

LUANA PORTES OZÓRIO COELHO

Predição do tropismo viral do HIV-1 utilizando plasma e célula em amostras de sangue de pessoas vivendo com HIV/AIDS.

SÃO PAULO

2013

LUANA PORTES OZÓRIO COELHO

Predição do tropismo viral do HIV-1 utilizando plasma e célula em amostras de sangue de pessoas vivendo com HIV/AIDS.

Defesa de dissertação apresentado ao Programa de Pós-Graduação em Ciências da Coordenadoria de Controle de Doenças da Secretaria de Estado da Saúde de São Paulo, para obtenção de Título de Mestre em Ciências.

Área de Concentração: Pesquisas Laboratoriais em Saúde Pública.

Orientador: Prof. Dr. Luis Fernando de Macedo Brígido

SÃO PAULO

2013

FICHA CATALOGRÁFICA

Preparada pelo Centro de Documentação – Coordenadoria de Controle de Doenças/SES-SP

©reprodução autorizada pelo autor, desde que citada a fonte

Coelho, Luana Portes Ozório

Predição do tropismo viral do HIV-1 utilizando plasma e células de amostras de sangue de pessoas vivendo com HIV/AIDS / Luana Portes Ozório Coelho - São Paulo, 2012.

Dissertação (mestrado) - Programa de Pós-Graduação em Ciências da Coordenadoria de Controle de Doenças da Secretaria de Estado da Saúde de São Paulo.

Área de concentração: Pesquisas Laboratoriais em Saúde Pública
Orientador: Luis Fernando de Macedo Brígido

- | | | |
|--------------------|--------------------------------------|---------------------|
| 1. HIV-1 | 2. Tropismo viral | 3. Receptores CXCR4 |
| 4. Receptores CCR5 | 5. Proteína gp120 do envelope de HIV | |

SES/CCD/CD-269/12

Aos meus pais, Izequiel e Luce Mary, pelo apoio incondicional em todas as etapas da minha vida e por todo carinho e dedicação que me auxiliaram a enfrentar todos os desafios. Vocês me ensinaram o valor do estudo, uma paixão que levarei pelo resto da vida.

Ao meu irmão, Ítalo, com quem compartilho esta conquista, pelo incentivo e apoio neste momento tão especial da minha vida.

Agradecimentos

À Deus, por ter me dado, entre outras dádivas, a oportunidade de vivenciar este processo, que tanto contribuiu para a minha formação profissional e pessoal.

Aos meus pais, **Izequiel Coelho** e **Luce Mary Coelho**, pela infinita dedicação, carinho e apoio incondicional.

Ao meu irmão **Ítalo Coelho**, pelo amor e exemplo de dedicação.

Aos meus melhores amigos, de perto e de longe, antigos e novos. Obrigada pelo apoio em todos os momentos e pelo companheirismo nesta vereda. Amo vocês.

Ao orientador Prof. Luis Fernando de Macedo Brígido. Seus conhecimentos e incentivos constantes foram fundamentais para a realização desta pesquisa. Admiro muito seu trabalho, sua força e sua determinação. Obrigada por tudo!

Aos amigos que fiz, muito mais que colegas, por partilharem seus saberes, experiências e opiniões com o grupo. Estes anos de convivência foram maravilhosos, com muitos momentos alegres e motivos para comemoração. Obrigada por este compartilhamento!

Aos funcionários do laboratório de genotipagem do HIV, pela atenção dispensada sempre com presteza e um sorriso no rosto.

Aos funcionários do Núcleo de Doenças Sanguíneas e Sexuais. Obrigada pelo apoio!

Às pessoas que participaram do estudo, por serem verdadeiros exemplos de vontade de viver e por me fazerem refletir sobre minhas próprias concepções sobre vários conceitos.

Ao Programa de Pós-Graduação da Coordenadora de Controle de Doenças (PPG-CCD) pelo incentivo e apoio financeiro.

À FAPESP, pela bolsa concedida e por proporcionar as melhores condições para a realização deste estudo.

A todos que contribuíram de alguma forma no trilhar desta pesquisa, acreditando no meu sonho. Ele tornou-se realidade por meio desta construção coletiva. Obrigada!

Resumo

O HIV-1 geralmente utiliza o receptor CD4 em conjunto com um correceptor principal, CCR5 ou CXCR4, na etapa fundamental de entrada na célula hospedeira, determinando assim o tropismo viral por células que expressam estas moléculas. Diferenças no tropismo de HIV-1 têm contribuído para a compreensão da patogênese do HIV e são necessárias para suportar a decisão médica em relação ao uso de antagonistas do CCR5. O teste fenotípico de tropismo determina diretamente esta característica biológica, porém os custos e a complexidade restringem a sua utilização, sendo assim, a predição genotípica tem sido proposta como uma alternativa. O objetivo desse estudo foi padronizar e avaliar um teste de predição genotípica do tropismo. Foi adaptado para o nosso laboratório o protocolo desenvolvido pelo Centro de Excelência em HIV/Aids da Columbia Britânica (CECB) usando sequência em replicatas, avaliamos os parâmetros que possam influenciar na predição do tropismo e comparamos a predição na célula e no plasma. Produtos de PCR e plasma enviados pelo CECB foram sequenciados e analisados de forma independente no CECB e no IAL, com alta concordância. Sequências da V3 do envelope viral de 81 pacientes com consentimento informado mostraram uma correlação entre o número de ambiguidades e o menor valor de FPR, uma tendência a ter uma maior distância genética inter-sequência entre os isolados preditos como X4 e sem correlação entre o número de repetições de sequências e o menor FPR. Estudamos também diferenças na origem do material genético viral (RNA viral ou DNA pró-viral) e sua influência sobre a predição de tropismo. Em 56 amostras pareadas observou-se uma tendência na ocorrência de mais X4 em DNA pró-viral de células. Nós observamos maior distância genética entre as sequências de célula em comparação com as de plasma, e documentamos as discrepâncias na determinação do tropismo viral utilizando diferentes *cut offs* de uso clínico. Nosso trabalho permitiu o desenvolvimento, padronização e validação de uma metodologia de baixo

custo para predição genotípica do tropismo. A metodologia desenvolvida está em uso na rede pública estadual de São Paulo, e a transferência da tecnologia para o Ministério da Saúde visa ajudar na organização de uma rede de genotipagem para novos alvos. O seguimento dos pacientes em acompanhamento clínico deve também contribuir para uma melhor compreensão do papel do tropismo de HIV-1 na patogênese viral.

Abstract

HIV-1 generally uses the CD4 receptor in conjunction with a major co-receptor, CCR5 or CXCR4, to enter the host cell, thereby determining the viral tropism for cells expressing these molecules. Differences in this key step of HIV-1 have contributed to understanding the pathogenesis of HIV and are required to support the medical decision regarding the use of CCR5 antagonists. Phenotypic tropism test directly determines this biologic characteristic but cost and complexity restrict its use, and genotypic prediction has been proposed as an alternative. Our goal was to standardize and evaluate a tropism prediction genotypic test. We adapted and validated a protocol developed by the British Columbia Excellence Center using sequence replicates, evaluated parameters that might influence tropism prediction and compared cell and plasma prediction. Plasma and PCR products sent by BCEC were sequenced and analyzed independently here and at BCEC, with high concordance. Viral envelope V3 sequences from 81 informed consent patients showed a correlation of sequence ambiguity to a low FPR rate, a tendency to have a higher inter-sequence genetic distance among predicted X4 isolates and no correlation of the number of replicates to a lower FPR. We also studied differences in viral genetic material (virion RNA or proviral DNA) and its influence on the prediction of tropism. In 56 paired analysis we observed a tendency to have more X4 predictions among proviral DNA from cells, a higher genetic distance in replicates from proviral and documented discrepancies in viral tropism prediction from paired cell/plasma determination at difference clinically used prediction cutoffs. Our work has allowed the development, standardization and validation of a low cost genotypic tropism methodology that is currently available at the public health system the state of Sao Paulo, the evaluation of genetic parameters associated with this prediction and the technology was transferred to the Ministry of Health for helping the organization of a network genotyping for new targets for. Long term follow-up of patients should also contribute to a better understanding of the role of tropism of HIV-1 in viral pathogenesis.

Lista de abreviaturas, siglas e símbolos

°C	graus Celsius
µL	microlitros (10^{-6} litro)
AIDS	Síndrome da Imunodeficiência Adquirida (<i>acquired immune deficiency syndrome</i>)
ARV	Terapia Anti-retroviral
CDC	Centro de Controle e Prevenção de Doenças (<i>Centers for Disease Control of Prevention</i>)
CECB	Centro de Excelência em HIV/Aids da Columbia Britânica
DNA	Ácido desoxirribonucleico (<i>deoxyribonucleic acid</i>)
EDTA	Ácido etilenediaminetetracético
<i>env</i>	Gene que codifica as glicoproteínas do envelope
<i>gag</i>	Gene que codifica proteínas estruturais
GENDIST	Distância genética entre as sequencias em nucleotídeos.
HAART	Terapia Anti-retroviral Altamente Ativa (<i>Highly Active Anti-Retroviral Therapy</i>)
HIV-1	Vírus da Imunodeficiência Humana tipo-1 (<i>Human Immunodeficiency Virus</i>)
HIV-2	Vírus da Imunodeficiência Humana tipo-2 (<i>Human Immunodeficiency Virus</i>)
mM	Milimolar (10^{-3} molar)

MNS	Mistura não sinônima
<i>nef</i>	Gene <i>negative factor</i>
ng	Nanogramas (10^{-9} gramas)
Pb	Pares de base
PCR	Reação em Cadeia da Polimerase
pM	picomol (10^{-12} mol)
<i>pol</i>	Gene polimerase, codifica as enzimas virais protease, transcriptase reversa e integrase
PRODIST	Distância de genética entre as sequencias em aminoácidos.
PT	Enzima viral protease
R5	Vírus com tropismo pelo correceptor CCR5
R5X4	Vírus com tropismo por ambos correceptores (CCR5 e CXCR4)
rcf	<i>Centrifugal Force Relative</i>
<i>rev</i>	Gene regulador da expressão viral
RNA	Ácido ribonucleico (<i>ribonucleic acid</i>)
RT	Enzima viral transcriptase reversa
SIV	<i>Simian immunodeficiency virus</i>
<i>tat</i>	Gene que codifica um transativador transcricional
U	Unidades
Vif	Proteína viral - Fator de infectividade viral – <i>viral infective factor</i>
Vpr	Proteína viral R

Vpu

Proteína viral U

X4

Vírus com tropismo pelo correceptor CXCR4

Lista de Figuras

Figura 1:	Representação esquemática da estrutura do HIV-1.	21
Figura 2:	Representação esquemática do genôma do HIV-1.	22
Figura 3:	Representação esquemática do gene envelope do HIV-1.	26
Figura 4A:	Esquema das alças da gp120.	29
Figura 4B:	Representação esquemática da sequência de aminoácidos no domínio V3.	29
Figura 5:	Etapas da replicação do HIV-1	30
Figura 6:	Esquema da ação do maraviroque	36
Figura 7:	Gel de agarose	42
Figura 8:	Exemplo de resultado emitido pelo geno2pheno	45
Figura 9:	Esquema de cálculo da distância genética	46
Figura 10:	Árvore filogenética representativa demonstrando a diversidade entre as replicatas.	52

Figura 11:	Correlação entre a contagem de CD4 (cel/mm ³) e carga viral (log ₁₀ /mL) em relação ao número de ambiguidades e a média das distâncias (GENDIST e PRODIST).	53
Figura 12:	Correlação linear entre a taxa de falso positivo	54
Figura 13:	Média da distância genética (PRODIST) nas amostras classificadas como X4 e R5 usando FPR 20%.	55
Figura 14:	Valores de FPR _{clonal} da sequência da alça V3 de acordo com o número de sequências obtidas.	55
Figura 15:	Correlação entre os valores de menor FPR _{clonal} das amostras de DNA proviral em relação aos valores de menor FPR _{clonal} das sequências obtidas a partir do RNA do vírus circulante no plasma.	58
Figura 16:	Teste <i>Wilcoxon signed rank</i> do desvio padrão dos valores de FPR _{clonal} das sequências obtidas de RNA viral e DNA pró viral pareadas.	59
Figura 17:	Correlação entre os valores de FPR _{clonal} no plasma e célula e o número de ambiguidade na V3	59

Figura 18:	Correlação entre os valores de menor FPR_{clonal} no plasma e célula e a média dos valores da distância genética.	60
Figura 19:	Média da Distância genética (PRODIST) entre as sequências em replicata obtidas a partir de virion e proviral.	60
Figura 20:	Correlação linear entre os valores de CD4, carga viral e distância genética (PRODIST).	61
Figura 21:	Árvore filogenética das amostras obtidas à partir de RNA viral e DNA proviral.	62

Lista de tabelas

Tabela 1:	Principais funções das proteínas codificadas pelo genoma do HIV-1.	24
Tabela 2:	Casos de AIDS por região geográfica do Brasil, período de 1980 a 2011.	33
Tabela 3:	Resultados dos valores de FPR-CECB e FPR- IAL .	49
Tabela 4:	Dados clínicos e demográficos. Características dos pacientes de acordo com o número de sequências obtidas no estudo de replicatas.	51
Tabela 5:	Dados clínicos e demográficos. Características dos pacientes de acordo com o número de sequências obtidas no estudo de comparação entre plasma e célula.	57
Tabela 6:	Comparação da predição de tropismo em células e plasma.	58

Lista de Anexos

- Anexo 1: Termo de Consentimento Livre e Esclarecido
- Anexo 2: Aprovação do Comitê de Ética
- Anexo 3: Teste genotípico para predição do tropismo do HIV-1
- Anexo 4: Concordance of HIV-1 tropism phenotype to predictions using web-based analysis of V3 sequences: Composite algorithms may be needed to properly assess viral tropism.
- Anexo 5: Evaluation of genotypic prediction of hiv-1 tropism using population (bulk) direct sequencing of replicates
- Anexo 6: Env V3 sequence based proviral DNA and virion RNA tropism prediction discordance using replicates

Índice

1. Introdução	18
1.1. Histórico	18
1.2. Classificação e Morfologia	20
1.3. Organização do genoma do HIV-1	22
1.4. Proteínas do envelope	25
1.5. Alça V3	27
1.6. Ciclo de replicação do HIV-1	29
1.7. Epidemiologia	32
2. Objetivos	37
2.1. Geral	37
2.2. Específicos	37
3. Metodologia	38
3.1. População Estudada	38
3.2. Coletas do Material Biológico	39
3.3. Extração do RNA virion e DNA proviral	39
3.4. Reação em Cadeia da Polimerase (PCR)	40
3.4.1. Amplificação do RNA viral	40
3.4.2. Amplificação do DNA proviral	41
3.4.3. Eletroforese em gel de agarose	41
3.4.4. Sequenciamento do produto amplificado	42
3.4.5. Predição genotípica do Tropismo Viral	43
3.4.6. Distância Genética	46
3.4.7. Análise Estatística	47

3.4.8. Aprimoramento do teste genotípico	47
4. Resultados	49
4.1. Aprimoramento do Teste genotípico para predição do tropismo do HIV-1 do Instituto Adolfo Lutz (IAL) por avaliação externa do Centro de Excelência em HIV/AIDS de <i>British Columbia</i> (CECB)	49
4.2. Avaliações da predição genotípica do tropismo do HIV-1 utilizando sequenciamento em replicata	50
4.2.1. Valores de FPR	54
4.3. Estudo comparativo entre plasma e célula na predição do tropismo viral utilizando sequências em replicata.	56
5. Discussão	63
5.1. Avaliações da predição genotípica do tropismo do HIV-1 utilizando sequenciamento simples e em replicata	65
5.2. Estudo comparativo entre plasma e célula na predição do tropismo viral utilizando sequências em triplicata	68
6. Conclusão	70
7. Referências	71

1. Introdução

1.1. Histórico

O HIV (*Human Immunodeficiency Virus*) é o agente etiológico da Síndrome da Imunodeficiência Adquirida (*acquired immune deficiency syndrome* - AIDS) em humanos. Essa doença é caracterizada por imunossupressão acompanhada de infecções oportunistas, ou pela ocorrência de um câncer raro, o Sarcoma de Kaposi. A infecção pelo HIV é crônica e a progressão da doença é lenta (Prusiner, 2002).

Os primeiros relatos de casos de AIDS surgiram em 1981 e eram relacionados a homens que faziam sexo com outros homens. A imunodeficiência era associada ao aparecimento de um tipo raro de câncer, o Sarcoma de Kaposi, e de uma série de outras infecções oportunistas que afetavam os indivíduos infectados. Logo, um grande número de casos da doença foi registrado e a AIDS atingiu proporções epidêmicas. O surgimento de casos em diferentes populações, incluindo homens heterossexuais, usuários de drogas, hemofílicos, crianças e receptores de transfusões de sangue, levou à identificação do mesmo agente etiológico, que era transferido através de fluídos corporais às pessoas infectadas (Prusiner, 2002).

Dois grandes grupos de pesquisa se esforçaram para realizar a identificação e o isolamento do HIV-1. O primeiro liderado por Luc Montagnier e seus colegas, do Instituto Pasteur, em Paris. O segundo, comandado por Robert Gallo, do Centro para Controle e Prevenção de Doenças (Center for Disease Control and Prevention - CDC), dos Estados Unidos. Além disso, outras descobertas da época foram cruciais para o isolamento do HIV, tais como, o isolamento da enzima transcriptase reversa (Baltimore, 1970; Temin & Mizutani, 1970) e a descoberta da interleucina-2 (Morgan et al., 1976), fator de crescimento essencial à cultura de célula de linfócitos T necessárias à replicação do HIV in vitro (Prusiner, 2002).

Luc Montagnier estudava os retrovírus causadores de câncer em humanos quando do surgimento da AIDS. Em 1982, começou a pesquisar a causa da doença, devido à hipótese levantada por Robert Gallo de que essa enfermidade seria provocada por um retrovírus. A partir de uma amostra da biópsia de um paciente com linfadenopatia, Montagnier e seus colaboradores fizeram um cultivo de linfócitos T. Em janeiro de 1983, foram descobertos no cultivo a atividade da transcriptase reversa, enzima presente apenas nos retrovírus, o que confirmou a hipótese de Gallo. Esse vírus foi isolado e denominado como vírus associado à linfadenopatia (*lymphadenopathy associated virus* - LAV) (Montagnier, 2002).

Em paralelo, o grupo de pesquisa de Gallo também estudava os retrovírus e foi o primeiro a isolar o HTLV (*humam T-lymphotropic virus*-HTLV), vírus causador de leucemia associada à células T em humanos. Ele defendia a hipótese de que a AIDS era causada por um vírus da mesma família do HTLV, pois o modo de transmissão era o mesmo e ambos se replicavam em linfócitos. Entretanto, o agente etiológico da AIDS possuía várias características distintas dos vírus da família HTLV. Esses não causam efeito citopático nas células e sim provocam sua imortalização. Assim, era pouco provável que a grande queda no número de células T fosse provocada por um HTLV. Posteriormente, foi comprovado que os dois grupos de pesquisa haviam descrito o mesmo vírus. Em 1984, Jay Levy e seus colaboradores da Universidade da Califórnia, isolaram o vírus a partir de indivíduos assintomáticos, e denominaram de retrovírus associado à AIDS (*AIDS-associated retrovirus*-ARV). Ainda em 1984, a associação do HIV como agente etiológico da AIDS foi aceita pela comunidade médica e científica (Gallo, 2002; Montagnier, 2002; Prusiner, 2002).

Em 1986, o Comitê Internacional de Taxonomia de Vírus (*International Commitee on Taxonomy of Viruses* – ICTV) recomendou a utilização da denominação de vírus da imunodeficiência humana (*Human immunodeficiency virus* – HIV). Nesse mesmo ano, outro tipo de HIV foi isolado e denominado vírus da imunodeficiência do tipo 2, *Human*

immunodeficiency virus type 2 – HIV-2. Esse vírus também causa a AIDS, contudo, de uma forma mais branda e restrita ao oeste da África.

Apesar da não existência de uma vacina preventiva contra a infecção por HIV, grandes avanços já foram alcançados no controle da AIDS. A partir de 1986, os testes sorológicos para detecção do HIV tornaram-se comercialmente disponíveis, o que reduziu significativamente a transmissão do vírus por transfusão de sangue em países desenvolvidos. Em 1987, foi desenvolvido o AZT, primeira droga anti-HIV. Em 1995, foi disponibilizada a terapia antirretroviral altamente ativa (HAART) que melhorou significativamente a vida dos portadores do HIV (Gallo & Montagnier, 2002).

1.2. Classificação e Morfologia

O HIV é um vírus com genoma de RNA de fita simples, pertencente ao gênero *Lentivirus* e à família *Retroviridae*. Os retrovírus utilizam a enzima transcriptase reversa para a transcrição do RNA em DNA. Após a síntese do DNA, o genoma viral se integra ao genoma da célula hospedeira e recebe o nome de provírus. Os retrovírus pertencentes ao gênero *Lentivirus*, do latim (lentos), receberam essa denominação porque possuem um longo período de incubação no hospedeiro, podendo persistir e replicar-se por muitos anos antes de causar sinais clínicos de doença. Os lentivírus produzem efeitos citopáticos, provocam deficiências imunológicas, desordens hepáticas e nervosas, além de doenças autoimunes. Os genomas desses vírus são relativamente grandes, apresentando um número maior de genes do que outros retrovírus. Além dos genes estruturais, *gag*, *pol* e *env*, comum a todos os retrovírus, os lentivírus codificam proteínas acessórias e regulatórias da replicação viral. A descoberta do HIV levou à caracterização de outros lentivírus, como aqueles isolados de primatas não humanos, SIV (*simian immunodeficiency virus*) (Janeway et al., 2002; Flint et al., 2004). Como já mencionado, estudos genômicos e sorológicos demonstraram a existência de dois tipos distintos de HIV: o HIV-1 e o HIV-2. A caracterização de vários isolados do HIV-1 permitiu identificação de quatro grupos: M

(*major*), O (*outlier*), N (*non M, non O*) e P (*putative*) (van Regenmortel et al., 2000; Janeway et al., 2002; Flint et al., 2004; Plantier et al., 2009).

A partícula do HIV-1 possui em seu interior duas cópias de RNA de fita simples. O virion possui um diâmetro de aproximadamente 100 nm, um genoma de cerca de 9,8 kb e um envelope, oriundo da célula hospedeira. As glicoproteínas virais são derivadas de um precursor gp160, que clivado por proteases celulares, produzindo as glicoproteínas gp120 de superfície (SU) e gp41, a glicoproteína transmembrana (TM). No interior do envelope encontra-se um segundo envoltório, a matriz, que é constituída pela proteína p17. O capsídeo é formado pela proteína p24, e no interior dele encontram-se as duas cópias do genoma de RNA fita simples, com polaridade positiva, associadas às proteínas p7 e p6, formando o nucleocapsídeo. As enzimas virais transcriptase reversa (RT), integrase (IN) e protease (PR) estão presentes dentro da partícula viral (Figura 1) (Larder et al., 2001).

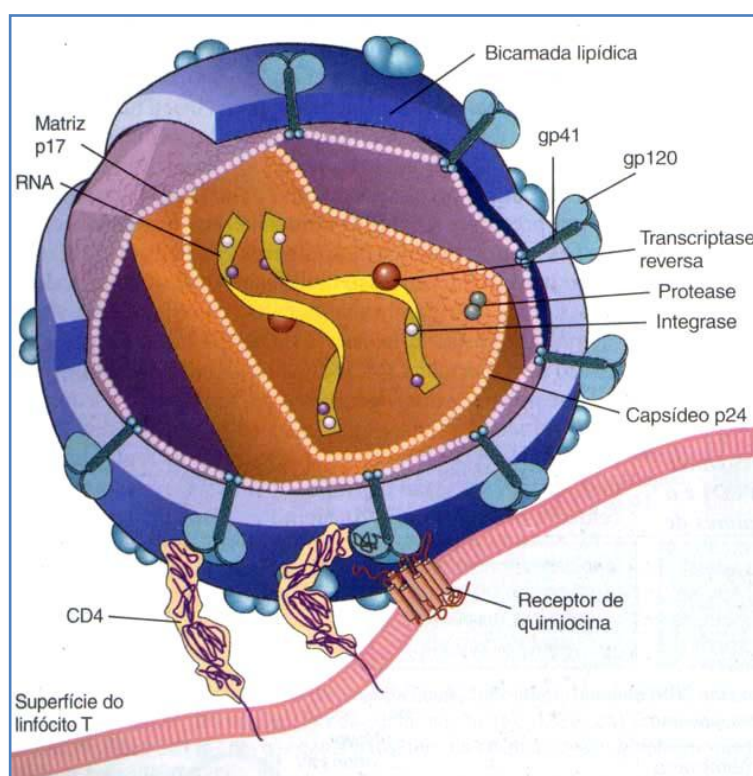


Figura 1: Representação esquemática da estrutura molecular do HIV-1 (Abbas et.al., 2005).

1.3. Organização do genoma do HIV-1

O genoma do HIV-1 contém pelo menos nove genes diferentes, sendo três deles encontrados em todos os retrovírus, são eles: *gag* (que codifica as proteínas estruturais), *pol* (que codifica as enzimas virais protease [PT], transcriptase reversa [RT] e integrase [IN]) e *env* (que codifica as glicoproteínas do envelope). Esses são os genes denominados estruturais. (Flint et al., 2004).

Há ainda os genes regulatórios, *tat* e *rev*, além de quatro genes acessórios: *nef*, *vif*, *vpr* e *vpu*. Uma diferença no genoma do HIV-1 e HIV-2 está na presença do gene *vpx* no HIV-2. Nas extremidades 3' e 5' do provírus são encontradas longas sequências repetidas denominadas de LTRs (*long terminal repeats*), que regulam a integração do vírus ao genoma hospedeiro, a expressão gênica e a replicação (Figura 2) (Larder et al., 2001; Wigg, 2002; McBurney & Ross, 2008). As funções desses genes estão resumidas na Tabela 1.

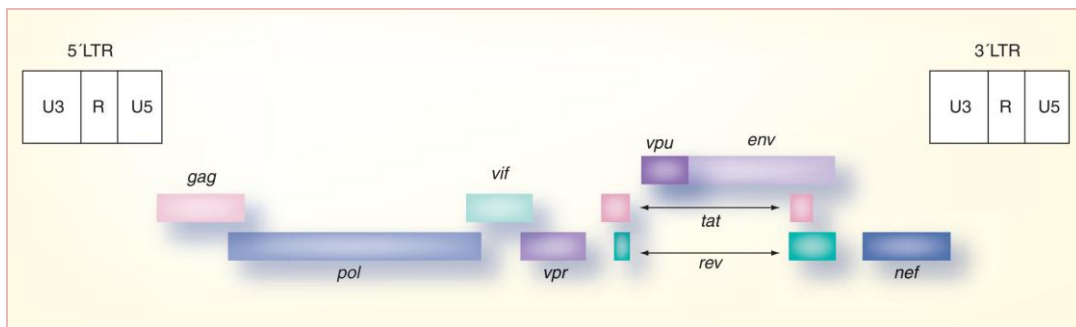


Figura 2: Representação esquemática da organização genômica do HIV-1 (McBurney & Ross, 2008).

O gene *gag* é traduzido em uma proteína precursora de 55kD, denominada p55 gag. Essa é clivada pela protease em quatro proteínas menores: p17 que constitui a matriz; p24 constitui o capsídeo; p7 constitui o nucleocapsídeo e p6 que auxilia a incorporação da proteína Vpr aos virions. Essas quatro proteínas estão presentes nas partículas maduras do HIV-1 (Göttlinger, 2001; Larder et al., 2001).

O gene *pol* codifica uma poliproteína precursora das enzimas virais PR, RT e IN. A poliproteína Pol é traduzida e fusionada a Gag, formando o precursor Gag-Pol. A protease viral cliva o polipeptídeo Pol, separando-o de Gag, e, posteriormente, cliva Pol para gerar PR, RT e IN. É somente no meio extracelular que, após a liberação do vírus da célula hospedeira, que a protease viral sofre auto-ativação e inicia o processo de clivagem, tornando infecciosa a partícula viral (Larder et al., 2001).

O gene *env* codifica uma poliproteína precursora, que é sintetizada na forma de trímero, com ligações não covalentes entre as subunidades gp120 e gp41. A função e a estrutura dessas proteínas serão apresentadas de forma detalhada adiante, na página 25 (Flint et al., 2004; Moore & Doms, 2003).

Tabela 1: Principais funções das proteínas codificadas pelo genoma do HIV-1 (adaptado de Watts et al., 2009)

Gene		Função
<i>Gag</i>	Antígeno grupo-específico	Proteínas do <i>core</i> da matriz.
<i>Pol</i>	Polimerase	Enzimas transcriptase reversa, proteínas e integrase.
<i>Env</i>	Envelope	Glicoproteínas: gp120 se liga ao CD4 e a receptores de quimiocinas e gp41 participa na fusão entre as membranas do vírus e da célula hospedeira.
<i>Tat</i>	Transativador	Regula positivamente a transcrição.
<i>Rev</i>	Regulador da expressão viral	Permite a exportação de transcritos não processados no núcleo da célula hospedeira.
<i>Vif</i>	Fator de infectividade viral	Aumenta a capacidade infectiva das partículas virais.
<i>Vpr</i>	Proteína viral R	Promove a importação nuclear de DNA viral e a parada do ciclo celular da célula hospedeira em G2.
<i>Vpu</i>	Proteína viral U	Promove a degradação intracelular de CD4 e aumenta a liberação de partículas virais na membrana celular da célula hospedeira.
<i>Nef</i>	Fator de regulação negativa	Regula negativamente a expressão de CD4 e de MHC classe I da célula hospedeira e aumenta a infectividade viral.

1.4. Proteínas do envelope

As glicoproteínas do envelope possuem papel crucial nos eventos iniciais da infecção pelo HIV, mediando a ligação do vírus à célula hospedeira e a fusão das membranas viral e celular. As proteínas do envelope também são os principais alvos para a resposta imune anti-HIV nos indivíduos infectados, devido a sua localização na superfície da partícula viral. Essas glicoproteínas são formadas por duas subunidades, gp120 e gp41, que são produto da clivagem por proteases celulares da poliproteína precursora gp160. As proteínas gp120 e gp41 formam um complexo trimérico no envelope viral, permanecendo ligadas não covalentemente após a clivagem proteolítica (Flint et al., 2004; Moore & Doms, 2003).

A glicoproteína gp120 permite o reconhecimento da célula alvo e determina o tropismo celular pela ligação ao receptor CD4 e a um dos vários correceptores de quimiocinas, mais frequentemente CCR5 ou CXCR4. A glicoproteína gp41 promove a fusão da membrana viral com a celular, processo que resulta na liberação do nucleocapsídeo viral no interior da célula hospedeira. Durante esse processo, gp41 sofre várias mudanças conformacionais que levam à exposição de um peptídeo de fusão, uma região hidrofóbica rica em resíduos de glicina, essencial ao processo de fusão das membranas viral e celular. A glicoproteína gp41 possui em sua estrutura uma porção citoplasmática, que ancora o complexo das proteínas do envelope na matriz do virion, uma região transmembrana hidrofóbica, e um ectodomínio onde se localiza o peptídeo de fusão (Chan et al., 1997; Golding et al., 2002; Kang et al., 2005).

A proteína gp120 é composta por cinco domínios constantes (C1-C5) e por cinco *alças* variáveis (V1-V5) intercalados. Esses domínios V1-V5 possuem grande variabilidade genética. A Figura 3 apresenta um esquema da organização estrutural de gp120 (Wyatt & Sodroski, 1998).

Análises estruturais da gp120 demonstraram a existência de pontes intramoleculares de dissulfeto, que permitem a incorporação das quatro primeiras regiões variáveis em estruturas em forma de alças expostas na

superfície da glicoproteína. A região mais conservada dessa proteína, o *core*, é formada por dois domínios distintos, um externo exposto na superfície viral, e um interno que se liga não covalentemente à gp41. No domínio externo, localizam-se os sítios de ligação à molécula de CD4 e aos correceptores, principalmente CCR5 e CXCR4 (Wyatt & Sodrosky, 1998).

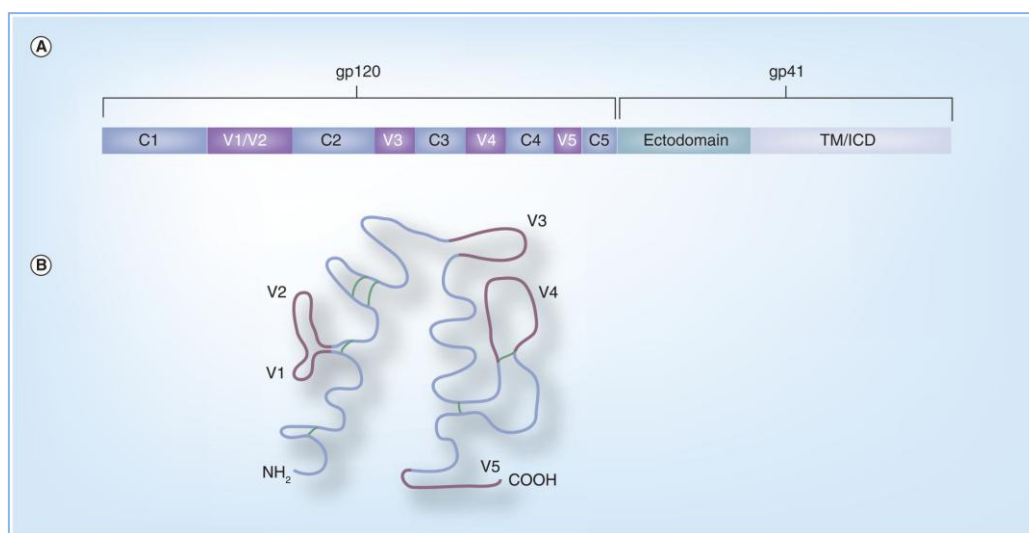


Figura 3: Representação esquemática do gene envelope do HIV-1 (A) representa a clivagem da glicoproteína gp160 em gp120 e gp41, bem como as respectivas regiões. (B) esquema das alças da gp120. As regiões constantes são representadas em azul e as regiões variáveis em roxo. Pontes dissulfeto são mostradas em verde. ICD: Domínio intracitoplasmático (McBurney & Ross, 2008)

A função que a gp120 desempenha no início da infecção viral e sua constante exposição ao sistema imune faz desta proteína um importante alvo para vacinas. No entanto, as primeiras vacinas baseadas em Env, compostas por subunidades monoméricas de gp120 se mostraram ineficientes (Burton et al., 2004; Herrera et al., 2005).

1.5. Alça V3

O alça V3 é uma região hipervariável, localizada na porção externa da glicoproteína do envelope, gp120 (Figura 4A). Essa região possui de 35 a 40 aminoácidos e varia muito entre diferentes linhagens virais. Entretanto, o alça V3 apresenta resíduos de aminoácidos relativamente conservados, em seu ápice se destacam GPGR ou GPGQ, que são essenciais ao reconhecimento pelo HIV-1 do correceptor. Além disso, existe outra região conservada, uma ponte dissulfeto, Cys296-Cys331 na base do alça V3 (Kwong, 2004; Stanfield et al., 2004). Vários estudos têm demonstrado a importância biológica da heterogeneidade do alça V3 em diversas funções virais, tais como, na ligação ao correceptor tanto CCR5 como CXCR4, que determina o tropismo celular. Nessa região também são encontrados muitos epitopos lineares e conformacionais indutores de anticorpos neutralizantes anti-HIV-1 e de resposta de linfócitos T citotóxicos. Devido à capacidade de indução de anticorpos neutralizantes que os peptídeos de V3 possuem, essa região foi conhecida por muito tempo como “determinante principal de neutralização” (*principal neutralizing determinant-PND*) (Casseb et al., 2002; Stanfield et al., 2004; Zhi-young et al., 2004).

Além do potencial imunogênico, o *alça* V3 tem papel importante no fenótipo indutor de sincícios, no tropismo viral e na determinação do tipo de correceptor a ser utilizado pelo vírus, se CCR5 ou CXCR4. Anticorpos anti-V3 impedem a entrada do vírus na célula porque interferem na ligação gp120-CCR5 (Fouchier, 1995; Trkola et al., 1996; Wyatt & Sodroski, 1998; Mouquet et al., 2012).

Os isolados de HIV-1 podem ser classificados em duas categorias de acordo com os efeitos citopáticos induzidos: indutor de sincício (SI) e não indutor de sincício (NSI). Essa característica está relacionada ao tipo celular infectado e às interações com o correceptor. Isolados primários de HIV-1, em sua maioria, utilizam o correceptor CCR5 para entrada na célula. As linhagens virais que utilizam esse correceptor são denominadas R5. Para a maioria desses isolados, que são transmitidos e predominam durante a fase inicial da doença, a presença de CCR5 é crucial para a infecção da célula. Existem relatos de indivíduos que apresentam resistência à infecção pelo HIV-1 devido às falhas na expressão de CCR5 em suas células. Essas linhagens virais infectam preferencialmente macrófagos e não induzem a formação de sincícios. Isolados de HIV-1, que surgem na fase mais tardia da infecção, utilizam, mais frequentemente, o correceptor CXCR4 ao invés de CCR5, ou ambos, e são denominados X4 ou R5/X4, respectivamente. Variantes virais X4 estão associadas à rápida progressão da AIDS, infectam os linfócitos T e produzem efeito citopáticos, formando sincícios (Milich et al, 1993; Fouchier, 1995).

Pouco se sabe sobre o mecanismo de variação desses vírus, como ocorre a mudança do uso do correceptor CCR5 para CXCR4, e as alterações na patogenicidade viral. No entanto, algumas alterações pontuais no *alça* V3 estão associadas à um determinado fenótipo, ou ao uso de CCR5 ou CXCR4. Aminoácidos nas posições 11 e/ou 25 (aminoácidos 306 e 320 na sequência referência HXB2) carregados positivamente são frequentemente encontrados nos vírus que utilizam CXCR4. Além disso, o aumento da carga positiva nos aminoácidos do *alça* V3 também está associado ao vírus X4. Essa relação foi encontrada nos *alça* V3 dos subtipos A,B,C,D, CRF01_AE, F e G. Dessa maneira, variações genéticas entre os subtipos podem alterar a patogenicidade viral., A Figura 4B apresenta um esquema da sequência de aminoácidos do *alça* V3. (Fouchier, 1995; Koning et al., 2002; Thomson et al.,2002; McBurney & Ross, 2008).

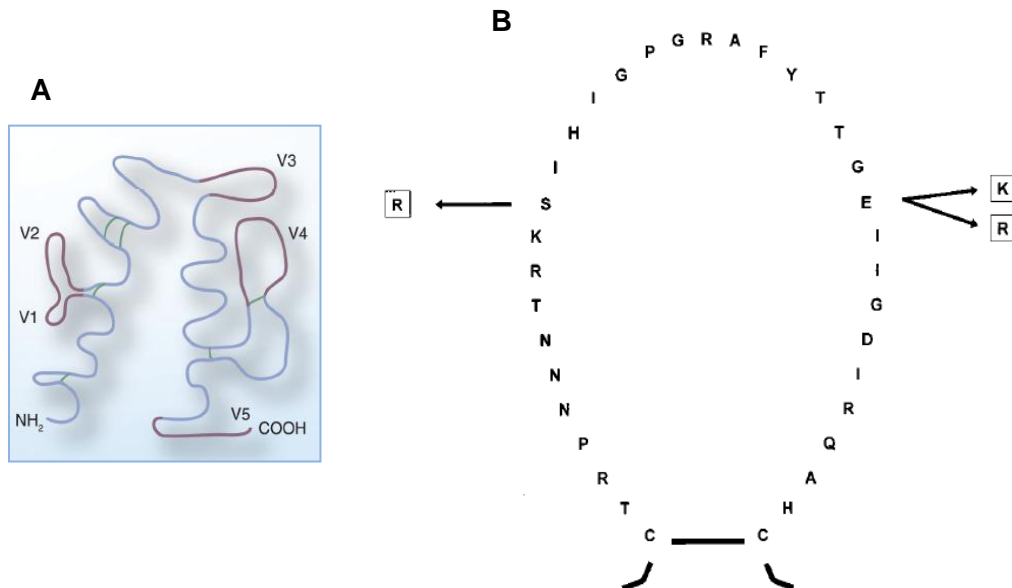


Figura 4: (A) esquema das alças da gp120. As regiões constantes são representadas em azul e as regiões variáveis em roxo. Pontes dissulfeto são mostradas em verde. (B) representação esquemática da sequência de aminoácidos no domínio V3, destacando a arginina (R) na posição 306 da V3 e/ou arginina ou lisina (K) na posição 320. (McBurney & Ross, 2008; Fouchier et al.,1995).

1.6. Ciclo de replicação do HIV-1

A infecção das células pelo HIV-1 tem início com a entrada do vírus na célula. Esse processo ocorre através de interações consecutivas entre as glicoproteínas do envelope (gp 120 e gp 41), os receptores de membrana CD4, além de um correceptor, CCR5 ou CXCR4, que pertencem à família dos receptores de quimiocinas. A expressão conjunta das moléculas de CD4 e um dos correceptores de quimiocinas possibilitaram a identificação de monócitos, macrófagos, células da glia, dendríticas, de Langerhans e linfóides, como alvos da infecção viral. Inicialmente gp120 se liga à molécula de CD4, tal ligação promove mudanças conformacionais nessa glicoproteína que expõe novos sítios de ligação e permite a interação da gp 120 com o

correceptor (CCR5 ou CXCR4). Essas interações promovem mudanças conformacionais na gp41 e a exposição de uma região hidrofóbica, chamada de peptídeo de fusão, que se insere na membrana celular e permite a fusão da membrana viral com a da célula-alvo. Esse processo é mediado pela formação de um complexo de seis hélices, três de origem C-terminal e três de origem N-terminal, organizado em um arranjo antiparalelo. Devido à essa transição, o peptídeo de fusão e o domínio transmembrana de gp 41 ficam próximos, assim como a membrana celular e a viral. O ganho de energia livre proveniente desse processo permite a fusão das duas membranas (Figura 5) (Moore & Doms, 2003; Abbas & Lichtman, 2005).

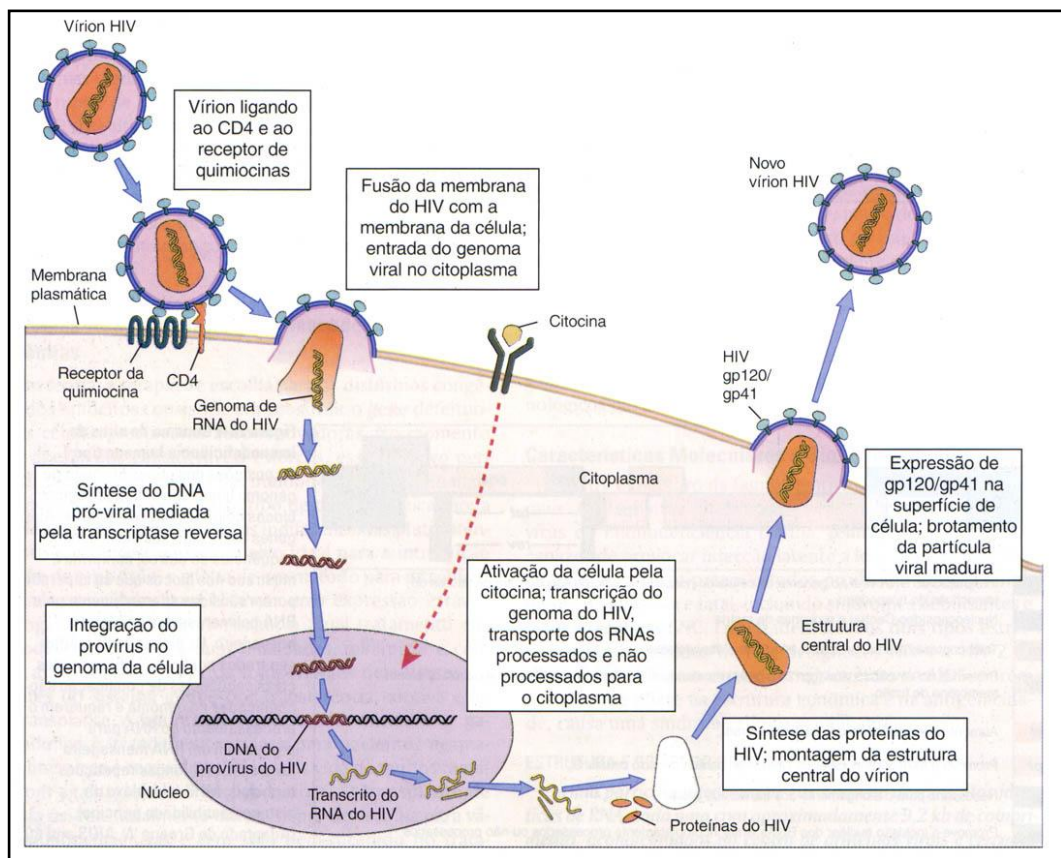


Figura 5: Etapas da replicação do HIV-1 (Abbas & Lichtman, 2005).

Após a entrada do *virion* na célula, as enzimas presentes no nucleocapsídeo se tornam ativas e iniciam o ciclo de replicação viral. O genoma é transcrito pela enzima viral transcriptase reversa. O produto desta enzima é um DNA linear de dupla fita que penetra no núcleo e catalisa a inserção do DNA viral no genoma da célula hospedeira (Flint et al., 2004; Abbas & Lichtman, 2005)

O início da transcrição dos genes do HIV-1 nas células T está associado à ativação fisiológica das células CD4 por antígenos e citocinas. A ativação dessas células induz a produção de diversos fatores de transcrição, que se ligam não somente aos promotores no DNA celular, mas também, a sítios localizados na LTR do provírus, iniciando a transcrição pela RNA polimerase celular. O produto da transcrição é processado de várias maneiras para produzir mRNAs que codificam todas as proteínas virais (Janeway et al., 2002; Abbas & Lichtman, 2005)

As proteínas virais Tat e Rev são essenciais à replicação viral nas células T ativadas, Tat é um potente regulador transcricional, que funciona como fator de alongamento, ligando-se ao mRNA incipiente e melhorando a transcrição deste pela RNA polimerase. A proteína Rev permite a exposição para fora do núcleo de mRNAs não processados para a tradução de diversas proteínas virais (Flint et al., 2004).

A expressão gênica do HIV pode ser dividida em fase precoce onde os genes regulatórios *rev*, *tat* e *nef* são expressos, e em uma fase tardia, em que os genes estruturais são expressos e genomas completos são produzidos. A mudança de fase de expressão gênica é iniciada pela produção e acúmulo de proteína Rev, que promove a exportação dos mRNAs tardios para fora do núcleo (Flint et al., 2004).

A montagem dos *virions* começa de um acúmulo de transcritos completos do genoma de RNA do provírus dentro de um complexo de nucleoproteínas, formado pelas proteínas básicas de *gag* e pelas enzimas codificadas por *pol*, necessárias ao ciclo seguinte de integração. Esse nucleocapsídeo é envolvido por um envelope, oriundo de membrana plasmática da célula no qual estão presentes tanto proteínas de membrana

celulares como glicoproteínas gp 120 e gp 41, e liberado por brotamento da célula hospedeira (Abbas & Lichtman, 2005; Janeway et al., 2002)

1.7. Epidemiologia

De acordo com os dados do Programa das Nações Unidas para HIV e aids, o UNAIDS (do inglês: *Joint United Programme on HIV/AIDS*), quase 60 milhões de pessoas foram infectadas pelo HIV e 25 milhões de pessoas morreram de causas relacionadas com o HIV. Estima-se que cerca de 33,4 milhões de indivíduos estejam infectados pelo HIV em todo globo terrestre. Apenas em 2008 houve 2,7 milhões (2,4 milhões - 3,0 milhões) de novas infecções e 2 milhões (1,7 a 2,4 milhões) de mortes em decorrência da AIDS (UNAIDS 2010).

Em 2008, cerca de 430.000 (240.000-610.000) crianças nasceram com HIV, elevando para 2,1 milhões (1,2 a 2,9 milhões) o número total de crianças menores de 15 anos vivendo com HIV. A África Subsaariana é a região mais afetada, respondendo por 67% das pessoas convivendo com HIV/aids no mundo e por 91% das novas infecções entre as crianças. Na África Subsaariana a epidemia deixou órfãs mais de 14 milhões de crianças.

No Brasil, desde a identificação dos primeiros casos de AIDS, de 1980 a 2011, foram notificados um total de 609 mil casos de AIDS. A tabela 2 representa o número de casos de AIDS nas regiões do Brasil, no período entre 1980 e 2011.

Tabela 2: Casos de AIDS por região geográfica do Brasil, período de 1980 a 2011 (Boletim Epidemiológico AIDS-DST 2011)

Região	Número
Norte	28, 248
Nordeste	78, 686
Centro-Oeste	35, 116
Sudeste	343, 095
Sul	123, 069
Total	608, 214

Atualmente, cerca de 220 mil pessoas estão submetidas ao tratamento antirretroviral no Brasil e cerca de 30 mil iniciam tratamento anualmente. Estudo realizado pelo Departamento DST, AIDS e Hepatites Virais (DDST/AIDS/HV) da Secretaria de Vigilância em Saúde (SVS) demonstrou maior tempo de sobrevida entre pacientes que vivem com HIV/AIDS cujo diagnóstico é estabelecido em períodos mais recentes. Aqueles com diagnóstico entre 1998 e 1999, apresentam média de sobrevida de 107 meses, bastante superior aos 58 meses daqueles com diagnóstico em 1996 (Boletim Epidemiológico, 2009).

Entre os determinantes de aumento da sobrevida inclui-se a introdução de antirretrovirais de terceira linha, especialmente novas classes de antirretrovirais. As recomendações de tratamento da multirresistência permitem obter supressão viral duradoura (carga viral indetectável), que está associada ao impacto favorável na mortalidade ao longo do tempo. O convívio com a replicação, resistência viral e falha terapêutica apresenta risco relativo

de morte de 1,21 para cada 3 meses de atraso em realizar a mudança de tratamento, além de ocasionar desenvolvimento de multirresistência (Petersen M et al., 2008).

Entre os medicamentos de novas classes estão os antagonistas de correceptor de quimiocina C-C tipo 5 que inibem especificamente a entrada na célula CD4 e, portanto, bloqueiam a replicação da variante R5 do HIV após ligação ao correceptor transmembrana CCR5. O inibidor de CCR5 disponível comercialmente é o maraviroque (MVQ), Figura 6.

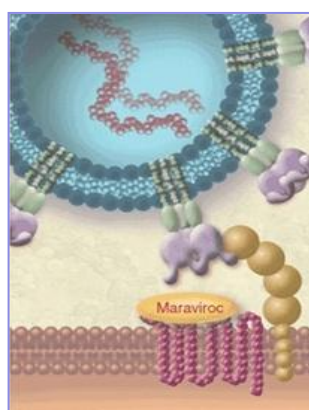


Figura 6: Maraviroque age bloqueando o correceptor CCR5, impedindo a ligação da gp120 viral e consequentemente a entrada e replicação do vírus.

Conforme mencionado, a entrada do HIV na célula envolve interações sequenciais entre o envelope viral e o receptor CD4, um correceptor de quimiocina, que pode ser o correceptor CCR5 ou correceptor CXCR4 e o domínio de fusão (Briz V et al., 2006; Este & Telenti; 2007). O uso de correceptor CCR5 ou CXCR4 pelo HIV é principalmente determinado por uma sequência de aminoácidos na alça V3 da gp120 (Chan et al., 1999; Comier & Dragic; 2002).

Podem também ocorrer misturas tanto de cepas virais R5 com X4 assim como cepas com tropismo duplo. O MVQ só deve ser utilizado por indivíduos infectados por vírus com tropismo exclusivo para R5.

As pessoas normalmente se infectam por vírus R5 e algumas podem selecionar naturalmente, durante o curso da infecção, vírus que utilizam o receptor X4. Aproximadamente 80% dos pacientes virgens de tratamento antirretroviral e 50-70% dos pacientes experimentados apresentam vírus exclusivamente R5 (Demares *et al.*, 2004; Lalezari *et al.*, 2007; Wilkin *et al.*, 2007; Nelson *et al.*, 2007; Moyle *et al.*, 2005).

Antagonistas de CCR5 foram reconhecidos como potencialmente úteis para terapêutica devido à evidência de que indivíduos homozigotos para a mutação no CCR5, com deleção do alelo delta 32 (CCR5 delta 32), são naturalmente resistentes à infecção pelo HIV (Paxton *et al.*, 1996; Liu *et al.*, 1996).

A avaliação do tropismo do HIV é essencial para a indicação do uso clínico desta classe de antirretrovirais. O tropismo pode ser observado fenotipicamente ou genotipicamente.

Alguns ensaios fenotípicos amplificam sequências do envelope do HIV-1 do plasma com objetivo de identificar a habilidade do vírus em infectar células que expressam os marcadores de superfície CD4 com o CCR5 ou CXCR4. Entretanto há fatores que limitam seu uso, como o elevado custo por teste, além dos custos com o transporte, pois há a necessidade de encaminhamento de amostra para fora do Brasil e grande complexidade de execução técnica. Um desses métodos de fenotipagem é o ensaio conhecido como *Trofile*, que foi desenvolvido pela *Monogram Bioscience* de São Francisco e é realizado apenas nos Estados Unidos da América. Já o teste genotípico, analisa proteínas do envelope do HIV-1 através da amplificação e análise do sequenciamento da alça V3 da gp120 (Mendoza *et al.*, 2008). Oferecem a vantagem de uma plataforma metodológica mais acessível, é realizado no Brasil, possui baixo custo (McGovern *et al.*, 2010).

Através da utilização de ferramentas de bioinformática foram desenvolvidos algoritmos de interpretação, para auxiliar na predição do tropismo por genotipagem, tais como geno2pheno e o escore da posição específica dos aminoácidos na matriz – PSSM. Essas ferramentas têm mostrado elevado valor preditivo positivo para resposta ao uso do MVQ

(McGovern et al., 2010).

A comparação entre as análises fenotípicas e genotípicas tem apresentado boa concordância (Poveda et al., 2010; Swenson et al., 2010).

Tendo em vista a introdução desse novo medicamento anti HIV assim como o interesse do estudo da patogênese viral, estabelecemos uma iniciativa para aprimorar metodologias que pudessem prever o fenótipo viral com base na seqüência do envelope viral.

2. Objetivos

2.1. Geral

Estudar a predição genotípica do tropismo viral em DNA proviral e RNA do HIV-1

2.2. Específicos

Os objetivos específicos incluem:

(i) aprimorar métodos de predição genotípica do tropismo do HIV-1 utilizando DNA proviral de células de sangue periférico e de RNA de virions plasmáticos;

(ii) avaliar os fatores associadas às diferenças na predição do tropismo viral a partir de células de sangue periférico em relação a aquela obtida a partir de virions plasmáticos.

3. Metodologia

3.1. População Estudada

Trata-se de um estudo transversal com análise de 81 amostras de plasma, incluindo 54 pareadas com análise adicional de células coletadas simultaneamente. Há um componente retrospectivo referente à coleta de dados de histórico da doença, por meio de revisão de Prontuário Médico, fichas clínicas preenchidas pelo médico assistente, e de bancos de dados do laboratório. Foram incluídos neste estudo pacientes com diagnóstico de Infecção por HIV, em acompanhamento clínico ambulatorial ou hospitalar em Unidades Públicas de Saúde (SUS): Instituto de Infectologia Emílio Ribas e Centro de Referência e Treinamento DST/AIDS-SP, no período de 2008-2010, que preencheram os critérios de inclusão listados abaixo, e que aceitaram e assinaram o Termo de Consentimento Livre e Esclarecido. A avaliação da indicação do indivíduo para participar do estudo foi realizada pelo profissional médico que presta assistência.

➤ **Critérios de inclusão:**

Pacientes em seguimento clínico regular na unidade de saúde, em uso de terapia antirretroviral combinada de alta atividade (HAART); que apresentem critérios de falência virológica, definida como Viremia Plasmática acima de 2.000 cópias/mL e que esteja iniciando ou considerando o uso de novas classes de drogas como parte do esquema terapêutico antirretroviral de resgate.

➤ **Critérios de Exclusão:**

Alterações clínicas ou laboratoriais severas que comprometam uso de medicação ARV ou seguimento clínico adequado.

3.2. Coletas do Material Biológico

Foram coletados de cada voluntário dois tubos de sangue total, aproximadamente 8 mL, em EDTA (ácido etilenediaminetetracético) colhidos por punção de veia periférica. O material biológico foi centrifugado a 2000 RPM (563 rcf) por 20 minutos, para que ocorresse a separação do plasma que foi armazenado em alíquotas de 500 uL. A centrifugação permite além da separação do plasma, a formação de um anel leucocitário contendo células mononucleares de sangue periférico (PBMC), que também foi armazenado em alíquotas de 500uL em tubo seco ou contendo 500 uL de soro fetal bovino e DMSO (Dimetil Sulfóxido) para posterior extração do DNA viral. O material foi armazenado a -70°C até a realização dos procedimentos laboratoriais.

3.3. Extração do RNA viral e DNA proviral

O RNA do HIV foi extraído a partir de plasma utilizando reagentes comercialmente disponíveis, seguindo as especificações do fabricante (*QIAamp Viral RNA Mini Kit, Qiagen, Hilden, Germany*). O DNA proviral foi extraído a partir de PBMC, através do *QIAamp DNA Mini Kit, Qiagen, Hilden, Germany*).

O material genético extraído (RNA viral ou DNA proviral) foi armazenado a -70°C até a realização dos procedimentos laboratoriais.

3.4. Reação em Cadeia da Polimerase (PCR)

3.4.1. Amplificação do RNA viral

A amplificação parcial do gene *env* do HIV, a partir do RNA extraído, foi realizada em uma etapa combinada de retrotranscrição e amplificação (*One-step RT-PCR*) em replicatas. Para cada amostra, foi realizado uma única extração, sendo as reações de amplificação realizadas independentemente com 2,5 uL desse extraído em três reações independentes.

A primeira etapa da reação foi realizada com um volume final de 25µL contendo, 1X de tampão 10X, 0,2mM de dNTPs, 2mM de MgSO₄, 0,625U da enzima *Platinum® Taq High Fidelity*, 2,5µL de RNA extraído, 15,775µL de água ultrapura, 0,4µM de cada *primer*, SQV3F1, *foward* 5' GAG CCA CCC ATA CAT ATT TAT TGT 3' (6858 → 6878) e CO602, *reverse* 5' GCC CAT AGT TGC TGC TGC TCC TCC CAA GAA CC 3' (7814 ← 7786). Os controles positivos e negativos foram incluídos em cada reação. As condições de termociclagem foram de 50°C 30 min, 94°C 2 min, seguido de 40 ciclos de 94°C 15 seg, 55°C 30 seg, 68°C 1 min 30 seg e 68°C 10 min. 2,5µL do produto da RT-PCR foram submetidos a uma segunda amplificação com os *primers* SQV3F2, *foward* 5' CCA TGC GCC TGT TTT GGT GCG AT 3' (6879 → 6898) e CD4R, *reverse* 5' TAT AAT TCA CTT CTC CAA TTG TCC 3' (7652 ← 7672) utilizando *Go Taq Master Mix* (Promega, EUA) com volume final de cada reação de 25µL. As condições de termociclagem foram de 10 ciclos de 94°C 2min, 94°C 15seg, 55°C 30seg, 72°C 1min, 25 ciclos de 94°C 15seg, 55°C 30 segundos, 72°C 1min (com acréscimo de 5 segundos a cada ciclo), 72°C 7min.

3.4.2. Amplificação do DNA proviral

A amplificação parcial do gene *env* do HIV a partir do DNA genômico também foi realizada em replicatas independentes. A primeira etapa da reação foi realizada com um volume final de 25µL contendo, 1X de tampão 10X, 0,2mM de dNTPs, 2mM de MgSO₄, 0,625U da enzima *Platinum® Taq High Fidelity*, 2,5µL de RNA extraído, 17,375µL de água ultrapura, 0,4µM de cada *primer*, ED5, *forward* 5' ATG GGA TCA AAG CCT AAA GCC ATG TG 3' (6557 → 6582) e ED12, *reverse* 5' AGT TGC TGC TGC TCC TCC CAA CCC GAA AAG 3' (7811 ← 7782). Os controles positivos e negativos foram incluídos em cada reação. As condições de termociclagem foram de 10 ciclos de 94°C 2 min, 94°C 15 seg, 68°C 30 seg, 68°C 1min, 25 ciclos de 94°C 15 seg, 55°C 30 seg, 68°C 1 min, adicionando 5 seg / ciclo, e 68°C 7min. 2,5µL do produto da primeira amplificação foram submetidos a uma segunda amplificação com os *primers* ED31, *forward* 5' CCT CAG CCA TTA AGG CAC CCT GTC CAA AG 3' (6817 → 6845) e ES8, *reverse* 5' CAC TTC TGT AAT TCC CCC TCA 3' (7668 → 7648). As condições da termociclagem foram de 10 ciclos de 94°C 2 min, 94°C 15 seg, 55°C 30 seg, 72°C 1 min, 25 ciclos de 94°C 15 seg, 55°C 30 seg, 72°C 1 min (adicionando 5 seg em cada ciclo).

3.4.3. Eletroforese em gel de agarose

O produto de todas as reações foram observados em eletroforese em gel de agarose 1,3%, TBE (tampão borato EDTA) 0,5%, corado com *SYBER™ Safe DNA Gel Stain (Invitrogen, California, USA)*, visualizado por luz ultra-violeta e fotografado. As bandas do material amplificado foram comparadas ao padrão de peso molecular *DNA Mass Ladder (Invitrogen, California, USA)*. Os fragmentos de proteínas esperados deveriam apresentar um tamanho em torno de 800 pares de base (pb) (Figura 7).

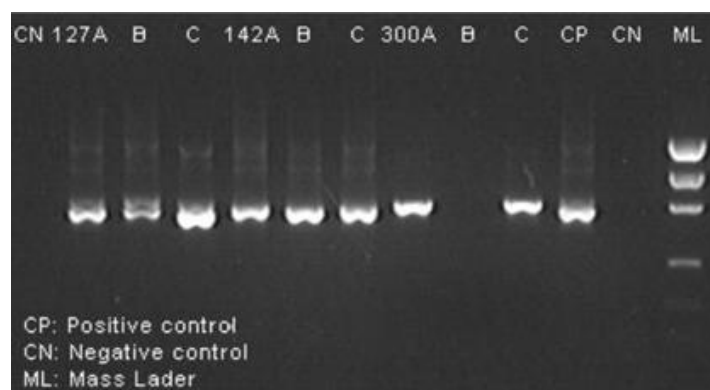


Figura 7: Gel de agarose representativo da reação de amplificação de RNA plasmático por RT-PCR em triplicatas independentes, juntamente com um quantificador (*mass ladder* - ML), controles positivo e negativo (CP, CN) e as amostras (127ABC, 142ABC, 300ABC), sendo que a amostra 300 não apresentou amplificação na replicata B.

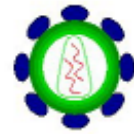
3.4.4. Sequenciamento do produto amplificado

Os produtos de PCR foram sequenciados utilizando os *primers* V3O2, *foward* 5' AAT GTC AGY ACA GTA CAA TGT ACA C 3' (6948 → 6969), ES7, *foward* 5' CTG TTA AAT GGC AGT CTA GC 3' (7005 → 7021), ED33 *reverse* 5' TTA CAG TAG AAA AAT TCC CCT C 3' (7360 ← 7378) e SQV3, *reverse* 5' GAA AAA TTC CCT TCC ACA ATT AAA 3' (7350 ← 7370). As reações de sequenciamento nucleotídico foram realizadas utilizando-se *Big Dye Kit* (*Applied Biosystems, Foster City, Califórnia*). As condições de ciclagem para o sequenciamento consiste em 25 ciclos de 96°C 10 seg, 50°C 5 seg, 60°C 4 min, 8°C infinito. Após a reação de sequenciamento, esse material foi precipitado utilizando uma solução de acetato de sódio e etanol seguido por uma etapa de desnaturação usando formamida Hi-Di a 94°C por 4 min e posteriormente submetido ao sequenciador automático ABI Prism 3130 XL *Genetic Analyzer* (*Applied Biosystems*).

3.4.5. Predição genotípica do Tropismo Viral

Os dados dos eletroferogramas foram analisados utilizando o *software Sequencher 4.6*. As sequências foram editadas manualmente e um consenso foi gerado com base em duas ou mais sequências de nucleotídeos, incluindo nucleotídeos ambíguos (misturas). Os arquivos gerados pela edição manual foram alinhados através do programa *BioEdit version 7.0.0*, com sequências consensos dos Subtipos B, C e a sequência referência HXB2 (subtipo B). O controle de qualidade das sequências e a subtipagem foram realizadas utilizando ferramentas baseadas na *Web* como NCBI *locator*, BLAST, com reavaliação do eletroferograma sempre que necessário. Após a correção do alinhamento no programa foi possível determinar o início da região codificante da alça V3 do gene envelope. O uso do correceptor pelo HIV-1 foi predita a partir da sequência de nucleotídeos da alça V3 através da ferramenta de bioinformática *Geno2pheno coreceptor (g2p)* URL: <http://coreceptor.bioinf.mpi-sb.mpg.de/cgi-bin/coreceptor.pl>. A determinação do tropismo foi obtida através da submissão das sequências (formato fasta) para *geno2pheno* usando a opção clonal. A ferramenta *g2p*, após a submissão da sequência, emite um resultado contendo um valor em porcentagem denominado de taxa de falsos positivos (FPR – *false positive rate*), que significa a probabilidade de classificar um vírus R5 falsamente como X4, sendo essa taxa atribuída a cada sequência submetida (Figura 8). Esta ferramenta pode gerar dois laudos: clonal (baseado na sequência genética) e clínico (utiliza além da sequência, valores de CD4 e carga viral disponíveis do paciente). Análises de tropismo com base no *geno2pheno*, opção clonal, é hoje recomendada pelo consenso europeu. Usamos, nas descrições e análises de tropismo o *geno2pheno*, opção clonal. Segundo esse consenso, o uso do ponto de corte de 20% representa um corte conservador aplicável à predição clínica baseada em uma única sequência genética, sendo usado o ponto de corte de 10% para triplicatas, assim, valores de FPRs acima do ponto de corte são classificadas como R5 e abaixo do ponto de corte, como X4. Para a emissão dos laudos clínicos, em alguns casos foram também utilizadas outras regras de análise (Cabral et

al., 2012) como o WebPSSM – *Position Specific Scoring Matrix* (uma ferramenta bioinformática que utiliza uma matriz para analisar os aminoácidos da região V3) e a regra 11/25, em que a presença de aminoácidos básicos (Arginina - R, Histidina – H, ou Lisina – K) nas posições 11 e/ou 25 na composição de aminoácidos da alça V3 são indicativos de vírus X4. Nessa análise apenas os resultados gerados pelo g2p estão sendo apresentados.



I. General information

Patient:	Study Id:
Birth date:	Viral load:
Sample received:	Sample collected:
Sample IDs: 218_11C plasma	Predicted V3-subtype: B
Sample type:	Report date:
Physician:	Reported by:

II. Additional clinical parameters used for the predictions:

Viral load:	100-1000
CCR5-genotype:	not determined
CD4-percentages:	not determined
CD4-cell counts:	not determined
CD8-cell counts:	not determined

III. Aligned V3 region

Consensus B: TTF ACA ABA CDE AAC AAC AAT ACA ABA AAA AST ATA CDT AEA BBA CCA BBA CAA BCA TTE TAT BGA ACA BBA BAC AFA ATA BBA BBT AFA ABA CAA BBA GGT TTF
 Query: C T R P N N N T R K S I H I G P G R A F Y T T G E I I G D I R Q A H C
 C T R P S T N I R R I H I G P G R A F Y T T G T I T G D V R K A Y C

IV. Coreceptor prediction

Model	Prediction	FPR	Remarks
Clonal	The CXCR4-coreceptor can be used.	0.2%	The 11/25 rule would predict this sequence as an X4-virus.
Clinical	The CXCR4-coreceptor can be used. CCR5-antagonists like Maraviroc (Selzenri/Selzenry) should not be administered	0.2%	

The significance level was set to: 10% false positive rate

Figura 8: Exemplo de resultado emitido pela ferramenta de bioinformática *geno2pheno coreceptor*. Na imagem é possível observar o valor de FPR_{clonal} e $FPR_{\text{clínico}}$ atribuído à sequência submetida ao *site*.

3.4.6. Distância Genética

A estimativa da diversidade de cada sequência da alça V3 foi avaliada como: (i) o número de nucleotídeos ambíguos, (ii) o número total de possíveis aminoácidos decorrentes dessas ambiguidades, denominadas de misturas não sinônimas (MNS) e (iii) o número de *codons* na V3 com uma ou mais MNS. A diversidade entre as sequências de uma mesma amostra (sequências em replicatas), foi quantificada por (iv) PROTDIST, que calcula uma matriz de distância a partir de sequências de aminoácidos, considerando todas as ambiguidades na sequência, (v) GENDIST que usa sequências de nucleotídeos para calcular uma matriz de distância, ambos usando o software *BioEdit* (Figura 9) e (vi) distância filogenética. A construção da árvore filogenética foi realizada utilizando o software PAUP v4.0b 10 (Swofford, 2005) aplicando o modelo de evolução selecionado pelo Modeltest (TVM+I+G).

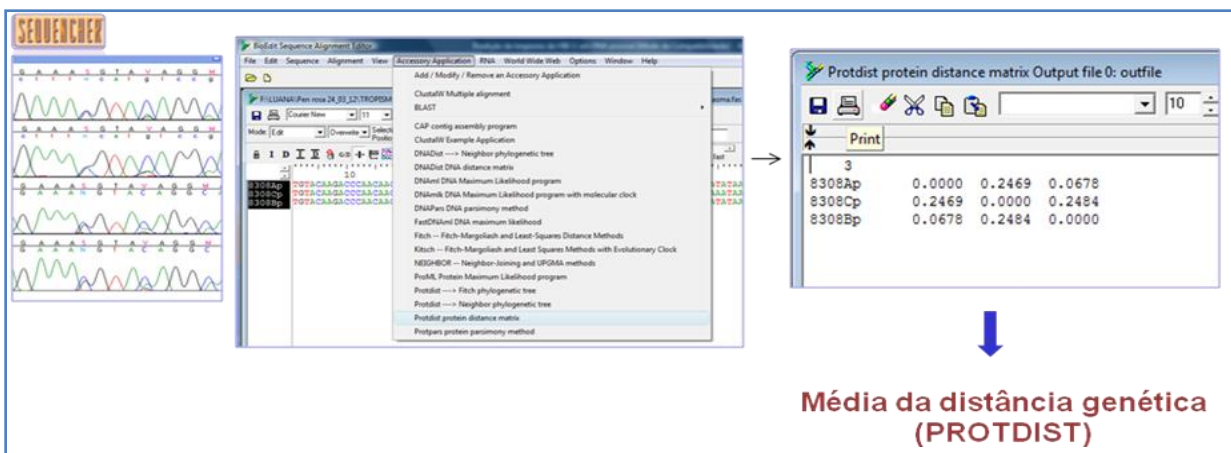


Figura 9: O esquema representa a maneira como foi calculado a distância genética em proteínas através da ferramenta *BioEdit*, onde é gerado uma matrix de valores de distâncias a partir do qual é calculado a média.

3.4.7. Análise Estatística

Os dados foram armazenados em banco de dados EpiInfo6 (CDC) e Excel. Testes t (Mann Whitney) e correlações lineares (Spearman) com o nível de significância de $p < 0,05$ foram feitos através dos *softwares* PrismaGraphPad. Os resultados foram expressos em medianas, descrevendo os percentis 25th e 75th (IQR).

3.4.8. Aprimoramento do Teste genotípico para predição do tropismo do HIV-1 do Instituto Adolfo Lutz (IAL) por avaliação externa do Centro de Excelência em HIV/AIDS da Columbia Britânica (CECB)

Através dos ensaios realizados em nosso laboratório, foi possível adaptar a metodologia para determinação do tropismo do HIV desenvolvido pelo Centro de Excelência em HIV/AIDS de *British Columbia*.

No laboratório de genotipagem do HIV do Instituto Adolfo Lutz (IAL) realizamos a metodologia para determinação do tropismo em amostras enviadas pelo Centro de Excelência em HIV/AIDS da Columbia Britânica (CECB) com adaptações desenvolvidas pelo CECB.

Foram enviadas 11 amostras de plasma e 11 produtos de PCR, *primers* para 1ºPCR (SQV3F 25uM, CO602R 25uM), 2º PCR (SQV3F2 25uM, CD4R 25uM) e sequenciamento (V3O2F 25uM, SQV3R 25uM). Diluímos os *primers* para PCR de 25uM para 10uM e os de sequenciamento de 25uM para 1uM.

Adaptamos o protocolo enviado às condições do nosso laboratório, reduzimos o volume final da reação de 50uL para 25uL mantendo a concentração final dos reagentes e incorporamos no primeiro PCR a utilização de enzima com ação corretiva.

1º- Com o produto de 1ºPCR enviado seguimos os procedimentos de “*Nested PCR*”, quantificação por eletroforese de proteínas em gel de agarose 1.3%, reação de sequenciamento, como recomendado no protocolo.

2° - Com o plasma seguiram o fluxo de extração de RNA utilizando “Kit” comercial seguindo as orientações do fabricante. Utilizando os *primers* enviados e previamente diluídos procedemos às reações de PCR, quantificação e sequenciamento.

4. Resultados

4.1. Aprimoramento do Teste genotípico para predição do tropismo do HIV-1 do Instituto Adolfo Lutz (IAL) por avaliação externa do Centro de Excelência em HIV/AIDS da Columbia Britânica (CECB)

Plasma: Seis amostras foram classificadas em nosso laboratório como R5 e quatro como X4, uma amostra foi o controle negativo da reação de PCR. O valor de falso positivo (FPR) nas amostras R5 variou de 5,9 a 96,4 e nas X4 de 0,7 a 2,5.

Produto de PCR: Nossas análises demonstraram a presença de variantes R5 em oito amostras, enquanto que a presença de variantes X4 foi encontrada em duas amostras e uma amostra não amplificou, sendo identificada posteriormente como o controle negativo das amostras.

Enviamos para Columbia Britânica os nossos resultados (Tabela 3), assim como um relatório descrevendo todo nosso procedimento.

Tabela 3: Resultados dos valores de FPR-CEBC e FPR- IAL

Amostra	FPR - CECB	Tropismo CECB	FPR - IAL	Tropismo IAL
PF0214	0,5	X4	0,7	X4
PF0248	1,7	X4	1,7	X4
PF0672	2,7	X4	3,8	X4
PF1483	55,7	R5	51,8	R5
PF1848	91,7	R5	84,9	R5
PF1974	86,7	R5	86,7	R5
PF2525	1,7	X4	1,7	X4
PF3248	96,4	R5	96,4	R5
PF3271	6,9	R5	1,9	X4
PF3384	71,8	R5	69,2	R5

A tabela demonstra as amostras com a identificação recebida pelo CECB e os respectivos valores de FPR. Os valores de FPR representados na tabela é o menor valor de FPR_{clonal} da amostra, realizado em triplicata.

4.2. Avaliações da predição genotípica do tropismo do HIV-1 utilizando sequenciamento em replicata

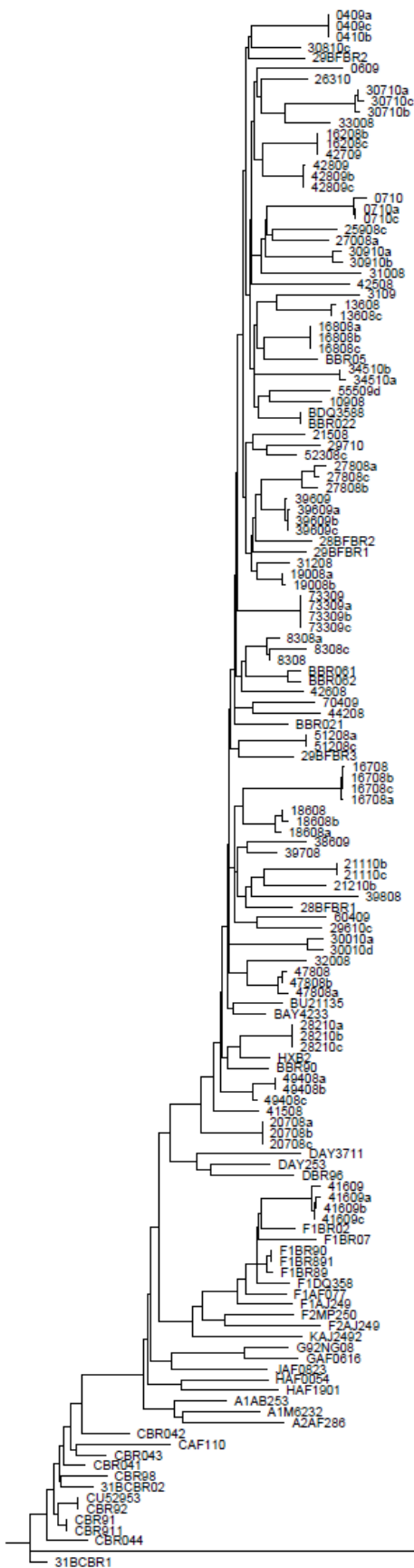
Para avaliar o uso de replicatas em nosso teste avaliamos em um grupo de 81 pacientes. Alguns dados clínicos, como os valores de CD4 e carga viral, podem ser vistos na tabela 4. Foram obtidas de 50 pacientes uma única sequência da região da alça V3 do gene envelope e de 31 pacientes, sequências em replicatas, de 2 a 4 sequências. Usando 20% como ponto de corte para a taxa de falso positivo (FPR), 41/81 (50,6%) foram preditos como X4 e 40/81 (49,4%) como R5. A mediana de linfócitos CD4 foi menor, 226cél/mm³ (IQR 82-378), em pacientes infectados com variantes X4; 324cél/mm³ (IQR 200-538, $p < 0,05$) em pacientes infectados com R5.

A diversidade entre as replicatas na V3 pode ser ilustrada na Figura 10 que demonstra uma árvore filogenética representativa onde as amostras de um mesmo paciente estão agrupadas dentro de *clusters* suportados por um *bootstrap* de 86% a 100%.

Dentro de cada sequência, (diversidade intra-sequência) a mediana de nucleotídeos ambíguos foi de 0,95, similar tanto para sequências únicas (IQR 0-2,85) como para replicatas (IQR 0-4,758). A mediana de misturas (ambiguidades) que leva a codificação de aminoácidos diferentes, ou seja, mistura não sinônima (MNS) ou códons com MNS foi 0 (IQR 0-2), um pouco mais elevada, mas não significativamente, nas amostras com sequências em replicatas ($p = 0,2$ e $p = 0,3$, respectivamente). Como seria de se esperar, o número de ambiguidades totais (mistura sinônima e mistura não sinônima) está fortemente correlacionado com o número de MNS ($p < 0,0001$). Não houve correlação entre diversidade intra-sequência (ambiguidades) e contagem de linfócito CD4 e nem com a carga viral. A distância genética em proteínas mostrou uma tendência de diminuição com o aumento da carga viral e de elevar com o aumento dos valores de linfócito CD4 (Figura 11).

Tabela 4: Dados clínicos e demográficos. Características dos pacientes de acordo com o número de sequências obtidas no estudo.

	Sequência única	Sequência replicata (2-4)
Idade (anos)	45 (39 - 49)	41 (34-47)
Gênero (%homens)	74%	74%
TCD4 na coleta (células/mm³)	241 (80 - 420)	305 (186 - 511)
Nadir TCD4 (células/mm³)	75 (24 - 128)	106 (38 - 161)
Viremia na coleta (Log₁₀/mL)	3,99 (3,36 - 4,73)	4,08 (3,54 - 4,47)
Número de tratamento	7 (4 - 9)	8 (5 - 9)
Tempo de tratamento (anos)	11,8 (10,2 - 13,5)	12,3 (10,6 - 13,9)
N	50	31



0.2

Figura 10: Árvore filogenética representativa demonstrando a diversidade entre as replicatas (inter - sequência) na V3 onde as amostras de um mesmo paciente estão suportadas por um *bootstrap* de 86% a 100%.

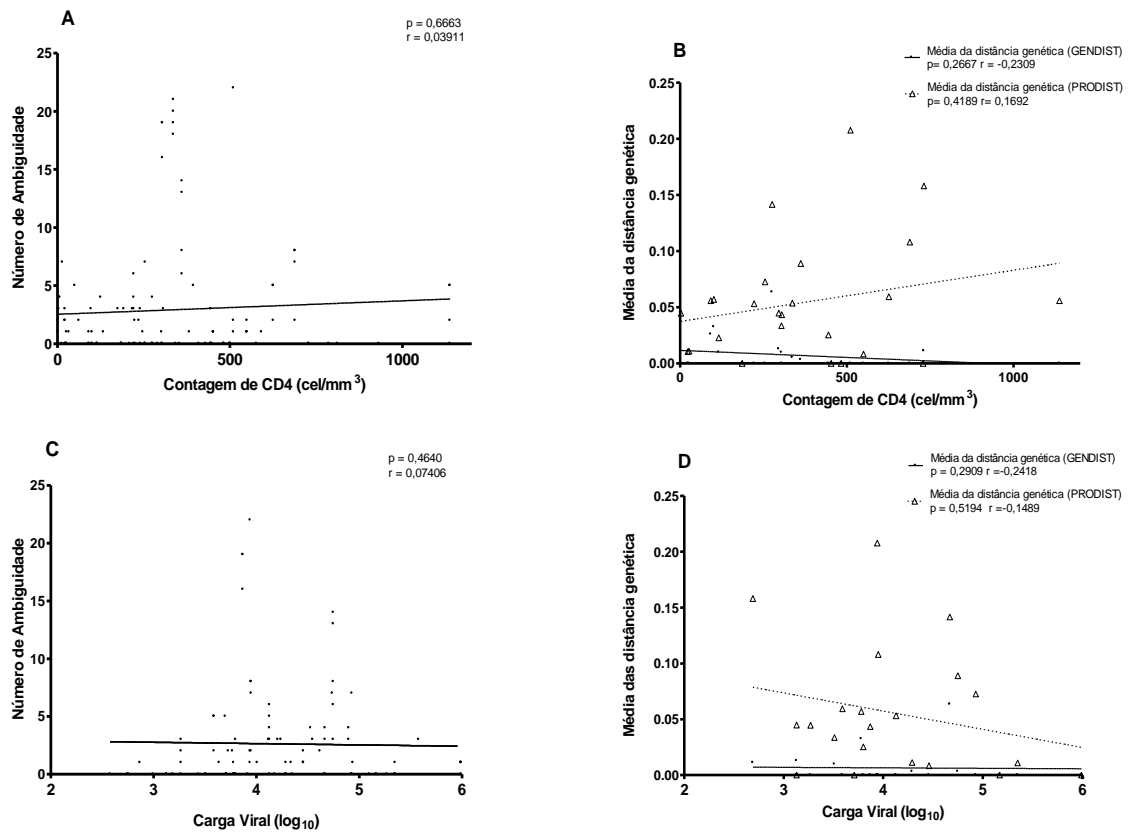


Figura 11: Contagem de CD4 (cel/mm³) (A,B) e carga viral (log₁₀/mL) (C,D) em relação ao número de ambiguidades (A,C) e a média das distâncias (GENDIST e PRODIST) (B,D). Correlação linear (Spearman), valores de r e p estão demonstrados na parte superior à direita.

4.2.1. Valores de FPR

O valor absoluto da taxa de falsos positivos (FPR), bem como a predição de X4 ou R5, considerando 20% o ponto de corte, foram inversamente correlacionados com as ambiguidades e MNS ($p < 0,0001$). Quando as sequências com elevado número de ambiguidades foram removidas da análise (sequências com 6 ou mais bases ambíguas), esta correlação foi ainda observada ($p = 0,0055$, Figura 12). A distância genética de proteína nas replicatas foi maior nas sequências com predição de X4 ($p = 0,055$, Figura 13). No entanto, o número de replicatas realizadas não está correlacionado com os valores de FPR (Figura 14).

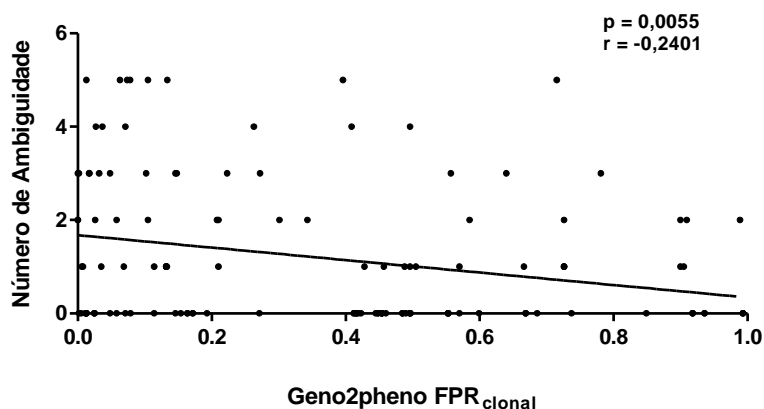


Figura 12: Correlação linear (Spearman) entre a taxa de falso positivo (FPR), opção clonal, obtida do website geno2pheno e número de ambiguidade na sequência da V3.

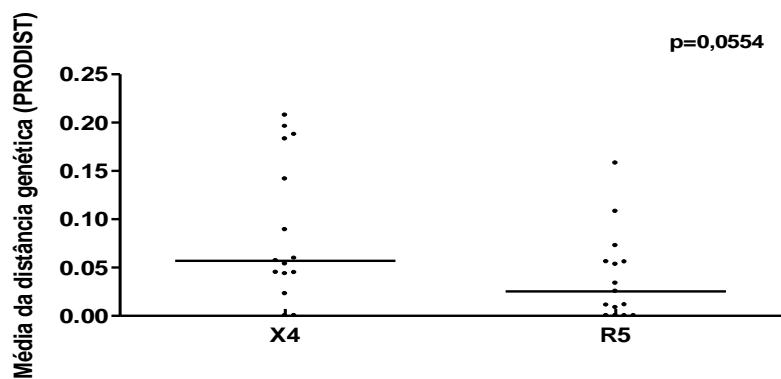


Figura 13: Média da distância genética (PRODIST) nas amostras classificadas como X4 e R5 usando FPR 20%.

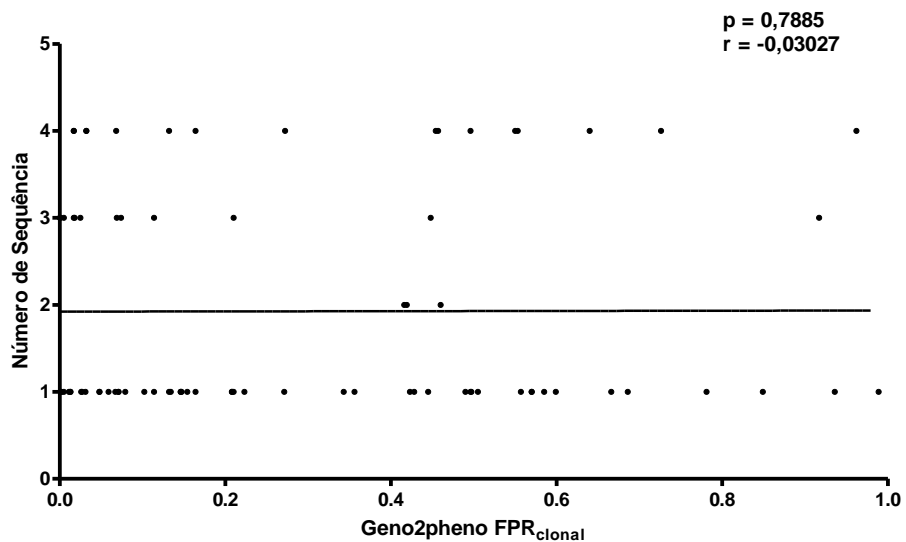


Figura 14: Menor valor de FPR_{clonal} da sequência da alça V3 da gp 120 de acordo com o número de sequência obtida de cada amostra.

4.3. Estudo comparativo entre plasma e célula na predição do tropismo viral utilizando sequências em replicata.

Para a comparação da predição do tropismo em dois diferentes compartimentos (plasma e célula), avaliamos o RNA extraído do plasma e do DNA genômico extraído do PBMC em paralelo, de uma mesma amostra, de 56 pacientes. Nessa análise também avaliamos tanto sequências individuais, em 28 pacientes, quanto amostras onde foram obtidas sequências em replicatas (2-5 replicatas), 26 pacientes. Nessa análise 56% dos pacientes eram do sexo masculino, 18% ARV naive, mediana de carga viral $4,05 \log_{10}$ e CD4 241cell/mm³, esses dados estão demonstrados na Tabela 5. De uma maneira geral existe boa correlação entre a predição do tropismo a partir do material genético oriundo do plasma e célula, e utilizando o menor valor de FPR_{clonal} obtidos nas sequências em replicata, o FPR_{clonal} do virion está fortemente correlacionado com menor de FPR proviral ($p < 0.0001$, $r = 0,9979$), Figura 15. Contudo, existem discordâncias quando se utiliza os diferentes pontos de cortes recomendados para uso clínico (geno2pheno 20%, 10% e 5,75%); entre as amostras pareadas com sequências únicas essa discordância foi de 10,7%, 10,7% e 3,2% e entre as replicatas foi 15,4%, 23%, 19,2%, respectivamente. A Tabela 6 resume esses achados. A predição usando replicatas levaria a uma mudança na predição do tropismo em 42% nas sequências de célula e 23% no plasma quando cada valor de FPR_{clonal} de cada sequência da replicata é considerado individualmente. O uso de outra ferramenta para a predição do tropismo, como o PSSM_{X4R5} (<http://fortinbras.us/fssm/index.html>) também mostrou um discordância na predição entre plasma e célula (14% em sequências únicas e 23% nas replicatas).

Quando todos os valores de FPRs das sequências em replicata foram avaliados, o desvio padrão dos FPRs derivados de célula foi maior do que os FPRs das sequências obtidas de plasma ($p=0,003$) (Figura 16). O número de ambiguidades presente em cada sequência está correlacionado com o valor de FPR_{clonal} , tanto no plasma como nas células ($p=0,0008$ e $p < 0,001$).

respectivamente), Figura 17. Nas amostras com sequências em replicatas, o menor valor de FPR_{clonal} em ambos, plasma e célula, apresentou uma correlação negativa, mas não significativa, com a média da distância genética em proteínas (PROTDIST) ($p=0,1776$ plasma e $p=0,0506$ célula), Figura 18. Entretanto, a mediana da distância genética foi significativamente maior na célula (0,144) quando comparado com as sequências de plasma (0,040) ($p<0,0001$), Figura 19. A média entre a distância genética nas replicatas, em ambos os compartimentos, não mostrou correlação com os valores de CD4 e carga viral Figura 20.

A distância entre as amostras e a relação filogenética dos isolados foi também avaliada com métodos de filogenia, utilizando o software PAUP v4.0b 10 , como pode ser observado na Figura 21. Amostras de um mesmo paciente formam *clusters* com *bootstrap* elevado; alguns desses clusters estão exemplificados nas Figura 21a e 21b. Contudo, não existe discriminação, usando o modelo (TVM+I+G), entre as amostras provenientes de célula ou plasma em alguns casos, assim como entre aquelas sequências que apresentam tropismo X4 ou R5. As Figuras 21A e 21B são destaques de diferentes agrupamentos de sequências de DNA e RNA.

Tabela 5: Dados clínicos e demográficos. Características dos pacientes de acordo com o número de sequências obtidas no estudo

	Sequência única	Sequência replicata (2-4)
Idade (anos)	45 (39 - 49)	41 (19 – 47)
Gênero (% masculino)	74%	72%
CD4 na coleta (células/mm ³)	241 (80 - 420)	305 (116 - 511)
Nadir CD4 (células/mm ³)	75 (24 – 128)	105 (27- 135)
Viremia na coleta(Log ₁₀ /mL)	3,99 (3,36 - 4,73)	4,17(3,77 - 4,53)
Número de tratamentos	7 (4 - 9)	8 (6 – 9)
N	28	26

Tabela 6: Comparação da predição de tropismo em células e plasma

	Sequência única Discordância	Proporção de Discordância	Replicatas Discordância	Proporção de Discordância
Geno2pheno₂₀	3/28		4/26	
	Plasma R5 n=3	10,7%	Plasma R5 n=2 Plasma X4 n=2	15,4%
Geno2pheno₁₀	3/28		6/26	
	Plasma R5 n=2 Plasma X4 n=1	10,7%	Plasma R5 n=3 Plasma X4 n=3	23%
Motivate (2% - 5,75%)	2/28		5/26	
	Plasma R5 n=1 Plasma X4 n=1	3,2%	Plasma R5 n=4 Plasma X4 n=1	19,2%

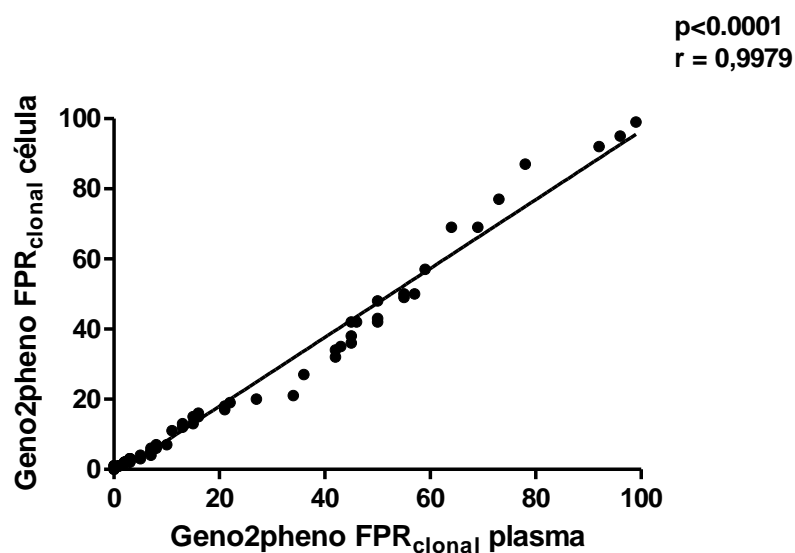


Figura 15: Correlação entre os valores de menor FPR_{clonal} das sequências obtidas a partir do DNA proviral nas células em relação aos valores de menor FPR_{clonal} das sequências obtidas a partir do RNA do vírus circulante no plasma.

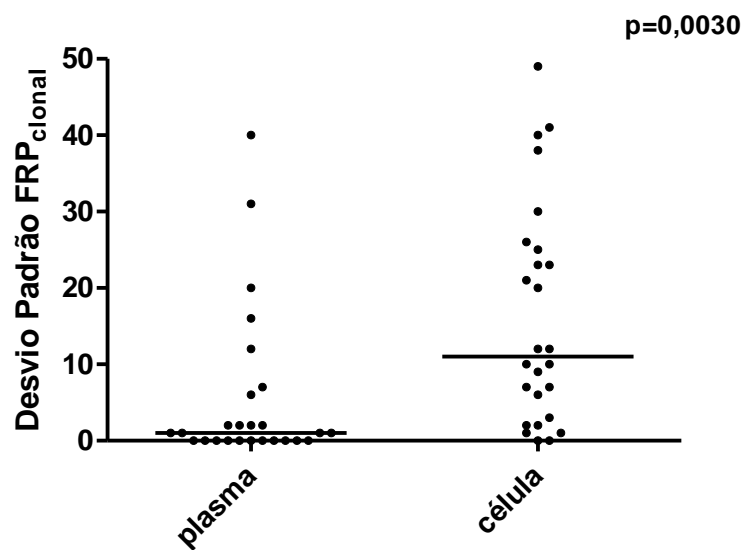


Figura 16: Teste *Wilcoxon signed rank* do desvio padrão dos valores de FPR_{clonal} das seqüências obtidas de RNA viral e DNA proviral pareadas.

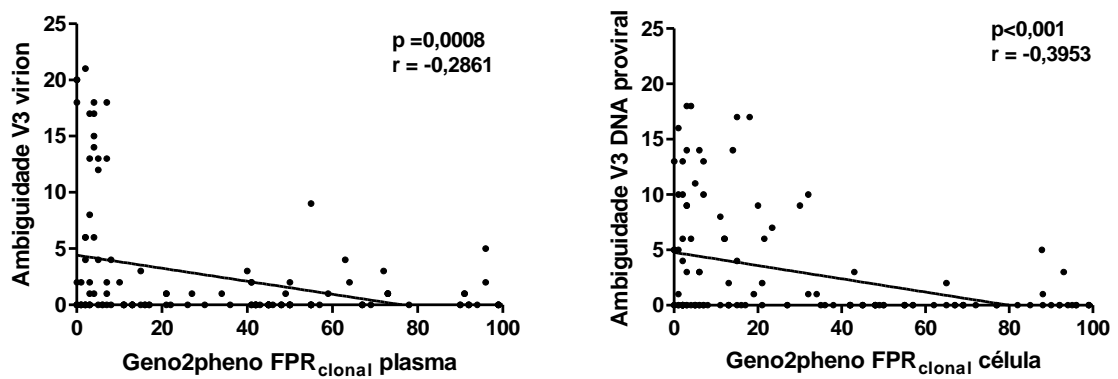


Figura 17: Correlação (Spearman) entre os valores de FPR_{clonal} no plasma e célula e o número de nucleotídeos ambíguos na sequencia da alça V3.

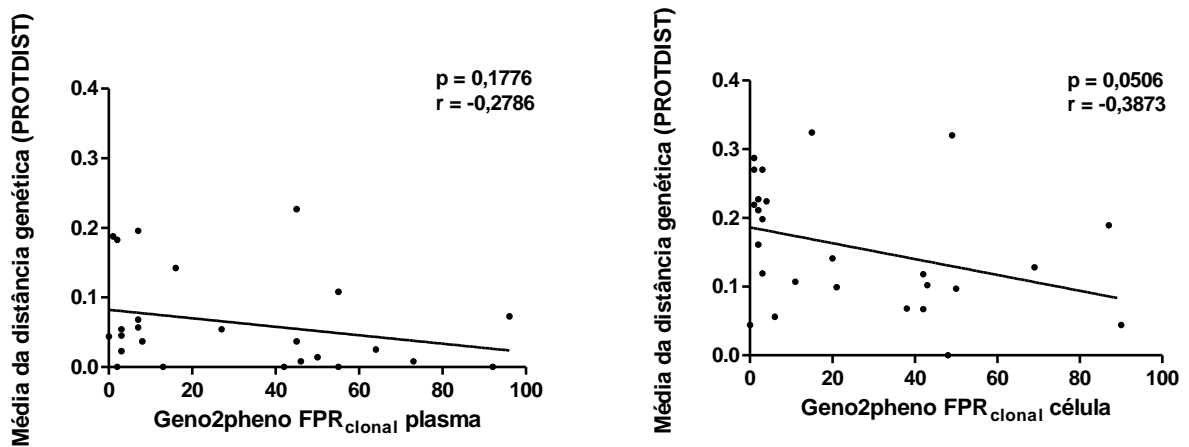


Figura 18: Correlação (Spearman) entre os valores de menor FPR_{clonal} no plasma e célula e a média dos valores da distância de proteínas (PROTDIST).

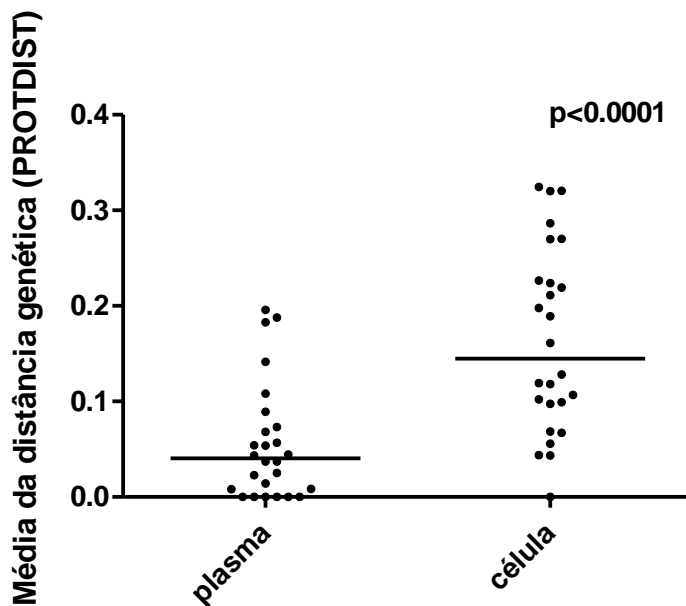


Figura 19: Distância genética (PROTDIST) entre as sequências em replicata obtidas a partir de plasma e célula. O traço representa a mediana desses valores.

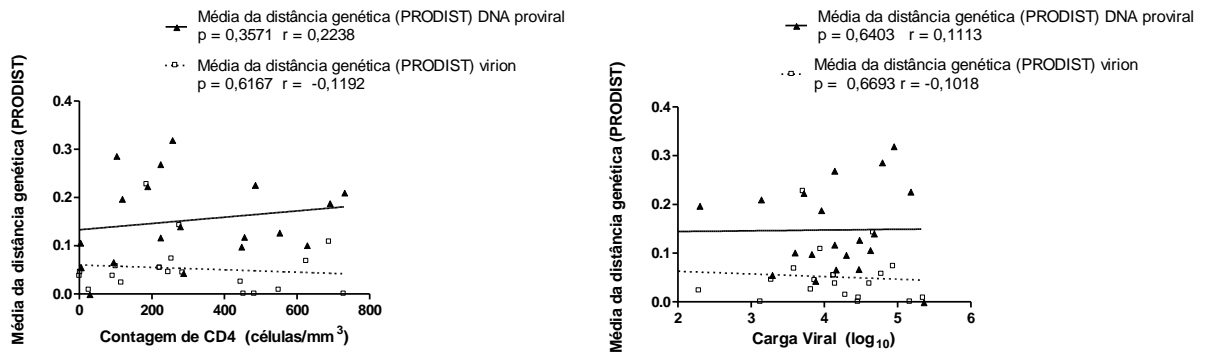


Figura 20: Correlação linear (Spearman) entre A - valores CD4 (células/mm³) e distância genética (PROTDIST), e B - Carga Viral (cópias / mL) e genética (PRODIST).

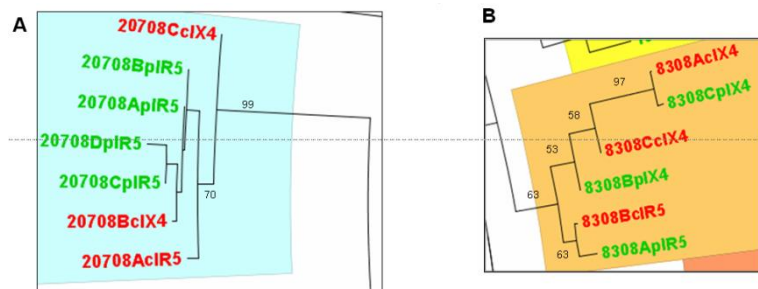
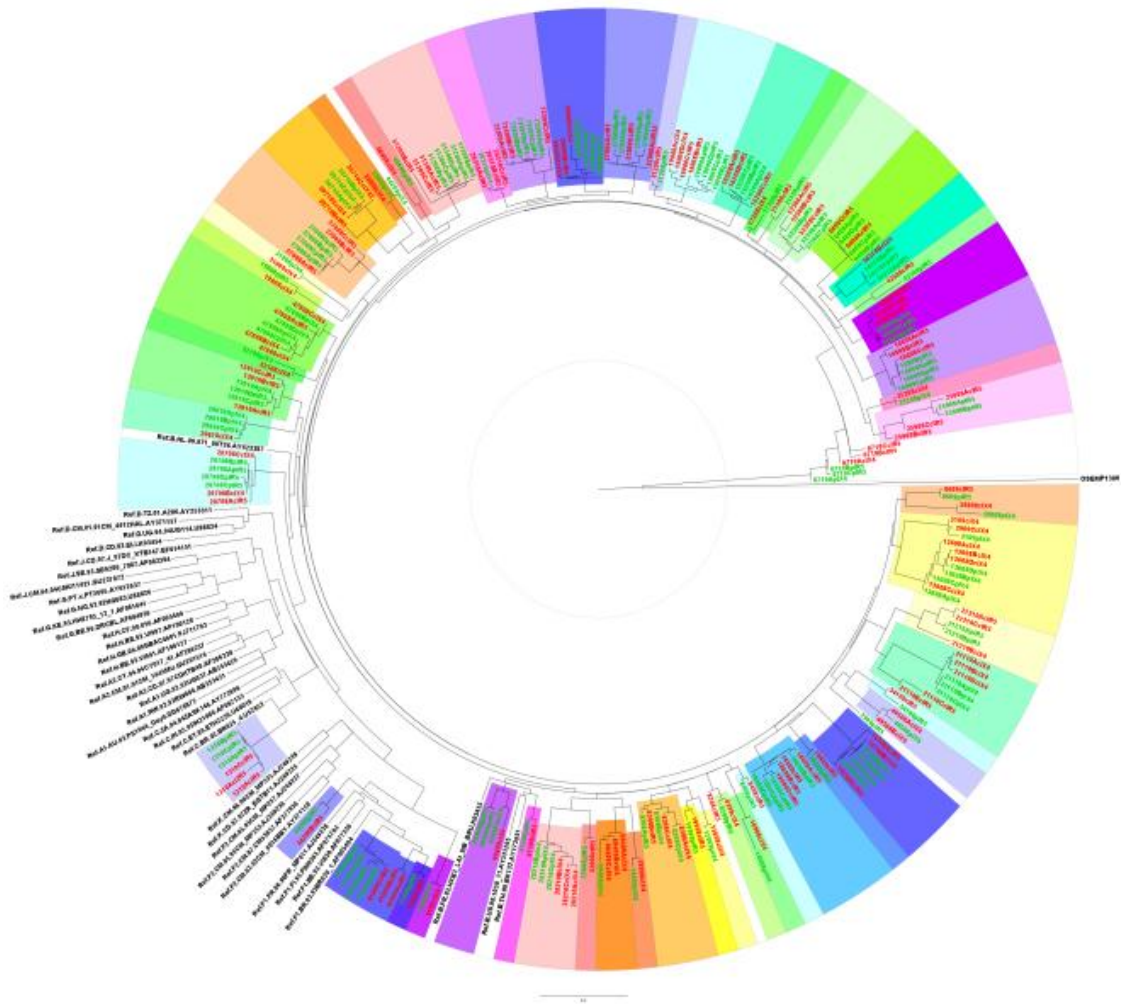


Figura 21 - Árvore filogenética construída usando PAUP *. Sequências obtidas a partir de DNA proviral (vermelho) e RNA viral (verde). Cada grupo de cor representa um *cluster* de sequência de um paciente, porém as cores não são únicas e, portanto, não representam qualquer correlação entre os ramos distantes ou amostras de pacientes diferentes. Sequência OSEMP1300 foi utilizada como outgroup. A e B são destaques de diferentes agrupamentos de sequências de DNA e RNA.

5. Discussão

Tropismo viral é a capacidade do vírus de penetrar e infectar células do hospedeiro por meio de ligação específica à receptores. No caso do HIV, essa habilidade de ligação à célula CD4 ocorre geralmente pelo correceptor CCR5 (vírus R5), correceptor CXCR4 (vírus X4) ou por ambos correceptores (tropismo duplo). Além de desempenhar importante papel na patogênese viral, a avaliação do tropismo tornou-se essencial para a indicação do uso clínico da nova classe de ARV, os antagonistas de correceptor de quimiocina CCR5, que inibem especificamente a entrada do vírus na célula e, bloqueiam a replicação da variante R5 do HIV após a ligação ao correceptor transmembrana CCR5. O inibidor de CCR5 disponível comercialmente é o maraviroque (Berger et al., 1998; Berger et al., 1999). O tropismo pode ser observado fenotipicamente ou predito genotipicamente (McGovern et al., 2010).

Ensaio fenotípicos comerciais analisam o envelope viral clonado em um arcabouço conhecido, permitindo automação, maior reprodutibilidade e comparabilidade, entretanto há fatores que limitam o seu uso, como o custo elevado, custos com o transporte da amostra e complexidade técnica.

O teste de tropismo genotípico tem sido proposto como alternativa para o teste fenotípico. Baseia-se na análise da sequência de aminoácidos na alça V3 da gp120 utilizando ferramentas de bioinformática e um algoritmo de predição do tropismo. O teste genotípico apresenta a vantagem de ser menos exigente tecnicamente, mais rápido e ter menor custo (Chan et al., 1999; Comier & Dragic, 2002; McGovern et al., 2010).

Um dos principais obstáculos da uma metodologia genotípica de predição é a baixa sensibilidade dos testes disponíveis na identificação de variantes X4. Portanto, um primeiro ponto do nosso trabalho foi o aprimoramento da metodologia já existente no nosso laboratório.

Com o objetivo de melhorar a metodologia para identificação de variantes X4, nós tivemos a possibilidade de intercâmbio com o Centro de Excelência em HIV/AIDS da Colúmbia Britânica (CECB). Onde foram

enviadas amostras para serem testadas em nosso laboratório. Recebemos um relatório da análise de nossas sequências, realizadas em nosso laboratório e enviadas ao CECB, sendo essas relacionadas a sequências realizadas anteriormente por eles, com concordância na maioria dos casos exceto um dos resultados. Esses dados estão demonstrados na Tabela 3. Nessa amostra, em nosso caso o FPR estava abaixo de 2%, levando a interpretação de tropismo X4, enquanto que no CECB o menor FPR foi 6.9%, levando a classificação de R5 pelo critério “Motive”, que considera o ponto de corte de 5,75%. É importante notar que o sequenciamento em nosso laboratório e o no CECB apresentam FPR abaixo de 10%, e seguindo as regras europeias a amostra seria classificada como X4 por ambos critérios, levando a 100% de concordância nos resultados.

Por não existir um critério definido, nós seguimos neste trabalho as regras europeias que recomendam, para o teste genotípico, o ponto de corte de 10% para amostras com mais de uma sequência e 20% para amostras com uma única sequência da alça V3 da gp120 (Vandekerckhove et al. 2009). Neste estudo nós procuramos avaliar fatores que poderiam influenciar na predição do tropismo viral, utilizando retro-transcrição, amplificação e sequenciamento em replicata.

Outra questão importante é a origem do material genético, sendo assim, nós avaliamos também o teste de tropismo genotípico no RNA do plasma e no DNA províral, identificamos eventuais diferenças na predição e se o uso de replicatas diminui a discordância na predição utilizando RNA ou DNA províral.

5.1 Avaliações da predição genotípica do tropismo do HIV-1 utilizando sequenciamento simples e em replicata

Este estudo avaliou um protocolo para a amplificação e sequenciamento em replicata da população (*bulk*) viral a partir de amostras de plasma. Nós descrevemos um método onde em uma única etapa é realizado a retro-transcrição e o primeiro PCR (RT-PCR) usando procedimentos que podem ser facilmente implementados em laboratórios com recursos limitados. Na predição genotípica do tropismo, a análise e interpretação da sequência é uma etapa crucial. A predição do tropismo pode ser feita usando ferramentas de diferentes *websites* e pela análise direta da sequência, como utilizando a “regra 11/25” , baseada na identificação da presença de aminoácidos básicos (arginina, lisina e/ou histidina) nas posições 11 e/ou 25 da sequência de aminoácidos da alça V3 da gp120. Aqui, nós não avaliamos detalhadamente todos os métodos de interpretação da sequência, e sim, para facilitar a comparabilidade realizamos as análises usando os recursos *Max Plank*, *Geno2pheno* (opção clonal). O *geno2pheno* é uma ferramenta de fácil utilização e sua aplicabilidade foi recentemente recomendado pelo consenso Europeu que estabeleceu dois pontos de corte dependendo do número de sequências obtidas para cada amostra (única sequência ou sequências em replicata) quando aplicado em teste para uso clínico, e considerando o menor valor de FPR, para a predição do tropismo viral nas sequências em replicata, já que estas implicam em diferentes valores de FPRs (Vandekerckhove et al.,2011). Um menor ponto de corte para definir a presença de variantes X4 nas amostras com sequências com replicatas foi proposto pelo grupo de Vancouver (McGovern et al., 2010), após uma avaliação retrospectiva dos resultados do estudo MOTIVATE. Este estudo classificou como X4, sequências com valores de FPR abaixo de 2%, possível X4, sequências com valores de FPR entre 2% e 5,75% e R5 sequências com valores de FPR acima de 5,75%. O Consenso Europeu sugere o ponto de corte de 10% nos casos de amostras com sequências em replicatas e um corte de 20% para

sequências individuais (únicas).

Neste estudo, onde avaliamos o impacto da replicata na predição do tropismo, foi utilizado o ponto de corte mais conservador de 20%. As sequências foram avaliadas individualmente e em replicatas. Nosso objetivo foi compreender melhor o papel das amplificações independentes e das replicatas no FPR. Assim, nós comparamos a predição inferida nas sequências únicas e nas replicatas (2-4) e avaliamos parâmetros que pudessem influenciar na predição do tropismo viral. Nós também avaliamos o valor de FPR como uma variável contínua para estimar a sua correlação com as métricas de diversidade viral.

O uso de replicatas na predição do tropismo viral aponta uma questão importante, ou seja, se o número de sequência por si só, aumentaria a probabilidade de encontrar variantes X4. Isto poderia sugerir que quanto maior o número de replicatas, maior a chance de detecção de vírus X4. De fato este é um problema no *deep sequencing* (Swenson et al., 2010). Como mostrado na figura 14, o número de replicatas não está correlacionado com o menor valor de FPR ou com a probabilidade de detecção de variantes X4. No entanto, o estudo encontrou forte correlação entre alguns parâmetros de diversidade e a predição de populações X4. Neste trabalho, a potencial diversidade intra-sequência (dentro da sequência de 35 aminoácidos da alça V3) gerada, assim como a diversidade inter-sequencia, como a distância genética de nucleotídeos e de proteínas entre as replicatas, demonstra uma tendência de correlação inversa com os menores valores de FPR. Quando os resultados de FPR foram dicotomizados em vírus X4/R5, foi observado uma discreta associação entre a predição de X4 e maior distância genética de proteínas ($p = 0,055$). Além disso, a presença de nucleótidos ambíguos, assim como ambiguidades que codificam aminoácidos não-sinônimos (mistura não-sinônima – MNS), apresentou uma forte correlação com FPR ($p < 0,0001$). Essas ambiguidades são também significativamente maior entre os vírus X4. A qualidade de uma sequência pode também aumentar a possibilidade de atribuir à uma base, durante a edição da sequência, um nucleotídeo ambíguo, o que levará a um aumento não relevante na

probabilidade de codificar aminoácidos diferentes, pois trata-se de um artefato. As sequências foram cuidadosamente revisadas e a qualidade foi avaliada para minimizar essa possibilidade. Para melhor avaliar esta questão, os casos com maior número de ambiguidades foram excluídos (sequências da V3 com mais de 6 ambiguidades), ainda sim, o número de ambiguidades mostrou-se associada ao vírus X4 e uma correlação inversa com FPR (figura 12). Ambiguidades atribuídas aos nucleotídeos na edição da sequência, e que podem codificar mais de um aminoácidos, são consideradas tanto na predição de resistência genotípica, no gene da polimerase, como na predição tropismo no gene envelope. Algumas ferramentas de bioinformática resolvem essas incertezas e descrevem todas as possibilidades de nucleotídeos, facilitando a interpretação da sequência. O sequenciamento da população viral permite o estudo da variante predominante nas *quasispecies* numa determinada amostra. A presença de nucleotídeos ambíguos pode sugerir a identificação de variação na população viral (Kouyos et al., 2011) ou co-infecção (Antoinette & Cornelissen, 2007). Além disso, a ambiguidade também tem sido correlacionada com variantes X4 em estudos mais detalhados utilizando *deep sequencing* (Prosperi et al., 2011). Um longo período de infecção está associado a um tropismo X4, e a maior diversidade dentro da *quasispecie*. Portanto, a identificação correta de nucleotídeos ambíguos parece refletir uma característica biológica relevante e é útil na correta identificação das diversas subpopulações presentes em uma amostra submetida ao sequenciamento populacional.

5.2. Estudo comparativo entre plasma e célula na predição do tropismo viral utilizando sequências em replicata.

Ao contrário do teste genotípico para resistências aos ARTs, o teste genotípico para de tropismo realizado a partir do DNA proviral, parece fornecer uma predição adequada e algumas redes nacionais tem estruturado a predição baseado no DNA proviral (Svicher et.al., 2012). Alguns trabalhos sugerem uma maior predição de vírus X4 em célula comparado com plasma (Verhofstede et al., 2008).

Como mencionado anteriormente, na ausência de um critério de comparação entre o teste realizado com uma única sequência e em replicatas, o consenso Europeu formulou um *guideline* que recomenda a utilização de um ponto e corte de 10% para sequências em replicata e 20% para sequências únicas. Essa recomendação é aplicado para o teste realizado em RNA e DNA proviral.

Nossa hipótese era de que o teste realizado em replicata poderia diminuir ou eliminar a discordância na predição do tropismo entre plasma e célula. As nossas análises não confirmam esta hipótese, e a utilização de replicatas também está assciado a discordâncias de predição de tropismo nos três valores de ponto de corte mais utilizados (5,75%, 10% e 20%). Árvores filogenéticas geradas usando o método NJ não ajudaram no entendimento destas diferenças. Sequências derivadas de plasma e célula, assim como sequências preditas como R5 e X4, não mostraram uma segregação clara em cluster na maior casos dos casos. Mesmo considerando essas diferenças, foi possível observar uma forte correlação entre a predição do tropismo viral, utilizando o menor valor de FPR nas replicatas, entre os dois diferentes compartimentos (Figura 15). Uma análise adicional das sequências, utilizando a ferramenta PSSM_{X4R5} (<http://fortinbras.us/fssm/index.html>), foi realizada para avaliar se essa discordância também estaria presente. As discordâncias foram similarmente observada.

Alguns parâmetros clínicos estudados, como CD4 ou carga viral, não

apresentaram associação com a predição. No entanto, o número de ambiguidades em cada sequência (Figura 17), assim como a distância genética em proteínas entre sequência (Figura 18), ou a dispersão (desvio padrão) dos valores de FPR obtidos com as replicatas, parece ser mais importante entre as sequências de DNA proviral. Isto poderia fornecer a base do aumento do número de predição de X4 na célula, e poderia significar que a população viral sequenciada a partir de DNA apresenta maior diversidade. A quantidade de material genético utilizado na reação de PCR, também poderia fornecer um impacto nessas diferenças (diversidade). Nós usamos uma quantidade de material extraído relativamente alto (2,5ul), estudos que avaliam diferentes quantidades de material na reação de PCR pode ajudar a esclarecer essa questão.

6. Conclusões

O estudo permitiu o desenvolvimento de um protocolo robusto para predição genotípica do tropismo do HIV-1.

Fatores relacionados a diversidade intra-sequência (número de nucleotídeos ambíguos e número de misturas não-sinônimas) e inter-sequência (árvore filogenética e distância genética de DNA e de proteínas) estão associados a valores baixos de FPR e consequentemente a predição de tropismo viral X4.

O número de replicatas não conduz necessariamente à obtenção de sequências com valores baixos de FPRs ou ao aumento da predição de tropismo X4, sugerindo que o uso de replicata não aumenta randomicamente a chance de identificar variantes virais com tropismo X4.

A discordância na predição genotípica entre plasma e célula foi frequente e o uso de sequências em replicata não resolve a discordância, com uma tendência para a predição maior de tropismo X4 entre as sequências provirais. Estudos de fatores que possam estar associados a essas discordâncias na predição genotípica do tropismo podem contribuir no entendimento da patogênese.

7. Referências

Abbas AK, Lichtman AH. Vírus da Imunodeficiência Humana Síndrome da Imunodeficiência Adquirida. *In: Imunologia Celular e Molecular*. 5º ed. Elsevier: 2005.

Antoinette C van der Kuyl and Marion Cornelissen: Identifying HIV-1 dual infections *Retrovirology*. 2007; 4: 67.

Berger EA, Doms RW, Fenyo EM et al. A new classification for HIV-1. *Nature* 1998; 391: 240.

Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol* 1999; 17: 657–700.

Boletim Epidemiológico–AIDS/DST. Ministério da Saúde - Secretaria de Vigilância em Saúde - Programa Nacional de DST e AIDS, ano 8, nº1, 2010/2011.

Briz V, Poveda E, Soriano V. HIV entry inhibitors: mechanisms of action and resistance pathways. *J Antimicrob Chemother*. 2006;57:619-27

Burton, DR, Desrosiers RC, Doms RW, et al. HIV vaccine design and the neutralizing antibody problem. *Nature Immunology*. 2004; 5 (3): 233-36.

Cabral GB, Ferreira JL, Coelho LP, Fonsi M, Estevam DL, Cavalcanti JS, Brígido LF. Concordance of HIV type 1 tropism phenotype to predictions using web-based analysis of V3 sequences: composite algorithms may be needed to properly assess viral tropism. *AIDS Res Hum Retroviruses*. 2012; 28(7):734-8

Casseb J, Katzenstein D, Winters M, Brígido LFM, et.al. Serotyping HIV-1 with V3 peptides: detection of high avidity antibodies presenting clade-specific reactivity. *Braz. J. Med. Biol. Res.* 2002; 35(3): 369-75.

Chan DC, Fass D, Berger JM, Kim PS. Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell.* 1997; 89: 263-73.

Chan S, Speck R, Power C, et al. V3 recombinants indicate a central role for CCR5 as coreceptor in tissue infection by HIV-1. *J Virol.* 1999;73:2350-8

Comier E, Dragic T. The crown and stem of the V3 loop play distinct roles in HIV-1 envelope glycoprotein interactions with CCR5 coreceptor. *J Virol.* 2002;76:8953-7

de Mendoza C, Van Baelen K, Poveda E, Rondelez E, Zahonero N, Stuyver L, Garrido C, Villacian J, Soriano V; Spanish HIV Seroconverter Study Group. 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; 48(3):241-4

Demares t J, Bonny T, Vavro C, et al. HIV-1 co-receptor tropism in treatment naive and experienced subjects. 44th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). October 30-November 2, 2004. Washington, DC. Abstract H-1136

Este JA, Telenti A. HIV entry inhibitors. *Lancet.* 2007;370:81-8

Fenyo, EM.; Morfeldt-Manson, L; Chiodi, F; et al. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J. Virol* 1988, 62:4414-4419.

Flint SJ, Enquist LW, Krug RM, Racanillo VR, Skalka AM. Principles of Virology: Molecular Biology, Pathogenesis and Control. 2004; Ed AMS Press, 627.

Fouchier RAM. Simple determination of Human Immunodeficiency Virus type 1 syncytium-inducing V3 genotype by PCR. *Journal of Microbiology*. 1995; 44(3): 906-11.

Gallo RC & Montagnier L. Prospects for the future. *Science*. 2002; 298:1730-1730.

Gallo RC. The early years of HIV-AIDS. *Science*. 2002; 298:1728-30.

Golding H, Zaitseva M, de Rosny E, et.al. Dissection of Human Immunodeficiency Virus Type 1 Entry with Neutralizing Antibodies to gp41 Fusion Intermediates. *J. Virol*. 2002; 76: 6780-90.

Göttlinger HG. HIV-1 Gag: A Molecular Machine Driving Viral Particle Assembly and Release. pp.2-28 *HIV Sequence Compendium*. Edited by: Kuiken C, Foley B, Hahn B, Marx P, McCutchan F, Mellors JW, Wolinsky S, Korber B. Publish by: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, LA-UR 0-2877. 2001.

Herrera MS, Klasse PJ, Michael E, et al. The impact of envelope glycoprotein cleavage on the antigenicity, infectivity and neutralization sensitivity of ENV-pseudotyped human immunodeficiency virus type 1 particles. *Virology*. 2005; 338(1): 154-172.

Janeway CA, Travers P, Walport M, Shlomchik M. *O Sistema Imune na Saúde e na Doença*. 5ª ed. São Paulo: Artmed; 2002.

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

Kang S, Quaan F.S, Hung C, et.al. Modified HIV proteins with enhanced binding to neutralizing monoclonal antibodies. *Virology* . 2005; 331: 20-32.

Koning F, Van Rij R, Schuitemaker H. Biological and Molecular Aspects of HIV-1 Coreceptor Usage. pp. 24-42 in *HIV Sequence Compendium*. Edited

by: Kuiken C, Foley B, Hahn B, Marx P, McCutchan F, Mellors JW, Wolinsky S, Korber B. Publish by: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, LA-UR 03-3564. 2002.

Kouyos RD, von Wyl V, Yerly S, Böni J, Rieder P, Joos B, Taffé P, Shah C, Bürgisser P, Klimkait T, Weber R, Hirschel B, Cavassini M, Rauch A, Battegay M, Vernazza PL, Bernasconi E, Ledergerber B, Bonhoeffer S, Günthard HF; Swiss HIV Cohort Study Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. *Clin Infect Dis.* 2011; 15;52(4):532-9

Kwong PD. Antibody: Hitting HIV-1 Where its Armor is Thickest. *Structure.* 2004; 12:173-74.

Lalezari J, et al. Efficacy and Safety of Maraviroc plus Optimized Background Therapy in Viremic ART-experienced Patients Infected with CCR5-tropic HIV-1: 24-Week Results of a Phase 2b/3 Study in the US and Canada. CROI 2007. Abstract 104bLB.

Larder B, Richman D, Vella S. HIV Resistance and Implications for Therapy, 2nd ed., MediCom Inc., Atlanta, 2001.

Liu R, Paxton WA, Choe S, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 1996;86:367-77

McBurney S. P & Ross T M. Viral sequence diversity: challenges for AIDS vaccine designs. *Expert Rev Vaccines.* 2008; 7(9): 1405–17.

McGovern RA, Harrigan PR, Swenson LC. Genotypic inference of HIV-1 tropism using population-based sequencing of V3. *J Vis Exp.* 2010 (46)

Milich L, Margolin B, Swanstrom R. V3 Loop of The Human Immunodeficiency Virus Type 1 Env Protein: Interpreting Sequence Variability. *J. Virol.* 1993; 67(9): 5623-34.

Montagnier L. A History of HIV Discovery. *Science*. 2002; 298: 1727-28.

Moore JP, Doms RW. The Entry Inhibitors : a fusion of science and medicine. *PNAS*. 2003; 100(19): 10508-602.

Mouquet H, Scharf L, Euler Z, Liu Y, Eden C, Scheid JF, Halper-Stromberg A, Gnanapragasam PN, Spencer DI, Seaman MS, Schuitemaker H, Feizi T, Nussenzweig MC, Bjorkman PJ. Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc Natl Acad Sci U S A*. 2012;20:109(47):E3268-77.

Moyle GJ, et al. Epidemiology and predictive factors for chemokine receptor use in HIV-1 infection. *J Infect Dis*. 2005;191:866-872

Nelson M, et al. Efficacy and Safety of Maraviroc plus Optimized Background Therapy in Viremic, ART-experienced Patients Infected with CCR5-tropic HIV-1 in Europe, Australia, and North America: 24-Week Results. *CROI 2007*. Abstract 104aLB

Paxton WA, Martin SR, Tse D, et al. Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposures. *Nat Med* 1996;2:412-7

Petersen ML, van der Laan MJ, Napravnik S, Eron JJ, Moore RD, Deeks SG. Long-term consequences of the delay between virologic failure of highly active antiretroviral therapy and regimen modification. *AIDS*. 2008; 18;22(16):2097-106.

Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, Damond F, Robertson DL, Simon F: A new human immunodeficiency virus derived from gorillas. *Nat Méd*. 2009, 15:871-872.

Poveda E, Alcamí J, Paredes R, Córdoba J, Gutiérrez F, Llibre JM, Delgado R, Pulido F, Iribarren JA, García Delatoro M, Hernández Quero J, Moreno S,

García F. Genotypic determination of HIV tropism - clinical and methodological recommendations to guide the therapeutic use of CCR5 antagonists. *AIDS Rev.* 2010; 12(3): 135-48.

Prosperi MCF, Prosperi L, Bruselles A, Abbate I, Rozera G, Vincenti D, MC, Capobianchi MR, Ulivi G. Combinatorial analysis and algorithms for quasispecies reconstruction using next-generation sequencing. *BMC Bioinformatics.* 2011; 12: 5.

Prusiner SR. Discovering the Cause of AIDS. *Science.* 2002; 298: 1726-49.

Stanfield RL, Gorny MK, Williams C, Zolla-Pazner S, Wilson IA. Structural Rationale for the Broad Neutralization of HIV-1 by Human Monoclonal Antibody. *Structure.* 2004; 12: 193-204.

Svicher V, Alteri C, Montano M et al. Performance of genotypic tropism testing on proviral DNA in clinical practice: results from the DIVA Study Group. *New Microbiologia* 2012; 35: 17-25.

Swenson LC, Moores A, Low AJ, Thielen A, Dong W, Woods C, Jensen MA, Wynhoven B, Chan D, Glascock C, Harrigan PR. Improved detection of CXCR4-using HIV by V3 genotyping: application of population-based and "deep" sequencing to plasma RNA and proviral DNA. *J Acquir Immune Defic Syndr.* 2010; 54(5):506-10

Thomson MM, Delgado E, Herrero I, et al. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. *J. Gen. Virol.* 2002; 83: 107-19.

Trkola A, Dragic T, Arthos J. CD4-dependent, antibody-sensitive interactions between HIV-1 and its correceptor CCR-5. *Nature.* 1996; 184-87.

UNAIDS: <http://www.unaids.org> (Acesso em 05/02/2013)

van Regenmortel MHV, et al. Virus Taxonomy. Seventh Report of International Committee on taxonomy of Viruses. San Diego; Academic Press.2000; 1162.

Vandekerckhove L P R, Wensing AMJ, Kaiser R, Brun-Vézinet F, Clotet B, De Luca B, Dressler S, Garcia F, Geretti A M, Klimkait T, Korn K, Masquelier B, Perno C F, Schapiro J M, Soriano V, Sönnnerborg A, Vandamme A-M, Verhofstede C, Walter H, Zazzi M, Boucher C A B, European Consensus Group on clinical management of tropism testing. European guidelines on the clinical management of HIV-1 tropism testing. AIDS Rev. 2011 ;13(2):77-108.

Vandekerckhove L, Wensing AMJ, Kaiser R, Brun-Vezinet F, Clotet B, De Luca A. European consensus on clinical use and interpretation of HIV-1 tropism testing. HIV Med 2009; 10: 48–49.

Verhofstede C, Vandekerckhove L, Van Eygen V et al. CXCR4-using HIV type 1 variants are more commonly found in peripheral blood mononuclear cell DNA than in plasmaRNA. J Acquir Immune Defic Syndr. 2009; 50: 126–136.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature.2009; 460(7256):711-6.

Wigg MD, Vírus da imunodeficiência humana. In: Santos NSO, Romanos MTV, Wigg MD. Introdução à Virologia Humana. Guanabara Koogan, Rio de Janeiro. 2002. p. 183-97.

Wilkin T, et al. HIV type 1 chemokine coreceptor use among antiretroviral experienced-patients screened for a clinical trial of a CCR5 inhibitor: AIDS Clinical Trial Group A5211. Clin Infect Dis. 2007; 44:591-595

Wyatt R, Sodroski J. The HIV-1 Envelope Glycoproteins: Fusegens, Antigens, and Immunogens. Science. 1998; 280:1884-88.

Zhi-yong Y, Chakrabarti BK, Xu L, et al. Selective Modification of Variable Loops Alters Tropism and Enhances Immunogenicity of Human Immunodeficiency Virus Type 1 Envelope. *Journal of Virology*. 2004; 78(8): 4029-36.

TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO

“Caracterização biológica e molecular do HIV: Identificação de epítomos relevantes ao controle da infecção e ao tropismo viral”

O Instituto Adolfo Lutz, em colaboração com o Programa Estadual de DST/AIDS está realizando um estudo para analisar as estruturas que o vírus HIV utiliza para infectar as células humanas.

A infecção pelo HIV é caracterizada por uma interação entre o vírus e o organismo humano e a evolução da doença está relacionada às características de ambos. Através de exames realizados no genoma humano e viral podemos conhecer alguns dos mecanismos envolvidos na infecção. Atualmente existem alguns medicamentos para HIV que agem bloqueando a entrada do vírus na célula humana, mas dependem de que tipo de estruturas (receptores) usadas pelo vírus para entrar nas células. Os testes de Tropismo do HIV são exames que procuram estudar quais as estruturas na célula o vírus usa na entrada, podendo auxiliar e eventualmente prever se esses novos medicamentos podem ser úteis ou não ao tratamento.

Este estudo prevê a realização de Testes de Tropismo, que inclui análises moleculares de regiões do vírus e das células humanas associadas à entrada do vírus nas células e eventualmente a caracterização biológica do tropismo do vírus através do isolamento do vírus.

As pessoas que participarem deste estudo serão acompanhadas normalmente por seus médicos; por um período de um ano será observada a evolução do tropismo viral a partir desses testes de tropismo e sua condição clínica a partir de informações sobre seu seguimento (como número de células CD4, carga viral, condição clínica), permitindo dessa forma avaliar se existe alguma relação entre os resultados desses testes e sua condição clínica. O término do estudo não interferirá no acompanhamento clínico ou no seu tratamento, que deverá ser continuado normalmente.

O objetivo do estudo é avaliar a aplicabilidade de testes de tropismo no entendimento da doença e a assistência médica dos pacientes. Esse estudo poderá eventualmente subsidiar a implantação de testes de tropismo na rede pública.

A sua participação no estudo implicará no seguinte:

- 1) Anotação pelo seu médico de informações clínicas de seu Prontuário Médico, de forma confidencial, não sendo identificado o seu nome em nenhum momento, durante ou após o estudo, garantindo a sua privacidade e anonimato. Serão utilizados códigos para sua identificação, sendo permitido o acesso apenas ao respectivo médico e aos responsáveis pelo estudo.
- 2) Coleta de material biológico (2 tubos de sangue), o correspondente a duas colheres de sopa, realizada no braço, podendo haver os inconvenientes dessas coletas, como leve dor no local da picada da agulha ou hematoma (mancha roxa). Todas as medidas serão tomadas para que não ocorram problemas. As amostras serão coletadas no Setor de coleta do seu serviço clínico ou local determinado pelo seu médico e sempre que possível a coleta será realizada junto com a coleta de seus exames de rotina, realizadas na entrada do estudo, no mês seis e no mês doze.
- 3) Caso você possua material biológico armazenado na Instituição que você é atendido este material também poderá ser analisado para a realização retrospectiva destas análises, se for mais adequado.

Embora esses exames adicionais possam não ter aplicabilidade prática hoje, os resultados serão enviados ao seu médico que informará você dos resultados, assim que estiverem concluídos. Será criado um banco de parte do sangue, que será guardado pelo período de 5 anos após o término do estudo. Caso algum exame não previsto seja realizado, você será consultado e novo Termo de Consentimento será solicitado, e realizado apenas após seu consentimento.

Se durante o estudo, surgir algum teste que possa ser feito com esse material armazenado, e que seja considerado por você ou por seu médico como benéfico á sua saúde, e que não seja parte do estudo, você poderá solicitar o material biológico coletado pelo Laboratório e, de acordo com sua vontade, destiná-lo a outro uso, respeitando as normas do Comitê de Ética das Instituições envolvidas e a Legislação em vigor.

Este estudo não garante que haja benefícios diretos para você, mas as informações coletadas e analisadas em conjunto, poderão contribuir muito para o combate da AIDS e no uso melhor e mais racional dos medicamentos anti-retrovirais.

Caso você não queira participar deste estudo, sua recusa não ocasionará em nenhum prejuízo ao seu atendimento médico.

Após leitura desse texto, compreendi os objetivos do estudo e concordo em participar como voluntário.

Assinatura do aplicador TCLE _____

Nome completo do paciente: _____

Assinatura: _____

Data: São Paulo, ____/____/____

Se você tiver qualquer dúvida sobre as condições aqui apresentadas, sobre a pesquisa e seus direitos enquanto participante, entre em contato com seu médico, ou com o Dr Luis Brígido, fone (11) 3068-2982, coordenador deste estudo e responsável pelo cumprimento das condições aqui apresentadas ou com o Dr. Eduardo Ronner Lagonegro, Coordenador do Comitê de Ética em Pesquisa, telefone: (011) 5087-9837. (Versão 7 abril 2008)



SECRETARIA DE ESTADO DA SAÚDE
COORDENADORIA DE CONTROLE DE DOENÇAS
CENTRO DE REFERÊNCIA E TREINAMENTO DST/AIDS
COMITÊ DE ÉTICA EM PESQUISA



São Paulo, 13 de Maio de 2008.

Ofício CRT-DST/AIDS – CEP n.º 075/2008.

PROTOCOLO CEP n.º 004/2008: "Caracterização Biológica e Molecular do HIV-1: Identificação de Epítomos Relevantes ao Controle da Infecção e ao Tropismo Viral".

- Projeto;
- Termo de Consentimento Livre Esclarecido – Versão datada de 07 de abril de 2008.


Prezado Investigador,

Após análise dos referidos documentos pelos membros deste Comitê, em reunião ordinária do dia 12/05/2008, foi emitido parecer: **APROVADO**.

Conforme a Resolução CNS 196/96, deve ser enviado um relatório semestral, a contar desta data, comunicando o andamento do estudo e um relatório de conclusão após o término do mesmo.

Solicitamos sejam rigorosamente observados os prazos para envio dos Relatórios Semestrais.

Atenciosamente,


Eduardo Ronner Lagonegro
Coordenador do Comitê de Ética em Pesquisa
CRT-DST/AIDS

Anexo: TCLE devidamente assinado.

Ilmo. Sr.
Prof.º Dr. Luis Fernando de Macedo Brígido
Investigador Principal

Teste genotípico para predição do tropismo do HIV-1

**Luana Portes Ozório COELHO, Gabriela Bastos CABRAL,
João Leandro de Paula FERREIRA, Luís Fernando de
Macedo BRÍGIDO**

*Núcleo de Doenças Sanguíneas e Sexuais, Centro de Virologia,
Laboratório de Genotipagem de HIV, Instituto Adolfo Lutz*

O uso adequado de terapias antirretrovirais é uma questão estratégica em Saúde Pública. Entre as inovações na área de medicamentos, encontram-se novos agentes que impedem a entrada do HIV na célula hospedeira, através da inibição da ligação do vírus ao correceptor CCR5¹. A eficácia destes medicamentos depende da necessidade do vírus de utilizar o correceptor CCR5, não atuando em variantes que utilizam correceptores CXCR4. Portanto, determinar o tropismo viral tem sido necessário para subsidiar a decisão clínica quanto ao uso desses novos medicamentos. A predição do tropismo também é importante no entendimento da patogênese viral, sendo o uso pelo vírus do correceptor CCR5 correlacionado com uma melhor evolução clínica^{2,3}. No laboratório de Retrovírus do Instituto Adolfo Lutz (IAL), o fenótipo viral tem sido determinado a partir da análise das sequências da região do envelope viral (V3), utilizando-se diferentes ferramentas de bioinformática como alternativa aos métodos fenotípicos como, por exemplo, o Trofile™. O teste de genotipagem

(Genotropismo) em triplicata tem se mostrado mais robusto e com melhor concordância com Trofile™⁴. O método compreende a extração do RNA, a partir de plasma. O RNA é retrotranscrito e amplificado em *One Step PCR* (Reação em Cadeia da Polimerase). A segunda amplificação (*Nested PCR*) permite a obtenção parcial da região do envelope viral do HIV-1. O produto dessa reação é sequenciado utilizando-se rodaminas marcadas (Big Dye) em sequenciador automático 3130 XL (Applied Biosystems). A edição das sequências é realizada em software Sequencher e analisada com ferramentas de bioinformática. As sequências são submetidas em *sites* que permitem a predição do tropismo viral (<http://coreceptor.bioinf.mpi-inf.mpg.de/index.php>) do Instituto Max Planck e (<http://indra.mullins.microbiol.washington.edu/pssm>) da Universidade de Washington.

Esta metodologia está sendo utilizada para subsidiar a decisão clínica em pacientes com múltiplas falhas terapêuticas que necessitam de terapia de resgate, assim como em estudos relacionados ao entendimento da patogênese viral

em pacientes atendidos em serviços do SUS no Estado de São Paulo.

REFERÊNCIAS

1. MacArthur RD, Novak RM. Maraviroc: The first of a New Class of Antiretroviral Agents. *Clinical Infectious Diseases*. 2008; 47:236-41.
2. Richman DD, Bozzette SA. The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. *J Infect Dis*. 1994; 169(5):968-74
3. Waters L, Mandalia S, Randell P, Wildfire A, Gazzard B, Moyle G. The Impact of HIV Tropism on Decreases in CD4 Cell Count, Clinical Progression, and Subsequent Response to a First Antiretroviral Therapy Regimen. *Clin Infect Dis*. 2008; 46:1617-1623
4. Swenson LC, Moores A, Low AJ, Thielen A, Dong W, Woods C et al. Improved Detection of CXCR4-Using HIV by V3 Genotyping: Application of Population- Based and "Deep" Sequencing to Plasma RNA and Proviral DNA. *Journal of AIDS*. 2010; 54(5):506-10.

Concordance of HIV Type 1 Tropism Phenotype to Predictions Using Web-Based Analysis of V3 Sequences: Composite Algorithms May Be Needed to Properly Assess Viral Tropism

Gabriela Bastos Cabral,¹ João Leandro de Paula Ferreira,¹ Luana Portes Osório Coelho,¹ Mylva Fonsi,² Denise Lotufo Estevam,² Jaqueline Souza Cavalcanti,¹ and Luis Fernando de Macedo Brígido¹

Abstract

Genotypic prediction of HIV-1 tropism has been considered a practical surrogate for phenotypic tests and recently an European Consensus has set up recommendations for its use in clinical practice. Twenty-five anti-retroviral-experienced patients, all heavily treated cases with a median of 16 years of antiretroviral therapy, had viral tropism determined by the Trofile assay and predicted by HIV-1 sequencing of partial *env*, followed by interpretation using web-based tools. Trofile determined 17/24 (71%) as X4 tropic or dual/mixed viruses, with one nonreportable result. The use of European consensus recommendations for single sequences (geno2pheno false-positive rates 20% cutoff) would lead to 4/24 (16.7%) misclassifications, whereas a composite algorithm misclassified 1/24 (4%). The use of the geno2pheno clinical option using CD4 T cell counts at collection was useful in resolving some discrepancies. Applying the European recommendations followed by additional web-based tools for cases around the recommended cutoff would resolve most misclassifications.

DETERMINATION OF VIRAL TROPISM is a necessary step prior to the use of CCR5 antagonists and may provide clues in HIV pathogenesis. Genotypic assays are interesting alternatives to phenotypic assays, and although different interpretations of genetic data have been suggested, this issue is still unresolved. Recently a European consensus proposed the use of genotypic data. They suggested the use of the geno2pheno clonal option, with a false-positive rate (FPR) of 20%, to predict tropism on a single population genome and a 10% cutoff is recommended to predict tropism based on replicates.¹

The evaluation of the concordance of phenotype determination of HIV-1 coreceptor usage to genotypic prediction has been an objective of different studies.²⁻⁶ The sensibility and specificity of genotype prediction are influenced by different factors including viral subtype and the prevalence of CXCR4 using variants in the population assessed.^{2,7,8} Among phenotypic assays, the Trofile is a reference. It is the only Clinical Laboratory Improvement Amendments (CLIA) certified assay, and its upgraded ESTA version has a high performance, with reported sensibility to X4 variants as low as 0.3% (http://www.trofileassay.com/what_is_trofile.html). Some aspects, however, limit its widespread use, including cost, sample

transport logistics, a number of nonassayed samples, the inability to assess cell-associated genomes, and level of viremia necessary to run the assay (1000 copies/ml).

Although the actual correlation of the genotypic prediction with clinical response should be the main objective of these tests, that is, the assays should predict the clinical usefulness of CCR5 antagonists and not just show intraassay comparability; these data are limited and complex to analyze, as therapy success depends on many factors.⁶ Therefore phenotypic assays are still important to validate genotypic predictions. In this study we compare the results of viral tropism as determined by the phenotypic Trofile ESTA assay to different genotypic tools.

Patients with virological failure on antiretroviral therapy, considering CCR5 antagonists as part of salvage therapy, were consecutively included. Two sets of patients were studied, paired samples, with two EDTA tubes obtained at the same blood drawn, at the clinical site and some additional unpaired cases, in which blood collection for the genotype assay was done before Trofile collection. Informed consent was obtained from all volunteers. V3 sequences were obtained with a nested PCR of the partial *env* genome as previously

¹Adolfo Lutz Institute, Virology Center, Retrovirus Laboratory, São Paulo, Brazil.

²Centro de Referência de Aids, São Paulo, Brazil.

described from bulk virion RNA ($n=19$) or cell DNA ($n=8$, samples 6, 7, 18, 19, 20, 21, 22, and 23).⁹ The V3 region was directly sequenced from PCR products in both directions. Fasta contigs were generated from two to four primers, with manual editing by Sequencher 4.6 software, considering nucleotide ambiguities (accession number JN 541241-65). The V3 sequence harboring nucleotide ambiguities was translated in all possible amino acids. Sequences with three or more amino acid mixtures due to nonsynonymous nucleotide ambiguities within the V3 region were not submitted to PSSM and net charge analyses. HIV-1 subtyping of partial *env* sequence was performed at NCBI and Rega websites.

The V3 sequences were interpreted according to different genotypic tropism predictions, including the 11/25 rule, which is based on the presence of arginine, histidine, or lysine at positions 11 or 25 of the V3 loop and two bioinformatics methods, PSSM (<http://ubik.microbiol.washington.edu/computing/pssm>), X4/R5 option, that analyze the composition and position of amino acids at the V3 sequence, generating a score concerning the likelihood of using the CXCR4 coreceptor and the Geno2pheno coreceptor (<http://coreceptor.bioinf.mpi-sb.mpg.de/cgi-bin/coreceptor.pl>), a statistical method based on the sequences FPRs, the likelihood of a sequence being mistakenly classified as a CXCR4.^{10,11}

Sequences FPR were obtained using both clonal and clinical options, the latter using recommended nadir CD4 T cell (TCD4) counts and also TCD4 at time of collection, along with viral load; cell counted with flow cytometry (BD, USA) and viral RNA with and B-DNA (Siemens, USA).

Tests were run blind to the other results, and both were delivered to a reference physician who could use either information to subsidize clinical management. Reports from Trofile were reported as R5, X4, or dual/mixed tropism. Our report to clinicians designated the predicted tropism as R5, X4, or possible X4, using a composite rule. Trofile results were reported as R5-tropic, X4, and dual/mixed-tropic. In this study we dichotomized the cases as X4 (including dual/mixed and possible X4) or R5.

A total of 25 patients were included in this study, mostly (76%) males, with a median age of 39 years. All patients were in virological failure under antiretroviral therapy, with a median TCD4 of 218 cells/mm³ and viral load of 10,000 copies/ml. Study cases consisted mostly of heavily treated patients, treated for a median of 16 years, having lived with HIV for a long time. The majority had a clade B envelope, with

one case clade F (Sample 21). Samples were collected from December 2008 to October 2009 for Trofile tests, with 17 cases of paired and 8 cases of unpaired samples, with collection for genotyping from 2007 to 2009. Table 1 describes patients' demographic and clinical data.

Most HIV-1 infections were of X4-tropic viruses, 17/24 (71%), using the Trofile assay, whereas the IAL criteria showed 16/25 (64%), with one nonreportable case by Trofile. The FPR geno2pheno were obtained in three options—one clonal and two clinical evaluations: (1) using the last TCD4 available at the time of tropism test collection and (2) using nadir TCD4, as recommended. Most cases showed an FPR below 20%: 15/25 (60%), 14/24 (58%), and 13/17 (77%), respectively. It should be noted that the lack of information at collection or the nadir TCD4 did not allow the use of the clinical option for 1/25 (4%) and 8/25 (32%) cases, respectively. According to the "11/25 rule" 10/25 (40%) and at PSSM 9/22 (41%) were classified as X4-tropic viruses. The net charge ranged from 4 to 9. However, three samples (12%) with more than three nonsynonymous amino acid possibilities due to nucleotide ambiguities were not evaluated by PSSM: samples 8, 10, and 19. Table 2 describes individual data of the study cases, with the V3 alignments along with the Trofile result and interpretations of major algorithms. Table 3 compares the Trofile results assuming it to be the gold standard for defining disease (X4 or dual/mixed) or "no disease," involving only the R5 viral populations, to evaluate different determination algorithms. The overall concordance between distinct genotypic tools and ESTA ranged from 50% (Geno2pheno_{10clinicalNadir}) to 95.8% (IAL criterion), with a sensitivity and specificity for the detection of X4 variants using genotypic-based algorithms, as compared to Trofile, varying according to genotypic algorithms. The lower sensitivity was 37.5% Geno2pheno_{10clinicalCollection}. However, the specificity was adequate for many algorithms, including the 11/25 rule, Motivate, Geno2pheno_{10clinicalCollection}, PSSM, and IAL criteria.

The population studied included patients with advanced disease, exposed to multiple ARV regimens that were considering the use of CCR5 antagonists as part of the salvage regimen. Therefore, this should explain the high prevalence of X4 and X4/R5 dual-tropic viruses, when compared with others studies.^{2,7} The study conclusions are limited by the small sample size and the use of a single sequence instead of replicates to predict tropism, but some points are relevant and deserve attention. The fact that the genotypic evaluation was

TABLE 1. CLINICAL AND DEMOGRAPHIC DATA OF PATIENTS INCLUDED IN THE STUDY ACCORDING TO TROFILE

	R5	DM/X4	Total
Age* (years)	45 (41–48)	18 (13–40)	39 (15–45)
Gender (% males)	86	76	76
TCD4* at collection (cells/mm ³)	226 (95–277)	184 (94–372)	218 (94–327)
Nadir TCD4* (cells/mm ³)	37 (17–79)	35 (27–97)	36 (23–108)
Viremia* at collection (log ₁₀ /ml)	3.43 (2.53–4.19)	4.47 (3.59–4.92)	4 (2.98–4.76)
Number of regimens*	8 (7–9)	6 (5–9)	8 (5–9)
Time on ARV* (years)	15 (13–16)	16 (4–17)	16 (13–17)
N (%)	7 (28)	17 (68)	25 ^a (100)

^aOne case of Trofile not reportable, reported as R5 by the IAL rule.

Patient characteristics according to Trofile result, R5 and dual/mixed or X4 tropism, including age at the time of collection, percentage of males, TCD4 counts at collection, and the nadir value, documented throughout follow-up, plasma viremia (viral load) at collection, number of different ARV combinations (regimens) used by the patient and the total time on ARV therapy in years, expressed as medians and 25th–75th IQR percentiles*, and the total number of cases in each group. ARV, antiretroviral.

TABLE 2. PROFILE AND GENOTYPIC TROPISM PREDICTION OF STUDY CASES

Samples	Profile	IAL criteria	G2F clonal FPR (%)	G2F clinical cFPR (%)	G2F clinical nFPR (%)	Rule 11/25	PSSM	Net charge	V3 sequence
1	R5	R5	44.8	49.1	23.2	R5	R5	5	CTRPNNNTRKSIHMGWGGRAFYATGDIIGDIRQAHC
2	D/M	R5	48.4	12.6	12.6	R5	R5	6	CVRPNNNTRKSIHMGWGGRAFYATGEIIGNIRQAHC
3	R5	R5	22.3	26.3	8.2	R5	R5	(5-4)	CTRPNNNTRKGIHMGJMVJGRAFYATGEIIGDIRQA[HY]C
4	D/M	X4	4.8	48.4	19	X4	X4	5	CTRPNNNTRKSIPIVSGRILYATGKLGDIRQAYC
5	Not reportable	R5	42.8	62.6	38.2	R5	R5	8	CTRPNNNTRRGIHMGPKAFYATGNIIGNIRKAHC
6	D/M	X4	0.5	0.2	0.1	X4	X4	6	CTRPNNNTRIRGIRIGPGRAVLATERIIGDIRRAHC
7	R5	R5	20.2	19.0	4.9	R5	R5	6	CTRPNNNTRKSVHIGPGSALYTTNIIGNIRRAHC
8	D/M	X4	3.2	11.9	6.8	R5	R5	—	CTRP x N x TRKSIH x G GRAFI x C x VI G x I x I QAHC
9	R5	R5	93.6	13.4	—	R5	R5	4	CTRLNNNTRRSIQMGWGGRAFYATGDIIGDIRLAHC
10	D/M	X4	7.8	71	—	R5	—	—	CARPNNNTRK S I Q H J MGW RG RAFYATGDI IV GDIRQAHC
11	D/M	X4	14.6	14.8	—	R5	R5	4	CERPNNNTRKGIHLGGRAFFATEAIGDIRQAHC
12	R5	R5	84.9	37.4	15.7	R5	R5	6	CTRPNNNTRKGIHMGWGRTLYATGAIIGDIRQAHC
13	D/M	X4	11.4	39.9	11.4	R5	R5	7	CTRPNNNTRKSIHMGAGRAFYTNGQIIGNIRQAHC
14	R5	R5	7.9	85.7	76.4	R5	R5	5	CTRPNNNTRKGIHIGPGRGAFYATDITGDIRQAHC
15	X4	X4	20.8	1.8	1.5	R5	X4	8	CTRPDYVAHKSINMLWGRRFHATGALKGNIKHL
16	R5	R5	57	51.4	8.8	R5	R5	5	CTRPNNNTRRSIHMAAGRALYTTDIIIGDIRQAHC
17	D/M	X4	2.5	13	13	X4	X4	8	CTRLNNLTRSIRIGPGGAWYAAAGRIVGKIRPAHC
18	X4	X4	0.7	0.1	—	X4	X4	6	CTRPNNNTRKRVTMGPRVWYTTGEIVGDIKRAHC
19	X4	X4	21.5	1.8	—	X4	—	—	CTRPNNNTRKSI HLGW x RT x I x I x I IGDI x I x A x C
20	X4	X4	15.6	15	—	R5	R5	4	CTRPNNNTRRSITIGPGRGAFYGTDIIGDIRQAHC
21	X4	X4	1.8	17	—	X4	X4	9	CTRPNNNTRKSIHIGLGHAFR T A TEKIIGNIRKAHC
22	X4	X4	0	—	—	X4	X4	8	CTRPGNKTGKRIRIGHIGPGRFTYTEKIRDIRQAHC
23	D/M	X4	18.3	23.2	23.2	R5	R5	(7-5)	CTRPNNNTRKGIHIGPGRSFYATG KE IIGDIRQAHC
24	D/M	X4	1.7	0.4	0	X4	X4	6	CSRPNNNTRKGISIGPGRAVYATEKIIGNIRQAHC
25	D/M	X4	0.1	0	0.4	X4	X4	7	CTRPNNNTRKRLYSSRWRTLYATE IV I GV GDIRQAHC

Data from each case including Profile result, IAL composite interpretation, geno2pheno false-positive rates (FPR) of clonal and clinical option using TCD4 at collection TCD4 (cFPR) and nadir TCD4 (nFPR), 11/25 rule, PSSM, net charge, and the V3 sequence. Possible amino acids at positions with ambiguous nucleotides are shown as [MV]; cases 8 and 19 were not resolved [X] due to excess ambiguities.

TABLE 3. CONCORDANCE, SENSITIVITY, AND SPECIFICITY OF GENOTYPING TOOLS TAKING AS REFERENCE THE TROFILE ASSAY

Genotypic algorithms	Concordance (%)	Sensitivity (%)	Specificity (%)
11/25 rule	70.8	59	100
Geno2pheno _{Motivate}	66.7	53	100
Geno2pheno _{20Clonal}	79.2	76	85.7
Geno2pheno _{20clfnicoC}	73.9	75	71.4
Geno2pheno _{20clfnicoN}	68.8	90	33.3
Geno2pheno _{10clonal}	66.7	58.8	85.7
Geno2pheno _{10clfnicoC}	69.6	37.5	100
Geno2pheno _{10clfnicoN}	50.0	50	50
PSSM	76.2	64	100
IAL criterion	95.8	94	100

Agreement between Trofile results with different genotypic algorithms; geno2pheno were evaluated using different cutoffs. Values of false-positive rate, including 5.75% (motivate), 10%, and 20%, using both clonal and clinical option, the latter applying both recommended nadir TCD4 (N) and TCD4 at collection (C).

conducted in a public health laboratory setting is notable and documents the feasibility of the test in resource-limited settings.

The reliability of the genotypic methods may be influenced by polymorphisms found in Brazilian clade B sequences.¹² HIV-1 clade B GWGR bearing isolates, an uncommon motif in the V3 loop worldwide but with an important prevalence in Brazil, was observed in 28% of isolates, including one case with discordant genotype/Trofile result.

The use of replicates, as suggested by the European consensus, should further improve prediction. It is important to note that this discordant case (sample 2, Table 2) was re-evaluated as replicates without resolving the discrepancy.

The use of individual prediction algorithms showed, as in previous studies, a robust specificity but a suboptimal sensitivity.^{2,13} Using the geno2pheno clinical option with nadir TCD4 we obtained the best sensitivity score, 90%. The specificity, however, is low (Table 3). A recent study also suggests the usefulness of geno2pheno clinical option prediction.⁴ In this study we further evaluated the clinical option using both the nadir and the TCD4 at the time of the collection. This has provided some additional information, as discussed below.

Overall, the analysis of the sequence set showed a high specificity with most web-based methods, higher than most reports in the literature.^{2,14} Although this could be secondary to the small sample size evaluated here, with our discordance increasing as more samples are analyzed, many studies did not use the new ESTA Trofile assay. The earlier version of Trofile, and possible other phenotypic assays, may lack the high sensibility of ESTA to X4. These tests would therefore miss some X4 in paired, comparative studies. The lack of one or more X4 predictions by the “gold standard” would imply that some real X4 cases, identified by a genotypic algorithm with high specificity as PSSM and 11/25 rule, would be considered false-positive results, when they were actually performing better than the comparable phenotypic assays. Although plausible, this cannot be confirmed at this point but should be considered when interpreting these studies.

In spite of the fact that individual algorithms do not show a good concordance to the Trofile results, mostly due to low sensibility, the use of a “composite” rule, as applied by our service, improves prediction. Previous studies have also sug-

gested that a combination of bioinformatics tools could improve the sensitivity and specificity.¹⁵ The use of the European Consensus in this study would lead to 4/24 (16.7%) misclassifications (samples 2, 14, 15, and 19). In this sense, the use of Geno2pheno_{20Clonal} could perform better if used in combination with other available genotypic prediction tools. In our analyses, a reliable prediction of X4 virus is obtained for samples with an FPR value below 5.75% (motivate cutoff). For R5, all but one case (sample 2) would be resolved with a cutoff of 22%; above 50% would yield a 100% concordance to Trofile. However, values within this range can benefit from other genotypic tools to improve the prediction, and the use of highly specific criteria, such as the “11/25 rule” and PSSM as well as the clinical option of geno2pheno, can be useful in this range. To illustrate, samples 3, 7, 15, and 19, all with a clonal FPR above 20%, would be considered R5 by European consensus. This would wrongly classified 2/4 (50%) (samples 15 and 19) (Table 2). The use of clinical geno2pheno (applying either nadir TCD4 or the last available TCD4 determination) may help resolve cases in this range. If clinical FPR is incorporated in the analysis, using collection TCD4, the two R5 have FPRs around 20% (26.3% and 19%), whereas the two X4 have an FPR below 5.75%. Additionally, these clinical FPRs are in concordance with PSSM and the results of the 11/25 rule. It is of note that the use of the clinical option with the nadir TCD4 would not resolve cases 3 and 7, as both show a low FPR at this option (8.2% and 4.9%).

The usefulness of these additional resources is also suggested by case 23, in which the clonal FPR, near the cutoff, can be further supported by incorporating the 11/25 rule. The most intriguing case is the discordant sample 2, with a high clonal FPR of 48.4%. This case has R5 in both the 11/25 rule and PSSM, but a low FPR (12.6%) at the clinical option that was not considered at the time of the test report.

Another situation is cases with an FPR below 20% but above 5.75%. Most (5/6) cases would be correctly classified as X4 by the European consensus recommendation, but by considering clinical option results, over 75%, along with R5 by PSSM and 11/25 rules, would resolve the single discordance. Although highly predicted for this study, these rules and cutoff suggestions should be improved and validated by larger studies. However, our data do not support the assumption that a single parameter, albeit practical, may be adequate for the prediction of genotypic tropism for Brazilian patients with advanced disease.

On the other hand, the study also suggests that a composite evaluation of existing tools may prove robust in predicting HIV coreceptor tropism.

Acknowledgments

This work was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) Grant 2009/8215-0. Additional unrestricted educational funding was received from ViiV HealthCare/GSK and Trofile tests were paid by Pfizer Inc.

Author Disclosure Statement

No competing financial interests exist. LFMB have received honoraria for lectures from pharmaceutical companies, including ViiV HealthCare/GSK and travel support from Pfizer.

References

1. Vandekerckhove L, Wensing A, Kaiser R, *et al.*: European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect Dis* 2011;10:70319–70324.
2. Garrido C, Roulet V, Chueca N, *et al.*: Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes. *J Clin Microbiol* 2008;46:887–891.
3. Qu S, Ma L, Yuan L, *et al.*: Co-receptor usage and prediction of V3 genotyping algorithms in HIV-1 subtype B from paid blood donors experienced anti-retroviral therapy in chinese central province. *Virology* 2010;7:280.
4. Prosperi M, Bracciale L, Fabbiani M, *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.
5. Svicher V, D'Arrigo R, Alteri C, *et al.*: Performance of genotypic tropism testing in clinical practice using the enhanced sensitivity version of Trofile as reference assay: Results from the OSCAR Study Group. *New Microbiol* 2010;33:195–206.
6. McGovern R, Thielen A, Mo T, *et al.*: Population-based V3 genotypic tropism assay: A retrospective analysis using screening samples from the A4001029 and MOTIVATE studies. *AIDS* 2010;24:2517–2525.
7. Raymond S, Delobe P, Mavigner M, *et al.*: CXCR4-using viruses in plasma and peripheral blood mononuclear cells during primary HIV-1 infection and impact on disease progression. *AIDS* 2010;24:2305–2312.
8. Seclén E, Garrido C, González MM, *et al.*: High sensitivity of specific genotypic tools for detection of X4 variants in anti-retroviral-experienced patients suitable to be treated with CCR5 antagonists. *J Antimicrob Chemother* 2010;65:1486–1492.
9. Ferreira JL, Thomaz M, Rodrigues R, *et al.*: Molecular characterization of newly identified HIV-1 infections in Curitiba, Brazil: Preponderance of clade C among males with recent infections. *Mem Inst Oswaldo Cruz* 2008;103:800–808.
10. Jensen M, Li Fu-Sheng, Wout A, *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;77:13376–13388.
11. Sing T, Low A, Beerenwinkel N, *et al.*: Predicting HIC coreceptor usage on the basis of genetic and clinical covariates. *Antiviral Ther* 2007;12:1097–1106.
12. Santoro-Lopes G, Harrison L, Tavares M, *et al.*: HIV disease progression and V3 serotypes in Brazil: Is B different from B-Br? *AIDS Res Hum Retroviruses* 2000;16:953–958.
13. Poveda E, Séclen E, González MM, *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;63:1006–1010.
14. Low AJ, Dong W, Chan D, *et al.*: Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. *AIDS* 2007;21:F17–F24.
15. Chueca N, Garrido C, Álvarez M, *et al.*: Improvement in the determination of HIV-1 tropism using the V3 gene sequence and a combination of bioinformatics tools. *J Med Virol* 2009;81:763–767.

Address correspondence to:
Luis Fernando de Macedo Brígido
Av Dr Arnaldo, no. 355
CEP 01246-902
Cerqueira César
Centro de Virologia
São Paulo, SP
Brazil
E-mail: lubrigido@gmail.com



Evaluation of genotypic prediction of HIV-1 tropism using population sequencing of replicates[☆]

Joao Leandro de Paula Ferreira, Luana Portes Ozorio Coelho, Rosangela Rodrigues, Gabriela Bastos Cabral, Jaqueline de Souza Cavalcanti, Paula Morena de Souza Guimaraes, Luis Fernando de Macedo Brigido*

Adolfo Lutz Institute, Virology Center, Retrovirus Laboratory, Sao Paulo, Brazil

A B S T R A C T

Article history:

Received 15 July 2011
Received in revised form 31 October 2011
Accepted 9 November 2011
Available online 26 November 2011

Keywords:

HIV-1
Viral tropism
PCR
Virion
Co-receptors
Brazil

Determination of human immunodeficiency virus tropism has contributed to the understanding of the pathogenesis of HIV and is necessary prior to the use of CCR5 antagonists. Replicate V3 sequences may generate different sequences and improve viral tropism prediction. The diversity of HIV was evaluated to access its influence on prediction. Plasma RNA was retro-transcribed and amplified using a one-step protocol, followed by nested PCR and sequencing using an ABI3130XL. Eighty-one patients, 74% male and 26% female, with a median age of 44 years had either a single sequence ($n=50$) or 2–4 replicates ($n=31$) evaluated. Most patients (92%) had used multiple anti-retroviral regimens. Tropism prediction was performed using the Geno2pheno clonal option. The number of ambiguous nucleotides, the deduced non-synonymous amino acids at V3 and the genetic distance were quantified. Using a 20% false positive rate (FPR) cut-off, 41/81 (50.6%) was predicted as X4. TCD4 was lower, 226 cells/mm³ (IQR 82–378), in patients infected with X4; TCD4 for R5 was 324 cells/mm³ (IQR 200–538, $p < 0.05$). The number of ambiguous nucleotides correlated with a lower FPR value ($p < 0.0027$). Although different sequences may be generated, the number of replicates was not associated to a lower FPR or X4 assignment, and may allow a better prediction of this biological characteristic. Ambiguous nucleotides correlate inversely to a lower FPR.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The study of HIV-1 (human immunodeficiency virus type 1) tropism has contributed to the understanding of viral pathogenesis (Fenyö et al., 1989; Koot et al., 1993; Shepherd et al., 2008). More recently, the development of CCR5 antagonists has made tropism determination necessary to prescribe these drugs as part of anti-retroviral combination therapies (FDA, 2007). Although viral phenotype assays are available commercially and at several research facilities, cost and other considerations make genotypic tropism prediction an interesting alternative.

Similar to many other RNA viruses, HIV biology is characterised by variability. Therefore, HIV should not be referred to as a single viral isolate but rather as a swarm of similar variants or a quasispecies (Nowak et al., 1990; Laming and Andino, 2010); a cloud of diversity around a master sequence. It is thought that only single or a few variants are able to establish infection in a

new host (Fischer et al., 2010). Genetic distance within the quasispecies population increases during the infection, both diverging from the initial inoculate as well as displaying more heterogeneity among the different variants (Shankarappa et al., 1999). The ability to identify this diversity in patients during the chronic stage of the disease is a limitation of genotypic tests that depend upon population sequencing.

Direct population sequencing allows the study of the major variant. The sequencing of HIV clones and, more recently, deep sequencing (Fischer et al., 2010) has allowed a much broader delineation of the variability present in a given HIV quasispecies. These methods, however, are complex and impractical for the clinical laboratories that perform sequencing assays to assist clinical decision, especially in settings with limited resources. An alternative would be to amplify, in independent reactions, the RNA present in the sample, allowing the sequencing of more variants and consequently better representing the original diversity. This methodology was proposed by the British Columbia Centre for Excellence in HIV/AIDS has proved useful for the prediction of viral tropism (McGovern et al., 2010). Factors that may influence viral tropism prediction were evaluated and a method is described based on parallel, independent retro-transcription, amplification and sequencing of viral RNA.

[☆] Supported by FAPESP 2009/08215-0.

* Corresponding author at: Av Dr Arnaldo 355, Virology Center, Retrovirus Laboratory, Sao Paulo 01246-902, Brazil. Tel.: +55 11 30682982; fax: +55 11 30682983.
E-mail address: lbrigido@gmail.com (L.F.d.M. Brigido).

Table 1
Clinical and demographic data.

	Single sequence	Replicate sequence	All cases
<i>n</i>	50	31	81
Age (years)	45 (39–49)	41 (34–47)	44 (38–49)
Gender (% males)	74%	74%	74%
TCD4 at collection (cells/mm ³)	241 (80–420)	305 (186–511)	291 (116–453)
Nadir TCD4 (cells/mm ³)	75 (24–128)	106 (38–161)	88 (24–161)
Viral load at collection (Log ₁₀ /mL)	3.99 (3.36–4.73)	4.08 (3.54–4.47)	4.05 (3.43–4.61)
Number of ARV regimens	7 (4–9)	8 (5–9)	8 (5–9)
Time on treatment (years)	11.8 (10.2–13.5)	12.3 (10.6–13.9)	12.0 (10.2–13.6)

Patient characteristics according to number of sequences obtained for this study. *n*, total number of cases in each group (cases with single sequence, cases with replicates, and all cases studied); age, at the time of collection for tropism test; percentage of males; TCD4 counts at collection; TCD4 nadir, value documented throughout follow-up; viral load; number of ARV, combinations used by the patient; time on ARV therapy in years.

2. Materials and methods

2.1. Study population

Biological samples included in the study were collected from patients followed at clinical sites in Sao Paulo state from 2008 to 2010. Patients provided written informed consent obtained at clinical sites by physicians. A report of genotypic tropism prediction was generated to support salvage therapy for patients with advanced disease. The study was approved by the Ethical Committees of the participating institutions. Table 1 lists the characteristics of the patients.

2.2. HIV-1 *env* genetic sequencing

Plasma collected in EDTA was separated by centrifugation within 6 h of collection and stored at -70°C until analysis. HIV RNA was extracted from plasma using the QIAmp Viral RNA Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. RNA was amplified in replicates using a nested one-step reverse transcription PCR (RT-PCR) method. Amplification of the HIV *envelope* (*env*) gene was performed using 2.5 μL of RNA template. The reaction consisted of MgSO₄ 2 mM, dNTP's 10 mM, DTT 0.1 M, Inhibitor RNase 10 U, Superscript III 20 U, Taq Platinum High Fidelity 0.625 U, buffer and RNase/DNase free water, all reagents from Invitrogen (USA) with 10 μM of each primer SQV3F1, forward position 5'-GAG CCA ATT CCC ATA CAT TAT TGT-3' (6858 → 6878), and CO602, reverse position 5'-GCC CAT AGT GCT TCC TGC TGC TCC CAA GAA CC-3' (7814 ← 7786). The final volume was 25 μL . Positive and negative controls were included in each experiment. PCR cycle conditions consisted of 50 $^{\circ}\text{C}$ for 30 min and 94 $^{\circ}\text{C}$ for 2 min, followed by 40 cycles of 94 $^{\circ}\text{C}$ for 15 s, 55 $^{\circ}\text{C}$ for 30 s, 68 $^{\circ}\text{C}$ for 1 min 30 s, and 68 $^{\circ}\text{C}$ for 10 min. Five microlitres of RT-PCR product was used in a second PCR (nested) with primers SQV3F2, forward position 5'-TGT GCC CCA GCT GGT TTT GCG AT-3' (6879 → 6898), and CD4R, reverse position 5'-TAT AAT TCA CTT CTC CAA TTG TCC-3' (7652 ← 7672), with Go Taq Green Master Mix (Promega, USA) in a final volume reaction of 50 μL . PCR cycle conditions consisted of 10 cycles of 94 $^{\circ}\text{C}$ for 2 min, 10 cycles 94 $^{\circ}\text{C}$ for 15 s, 55 $^{\circ}\text{C}$ for 30 s, 72 $^{\circ}\text{C}$ for 1 min, 25 cycles of 94 $^{\circ}\text{C}$ for 15 s, 55 $^{\circ}\text{C}$ for 30 s, 72 $^{\circ}\text{C}$ for 1 min (increasing by 5 s in each sequential elongation), with a final elongation period of 72 $^{\circ}\text{C}$ for 7 min. PCR reactions were considered valid when both the PCR products and the positive control bands were observed at approximately 700 bp when separated on a 1.3% agarose gel and visualised with SYBR Safe DNA gel staining (Invitrogen, USA) along with controls. Some of the single sequences were obtained using a previously published nested PCR protocol based on the WHO guideline recommendations (Ferreira et al., 2008). Nested PCR products were quantified using a low DNA mass ladder (Invitrogen, USA) diluted (1:10) in water to allow an input of 5–10 ng of DNA for rhodamine incorporation. The Big Dye Kit (Ready Reaction

Mix, Applied Biosystems, Foster City, CA) was used for sequencing with the primers V3O2, forward position 5'-AAT GTC AGY ACA GTA CAA TGT ACA C-3' (6948 → 6969), ES7, forward position 5'-CTG TTA AAT GGC AGT CTA GC-3' (7005 → 7021), ED33, reverse position 5'-TTA CAG TAG AAA AAT TCC CCT C-3' (7360 ← 7378), and SQV3, reverse position 5'-GAA AAA TTC CCT TCC ACA ATT AAA-3' (7350 ← 7370). Cycle sequencing conditions consisted of 25 cycles at 96 $^{\circ}\text{C}$ for 10 s, 50 $^{\circ}\text{C}$ for 5 s, 60 $^{\circ}\text{C}$ for 4 min, 8 $^{\circ}\text{C}$ for an undetermined length of time. After cycling, DNA was precipitated using a solution of sodium acetate and ethanol, followed by denaturation using Hi-Di Formamide at 94 $^{\circ}\text{C}$ for 4 min. Sequences were determined with an automated sequencer (Genetic Analyzer 3130 XL, Applied Biosystems).

2.3. Tropism prediction

Electropherogram data were analysed using Sequencher 4.6 software. Sequences were edited manually, and a consensus was generated based on two or more sequences and included ambiguous nucleotides (mixtures). Sequence quality control and preliminary subtyping was performed using the web-based tools NCBI locator, BLAST and NCBI Genotyping, referring back to the electropherograms as needed. The co-receptor used by HIV-1 was predicted from the V3 *env* region nucleotide sequence by Geno2pheno (URL: <http://coreceptor.bioinf.mpi-sb.mpg.de/cgi-bin/coreceptor.pl>). The determination of tropism was obtained by submitting the sequences (FASTA format) to Geno2pheno using the clonal option, which relies on ranking of the submitted sequences to a curate data set. A false positive rate (FPR), the probability of falsely classifying an R5-virus as X4, was assigned to each submitted sequence. These rates, using the lower FPR from all sequences obtained, allowed us to determine the likelihood of each sequence tropism prediction as well as the patients' HIV infection population tropism prediction as recommended by the European consensus (Vandekerckhove et al., 2011).

2.4. Distance metrics

An estimation of diversity at V3 of each sequence (intra-sequence) was evaluated as follows: (i) the number of ambiguous nucleotides; (ii) the number of all possible amino acids coded by ambiguities allowing for non-synonymous mixtures (NSM); and (iii) the number of amino acid positions (codons) at V3 with NSM. The diversity of sequences between replicates (inter-sequence) was quantified by Prodist, which computes a distance matrix from protein sequences taking into account a variety of sequence ambiguities, and Gendist, which uses nucleotide sequences to compute a distance matrix, using Bioedit software. Bootstrap support was used for phylogenetic tree nodes, and 1000 replicates were constructed using Paup4 by applying the evolution model selected using Modeltest (GTR+I+G). Epi6 was used for data management,

Student's *t*-test, Mann–Whitney *U*-test and linear correlations (Spearman) were performed using GraphPad Prism. The results are expressed as medians, 25th and 75th percentiles (IQR), with a level of significance of $p < 0.05$.

3. Results

Env sequences were obtained from 81 patients, 74% male, with a median age of 44 years. Most (92%) were on anti-retroviral treatment (ARV) but were not exposed to CCR5 antagonists, and 8% were naïve to all ARV. Patients on ARV were heavily treated, exposed at collection to a median of 8 different regimens (3–14) during a median treatment period of 12 years (10–13). The median viral load was 4.05 Log₁₀ and TCD4 263 cells/mm³. From 50 patients' cases, a single *env* sequence was obtained, and from 31 cases a median of 4 replicates (2–4) were obtained (Table 1). The use of a one-step protocol, using three independent PCR reactions, carried out on plasma RNA extracted from samples with viraemia (over 2000 copies/mL), allowed the evaluation of three sequences in 78% of cases, two sequences in 16% of cases, and one sequence in 6% of cases. Overall, 41/81 (50.6%) were predicted as X4 tropic using a 20% FPR cut-off. TCD4 was significantly lower among cases infected with variants predicted as X4, TCD4 226 cells/mm³ (IQR 82–378), compared with R5, TCD4 324 cells/mm³ (IQR 200–538, $p < 0.05$).

The diversity between replicates (inter-sequence) at V3 varied; a representative phylogenetic tree is shown in Fig. 1. Samples from the same patient grouped within clusters that were supported by bootstrap from 74% to 100%. Out of 31 V3 sequence replicates, 80.6% had some nucleotide or nucleotide ambiguity difference, and 64.5% of replicate sets had one or more amino acid or NSM difference. The median genetic distance was zero (IQR 0.0–0.0098) and the median protein distance was 0.045 (IQR 0.011–0.089). Within each sequence (intra-sequence diversity), the median number of ambiguous nucleotides was 0.95 (IQR 0–2.85) and was similar for both singles and replicates. The median number of NSM or codons with NSM was zero (IQR 0–2) and was slightly higher, but not significantly, for replicates ($p = 0.2$ and $p = 0.3$, respectively). As expected, the number of ambiguities correlated strongly to the number of NSM ($p < 0.0001$). There was no correlation between either inter or intra-sequence diversity and TCD4 values or viral load; however, although it was not significant ($p = 0.3$), the protein distance appeared to decrease with viral load and increase with TCD4 counts (Fig. 2).

3.1. Value FPR

False positive rates, as well as X4 or R5 prediction by dichotomisation at 20% FPR, were inversely correlated to both ambiguities and NSM ($p < 0.0001$). When sequences with a high number of ambiguities (6 or more ambiguous bases) were removed from the analysis, this correlation was still observed ($p = 0.0027$, Fig. 3). The inter-replicate protein distance showed some association to X4 prediction ($p = 0.055$, Fig. 4). However, the number of replicates did not correlate to either a lower predicted FPR of a replicate set or to the individual FPR obtained for each sequence (Fig. 5).

4. Discussion

This study evaluated a protocol for the amplification of replicate population (bulk) sequencing from plasma samples. We describe a method for one-step RT-PCR using reagents that may facilitate its implementation in settings with limited resources. The analysis and the sequence interpretation of genotypic tropism is a crucial step. Tropism prediction may be performed using tools on different web-sites and by inspection of the sequences, as using the “11/25 rule”,

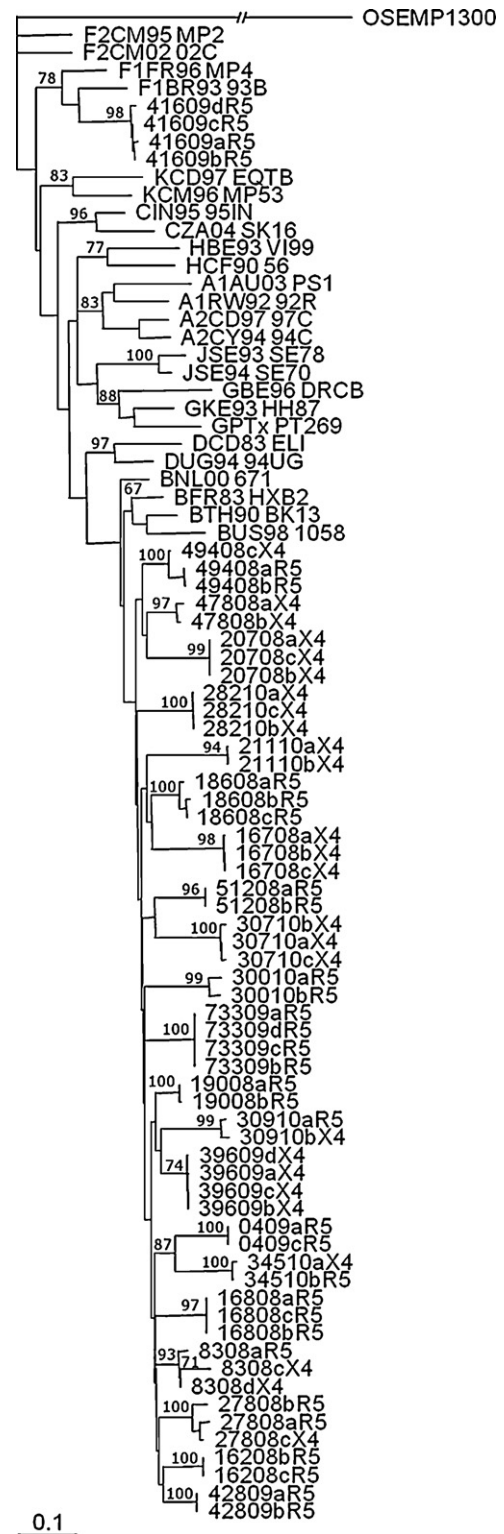


Fig. 1. Representative maximum likelihood phylogenetic tree of partial *env*-*lope* region sequences of the study, constructed according to evolutionary model (GTR+I+G) as selected at Modeltest (PAUP* v4.0b10). References HIV-1 subtypes were obtained from the Los Alamos HIV Sequence Database. Bootstrap values (1000 replicates) at key branches are depicted. Sequences from study are designated by ID number, a letter (e.g. a,b,c) for each sequence in cases with replicates and designated as X4 or R5 according to the Geno2pheno tropism prediction of each sequence using a false positive rate of 20%, clonal option.

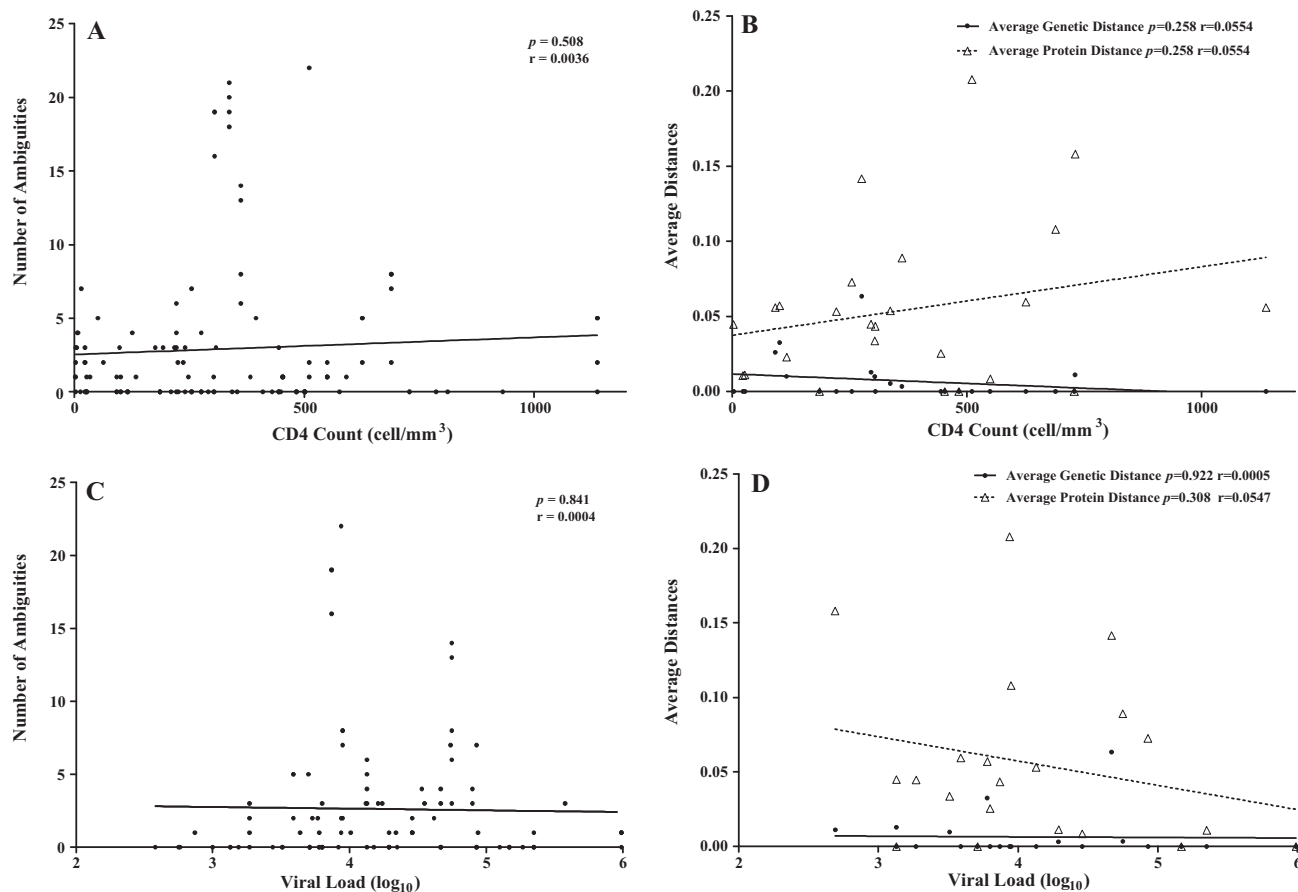


Fig. 2. TCD4 values, cells/mm³ (A, B) and viral load, Log₁₀/mL, (C, D) along with number of ambiguities (A, C) and genetic and protein distance (B, D). Linear correlation (Spearman) *r* and *p* value is shown at top right.

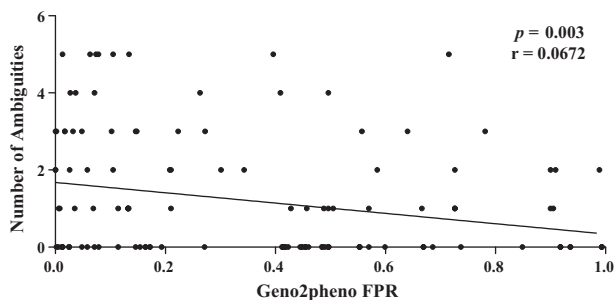


Fig. 3. False positive rate (FPR), obtained at clonal option of Geno2pheno website along with number of ambiguities. Only sequences with less than 6 ambiguities at V3 loop are shown. Linear correlation (Spearman) *r* and *p* value is shown at top right.

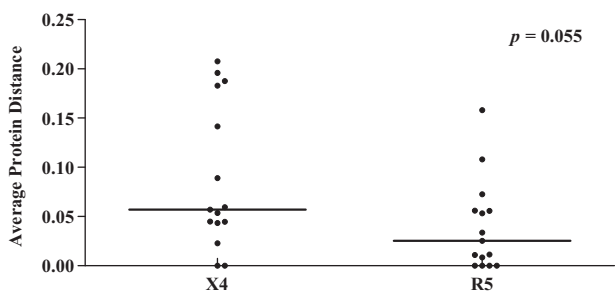


Fig. 4. Average protein distance according to assigned predicted viral phenotype CXCR4 (X4) or CCR5 (R5) using a false positive rate (FPR) of 20% cut-off for categorization.

based in the identification of basic side chains in residues 11 or 25. Here, we did not evaluate all methods of sequence interpretation, but instead aimed to simplify the evaluation and analysis of the sequence parameters described here with the Max Planck resource, Geno2pheno (clonal option). It is a user-friendly tool, allowing the evaluation of sequences with ambiguous nucleotides. A European consensus has recently recommended its use, establishing differential cut-offs for single and replicate sequences when applying the test for clinical use, and considering the lower FPR for viral tropism prediction when replicates implicate different FPRs for a sequence set (Vandekerckhove et al., 2011).

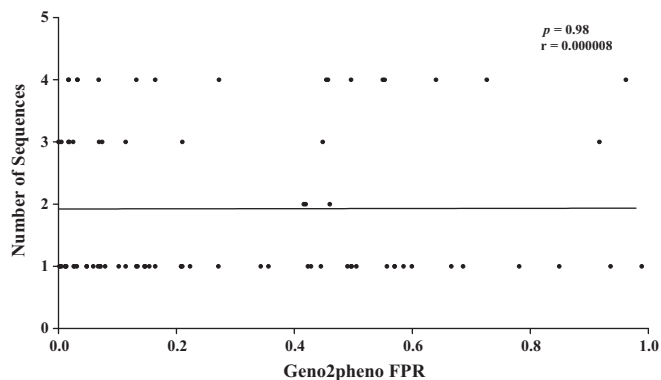


Fig. 5. V3 sequence false positive rate (FPR), or the lower FPR for replicates, according to the number of sequence analysed in each case, as one (single sequence) up to four replicates from same plasma sample.

The lowest cut-off for defining the presence of X4 variants with replicates was proposed by the Vancouver group (McGovern et al., 2010) after a retrospective evaluation of the Motivate study results. This study defined X4 cases as having an FPR below 2%, possible X4 with an FPR from 2% to 5.75% and R5 for cases with an FPR above 5.75%. The European consensus suggests a cut-off of 10% for cases with replicates and a 20% cut-off for single sequences. We used the conservative 20% cut-off in this study. Although some of the single sequences were obtained using different primers, as suggested by WHO guidelines, edition and tropism interpretation were the same as those used for replicates. Sequences were evaluated both individually and as replicates. Our aim was to better understand the role of independent amplification of replicates on the FPR. Therefore, we compared predictions inferred from both single and replicate (2–4) sequences to evaluate parameters that may influence viral tropism prediction. We also evaluated the FPR value as a continuous variable to evaluate its correlation to viral diversity metrics.

The use of replicates to predict viral tropism poses the question of whether the number of sequences alone could increase the likelihood of finding an X4 tropic variant. It may be suggested that the higher the replicate number, the more likely is the chance of detecting an X4 virus. This would represent an X4 classification secondary to the number of replicates. Indeed, this is a problem in “deep sequencing”, where a cut-off is defined (Swenson et al., 2011), above which the identification X4 is assumed to be “real”. As shown in Fig. 5, the number of replicates does not correlate to a lower FPR or to the likelihood of detecting X4 viruses. We did, however, find a strong correlation between some diversity parameters and the prediction of X4 populations. In this study, the potential diversity generated intra-sequence, as the genetic and protein distance among replicates, suggested an inverse correlation to FPR. When FPR results were separated as either X4/R5 viruses, a weak association of X4 to higher protein distance was observed ($p=0.055$). Moreover, the presence of ambiguous nucleotides, as well as ambiguities that would code for non-synonymous amino acid mixtures, shows a strong inverse correlation to a lower FPR ($p<0.0001$). These ambiguities are also significantly higher among X4 viruses. The quality of a sequence may also increase the chance of assigning a base as an ambiguous nucleotide, and consequently a non-biological increase in the probability of different deduced amino acids. Sequences were carefully revised and evaluated for quality to minimise this possibility. Furthermore, when cases with higher numbers of ambiguities were excluded (e.g. over 6 ambiguities at V3), the number of ambiguities still associated to X4 viruses and showed an inverse correlation to FPR (Fig. 3).

Ambiguities in nucleotide assignments at sequences that may code for more than one amino acid are considered both in the prediction of genotype resistance from the polymerase gene, as well as in tropism prediction from envelope genes. Some bioinformatics tools resolve these uncertainties and describe all possible nucleotide translations, thereby facilitating sequence interpretation. Manual revision further allows the correct interpretation of sequence information. Population sequencing allows the study of the most represented variant within the actual quasispecies in a given individual, and the identification of ambiguous nucleotides suggests the presence of more diverse populations (Kouyos et al., 2011) or of a co-infection (Cornelissen et al., 2007). Moreover, ambiguity has also been correlated to X4 variants using deep sequencing (Abbate et al., 2011). A longer period of infection is associated to an X4 tropism, and the larger diversity within the quasispecies of an older infection could increase the probability of having X4 variants. It is possible that the weighing of ambiguities by the time of infection might improve prediction; our study cannot discount this possibility.

The fact that we compared replicates with singles from different patients hinders comparability, and paired experiments amplifying

unique or multiple sequences would be more precise. However, as the conditions of amplification, sequencing and interpretation were similar in both groups, the evaluation of individual sequences from cases with more than one sequence and the random assignments of patients to either single or replicate groups compensate in part to overcome this limitation.

In conclusion, the presence of ambiguous nucleotides and non-synonymous amino acid mixtures at the sequence level, and to some extent the distance between replicate sequences, correlates to a lower FPR and consequently to X4 viral tropism prediction. More importantly, the number of replicates did not lead to a lower FPR or an increase in X4 prediction, suggesting that replicates do not randomly increase the chance of identifying an X4 tropic variant, but rather allow a better identification of this biological characteristic. However, although these clinical cut-offs are useful to guide CCR5 antagonist use in clinical practice, the extent to which the quasispecies must be assessed to predict adequately X4 tropic behaviour requires further investigation.

References

- Abbate, I., Vlassi, C., Rozera, G., Bruselles, A., Bartolini, B., Giombini, E., Corpolongo, A., D'Offizi, G., Narciso, P., Desideri, A., Ippolito, G., Capobianchi, M.R., 2011. Detection of quasispecies variants predicted to use CXCR4 by ultra-deep pyrosequencing during early HIV infection. *AIDS* 25, 611–617.
- Approval of Maraviroc, CCR-5 Co-receptor Antagonist for Treatment of HIV. Available from: <http://www.fda.gov/ForConsumers/ByAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm124198.htm> (last accessed 06.03.11).
- Cornelissen, M., Jurriaans, S., Kozaczynska, K., Prins, J.M., Hamidjaja, R.A., Zorgdrager, F., Bakker, M., Back, N., van der Kuyl, A.C., 2007. Routine HIV-1 genotyping as a tool to identify dual infections. *AIDS* 23, 807–811.
- Fenyö, E.M., Albert, J., Asjö, B., 1989. Replicative capacity, cytopathic effect and cell tropism of HIV. *AIDS* 3, 5–12.
- Ferreira, J.L.P., Thomaz, M., Rodrigues, R., Harrad, D., de Oliveira, C.M., Oliveira, C.A.F., et al., 2008. Molecular characterizations of newly identified HIV-1 infections in Curitiba, Brazil: preponderance of clade C among males with recent infections. *Mem. Inst. Oswaldo Cruz* 103, 800–808.
- Fischer, W., Ganusov, V.V., Giorgi, E.E., Hraber, P.T., Keele, B.F., Leitner, T., Han, C.S., Gleasner, C.D., Green, L., Lo, C.C., Nag, A., Wallstrom, T.C., Wang, S., McMichael, A.J., Haynes, B.F., Hahn, B.H., Perelson, A.S., Borrow, P., Shaw, G.M., Bhattacharya, T., Korber, B.T., 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One* 5 (8), e12303.
- Koot, M., Keet, I.P., Vos, A.H., de Goede, R.E., Roos, M.T., Coutinho, R.A., Miedema, F., Schellekens, P.T., Tersmette, M., 1993. Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. *Ann. Intern. Med.* 118, 681–688.
- Kouyos, R.D., von Wyl, V., Yerly, S., Böni, J., Rieder, P., Joos, B., Taffé, P., Shah, C., Bürgisser, P., Klimkait, T., Weber, R., Hirschel, B., Cavasini, M., Rauch, A., Battegay, M., Vernazza, P.L., Bernasconi, E., Ledergerber, B., Bonhoeffer, S., Günthard, H.F., Swiss HIV Cohort Study, 2011. Ambiguous nucleotide calls from population based sequencing of HIV-1 are a marker for viral diversity and the age of infection. *Clin. Infect. Dis.* 15, 532–539.
- Lauring, A.S., Andino, R., 2010. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6 (7), e1001005.
- McGovern, R.A., Harrigan, P.R., Swenson, L.C., 2010. Genotypic inference of HIV-1 tropism using population-based sequencing of V3. *J. Vis. Exp.* 27, 46.
- Nowak, M.A., May, R.M., Anderson, R.M., 1990. The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS* 4, 1095–1103.
- Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C.R., Learn, G.H., He, X., Huang, X.L., Mullins, J.I., 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* 73, 10489–10502.
- Shepherd, J.C., Jacobson, L.P., Qiao, W., Jamieson, B.D., Phair, J.P., Piazza, P., Quinn, T.C., Margolick, J.B., 2008. Emergence and persistence of CXCR4-tropic HIV-1 in a population of men from the Multicenter AIDS Cohort Study. *J. Infect. Dis.* 199, 1104–1112.
- Swenson, L.C., Mo, T., Dong, W.W., Zhong, X., Woods, C.K., Jensen, M.A., Thielen, A., Chapman, D., Lewis, M., James, I., Heera, J., Valdez, H., Harrigan, P.R., 2011. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. *J. Infect. Dis.* 203, 237–245.
- Vandekerckhove, L., Wensing, A., Kaiser, R., Brun-Vézinet, F., Clotet, B., De Luca, A., Dressler, S., Garcia, F., Geretti, A., Klimkait, T., Korn, K., Masquelier, B., Perno, C., Schapiro, J., Soriano, V., Sönnnerborg, A., Vandamme, A.M., Verhofstede, C., Walter, H., Zazzi, M., Boucher, C., 2011. European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect. Dis.* 11, 394–407.

Manuscript Number: DMID-13-150

Title: ENV V3 SEQUENCE BASED PROVIRAL DNA AND VIRION RNA TROPISM
PREDICTION DISCORDANCE USING REPLICATES

Article Type: Original Article

Keywords: Viral Tropism, PCR, HIV-1, virion, provirus, CCR5, CXCR4, loop
V3

Corresponding Author: Dr. Luis Fernando Brígido,

Corresponding Author's Institution:

First Author: Luana Portes Ozório Coelho

Order of Authors: Luana Portes Ozório Coelho; Gabriela Bastos Cabral;
João Leandro de Paula Ferreira; Paula Morena de Souza Guimarães; Luis
Fernando Brígido

Abstract: Analyses of HIV-1 Envelope V3 sequences are used in studies of viral pathogenesis and prior to CCR5 antagonist prescription to predict viral tropism. Many studies reported discordances when sequences derived from cell DNA or plasma RNA are compared. We assessed if the use of replicate sequences, analyzed in two commonly used bioinformatics tools, would decrease this discordance. HIV-1 partial env region from paired proviral DNA and virion RNA sequenced and tropism predicted with Geno2Pheno (FPRclonal) and PSSM. Paired plasma/cell single (n=28) or replicates (n=26) sequences were obtained from 54 patients. The lower virion FPRclonal value strongly correlated (Spearman $r=0.99$ $p<0.0001$). However, discordance in tropism prediction was observed both in singles and replicates at different geno2pheno cutoffs, and PSSM did not decrease tropism prediction discordance. Provirus sequences had a higher genetic distance (Protdist) and FPRclonal prediction values dispersion (Wilcoxon signed-rank test), suggesting that replicate assays may better represent the diversity of HIV-1 variants.



SECRETARIA DE ESTADO DA SAÚDE
COORDENAÇÃO DOS INSTITUTOS DE PESQUISA
INSTITUTO ADOLFO LUTZ
CENTRO DE VIROLOGIA
Laboratório de Retrovírus



January 23, 2013

Dr Ronald N.Jones, M.D
Editorial Office, Diagnostic Microbiology and Infectious Disease
Suit A
345 Beaver Kreek Centre
North Liberty
Iowa 52317
USA

Dear Dr Ronald N. Jones, M.D

We would like to submit the manuscript entitled "**Env v3 sequence based proviral DNA and virion RNA tropism prediction discordances using replicates**" for consideration for publication at Diagnostic Microbiology and Infectious Disease.

Thank you for your attention,

Luis Fernando de Macedo Brígido
Laboratório de Retrovírus - Virologia
Instituto Adolfo Lutz
RG 1.733.1.0/SSP-PE
CRM/SP 45.482

Luis Fernando de Macedo Brígido, M.D, Ph.D.
Instituto Adolfo Lutz, Retrovirus Laboratory
Virology Center. Ave. Dr. Arnaldo 355
São Paulo- SP 01246-902 Brazil
Phone: 55 11 3068 2982
E-mail: lubrigido@gmail.com

Supplementary Materials

[Click here to download Supplementary Materials: Suggestion of Reviewers.doc](#)

**ENV V3 SEQUENCE BASED PROVIRAL DNA AND VIRION RNA TROPISM
PREDICTION DISCORDANCE USING REPLICATES**

1
2
3
4 Luana Portes Ozório Coelho¹, João Leandro de Paula Ferreira¹, Gabriela Bastos Cabral¹, Paula
5
6 Morena de Souza Guimaraes¹, Luis Fernando de Macedo Brigido¹.
7
8

9
10 1 Adolfo Lutz Institute, Virology Center, Retrovirus Laboratory, Sao Paulo, Brazil.
11

12
13 Disclaimer: The authors did not identified any potential conflicts of interest
14
15

16
17 **Contact information for corresponding author**
18

19 Luis Fernando de Macedo Brigido
20

21 Av Dr Arnaldo 355, 01246-902
22

23 Virology Center, Retrovirus Laboratory, Sao Paulo, Brazil
24

25 Fone +55 11 30682982
26

27 Fax +55 11 30682983
28

29 E-mail address: lubrigido@gmail.com
30
31

32
33 Supported by FAPESP 2009/08215-0 and 2011/21958-2
34
35

36
37 Running head: **Cell and plasma tropism prediction using replicates**
38
39

40 **Manuscript word count:** 2782
41

42
43 **Abstract word count:** 152
44

45
46 **Figures:** 3
47

48
49 **Tables:** 2
50

51
52 **Supplementary:** 2
53

54
55 Part of that content of this manuscript was presented at the XIX International AIDS
56
57 Conference, poster discussion sessions as abstract THPE 074
58
59
60
61
62
63
64
65

Abstract

Analyses of HIV-1 Envelope V3 sequences are used in studies of viral pathogenesis and prior to CCR5 antagonist prescription to predict viral tropism. Many studies reported discordances when sequences derived from cell DNA or plasma RNA are compared. We assessed if the use of replicate sequences, analyzed in two commonly used bioinformatics tools, would decrease this discordance. HIV-1 partial *env* region from paired proviral DNA and virion RNA sequenced and tropism predicted with Geno2Pheno (FPR_{clonal}) and PSSM. Paired plasma/cell single (n=28) or replicates (n=26) sequences were obtained from 54 patients. The lower virion FPR_{clonal} value strongly correlated (Spearman $r=0.99$ $p<0.0001$). However, discordance in tropism prediction was observed both in singles and replicates at different geno2pheno cutoffs, and PSSM did not decrease tropism prediction discordance. Provirus sequences had a higher genetic distance (Protdist) and FPR_{clonal} prediction values dispersion (Wilcoxon signed-rank test), suggesting that replicate assays may better represent the diversity of HIV-1 variants.

Key words: Viral Tropism, PCR, HIV-1, virion, provirus, CCR5, CXCR4, loop V3

Introduction

Human immunodeficiency virus type 1 (HIV-1) entry into host cells generally requires interactions of the envelope glycoprotein gp 120 with the CD4 receptor and with one of the chemokine receptors, CCR5 or CXCR4. Virus strains are classified as CCR5 tropic (R5), CXCR4 tropic (X4) or dual tropic (R5/X4), according to their coreceptor use (Berger et al., 1998; Berger et al., 1999). The study of HIV-1 coreceptor use has implications in viral pathogenesis, due to its association to disease progression in HIV-1 infected individuals (Regoes and Bonhoeffer, 2005; Waters et al., 2008; Raymond et al., 2008). Determining HIV-1 coreceptor usage is also critical as CCR5 coreceptor has become the target for a new class of anti-HIV-1 drugs that specifically inhibit the entry of CCR5-tropic HIV-1 strains into the target cells. Maraviroc is the first approved CCR5 antagonist that entered clinical practice in 2007. Assessment of HIV-1 coreceptor usage is mandatory for the clinical use of this drug (MacArthur and Novak, 2008; Vandekerckhove et al., 2011).

The determination of coreceptor use is based mainly on phenotypic assays, one CLIA approved, Trofile™, and ‘in-house’ tests (Whitcomb et al., 2007; Trouplin et al., 2001). Genotypic prediction of viral tropism has been proposed as an alternative to phenotypic assays (Vandekerckhove et al., 2009; Harrigan and Goretti, 2010). Genotypic testing is based on the analysis of the V3-loop *envelope* (*env*) sequence of HIV-1, using sequence alignment motifs or bioinformatics prediction tools to predict coreceptor usage. In spite the fact that other gp120 regions (as V1, V2, and C4), and the gp41 transmembrane protein (Thielen et al., 2011; Dimonte et al., 2011) may be involved in viral tropism, V3 based genotypic testing is commonly used. It is less technically demanding, more rapid and less expensive than phenotypic testing, thereby meeting today’s need for a fast and reliable assay for the routine diagnostics practice. Emerging data provide evidence for the reliability of bulk genotyping testing to predict virological responses to Maraviroc (McGovern et al., 2010). Both genotypic and phenotypic testing rely on the initial amplification of the HIV-1 glycoprotein gp120 coding

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

sequence from plasma viral RNA. Detectable HIV-1 RNA is needed for successful virion based assay. In many patients with early failure, for whom a treatment change is considered, the HIV-1 RNA will take some time reach this level, allowing resistance accumulation to drugs composing a HAART regimen. In addition, for some patients a treatment change may be considered when the viral load is suppressed, as to address problems of intolerance to one of the HAART components. Provirus DNA therefore is considered a potential alternative source of viral genetic for tropism testing in patients with low or undetectable viral load (Frange et al., 2009; Soulie et al., 2007; Verhofstede et al., 2009). RNA virus are known to consist of a group of similar variants known as quasispecies (Domingo et al., 2012). The use of ultra deep sequencing better grasp this variability, but its application on clinical routine is not yet feasible. The analysis of more than one sequence may improve the identification of variations of viral tropism within the HIV quasispecies, and the use of triplicates has shown to better predict Maraviroc response (McGovern et al., 2010). Moreover, the use of replicates has been proposed for use in clinical practice for tropism prediction in Europe (Vandekerckhove et al., 2011). Tropism prediction based on Provirus DNA or Virion RNA has been shown to diverge in many studies that compare the two compartments, with a tendency for a higher X4 prediction in cell derived DNA (Verhofstede et al., 2009). The aim of this study was to evaluate genotypic tropism testing from plasma virion and provirus DNA to investigate differences in the prediction and whether the use of replicates would minimize these discrepancies.

Material and Methods

Study population

Biological samples included in the study were collected from HIV infected patients followed at clinical sites in Sao Paulo state from 2008 to 2010. Patients provided written informed consent, obtained at clinical sites by physicians. A report of genotypic tropism prediction was generated

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

to subsidize salvage therapy. The study was approved by Ethical Committees of the participating institutions.

Nucleic acid isolation

Blood collected with EDTA (BD, USA) was separated by centrifugation within 6 hours of collection and stored at -70 °C. HIV RNA was extracted from plasma with QIAmp Viral RNA Kit (Qiagen Hilden Germany). Provirus DNA was extracted from peripheral blood mononuclear cell with QIAamp DNA Blood Mini Kit (Qiagen Hilden Germany), both following the manufacturer's instructions. For each sample, one extraction from plasma and one from cell was performed.

Viral RNA amplification

RNA was independently amplified in replicates using nested one-step reverse transcription PCR (RT-PCR) method, (Ferreira et al., 2012) adapted from a British Columbia Excellence Center protocol. Briefly, the amplification of partial HIV *envelope* (*env* V3) gene was performed with 2.5µL of RNA input, Superscript III and Taq Platinum High Fidelity (Life Technologies). The second PCR was done using 2.5µL of RT-PCR products with Go Taq Green Master Mix (Promega, USA).

Proviral DNA amplification

The HIV-1 *env* fragment (circa 960 bp) including the V3 region was performed with 2.5µL of DNA input in each replicate. Reaction mixture include 10uM of each primer ED5, forward position ATG GGA TCA AAG CCT AAA GCC ATG TG (6557 → 6582) and ED12, reverse position AGT GCT TCC TGC TGC TCC CAA GAA CCC AAG (7811 ← 7782) with Taq Platinum High Fidelity (Life Technology) to a final volume reaction of 25uL. Positive and negative controls were included in each run. PCR conditions consisted of 10 cycles 94°C 2min, 94°C 15sec, 55° C 30sec, 68° 1min, 25 cycles 94°C 15sec, 55°C 30sec, 68°C 1 min add 5 sec/cycle, and 68°C 7min. A second PCR (nested) was performed with 2.5 µL of first round

1 PCR products and primers ED31, forward position CCT CAG CCA TTA CAC AGG CCT GTC
2 CAA AG (6817 → 6845) and ES8, reverse position CAC TTC TCC AAT TGT CCC TCA
3 (7668 ← 7648), using Taq DNA Polymerase (Invitrogen), to a final volume reaction of 25µL.
4 PCR conditions consisted of 10 cycles 94°C 2min, 94°C 15sec, 55° C 30sec, 72°C 1min, 25
5 cycles 94°C 15sec, 55°C 30 sec, 72°C 1min (increasing by 5sec in each sequential elongation),
6 with a final elongation period with 72°C 1min add 5 sec/cycle. PCR reaction was considered
7 valid when the PCR products and positive control bands were visualized around 700 bp at 1.3%
8 agarose gel with SYBR Safe DNA gel staining (Invitrogen, USA), along with negative controls.
9

10 **Partial *env* region sequencing**

11 Each amplified product from either virion and provirus replicates were marked with Big Dye
12 Kit (Ready Reaction Mix, Applied Biosystems, Foster City, Calif) with the primers V3O2,
13 forward position 5' AAT GTC AGY ACA GTA CAA TGT ACA C 3' (6948 → 6969), ES7,
14 forward position 5'CTG TTA AAT GGC AGT CTA GC 3' (7005 → 7021), ED33 reverse
15 position 5'TTA CAG TAG AAA AAT TCC CCT C 3' (7360 ← 7378) and SQV3, reverse
16 position 5'GAA AAA TTC CCT TCC ACA ATT AAA 3' (7350 ← 7370). Cycle sequencing
17 conditions consisted of 25 cycles 96°C 10sec, 50°C 5sec, 60°C 4min, 8°C infinity.
18

19 After cycle sequencing RNA and Proviral DNA precipitation using a solution of sodium acetate
20 and ethanol was followed by denaturation using Hi-Di Formamide at 94°C 4min and resolved at
21 an automated sequencer Genetic Analyzer 3130 XL (Applied Biosystems).
22

23 **Viral tropism determination**

24 Electropherogram data were analyzed using the Sequencher 4.6 software. Sequences were
25 edited manually and a contig was generated based in two or more sequences, including
26 ambiguous nucleotides (mixtures). The quality control of sequences and preliminary subtyping
27 were performed by using web-based tools as NCBI locator, BLAST and genotyping, with
28 reevaluation of electropherogram whenever needed. The coreceptor usage of HIV-1 was
29

1 predicted from the V3 *env* region nucleotide sequence by Geno2pheno URL:
2 <http://coreceptor.bioinf.mpi-sb.mpg.de/cgi-bin/coreceptor.pl> The determination of tropism
3
4 was obtained through the submission of the sequences (fasta format) to geno2pheno using
5
6 clonal option. That relies on ranking of submitted sequences to curate data set. A False positive
7
8 rate (FPR), probability of classifying an R5-virus falsely as X4, is assigned to each submitted
9
10 sequence. These rates allowed both the evaluation of the likelihood of each sequence tropism
11
12 prediction as well as the patients' HIV infection population tropism prediction, using the lower
13
14 FPR from all sequences obtained, as recommended by the European consensus.⁷
15
16 Sequences were also evaluated at the PSSM_{X4R5} algorithm site [http://fortinbras.us/cgi-](http://fortinbras.us/cgi-bin/fssm/fssm.pl)
17
18 [bin/fssm/fssm.pl](http://fortinbras.us/cgi-bin/fssm/fssm.pl). This bioinformatics method scores V3 amino acid sequences or
19
20 nucleotide and predicts CXCR4 usage This score, determined on the basis of position-
21
22 specific scoring matrices [PSSM], can be interpreted as revealing a propensity to use
23
24 CXCR4 as follows: known R5 viruses had low scores, R5X4 viruses had intermediate
25
26 scores, and X4 viruses had high scores.
27
28
29
30
31

32 **Distance metrics**

33
34
35
36 Estimation of diversity at V3 of each sequence (intra-sequence) was evaluated as the number of
37
38 non synonymous ambiguities, nucleotide mixtures that would lead to codification of more than
39
40 one amino-acid (NSA). The diversity inter-sequences (between replicates) was quantified by
41
42 Protdist (Protein distance matrix - version 3.5c, BioEdit Sequence Alignment Editor), that
43
44 computes a distance matrix from protein sequences taking into account a variety of sequence
45
46 ambiguities; dispersion was quantified by comparing the paired standard deviation of the
47
48 sequence's FPR among replicates. The relationship of the sequences was further evaluated
49
50 with Neighbor Joining (NJ) phylogenetic tree, using the evolutionary model selected for the
51
52 dataset by modeltest TVM + I + G with PAUP software v4.0b.
53
54
55
56
57
58
59
60
61
62
63
64
65

Statistical analyses

Epi6 and excel were used for data storage. T tests, Wilcoxon signed rank test, linear regression and correlation (Spearman) were done with and Prism GraphPad5.0. Results are expressed as medians, level of significance of $p < 0.05$.

Results

The population characteristics are reported in Table 1. Of the 54 sample patients we have analyzed 28 single sequences paired (RNA/DNA) and 26 sequences paired in replicates (2 to 5).

The median number of NSA was similar in sequences originated from virions or provirus and both plasma and cell derived FPR_{clonal} had a negative correlation to the number of NSA ($p = 0.0009$ and $p < 0.0001$ respectively) supplementary material 1A and 1B. The paired plasma / cell lowest FPR_{clonal} value (used for prediction) showed a strong correlation ($p < 0.0001$, $r = 0.9979$) (Figure 1). However, there were discordances when the different cutoffs recommended for clinical use (geno2pheno 20%, 10% and 5.75%) are applied. Paired plasma/cell prediction for single sequences had a discordance of 10.7%, 10.7% and 7.2% and replicates 15.4%, 23%, 19.2%, respectively (Table 2). The prediction using replicates would lead to a change in tropism prediction in 42% of cell and in 23% of plasma predictions when each sequence replicate is considered individually and its FPR_{clonal} compared to the other replicated sequences. The use of PSSM also showed comparable plasma/cell discordance (14% for singles and 23% for replicates). When all replicate sequences' FPR are evaluated, the paired standard deviation of cell derived FPR is significantly greater than that of plasma derived sequences ($p = 0.0030$) (Figure 2). In samples with sequences replicates, the lowest value FPR_{clonal} , from both plasma and cell, showed a negative correlation, but not significant, with the average genetic distance ($p = 0.178$ plasma and $p = 0.051$ cell). However, the genetic distance was significantly higher in provirus as compared to that from virions ($p < 0.0001$) (Figure 3). The genetic distance from both cell and plasma replicates showed no correlation to CD4 counts or viral load (data not shown). The distance between the samples and the phylogenetic relationship of the strains was also evaluated with phylogeny methods, NJ (evolutionary model TVM + I + G) using the PAUP

1 software v4.0b 10, as can be seen in supplementary data. Cell and plasma derived sequences
2 from a same patient form clusters with high bootstrap, some of these clusters are exemplified in
3 supplementary material 2, but sequences from plasma or cell, as well as those predicted as X4 or
4 R5 tropism, did not shown a clear segregation in most cases.
5
6
7
8
9

10 **Discussion**

11
12 The ability to enter a permissive cell is a major step in viral life cycle. The V3 region of the
13 HIV-1 envelope not only contributes with important structural elements to viral entry but also
14 may define cell type specificity, as by differential chemokine coreceptor use. Tropism
15 determination using V3 sequencing analyzed at available bioinformatics tools, is practical and
16 generates a prediction that have been shown to be associated to CCR5 antagonists efficacy
17 (McGovern et al., 2010) and to the pathogenic potential of virus, being X4 variants associated
18 to disease progression (Regoes and Bonhoeffer., 2005; Waters et al., 2008; Lanca et al., 2012).
19 When tropism test is used to indicate the use of CCR5 antagonists, the sensitivity to detect a X4
20 variant becomes a critical issue, as this class of antiretroviral is only indicated for patients
21 harboring only R5 tropic viruses. The use of single sequence for prediction have been
22 challenged by studies showing a better accuracy of triplicate as compared to single sequences
23 using the geno2pheno coreceptor algorithm for predicting HIV tropism (Swenson et al., 2010)
24 and more importantly from the clinical correlation of tropism prediction using triplicates in
25 therapy-experienced patients, where a re-analysis of clinical trials samples demonstrated that
26 replicate genotypic tropism prediction better correlate to clinical response to maraviroc-
27 containing regimens (McGovern et al., 2010). The fact that single *pol* sequence has provided
28 valuable antiretroviral resistance information in guiding salvage therapy may be cited as an
29 argument for the redundancy of replicate sequence analysis. Differences in selective pressure
30 on the viral envelope protein compared with *pol*, which are reflected by nine-fold higher
31 nucleoside substitutions/site/year in *env* (Hahn at al., 1986) may suggest that the situation of
32 resistance test might not apply directly to tropism prediction, but the issue is far from settled. In
33 the absence of a direct comparison of single and replicate test procedures in clinical practice, the
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 recently formulated European guidelines advise triplicate testing with false-positive rate
2 (FPR_{clonal}) of 10%. If single sequence is used then a more conservative FPR_{clonal} of 20% for
3
4 RNA samples, especially with lower viral load, and for DNA sample is recommended
5
6 (Vandekerckhove et al., 2011). Another issue that distinguished resistance of genotype test to
7
8 tropism prediction is the fact that, contrary to resistance test, provirus cell DNA seems to
9
10 provide adequate prediction and some national networks have been structured based on provirus
11
12 prediction (Svicher et.al., 2012). Some articles (Verhofsted et al., 2009) but not others
13
14 (Raymond et al., 2008) have suggested that cell prediction of X4 viruses is more common than a
15
16 plasma prediction. This may be due to variables as stage of disease and methods of analysis.
17
18
19

20 We hypothesized that replicate would diminished or eliminate cell-plasma prediction
21
22 discrepancies. Our findings did not support this assumption, and the use of replicate was
23
24 actually associated to some increase the discrepancy at the three most used cutoff values
25
26 (5.75%, 10% and 20%). Phylogenetic trees generated using NJ method did not help in the
27
28 understanding this differences. Cell and plasma derived sequences, as well as X4 and R5
29
30 predicted sequences, did not show any clear clustering for most samples. Even considering
31
32 these differences, there is a strong correlation of viral tropism prediction, using the
33
34 recommended lower FPR value, between both compartments (Figure 1). Further analysis of
35
36 sequences with PSSM was used to evaluate if this discordance would also be present, and they
37
38 are similarly observed. Some parameters, as CD4 or viral load, did not show an association to
39
40 the prediction. However, number of ambiguities in each sequence (Supplementary material 1A
41
42 and 1B), as well as inter sequence distance (Figures 3 and 2), or the dispersion (standard
43
44 deviation) in the FPR obtained with replicates, seem to be more important among provirus
45
46 DNA. This could provide the basis of the increase number of X4 in cell prediction, and could
47
48 mean that a more diverse population is sequenced from cell DNA. Sample input may have an
49
50 impact in these differences. We used a relatively high input (2.5 uL) of extracted genetic
51
52 material, but studies that assess different sample input may help to clarify this issue. The
53
54 presence of defective genomes may have a complicating role in this interpretation, as
55
56 nonfunctional genomes may be included in such predictions. On the other hand, defective
57
58
59
60
61
62
63
64
65

1 genomes may provide indication of related, viable virus, which may be of relevance. Our work
2 does not address this issue. One important observation is the rather large percentage of change
3 in prediction when cell or plasma as considered, up to 12% for plasma prediction and 48% for
4 cell prediction. Single sequencing is preferred for many as it is cheaper. Our study support the
5 notion the replicate evaluation may improve non-R5 prediction sensitivity, is applicable to
6 clinical laboratories even in resource constrained environments and provides a broader
7 observation of the viral variability characteristics of chronic HIV-1 infection.
8
9
10
11
12
13
14
15
16
17
18
19

20 **Funding**

21
22 This work was supported by Foundation of Support to Research the State of Sao Paulo -
23 FAPESP 2009/08215-0 and FAPESP 2011/21958-2.
24
25
26

27 **Transparency declarations**

28 The authors have perceived no conflict of interest to declare.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 **References**
4
5

6 Berger EA, Doms RW, Fenyo EM et al. A new classification for HIV-1. *Nature* 1998; 391: 240.

7
8
9 Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral
10 entry, tropism, and disease. *Annu Rev Immunol* 1999; 17: 657–700.
11

12
13 Dimonte S, Mercurio F, Svicher V, D'Arrigo R, Perno CF, Ceccherini-Silberstein F. Selected
14 amino acid mutations in HIV-1 B subtype gp41 are associated with specific gp120v₃ signatures
15 in the regulation of co-receptor usage. *Retrovirology* 2011, 8-33.
16

17
18 Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Ver* 2012;
19 76:159-216.
20

21
22 Ferreira JLP, Coelho LPO, Rodrigues R et al. Evaluation of genotypic prediction of HIV-1
23 tropism using population sequencing of replicates. *Journal of Virological Methods* 2012; 179:
24 325–329.
25

26 Frange P, Galimand J, Goujard C et al. High frequency of X4/ DM-tropic viruses in PBMC
27 samples from patients with primary HIV-1 subtype-B infection in 1996–2007: the French
28 ANRS CO06 PRIMO Cohort Study. *J Antimicrobial Chemother* 2009; 64: 135–141.
29

30
31 Hahn BH, Shaw GM, Taylor ME et al. Genetic variation in HTLV-III/ LAV over time in
32 patients with AIDS or at risk for AIDS. *Science* 1986; 232: 1548–1553.
33

34 Harrigan PR, Geretti AM. Genotypic tropism testing: evidence based or a leap of faith?. *AIDS*
35 2010; 25: 257–264.
36

37
38 Lanca AM, Collares JK, Ferreira JL et al. HIV-1 tropism and CD4 T lymphocyte recovery in a
39 prospective cohort of patients initiating HAART in Ribeirão Preto, Brazil. *Mem Inst Oswaldo*
40 *Cruz* 2012; 107: 96-101.
41

42 MacArthur R.D, Novak R.M. Reviews of antiinfective agents: maraviroc: the first of a new
43 class of antiretroviral agents. *Clin. Infect. Dis* 2008; 47: 236- 241.
44

45 McGovern RA, Thielen A, Mo T et al. Population-based V3 genotypic tropism assay: a
46 retrospective analysis using screening sample from the A4001029 and MOTIVATE studies.
47 *Aids* 2010; 24: 2517-2525.
48

49
50 Raymond S, Delobela P, Mavignera M et al. Correlation between genotypic predictions based
51 on V3 sequences and phenotypic determination of HIV-1 tropism. *AIDS* 2008; 22: F11-F16.
52

53 Regoes R.R, Bonhoeffer S. The HIV coreceptor switch: a population dynamical perspective.
54 *Trends. Microbiol* 2005; 13: 269-277.
55

56
57 Soulie C, Marcelin AG, Ghosn J et al. HIV-1 X4/R5 coreceptor in viral reservoir during
58 suppressive HAART. *AIDS* 2007; 21: 2243–2245.
59
60
61
62
63
64
65

1 Svicher V, Alteri C, Montano M et al. Performance of genotypic tropism testing on proviral
2 DNA in clinical practice: results from the DIVA Study Group. *New Microbiologia* 2012; 35:
3 17-25.

4 Swenson LC, Knapp D, Harrigan PR. Calibration and accuracy of the geno2pheno coreceptor
5 algorithm for predicting HIV tropism for single and triplicate measurements of V3 genotype.
6 2010. Tenth International Congress on Drug Therapy in HIV Infection Glasgow. Abstract O122.
7 Trouplin V, Salvatori F, Cappello F et al. Determination of coreceptor usage of human
8 immunodeficiency virus type 1 from patient plasma samples by using a recombinant phenotypic
9 assay. *J Virol* 2001; 75: 251-259.

10
11
12 Thielen A, Lengauer T, Swenson LC et al. Mutations in gp41 are correlated with coreceptor
13 tropism but do not improve prediction methods substantially. *Antiviral Therapy* 2011; 16:319-
14 328.

15
16 Vandekerckhove L, Wensing A, Kaiser R et al. European guidelines on the clinical
17 management of HIV-1 tropism testing. *The Lancet infectious diseases* 2011; 11: 394-407.

18
19
20 Vandekerckhove L, Wensing AMJ, Kaiser R, Brun-Vezinet F, Clotet B, De Luca A. European
21 consensus on clinical use and interpretation of HIV-1 tropism testing. *HIV Med* 2009; 10: 48-
22 49.

23
24 Verhofstede C, Vandekerckhove L, Van Eygen V et al. CXCR4-using HIV type 1 variants are
25 more commonly found in peripheral blood mononuclear cell DNA than in plasmaRNA. *J*
26 *Acquir Immune Defic Syndr.* 2009; 50: 126-136.

27
28
29 Waters L, Mandalia S, Randell P et al. The impact of HIV tropism on decreases in CD4 cell
30 count, clinical progression, and subsequent response to a first antiretroviral therapy regimen.
31 *Clin Infect Dis* 2008; 46: 1617-1623.

32
33 Whitcomb JM, Huang W, Fransen S et al. Development and characterization of a novel single-
34 cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor
35 tropism. *Antimicrob Agents Chemother* 2007; 51: 566-575.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

FIGURE LEGENDS

FIGURE 1: Correlation (Spearman) of the plasma and cell derived lowest value FPR_{clonal} value, used for patients' Tropism prediction.

FIGURE 2: Wilcoxon signed-rank test of the standard deviation of paired plasma and cell derived replicate sequences' FPR_{clonal} .

FIGURE 3: Wilcoxon signed-rank test of the paired cell/plasma genetic distance (Protdist).

SUPPLEMENTARY 1 (1A and 1B): Correlation (Spearman) of sequences' FPR_{clonal} from plasma (1A) and cell (1B) with the number of amino acids at codons with nucleotide ambiguities (non synonymous nucleotide ambiguities).

SUPPLEMENTARY 2: Neighbour Joining phylogenetic tree reconstructed using Paup*. Sequences obtained from proviral DNA are (red) and viral RNA (green) along with references (no color). Colored groups depicts each patient sequence cluster, however colors are not unique and therefore do not represent any correlation between distant branches or different patient samples. OSEMP1300 sequence was used as outgroup

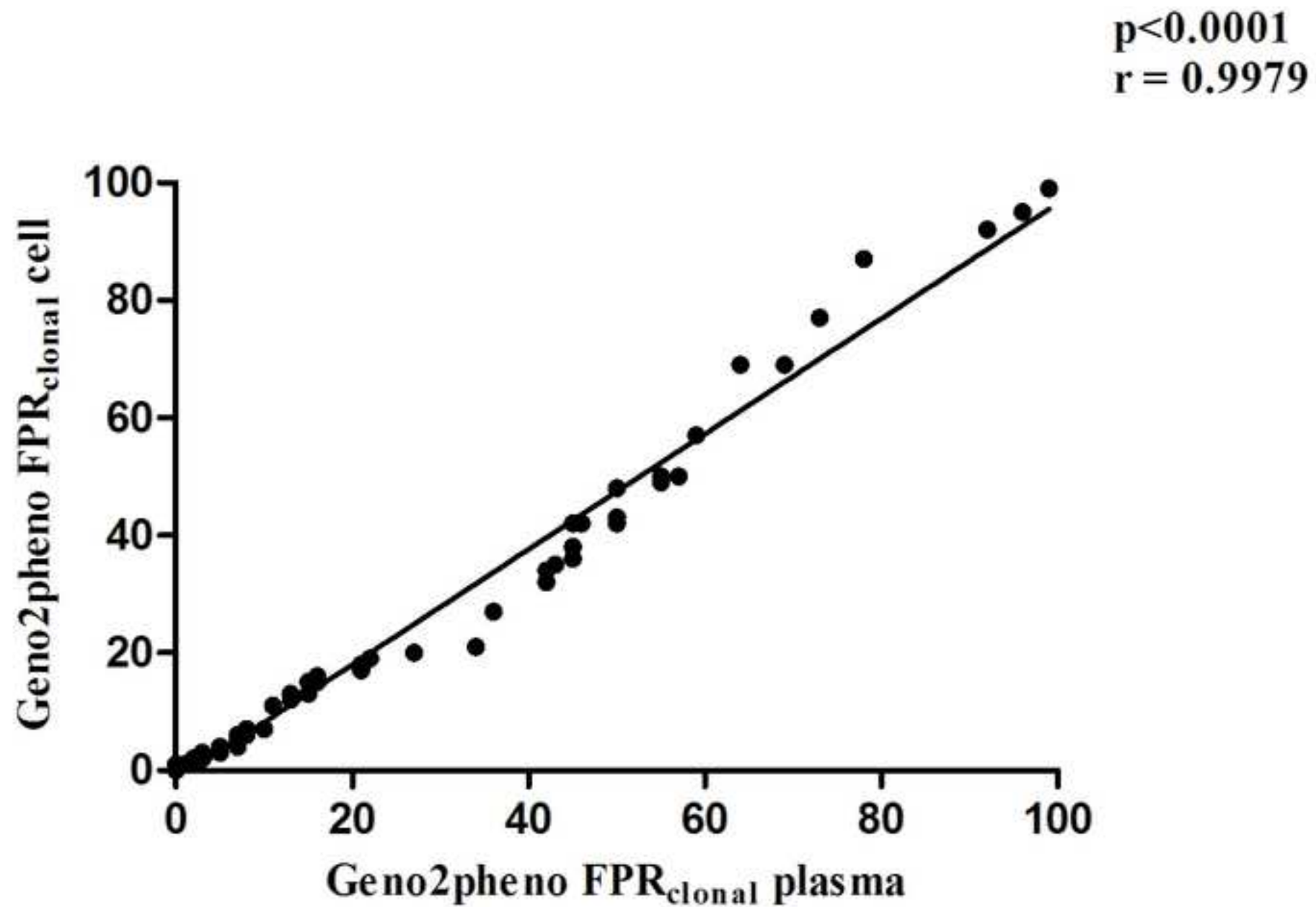
Table 1. Patients' demographic characteristics according to the number of sequences obtained (single sequence or more than one sequence).

	Single	Replicates
Age (years), median (IQR)	45 (36-48)	40 (19-45)
Gender (% males)	71%	73%
CD4 at collection (cells/mm ³), median (IQR)	245 (134-430)	337 (186-483)
Nadir CD4 (cells/mm ³), median (IQR)	75 (24-185)	107 (39-161)
RNA viral load (log ₁₀ copies/mL), median (IQR)	3.8 (3.36-4.73)	4.2 (3.84-4.61)
Number of ARV regimens (IQR)	8 (4-10)	8 (5-10)

Table 2. Number of tropism predictions discordances between plasma RNA and cell provirus DNA in single and replicates sequences.

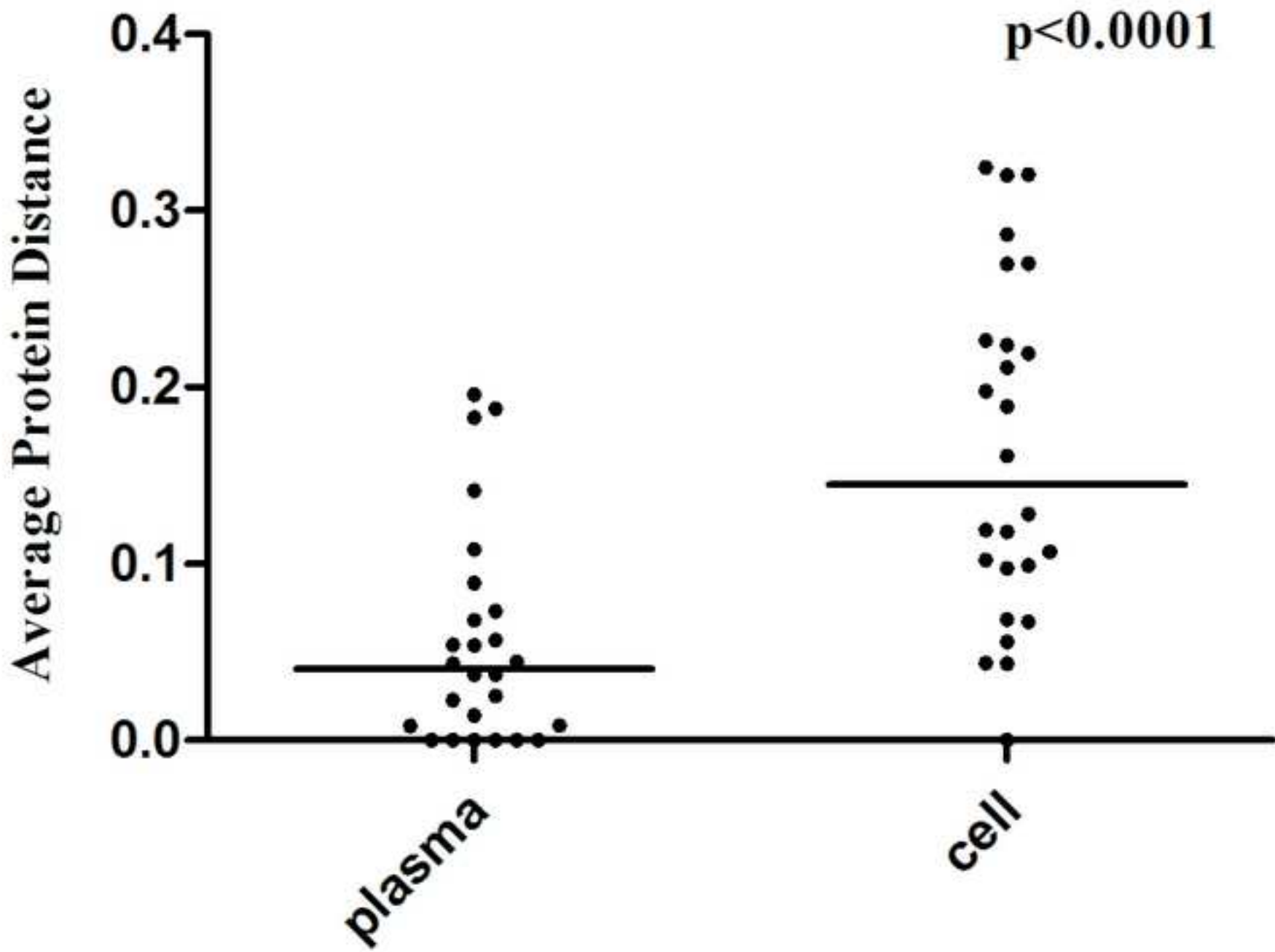
	Single	Discordances (%)	Replicates	Discordances (%)
Geno2pheno ₂₀	3/28	10.7%	2/26	15.4%
	n=3 pR5/cX4		n=2 pX4/cR5 n=2 pR5/cX4	
Geno2pheno ₁₀	3/28	10.7%	6/26	23%
	n=2 pR5/cX4		n=3 pX4/cR5	
	n=1 pX4/cR5		n=3 pX4/cR5	
Motivate _{5,75}	2/28	7.2%	5/26	19.2%
	n=1 pR5/cX4		n=1 pX4/cR5	
	n=1 pX4/cR5		n=4 pR5/cX4	

c, cell
p, plasma



Figure(s)

[Click here to download high resolution image](#)



Supplementary Materials

[Click here to download Supplementary Materials: Supplementary1A.tif](#)

Supplementary Materials

[Click here to download Supplementary Materials: Supplementary 1B.tif](#)

Supplementary Materials

[Click here to download Supplementary Materials: Supplementary 2.doc](#)