Sensitive Deep Sequencing-based HIV-1 Genotyping Assay to Simultaneously
Determine Susceptibility to Protease, Reverse Transcriptase, Integrase, and
Maturation Inhibitors, as well as HIV-1 Coreceptor Tropism

Richard M. GIBSON 1, Ashley M. MEYER 1, Dane WINNER 1, John ARCHER 2, Felix FEYERTAG 2,
Ezequiel RUIZ-MATEOS 3, Manuel LEAL 3, David L. ROBERTSON 2, Christine L. SCHMOTZER 1, and
Miguel E. QUIÑONES-MATEU 1,4,*

1 University Hospital Translational Laboratory, University Hospitals Case Medical Center, Cleveland, Ohio, USA;
2 Computational and Evolutionary Biology, Faculty of Life Sciences, University of Manchester, Manchester, United
Kingdom; 3 Laboratory of Immunovirology, Biomedicine Institute of Seville, Service of Infectious Diseases, Virgen
del Rocio University Hospital, Seville, Spain; and 4 Department of Pathology, Case Western Reserve University,
Cleveland, Ohio, USA

This research was presented in part at the International HIV & Hepatitis Virus Drug Resistance Workshop and
Curative Strategies, Toronto, Canada (4 to 8 June 2013)

* Corresponding Author and request for reprints: Department of Pathology, Case Western Reserve University /
University Hospitals Case Medical Center, 10900 Euclid Avenue, Cleveland, Ohio 44106-7288, Phone: (216) 286-
4283, Fax: (216) 201-5432. E-mail: meq@case.edu
ABSTRACT

With twenty-nine individual antiretroviral drugs, from six classes, approved for the treatment of human immunodeficiency virus type 1 (HIV-1) infection, a combination of different phenotypic and genotypic tests is currently needed to monitor HIV-infected individuals. In this study, we have developed a novel HIV-1 genotypic assay based on deep sequencing (DEEPGEN™HIV) to simultaneously assess HIV-1 susceptibility to all drugs targeting the three viral enzymes as well as to predict HIV-1 coreceptor tropism. Patient-derived gag-p2/NCP7/p1/p6/PR/RT/IN- and env-C2V3 PCR products were sequenced using the Ion Torrent Personal Genome Machine™. Reads spanning the 3’ end of Gag, protease (PR), reverse transcriptase (RT), integrase (IN), and V3 regions were extracted, truncated, translated, and assembled for genotype and HIV-1 coreceptor tropism determination. DEEPGEN™HIV consistently detected both minority drug-resistant viruses and non-R5 HIV-1 variants from clinical specimens with viral loads ≥1,000 copies/ml, and from B and non-B subtypes. Additional mutations associated with resistance to PR, RT, and IN inhibitors, previously undetected by standard (Sanger) population sequencing, were reliably identified at frequencies as low as 1%. DEEPGEN™HIV correlated with phenotypic (original Trofile, 92%; ESTA, 80%; Trocai, 81%; and VERITROP, 80%) and genotypic (population sequencing/Geno2Pheno 10% FPR, 84%) HIV-1 tropism tests. DEEPGEN™HIV (83%) and Trofile (85%) showed similar concordance with clinical response following an 8-day maraviroc monotherapy (MCT). In summary, this novel all-inclusive HIV-1 genotypic and coreceptor tropism assay, based on deep sequencing of the PR, RT, IN, and V3 regions, permits the multiplex detection of low level drug-resistant and/or non-R5 viruses in up to 96 clinical samples simultaneously. This comprehensive test, the first of its class, will be instrumental in the development of new antiretroviral drugs and, more important, will aid in the treatment and management of HIV-infected individuals.
INTRODUCTION

According to UNAIDS, at the end of 2012 close to ten million of the approximately 34 million people living with HIV had access to antiretroviral therapy, with the goal being to reach 15 million people with HIV treatment by 2015 (1). This broader access to antiretroviral drugs has led not only to considerable reductions in morbidity and mortality (2, 3) but, unfortunately, has increased the risk of virologic failure due to emergence (3, 4), and potential transmission (5), of drug-resistant viruses. Therefore, detecting and quantifying drug resistance, and in the case of CCR5 receptor antagonists determining HIV-1 coreceptor tropism, has become the standard of care prior to designing new antiretroviral regimens (3, 6-8).

To date, twenty-nine individual antiretroviral drugs from six drug classes have been approved by the U.S. Food and Drug Administration (FDA) to be used in the treatment of HIV-1 infection, including protease (PI), nucleoside/nucleotide reverse transcriptase (NRTI), nonnucleoside reverse transcriptase (NNRTI), integrase (INI), fusion (F), and entry (E), inhibitors (www.fda.gov). HIV-1 resistance to PI, NRTI, NNRTI, and INI can be determined using (i) indirect methods based on detection of specific amino acid substitutions (due to underlying nucleotide mutations) in the respective coding regions previously associated with resistance to specific antiretroviral drugs (i.e., genotyping) (6, 9, 10), (ii) more direct methods that test the ability of a patient-derived virus to replicate in the presence of antiretroviral drugs in a cell-based assay (i.e., phenotyping) (11-13), or (iii) a combination of both approaches that takes advantage of a large database to infer the level of HIV-1 drug-resistance based on genotyping and its relationship with matched phenotypic data (14). Similarly, since treatment with CCR5 antagonists requires the prior knowledge of the HIV-1 coreceptor tropism in the patient, i.e., CCR5- or CXCR4-tropic viruses (R5 and X4, respectively), dual tropic (R5/X4), or a mixture of both R5 and X4 viruses (7, 15), a multitude of phenotypic and genotypic approaches to determine HIV-1 coreceptor tropism have been developed (8, 16, 17). Phenotypic assays to determine HIV-1 drug resistance or tropism usually involve the generation of patient-derived pol- (11-13) or env-recombinant (18-20) viruses, respectively, to
quantify their ability to infect susceptible cell lines expressing the appropriate HIV-1 receptors and coreceptors
or, in the case of HIV-1 tropism, may also be based on the quantification of cell-to-cell fusion events (21-23).
Whereas, genotypic HIV-1 tropism tests (8, 15, 16, 24-26) take advantage of the properties of specific regions in
the env gene as determinants of CCR5 or CXCR4 tropism, mainly the V3 region of the gp120, and their
interpretation based on a series of bioinformatic methods to infer the ability of HIV-1 to use any or both
coreceptors to enter host cells (27-30).

As expected, phenotypic (experimental) and genotypic (computational) approaches to determine HIV-1 drug
resistance or HIV-1 coreceptor tropism have some disadvantages, including the longer turnaround time and
higher cost of the phenotypic assays or the intrinsic predictive nature of the genotypic tests. Particular
emphasis has been made on the limited sensitivity of genotypic HIV-1 tropism assays to detect minor non-R5
variants (16, 31), and to a lesser extent on the ability of genotypic HIV-1 drug resistance tests to detect minority
drug resistant variants (32-34). In the case of HIV-1 drug resistance, the vast amount of information
accumulated during the last two decades by correlating mutations with phenotypic data has led to the almost
exclusive use of genotypic antiretroviral testing based on population (Sanger) sequencing to manage patients
infected with HIV-1 (2, 35). In contrast, although several studies have shown significant concordance and similar
predictive values (36-40), genotypic HIV-1 tropism assays based on population sequencing seem to be less
sensitive and specific than phenotypic assays (8, 16, 17, 41). Thus, a cell-based assay (Trofile, Monogram
Biosciences)(19, 42) is currently the standard method to determine HIV-1 coreceptor tropism in the U.S., while
genotypic HIV-1 tropism tests are largely used in Europe (16, 31).

To date, all current commercial genotypic HIV-1 drug resistance assays are based on population sequencing
(10, 43, 44), which can only detect minority variants present above 20% of the viral population (44-48).
However, and although still uncertain, drug resistant HIV-1 minority variants (i.e., as low as 1% of the viral
population) have been suggested to be clinically relevant as they have a high chance of being selected for under antiretroviral drug pressure (49-57). For that reason, a series of ultrasensitive assays have been developed to detect drug resistant HIV-1 minority variants, e.g., allele-specific polymerase chain reaction (49, 58), oligonucleotide ligation assays (33, 59), and deep (next-generation) sequencing (60-62). On the other hand, as described above, the adoption of genotypic HIV-1 tropism assays in the clinical setting has been hampered by the limited sensitivity of the population-based sequencing assays to detect minor non-R5 variants. Therefore, more sensitive genotypic HIV-1 tropism assays based on deep sequencing have been developed to detect non-R5 variants present at frequencies below 20% of the population, which have been shown to correlate well with both phenotypic assays (36, 63-67) and virological response to CCR5-receptor antagonists such as maraviroc (Selzentry/Celsentri, Pfizer, NY) (36, 63, 66). Nevertheless, a combination of at least two different genotypic assays are still needed to assess the susceptibility of a patient-derived HIV-1 to all FDA-approved antiretroviral drugs, including CCR5 antagonists. Therefore, in this study we have developed, characterized, and validated a novel HIV-1 genotyping assay based on deep sequencing to simplify the monitoring of patients infected with HIV-1. This all-inclusive, sensitive methodology accurately provides drug resistance information for all protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism, in a single, more efficient, rapid, and affordable clinical assay.
MATERIALS and METHODS

Viruses and plasmids. The following viruses were obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 A-92RW009, HIV-1 A-93RW020, HIV-1 A-92UG029, HIV-1 B-92BR014, HIV-1 B-92TH539, HIV-1 B-US714, HIV-1 B-US727, HIV-1 C-92BR025, HIV-1 D-94UG106, HIV-1 D-92UG038, HIV-1 D-93UG065, HIV-1 F-93BR029, and HIV-2CBL-20 or Dr. Eric J. Arts’ laboratory at Case Western Reserve University (CWRU), Cleveland, OH: HIV-1 A-V115, HIV-1 A-V120, HIV-1 C-C18, HIV-1 C-C20, HIV-1 C-CZ1, HIV-1 D-V89, HIV-1 D-V122, HIV-1 D-V126, HIV-1 F-V820, HIV-1 F-V164, HIV-1 F-CA16, and HIV-1 F-CA20. Aliquots of additional RNA or DNA viruses were obtained from the Molecular Diagnostics or Medical Microbiology laboratories at University Hospitals Case Medical Center (UHCMC), Cleveland, OH (BK virus, BKV; Cytomegalovirus, CMV; Herpes simplex virus 1 and 2, HSV-1 and HSV-2; and Varicella zoster virus, VZV) or the Division of Infectious Diseases, School of Medicine at CWRU (Hepatitis B virus, HBV; Hepatitis C virus, HCV; and Epstein-Barr virus, EBV). Plasmids containing patient-derived HIV-1 gag-p2/NCP7/p1/p6/pol-PR/RT/IN-coding sequences from multidrug-resistant viruses, i.e., 08-180 and 08-194 have been previously described (68), as well as a pNL4-3 plasmid containing the env gene from the R5 HIV-1 virus (69). Five plasmid mixtures containing drug resistance mutations in the HIV-1 pol gene, i.e., K65R (5%) + wild type (95%), K103N (5%) + wild type (95%), K101E (5%) + E138K (5%) + wild type (90%), K101E + E138K (10%) + wild type (90%), and M184V (RT) + E92Q (IN) (10%) + wild type (90%) were obtained from Gilead Sciences, Inc. (Foster City, CA).

Clinical samples. Plasma samples for the characterization and verification of the novel HIV-1 genotypic and coreceptor tropism assay were obtained during routine patient monitoring from a well-characterized cohort of HIV-infected individuals at the AIDS Clinical Trials Unit (ACTU) at CWRU/UHCMC, with the understanding and written consent of each participant. RNA specimens, derived from plasma samples collected from HIV-infected individuals enrolled in the (i) maraviroc expanded access program in Europe or (ii) ALLEGRO trial, were obtained.
written informed consent was obtained from the patients before participation in the study as previously described (40, 70). HIV-1 coreceptor tropism was determined at baseline using two phenotypic assays, i.e., the original version of the Trofile™ (19) and VERITROP™ (23) and by population sequencing analyzed with Geno2Pheno (27), with a false positive rate (FPR, predicted frequency of classifying an R5 sequence as non-R5 virus) based on optimized cutoffs associated with the analysis of clinical data from MOTIVATE (2.5% and 5.75%) (63). Finally, plasma samples were obtained from HIV-infected individuals at the Infectious Diseases Unit Virgen del Rocio University Hospital (Sevilla, Spain) participating in a study to evaluate the use of an 8-day maraviroc monotherapy clinical test (MCT) (71, 72). Patients provided written informed consent and the ethical committee of the hospital approved the study (72). HIV-1 coreceptor tropism in these samples was determined at baseline using two different phenotypic assays, i.e., the enhanced sensitivity Trofile assay (ESTA) (42) and Trocai (73) and by population sequencing analyzed with Geno2Pheno (27), with a FPR of 10% following the recommendations from the European Consensus Group on clinical management of HIV-1 tropism testing as described in the Geno2Pheno website (http://coreceptor.bioinf.mpi-inf.mpg.de/index.php).

Reverse transcription (RT)-PCR amplification of gag-p2/NCp7/p1/p6/pol-PR/RT/IN- and env-C2V3-coding regions. Plasma viral RNA was purified from pelleted virus particles by centrifuging one milliliter of plasma at 18,000g x 60 min at 4°C, removing 860 µl of cell-free supernatant and resuspending the pellet in the remaining 140 µl, to finally extract viral RNA using QIAamp Viral RNA Mini kit (Qiagen; Valencia, CA). Viral RNA was reverse-transcribed using AccuScript High Fidelity Reverse Transcriptase (Stratagene Agilent; Santa Clara, CA) and the corresponding antisense external primers in 20 µl reaction mixture containing 1 mM dNTPs, 10 mM DTT and 10 units of RNase inhibitor. The HIV-1 genomic region encoding the Gag proteins p2, p7, p1 and p6, and the protease, reverse transcriptase, and integrase enzymes was amplified as two overlapping fragments (1,657 nt and 2,002 nt corresponding to the p2-5’half RT and 3’half RT-INT, respectively) using a series of external and
nested primers with defined cycling conditions (13). External PCR reactions were carried out in a 50-µl mixture containing 0.2 mM dNTPs, 1 mM MgCl₂ and 2.5 units of Pfu Turbo DNA Polymerase (Stratagene). Nested PCR reactions were carried out in 50-µl mixture containing 0.2 mM dNTPs, 0.3 units of Pfu Turbo DNA Polymerase and 1.9 units of Taq Polymerase (Denville Scientific; Metuchen, NJ). A fragment corresponding to the C2V3 region (480 nt) of the surface glycoprotein (gp120) in the envelope gene was amplified using a series of external and nested primers with defined cycling conditions as previously described (74).

**Population (Sanger) sequencing analysis.** PCR products corresponding to the gag-p2/NCp7/p1/p6/pol-PR/RT/IN- and env-C2V3-coding regions of HIV-1 were purified with the QIAquick PCR Purification kit (Qiagen) and sequenced (Sanger, population, or global sequence) using AP Biotech DYEnamic ET Terminator cycle with Thermosequenase II (Davis Sequencing LCC, Davis, CA). Nucleotide sequences were analyzed using DNASTAR Lasergene Software Suite v.10.0.1 (Madison, WI).

**Deep sequencing of gag-p2/NCp7/p1/p6/pol-PR/RT/IN- and env-C2V3-coding regions.** The three PCR products corresponding to the gag-p2/NCp7/p1/p6/pol-PR/RT/IN- (1,657 nt and 2,002 nt fragments) and env-C2V3 (480 nt fragment) coding regions of HIV-1 were purified (Agencourt AMPure XP, Beckman Coulter) and quantified (2100 Bioanalyzer DNA 7500, Agilent Technologies) prior to using the Ion Xpress™ Fragment Library Kit (Life Technologies, Carlsbad CA) to construct a multiplexed library for shotgun sequencing on the Ion Personal Genome Machine™ (PGM, Life Technologies)(Figure 1). Briefly, a mixture of all three purified DNA amplicons (33 ng each) was randomly fragmented and blunt-ends repaired using the Ion Shear™ Plus Reagent (Life Technologies) followed by DNA purification (Agencourt AMPure XP, Beckman Coulter). The P1 adapter (5'-CCA CTA CGC CGC TTT CCT CTC TAT GGG CAG TCG GT G AT; 5'-ATC ACC GAC TGC CCA TAG AGA GGA AAG CGG AGG CGT AGT GG*T*T) and one of 96 barcodes were ligated to the repaired fragment ends prior to DNA purification (Agencourt AMPure XP, Beckman Coulter). DNA fragments were then selected by size (i.e., 300 bp;
Pippin Prep™, Life Technologies) and each barcoded library, i.e., a mixture of all three amplicons per sample, was purified (Agencourt AMPure XP, Beckman Coulter) and normalized using the Ion Library Equalizer™ Kit (Life Technologies). All barcoded DNA libraries, corresponding to patient-derived amplicons plus the HIV-1 NL4-3 control, were pooled in equimolar concentrations and templates prepared and enriched for sequencing on the Ion Sphere Particles™ (ISPs) using the Ion OneTouch™ 200 Template Kit v2 (Life Technologies) in the Ion OneTouch™ 2 System (Life Technologies). Templated ISPs were quantified (Qubit® 2.0, Life Technologies) and loaded into an Ion 318™ Chip (Life Technologies) to be sequenced on the Ion PGM™ using the Ion PGM™ Sequencing 200 Kit v2 (Life Technologies). Following a 4 hours and 20 minutes sequencing run, signal processing and base calling was performed with Torrent Analysis Suite version 3.4.2.

**Read mapping, variant calling, and phylogenetic analysis.** As part of the novel HIV-1 genotypic and coreceptor tropism assay we developed the DEEPGEN™ Software Tool Suite for the processing of HIV-1 deep sequencing data and HIV-1 drug resistance determination. DEEPGEN™ uses two main tools: **Viral Read Mapper** and **Variant Caller**.

- **Viral Read Mapper:** To minimize the amount of data loss during mapping due to the high HIV-1 sequence variability and to allow for inter-patient indel variation across the gag-p2/NCp7/p1/p6/pol-PR/RT/IN- and env-C2V3-coding regions, sample-specific reference sequences were constructed for each one of these two genomic regions, i.e., positions 1,807 to 5,096 and 6,900 to 7,400, in the HXB2 reference strain (GenBank accession no. K03455), respectively. Mapping of reads from each sample/region occurred in three stages. First, a guide template for mapping was selected from the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html) by comparing 100 randomly selected reads to the corresponding region within all full-length sequences present within the HIV Sequence Database. This comparison was performed using a k-mer approach that rapidly identifies regions of similarity between any two sequences to select a guide sequence for mapping with minimal divergence from the read data, as such
divergence is the primary cause of biased data loss (75). Following the selection of a guide sequence, reads were mapped and aligned using the mapping algorithm previously described (76). During mapping site indexes in relation to HXB2 were also maintained. Next to reduce diversity between reads and reference, a consensus was generated across each site of the guide sequence and reads re-mapped to this final consensus template. Reads spanning the 3’end of Gag, PR, RT, and INT were then translated and assembled for genotyping.

- **Variant Caller**: Variant calling, the identification and calculation of the frequency of each amino acid present in each genomic position, was calculated using as input a table generated by the Viral Read Mapper, which includes the nucleotide frequencies at each position relative to the reference sequence and with numbering relative to the HIV-1b_HXB2 reference strain. Coverage, indel, codon, and residue frequencies at each position were also listed. *Variant Caller* summarized the results in a graphical interface with particular focus on sites of known drug resistance based on the latest edition of the IAS-USA HIV Drug Resistance Mutations list (77). A list of the amino acids at these positions, and their frequencies, was exported as a tabulated text file and used with the HIVdb Program Genotypic Resistance Interpretation Algorithm from the Stanford University HIV Drug Resistance Database ([http://hivdb.stanford.edu](http://hivdb.stanford.edu)) to infer the levels of susceptibility to protease, reverse transcriptase, and integrase inhibitors.

In addition, for each dataset, reads spanning amino acid positions (i) 50 to 85 in the protease (HXB2 2,400 to 2,508), (ii) 180 to 215 in the RT (HXB2 3,087 to 3,195), (iii) 130 to 165 in the integrase (HXB2 4,617 to 4,725), and (iv) 1 to 35 in the V3 region (HXB2 7,110 to 7,217) were extracted, truncated and translated for phylogenetic analysis and HIV-1 coreceptor tropism prediction as described below. Within each dataset only one representative of any identical variant was maintained, but the overall frequency stored. All variants with a frequency >1 within the population were aligned using ClustalW (78) and phylogeny reconstructed using the neighbor-joining statistical method as implemented within MEGA 5.05 (79). In this study, minority variants were defined as variation detected at >1% (based on the intrinsic error rate of the system as described below) and
<20% of the virus population, corresponding to those mutations that cannot be determined using population sequencing (44-48).

**Genotypic HIV-1 coreceptor tropism determination.** HIV-1 co-receptor tropism was predicted from population and deep sequencing V3 sequences using Geno2Pheno (27). Regarding population V3 sequences, nucleotide mixtures were considered when the second highest peak in the electropherogram was above 25%, and the nucleotide mixtures translated into all possible permutations. Geno2Pheno with a FPR of 2.5% and 5.75% based on optimized cutoffs associated with the analysis of clinical data from MOTIVATE (2.5% and 5.75%) (63) or a FPR of 10% following the recommendations from the European Consensus Group on clinical management of HIV-1 tropism testing as described in the Geno2Pheno website (http://coreceptor.bioinf.mpi-inf.mpg.de/index.php), were used for the clinical samples obtained from the Madrid and Seville cohorts, respectively. In the case of deep sequencing V3 sequences, reads spanning amino acid positions 1 to 35 in the V3 region (HXB2 7,110 to 7,217) were extracted and truncated for HIV-1 coreceptor tropism determination using Geno2Pheno (27) with a FPR of 3.5% based on optimized cutoffs for determining HIV-1 coreceptor usage as previously described (36, 63, 80). Deep sequencing V3 sequences usually spanned 105 nucleotides (35 amino acids), with some minor discrepancies associated with natural HIV-1 variation (81, 82), which led to V3 sequences with an open reading frame of 96, 99, 102, 108, or 111 nucleotides, all starting and ending with a cysteine codon, i.e., TG(T/C). V3 reads with stop codons (TGA, TAA, or TAG) and/or where the nucleotide length was not a multiple of 3 (e.g., 101, 103, 104, 106, etc.), mostly associated with natural or methodology (PCR or sequencing)-induced insertions and/or deletions, were not included in the analysis. Deep sequencing of the V3 region was considered unsuccessful if reads from the majority variants had to be omitted from the analysis. Finally, plasma samples were classified as containing non-R5 viruses if at least 2% of the individual sequences, as determined by deep sequencing, were predicted to be non-R5 (36, 63).
**Statistical Analyses.** Descriptive results are expressed as median values, interquartile ranges, standard deviations, and confidence intervals. Pearson’s correlation coefficient was used to determine the strength of association between categorical variables. A paired t-test was used to compare the number of drug resistance mutations detected by population and deep sequencing in the same sample. All differences with a \( P \) value of <0.05 were considered statistically significant. The kappa coefficient, which assesses a chance-adjusted measure of the agreement between any number of categories, was calculated using ComKappa3 v.3.0.1 (83) to quantify the concordance among the different HIV-1 tropism determinations. All statistical analyses were performed using GraphPad Prism v.6.0b (GraphPad Software, La Jolla, CA) unless otherwise specified. gag-p2/NCp7/p1/p6/pol-PR/RT/IN and/or env-V3 nucleotide sequences obtained by deep sequencing in this study have been submitted to the Los Alamos National Laboratory HIV-DB Next Generation Sequence Archive.
RESULTS

Characterization of the RT-PCR amplification step. As described in Materials and Methods and shown in Figure 1, the novel HIV-1 genotyping and coreceptor tropism assay requires the RT-PCR amplification of three amplicons covering the HIV-1 gag-p2/NCp7/p1/p6/pol-PR/RT/IN- and env-C2V3-coding regions. The characterization of the two overlapping PCR products (1,657 nt and 2,002 nt) comprising the sub-genomic HIV-1 region spanning the Gag proteins p2, p7, p1 and p6, and the protease, reverse transcriptase, and integrase coding regions was detailed previously as part of a study describing VIRALARTSTM HIV, an HIV-1 phenotyping assay (13). Thus, here we focused in the characterization of the 480 nt fragment of the HIV-1 env gene, corresponding to the C2V3 region of the surface glycoprotein (gp120). The sensitivity of the RT-PCR amplification step was tested by analyzing 79 plasma samples obtained from the ACTU (Cleveland, Ohio). Blood samples from HIV-infected individuals with plasma viral loads ranging from 1,000 to >10,000 copies of viral RNA/ml were used to PCR amplify the C2V3 fragment. Similar to results observed with the two overlapping fragments spanning the gag-p2/NCp7/p1/p6/pol-PR/RT/IN-coding region (13), RT-PCR products of the correct size were consistently obtained (92%, 73/79) in plasma samples with ≥1,000 copies/ml of HIV RNA (Table 1).

Highly reproducible success in RT-PCR amplification of the specific HIV-1 gag-p2/NCp7/p1/p6/pol-PR/RT/IN- and env-C2V3 products was obtained when testing fifteen plasma samples with different viral loads. Details of these tests using two different operators, with different lots of critical reagents, and over a seven-day period are described in Weber et al (13), for the gag-p2/NCp7/p1/p6/pol-PR/RT/IN fragments, and in Supplemental Table 1 for the env-C2V3 fragment. The specificity of the RT-PCR primers and reactions for the env-C2V3 fragment was analyzed using nucleic acids from a series of RNA and DNA viruses (i.e., BKV, CMV, HSV-1, HSV-2, VZV, HBV, HCV, and EBV). As expected, no cross-reactivity was observed with any of these viruses as all RT-PCR reactions failed to generate any detectable amplicons (Supplemental Fig. 1). Similar results were obtained for the gag-p2/NCp7/p1/p6/pol-PR/RT/IN fragments as previously described (13).
Finally, although most of the HIV-1 genotyping and coreceptor tropism determinations are performed in North America, Europe, and Australia where subtype HIV-1 strains are predominant (http://www.who.int/hiv/pub/global_report2010/en/index.html), it was important to test the ability of the assay to work with more worldwide prevalent non-B HIV-1 variants. For that, the env-C2V3 fragment was RT-PCR amplified from 33 diverse HIV-1 isolates, including five subtype A (HIV-1_A-92RW009, HIV-1_A-93RW020, HIV-1_A-92UG029, HIV-1_A-V115, and HIV-1_A-V120), five subtype B (HIV-1_B-92BR014, HIV-1_B-92TH593, HIV-1_B-US714, HIV-1_B-92US727, and HIV-1_B-92US076), five subtype C (HIV-1_C-92BR025, HIV-1_C-C18, HIV-1_C-C20, HIV-1_C-C21, and HIV-1_C-C22), six subtype D (HIV-1_D-94UG108, HIV-1_D-92UG038, HIV-1_D-93UG065, HIV-1_D-V89, HIV-1_D-V122, and HIV-1_D-V126), six subtype F (HIV-1_F-93BR029, HIV-1_F-93BR020, HIV-1F-V1820, HIV-1_F-V164, HIV-1_F-C16, and HIV-1_F-C20), two subtype G (HIV-1G-RU570 and HIV-1G-RU132), and four circulating recombinant forms (HIV-1_AE-CMU02, HIV-1_AE-CMU06, HIV-1_AE-92TH021, and HIV-1_BF-93BR029). Amplicons of the correct size were obtained for the env-C2V3 (Supplemental Table 2) and gag-p2/NCp7/p1/p6/pol-PR/RT/IN fragments (13) from all HIV-1 group M isolates analyzed, while negative or inconclusive results were obtained with the HIV-2_CBL-20 strain (data not shown).

**Estimation of the intrinsic error rate of the assay.** Point mutations, insertions, and deletions (indels) can be introduced in the PCR amplification and sequencing steps of any deep sequencing-based assay (84). Therefore, it was important to calculate the intrinsic (combined) error rate of our novel HIV-1 genotyping and coreceptor tropism assay since this value could affect the practical limit of detection of the assay. For that, the pNL4-3-hRluc plasmid containing the entire genome of the wild type HIV-1 group M strain (69) was transformed into Electrocomp TOP10 bacteria (Invitrogen). One bacteria colony was grown overnight in 10 ml of bacteria culture, plasmid DNA was purified, and transformed again into bacteria. Ten individual, theoretically identical, bacteria colonies were used for the direct PCR amplification of the gag-p2/NCp7/p1/p6/pol-PR/RT/IN and env-C2V3 fragments (Fig. 2A), and sequenced using the same protocol utilized with the clinical samples. The quality of the DNA sequences was analyzed, and reads filtered, in the Ion Torrent server using a Phred quality score of 20.
(Q20), which provides a base call accuracy of 99% (i.e., a 1 in 100 probability of an incorrect base call). The average coverage (sequencing depth) per nucleotide position for the ten clones was 5,750 (range 681 to 15,614) and 3,797 (range 942 to 9,981) for the gag-p2/NCp7/p1/p6-pol-PR/RT/IN and env-V3 regions, respectively (Fig. 2B). For each individual NL4-3 clone, reads were independently mapped to the pNL4-3-hRluc reference sequence (13, 69) and all point mutation and indel information in relation to the reference was analyzed using Segminator II (76).

Although all ten NL4-3 clones were expected to have no mutations (i.e., point mutation and/or indels) relative to the pNL4-3-hRluc reference sequence, a number of errors were observed throughout the p2/NCp7/p1/p6-pol-PR/RT/IN and V3 regions, ranging from 0% to 29% (mean 0.39%) and 0% to 9.5% (mean 0.37%), respectively (Fig. 2C). The average error frequency due to point mutation was 0.17% (range 0% to 2.5%) and 0.12% (0% to 0.3%) for the p2/NCp7/p1/p6-pol-PR/RT/IN and V3 regions, respectively, whereas the average error rate associated with indels was 0.22% (range 0% to 28%) and 0.25% (0% to 9.2%). Most of the positions with a total (point mutation + indels) error rate above 1% corresponded to the last nucleotide of a homopolymeric region, defined as four or more identical consecutive nucleotides (data not shown). Some of these nucleotide positions corresponded to codons that have been associated with resistance to antiretroviral drugs, e.g., L10 in the protease (3.5%), K101 in the RT (3%), and G193 in the integrase (10.5%) or with coreceptor tropism, e.g., position 11 in the V3 region (2.6%)(Fig. 2D, Table 2). Interestingly, most of the errors in these (homopolymeric) positions corresponded to indels, with a limited number of point mutation errors, e.g., L10 (3.3% vs. 0.22%), K101 (2.7% vs. 0.24%), G193 (9.6% vs. 0.89%), and position 11 in the V3 (2.3% vs. 0.25%), respectively (Table 2). Therefore, considering that (i) the overall error rates for the gag-p2/NCp7/p1/p6-pol-PR/RT/IN and env-V3 regions were 0.39% and 0.37%, respectively, (ii) the point mutation error rates were below 1% for all the codons associated with drug resistance, and (iii) the Variant Caller in the DEEPGEN™ Software...
Tool Suite identifies and filters out the indels, it was reasonable to define a frequency of 1% as the minimum threshold to detect mutations in minority HIV-1 variants with this novel assay.

**Performance of the novel deep sequencing-based HIV-1 genotypic and coreceptor tropism assay.** The measure of success for any deep sequencing-based assay depends on its ability to generate the maximum number of reads per sequencing run (individual sequences), which then allows the detection of minority variants within the HIV-1 population. This inherent quality is the sum of a series of metrics including, but not limited to, (i) number of samples multiplexed and sequenced per run, (ii) chip loading efficiency, (iii) total number of quality reads, (iv) mean read length, and (v) sequencing coverage at each nucleotide position. Most of the deep sequencing runs described in this study involved multiplexing up to 96 individual samples per sequencing reaction, a number that ensured the minimum coverage of 1,000 per nucleotide position sequenced required to secure the detection of a minor variant present at least at 1% of the population (85). Efficient loading of ion sphere particles into the Ion 318™ chip proved to be user-dependent (mean 72%, range 60% to 84%). The total number of quality reads was proportional to chip loading efficiency (empty wells), with other parameters such as enrichment (no template), polyclonality (ISPs with excess DNA library), test fragments, and primer dimers potentially affecting the final number of total reads in this study (mean 3,827,323; range 3,051,463 to 4,936,375). We used the Ion PGM™ Sequencing 200 Kit v2 in all sequencing runs, generating an average read length of 147 bp (range, 119 bp to 178 bp). As expected, the average coverage varied with each sequencing run, correlating mostly with the number of multiplexed samples per sequencing reaction, e.g., 20 samples (mean 9,008; range 3,776 to 15,458 and 6,494; range 2,322 to 8,599 in the gag-p2/NCp7/p1/p6/pol-PR/RT/IN and env-V3 region, respectively) or 96 samples (4,485; 1,612 to 7,274 and 1,017; 966 to 1,070).

As described above, we calculated the error rate of the HIV-1 genotyping and coreceptor tropism assay to be below 1% and incorporated this error-defined cutoff into the evaluation of the analytical sensitivity and limit of
detection for minority HIV-1 variants. For that, we evaluated extensively the analytical sensitivity of the test to detect and quantify drug resistance mutations (\textit{gag-p2/NCp7/p1/p6/pol-PR/RT/IN}) and non-R5 variants (\textit{env-V3}) within mixtures of viral populations. First, we sequenced five plasmid mixtures that contained one or two drug resistance mutations in the RT- and/or IN-coding regions at a frequency of 5% or 10% (i.e., mixture of plasmids containing the respective mutations with a plasmid comprising the wild-type HIV-1 \textit{HXB2} sequence). As shown in Supplemental Figure S2, all mutations were detected and quantified at the expected proportions, including those at a frequency of 5% of the total population. Next, in order to quantify more accurately the analytical sensitivity of the assay, we mixed DNA from a plasmid containing a patient-derived multidrug-resistant \textit{gag-p2/NCp7/p1/p6/pol-PR/RT/IN} fragment in the X4 HIV-1\textit{NL4-3} backbone (08-180)(68) with DNA from a plasmid containing the genome of the wild-type HIV-1\textit{NL4-3} virus carrying the \textit{env} gene from the R5 HIV-1\textit{YU2} virus (69). Plasmid DNA was quantified and dilutions used to prepare eight mixtures containing the X4 multidrug-resistant 08-180 plasmid at 0%, 0.1%, 1%, 2%, 3%, 5%, 10%, and 100% at a final concentration of 0.1 ng/ml. This total plasmid concentration (0.1 ng/ml or 100,000 femtograms/ml) theoretically allowed the detection of 100 fg of the plasmid when diluted to 0.1% of the population using nested PCR (86). Plasmid 08-180\textit{pol/NL43-(X4)env} was generated by the yeast cloning method, which allows a better representation of the \textit{in vivo} HIV-1 quasispecies (13). It contained numerous drug resistance mutations in the protease, RT, and integrase, most of them as majority members of the quasispecies (>99% of the population); however, two amino acid substitutions in the protease were present as minority variants, i.e., L33F at 21.9% and F53Y 1.7%. Interestingly, most drug resistance mutations were detected in the plasmid mixtures containing approximately 1% of the 08-180\textit{pol/NL43-(X4)} plasmid with the exception of substitution T215Y in the RT, which was identified as 0.95% (Fig. 3A and Supp. Fig. S3). As expected, detection of minority mutations leading to two amino acid substitutions (L33F and F53Y) faded quickly and proportional to their frequency in the original population. Similar results were observed during the detection of X4 (\textit{NL4-3}) V3 sequences, which were detected in a plasmid mixture containing approximately 1% of the 08-180\textit{pol/NL43-(X4) env} plasmid (Fig. 3A and Supp. Fig. S3). Unfortunately,
and most likely due to the need to remove V3 sequences with odd open reading frames (as described in Materials and Methods), quantification of X4 sequences in the mixtures containing the 08-180pol/NL43-(X4) env plasmid at 2% and 3% of the population failed or was not accurate, respectively (Fig. 3A and Supp. Fig. S3).

Finally, in order to mimic the first steps of the assay (i.e., RNA purification and RT-PCR) under controlled conditions, HIV-1-seronegative plasma samples were spiked with two viruses, the first one a patient-derived multidrug-resistant gag-p2/NCp7/p1/p6/pol-PR/RT/IN recombinant virus constructed using the X4 HIV-1NL4-3 backbone (08-194)(68) and a wild-type HIV-1NL4-3 virus carrying the env gene from the R5 HIV-1YU2 virus (69).

Plasma HIV-1 RNA (viral) load was determined (COBAS® AmpliPrep/COBAS® TaqMan® HIV-1 test v2.0, Roche) and dilutions used to prepare four mixtures containing the X4 multidrug-resistant 08-194 virus at 0%, 1%, 5%, and 100% in a final viral load of 100,000 copies/ml. This was the average viral load in plasma samples obtained from highly antiretroviral-experienced patients, usually carrying multidrug-resistant viruses, from recent previous studies in our laboratory (13, 68). Viral RNA was purified, RT-PCR amplified, barcoded in quadruplicate and deep sequenced as described in Materials and Methods. As expected, all drug resistance mutations from 08-194 were detected when the mixture contained 100% of this virus (Fig. 3B). Interestingly, a few amino acid substitutions were identified by deep sequencing that were not detected in the original study using Sanger sequencing (68), e.g., F53L (2.2%), V77I (17.4%), I93V (19.7%) in the protease and L100I (3.5%) in the RT coding regions (Fig. 3B). All mutations present at a frequency of >50% in the original 08-194 virus were detected in the 5%-95% (08-194:wild type) mixture; however, none of these mutations were identified when the mixture included 1% of the 08-194 virus (Fig. 3B). Similar results were observed in the env gene, i.e., X4 sequences corresponding to the V3 region of the HIV-1NL4-3 were detected when present at a frequency of 100% and 5% in the viral mixture (Fig. 3B).
Reproducibility of the HIV-1 genotyping and coreceptor tropism assay was evaluated by testing samples from the wild-type HIV-1 control strain (NL4-3), an antiretroviral-naive (12-596) and two antiretroviral-experienced (08-198 and 12-069) individuals. The four samples were RT-PCR amplified in triplicate (3x), each amplicon was barcoded in quadruplicate (4x), DNA libraries prepared in duplicate (2x), and then sequenced twice (2x) for a total of 48 sequences per virus (Fig. 4A). First, reads with a frequency >1 corresponding to 105 bp fragments from the protease, RT, integrase, and V3 regions were used to construct neighbor-joining phylogenetic trees to quantify intra- and inter-patient genetic distances and rule out any potential cross-contamination. Figure 4B shows a clear virus-dependent clustering of sequences in all four HIV-1 regions. As expected, interpatient genetic distances were larger than the range of intrapatient genetic diversity in the four HIV-1 regions, i.e., 0.0495 (0.0023 to 0.0113), 0.0698 (0.0019 to 0.0150), 0.0554 (0.0001 to 0.0145), and 0.3671 (0.0046 to 0.1067) substitutions per site in the protease, RT, integrase, and V3, respectively. Next, the frequency of each nucleotide at each position was compared among the 16 sequences obtained for each one of the triplicate amplicons (n = 48) for all four viruses in the gag-p2/NCp7/p1/p6/pol-PR/RT/IN and env-V3 regions. Statistically significant correlations were observed when the three sets of 16 sequences were compared for each virus, with r values ranging from 0.9857 to 0.9996 (P <0.0001, Pearson coefficient correlation)(Fig. 4C). More important, all 48 sequences detected the same amino acids, with similar frequency, in each position in all four viruses. This was evident when only positions associated with drug resistance in the protease, RT, and integrase were evaluated (Fig. 4D). Wild type amino acids were basically the only ones identified in these positions (range, 99.6% to 100%) in the NL4-3 reference virus, while various mutations were detected and different frequencies in the patient-derived samples (Fig. 4D and 4E). Finally, a series of minor amino acid substitutions were identified repeatedly in the patient-derived samples (08-198, 12-069, and 12-596) at frequencies below the limit of detection of Sanger sequencing, e.g., A98G (3.3% ± 0.8, mean ± standard deviation) and K103N (8.5% ± 2.6) in virus 12-069 (Fig. 4E).
Comparison of drug susceptibility determination using deep sequencing (DEEPGEN™HIV) to the current standard HIV-1 genotypic assays based on population sequencing. As described above, a multitude of HIV-1 drug-resistance methods have been developed but only a few have been deployed in the clinical setting, including several genotypic tests based on population sequencing (10, 43, 44). Here, plasma samples from 166 treatment-experienced HIV-infected individuals from two cohorts of patients (Seville and Madrid) were analyzed using standard population-based HIV-1 genotyping and the novel deep-sequencing assay. The mean CD4+ T-cell count in these patients was 353 cells/µl (interquartile range, IQR: 190 - 488) and their mean plasma viral load was 69,459 copies/ml (IQR: 5,218 - 87,000). The length of the antiretroviral treatment varied among individuals, averaging 8.2 years (ranging from 1989 to 2012) and included a diversity of treatment regimens using a multitude of PIs, NRTIs, NNRTIs, raltegravir and/or maraviroc. Altogether a total of 1,701 mutations (379 and 1,322 in the Seville and Madrid cohorts, respectively) in positions associated with drug resistance were detected by both methodologies (i.e., 954 in the protease, 613 in the RT, and 134 in the integrase)(Fig. 5A). As expected, all the drug resistance mutations identified by population sequencing were also detected by deep sequencing, while an additional 1,073 drug resistance mutations (337 and 736 in the Seville and Madrid cohorts, respectively) were detected only by deep sequencing (i.e., 511 in the protease, 1,015 in the RT, and 97 in the integrase)(Fig. 5A). Overall, the difference in the numbers of drug resistance mutations detected by both methods was significant, even when the mutations were quantified by drug class, i.e., an average of 3.1, 2.8, and 0.6 additional mutations associated with PI, RTI, and INI, respectively, were detected by deep sequencing compared to population sequencing (Paired t test, p < 0.0001)(Fig. 5A). Interestingly, additional PI (mean 5.3 vs. 3.1) and RTI (3.1 vs. 1.1) but not INI (0.9 vs. 0.8) resistance mutations were identified by deep sequencing in patients from the Seville cohort (Fig. 5A). Unlike some of the HIV-infected individuals from the Madrid cohort, these patients from Seville were not treated with raltegravir. The slight difference in the number of INI mutations detected by deep sequencing in the Seville patients corresponded to the identification of the L101I polymorphism, which has been associated with decreased susceptibility to the INI dolutegravir (87).
Supplemental Figure S4 shows a comparison of the frequency of amino acids detected by population and deep sequencing in a virus carrying multiple mutations associated with resistance to PI, RTI, and INI.

Comparison of HIV-1 coreceptor tropism data obtained with DEEPGEN™HIV to other phenotypic or genotypic HIV-1 coreceptor tropism assays. Plasma samples from 114 HIV-infected individuals screened to be treated with a maraviroc-containing regimen, a subset of samples from the same Seville and Madrid cohorts of patients, were analyzed using the novel deep sequencing-based HIV-1 coreceptor tropism assay. These results were compared with a series of genotypic (population sequencing) and phenotypic [ESTA (42), Trocai (73), and VERITROP (23)] HIV-1 tropism tests. Only samples with results from all the different tests (i.e., 38 and 76 from Seville and Madrid, respectively) were included. Hierarchical clustering analysis grouped the different HIV-1 coreceptor tropism determinations based on their ability to detect R5 and non-R5 (X4, dual tropic, and/or dual mixed, D/M) sequences (Fig. 5B). Plasma samples from the Seville cohort were from patients participating in a study to evaluate the use of an 8-day maraviroc monotherapy clinical test (MCT), the rationale being that plasma viral load in patients carrying non-R5 viruses will fail to decrease at least one log10, or being undetectable in subjects with <1,000 copies/ml, after receiving maraviroc in this period of time (71, 72). Overall, in this cohort of patients, the concordance and agreement was high among the different HIV-1 tropism methods, with DEEPGEN™HIV (Geno2Pheno FPR 3.5%) showing good agreement with population sequencing analyzed using Geno2Pheno/FPR 10% (84.4%, kappa = 0.37), MCT (82.9%, kappa = 0.44), and ESTA (80%, kappa = 0.47) (Fig. 5B). Similar concordance was observed between ESTA and MCT (85%, kappa = 0.64) and between population sequencing/Geno2Pheno/FPR 10% and MCT (82.9%, kappa = 0.51). Interestingly, a perfect agreement (100%, kappa = 1) was observed between MCT and Trocai, the phenotypic HIV-1 tropism assay performed in Seville (data not shown). Using slightly older samples from the Madrid cohort, DEEPGEN™HIV showed excellent agreement with the original Trofile assay (91.7%, kappa = 0.79) and good concordance with VERITROP (79.8%, kappa = 0.58) or population sequencing/Geno2Pheno/2.5%-5.75% (74.4%, kappa = 0.37) (Fig. 5B).
Concordance between the original Trofile assay and population sequencing/Geno2Pheno/2.5%-5.75% was comparable (80.3%, kappa = 0.54), while a 73.7% (kappa = 0.5) agreement was observed between the two phenotypic tests (VERITROP and Trofile) in baseline samples from these patients (Fig. 5B).
DISCUSSION

Almost twenty-five years since drug-resistant HIV-1 strains were first reported following initial treatments with AZT (88), and twenty-eight additional antiretroviral drugs later, detection and characterization of drug-resistant variants has become essential in the clinical care of HIV-infected patients (3, 6-8). A multitude of genotypic (sequencing-based) and phenotypic (cell-based) assays have been developed to identify and quantify HIV-1 drug resistance, and in the case of CCR5 receptor antagonists, determining HIV-1 coreceptor tropism (8, 34, 35, 89). Most of these HIV-1 drug resistance or tropism tests correspond to methods developed and used in research settings; however, a few selected assays are commercially available. All current commercial genotypic HIV-1 drug resistance assays are based on population (Sanger) sequencing (10, 43, 44), limiting the detection of minority variants to 20% of the viral population (44-48). In the case of HIV-1 tropism determination, only a couple of ultrasensitive deep sequencing-based genotypic tests are currently used in the clinical setting (63, 66).

Here, we have developed characterized, and validated the first all-inclusive HIV-1 genotyping and tropism test (DEEPGEN™ HIV) based on deep sequencing, which is able to detect minority drug resistant and non-R5 HIV-1 variants in a single assay.

Deep sequencing is here to stay. The efficiency, sensitivity, and cost-effectiveness of deep sequencing-based methodologies have allowed their use in a multitude of biological fields, including the sequencing of whole genomes of animals and plants (90, 91), targeted studies on polymorphisms related to various genetic disorders (92) and cancer (93, 94), and research aimed to decode the genome of bacteria (95, 96), fungi (97), and viruses (98, 99). We have also observed an exponential increase in the number of HIV studies based on deep sequencing during the last five years. Many of them related to HIV-1 pathogenesis, transmission, and evolution (98, 100) but the majority aimed to the detection of minority non-R5 HIV-1 variants (24, 26, 36, 63, 65-67, 75, 101-104) or low frequency drug-resistant variants that could lead to treatment failure (51, 60-62, 84, 105-108). However, none of the methods used in these studies have been able to simultaneously evaluate
resistance to antiretroviral drugs targeting Gag, protease, reverse transcriptase, and integrase and the ability of the virus to use the CCR5 coreceptor to enter the host cell, in a single assay. DEEPGEN™HIV was designed to use the same two overlapping amplicons covering the 3’end of Gag and the entire pol gene used in the phenotypic HIV-1 assay VIRALARTS™HIV (13), with the addition of the V3 region of the gp120 in the env gene, which allows the possibility to reflex from the ultrasensitive genotypic test (DEEPGEN™HIV) to the comprehensive cell-based VIRALARTS™HIV using the same PCR products. The 97% amplification success of the two overlapping fragments covering the gag-p2/NCp7/p1/p6/pol-PR/RT/IN region from plasma samples with ≥1,000 copies/ml of HIV RNA (13) was here matched by the successful amplification (92%) of the env-V3 fragment from plasma samples with similar viral load, which is comparable to the success rate observed in other studies (23, 62, 109). Moreover, we have been able to PCR amplify intermittently these long fragments from plasma samples with viral loads between 50 to 1,000 copies/ml, depending on the quality of the clinical specimen (data not shown). The RT-PCR step was not only reproducible but ensured successful amplification of samples from diverse HIV-1 subtypes while avoiding amplification of non-specific products from endogenous or any of the related viruses tested.

Any methodology based on PCR amplification and deep sequencing endures the same fundamental problem, that is, errors are introduced during the process (67, 84, 85, 110, 111). In fact, a limited number of errors are introduced during the PCR step, with most of the errors produced during deep sequencing, mainly insertions and deletions (84). Thus, here it was important to calculate the intrinsic error rate of the entire system since it would undoubtedly affect the limit of detection of the assay. The combined error rate, i.e., point mutations, insertions, and deletions, of our new HIV-1 genotyping and coreceptor tropism assay was 0.39% and 0.37% for the gag-p2/NCp7/p1/p6/pol-PR/RT/IN and env-V3 regions, respectively. These values were similar to average error rate rates previously reported in the HIV-1 pol or env genes using other deep sequencing platforms, ranging from 0.3% to 0.98% (62, 66, 67, 84, 85, 110, 112-115). Approximately a 10-fold higher combined error rate was observed in nucleotide positions associated with homopolymeric regions, resembling
findings by other studies describing similar difficulties when sequencing regions with identical consecutive nucleotides (61, 62, 84, 85, 113). More importantly, some of these homopolymeric regions encompass positions associated with resistance to antiretroviral drugs, which could represent a challenge during the interpretation of minority mutations detected at these positions. Using computational methods we were able to discern between genuine genetic variation and errors introduced during the sequencing process. For example, codon 193 in the integrase showed the highest error rate in the entire HIV-1 genomic region analyzed, i.e., 9.6% indels and 10.5% overall error rate. In HIV-1NL4-3 the 25 nucleotides upstream of this position correspond to a series of identical consecutive nucleotides (Fig. 6), which could explain the elevated error rate in this particular region. However, the Ion Torrent software was able to filter most of the reads with sequencing errors, at the expense of reducing approximately 5-fold the coverage around this region (Fig. 2B, Fig. 6) but still above the 1,000 reads required to guarantee the detection of a minor variant present at least at 1% of the population (85). Moreover, the Variant Caller in the DEEPGEN™ Software Tool Suite filtered out all the indels at position 193 and, as a consequence, we were able to accurately and repeatedly determine the correct amino acid in all the wild-type HIV-1NL4-3 sequences and the different variants from patient-derived samples (Fig. 6).

The approach of using a deep sequencing-based assay to detect minority HIV-1 variants relies on the capacity of the methodology to truthfully detect low-frequency sequences from usually complex virus populations. The limit of detection of these assays is a combination of the intrinsic error rate of the method and the analytical sensitivity of the test. As described above, the ability of our HIV-1 genotyping and coreceptor tropism assay to discern between genetic variation and sequencing error was calculated as 1%. The analytical sensitivity of the assay, determined using mixtures of plasmid DNA, seems to suggest that the deep sequencing-based test is able to detect minority variants when present as low as 1% of the population; however, mutations were detected in the HIV-infected plasma mixtures at a 5% frequency but not at 1%. Five percent is the limit of detection previously calculated studying HIV with 454 pyrosequencing (62, 66). It is possible that, although
theoretically sequences at a frequency of 1% could be identified, bias for the majority members of the population during RT-PCR amplification, library preparation, and/or sequencing could be limiting the ability of the assay to efficiently detect minority variants below 5%. Additional studies, using a larger number of clinical samples will be needed to verify this limit of detection. Nevertheless, our novel assay was able to reproducibly detect HIV-1 minority variants at frequencies at least four-fold lower than the 20% reported for the standard Sanger-based sequencing (16, 44-48).

A multitude of studies have compared the efficacy of phenotypic and genotypic assays to detect and quantify HIV-1 drug resistance (116-120) or coreceptor tropism (38, 121-124). While there is usually high concordance between drug-resistance methodologies (13, 119, 120), in general, population-based sequencing HIV-1 tropism tests are less sensitive and less specific than phenotypic assays (8, 41). Here, DEEPGEN™HIV identified all the drug resistance mutations detected by the standard population sequencing-based HIV-1 genotyping tests used to monitor 166 individuals in two well-established cohorts of patients. More important, over 1,000 additional minor drug resistance mutations, at frequencies between 1% and 20%, were detected using the novel deep sequencing-based assay. In the case of the detection of minority non-R5 variants and determination of HIV-1 coreceptor tropism, DEEPGEN™HIV showed good concordance with both ESTA and the original Trofile assay (80% and 92%, respectively), similar to the agreement of these phenotypic tests with 454-based sequencing systems used to determine HIV-1 tropism (80% to 87%)(24, 36, 63-66, 104, 125, 126). Deep sequencing-based HIV-1 coreceptor tropism assays have shown good concordance with virological response to CCR5 antagonists such as maraviroc (36, 63); thus, more sensitive HIV-1 genotyping methods based on deep sequencing may also help detecting minority drug-resistant variants earlier, possibly improving the designing of antiretroviral treatment schemes.

As described above, DEEPGEN™HIV detected additional (minority) HIV-1 drug resistance mutations and non-R5 variants compared to population sequencing; however, the clinical relevance of these minority members of
the viral population is still under debate (33, 51, 55, 57, 127-129). It is reasonable to assume that, under the appropriate selection (i.e., drug pressure), minority variants carrying drug resistance mutations will eventually outcompete other members of the viral population, and that the earlier these mutations are detected, the sooner the proper strategy can be defined to control the growth of these viruses. Although, the threshold for the identification of significant low-level variants may be mutation and/or antiretroviral drug dependent (32), early detection of drug resistant HIV-1 minority variants, or non-R5 in the case of CCR5-antagonists, may contribute to the prediction of clinical or treatment outcome (32, 36, 57, 63, 127). Additional effort, most likely in the form of longitudinal studies, is needed to better understand the clinical relevance of these minority HIV-1 variants. Here we have developed, characterized, and validated a novel all-inclusive and ultrasensitive HIV-1 genotyping assay based on deep sequencing to simplify the monitoring of patients infected with HIV-1. This first-in-class methodology accurately provides drug resistance information for all protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism, in a single, more efficient, rapid, and affordable test. This unique assay may provide the platform not only to predict clinical outcome in HIV-infected individuals -whom otherwise could fail treatment regimens- but in the development and clinical assessment of novel antiretroviral drugs.
ACKNOWLEDGEMENTS

We thank Dr. Eric J. Arts (HIV-1 isolates), Dr. Michael R. Jacobs (BKV, CMV, HSV-1, HSV-2, and VZV), Dr. David Canaday (EBV), and Dr. Donald Anthony (HBV and HCV) from Case Western Reserve University (Cleveland, OH) for providing access to the respective RNA or DNA viruses. We also thank Dr. Benigno Rodriguez and the Case Western Reserve University/University Hospitals Center for AIDS Research (NIH P30 AI036219) and Dr. Vicente Soriano (Hospital Carlos III, Madrid, Spain) for providing clinical samples for the characterization and validation studies. We are grateful to Dr. Michael D. Miller and Dr. Kirsten L. White (Gilead Sciences, Inc., Foster City, CA) for providing some of the plasmid mixtures used in the analytical sensitivity experiments. E.R-M and M.L. were supported by Redes Telemáticas de Investigación Cooperativa en Salud (RETICS), Red de Investigación en SIDA (RIS RD12/0017/0029).

M.E.Q-M designed the study, collected and assembled the data, wrote and drafted the manuscript. R.M.G., A.M.M., and D.W. performed all cell culture, molecular, and sequencing experiments. J.A., F.F., M.E.Q-M and D.L.R. designed the DEEPGEN™ Software Tool Suite, while J.A. and F.F. developed it. E.R-M and M.L. performed the studies associated with the Seville cohort, including population sequencing, MCT, and Trocaí analyses. R.M.G, C.L.S., and M.E.Q-M contributed to the overall analysis of the data. All authors read and approved the final manuscript.
REFERENCES


915 91. Scally A, Duthieil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T,
916 Marques-Bonet T, McCarthy S, Montgomery SH, Schwalie PC, Tang YA, Ward MC, Xue Y, Yngvadottir
917 B, Alkan C, Andersen LN, Ayub Q, Ball EV, Beal K, Bradley BJ, Chen Y, Clee CM, Fitzgerald S, Graves TA,
919 JC, Munch K, O'Connor TD, Phillips AD, Prado-Martinez J, Rogers AS, Sajjadian S, Schmidt D, Shaw K,
921 Dermitzakis ET, Eichler EE, Flicek P, Goldman N, Mundy NI, Ning Z, Odom DT, Ponting CP, Quail MA,
923 Insights into hominid evolution from the gorilla genome sequence. Nature 483:169-175.
925 Integrated with Clinicopathological Information Reveals Novel and Rare Mutations in Atypical,
926 Suspected and Unknown Cases of Alport Syndrome or Proteinuria. PLoS One 8:e76360.
927 93. Matsushita H, Vesely MD, Koboldt DC, Rickert CG, Uppaluri R, Magrini VJ, Arthur CD, White JM, Chen
931 94. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, Kalyana-Sundaram S, Sam L,
932 Balbin OA, Quist MJ, Barrette T, Everett J, Siddiqui J, Kunju LP, Navone N, Araujo JC, Troncoso P,
933 Logothetis CJ, Innis JW, Smith DC, Lao CD, Kim SY, Roberts JS, Gruber SB, Pienta KJ, Talpaz M,
934 Chinnaiyan AM. 2011. Personalized oncology through integrative high-throughput sequencing: a pilot
938 during Staphylococcus aureus infection involves mutation of the essential two component regulator
940 96. Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR,
941 Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK.
944 sequencing of a fungal genome using only SOLiD short read data: verification on Aspergillus oryzae
948 SH, Sam NE, Haynes BF, Simek M, Burton DR, Koff WC, Doria-Rose NA, Connors M, Mullikin JC, Nabel
949 GJ, Roederer M, Shapiro L, Kwong PD, Mascola J, Simek M, Burton DR, Koff WC, Doria-Rose NA, Connors M, Mullikin JC, Nabel
950 99. Quan PL, Firth C, Conte JM, Williams SH, Zambrana-Torrelio CM, Anthony SJ, Ellison JA, Gilbert AT,
951 Kuzmin IV, Nieszoda M, Osinubi MO, Recueno S, Markotter W, Breiman RF, Kalemba L, Malekani J,
953 DA, Moran D, Ngam S, Akaibe D, Agwanda B, Briese T, Epstein JH, Daszak P, Rupprecht CE, Holmes EC,
954 Lipkin WI. 2013. Bats are a major natural reservoir for hepaciviruses and pegiviruses. Proceedings of the
956 100. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, Berlin AM, Malboeuf CM,
958 Dudek T, Tully D, Newman R, Axtens KL, Gralenko AD, Battis L, Kemper M, Zeng Q, Shea TP, Guuja S,
959 Zedlack C, Gasser O, Brander C, Hess C, Gunthard HF, Brumme ZL, Brumme CJ, Bazner S, Rychert J,


1056


Table 1. Sensitivity of RT-PCR amplification of the C2V3 region of the HIV-1 env gene (480 nt)

<table>
<thead>
<tr>
<th>Viral load (copies/ml)</th>
<th>% Positive samples by RT-PCR (No. of positive samples/total No. of samples tested)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,001 – 5,000</td>
<td>95 (19/20)</td>
</tr>
<tr>
<td>5,001 – 10,000</td>
<td>85 (17/20)</td>
</tr>
<tr>
<td>&gt; 10,000</td>
<td>95 (37/39)</td>
</tr>
</tbody>
</table>

RT-PCR amplification of patient-derived env fragments was performed with plasma samples (n = 79) from HIV-infected individuals with viral loads ranging from 1,000 to >10,000 copies of viral RNA/ml as described in Materials and Methods.
### Table 2. Error rate distribution

<table>
<thead>
<tr>
<th>Genomic region</th>
<th>Codon</th>
<th>Sequence</th>
<th>Total Error Rate (mean ± SD)</th>
<th>Point Mutations (mean ± SD)</th>
<th>Indels (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Gag/Pol</em></td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.0039 ± 0.0001</td>
<td>0.0017 ± 0.0001</td>
<td>0.0022 ± 0.0001</td>
</tr>
<tr>
<td><em>env-V3</em></td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.0037 ± 0.0002</td>
<td>0.0012 ± 0.0001</td>
<td>0.0025 ± 0.0002</td>
</tr>
<tr>
<td>Protease</td>
<td>L10</td>
<td>cgacccCTgct</td>
<td>0.0353 ± 0.0086</td>
<td>0.0022 ± 0.0002</td>
<td>0.0331 ± 0.0051</td>
</tr>
<tr>
<td></td>
<td>M46</td>
<td>accaaaaATGata</td>
<td>0.0920 ± 0.0056</td>
<td>0.0013 ± 0.0001</td>
<td>0.0907 ± 0.0043</td>
</tr>
<tr>
<td></td>
<td>F53</td>
<td>agagTTTatc</td>
<td>0.0436 ± 0.0034</td>
<td>0.0021 ± 0.0002</td>
<td>0.0415 ± 0.0022</td>
</tr>
<tr>
<td>RT</td>
<td>F77</td>
<td>agatTTTCag</td>
<td>0.0205 ± 0.0028</td>
<td>0.0018 ± 0.0001</td>
<td>0.0187 ± 0.0011</td>
</tr>
<tr>
<td></td>
<td>K101</td>
<td>gtttaAACag</td>
<td>0.0295 ± 0.0044</td>
<td>0.0024 ± 0.0002</td>
<td>0.0271 ± 0.0032</td>
</tr>
<tr>
<td></td>
<td>V179</td>
<td>atatCTCtc</td>
<td>0.0756 ± 0.0122</td>
<td>0.0036 ± 0.0002</td>
<td>0.0720 ± 0.0091</td>
</tr>
<tr>
<td></td>
<td>G190</td>
<td>gtaaGATcct</td>
<td>0.0221 ± 0.0036</td>
<td>0.0012 ± 0.0001</td>
<td>0.0209 ± 0.0021</td>
</tr>
<tr>
<td>Integrase</td>
<td>G193</td>
<td>attgggGGGct</td>
<td>0.1048 ± 0.0277</td>
<td>0.0089 ± 0.0003</td>
<td>0.0959 ± 0.0145</td>
</tr>
<tr>
<td></td>
<td>V3</td>
<td>R11</td>
<td>0.0256 ± 0.0037</td>
<td>0.0025 ± 0.0002</td>
<td>0.0231 ± 0.0020</td>
</tr>
</tbody>
</table>

* Number of combined PCR and sequencing errors, i.e., point mutations, insertions and deletions (indels) per read calculated using Segminator II (76). Mean and standard deviation (SD) values obtained from 10 independent sequences are indicated.

*b* HIV-1 genomic region analyzed. *Gag/Pol* and *env-V3* correspond to the *gag*-p2/NCp7/p1/p6/PR/RT/IN and V3 region of the gp120 in the envelope gene, respectively.

*c* Codons associated with resistance to antiretroviral drugs determined to have total error rate values above 1%.

*d* Nucleotide sequence based on the population sequencing of the HIV-1 NL4-3 clone (13), which around these codons was identical to the HIV-1 HXB2 reference sequence (GenBank accession number K03455). The respective codon is indicated in uppercase while the nucleotide position associated with the elevated error rate (>1%) is shown in **bold** and *italic*. Position numbering is relative to the HIV-1 HXB2 reference sequence.
FIGURE LEGENDS

Figure 1. Overview of the protocol for the novel HIV-1 genotyping and coreceptor tropism assay (DEEPGEN™HIV). Three PCR products corresponding to the gag-p2/NCp7/p1/p6/pol-PR/RT/IN- (1,657 bp and 2,002 bp) and env-C2V3- (480 bp) coding regions of HIV-1 were used to construct a multiplexed library for shotgun sequencing on the Ion PGM™. Signal processing and base calling was performed with Torrent Analysis Suite version 3.4.2 and sequences analyzed using DEEPGEN™ Software Tool Suite. The HIVdb Program Genotypic Resistance Interpretation Algorithm from the Stanford University HIV Drug Resistance Database (http://hivdb.stanford.edu) and Geno2Pheno (27) were used to infer the levels of susceptibility to PI, RTI, and INI and for HIV-1 coreceptor tropism determination, respectively.

Figure 2. Error rate determination. (A) The pNL4-3-hRluc plasmid containing the entire genome of the wild type HIV-1NL4-3 strain (69) was transformed into bacteria and the gag-p2/NCp7/p1/p6/pol-PR/RT/IN and env-C2V3 fragments were PCR amplified and deep sequenced from ten individual colonies. Reads from each individual NL4-3 clone were independently mapped to the pNL4-3-hRluc reference sequence (13, 69) using Segminator II (76). (B) Coverage, i.e., number of reads per nucleotide position, for the ten NL4-3 clones. (C) Overall (point mutation, insertions, and deletions) error rate per nucleotide position calculated using a Phred Quality Score of 20. The mean error rates for the gag-p2/NCp7/p1/p6/pol-PR/RT/IN and env-V3 regions are indicated in green, while the minimum threshold to detect mutations in minority HIV-1 variants with this novel assay (1%) is indicated by a dashed red line. (D) Overall error rate in positions associated with drug resistance. Only codon changes with error rates above 1% are indicated, i.e., L10, M46, and F53 in the protease; F77, K101, V179, and G190 in the RT; G193 in the integrase; and amino acid 11 in the V3 region. Homopolymeric regions, defined as four or more identical consecutive nucleotides, are indicated as vertical grey bars.
Figure 3. The ability of the novel HIV-1 genotyping and coreceptor tropism assay to detect drug resistance mutations (in the gag-p2/NCp7/p1/p6/pol-PR/RT/IN fragment) and non-R5 variants (in the env-V3 region) within mixtures of viral populations was used to calculate the analytical sensitivity of the test. (A) A gag-p2/NCp7/p1/p6/pol-PR/RT/IN PCR product was obtained from an antiretroviral-experienced patient (08-180) and used to construct p2-INT recombinant viruses based on the yeast cloning method to maintain the HIV-1 quasispecies (13, 68). This plasmid preparation contained the pol gene from the patient and the env gene from the CXCR4-tropic HIV-1$_{194}$ strain, i.e., **08-180 pol/NL43-(X4)env**. Plasmid **NL4-3 pol/YU2(R5)env** contains the genome of the wild-type HIV-1$_{NL4-3}$ virus carrying the env gene from the R5 HIV-1$_{YU2}$ virus (69). A series of plasmid mixtures were created by mixing 0.1%, 1%, 2%, 3%, 5% and 10% of the 08-180pol/NL43-(X4)env plasmid with the corresponding amount of the NL4-3pol/YU2(R5)env plasmid at a final concentration of 0.1 ng/ml. DNA from the entire plasmid mixtures, together with the two individual plasmids as controls (100%), was purified and deep sequenced as described in Materials and Methods. (B) Plasma containing a patient-derived multidrug-resistant gag-p2/NCp7/p1/p6/pol-PR/RT/IN recombinant virus constructed using the X4 HIV-1$_{NL4-3}$ backbone (08-194) and a wild-type HIV-1$_{NL4-3}$ virus carrying the env gene from the R5 HIV-1$_{YU2}$ virus were mixed at 0%, 1%, 5% and 100% of the 08-194 virus at a final concentration of 100,000 copies/ml. The frequency of each mutation detected in the original population at ≥1% of the population, threshold calculated based on the intrinsic error rate of the assay, is indicated (mean ± standard deviation, from quadruplicate experiments in the case of the experiment using mixtures of HIV-infected plasma). Amino acid substitutions detected at a frequency below 90% of the population are indicated in red.

Figure 4. Assay reproducibility. (A) Samples from two antiretroviral-naïve (NL4-3 and 12-596) and two antiretroviral-experienced (08-198 and 12-069) individuals were RT-PCR amplified in triplicate, each amplicon barcoded four times, two DNA libraries prepared, and sequenced in duplicated for a total of 48 sequences per sample. (B) Neighbor-joining phylogenetic trees constructed using reads with a frequency >1 corresponding to
105 bp fragments from the protease, RT, integrase, and V3 regions. Each color-coded dot represents a unique variant, frequency is not depicted. Bootstrap resampling (1,000 data sets) of the multiple alignments tested the statistical robustness of the trees, with percentage values above 75% indicated by an asterisk. s/nt, substitutions per nucleotide. (C) Pearson correlation coefficient was used to determine the strength of association between the frequency of each nucleotide at each position among the 16 sequences obtained for each one of triplicate amplicons (n = 48) for all four viruses in the gag-p2/NCp7/p1/p6/pol-PR/RT/IN and env-V3 regions. Over 135,000 and 5,000 points are included in each one of the gag-p2/NCp7/p1/p6/pol-PR/RT/IN and env-V3 plots, respectively. r, correlation coefficient; p, two-tailed p value. (D) Amino acids detected in codons associated with drug resistance in the protease, RT, and integrase regions according to the IAS-USA (77). Drug resistance mutations with a frequency ≥20% (blue, green, or purple), <20% (yellow) or any other amino acid changes (grey) are indicated. Only amino acid substitutions with a frequency >1% are depicted. (E) Frequency of amino acids in positions associated with drug resistance (gag-p2/NCp7/p1/p6/pol-PR/RT/IN) or X4 variants (env-V3) found in any of the four samples. Each dot represents the mean and 95% confidence intervals, with the exception of the insert (sample 12-069) where each dot indicates the frequency of amino acids detected in each of the 48 replicates, including their mean ± standard deviation.

Figure 5. Comparison of the novel HIV-1 genotyping and coreceptor tropism assay with other HIV-1 genotypic phenotypic tests. Plasma samples from 166 treatment-experienced HIV-infected individuals from two cohorts of patients (Seville and Madrid) were analyzed as described in Materials and Methods. (A) Top two plots compare the number of drug resistance mutations detected by standard population (Sanger) and deep sequencing in each patient. The total numbers of drug resistance mutations identified by each sequencing method are indicated. The difference in the numbers of drug resistance mutations (mean ± standard deviation) detected by population and deep sequencing in the protease (PR), reverse transcriptase (RT), and integrase (INT) regions is indicated in the bottom two plots. Mean values are indicated. Statistically significantly differences are marked.
with an asterisk (Paired t test, \( p < 0.0001 \)). Deep, deep sequencing; Pop, population sequencing. (B) Hierarchical clustering analysis was used to group the different HIV-1 coreceptor tropism determinations by similarity. Dendograms were calculated using the Euclidean distance and Complete cluster methods with 100 bootstrap iterations as described (http://www.hiv.lanl.gov/content/sequence/HEATMAP/heatmap.html). Bootstrap values above 60% are indicated with an asterisk. Green and red blocks indicate the absence or presence of non-R5 (X4) viruses, respectively, as determined by each assay. Concordance between DEEPGEN™HIV and the other HIV-1 coreceptor tropism assays are indicated. G2Pclin, Geno2Pheno with a FPR of 10%; MCT, 8-day maraviroc monotherapy clinical test (71, 72); ESTA, enhanced sensitivity Trofile assay (42); Trofile, the original version of the Trofile™ assay (19); VeriTrop, phenotypic HIV-1 tropism assay (23); G2Pmot, Geno2Pheno with a FPR of 2.5% and 5.75% based on optimized cutoffs associated with the analysis of clinical data from MOTIVATE (63).

**Figure 6.** Coverage around codon 193 of the HIV-1 integrase. First panel summarizes the coverage of the 10 HIV-1\textsubscript{NL4-3} sequences obtained from the clones used to calculate the intrinsic error rate of the assay (grey lines) and 6 HIV-1\textsubscript{NL4-3} sequences obtained from the positive controls in six regular sequencing runs (blue lines). The second panel includes the coverage of 8 patient-derived HIV-1 sequences (clinical samples labeled as 13-xxx). Grey bar indicates the homopolymeric region at and upstream of codon 193, i.e., 25 nucleotides in bold in the nucleotide alignment. Red vertical line depicts codon 193.
Figure 1. Gibson et al 2013
Figure 2. Gibson et al 2013
A. Analytical Sensitivity Mixing Plasmid DNA

B. Analytical Sensitivity Mixing Plasma Spiked with HIV-1

Figure 3. Gibson et al 2013
A. Experiment Design

- Sample: 1
- PCR (3x): 3
- Barcoding (4x): 12
- Emulsion (2x): 24
- Sequencing (2x): 48

B. Genetic Diversity

PR, RT, INT, V3

C. Reproducibility (nt)

gag-p2/NCP7/p1/p6/pol-PR/RT/IN, V3

Figure 4. Gibson et al 2013
A. Correlation of DEEPGEN with population sequencing

Seville Cohort (n = 75)

Madrid Cohort (n = 91)

B. Correlation of DEEPGEN with other HIV-1 tropism tests

Seville Cohort

Madrid Cohort

DEEPGEN

G2Pclin 84.4%
MCT 82.9%
ESTA 80%

DEEPGEN

Trofile 91.7%
VeriTrop 79.8%
G2Pmot 74.4%