W, Heyndrickx L, Motte J, Fransen K, Peeters M, et al. Genetic and antigenic variability of HIV type 1 in Brazil. *AIDS Res Hum Retroviruses.* 1994 Sep;10(9):1157-63.

Cullen B. Species and Tissue Tropisms of HIV-1: Molecular Basis and Phenotypic Consequences. *HIV Sequence Compendium.* 2001:1-12.

Curlin ME, Zioni R, Hawes SE, Liu Y, Deng W, Gottlieb GS, et al. HIV-1 envelope subregion length variation during disease progression. *PLoS Pathog.* 2010;6(12):e1001228.

da Costa SM, Schechter M, Shindo N, Vicente AC, Oliveira EF, Pinto ME, et al. Sequence and phylogenetic analysis of glycoprotein 120 of an HIV type 1 variant (GWGR) prevalent in Brazil. *AIDS Res Hum Retroviruses*. 1995 Sep;11(9):1143-5.

Dash PK, Siddappa NB, Mangaiarkarasi A, Mahendarkar AV, Roshan P, Anand KK, et al. Exceptional molecular and coreceptor-requirement properties of molecular clones isolated from an Human Immunodeficiency Virus Type-1 subtype C infection. *Retrovirology*. 2008;5:25.

de Brito A, Komninakis SC, Novoa P, de Oliveira RM, Fonseca LA, Duarte AJ, et al. Women infected with HIV type 1 Brazilian variant, subtype B (B'-GWGR motif) have slower progression to AIDS, compared with patients infected with subtype B (B-GPGR motif). *Clin Infect Dis*. 2006 Dec 1;43(11):1476-81.

de Mendoza C, Van Baelen K, Poveda E, Rondelez E, Zahonero N, Stuyver L, et al. Performance of a population-based HIV-1 tropism phenotypic assay and correlation with V3 genotypic prediction tools in recent HIV-1 seroconverters. *J Acquir Immune Defic Syndr*. 2008 Jul 1;48(3):241-4.

Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science*. 1996 Sep 27;273(5283):1856-62.

Deforche K, Camacho R, Grossman Z, Silander T, Soares MA, Moreau Y, et al. Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. *Infect Genet Evol*. 2007 Jun;7(3):382-90.

Deforche K, Camacho RJ, Grossman Z, Soares MA, Van Laethem K, Katzenstein DA, et al. Bayesian network analyses of resistance pathways against efavirenz and nevirapine. *AIDS*. 2008 Oct 18;22(16):2107-15.

Deforche K, Silander T, Camacho R, Grossman Z, Soares MA, Van Laethem K, et al. Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. *Bioinformatics*. 2006 Dec 15;22(24):2975-9.

Delwart EL, Herring B, Rodrigo AG, Mullins JI. Genetic subtyping of human immunodeficiency virus using a heteroduplex mobility assay. *PCR Methods Appl*. 1995 Apr;4(5):S202-16.

Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, et al. Identification of a major co-receptor for primary isolates of HIV-1. *Nature*. 1996 Jun 20;381(6584):661-6.

Dorr P, Westby M, Dobbs S, Griffin P, Irvine B, Macartney M, et al. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob Agents Chemother*. 2005 Nov;49(11):4721-32.

Emmelkamp JM, Rockstroh JK. CCR5 antagonists: comparison of efficacy, side effects, pharmacokinetics and interactions--review of the literature. *Eur J Med Res*. 2007 Oct 15;12(9):409-17.

Eshleman SH, Hudelson SE, Redd AD, Wang L, Debes R, Chen YQ, et al. Analysis of genetic linkage of HIV from couples enrolled in the HIV Prevention Trials Network 052 trial. *J Infect Dis*. 2011 Dec 15;204(12):1918-26.

Este JA, Telenti A. HIV entry inhibitors. *Lancet*. 2007 Jul 7;370(9581):81-8.

Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*. 1996 May 10;272(5263):872-7.

Ferraro GA, Mello MA, Suttmoller F, Van Weyenbergh J, Shindo N, Galvao-Castro B, et al. Biological characterization and chemokine receptor usage of HIV type 1 isolates prevalent in Brazil. *AIDS Res Hum Retroviruses*. 2001 Sep 1;17(13):1241-7.

Fouchier RA, Groenink M, Kootstra NA, Tersmette M, Huisman HG, Miedema F, et al. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol*. 1992 May;66(5):3183-7.

Franca RF, Castro-Jorge LA, Neto RJ, Jorge DM, Lima DM, Colares JK, et al. Genotypic characteristics of HIV type 1 based on gp120 hypervariable region 3 of isolates from Southern Brazil. *AIDS Res Hum Retroviruses*. 2011 Aug;27(8):903-9.

Freed EO, Myers DJ, Risser R. Identification of the principal neutralizing determinant of human immunodeficiency virus type 1 as a fusion domain. *J Virol*. 1991 Jan;65(1):190-4.

Goetz MB, Leduc R, Kostman JR, Labriola AM, Lie Y, Weidler J, et al. Relationship between HIV coreceptor tropism and disease progression in persons with untreated chronic HIV infection. *J Acquir Immune Defic Syndr*. 2009 Mar 1;50(3):259-66.

Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, et al. *Pneumocystis carinii* pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med*. 1981 Dec 10;305(24):1425-31.

Haqqani AA, Tilton JC. Entry inhibitors and their use in the treatment of HIV-1 infection. *Antiviral Res*. 2013 May;98(2):158-70.

Ho YS, Abecasis AB, Theys K, Deforche K, Dwyer DE, Charleston M, et al. HIV-1 gp120 N-linked glycosylation differs between plasma and leukocyte compartments. *Virol J*. 2008;5:14.

Hoffmann C. The epidemiology of HIV coreceptor tropism. *Eur J Med Res*. 2007 Oct 15;12(9):385-90.

Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, et al. Structure of a V3-containing HIV-1 gp120 core. *Science*. 2005 Nov 11;310(5750):1025-8.

Hughes A, Nelson M. HIV entry: new insights and implications for patient management. *Curr Opin Infect Dis.* 2009 Feb;22(1):35-42.

Hwang SS, Boyle TJ, Lyerly HK, Cullen BR. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science.* 1991 Jul 5;253(5015):71-4.

Ince WL, Harrington PR, Schnell GL, Patel-Chhabra M, Burch CL, Menezes P, et al. Major coexisting human immunodeficiency virus type 1 env gene subpopulations in the peripheral blood are produced by cells with similar turnover rates and show little evidence of genetic compartmentalization. *J Virol.* 2009 May;83(9):4068-80.

Iwamoto A, Hosoya N, Kawana-Tachikawa A. HIV-1 tropism. *Protein Cell.* 2010 Jun;1(6):510-3.

Jansson M, Backstrom E, Scarlatti G, Bjorndal A, Matsuda S, Rossi P, et al. Length variation of glycoprotein 120 V2 region in relation to biological phenotypes and coreceptor usage of primary HIV type 1 isolates. *AIDS Res Hum Retroviruses.* 2001 Oct 10;17(15):1405-14.

Jensen MA, Coetzer M, van 't Wout AB, Morris L, Mullins JI. A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *J Virol.* 2006 May;80(10):4698-704.

Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J Virol.* 2003 Dec;77(24):13376-88.

Jensen MA, van 't Wout AB. Predicting HIV-1 coreceptor usage with sequence analysis. *AIDS Rev.* 2003 Apr-Jun;5(2):104-12.

Jiang X, Burke V, Totrov M, Williams C, Cardozo T, Gorny MK, et al. Conserved structural elements in the V3 crown of HIV-1 gp120. *Nat Struct Mol Biol.* 2010 Aug;17(8):955-61.

Kagan RM, Johnson EP, Siaw MF, Van Baelen B, Ogden R, Platt JL, et al. Comparison of genotypic and phenotypic HIV type 1 tropism assay: results from the screening samples of Cenicriviroc Study 202, a randomized phase II trial in treatment-naïve subjects. *AIDS Res Hum Retroviruses.* 2014 Feb;30(2):151-9.

Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science.* 2006 Jul 28;313(5786):523-6.

Klibanov OM, Williams SH, Iler CA. Cenicriviroc, an orally active CCR5 antagonist for the potential treatment of HIV infection. *Curr Opin Investig Drugs.* 2010 Aug;11(8):940-50.

Kowalski M, Potz J, Basiripour L, Dorfman T, Goh WC, Terwilliger E, et al. Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science.* 1987 Sep 11;237(4820):1351-5.

LaRosa GJ, Davide JP, Weinhold K, Waterbury JA, Profy AT, Lewis JA, et al. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. *Science*. 1990 Aug 24;249(4971):932-5.

Leal E, Silva WP, Sucupira MC, Janini LM, Diaz RS. Molecular and structural characterization of HIV-1 subtype B Brazilian isolates with GWGR tetramer at the tip of the V3-loop. *Virology*. 2008 Nov 25;381(2):222-9.

Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol*. 2007 Dec;25(12):1407-10.

Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem*. 1990 Jun 25;265(18):10373-82.

Li Y, Yang D, Wang JY, Yao Y, Zhang WZ, Wang LJ, et al. Critical Amino Acids within the Human Immunodeficiency Virus Type 1 Envelope Glycoprotein V4 N- and C-Terminals Contribute to Virus Entry. *PLoS One*. 2014 Jan 21;9(1):e86083.

Lin NH, Kuritzkes DR. Tropism testing in the clinical management of HIV-1 infection. *Curr Opin HIV AIDS*. 2009 Nov;4(6):481-7.

Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*. 1996 Aug 9;86(3):367-77.

Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol*. 1999 Jan;73(1):152-60.

Low AJ, Dong W, Chan D, Sing T, Swanstrom R, Jensen M, et al. Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. *AIDS*. 2007 Sep 12;21(14):F17-24.

Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*. 1986 Nov 7;47(3):333-48.

Marier JF, Trinh M, Pheng LH, Palleja SM, Martin DE. Pharmacokinetics and pharmacodynamics of TBR-652, a novel CCR5 antagonist, in HIV-1-infected, antiretroviral treatment-experienced, CCR5 antagonist-naïve patients. *Antimicrob Agents Chemother*. 2011 Jun;55(6):2768-74.

Meini G, Rossetti B, Bianco C, Ceccherini-Silberstein F, Di Giambenedetto S, Sighinolfi L, et al. Longitudinal analysis of HIV-1 coreceptor tropism by single and triplicate HIV-1 RNA and DNA sequencing in patients undergoing successful first-line antiretroviral therapy. *J Antimicrob Chemother*. 2014 Mar;69(3):735-41.

Mild M, Gray RR, Kvist A, Lemey P, Goodenow MM, Fenyo EM, et al. High inpatient HIV-1 evolutionary rate is associated with CCR5-to-CXCR4 coreceptor switch. *Infect Genet Evol.* 2013 Oct;19:369-77.

Mild M, Kvist A, Esbjornsson J, Karlsson I, Fenyo EM, Medstrand P. Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection. *Infect Genet Evol.* 2009 May 14.

Mild M, Kvist A, Esbjornsson J, Karlsson I, Fenyo EM, Medstrand P. Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection. *Infect Genet Evol.* 2010 Apr;10(3):356-64.

Morgado MG, Guimaraes ML, Gripp CB, Neves Junior I, Costa CI, dos Santos VG, et al. Polymorphism of the predictive antigenic sites on the V3 loop of Brazilian HIV-1 subtype B strains. HEC/FIOCRUZ AIDS Clinical Research Group. *Mem Inst Oswaldo Cruz.* 1996 May-Jun;91(3):339-42.

Morgado MG, Sabino EC, Shpaer EG, Bongertz V, Brigido L, Guimaraes MD, et al. V3 region polymorphisms in HIV-1 from Brazil: prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. *AIDS Res Hum Retroviruses.* 1994 May;10(5):569-76.

Mueller MC, Bogner JR. Treatment with CCR5 antagonists: which patient may have a benefit? *Eur J Med Res.* 2007 Oct 15;12(9):441-52.

Myllymäki P, Silander T, Tirri H, Uronen P. B-Course: A Web-Based Tool for Bayesian and Causal Data Analysis. *International Journal on Artificial Intelligence Tools.* 2002;11(3):18.

Nichols WG, Steel HM, Bonny T, Adkison K, Curtis L, Millard J, et al. Hepatotoxicity observed in clinical trials of aplavivoc (GW873140). *Antimicrob Agents Chemother.* 2008 Mar;52(3):858-65.

Obermeier M, Symons J, Wensing AM. HIV population genotypic tropism testing and its clinical significance. *Curr Opin HIV AIDS.* 2012 Sep;7(5):470-7.

Ogert RA, Lee MK, Ross W, Buckler-White A, Martin MA, Cho MW. N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. *J Virol.* 2001 Jul;75(13):5998-6006.

Olshevsky U, Helseth E, Furman C, Li J, Haseltine W, Sodroski J. Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. *J Virol.* 1990 Dec;64(12):5701-7.

Paar C, Geit M, Stekel H, Berg J. Genotypic prediction of human immunodeficiency virus type 1 tropism by use of plasma and peripheral blood mononuclear cells in the routine clinical laboratory. *J Clin Microbiol.* 2011 Jul;49(7):2697-9.

Palmer C, Balfe P, Fox D, May JC, Frederiksson R, Fenyo EM, et al. Functional characterization of the V1V2 region of human immunodeficiency virus type 1. *Virology.* 1996 Jun 15;220(2):436-49.

Peeters M. Recombinant HIV sequences: their role in the global epidemic. HIV Sequence Compendium 2000. Los Alamos: Los Alamos National Laboratory; 2000. 15p.

Pillai S, Good B, Richman D, Corbeil J. A new perspective on V3 phenotype prediction. *AIDS Res Hum Retroviruses*. 2003 Feb;19(2):145-9.

Pimentel VF, Morgado MG, Bello G, Guimaraes MD, Castilho EA, Veloso VG, et al. Temporal trends and molecular epidemiology of HIV type 1 infection in Rio de Janeiro, Brazil. *AIDS Res Hum Retroviruses*. 2013 Dec;29(12):1553-61.

Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemee V, et al. A new human immunodeficiency virus derived from gorillas. *Nat Med*. 2009 Aug;15(8):871-2.

Pollakis G, Abebe A, Kliphuis A, Chalaby MI, Bakker M, Mengistu Y, et al. Phenotypic and genotypic comparisons of CCR5- and CXCR4-tropic human immunodeficiency virus type 1 biological clones isolated from subtype C-infected individuals. *J Virol*. 2004 Mar;78(6):2841-52.

Poon AF, Lewis FI, Pond SL, Frost SD. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLoS Comput Biol*. 2007 Nov;3(11):e231.

Poveda E, Briz V, de Mendoza C, Benito JM, Corral A, Zahonero N, et al. Prevalence of X4 tropic HIV-1 variants in patients with differences in disease stage and exposure to antiretroviral therapy. *J Med Virol*. 2007 Aug;79(8):1040-6.

Poveda E, Seclen E, Gonzalez Mdel M, Garcia F, Chueca N, Aguilera A, et al. Design and validation of new genotypic tools for easy and reliable estimation of HIV tropism before using CCR5 antagonists. *J Antimicrob Chemother*. 2009 May;63(5):1006-10.

Prosperi MC, Bracciale L, Fabbiani M, Di Giambenedetto S, Razzolini F, Meini G, et al. Comparative determination of HIV-1 co-receptor tropism by Enhanced Sensitivity Trofile, gp120 V3-loop RNA and DNA genotyping. *Retrovirology*. 2010;7:56.

Raymond S, Delobel P, Mavigner M, Cazabat M, Souyris C, Sandres-Saune K, et al. Correlation between genotypic predictions based on V3 sequences and phenotypic determination of HIV-1 tropism. *AIDS*. 2008 Sep 12;22(14):F11-6.

Raymond S, Recordon-Pinson P, Saliou A, Delobel P, Nicot F, Descamps D, et al. Improved V3 genotyping with duplicate PCR amplification for determining HIV-1 tropism. *J Antimicrob Chemother*. 2011 Sep;66(9):1972-5.

Recordon-Pinson P, Soulie C, Flandre P, Descamps D, Lazrek M, Charpentier C, et al. Evaluation of the genotypic prediction of HIV-1 coreceptor use versus a phenotypic assay and correlation with the virological response to maraviroc: the ANRS GenoTropism study. *Antimicrob Agents Chemother*. 2010 Aug;54(8):3335-40.

Regoes RR, Bonhoeffer S. The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol*. 2005 Jun;13(6):269-77.

Repits J, Sterjovski J, Badia-Martinez D, Mild M, Gray L, Churchill MJ, et al. Primary HIV-1 R5 isolates from end-stage disease display enhanced viral fitness in parallel with increased gp120 net charge. *Virology*. 2008 Sep 15;379(1):125-34.

Resch W, Hoffman N, Swanstrom R. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. *Virology*. 2001 Sep 15;288(1):51-62.

Rizzuto CD, Wyatt R, Hernandez-Ramos N, Sun Y, Kwong PD, Hendrickson WA, et al. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science*. 1998 Jun 19;280(5371):1949-53.

Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, et al. HIV-1 nomenclature proposal. *Science*. 2000 Apr 7;288(5463):55-6.

Rosen O, Sharon M, Quadt-Akabayov SR, Anglist J. Molecular switch for alternative conformations of the HIV-1 V3 region: implications for phenotype conversion. *Proc Natl Acad Sci U S A*. 2006 Sep 19;103(38):13950-5.

Sagar M, Wu X, Lee S, Overbaugh J. Human immunodeficiency virus type 1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. *J Virol*. 2006 Oct;80(19):9586-98.

Sanabani SS, Pastena ER, da Costa AC, Martinez VP, Kleine-Neto W, de Oliveira AC, et al. Characterization of partial and near full-length genomes of HIV-1 strains sampled from recently infected individuals in Sao Paulo, Brazil. *PLoS One*. 2011;6(10):e25869.

Sanabani SS, Pastena ER, Neto WK, Martinez VP, Sabino EC. Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in Sao Paulo, Brazil. *Virol J*. 2010;7:74.

Sander O, Sing T, Sommer I, Low AJ, Cheung PK, Harrigan PR, et al. Structural descriptors of gp120 V3 loop for the prediction of HIV-1 coreceptor usage. *PLoS Comput Biol*. 2007 Mar 30;3(3):e58.

Santoro-Lopes G, Harrison LH, Tavares MD, Xexeo A, Dos Santos AC, Schechter M. HIV disease progression and V3 serotypes in Brazil: is B different from B-Br? *AIDS Res Hum Retroviruses*. 2000 Jul 1;16(10):953-8.

Saracino A, Monno L, Cibelli DC, Punzi G, Brindicci G, Ladisa N, et al. Co-receptor switch during HAART is independent of virological success. *J Med Virol*. 2009 Dec;81(12):2036-44.

Saracino A, Monno L, Punzi G, Cibelli DC, Tartaglia A, Scudeller L, et al. HIV-1 biological phenotype and predicted coreceptor usage based on V3 loop sequence in paired PBMC and plasma samples. *Virus Res*. 2007 Dec;130(1-2):34-42.

Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK, et al. In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nat Med*. 1997 Nov;3(11):1259-65.

Schreiber M, Wachsmuth C, Muller H, Hagen C, Schmitz H, van Lunzen J. Loss of antibody reactivity directed against the V3 domain of certain human immunodeficiency virus type 1 variants during disease progression. *J Gen Virol*. 1996 Oct;77 (Pt 10):2403-14.

Schuitmaker H, Koot M, Kootstra NA, Dercksen MW, de Goede RE, van Steenwijk RP, et al. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocyctotropic to T-cell-tropic virus population. *J Virol*. 1992 Mar;66(3):1354-60.

Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol*. 1999 Dec;73(12):10489-502.

Shioda T, Oka S, Xin X, Liu H, Harukuni R, Kurotani A, et al. In vivo sequence variability of human immunodeficiency virus type 1 envelope gp120: association of V2 extension with slow disease progression. *J Virol*. 1997 Jul;71(7):4871-81.

Sierra S, Kaiser R, Thielen A, Lengauer T. Genotypic coreceptor analysis. *Eur J Med Res*. 2007 Oct 15;12(9):453-62.

Simon-Loriere E, Galetto R, Hamoudi M, Archer J, Lefeuvre P, Martin DP, et al. Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. *PLoS Pathog*. 2009 May;5(5):e1000418.

Sing T, Low AJ, Beerenwinkel N, Sander O, Cheung PK, Domingues FS, et al. Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. *Antivir Ther*. 2007;12(7):1097-106.

Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. *Virus Res*. 2012 Nov;169(2):415-29.

Soulie C, Fourati S, Lambert-Niclot S, Malet I, Wirden M, Tubiana R, et al. Factors associated with proviral DNA HIV-1 tropism in antiretroviral therapy-treated patients with fully suppressed plasma HIV viral load: implications for the clinical use of CCR5 antagonists. *J Antimicrob Chemother*. 2010 Apr;65(4):749-51.

Su Z, Gulick RM, Krambrink A, Coakley E, Hughes MD, Han D, et al. Response to vicriviroc in treatment-experienced subjects, as determined by an enhanced-sensitivity coreceptor tropism assay: reanalysis of AIDS clinical trials group A5211. *J Infect Dis*. 2009 Dec 1;200(11):1724-8.

Subbarao S, Vanichseni S, Hu DJ, Kitayaporn D, Choopanya K, Raktham S, et al. Genetic characterization of incident HIV type 1 subtype E and B strains from a prospective cohort of injecting drug users in Bangkok, Thailand. *AIDS Res Hum Retroviruses*. 2000 May 20;16(8):699-707.

Sucupira MC, Sanabani S, Cortes RM, Giret MT, Tomiyama H, Sauer MM, et al. Faster HIV-1 disease progression among Brazilian individuals recently infected with CXCR4-utilizing strains. *PLoS One*. 2012;7(1):e30292.

Swenson LC, Mo T, Dong WW, Zhong X, Woods CK, Jensen MA, et al. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. *J Infect Dis*. 2011 Jan 15;203(2):237-45.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011 Oct;28(10):2731-9.

Tarosso LF, Sanabani SS, Ribeiro SP, Sauer MM, Tomiyama HI, Sucupira MC, et al. Short communication: HIV type 1 subtype BF leads to faster CD4+ T cell loss compared to subtype B. *AIDS Res Hum Retroviruses*. 2014 Feb;30(2):190-4.

Trabaud MA, Icard V, Scholtes C, Perpoint T, Koffi J, Cotte L, et al. Discordance in HIV-1 Co-receptor use prediction by different genotypic algorithms and phenotype assay: Intermediate profile in relation to concordant predictions. *J Med Virol*. 2012 Mar;84(3):402-13.

Trouplin V, Salvatori F, Cappello F, Obry V, Brelot A, Heveker N, et al. Determination of coreceptor usage of human immunodeficiency virus type 1 from patient plasma samples by using a recombinant phenotypic assay. *J Virol*. 2001 Jan;75(1):251-9.

Tscherning C, Alaeus A, Fredriksson R, Bjorndal A, Deng H, Littman DR, et al. Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology*. 1998 Feb 15;241(2):181-8.

Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, Leitner T, et al. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. *PLoS One*. 2009;4(5):e5683.

Turner BG, Summers MF. Structural biology of HIV. *J Mol Biol*. 1999 Jan 8;285(1):1-32.

Van Baelen K, Vandenbroucke I, Rondelez E, Van Eygen V, Vermeiren H, Stuyver LJ. HIV-1 coreceptor usage determination in clinical isolates using clonal and population-based genotypic and phenotypic assays. *J Virol Methods*. 2007 Dec;146(1-2):61-73.

Van Laethem K, Schrooten Y, Lemey P, Van Wijngaerden E, De Wit S, Van Ranst M, et al. A genotypic resistance assay for the detection of drug resistance in the human immunodeficiency virus type 1 envelope gene. *J Virol Methods*. 2005 Jan;123(1):25-34.

Vandekerckhove LP, Wensing AM, Kaiser R, Brun-Vezinet F, Clotet B, De Luca A, et al. European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect Dis*. 2011 May;11(5):394-407.

Vandenbroucke I, Eygen VV, Rondelez E, Vermeiren H, Baelen KV, Stuyver LJ. Minor Variant Detection at Different Template Concentrations in HIV-1 Phenotypic and Genotypic Tropism Testing. *Open Virol J*. 2008;2:8-14.

Velasco-de-Castro CA, Grinsztejn B, Veloso VG, Bastos FI, Pilotto JH, Fernandes N, et al. HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary counseling and testing sites in Rio de Janeiro, Brazil. *PLoS One*. 2014;9(1):e87622.

Verhofstede C, Nijhuis M, Vandekerckhove L. Correlation of coreceptor usage and disease progression. *Curr Opin HIV AIDS*. 2012 Sep;7(5):432-9.

Verhofstede C, Vandekerckhove L, Eygen VV, Demecheleer E, Vandenbroucke I, Winters B, et al. CXCR4-using HIV type 1 variants are more commonly found in peripheral blood mononuclear cell DNA than in plasma RNA. *J Acquir Immune Defic Syndr*. 2009 Feb 1;50(2):126-36.

Vrancken B, Lequime S, Theys K, Lemey P. Covering all bases in HIV research: unveiling a hidden world of viral evolution. *AIDS Rev*. 2010 Apr-Jun;12(2):89-102.

Wang N, Zhu T, Ho DD. Sequence diversity of V1 and V2 domains of gp120 from human immunodeficiency virus type 1: lack of correlation with viral phenotype. *J Virol*. 1995 Apr;69(4):2708-15.

Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, et al. Antibody neutralization and escape by HIV-1. *Nature*. 2003 Mar 20;422(6929):307-12.

Whitcomb JM, Huang W, Fransen S, Limoli K, Toma J, Wrin T, et al. Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. *Antimicrob Agents Chemother*. 2007 Feb;51(2):566-75.

Wilén CB, Tilton JC, Doms RW. HIV: cell binding and entry. *Cold Spring Harb Perspect Med*. 2012;2(8).

Wilkin TJ, Gulick RM. CCR5 antagonism in HIV infection: current concepts and future opportunities. *Annu Rev Med*. 2012;63:81-93.

Willey RL, Rutledge RA, Dias S, Folks T, Theodore T, Buckler CE, et al. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. *Proc Natl Acad Sci U S A*. 1986 Jul;83(14):5038-42.

Xiao L, Owen SM, Goldman I, Lal AA, deJong JJ, Goudsmit J, et al. CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global HIV-1 isolates: delineation of consensus motif in the V3 domain that predicts CCR-5 usage. *Virology*. 1998 Jan 5;240(1):83-92.

Xu S, Huang X, Xu H, Zhang C. Improved prediction of coreceptor usage and phenotype of HIV-1 based on combined features of V3 loop sequence using random forest. *J Microbiol*. 2007 Oct;45(5):441-6.

Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, et al. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology*. 2004 Dec;14(12):1229-46.

Zimmerman PA, Buckler-White A, Alkhatib G, Spalding T, Kubofcik J, Combadiere C, et al. Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. *Mol Med.* 1997 Jan;3(1):23-36.

Zolla-Pazner S. Identifying epitopes of HIV-1 that induce protective antibodies. *Nat Rev Immunol.* 2004 Mar;4(3):199-210.

APÊNDICES

Apêndice A – Termo de Consentimento Livre e Esclarecido

**HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA DA UNIVERSIDADE
DE SÃO PAULO-HCFMUSP
TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
Processo nº 0108/08**

**DADOS DE IDENTIFICAÇÃO DO SUJEITO DA PESQUISA OU RESPONSÁVEL
LEGAL**

1. NOME:
DOCUMENTO DE IDENTIDADE Nº : SEXO : .M ☐ F ☐
DATA NASCIMENTO:/...../.....
ENDEREÇO Nº APTO:
BAIRRO: CIDADE:
CEP: TELEFONE: DDD(.....).....
2. RESPONSÁVEL LEGAL
NATUREZA (grau de parentesco, tutor, curador etc.)
DOCUMENTO DE IDENTIDADE : SEXO: M ☐ F ☐
DATA NASCIMENTO:/...../.....
ENDEREÇO: Nº APTO:
BAIRRO: CIDADE:
CEP: TELEFONE: DDD (.....).....

DADOS SOBRE A PESQUISA

1. TÍTULO DO PROTOCOLO DE PESQUISA: APLICAÇÃO DE INSTRUMENTOS DE BIOINFORMÁTICA NA AVALIAÇÃO DE NOVAS CLASSES DE ANTIRRETROVIRAIS: ANTAGONISTAS DE CCR5 E INIBIDORES DE INTEGRASE
PESQUISADOR : Dr. Jorge Simão do Rosário Casseb
CARGO/FUNÇÃO: Médico-Pesquisador INSCRIÇÃO
CONSELHO REGIONAL Nº : CRM/SP 67217
UNIDADE DO HCFMUSP: Departamento de Dermatologia do Hospital das Clínicas da Faculdade de São Paulo

3. AVALIAÇÃO DO RISCO DA PESQUISA:

RISCO MÍNIMO ☐	RISCO MÉDIO ☐
RISCO BAIXO x	RISCO MAIOR ☐

4. DURAÇÃO DA PESQUISA : 24 meses

O(a) senhor(a) está sendo convidado a participar de um estudo sobre a utilização de instrumentos no laboratório para avaliar o uso de novos medicamentos contra o Vírus da Imunodeficiência Humana (HIV).

Para infectar uma pessoa, o HIV precisa se ligar a célula CD4, utilizando diferentes “portas de entrada”. A determinação de qual destas portas de entrada é utilizada pelo vírus é importante para a aplicação de novos tratamentos. O vírus apresenta afinidade (tropismo) por diferentes “portas de entrada”. Assim, o objetivo do nosso

estudo é avaliar os testes de computador podem determinar o tropismo do HIV e com isso, facilitar a seleção de possíveis candidatos ao uso de medicamentos que bloqueiam uma dessas portas de entrada, chamados de antagonistas de CCR5.

Mais recentemente, novos medicamentos que agem no ciclo de vida do vírus, impedindo a sua multiplicação foram lançados no mercado. Neste estudo, estamos avaliando a presença de mutações dos vírus que podem diminuir a eficácia dos inibidores de integrase, que é um medicamento contra o HIV. Em conjunto com outras medicações já usadas na rotina, esses medicamentos também podem interferir na atividade do HIV.

Para participar do estudo, serão solicitadas algumas informações de identificação pessoal (nome, sobrenome, data de nascimento, local de nascimento, endereço e outros). Estas informações serão mantidas em sigilo, ou seja, o seu nome não será divulgado e as amostras receberão um código para a identificação. Após esta etapa, o(a) senhor(a) coletará 10 mL (duas colheres de sopa) através de punção venosa do sangue periférico.

Esta coleta será realizada com uma agulha esterilizada conectada em um tubo de vidro contendo anticoagulante (EDTA) e será introduzida em uma veia do seu antebraço, aspirando o sangue para dentro do tubo por um processo a vácuo.

A coleta pode acarretar em baixo risco a sua saúde e o desconforto gerado consiste em uma picada durante a introdução da agulha no antebraço, idêntica ao procedimento normal de coleta de sangue.

O(A) senhor(a) não receberá nenhum benefício direto por fazer parte deste estudo. As informações obtidas poderão ajudar a outras pessoas que tenham o HIV.

O (A) senhor (a) também tem o direito de optar por não participar do estudo. Caso o(a) senhor(a) decida não participar, isto em nada afetará o seu direito a receber tratamento na rede pública de saúde.

Em qualquer etapa do estudo, o(a) senhor(a) terá acesso aos profissionais responsáveis pela pesquisa para esclarecimento de eventuais dúvidas. O principal investigador é o Dr. Jorge Casseb que pode ser encontrados no Laboratório de Alergia e Imunologia Clínica e Experimental - LIM/56 Av. Enéas de Carvalho Aguiar, 470 / prédio IMT II / 3º andar ou pelos telefones: (11) 3061-7193 / 7457. Se o(a) senhor(a) tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) – Rua Ovídio Pires de Campos, 225 – 5º andar – tel: 3069-6442 ramais 16, 17, 18 ou 20, FAX: 3069-6442 ramal 26 – E-mail: cappelq@hcnet.usp.br

Reforçamos que participar do estudo é um ato inteiramente voluntário. O(A) senhor(a) pode preferir não participar do estudo ou abandoná-lo a qualquer momento. O(A) senhor(a) será tratado da mesma forma, independentemente de sua decisão. Nós prestaremos todos os serviços clínicos e profissionais de diagnóstico e de laboratório que fizerem parte do estudo, sem que isso incorra em custo algum para o(a) senhor(a).

Uma amostra do seu sangue será enviada para o exterior para a realização de um exame para a confirmação dos resultados que auxiliarão na avaliação deste estudo.

As informações obtidas serão analisadas em conjunto com as informações de outros pacientes, não sendo divulgado a identificação de nenhum paciente. Os documentos contendo suas informações serão guardados em local seguro e somente o pesquisador responsável e o executante terão acesso.

Nós lhe comunicaremos sobre novas informações deste ou de outros estudos que possam afetar sua saúde, bem-estar ou vontade de continuar participando deste estudo. Quando o estudo terminar, nós lhe informaremos os resultados.

Não há despesas pessoais para o participante em qualquer fase do estudo, incluindo exames e consultas. Também não há compensação financeira relacionada à sua participação. Se existir qualquer despesa adicional, ela será absorvida pelo orçamento da pesquisa.

Em caso de dano pessoal relacionado com os procedimentos ou tratamentos propostos neste estudo (nexo causal comprovado), o participante tem direito a tratamento médico na Instituição, bem como às indenizações legalmente estabelecidas.

Todos os dados e o material coletado serão utilizados somente para esta pesquisa.

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Acredito ter sido suficientemente informado a respeito das informações que li ou que foram lidas para mim, descrevendo o estudo “APLICAÇÃO DE INSTRUMENTOS DE BIOINFORMÁTICA NA AVALIAÇÃO DE NOVAS CLASSES DE ANTIRRETROVIRAIS: ANTAGONISTAS DE CCR5 E INIBIDORES DE INTEGRASE”.

Eu discuti com o Dr. Jorge Casseb sobre a minha decisão em participar nesse estudo. Ficaram claros para mim quais são os propósitos do estudo, os procedimentos a serem realizados, seus desconfortos e riscos, as garantias de confidencialidade e de esclarecimentos permanentes. Ficou claro também que minha participação é isenta de despesas e que tenho garantia do acesso a tratamento hospitalar quando necessário. Concordo voluntariamente em participar deste estudo e poderei retirar o meu consentimento a qualquer momento, antes ou durante o mesmo, sem penalidades ou prejuízo ou perda de qualquer benefício que eu possa ter adquirido, ou no meu atendimento neste Serviço.

Assinatura do paciente/representante legal Data ____ / ____ / ____

Declaro que obtive de forma apropriada e voluntária o Consentimento Livre e Esclarecido deste paciente ou representante legal para a participação neste estudo.

Assinatura do responsável pelo estudo Data ____ / ____ / ____

Apêndice B – Script de programação em *Python* para cálculo da carga elétrica líquida de uma sequência de aminoácidos

```
s = raw_input('Enter your sequence: ')
# s = sequência de aminoácidos da região V3 ou da gp120
do HIV
p = 'HKRhkr'
n = 'DEde'
x = 0
count = 0
for x in s:
    if x in p:
        count +=1
    elif x in n:
        x = n
        count -=1
print 'net charge:', count
```

Apêndice C – Artigo científico aceito para publicação na Revista do Instituto de Medicina Tropical de São Paulo

Rev. Inst. Med. Trop. São Paulo
56(4): 1-xxx, July-August, 2014
doi: 10.1590/S0036-46652014000400077

DETERMINATION OF VIRAL TROPISM BY GENOTYPING AND PHENOTYPING ASSAYS IN BRAZILIAN HIV-1-INFECTED PATIENTS

Lilá Bárbara ARRUDA(1,2), Marília Ladeira de ARAÚJO(2), Maira Lucila MARTINEZ(2), Claudio Roberto GONSALEZ(3), Alberto José da Silva DUARTE(2), Eoin COAKLEY(4), Yolanda LEE(4) & Jorge CASSEB(1,2)

SUMMARY

The clinical application of CCR5 antagonists involves first determining the coreceptor usage by the infecting viral strain. Bioinformatics programs that predict coreceptor usage could provide an alternative method to screen candidates for treatment with CCR5 antagonists, particularly in countries with limited financial resources. Thus, the present study aims to identify the best approach using bioinformatics tools for determining HIV-1 coreceptor usage in clinical practice. Proviral DNA sequences and Trofile results from 99 HIV-1-infected subjects under clinical monitoring were analyzed in this study. Based on the Trofile results, the viral variants present were 81.1% R5, 21.4% R5X4 and 1.8% X4. Determination of tropism using a Geno2pheno^(computer) analysis with a false positive rate of 10% gave the most suitable performance in this sampling: the R5 and X4 strains were found at frequencies of 78.5% and 28.4%, respectively, and there was 78.6% concordance between the phenotypic and genotypic results. Further studies are needed to clarify how genetic diversity amongst virus strains affects bioinformatics-driven approaches for determining tropism. Although this strategy could be useful for screening patients in developing countries, some limitations remain that restrict the wider application of coreceptor usage tests in clinical practice.

KEYWORDS: HIV-1; Tropism; Brazil; V3 loop.

INTRODUCTION

Maraviroc, which blocks virus entry by binding to the CCR5 coreceptor, was the first CCR5 antagonist approved for clinical use. This molecule has been widely recommended for controlling HIV infection with a low level of drug resistance^{6,14,28}. However, the use of this drug class in clinical practice requires the assessment of the coreceptor usage before and during the treatment²⁰. Studies comparing the efficiency of phenotypic and genotypic assays, in order to develop widely applicable strategies useful in developing countries, have raised several issues regarding the determination of HIV-1 tropism^{7,18}.

Although phenotypic tropism tests (PTT) are considered the gold standard for determining coreceptor usage, their wide application in clinical practice is limited mainly by their high cost and the logistical restrictions of the procedures¹⁷. In this context, programs able to predict coreceptor usage based on V3 sequences have provided a suitable alternative for screening candidates for CCR5 antagonist therapy²⁵.

Several bioinformatics protocols for tropism prediction have been published. They assess tropism based on amino acids sequences from

the V3 loop, which is known to be an important binding region on the gp120 envelope protein^{16,23}. Because the use of phenotypic assays is still limited, the European Guidelines have encouraged the application of bioinformatics programs in coreceptor usage determination²⁷. However, due to the intrinsic differences of each predictive system, divergent outputs are expected and remain a reason for concern in the wider application of this approach^{9,15,30,23-24}.

The aim of this study was to examine the use of the Geno2pheno^(computer) program for tropism prediction and determine its usefulness as an alternative method in clinical practice for screening CCR5 antagonist therapy candidates for the first time in Brazil.

MATERIAL AND METHODS

Sampling: Sample collection occurred between July 2009 and October 2010, from the HIV outpatient clinic ADEE3002/HCFMUSP. Only 99 from this cohort showed the minimum plasmatic RNA viral load necessary to qualify for the Trofile assay (> 1000 copies/mL, as recommended by Monogram Biosciences, San Francisco, CA). The written informed consent for collecting blood samples and the protocol for this study were approved by the Ethical Research Board of the

(1) Institute of Tropical Medicine of São Paulo, University of São Paulo, Dr. Eneas de Carvalho Aguiar Avenue 470, Building I, São Paulo, SP, Brazil.

(2) Laboratory of Investigation in Dermatology and Immunodeficiencies, LIM56, Department of Dermatology School of Medicine at University of São Paulo, Dr. Eneas de Carvalho Aguiar Avenue 470, Building II, 3rd floor, São Paulo, SP, Brazil.

(3) HIV Out-clinic, Ambulatory of Secondary Immunodeficiencies, ADEE3002, Department of Dermatology, Hospital of Clinics at School of Medicine at University of São Paulo.

(4) Monogram Biosciences, Inc., 345 Oyster Point Blvd. South San Francisco, CA 94080-1913, USA.

Correspondence to: Jorge Casseb, Laboratório de Investigação em Dermatologia e Imunodeficiências, LIM56, Av. Dr. Eneas de Carvalho Aguiar 500, 3^o andar, Prédio 2, 05403-000 São Paulo, SP, Brazil. Tel.: +55 11 3061 7194. Fax: +55 11 3081 7190. E-mail: jcasseb10@gmail.com

Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo (Cappesq - HC/FMUSP, number 08/0108).

Subjects: A total of 99 patients under clinical monitoring were included in this study, of whom 29 (29.3%) were females and 70 (70.7%) were males. With respect to antiretroviral therapy, 62 (62.6%) patients were drug-naïve, and 37 (37.4%) were under HAART. Only subjects being treated with nucleoside and non-nucleoside reverse transcriptase inhibitors (NRTI and NNRTI) and protease inhibitors (PI) were studied.

After signing the informed consent form, 15 mL of venous blood was collected in EDTA-treated tubes from each patient. Plasma samples were separated and stored at -70 °C until their shipment to Monogram Biosciences to perform the Trofile assay. Peripheral blood mononuclear cells (PBMCs) were isolated from blood samples by Ficoll-Hypaque density gradient centrifugation and stored at -70 °C until use. DNA was extracted from PBMCs using the QIAx Genomic Blood DNA Purification Kit (Qiagen, Crawley, UK) according to the manufacturer's instructions.

The HIV-1 plasmatic viral load was measured using the Versant HIV RNA 3.0 Assay (bDNA) (Siemens Healthcare Diagnostics, San Francisco, USA), and CD4+CD8+ T cell counts were determined by flow cytometry. These data were collected during routine clinical monitoring of patients and were obtained from the laboratory database.

In addition, 185 sequences containing known coreceptor sequence variations were obtained from the HIV Sequence Database (www.hiv-web.lanl.gov) to assess the application of genotypic tropism test (GTT) in identifying different HIV-1 variants. Only V3 sequences with 35 amino acids were evaluated; sequences with 34 or 36 amino acids were excluded. All deposited sequences originating in Brazil were included, and at least 5% of the sequences for each group were selected from other countries of origin (Table 1).

Table 1
Distribution of HIV database sequences by subtype and phenotypic tropism

Subtype	R5	R5X4	X4	Total
B	115	18	13	146
C	23	7	5	35
FI	2	1	1	4
Total	140	26	19	185

V3 region amplification and sequencing: The V3 region was amplified by nested polymerase chain reaction (PCR) using the previously described outer primers ED5/ED12 and inner primers ED31/ED33 (4). The PCR using the outer primers was performed with the following steps: one cycle of 94 °C for one min; 35 cycles of 94 °C for 45 sec, 55 °C for 45 sec, and 72 °C for two min; and a final extension cycle of 72 °C for 10 min. The PCR using the inner primers was performed with the following steps: one cycle of 94 °C for one min; 35 cycles of 94 °C for one min, 55 °C for one min and 72 °C for one min and 30 sec; and a final extension cycle of 72 °C for 10 min. The 50-µL reaction mixtures contained 2 mM MgCl₂, 2 µM of each primer, 0.2 mM of each dNTP, 1 U Taq DNA polymerase and 5 µL of extracted DNA for the outer PCR or 5 µL of PCR product for the inner PCR.

The inner PCR products were purified using the QIAquick PCR fragment purification kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The sequencing reaction was carried out using the ABI Prism Big Dye Terminator v.3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, USA), adding 20-100 ng of amplified and purified product and each inner PCR primer at 1 µM. The reaction was carried out according to the manufacturer's instructions, and the sequencing was performed using the ABI Prism 3100 system (Applied Biosystems, Foster City, USA). The resulting nucleotide sequences were edited using SeqScape Software v.2.6 (Applied Biosystems).

HIV-1 envelope subtyping was performed by manual phylogenetic analysis using the 2008 Alamos reference dataset (available at <http://www.hiv.lanl.gov/content/sequence/NEUALIGN/align.html>). Sequences were aligned and edited, and a tree was constructed with the Neighbor-Joining method and the Maximum Composite Likelihood model with 1000 bootstrap replicates using the MEGA v.5 package (25).

Tropism determination: Coreceptor usage was determined using the Geno2pheno_[coreceptor] bioinformatics tool (G2P) (10). False positive rates (FPR) were 10% for the proviral DNA sequences obtained in this study or 20% for clonal and clinical data.

The Trofile Coreceptor Tropism Assay (Monogram Biosciences, San Francisco, USA) was performed on plasma samples from patients showing viral load > 1000 copies/mL.

RESULTS

Comparison between PTT and GTT: From the 99 samples sent for analysis with the Trofile assay, 55 (55.6%) were identified as R5 viruses, 22 (22.2%) were R5X4, and only one (1%) sample showed tropism to the CXCR4 coreceptor. The Trofile assay was unable to generate results for 21 (21.2%) samples. Therefore, the coreceptor prevalence within the assayable sample population was 70.5% (55/78) R5, 28.3% (22/78) R5X4 and 1.3% (1/78) X4 strains.

Readable sequences were obtained from 70 proviral DNA sequences, resulting in the generation of both Trofile and G2P results for 56 samples. These matched sample measures allowed for sensitivity and specificity testing, which showed that a FPR of 10% gave the highest sensitivity and specificity (Table 2).

Table 2
Sensitivity and specificity test for Geno2pheno_[coreceptor] false positive rates in prediction of CXCR4 coreceptor usage

G2P FPR	Sensitivity (%)	Specificity (%)
10%	54.5	84.4
20% clinical data	50.0	82.6
20% clonal data	34.8	68.1

Note: G2P FPR: Geno2pheno_[coreceptor].

Based on the G2P FPR of 10%, 55 (78.5%) and 15 (21.4%) of the 70 sequences obtained in this study were predicted to be R5 and X4 strains, respectively. There were no differences between the Trofile

ARRUDA, L.B.; ARAÚJO, M.L.; MARTINEZ, M.L.; GONSALEZ, C.R.; DUARTE, A.J.S.; COAKLEY, E.; LEE, Y. & CASSEB, J. - Determination of viral tropism by genotyping and phenotyping assays in Brazilian HIV-1-infected patients. *Rev. Inst. Med. Trop. São Paulo*, 56(4): xxx, 2014.

results (R5 = 70.5%; X4 and R5X4 = 29.5%) and G2P results found for this population.

Envelope subtyping was performed on the sequences obtained in this study. Only B (n = 46), C (n = 2) and F1 (n = 8) subtypes were found, and it was not possible to identify any recombinant forms. Tropism prediction showed good accuracy regardless of the subtype, generating concordances between the Trofile and G2P results of 76.1% for subtype B samples, 100% for subtype C and 87.5% for subtype F1. Due to the small number of clinical samples in our study, a dataset containing 185 sequences from the Los Alamos HIV-1 sequence database was also analyzed. The prediction accuracy for the samples in this dataset was represented by concordances between the Trofile and G2P results of 88.3% for subtype B, 88.6% for subtype C and 100% for subtype F1 both sequences available in this database.

DISCUSSION

Phenotypic tropism tests (PTT) still represent the gold standard in tropism determination, but high cost and logistical restrictions are significant barriers preventing the wide application of these assays, particularly in developing countries²³. Thus, the adoption of bioinformatics tools for determining coreceptor usage could provide a more accessible alternative for screening candidates for CCR5 antagonist therapy^{19,20}.

The phenotyping test has some limitations in the clinical setting. While the commercially available Trofile assay is the most widely used phenotypic test¹⁷ and provides accurate determination of HIV tropism, it was unable to generate results for 21% of the samples in this study. An additional important limitation of the Trofile assay is the requirement that samples have a minimum plasma viral load of 1000 copies/mL. Based on studies that identified similar tropism behavior in 75% to 90% of sequences from plasma and PBMCs, using proviral DNA as the source of viral genetic material could be the most reliable option for obtaining V3 sequences. Moreover, X4 viruses were detected more frequently in DNA samples from PBMCs than in RNA samples from plasma^{13,20,21}. In fact, the Trofile test is mostly used in clinical trials and, in the meantime, has been replaced in the clinical setting by genotyping tests (at least in Europe). Furthermore, the Trofile test is not an entirely phenotypic assay, as a library of V3 loop sequences of HIV-RNA from the test sample is prepared by PCR amplification. There is bias in this PCR amplification, as is in any PCR. Only the read out of the Trofile test is phenotypic. In contrast, the MT2 assay is entirely phenotypic. Bias is also present in population based genotypic testing by the PCRs and the sequencing reactions used.

Bulk sequencing of proviral DNA is an intrinsically limited technique, because it produces a consensus sequence from dominant strains within the viral quasispecies^{7,22}. Although it has been suggested that this problem could be mitigated by increasing the number of replicate sequences produced from each sample, this modification also increases the method's costs and time required and could further limit its use, particularly in countries with restricted financial resources^{7,13}. The present study aimed to develop an accurate and cost-effective strategy for applying tropism testing in clinical practice. Although the European Guidelines recommend performing G2P with an FPR of 10% for triplicate samples and 20% for samples with a single sequence²⁷, the present study's

sensitivity and specificity tests showed that a G2P FPR = 10% is the best significance level, even for single sequences (Table 2).

This study showed that the concordance between PPT and GTT was also high for different subtypes (mean of concordance = 88%), and this finding was confirmed by testing a large dataset from the Los Alamos Sequence Database (mean of concordance = 92%). These results corroborate previous studies that showed 80-90% agreement between phenotypic and genotypic tests for determining tropism^{13,16,17,26}. Despite the high concordance between the phenotypic and genotypic assays, there has been disagreement between the results generated by different bioinformatics-based assays, and several approaches to improving these predictions have been proposed^{23,24,22,28}. Additional studies are needed to determine if viral sequences in the Brazilian patient population might be so distinct as to invalidate their analysis with available bioinformatics tools, which derive their datasets mainly from sequences found in patients from the USA and European countries.

Furthermore, while the determination of HIV-1 tropism has focused on the V3 loop, other features of gp120 could influence virus affinity for a particular coreceptor, including N-glycosylation sites and variations in extensions of the V1/V2 region^{8,12}. Information on the three-dimensional structure of the V3 loop and clinical and laboratory data from patients, such as T cells counts and viral load, could also substantially improve tropism predictions²⁹.

CONCLUSIONS

The application of bioinformatics tools in coreceptor usage determination has been widely studied and could provide an alternative approach in clinical practice for screening candidates for CCR5 antagonist therapy, especially in cases where confirmatory assays for coreceptor usage determination are unavailable. This study showed that G2P analysis of sequences from PBMCs with an FPR of 10% could be the most suitable alternative approach for determining tropism, although this approach should be tested in a nationwide study to determine how HIV genetic diversity influences the findings.

RESUMO

Determinação do tropismo viral por ensaios genotípicos e fenotípicos em pacientes brasileiros infectados por HIV-1

A aplicação clínica dos antagonistas de CCR5 envolve em primeiro lugar determinar o uso de co-receptor pela cepa viral infectante. Programas de bioinformática que prevêem o uso co-receptor poderiam fornecer um método alternativo para selecionar candidatos para o tratamento com os antagonistas do CCR5, particularmente em países com poucos recursos financeiros. Assim, o presente estudo teve por objetivo identificar a melhor abordagem utilizando ferramentas de bioinformática para determinar qual o tipo de co-receptor do HIV-1 que poderia ser usado na prática clínica. Sequências de DNA proviral e Trofile resultados a partir de 99 pacientes infectados pelo HIV-1 sob monitorização clínica foram avaliadas. Com base nos resultados do Teste Trofile, as variantes virais presentes eram R5 (81.1%), R5X4 (21.4%) e X4 (1.8%). Determinação do tropismo pela análise do Geno2pheno, com uma taxa de falso positivos de 10% apresentou desempenho mais adequado para esta amostragem: as cepas R5 e X4 foram encontrados

em frequências de 78.5% e 28.4%, respectivamente, e foi de 78.6% a concordância entre os resultados fenotípica e genotípica. Mais estudos são necessários para esclarecer como a diversidade genética entre as cepas do vírus afeta abordagens baseadas na determinação do tropismo pelas ferramentas de bioinformática. Embora esta estratégia possa ser útil para o rastreamento de pacientes em países em desenvolvimento, permanecem algumas limitações que restringem a aplicação mais ampla utilização de testes de co-receptor na prática clínica.

ACKNOWLEDGEMENTS

We thank all patients who participated in this study. We also thank Rosângela Araújo and Noêmia Orii for performing the flow cytometry experiments, José Eduardo Martins for the determination of HIV-1 viral load and the financial support from FAPESP (08/58138-0; 08/51265-6; 10/00222-5); FFM; CNPq.

REFERENCES

1. Braun P, Wasmann F. Phenotypic assays for the determination of coreceptor tropism in HIV-1 infected individuals. *Eur J Med Res*. 2007;12:463-72.
2. Choea N, Garrido C, Alvarez M, Poveda E, de Dios Lara J, Zahonero N, et al. Improvement in the determination of HIV-1 tropism using the V3 gene sequence and a combination of bioinformatic tools. *J Med Virol*. 2009;81:763-7.
3. Delobel P, Nugeyre MT, Cazabat M, Pasquier C, Marchoux B, Mascop P, et al. Population-based sequencing of the V3 region of env for predicting the coreceptor usage of human immunodeficiency virus type 1 quasispecies. *J Clin Microbiol*. 2007;45:1572-80.
4. Delwart EL, Herring B, Rodrigo AG, Mullins JE. Genetic subtyping of human immunodeficiency virus using a heteroduplex mobility assay. *PCR Methods Appl*. 1995;4:S202-16.
5. Dybowski JN, Heider D, Hoffmann D. Prediction of co-receptor usage of HIV-1 from genotype. *PLoS Comput Biol*. 2010;6:e1000743.
6. Emmelkamp JM, Rockstroh JK. CCR5 antagonists: comparison of efficacy, side effects, pharmacokinetics and interactions—review of the literature. *Eur J Med Res*. 2007;12:409-17.
7. Harrigan PR, Geretti AM. Genotypic tropism testing: evidence-based or leap of faith? *AIDS*. 2011;25:257-64.
8. Hughes A, Nelson M. HIV entry: new insights and implications for patient management. *Curr Opin Infect Dis*. 2009;22:35-42.
9. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Steiner D, Ho HK, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J Virol*. 2003;77:13376-88.
10. Lengauer T, Sander O, Sierra S, Thielens A, Kaiser R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol*. 2007;25:1407-10.
11. McGovern RA, Harrigan PR, Swenson LC. Genotypic inference of HIV-1 tropism using population-based sequencing of V3. *J Virol*. 2010;84(46):pii:2531. doi:10.7791/2531.
12. Mild M, Kvist A, Eide-Jensen J, Karlsson I, Fenyo EM, Modstrand P. Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection. *Infect Genet Evol*. 2010;10:356-64.
13. Paar C, Geit M, Stichel H, Berg J. Genotypic prediction of human immunodeficiency virus type 1 tropism from plasma and peripheral blood mononuclear cells in the routine clinical laboratory. *J Clin Microbiol*. 2011;49:2697-9.
14. Perea J, Pettilla J, Palido F, Sanchez-de la Rosa R, Alorin-Villaverde C, Berenguer J, et al. Clinical utility of maraviroc. *Clin Drug Investig*. 2011;31:527-42.
15. Pillai S, Good B, Richman D, Corbett J. A new perspective on V3 phenotype prediction. *AIDS Res Hum Retroviruses*. 2003;19:145-9.
16. Prospero MC, Fanti I, Ulivi G, Micarelli A, De Luca A, Zazzi M. Robust supervised and unsupervised statistical learning for HIV type 1 coreceptor usage analysis. *AIDS Res Hum Retroviruses*. 2009;25:305-14.
17. Raymond S, Delobel P, Marigner M, Cazabat M, Soeys C, Sanchez-Saune K, et al. Correlation between genotypic predictions based on V3 sequences and phenotypic determination of HIV-1 tropism. *AIDS*. 2008;22:F11-6.
18. Recorcon-Finon P, Soulie C, Flandre P, Descamps D, Larzek M, Charpentier C, et al. Evaluation of the genotypic prediction of HIV-1 coreceptor use versus a phenotypic assay and correlation with the virological response to maraviroc: the ANRS GenTropism study. *Antimicrob Agents Chemother*. 2010;54:3335-40.
19. Sander O, Sing T, Sommer I, Low AJ, Cheung PK, Harrigan PR, et al. Structural descriptors of gp120 V3 loop for the prediction of HIV-1 coreceptor usage. *PLoS Comput Biol*. 2007;3(3):e58.
20. Saracino A, Monno L, Cibelli DC, Punzi G, Brindicci G, Ladina N, et al. Co-receptor switch during HAART is independent of virological success. *J Med Virol*. 2009;81:2036-44.
21. Seclén E, Garrido C, González M del M, González-Lahoz J, de Mendoza C, Soriano V, et al. High sensitivity of specific genotypic tools for detection of X4 variants in antiretroviral-experienced patients suitable to be treated with CCR5 antagonists. *J Antimicrob Chemother*. 2010;65:1486-92.
22. Seclén E, Soriano V, González MM, Gómez S, Thielens A, Poveda E. High concordance between the position-specific scoring matrix and geno2pheno algorithms for genotypic interpretation of HIV-1 tropism: V3 length as the major cause of disagreement. *J Clin Microbiol*. 2011;49:3380-2.
23. Sierra S, Kaiser R, Thielens A, Lengauer T. Genotypic coreceptor analysis. *Eur J Med Res*. 2007;12:453-62.
24. Sing T, Low AJ, Boerenwinkel N, Sander O, Cheung PK, Domingues FS, et al. Predicting HIV coreceptor usage on the basis of genetic and clinical correlates. *Antivir Ther*. 2007;12:1097-106.
25. Tamura K, Peterson D, Peterson N, Staecher G, Nei M, Kumar S. MEGA3: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731-9.
26. Trabaud MA, Icard V, Scholtes C, Porpoint T, Koffi J, Cotte L, et al. Discordance in HIV-1 co-receptor use prediction by different genotypic algorithms and phenotype assay: intermediate profile in relation to concordant predictions. *J Med Virol*. 2012;84:402-13.
27. Vandekerckhove LP, Wensing AM, Kaiser R, Brun-Vezinet F, Clotet B, De Luca A, et al. European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect Dis*. 2011;11:394-407.
28. Verhofstede C, Vandekerckhove L, Eygen VV, Demicheleer H, Vandenbroucke I, Winters B, et al. CXCR4-using HIV type 1 variants are more commonly found in peripheral blood mononuclear cell DNA than in plasma RNA. *J Acquir Immune Defic Syndr*. 2009;50:126-36.

Received: 28 July 2013

Accepted: 30 January 2014

Apêndice D – Resumo do trabalho apresentado oralmente na 4ta. Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio)

Martes 29

INDEAR

Horario	Autores	Título
17:00-17:15	Diego Nicolás De Panis , Pedro Furió-Tari, Julián Padró, Ignacio Soto, Sonia Tarazona, Hernán Dopazo, Ana Conea and Esteban Hasson.	Transcriptomics of host adaptation, early results: Gene expression patterns of the cactophilic fly <i>Drosophila buzzatii</i> in its natural breeding and feeding resources
17:20-17:35	Gabriela Aguilera , Damien M. de Vienne, Oliver N. Ross, Elsa Petit, Tatiana Giraud, Michael E. Hood and Toni Gabaldon.	High variability of mitochondrial gene order among Fungi
17:40-17:55	Sergio Gonzalez , Bernardo Clavijo, Patricio Moreno, Maximo Rivarola, Paula Fernandez, Marisa Farber and Norma Paniego.	ATGC Transcriptomics: An application to explore and analyze de novo transcriptomic data.
18:00-18:15	Liã Bárbara Amuda , Gertjan Beheydt, Priscila Benevides, Bruno Aguiar, Andrea Clemencia Pineda, Kristel Van Laethem, Alberto Duarte and Jorge Casseb.	Bayesian network approaches to evaluate novel V3 mutations associated to HIV-1 tropism

Bayesian network approaches to evaluate novel V3 mutations associated to HIV-1 tropism

Liã Bárbara Arruda^{1,2}, Gertjan Beheydt³, Priscila Benevides², Bruno Aguiar², Andrea Clemencia Pineda^{3,4}, Kristel Van Laethem³, Alberto Duarte², Jorge Casseb^{1,2}

¹Institute of Tropical Medicine of Sao Paulo – USP, Sao Paulo, SP

²Laboratory of Medical Investigation in Dermatology and Immunodeficiencies – LIM56, Department of Dermatology, FMUSP, Sao Paulo, SP

³Laboratory for Clinical and Evolutionary Virology, Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

⁴Clinical and Molecular Aspects of Infectious Diseases Group, Universidad del Rosario, Bogotá, Colombia

Background

Genotypic tropism tests (GTT) are a suitable alternative for screening HIV-1 infected subjects before starting treatment with CCR5 antagonists. GTT are based on sequencing analysis of the V3 loop, a hypervariable region encompassing 35 amino acids in the HIV envelope protein.

Materials and methods

V3 amino acid sequences, phenotypic tropism commercial test (Trofile, Monogram Biosciences) and genotypic tropism test (Geno2pheno[coreceptor] - G2P, false positive rate = 10%) were obtained from 75 HIV-1-infected subjects. A dataset of 1208 V3 sequences from Los Alamos database was included to evaluate the findings obtained from the clinical samples. Bayesian network (BN) learning was used to evaluate the dependency of V3 loop mutations and HIV tropism.

Results

Trofile results showed 72% (n=54), 26.7% (n=20) and 1.3% (n=1) of R5, R5X4 and X4 strains, respectively. G2P showed 76% (n=57) of R5 and 24% (n=18) of R5X4/X4 strains. Sensitivity and specificity of G2P were 80% and 55%, respectively. Los Alamos database sequences included 82.6% (n=997), 11.4% (n=138) and 6% (n=73) of R5, R5X4 and X4 strains, respectively. BN learning showed no strong dependency between V3 mutations and Trofile results from clinical sequences, although, for G2P results it was showed a strong dependency of R5 strains and mutations 9S, 11R, 19V and 25K. Among Los Alamos sequences mutations 9R, 18Q, 11R, 24G and 25D also showed strong dependency with R5 strains. No mutations showed dependency with R5X4 or X4 strains. The frequency of the mutations 11R, 11S, 16P, 16W, 24G, 25D, 25E and 25Q was statistically different between clinical and Los Alamos sequences.

Conclusion

Our findings corroborate that the classical V3 positions 11, 24 and 25 are related with tropism [1-3], when genotypic and database sequences were evaluated. However, phenotypic tropism showed no strong dependency with

any mutation. Thus, frequency differences of some mutations could indicate a prediction bias, suggesting that mutations observed in virus populations from a specific area could influence the viral tropism. Regional studies should be encouraged to confirm the role of area-specific viral polymorphisms in the worldwide application of bioinformatics tools for tropism determination.

Acknowledgments

We thank Monogram Biosciences for performing the Trofile Assay and FAPESP for the financial support (2008/58138-0, 2010/00222-5 and 2011/17781-4).

References

1. De Jong JJ, De Ronde A, Keulen W, Tersmette M, Goudsmit J: **Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution.** *J Virol* 1992, 66:6777-6780.
2. Fouchier RA, Groenink M, Kootstra NA, Tersmette M, Huisman HG, Miedema F, Schuitemaker H: **Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule.** *J Virol* 1992, 66:3183-3187.
3. Traub MA, Icard V, Scholtes C, Perpoint T, Koffi J, Cotte L, Makhoulfi D, Tardy JC, Andre P: **Discordance in HIV-1 Co-receptor use prediction by different genotypic algorithms and phenotype assay: Intermediate profile in relation to concordant predictions.** *J Med Virol* 2012, 84:402-413.

ANEXOS

Anexo A – Certificado emitido pela Comissão de Ética

APROVAÇÃO

A Comissão de Ética para Análise de Projetos de Pesquisa - CAPPesq da Diretoria Clínica do Hospital das Clínicas e da Faculdade de Medicina da Universidade de São Paulo, em sessão de 04/06/2008, **APROVOU** o Protocolo de Pesquisa nº **0108/08**, intitulado: "**DETERMINAÇÃO DO TROPISMO DO HIV-1 PELOS CO-RECEPTORES CCR5 E CXCR4 PELO USO DE FERRAMENTAS DE BIOINFORMÁTICA**" apresentado pelo Departamento de **DERMATOLOGIA**, inclusive o Termo de Consentimento Livre e Esclarecido.

Cabe ao pesquisador elaborar e apresentar à CAPPesq, os relatórios parciais e final sobre a pesquisa (Resolução do Conselho Nacional de Saúde nº 196, de 10/10/1996, inciso IX.2, letra "c").

Pesquisador (a) Responsável: **Dr. Jorge Simão do Rosário Casseb**

Pesquisador (a) Executante: **Liã Bárbara Arruda**

CAPPesq, 06 de Junho de 2008

Prof. Dr. Eduardo Massad
Presidente da Comissão de
Ética para Análise de Projetos
de Pesquisa

Anexo B – Aprovação da alteração de título

Ao
Departamento de Dermatologia

A Comissão de Ética para Análise de Projetos de Pesquisa - CAPPesq da Diretoria Clínica do Hospital das Clínicas e da Faculdade de Medicina da Universidade de São Paulo, em sessão de 25.03.09, **APROVOU** alteração de título do Protocolo de Pesquisa nº **0108/08** "Determinação do tropismo do HIV-1 pelos co-receptores CCR5 e CXCR4 pelo uso de ferramentas de bioinformática" para "Aplicação de instrumentos de bioinformática na avaliação de novas classes de antiretrovirais: Antagonistas de CCR5 e inibidores de Integrase", bem como o Termo de Consentimento Livre e Esclarecido.

Pesquisador (a) Responsável: **Dr. Jorge Simão do Rosário Casseb**

Pesquisador (a) Executante: **Lia Bárbara Arruda**

CAPPesq, 27 de Março de 2009.

PROF. DR. EDUARDO MASSAD
Presidente da Comissão Ética para Análise
de Projetos de Pesquisa

Anexo C – Carta de aceite para publicação de artigo na Revista do Instituto de Medicina Tropical de São Paulo

9/4/2014

Gmail - Fwd: Revista do Instituto de Medicina Tropical de São Paulo - Decision on Manuscript ID RIMTSP-2013-0113.R2

Liã Bárbara Arruda <libarbara@gmail.com>

Fwd: Revista do Instituto de Medicina Tropical de São Paulo - Decision on Manuscript ID RIMTSP-2013-0113.R2

3 messages

jorge Casseb <jcasseb10@gmail.com>
To: Liã Bárbara Arruda <libarbara@gmail.com>

30 January 2014 14:27

----- Forwarded message -----

From: <thalesdebrito.revimtsp@gmail.com>

Date: Thursday, January 30, 2014

Subject: Revista do Instituto de Medicina Tropical de São Paulo - Decision on Manuscript ID RIMTSP-2013-0113.R2

To: jcasseb@usp.br, jcasseb10@gmail.com

30-Jan-2014

Dear Prof. Casseb:

It is a pleasure to accept your manuscript entitled "Determination of viral tropism by genotyping and phenotyping assays in Brazilian HIV-1-infected patients" in its current form for publication in the Revista do Instituto de Medicina Tropical de São Paulo.

Thank you for your fine contribution. On behalf of the Editors of the Revista do Instituto de Medicina Tropical de São Paulo, we look forward to your continued contributions to the Journal.

Sincerely,
Prof. Thales de Brito
Editor-in-Chief, Revista do Instituto de Medicina Tropical de São Paulo
thalesdebrito.revimtsp@gmail.com

Anexo D – Certificado de apresentação oral na 4ta. Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio)

<p>Rosario, 29 al 31 de Octubre de 2013</p>		
<p>A2B2C, Asociación Argentina de Bioinformática y Biología Computacional, and SolBio, Sociedad Iberoamericana de Bioinformática, certify that</p>		
<p>Liã Bárbara Arruda</p>		
<p>has presented a <i>conference</i> titled</p>		
<p>Bayesian network approaches to evaluate novel V3 mutations associated to HIV-1 tropism</p>		
<p>in the IV Argentinean Conference on Computational Biology and Bioinformatics & the IV Conference of the Iberoamerican Society for Bioinformatics (SOIBIO).</p>

<p>Dr. Ignacio Ponzoni <i>President A2B2C</i></p>	<p>Dra. Elizabeth Tapia <i>Executive Chair, 4CAB2C & SolBio</i></p>	<p>Dr. Javier de las Rivas <i>Executive Board Member, SolBio</i></p>

Anexo E – Relatórios das corridas de sequenciamento da plataforma *Ion Torrent* (Life Technologies)

Chip 1

Run Summary

Barcode Name	Sample	Bases	$\geq Q20$	Reads	Mean Read Length
No barcode	None	551,771	450,904	3,468	159 bp
IonXpress_020	006B	6,596,777	5,646,158	41,797	157 bp
IonXpress_021	110	12,911,857	10,871,239	77,196	167 bp
IonXpress_022	113	25,214,836	21,227,312	149,106	169 bp
IonXpress_023	115	6,791,714	5,744,511	39,582	171 bp
IonXpress_024	129	6,364,321	5,369,314	38,790	164 bp

Test Fragment	Reads	Percent 50AQ17	Read Length Histogram
TF_A	2,792	87%	

Chip 2

Run Summary

Barcode Name	Sample	Bases	$\geq Q20$	Reads	Mean Read Length
No barcode	None	540,697	453,541	3,313	163 bp
IonXpress_023	None	5,637,266	4,623,735	33,477	168 bp
IonXpress_024	None	638,035	547,404	4,133	154 bp
IonXpress_025	130	12,841,195	11,028,321	74,126	173 bp
IonXpress_026	135	20,516,794	17,758,470	117,192	175 bp
IonXpress_027	149	9,471,428	8,208,579	54,629	173 bp
IonXpress_028	240	15,004,551	12,921,478	85,080	176 bp
IonXpress_029	271	16,081,235	13,742,254	89,368	179 bp

Test Fragment	Reads	Percent 50AQ17	Read Length Histogram
TF_A	4,403	90%	

Chip 3

Run Summary

Barcode Name	Sample	Bases	$\geq Q20$	Reads	Mean Read Length
No barcode	None	423,622	333,048	3,073	137 bp
IonXpress_020	061B	8,587,077	7,127,196	52,127	164 bp
IonXpress_021	077B	9,373,689	7,693,705	58,526	160 bp
IonXpress_022	127	12,466,826	10,256,032	75,349	165 bp
IonXpress_023	138	16,228,113	13,340,495	99,673	162 bp
IonXpress_024	145	10,008,407	8,307,503	60,910	164 bp

Test Fragment	Reads	Percent 50AQ17	Read Length Histogram
TF_A	4,033	82%	

Chip 4

Run Summary

Barcode Name	Sample	Bases	≥ Q20	Reads	Mean Read Length
No barcode	None	367,348	276,836	2,794	131 bp
IonXpress_025	212	20,927,851	16,745,783	126,024	166 bp
IonXpress_026	219	5,664,687	4,616,926	34,836	162 bp
IonXpress_027	243	7,989,996	6,564,439	47,432	168 bp
IonXpress_028	251	17,745,595	14,206,702	102,540	173 bp
IonXpress_029	256	11,260,580	9,007,042	66,597	169 bp

Test Fragment	Reads	Percent 50AQ17	Read Length Histogram
TF_A	4,239	83%	

Anexo F - Predição de sítios de N-glicosilação das sequências geradas por sequenciamento massivo paralelo

	10	20	30	40	50	60	70	80	90	100
E1006B	?*K*RRSGRI	GSAAWIGGTM	LLGLMICS	VDQLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1061B	MRVKENYQ*	WTALWRGGTL	LLGLMICS	ITELWVIVY	GVPMKEAT	TLFCASDAKA	SDTEAHNV	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1077B	-?ESEGDEE	LSVLWRGTM	LLGLMICS	TEKLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1110	MRVKEEYQHL	WRWGRWGT	LLGLMICS	AGQWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1113	MRVKENCQHL	WRWGRWGT	LLGLMICS	TEQLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1115	?*ERRGRRI	VSAAWRGGTM	LLGLMICS	AGQWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1127	MRVKENYQPL	WRWGRWGT	LLGLMICS	AGQWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1129	?*E*RRSGRI	VSTWRGGTL	LLGLMICS	TEQLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1130	-?AKNQCQHL	WRWGRWGT	LLGLMICS	TNNLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1135	MRVKIGRKL	SREWRGGTM	LLGLMICS	AEPLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1138	-?ESEGDEE	LLVWRWGT	LLGLMICS	AEQLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1149	MRVKENYQHL	WRWGRWGT	LLGLMICS	AEQLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1145	-MRVEMQRS	WPPWGGWGLL	FLGLIICA	-EDLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1212	-MRVKENYQ	LMSWRGGTL	FLGLMICS	TEDLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1219	?*E*GGCGGI	VSTWGGWGLL	FLGLMICS	TEKLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1240	?ESEGDEE*	WKLWRWGT	LLGLMICS	VDQLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1243	---MRVKENC	QHLWRWGT	LLGLMICS	TEKSWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1251	?*ERRRGRI	WKLWGWGT	LLGLMICS	EEKLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1266	?*E*GGCGGI	GSTWGGWGLL	FLGLIICA	-ENLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1271	-----	?GDQDGGWGLL	FLGLIICA	-ENLWVIVY	GVPMKEAT	TLFCASDAKA	YDTEVHNW	THACVPTDPN	POEVFLGNT	ENFNMKNNM
E1006B	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTDVK--RND	STI---ENDI	MGEMKNCSEF	ITITIKNRH	KEYALFYKLD	LEPIENNST	YTLITCNTSV
E1061B	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTDVKNGINT	TSNRTIGEE	-ADMRNCSFN	ITITIGNRKH	KEYALFYKLD	LTKIEGNGIS	YMLINCNTSV
E1077B	VEQMHTDIIS	LWDQSLKPCV	KLPLCVTLN	CTNVNITNT	NSN-SLI---	-----	-----	-----	-----	YRLINCNTSV
E1110	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTDVNITNT	NANTITII*MK	*GEIKNCSFN	ITISIGNKLQ	KDYALFNLD	VVPIDNDNTS	YMLTKCNTSV
E1113	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTDMKCNSTI	NNSSGIMVMM	KEIKKNCSEF	VTSIGKRMQ	KEYALFYKLD	LEPIDNDNTS	YTLISCNTSV
E1115	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTNMKNNTN	NSSSGMILEP	G-EIKNCSFN	VTSIGNRKM	KEYALFYRLD	VVPIDNDNTS	YMLINCNTSV
E1127	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTDVKNVNT	TSNNTITLVL	*GNMKNCSEF	VTSIEKKVR	KEYALFYKLD	VVPIDNDNTS	YMLINCNTSV
E1129	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTNMKNNTN	NSSSWEKMEG	-----	-----	-----	-----	YMLINCNTSV
E1130	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTAEKNTN-	-----	-----	-----	-----	-----	YMLINCNTSV
E1135	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CDNLNNTVNP	NTSKIFMIKM	IGEIKNCSFN	ITTSIQKVK	KEYALFNLD	VVPIDNDNTS	YMLINCNTSV
E1138	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CNNYNGTNYN	STIET--IIVE	KEEVKNCSEF	VTSIGKRVK	KEYALFNLD	VVPIDNDNTS	YMLINCNTSV
E1149	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTNVNNTST	NSNNGLRME	EGEMKNCSEF	ITTSIGNRKM	KEYALFYRLD	VVPIDNDNTS	YMLINCNTSV
E1145	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTDVNNNTN	NNTN--TIKE	EGAIKNCSEF	IT-----	-----	-----	YMLINCNTSV
E1212	VEQMHEDIIN	LWDQSLKPCV	KLPLCVTLN	CNNFNNTSF	NNSR*-----	IKK*KNCSEF	ITTSIRDKVK	KEYALFYKLD	VVPIDNNST	YMLINCNTSV
E1219	VEQMHEDIIN	LWDQSLKPCV	KLPLCVTLN	CTDAGNDNT	NSSTRNRVKE	KEIKKNCSEF	VTSIGKRVK	KEYALFYKLD	VVPIDNNST	YMLINCNTSV
E1240	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CADYNNNTN	NTSGNWMK	KEIKKNCSEF	ITTDQKQKQ	KEYALFYRLD	IEPIENNTI	YMLINCNTSV
E1243	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTN**NSTN	NNSTARMLVQ	KEIKKNCSEF	ATTIGKQVK	KEYALFNLD	VVPIDNDNTS	YMLINCNTSV
E1251	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTNVNSNTN	SSSSIIKEM	-----	-----	-----	-----	YMLINCNTSV
E1256	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTNANSEIT	NDKNSLKVE	QDAVQNCSEF	MT-----	-----	-----	YMLINCNTSV
E1271	VEQMHEDIIS	LWDQSLKPCV	KLPLCVTLN	CTNVKNNTI	NTISTMLKN	KEIKKNCSEF	ITIGIADKIK	KEYALFYSLD	VT--KVDNTS	YMLINCNTSV

E106B	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEGDI	VIKSENFNN	VKTIIIVQLNS	SVEINCTRP	300
E1061B	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	299
E1077B	ITQACPKVSW	DPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	298
E110	ITQACPKVTF	DPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	297
E1113	ITQACPKVSW	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	296
E1115	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	295
E1127	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	294
E1129	ITQACPKVSW	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	293
E1130	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	292
E1135	ITQACPKVSW	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	291
E1138	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	290
E1149	ITQACPKVSW	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	289
E1145	ITQACPKVSW	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	288
E1212	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	287
E1219	ITQACPKVSW	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	286
E1240	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	285
E1243	ITQACPKVSW	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	284
E1251	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	283
E1256	ITQACPKVTF	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	282
E1271	ITQACPKVSW	EPIPIHYCAP	AGFAILKCN	KEFNGTGPC	NVSTVH--G	IRPVVSTQ	LNGSLAEEV	IIRSQNIDN	TKTIIIVHLNE	SVQINCTRP	281
E1006B	NNTKRSISLA	WGRAIYMEGE	IGDIRQAHN	ISKAKWNNTL	QJIAKILREQ	FKNKTIIFNQ	SAGGDEPIVL	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	400
E1061B	NNTKRSIHLG	PGRAFYATGE	IGDIRQAFCT	LNAIKWNNTL	QJIAKILREH	FKNKTIIFNQ	SAGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	399
E1077B	NNTKRSIHLG	PGRAFYATGE	IGDIRQAHN	VSRQWNNTL	K-----	---NKTIEFT	SSGGDEPIIT	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	398
E110	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	INADWNNTL	GEIVKILGER	FKNKTIIFNQ	SSGGDEPIIT	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	397
E1113	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	LSEAKWNNTL	KIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	396
E1115	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	ISRTAWNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	395
E1127	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	LSKIDWNNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	394
E1129	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	ISKVQWNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	393
E1130	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	LSKIDWNNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	392
E1138	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	VSKDKWSRIL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	391
E1149	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	LSRGNWEDTL	KJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	390
E1145	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	ISKAKWNNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	389
E1212	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	INRTWNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	388
E1219	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	LSAQNNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	387
E1240	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	VSRQWNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	386
E1243	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	LSRGNWEDTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	385
E1251	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	ISREKWNNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	384
E1256	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	ISREKWNNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	383
E1271	NNTKRSIHLG	WGRAFYATGE	IGDIRQAHN	LSAQNNTL	QJIAKILREH	FKNKTIIFNQ	SSGGDEPIVM	HSFNCGGEFF	YCNITIGLFS	SWFNSTWSTE	382