Characterization of HIV-1 near full-length proviral genome quasispecies from patients with undetectable viral load undergoing first-line HAART therapy

Brunna M. Alves¹, Juliana D. Siqueira¹, Marianne M. Garrido², Ornella M. Botelho¹, Isabel M. Prellwitz¹, Sayonara R. Ribeiro², Esmeralda A. Soares¹ and Marcelo A. Soares¹,³*

1 Programa de Oncovirologia, Instituto Nacional de Câncer, Rio de Janeiro, Brazil;
2 Serviço de Doenças Infecciosas, Hospital Federal de Ipanema, Rio de Janeiro, Brazil;
3 Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
* Correspondence: masoares@inca.gov.br; Tel.: +55-21-3207-6591

Abstract: Increased access to highly active antiretroviral therapy (HAART) by HIV+ individuals has become a reality worldwide. In Brazil, ART currently reaches over half of the HIV-infected subjects. In the context of a remarkable HIV-1 genetic variability, highly related variants, called quasispecies, are generated. HIV quasispecies generated during infection can influence virus persistence and pathogenicity, representing a challenge to treatment. However, the clinical relevance of minority quasispecies is still uncertain. For this study, we have determined the archived proviral sequences, viral subtype and drug resistance mutations from a cohort of HIV+ patients with undetectable viral load undergoing HAART as first-line therapy using next-generation sequencing for near full-length virus genome (NFLG) assembly. HIV-1 consensus sequences representing NFLG were obtained for eleven patients, while for another twelve varying genome coverage rates were obtained. Phylogenetic analysis showed the predominance of subtype B (83%; 19/23). Considering the minority variants, 18 patients carried archived virus harboring at least one mutation conferring antiretroviral resistance; for six patients, the mutations correlated with the current ARVs used. These data highlight the importance of monitoring HIV minority drug resistant variants and their clinical impact, to guide future regimen switches and improve HIV treatment success.

Keywords: HIV-1; quasispecies; minority resistance mutations; HAART; drug resistance; undetectable viral load.

1. Introduction

According to UNAIDS, approximately 36.7 million people were living with the human immunodeficiency virus (HIV) worldwide at the end of 2016, making HIV infection a major public health problem [1]. One of the factors related to the increased number of people living with HIV is the greater access to highly active antiretroviral therapy (HAART), which is strongly associated with an increase in the expectancy and quality of life of this population. By 2016, 19.5 million people had access to HAART, an increase of 6% over 2015 [2]. In Brazil, it has been estimated that about 830,000 individuals were living with HIV/AIDS by the end of 2016, representing an HIV prevalence rate of 0.24%. The access to HAART reached approximately 490,000, which represents more than half of the estimated Brazilians HIV-positive individuals. Of this total, approximately 450,000 had an undetectable viral load at least six months after HAART initiation, one indicator of therapeutic success [1]. According to the Brazilian Ministry of Health, virologic failure occurs when HAART fails to suppress and sustain a person’s undetectable viral load after six months of initiating or modifying treatment, in addition to the detection of viral load in patients who were previously undetectable [3]. Brazil has been cited as a reference in access to HAART since November 1996, when the government guaranteed universal and free access to therapy to the Brazilian HIV-
population. New recommendations that stimulate the initiation of HAART for all HIV-positive individuals, independent of CD4+ T lymphocyte counts, were implemented in October 2013 in order to reduce the transmission of the virus [4].

A remarkable HIV-1 genetic variability results from mutational events associated with the high error-prone rate of the viral reverse transcriptase (RT), high virus replication rates and homologous recombination events. HIV-1 diversity imposes an important clinical challenge, as it allows the virus to adapt to and evade immune responses and antiretroviral therapy [5], therefore influencing diagnosis and treatment [6]. In the context of continuous HIV-1 genetic variability within each individual, several highly related but genetically distinct variants are generated and referred to as viral quasispecies [7]. The quasispecies heterogeneity associated with the selective pressure exerted by the immune system will influence virus persistence and pathogenicity, allowing the adaptation of the quasispecies through intrahost persistence and the ability to outgrow other less adapted variants [8].

The HIV evolutionary dynamics associated with ART pressure allows the appearance of drug-resistant variants [9, 10]. Antiretroviral resistance is an important public health concern since it limits therapeutic options, causes treatment failure and can be transmitted, compromising future treatment option in untreated individuals [11, 12]. Approximately 10% of newly diagnosed patients are infected with strains that have at least one transmitted drug-resistance mutation (TDRM) [13].

Aiming to establish more efficient ART regimens, genotypic assays to detect drug resistance-associated mutations have been widely used in the clinical setting [14]. This practice has lately been greatly benefited by the use of Next-Generation Sequencing (NGS), which provides a large data volume in a cost-effective and highly sensitive way. Of note, NGS allows the detection of HIV minority variants, previously undetectable by Sanger sequencing, which has a detection limit of 10-25% frequency in the viral population [15-17]. Recent studies using NGS for HIV-1 sequencing were able to detect minority variants below 1% in frequency in the viral population, allowing the identification of drug-resistant minority variants, the study of transmitted resistant viruses and the impact of those minority variants on treatment efficacy [18-22]. Moreover, HIV-1 sequencing by NGS significantly contributes to the analysis of viral genetic diversity, evolutionary and epidemic processes, since near full-length genomes (NFLG) can be obtained [23]. These genomes contribute to the growing interest in tests that simultaneously probe multiple genomic regions that are targeted by antiretroviral drugs acting on different steps of the virus replicative cycle [24].

HIV minor drug resistant variants may persist in the infected individual [25, 26]. Despite their low replicative capacity due to the presence of transmitted drug-resistance mutations (TDRM), they may persist as archived proviruses in PBMCs for several years and may have a long-term impact on therapeutic response [27-29]. However, the clinical significance of these minority drug resistant variants is still uncertain, as their role in the future response to treatment is not fully understood. Some studies have shown the association of such variants with increased risk of treatment failure in treated patients, as well as in patients with no previous ART history [18, 30-35]. On the other hand, different studies did not find any influence of these minority resistant variants on treatment response [36-42]. The association between HIV minor non-nucleoside reverse transcriptase inhibitors (NNRTI) resistant variants and a worse treatment prognosis has been described [22, 26, 43-45]. Patients carrying those variants had a three-fold higher risk of treatment failure compared to patients without NNRTI-resistance mutations when subjected to therapeutic regimens based on this ARV class [46, 47]. With respect to patients undergoing therapeutic success, a single study was conducted that showed the presence of minority drug-resistance mutations in five out of eleven patients and a large variability of the archived proviral epitopes [48]. This study highlights the importance of more sensitive genotypic resistance tests and further studies on the influence of these mutations on the outcome of HAART, especially in patients undergoing therapeutic success, since this cohort is still scarcely discussed.

In the present study, we have assessed the presence of HIV minority drug-resistant variants archived in PBMC samples in a Brazilian cohort of chronic HIV patients undergoing HAART as first-line therapy and undetectable viral load. Upon analyzing the HIV antiretroviral resistance
profiles of the patients by NGS, our study was able to evidence a high prevalence of drug-resistance mutations in this cohort, despite the therapeutic success achieved by their carriers. This is the first study investigating the presence of HIV minority drug-resistance mutations among Brazilian patients under virologic control.

2. Materials and Methods

2.1. Study Population and Sample Collection

A cross-sectional study was carried out among patients attending a sexually transmitted diseases/ HIV ambulatory at Ipanema’s Federal Hospital, Rio de Janeiro, Brazil, between February and July 2016. These patients were recruited during the clinical follow-up routine and had a 10 ml sample of whole peripheral blood collected. A questionnaire was applied to collect epidemiological (date of birth, sex, risk behavior) and clinical (date of HIV diagnosis, HIV-1 viral load, CD4+ T-cell counts, CD8+ T-cell counts, treatment history) data. Written informed consent was obtained from all participants and data were processed using unique identifiers to ensure confidentiality.

Inclusion criteria included age equal or greater than 18 years, being under first-line HAART and having undetectable HIV viral load for at least 12 months prior to collection date. Patients who had a history of previous virological failure, underwent a change in the therapeutic regimen due to intolerance or poor adherence, that were classified into clinical and/or immunological AIDS (had an AIDS defining disease or CD4+ counts equal or less than 200 cels/mm³, according to CDC criteria, 2014) [49] or that were under follow-up at the reference center for less than 12 months were excluded from the analysis. This research was approved by the Ethics Committees in Research of the Brazilian National Cancer Institute – INCA and of Ipanema’s Federal Hospital (CAAE 52862016.9.0000.5274, approved on 26 March 2016).

2.2. DNA Extraction and PCR of proviral DNA

Plasma and buffy coat were separated by centrifugation of the whole blood, and the latter was used for genomic DNA extraction with the Genomic DNA Extraction Kit (Real Genomics, BioAmerica, Inc.) following manufacturer’s specifications.

Nested PCR was carried out for the amplification of near full-length HIV genome using a set of five overlapping fragments, of approximately 2 kb each, or alternatively four fragments, with ~3 kb each, as previously described [50, 51]. Each fragment overlapped the adjacents by an average of 400 bp (minimum 89 bp, maximum 555bp). All reactions were performed in a Veriti® 96 Well Thermal Cycler (Life Technologies, Carlsbad, U.S.A.) using Platinum™ Taq DNA Polymerase High Fidelity (Life Technologies) in a final volume of 25 µL. An overview of all PCR fragment coverage regions and list of primers used are provided in Supplementary Table S1.

PCR products were visualized in 1% agarose gels. Duplicates were made for each PCR and pooled before proceeding to the purification step with the GFXTM PCR DNA and Gel Band Purification Kit (GE Healthcare, Massachusetts, USA). The purified fragments were quantified in a NanoDrop ND 1000 apparatus (Thermo Scientific, Massachusetts, USA) and diluted to 4 ng/µL before pooled per sample. The final product pool per each sample was diluted to 0.4 ng/µL for library construction.

2.3. Library Construction and NGS

Libraries were prepared using the Nextera XT DNA Sample Preparation kit (Illumina Inc., SanDiego, USA) according to the manufacturer’s protocol. Briefly, after the fragmentation step using transposon technology, the libraries were subjected to a PCR where they were tagmentated with Illumina sequencing adaptors and molecular tags (indexes) to identify each sample. A purification step was performed to select fragments of ~800bp and libraries were quantified by qPCR with the
KAPA library quantification kit (Kapa Biosystems, Massachusetts, USA). Individual libraries were then diluted to 4 nM, considering the mean size of each library and the quantification performed, and pooled. The final product was diluted to 12 pM and sequenced in a MiSeq Illumina platform (2 x 301 paired-end run) (Illumina) with 1% denatured PhiX DNA as a sequencing control.

2.4. Data Analysis

The analysis of the obtained files was performed in Geneious v.9.1.3 as described by Dudley et al. [52]. Briefly, the two fastq files generated per sample after the demultiplexing process were paired and trimmed at the ends to an error rate below 0.1%. The products were then used in the assembly of the viral genome sequence using an annotated HIV-1 HXB2 reference sequence made available by Dudley and colleagues, including information on resistance mutation-associated positions, genes, CDS and mature peptide positions (https://dholk.primate.wisc.edu/wiki/dho/public/page.view?name=default). The alignment parameters used in the assembly were as previously described [52]. For this alignment, ten iterations were used, where the first of the ten alignments was performed against the annotated reference sequence and the remaining nine used the consensus obtained in the previous step as reference, in order to reduce the influence of the reference used in the alignment product.

2.4.1. Analysis of resistance mutations

The annotated drug resistance mutations were based on the consensus of the International Antiviral Society [53] and the Stanford HIV Drug Resistance Database (available at http://hivdb.stanford.edu/hiv/). Transmitted drug resistance mutations (TDRM) were also included and defined according to the classification of the World Health Organization established by Bennett et al. [11] and in the TDRM database of the Stanford HIV Database. The HXB2 reference annotation was manually updated to include new positions at the lists mentioned above and resistance mutations described in the literature at the C-terminal region of RT covering the connection (CN) and RNase H (RH) subdomains [54-61] The mutations evaluated in the RT C-terminal region are listed in detail in Table 1.

The variant finder of Geneious v.9.1.3 was used to call nucleotide variants from the reference sequence at frequency higher than 1%. This tool evaluates nucleotide substitutions found in the alignment along with their frequency and whether they are synonymous or non-synonymous. The frequency and the number of reads representing each ARV-resistant variant was determined in relation to the annotated reference and exported to an Excel file. Variants between 20 and 1% of frequency were considered as minor variants, since 20% is the minimum frequency associated with detection of mutations by commercial genotype resistance assays that use Sanger sequencing [62].

2.4.2. Phylogenetic analysis

A consensus sequence was derived for each sample from the reference assembly with Geneious using the 50% stringency setting. HIV-1 subtype classification of each query sequence was inferred through phylogenetic analysis performed with the maximum likelihood (ML) method using PhyML v.3.0 [63] and the best model of nucleotide substitution was inferred with Model Generator [64]. Sequences suggestive of intersubtype recombination were further analyzed with the bootscanning tool of Simplot v.3.5.1 [65] for determining patterns of recombination and the HIV-1 subtypes involved in the recombination event. The following parameters were used: window = 400 pb; steps = 40 pb; T/t = 2.0; gapstrip = on; replicas = 100; nucleotide substitution model = F84; method = Maximum Likelihood. Recombinant strains were further confirmed by phylogenetic analysis of individual HIV-1 subtype genomic fragments as suggested by the bootscanning breakpoint analysis (data not shown).
**Table 1.** Mutations analyzed in the C-terminal domains of reverse transcriptase covering the connection (CN) and RNase H (RH) subdomains and the respective classes of antiretrovirals associated with resistance as described in the literature*

<table>
<thead>
<tr>
<th>Subdomain</th>
<th>Mutation</th>
<th>ARV Class Associated with Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>E312Q</td>
<td>NRTI</td>
</tr>
<tr>
<td></td>
<td>Y318F/W</td>
<td>NNRTI</td>
</tr>
<tr>
<td></td>
<td>G335D/C</td>
<td>NRTI (polymorphism)</td>
</tr>
<tr>
<td></td>
<td>N348I</td>
<td>NRTI and NNRTI</td>
</tr>
<tr>
<td></td>
<td>A360I/V</td>
<td>NRTI</td>
</tr>
<tr>
<td></td>
<td>V365I</td>
<td>NRTI</td>
</tr>
<tr>
<td></td>
<td>T369I/V</td>
<td>NRTI and NNRTI</td>
</tr>
<tr>
<td></td>
<td>A371V</td>
<td>NRTI</td>
</tr>
<tr>
<td></td>
<td>A376S</td>
<td>NRTI and NNRTI</td>
</tr>
<tr>
<td></td>
<td>E399D/G</td>
<td>NRTI and NNRTI</td>
</tr>
<tr>
<td></td>
<td>A400T</td>
<td>NRTI</td>
</tr>
<tr>
<td></td>
<td>N348I</td>
<td>NRTI and NNRTI</td>
</tr>
<tr>
<td>RH</td>
<td>Q509L</td>
<td>NRTI and NNRTI</td>
</tr>
<tr>
<td></td>
<td>Q547K</td>
<td>NRTI</td>
</tr>
</tbody>
</table>

NRTI, nucleotide/nucleoside reverse transcriptase inhibitors; NNRTI, non-nucleoside reverse transcriptase inhibitors.

* References [54-61].

**2.4.3. Analysis of viral tropism**

For the phenotypic prediction of HIV co-receptor usage, the reads spanning the complete region of V3 loop of the env gene were selected and unique haplotypes were identified and their relative frequency considered. Amino acid sequences of the haplotypes were analyzed only if started and ended with cysteine residues and contained 32-38 amino acid residues, a pattern consistent with functional sequences [66]. Valid haplotypes were analyzed with the Geno2Pheno algorithm (available at http://coreceptor.geno2pheno.org) [67] using a false positive rate cut-off of 10% for classification as R5- or non-R5-using viruses, as recommended by the European Consensus Group on Clinical Management of HIV-1 Tropism Testing and used in several studies [51, 68-71].

**3. Results**

The median age of the patients included in the study was 38 years and 75% were male (Table 2). The median time of antiretroviral treatment at enrollment was approximately three years and the median baseline CD4+ T-cell count was 712.5 cells/mm³. All patients have reported being infected by sexual transmission, 56% of which were men who have sex with men. Most patients (19; 59%) were under HAART composed of tenofovir (TDF), lamivudine (3TC) and efavirenz (EFV) at the time of sample collection.

**Table 2.** Demographic and clinical characteristics of the 32 HIV-positive participants of the study
Nine samples (28%) failed to have more than one virus DNA fragment PCR-amplified and were excluded from further analyses, while 23 samples amplified at least two fragments and were sequenced using NGS. Of the latter, eleven samples had the near full-length genome (NFLG) sequenced, and the remaining samples were missing one or two DNA fragments over the genome. The average number of reads obtained per sample were 774,536 (374,780 – 1,647,090). After assembling, the average coverage per nucleotide position was 7,193 and it was homogeneous over the regions sequenced. The Gag CDS was complete for 22 samples (96%), the Pol CDS for 15 (65%) and the Env CDS for 17 (75%).

Most samples (18; 78%) carried archived viruses harboring at least one mutation associated with antiretroviral resistance (Table 3). Among them, 13 samples had mutations in the RT region, including the connection domain. Four carried resistance only to NRTI, six only to NNRTI, while three presented resistance mutations to both NRTI and NNRTI. The ART regimen of these 13 patients included two NRTI and one NNRTI. One sample (patient #28) presented four thymidine analogue-associated mutations (TAMs), with frequency varying from 38.7 to 99.5%, that are associated with intermediate resistance to TDF, part of the patient’s current therapy regimen. Three samples presented viruses harboring the M184V mutation (frequency varying from 1.3 to 8%) that confers high-level resistance to 3TC, used by those patients. However, this mutation is known to increase the susceptibility to zidovudine (ZDV) and TDF, that were also part of their ART regimen. Two samples presented the E399G mutation with frequency 1.5 of and 4.9%. This mutation is associated with resistance to EFV which was included in these patients’ therapy. Three samples presented major resistance mutations in the protease sequence, and while two of them used protease inhibitors (PI), the mutations observed were not related to the PI used. Resistance mutations in the integrase region were found in five samples. Most of them were minority variants, present in frequency below 11%, while one (T66I) was found with frequency of 31% in sample 29. Only one mutation associated with resistance to entry inhibitors was found in the env gene, with a frequency of 3% in sample 13.

Follow-up data of the patients under study (i.e. HIV viral load and CD4+ T-cell counts) were obtained during the preparation of this report to evaluate the clinical progress of the viral infection. The time information varied between 9 and 17 months after the initial collection time. Median baseline CD4+ T-cell counts were 758 cells/mm³ (n = 31; IQR675-963) and median baseline CD8+ T-cell counts were 867.5 cells/mm³ (n = 25; IQR657.75-1035), higher compared to the values observed at the time of patient inclusion. With respect to HIV viral load (n = 31), only one patient presented a detectable load (patient #8; 222 copies/ml) approximately one year after inclusion in the study. There was no change in the antiretroviral regimen of any patient, except for patients using LPV/r, who were switched to DRV/r due to changes in first-line regimens recommended by the Brazilian Ministry of Health.
The HIV-1 consensus sequence was generated for each sample for phylogenetic analysis. Most NFLG obtained were from HIV-1 subtype B (9; 82%; Fig. 1). The two remaining were classified as distinct unique recombinant forms (URF) comprising subtypes B and F1 based on Simplot analysis (Fig. 2). The sequences from the twelve non-complete genomes were predominantly of subtype B (n = 10), and two URF were found, one URF-BF1 and one URF-BC (Fig. 2).

Figure 1. Phylogenetic maximum likelihood analysis of HIV near full-length genomes obtained in this study. The analysis was conducted with 1000 bootstrap iterations and included eleven HIV-1 proviral sequences from Ipanema’s Federal Hospital, Rio de Janeiro (represented in bold) and references from HIV-1 subtype (represented by the subtype, country, year and GenBank accession number). Only bootstrap values greater than 0.7 are shown. The gray box highlight the sequences classified as subtype B, among which nine of the eleven sequences were placed. The two remaining sequences represented unique recombinant forms comprising subtype B and other subtypes.

Virus envelope tropism analysis was performed with the reads spanning the V3 env region of the sequences and the results are depicted in Table 3. All but one sample contained the V3 loop region sequenced and were analyzed. We found that 17 (77%) patients showed CCR5 usage as the major tropism profile (R5), while 5 patients (23%) presented predominantly the CXCR4 tropism profile (X4 and/or R5X4).
Figure 2. Phylogenetic classification of recombinant viruses considering the phylogeny and similarity analyses. The gray shading patterns represent the different subtypes: black for subtype B, gray for subtype F1 and white for subtype C. Sample IDs are represented at the left of each virus structure, which is in-scale relative to the genomic coordinates of the reference HXB2 genome at the top.

4. Discussion

Several viral factors are known to affect and modulate progression of HIV-1-related AIDS from the early asymptomatic phase of the disease, like the virus genetic diversity, viral fitness and co-receptor tropism [72-76]. In this context, we analyzed the archived HIV-1 proviral sequences from HIV patients in the early chronic phase of infection under HAART and with undetectable HIV viral load through NGS. This type of population is poorly explored in high-throughput studies, despite being of great relevance due the increased access to HAART worldwide. In the present report, we describe the genetic variability, prevalence of drug resistance-associated mutations and viral tropism of those patients at an ultra-deep level. It is noteworthy that this is the first study investigating the presence of minority drug resistance mutation among Brazilian patients under virologic control.

The majority of our patients self-reported as MSM (56%), the median time to start the ART after the HIV diagnostic was about 1.2 year and they have been successfully treated for approximately three years. With the new recommendations of Brazilian Ministry of Health implemented in the end of 2013, it is expected that the time lapse between HIV diagnosis and HAART initiation approaches zero, which contributes to reducing virus transmission. On the other hand, the test-and-treat approach highlights the importance of monitoring the prevalence of antiretroviral resistance and TDRM at population level, to avoid resistance dissemination and therapeutic failures due to TDR.

As seen in Latin America and the Caribbean countries, our study found a prevalence of subtype B infections (19; 83%) [77-79]. Although subtype B is the major HIV-1 genetic clade circulating in the Brazilian epidemic, the overall prevalence of non-B strains, like URF BF1, URF BC, and particularly of subtype C and CRF31_BC in the South of Brazil, is increasing [80-89]. None of the BF1 (3, 13%) or the BC recombinant structure (1, 4%) identified in this study showed similarity with the recombination patterns of known CRFs or other recombinants. CRFs were also not detected among HIV-1 NFLG obtained from children and adolescents from São Paulo, Brazil, which corroborate the hypothesis that novel recombinants are continually arising at the Brazilian HIV-1 epidemic scenario [86]. The analysis of HIV-1 NFLG, as opposed to most previous Brazilian HIV subtype studies that analyzed HIV sequence information only from shorter fragments, may unveil an underestimation of the occurrence of recombinant viruses in the country.
Table 3. Distribution of the antiretroviral resistance mutations and major envelope tropism found across the 23 HIV-1 genome sequences analyzed

<table>
<thead>
<tr>
<th>Patient</th>
<th>Protease mutations (coverage; frequency)</th>
<th>Reverse transcriptase mutations (coverage; frequency)</th>
<th>RT Connection mutations (coverage; frequency)</th>
<th>RT RNase H mutation (coverage; frequency)</th>
<th>Integrase mutations (coverage; frequency)</th>
<th>Envelope mutations (coverage; frequency)</th>
<th>ART regimen</th>
<th>Tropism ¹</th>
<th>Subtype/ URF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>V82I (6,081; 95.6%)</td>
<td>M184V (11,979; 2.0%)</td>
<td>-</td>
<td>-</td>
<td>S147G (8,945; 1.1%)</td>
<td>AZT+3TC+NVP</td>
<td>100.0%</td>
<td>X4/R5X4</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>R263K (2,201; 6.2%)</td>
<td>-</td>
<td>AZT+3TC+EFV</td>
<td>63.0%</td>
<td>X4/R5X4</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>97.2%</td>
<td>R5</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>I47V (3,589; 10.3%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>AZT+3TC+ATV</td>
<td>100%</td>
<td>R5</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>NA</td>
<td>E399D (2,701; 99.2%)</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>97.1%</td>
<td>R5</td>
<td>BC</td>
</tr>
<tr>
<td>11</td>
<td>D30N (8,012; 2.8%)</td>
<td>M41L (5,444; 99.8%)</td>
<td>T369V (9,726; 39.6%)</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>96.1%</td>
<td>R5</td>
<td>BF</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>84.8%</td>
<td>R5</td>
<td>B</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>E138K (4,909; 1.3%)</td>
<td>-</td>
<td>V38A (2,151; 3.3%)</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>99.0%</td>
<td>X4/R5X4</td>
<td>B</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>R263K* (1,776; 11.0%)</td>
<td>-</td>
<td>AZT+3TC+LPV/r</td>
<td>97.8%</td>
<td>R5</td>
<td>B</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>E399G (14,872; 1.5%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>97.7%</td>
<td>R5</td>
<td>B</td>
</tr>
<tr>
<td>16</td>
<td>D30N (4,574; 49.3%)</td>
<td>M46I (4,378; 45.6%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>AZT+3TC+FPV/r</td>
<td>NA</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>A62V (4,116; 1.0%)</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>TDF+3TC+EFV</td>
<td>91.6%</td>
<td>R5</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>NA</td>
<td>A376S (6,663; 99.9%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>AZT+3TC+EFV</td>
<td>99.6%</td>
<td>R5</td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>TDF+3TC+EFV</td>
<td>100.0%</td>
<td>R5</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>L210W (12,583; 100.0%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TDF+EFV+FTC</td>
<td>99.3%</td>
<td>R5</td>
</tr>
<tr>
<td>22</td>
<td>-</td>
<td>NA</td>
<td>A376S* (1,082; 94.9%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>100.0%</td>
<td>R5</td>
</tr>
<tr>
<td>23</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>T97A (4,113; 2.1%)</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>96.0%</td>
<td>R5</td>
</tr>
<tr>
<td>26</td>
<td>-</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>100.0%</td>
</tr>
<tr>
<td>27</td>
<td>-</td>
<td>NA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>99.9%</td>
<td>R5</td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>M41L (5,791; 99.5%)</td>
<td>-</td>
<td>D67N (6,465; 72.9%)</td>
<td>E399D (2,051; 99.9%)</td>
<td>NA</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>99.1%</td>
</tr>
<tr>
<td>29</td>
<td>-</td>
<td>-</td>
<td>E399D (5,209; 100.0%)</td>
<td>T66I (7,444; 31.0%)</td>
<td>NA</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>99.7%</td>
<td>X4/R5X4</td>
</tr>
<tr>
<td>31</td>
<td>-</td>
<td>M184V (2,708; 1.3%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>100.0%</td>
<td>R5</td>
</tr>
<tr>
<td>32</td>
<td>-</td>
<td>V179D (7,146; 99.3%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TDF+3TC+EFV</td>
<td>100.0%</td>
<td>R5</td>
</tr>
</tbody>
</table>

#The near full-length genomes are in bold; NA, not available; -, no mutations found; *, only partial sequence available; 1X4/R5X4: CXCR4 and/or CXCR4/CCR5 tropism profile; R5: CCR5 tropism profile.
The integrated state of HIV allows the persistence of HIV, either as wild-type or drug-resistant variants, as archived proviruses in PBMCs. Consequently, this persistence may potentially compromise the efficacy of targeted antiretroviral drugs exerting a long-term impact on responses to HAART [28, 90]. Also, it should be note that routine genotyping tests generally focus on currently plasma circulating viral variants and on a limited number of short genomic fragments, despite the growing need to simultaneously probe multiple genomic regions that target different steps in the viral life cycle. These facts, coupled with previous observations that standard bulk sequencing cannot fully access the spectrum of viral variants archived in the proviral DNA, justify the use of NGS technologies to study proviral DNA from PMBCs as a valuable source for resistance analysis [14, 91-93].

We found 18 patients (78%) who were successful responders to first-line ART (undetectable HIV viral load) that carried archived viruses harboring at least one drug resistance-associated mutation. Only a single study has been previously published in this regard on HIV patients undergoing therapeutic success [48]. The authors evidenced the presence of antiretroviral resistance mutations above 1% of frequency in five of the eleven patients analyzed (45%). The difference in prevalence found in both studies can be attributed to differences in the ultradepth sequencing method used (MiSeq - Illumina versus 454 Life Science - Roche GS Junior), the regions covered by the study (NFLG versus Gag, Pol and Nef regions) and the associated coverage obtained, as two patients of their cohort only presented mutations under 1%.

Eight of our proviral sequences (8/23, 22%) harbored only minority variants with drug resistance-associated mutations. This finding underscores the use of more sensitive HIV genotyping techniques, since the standard genotypic resistance testing using bulk sequencing of the viral population can only detect viral variants that constitute over 15-20% of the total viral population [62], and likely underestimates the overall prevalence of resistant variants and may impact on the surveillance of HIV resistance and on the clinical management of treated patients.

The clinical significance of drug-resistant minority variants is still uncertain and has been addressed by many studies, due to its potential impact on the response to future treatment schemes. Several studies point out that these resistance variants would be favored by the ARV selective pressure, leading to a substitution of the wild-type virus and consequently to therapeutic failure [18, 30-35]. However, other studies found no influence of these minority drug-resistance variants on treatment response, highlighting the role of the majority, not the minority variants, in this outcome [36-42].

Despite the high prevalence of drug resistance mutations (18/23) and of mutations associated with the current ARV therapy of the patients (6/23) in our study, only one patient had detectable viral load 15 months after inclusion in the study. Unfortunately, we failed to PCR-amplify the HIV RT polymerase domain of this patient, and only the E399D mutation was found in RT connection domain of that virus. Most of the drug resistance mutations associated with the current therapy regimens of the patients analyzed were found at low prevalence at the viral population (1.3%-8%), except for the patient harboring TAMs (patient #28; Table 3). Surprisingly, this patient maintained an indetectable viral load 14 months after inclusion in the study, paralleled by an increase in the CD4+ T-cell counts (681 cells/mm³) compared to those at the time of enrolment (535 cells/mm³). The low prevalence observed in most of the mutations found and the high adherence reported by the clinicians of the program may have contributed to the high therapeutic success rate observed in these patients one year after inclusion in this study.

In Brazil, two entry inhibitors/antagonists, enfuvirtide (ENF) and maraviroc (MVC), have been used since 2005 and 2007, respectively, in therapeutic rescue strategies for patients failing previous ARV regimens [94, 95]. The resistance mutation to ENF observed in one of our patients (#13; Table 3) emphasizes the importance of including this region at the genotypic resistance analyses, since resistance to ENF is characterized by a low genetic barrier [96-98] and can lead to therapeutic failure if used in patients carrying resistance to that drug.

Two patients (#5 and 11) were found to harbor proviruses with three transmitted drug resistance mutations to protease inhibitors (PI), all of them as minority variants (frequency between
1.9% and 10.3% of the virus population). The treatment history of each patient was considered to define the transmitted resistance. All patients were under NRTI use, so it was not possible to characterize NRTI-associated TDRM. Only two patients were under PI use in their treatments. The prevalence of TDR found in this study (2/21; 9%) was in agreement with the moderate TDR prevalence reported by other Brazilian studies, usually ranging from 5% to 10% [99-104].

With respect to HIV coreceptor tropism, R5 viruses usually predominate during primary HIV infection, whereas the transition to X4 viruses occurs in later stages of HIV disease, being associated with more rapid CD4+ T-cell depletion and consequently to AIDS progression [105-109]. Genotypic predictors prove to be highly concordant with phenotypic data and can be reliably used to determine viral tropism with better results in PBMC than in plasma samples [110]. In this study, we used the Geno2pheno algorith with a false positive rate (FPR) cutoff of 10% [51, 68-71]. Prediction of coreceptor usage showed that most individuals (17/22; 77%) presented only R5-tropic viruses, while five individuals presented X4/R5X4-tropic viruses. The prevalence of R5-tropic viruses was similar to the 78% prevalence found by de Azevedo et al. [111] in another Brazilian cohort.

Hypermutated proviral sequences were detected in only one individual (patient #26) through the identification of an excessive G → A change pattern, consistent with APOBEC3G/F signature. Several early stop codons resulted from those nucleotide substitutions were observed along that HIV genome. The evidence of hypermutation as an APOBEC action was also highlighted by the Stanford HIV drug resistance database during the analysis of resistance mutations for that virus.

We are aware that additional studies should be carried out to validate our findings in a larger population. Our strict inclusion criteria did not allow the enrollment of a large cohort for this study. Another limitation was the difficulty of PCR-amplifying the archived proviral genomes from some patients in a setting of undetectable HIV viral load and early chronic infection, where PBMC archived HIV reservoirs are thought to be small. Even with the utilization of diverse strategies, some genomic regions could not be amplified for some patients.

The analysis of resistance-associated mutations in HIV-positive patients is an important issue when considering the context of broad access to antiretroviral treatment and high rates of therapeutic success, one of the main goals to be achieved worldwide against HIV infection in the coming years. The high rate of resistance-associated mutations found in our cohort, composed of patients with undetectable viral load undergoing HAART as first-line therapy, directs attention to the selection of antiretroviral resistant variants in a context of therapeutic success and early in chronic infection. These observations underscore the importance of further studies in order to better correlate the presence of these drug resistant mutations and response to ART to investigate their potential association with therapeutic failure and to establish effective public policies to decrease their prevalence and transmission.

**Supplementary Materials:** The HIV-1 consensus sequences were deposited at the GenBank nucleotide database under the accession numbers MG571979-MG572011. The raw NGS reads were submitted to the Sequence Read Archive (SRA) under the numbers SRR6324863-SRR6324885.

**Acknowledgements:** This work was supported by the Rio de Janeiro State Science Foundation – FAPERJ (grant # E-26/201.331/2014 to M.A.S.) and by the Brazilian National Research Council – CNPq (grants # 454623/2014-3 to M.A.S. and # 460346/2014-8 to E.A.S.). B.M.A. was recipient of a scholarship from the Brazilian Ministry of Education (CAPES) to spend six months at the Lady Davis Institute for Medical Research, McGill University, Montreal, Canada. This work is part of the requirements for B.M.A. to obtain her Ph.D. title at the Graduate Program in Oncology of the Brazilian National Cancer Institute (INCA).

**Author Contributions:** E.A.S. and M.A.S. conceived and financed the study. B.M.A., J.D.S., O.M.B. and I.M.P. performed all the molecular biology and the computer analyses of the work. M.M.G. and...
S.R.R. were responsible for patient recruitment and follow-up, and collected all clinical and laboratory data of the patients. All authors wrote and revised the manuscript, and agreed with the submission of its final version.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


49. CDC, Revised surveillance case definition for HIV infection - United States. *2014*, *63*, (RR-03), 1-10.


64. Keane, T. M.; Creevey, C. J.; Pentony, M. M.; Naughton, T. J.; McLnerney, J. O., Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology* 2006, 6, 29.


