

# Results of External Quality Assessment for Proviral DNA Testing of HIV Tropism in the Maraviroc Switch Collaborative Study

Elise Tu,<sup>a</sup> Luke C. Swenson,<sup>b</sup> Sally Land,<sup>d</sup> Sarah Pett,<sup>a,c</sup> Sean Emery,<sup>a</sup> Kat Marks,<sup>e</sup> Anthony D. Kelleher,<sup>a,c,e</sup> Steve Kaye,<sup>f</sup> Rolf Kaiser,<sup>g</sup> Eugene Schuelter,<sup>g</sup> Richard Harrigan,<sup>b</sup> on behalf of the MARCH Laboratory Group and the MARCH Study Group

The Kirby Institute, University of New South Wales, Sydney, NSW, Australia<sup>a</sup>; BC Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada<sup>b</sup>; Immunology, HIV and Infectious Diseases Clinical Services Unit, St Vincent's Hospital, Sydney, Australia<sup>c</sup>; NRL, Fitzroy, Victoria, Australia<sup>d</sup>; Centre for Applied Medical Research (AMR), St Vincent's Hospital, Sydney, NSW, Australia<sup>e</sup>; Imperial College London, Norfolk Place, London, United Kingdom<sup>f</sup>; Institut für Virologie, der Universität zu Köln, Cologne, Germany<sup>g</sup>

The **Maraviroc Switch** collaborative study (MARCH) is a study in aviremic patients on stable antiretroviral therapy and utilizes population-based sequencing of proviral DNA to determine HIV tropism and susceptibility to maraviroc. An external quality assessment (EQA) program was implemented to ensure competency in assessing the tropism of clinical samples conducted by MARCH laboratories ( $n = 14$ ). The MARCH EQA has three prestudy phases assessing V3 loop sequencing and tropism determination using the bioinformatic algorithm *geno2pheno*, which generates a false-positive rate (FPR). DNA sequences with low FPRs are more likely to be from CXCR4-using (X4) viruses. Phase 1 of the EQA involved chromatogram interpretation. Phases 2, 2/3, and 3 involved patient and clonal samples. Clinical samples used in these phases were from treatment-experienced HIV-infected volunteers; 18/20 had viral loads of  $<50$  copies/ml, and 10/15 were CXCR4-tropic on prior phenotyping. All samples were tested in triplicate, and any replicate with a *geno2pheno* FPR of  $<10\%$  was designated X4. Performance was deemed adequate if  $\leq 2$  R5 and  $\leq 1$  X4 specimens were miscalled. For several clinical samples in the EQA, triplicate testing revealed marked DNA variability (FPR range, 0 to 96.7%). Therefore, a consensus-based approach was employed for each sample, i.e., a median FPR across laboratories was used to define sample tropism. Further sequencing analysis showed mixed viral populations in the clinical samples, explaining the differences in tropism predictions. All laboratories passed the EQA after achieving predefined competence thresholds in either of the phase 2 rounds. The use of clinical samples from patients resembling those who were likely to be screened in the MARCH, coupled with triplicate testing, revealed inherent DNA variability that might have been missed if single or duplicate testing and/or clonal samples alone were used. These data highlight the importance of intensive EQA of tropism laboratories before embarking on clinical studies. (This study has been registered at [ClinicalTrials.gov](http://ClinicalTrials.gov) under registration no. NCT01384682 [<http://www.clinicaltrials.gov/ct2/show/study/NCT01384682?term=NCT01384682&rank=1>].)

**M**araviroc (MVC), an approved antiviral agent for the treatment of HIV-1 infection, is a small molecule blocker to receptor 5 of the CC chemokine (CCR5) group (1). MVC is a host-directed therapy and is an effective antiviral agent in combination with other antiretroviral agents in patients identified as having a CCR5-tropic (R5) virus (2).

The sensitivity of the host's HIV to CCR5 antagonists, including MVC, has been determined using a variety of testing platforms for the assessment of viral tropism. The most widely utilized test for tropism is the phenotypic recombinant virus assay, the enhanced-sensitivity Trofile assay (ESTA) (3). The assay is typically performed in viremic patients (ideally those with  $>1,000$  copies/ml) and can accurately discriminate between MVC responders and nonresponders (4).

Genotypic tropism testing using proviral DNA is increasingly available. A prediction of the likely coreceptor usage of a patient's viral population is determined through the amplification, sequencing, and analysis of the HIV *env* V3 loop followed by Web-based bioinformatic algorithms, e.g., *geno2pheno* [coreceptor] (5) or the position-specific scoring matrix (PSSM) (6). However, there is still a paucity of prospective data on the clinical cutoffs that should be applied using this tropism testing platform. Data from retrospective reanalysis of the MVC licensing studies suggest a *geno2pheno* false-positive rate (FPR) cutoff of approximately 5.75% (7, 8), while the European Consensus Group guidelines (9) recommend cutoffs as high as 20% in some cases. Furthermore,

these cutoffs were optimized with data from plasma-based approaches, but genotypic tropism testing can also be performed using proviral DNA (10–14).

Despite the lack of validation for tropism testing of proviral DNA in aviremic patients exposed to MVC, recent European Consensus Group guidelines (9) have been published recommending how many times a patient sample should be sequenced (single, duplicate, or triplicate) and how the *geno2pheno* FPR should be applied to define tropism in aviremic individuals. For population genotyping using proviral DNA, the study group recommended FPR cutoffs of 10% or 20% in triplicate and single testing, respectively (9). DNA sequences with FPRs below the cutoff are classified as being from CXCR4-using (X4) viruses.

The **Maraviroc Switch** collaborative study (MARCH) ([ClinicalTrials.gov](http://ClinicalTrials.gov) identifier: NCT01384682; see <http://www>

Received 24 February 2013 Returned for modification 25 March 2013

Accepted 10 April 2013

Published ahead of print 17 April 2013

Address correspondence to Elise Tu, [etu@kirby.unsw.edu.au](mailto:etu@kirby.unsw.edu.au), or Luke C. Swenson, [lsvenson@cfcenet.ubc.ca](mailto:lsvenson@cfcenet.ubc.ca).

E.T. and L.C.S. contributed equally to this article.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.00510-13

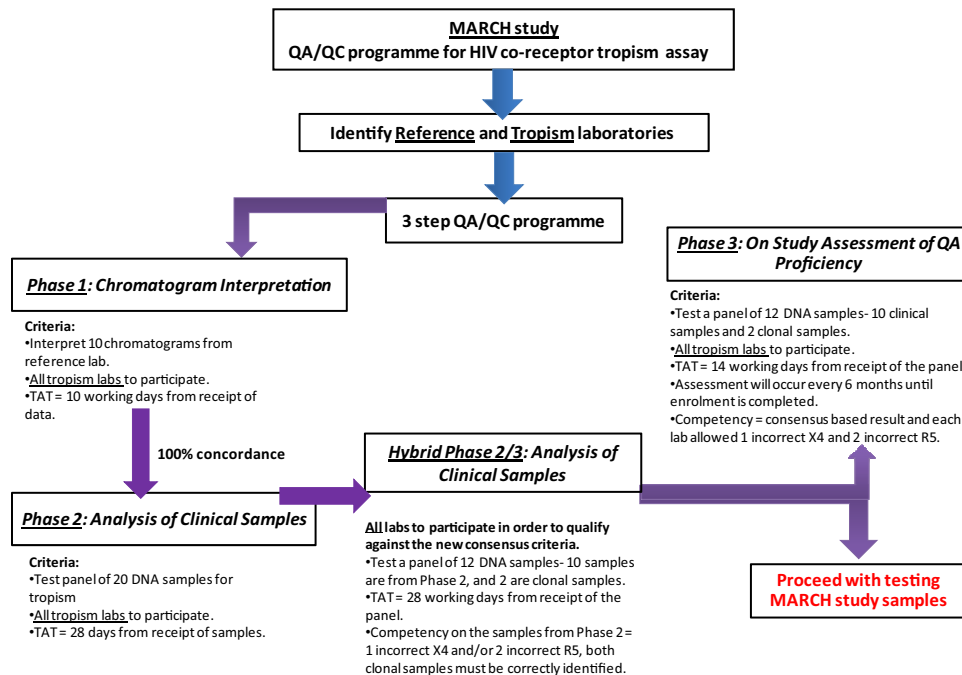


FIG 1 Summary of the EQA program for the MARCH HIV coreceptor tropism assay. TAT, turnaround time.

[.clinicaltrials.gov/ct2/show/study/NCT01384682?term=NCT01384682&rank=1](http://clinicaltrials.gov/ct2/show/study/NCT01384682?term=NCT01384682&rank=1)) is an ongoing international randomized clinical trial of MVC as a replacement for the current nucleoside or nucleotide analogue reverse transcriptase inhibitor or boosted protease inhibitor (PI/r) in virologically suppressed subjects (plasma viremia <200 copies/ml) on a stable PI/r-based therapy. Participants are only eligible if they have R5 HIV as assessed by genotypic testing using proviral DNA. The study is being conducted in 62 sites across Europe ( $n = 27$ ), Australia ( $n = 11$ ), Asia ( $n = 2$ ), and North ( $n = 4$ ) and South ( $n = 18$ ) America, and 14 laboratories are involved in the study.

When the study was established in 2011, one of the challenges was the lack of a standardized international quality assurance or quality control program for the assessment of tropism using proviral DNA across the network of international laboratories serving the sites in which the MARCH was being conducted. The aim of this paper is to describe the experience of the MARCH laboratory group in establishing and completing an external quality assessment (EQA) program prior to testing participant samples in the MARCH.

**Laboratory methods.** Each of the 14 laboratories performing genotypic tropism testing for the MARCH uses its own in-house methodology. There are, however, some key commonalities among these methods, and a representative method is detailed in Swenson et al. (13). In this EQA program, the starting material is DNA (see below). The V3 region of HIV *gp120* is amplified in triplicate by PCR and is sequenced by standard population-based sequencing. The resulting chromatograms are analyzed either manually by technicians or automatically using the freely accessible software RECall (15), with no manual inspection. These sequences are then submitted to the geno2pheno algorithm for tropism interpretation. Geno2pheno converts an input V3 sequence to an FPR value. This FPR is related to the likely phenotypic tro-

pism that is associated with the submitted sequence (5). Initially, an FPR of  $\geq 20\%$  was used to determine R5 versus X4, but this was lowered to  $\geq 10\%$  after completion of phase 2/3 (see below). In summary, samples were classified as R5 only if all three sequences had an FPR of  $\geq 10\%$ , and the samples were classified as X4 if they had an FPR of  $< 10\%$ ; samples that were unable to be amplified were classified as a fail.

**External quality assessment methods.** The objective of the EQA was to ensure uniformity, reproducibility, and competency in DNA-based HIV tropism testing by the laboratories involved in the MARCH. There were four phases of EQA for the tropism laboratories in the MARCH (Fig. 1), and all phases were overseen by NRL, Victoria, Australia. Laboratories having performance issues in any of the phases were contacted to identify areas of improvement (e.g., were encouraged to use RECall), such that an increasing number of laboratories were able to meet competency criteria in subsequent EQA phases.

The purpose of phase 1 was to assess laboratory proficiency in the sequence interpretation steps, from processing the raw sequence chromatogram to interpreting with the geno2pheno algorithm. The NRL distributed identical sets of sequence chromatogram files corresponding to 10 samples sourced from the Vancouver and Cologne laboratories. The samples were selected such that five had FPRs of  $< 20\%$  and five had FPRs of  $> 20\%$ . The laboratories then analyzed the chromatograms manually or automatically, and they submitted the resulting sequences to geno2pheno for interpretation. A tropism classification was determined for each of the 10 specimens, and these results were returned to NRL. To pass phase 1, each laboratory was required to achieve the same results as those of the reference laboratory on all 10 sets of chromatograms.

The aim of phase 2 was to assess all the steps required for DNA-based genotypic tropism testing. This phase involved test-

ing 20 DNA samples derived from clinical specimens. Eighteen of the samples were from treatment-experienced patients with virologic suppression (plasma viremia, <50 copies/ml) on combination antiretroviral therapy (cART) for  $\geq 6$  months. Patient samples for phase 2 were selected from a patient population mirroring the profiles of likely MARCH participants. Moreover, the selection was weighted toward “known” X4 tropism (in 10/15) on prior phenotypic testing prior to virologic suppression, as it was felt that the ability to detect X4 was more important than the ability to detect R5, the former being a safety issue for a potential MVC switch.

To have sufficient starting material for all laboratories, each was sent specimens containing 2 to 8  $\mu\text{g}$  purified DNA in 50  $\mu\text{l}$  of nuclease-free water. All laboratories were instructed to test each sample in triplicate. The V3 region was PCR amplified, sequenced, interpreted, and analyzed using geno2pheno. A sample was classified as X4 if any of the 3 amplicons had a geno2pheno FPR below the classification cutoff (either 10% or 20%). For comparisons among the laboratories, a consensus-based approach was used to establish the “correct” classification of a sample; this was necessary due to the inherent variabilities of the clinical specimens. In this scheme, the consensus value for each sample was set as the median of the lowest FPR from each laboratory, and if this fell below the cutoff, then the consensus classification for that sample was X4. For example, if a sample had a median FPR of 5%, the reference tropism classification for this sample would be X4.

For phase 2, samples for which the median interlaboratory FPR was <20% were called X4, but this cutoff was subsequently revised to <10% (see Results). Of the 20 samples, a maximum of 3 incorrect tropism classifications were permitted: up to 2 R5 samples could be incorrectly called X4, and 1 X4 sample could be incorrectly called R5. These criteria allowed for the variability of clinical specimens (i.e., permitting 3 samples per laboratory to be classified differently from the consensus), while remaining conservative (i.e., only one “missed” X4 was permitted). Laboratories were also required to successfully obtain sequences from 80% of samples (16 of 20).

Because of the revised cutoff and the establishment of competency criteria in phase 2, another EQA phase (called hybrid phase 2/3) was implemented to test these as prespecified criteria. The phase included 10 of the original clinical specimens, plus 2 clonal specimens (1 X4 and 1 R5). To pass hybrid phase 2/3, laboratories were required to pass the prespecified criteria for the clinical specimens and to correctly classify both clonal samples. Laboratories that successfully passed phase 2 and/or the hybrid phase could begin testing the enrolled participants in the MARCH. An ongoing phase 3 will occur every 6 months during the screening phase of the MARCH and will comprise testing of the original clinical samples, plus 2 clonal samples (Fig. 1). Data from phase 3 are not currently available and are not discussed here.

**Further investigations.** In addition to the EQA methodology, further investigations were performed to estimate the variation within samples, as well as to estimate the effect of manual versus automated sequence analysis. For estimating variation, all 20 phase 2 samples were repeatedly amplified to a maximum of 32 sequences per sample. Each sequence was scored by geno2pheno. The median and interquartile range of the FPR for each sample were also quantified. Furthermore, seven samples from the hybrid phase 2/3 underwent deep sequencing on a Roche/454 Genome Sequencer Junior.

To determine whether manual sequencing analysis had an impact on the results, raw chromatogram files were requested from the laboratories participating in the hybrid phase 2/3 and were processed using RECall. These chromatogram files were generated independently at each laboratory using their own methodologies. Processing the raw files with RECall eliminated variability in the sequencing analysis and in the interpretation step of the protocol, such that any variation observed could be attributed only to the amplification and the sequencing steps.

The nucleotide sequences obtained by all laboratories participating in phase 2 (triplicate testing) were analyzed using phylogenetic analysis. The alignment and phylogenetic analysis were performed using ClustalX.

## RESULTS

All laboratories met competency criteria in phase 1 of the EQA for tropism testing in the MARCH, confirming their proficiency in obtaining inferred tropism results from raw sequence chromatogram files. This phase also confirmed that these steps of genotypic tropism testing are reproducible across different settings.

Subsequent phases of the EQA showed more variability between laboratories, which is likely due to variation introduced by the sample handling, amplification, sequencing, interpretation steps, and biological variation (16). Of the 15 laboratories that passed phase 1, 14 laboratories participated in subsequent EQA phases, and one laboratory withdrew participation from the EQA program.

Laboratories participating in phase 2 returned more variable results than those in phase 1 (Fig. 2). Twenty DNA samples were distributed, of which 19 were successfully amplified by the majority of laboratories. One laboratory only successfully amplified and sequenced 15/19 samples. Phase 2 initially used an FPR cutoff of 20% and allowed 3 incorrectly classified samples (2 R5 samples called X4 and 1 X4 sample called R5).

Based primarily on the analysis of a subset of patients in the Maraviroc Therapy in Antiretroviral Treatment-Experienced (MOTIVATE) and A4001029 trials of maraviroc (17), the geno2pheno cutoff was lowered to 10%. This was based partially on the finding that an FPR of 10% outperformed an FPR of 20% for distinguishing between the responders and nonresponders to maraviroc (see supplementary Fig. 1 of reference 17). Using the altered cutoff, there were fewer incorrect X4 classifications per laboratory (mean of 1.54 with a cutoff of 20% versus a mean of 1.38 with cutoff of 10%). On average, each laboratory misclassified a median of 1 R5 sample (interquartile range [IQR], 1 to 1.75) and 1 X4 sample (IQR, 1 to 3) in phase 2. A total of 7 laboratories were deemed appropriate to open for tropism testing after meeting the competency criteria in phase 2.

The competency criteria established in phase 2 were then applied in hybrid phase 2/3 (Table 1). In this phase, laboratories misclassified a median of 0 R5 samples (IQR, 0 to 0), and 1 X4 sample (IQR, 0 to 1). All laboratories except one were able to correctly classify both clonal specimens, a requirement in hybrid phase 2/3. Based on the phase 2/3 results, a further 3 laboratories were deemed appropriate to open for testing, in addition to the 7 that were deemed appropriate after phase 2, for a total of 10.

**Subsequent analyses.** Specimens distributed in phase 2 underwent repeated amplification and sequencing to a maximum of 32 amplifications per sample in the reference lab, with a median of 17

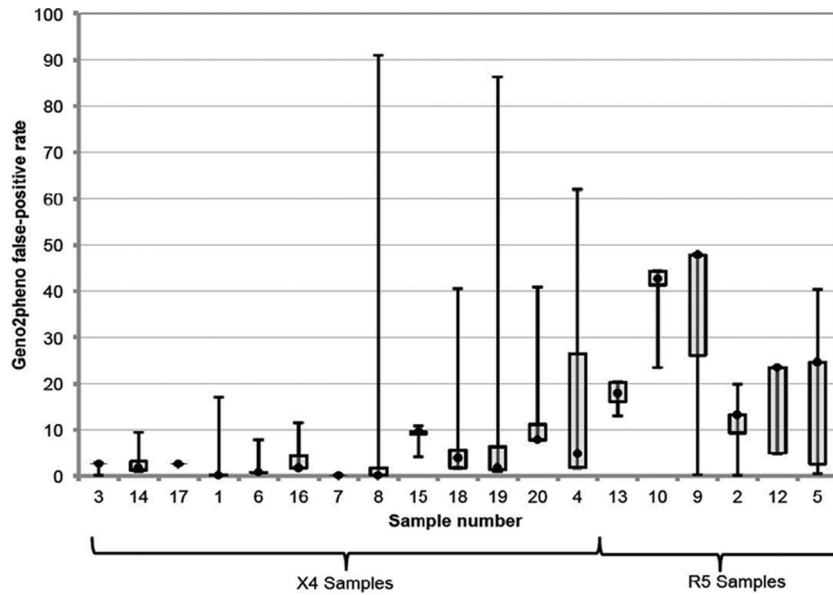


FIG 2 Interlaboratory variation in minimum FPR. There was significant variation across the laboratories for each sample, even when the lowest geno2pheno false-positive rates (FPRs) were examined. Samples were sorted by consensus tropism and then by decreasing interlaboratory agreement. The minimum FPRs obtained from all laboratories from each of the 20 specimens in phase 2 were used to generate box-and-whisker plots for each specimen. Horizontal lines represent the median FPR, boxes represent the interquartile range of the values, and whiskers represent the minimum and maximum values obtained across the laboratories. One laboratory (laboratory B) gave outlier results for 4 samples (with outlier defined as an FPR outside the range mean  $\pm$  1.96 standard deviations). These were samples 1, 3, 8, and 15. The other laboratories contributed 0 to 2 outlier results.

(IQR, 10.5 to 23). The variation detected in these samples is displayed in Fig. 3. Twelve of 19 samples appeared to have mixed R5 and X4 populations, according to this repeated sequencing, with a median of 55% of sequences below an FPR of 10% (IQR, 28 to 75%). These mixed-tropism samples were associated with a greater number of laboratories giving classifications that were discordant from the consensus. A median of 14% of laboratories (IQR, 7 to 10%) gave discordant results in the mixed-tropism samples, versus a median of 0% discordant laboratories (IQR, 0 to 21%) where repeated sequencing indicated “pure” R5 or X4 populations. Figure 3 shows the sample variation detected by this method. Seven samples had consistent tropism classifications by all laboratories, and these appeared to have small interquartile ranges in the FPRs as detected by repeated sequencing, suggesting

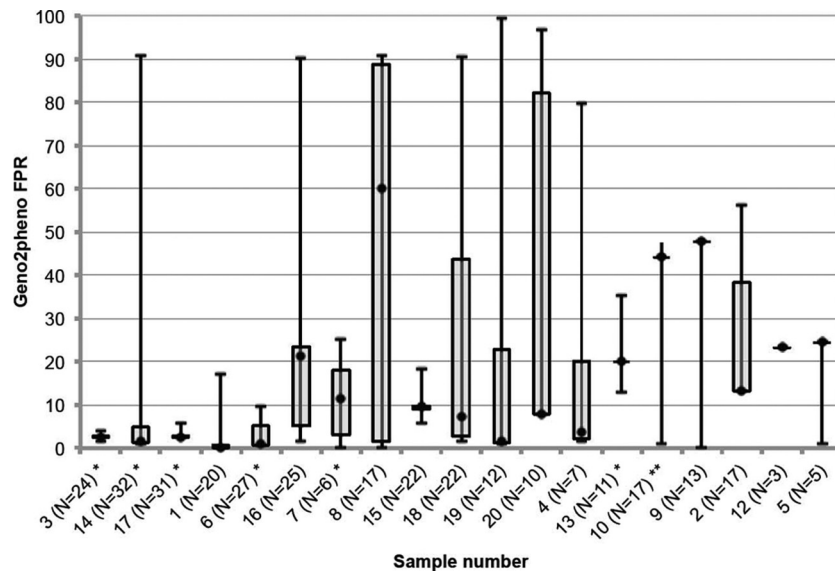
that lower variability might have contributed to the higher interlaboratory agreement. Further analysis of the sequences using phylogenetic analysis revealed that where tropism results disagreed with the consensus, the sequences still clustered closely with other sequences from the same specimen, suggesting that sample mix-ups and to a lesser extent contamination, were not major issues in the EQA. No samples had apparently been switched at any laboratory, and there were 10 possible contamination events identified out of a total of 685 sequences that were obtained for phase 2 (Fig. 4).

For phase 2/3, laboratories that classified samples as having discordant tropism results compared to the consensus were asked to submit the raw chromatogram files obtained for these samples. All chromatograms were then interpreted automatically with

TABLE 1 EQA hybrid phase 2/3 results for the 14 laboratories<sup>a</sup>

| Clone or sample no. | Consensus classification | FPR (%) at the indicated laboratory |      |      |      |      |      |      |      |      |      |      |      |      |      |
|---------------------|--------------------------|-------------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|
|                     |                          | A                                   | B    | C    | D    | E    | F    | G    | H    | I    | J    | K    | L    | M    | N    |
| Clone 1             | R5                       | 17                                  | 8.5  | 17   | 17   | 17   | 17   | 17   | 17   | 17.9 | 17   | 17.9 | 17   | 17   | 17   |
| Sample 3            | R5                       | 16.6                                | 5.8  | 13.2 | 20.8 | 13.2 | 13.2 | 13.2 |      | 16.9 | 20.7 | 13.2 | 16.6 | 13.1 | 14.7 |
| Sample 5            | R5                       | 47.8                                | 6.8  | 5.3  | 47.8 | 0.4  | 47.8 | 47.8 | 47.8 | 44.5 | 47.8 | 47   | 47.8 |      | 47.8 |
| Sample 6            | R5                       | 44.2                                | 5.9  | 42.6 | 44.2 | 44.2 | 44.2 | 44.2 | 44.2 | 44.2 | 44.2 | 44.2 | 44.2 |      |      |
| Sample 7            | R5                       | 20.2                                | 15.4 | 13   | 13   | 13   | 20.2 | 20.2 | 20.2 | 20.2 | 17.3 | 13   | 20.2 |      | 16.4 |
| Clone 2             | X4                       | 0.1                                 | 0    | 0    | 0.2  | 0.1  | 0    | 0    | 0.1  | 0.2  | 0.2  | 0    | 0    | 0.2  | 0.2  |
| Sample 4            | X4                       | 2.7                                 | 1.7  | 2.6  | 2.5  | 2.6  | 2.6  | 2.6  | 2.6  | 2.6  | 2.8  | 2.6  | 2.6  | 2.7  | 2.6  |
| Sample 8            | X4                       | 3.2                                 | 1.3  | 1.7  | 3.2  | 19.1 | 3.2  | 3.2  | 34.3 | 3.8  | 1.3  | 1.8  | 1.2  | 1.2  | 1.3  |
| Sample 9            | X4                       | 3.1                                 | 1.7  | 1.8  | 22.4 | 1.9  | 2.9  | 21.2 | 21.3 | 6    |      | 16.2 | 1.7  |      | 0.7  |
| Sample 10           | X4                       | 2.6                                 | 1.7  | 2.6  | 2.6  | 2.6  | 2.6  | 2.6  | 3.5  | 3.2  | 2.6  | 3.2  | 2.6  | 2.6  | 2.6  |
| Sample 11           | X4                       | 5.3                                 | 3.2  | 4.8  | 4.1  | 24   | 1.7  | 16.4 |      | 1.3  | 24   | 4.6  | 6.7  |      | 16.4 |
| Sample 12           | X4                       | 7.8                                 | 7.8  | 7.8  | 7.8  | 7.8  | 7.4  | 7.8  |      | 31.1 | 7.8  | 7.8  | 7.8  | 14.3 | 7.8  |

<sup>a</sup> The lowest FPR results from the triplicate testing for each sample by each laboratory are shown here. The FPR cutoff was 10%. R5 was determined if the FPR was  $\geq$ 10%, and X4 (shaded cells) was determined if the FPR was  $<$ 10%. Blank boxes indicate that the sample could not be amplified.



**FIG 3** Variation by repeated amplifications. The same specimens shown in Fig. 2 were repeatedly sequenced up to 32 times. For comparison, they are displayed in the same order as in Fig. 2. The intensive sequencing revealed marked variation within the samples. Horizontal lines represent the median FPR obtained across all repeated sequences, boxes represent the interquartile ranges of the values, and whiskers represent the minimum and maximum values obtained by repeated sequencing. Samples for which all laboratories gave the same tropism determination often had lower variability according to repeated sequencing analysis (see specimens with asterisks). This approach shows the importance of triplicate sequencing, since many samples had large spreads of FPR values (e.g., sample 8). The method of taking the minimum FPR detected is also key to a sensitive detection of the X4 virus, since higher FPRs might obscure X4 detection (e.g., sample 19). \*, all laboratories called X4; \*\*, all laboratories called R5.

RECall and were scored with geno2pheno. Applying consistent base-calling interpretation software resulted in two additional laboratories meeting the competency criteria in phase 2/3, for a total of 11. RECall reanalysis corrected a mean of 0.6 sample per laboratory (range, 1 incorrect sample added by RECall reanalysis to 3 corrected samples). The coefficient of variation of the geno2pheno FPRs of the samples was significantly lower after RECall reanalysis ( $t$  test,  $P = 0.04$ ), suggesting decreased inter-laboratory variability. As a result, all laboratories are now encouraged to use RECall in their sequencing analyses.

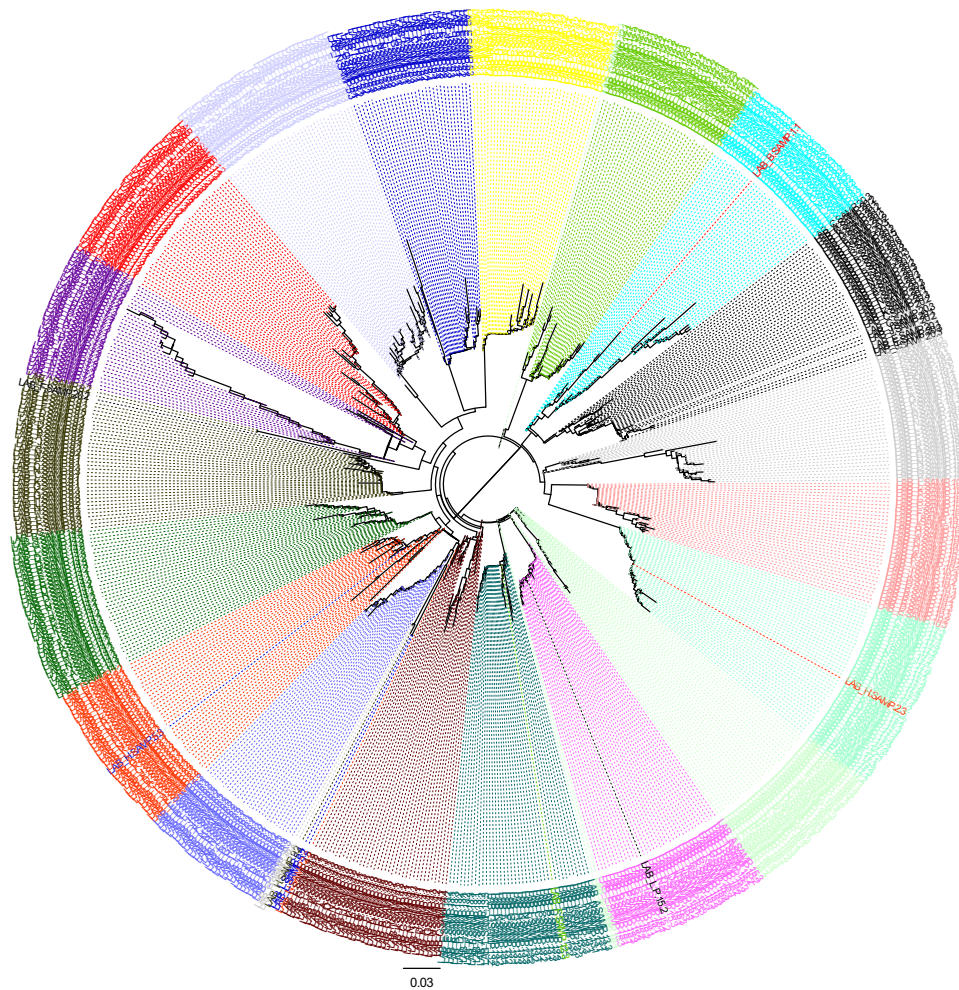
To further estimate the effect of intrasample diversity on inter-laboratory concordance, samples from hybrid phase 2/3 (clonal sample 2, sample 4, and samples 7 through 11) underwent deep sequencing on a Roche/454 Genome Sequencer Junior. Sequence depth was a median of 1,877 sequences/sample (IQR, 1,200 to 2,326). The percentage of sequences in the sample scoring  $<10\%$  by geno2pheno was calculated to quantify the sample diversity. There were three samples for which deep sequencing indicated 100% of sequences as scoring below a geno2pheno cutoff of 10% (sample 4, sample 10, and clone 2). Interestingly, these samples had the highest interlaboratory agreement, with all laboratories returning X4 results. Similarly, a sample consistently called R5 by all laboratories (sample 7) had a very low prevalence (2.5%) of sequences that were  $<10\%$ . The remaining samples had both lower interlaboratory agreement and mixed populations of sequences above and below the FPR cutoff of 10%, which might account for the lower agreement. For example, sample 8 had 47.5% of sequences that were  $<10\%$ , with 86% of laboratories giving X4 results. Two-thirds of laboratories returned X4 results for samples 9 and 11, and these samples had 25% and 1.2% of sequences that were  $<10\%$ , respectively.

## DISCUSSION

The traditional structure of an EQA is to discern the ability of the laboratories to perform a test and to reveal the correct result as already determined by a reference laboratory or by a “gold standard.” Tropism laboratories chosen to participate in the MARCH are experienced in genotypic assessments of tropism using proviral DNA, and they operated using their normal methods. The major finding through the MARCH EQA was the high variability within proviral DNA samples when these are derived from patients with chronic infection who have had prolonged periods of viral suppression.

In its purest form, an EQA assessing tropism using proviral DNA could have been performed using only clonal samples to assess the ability of the laboratories to adequately perform each step of the method. One might argue that an EQA using only clonal samples, followed by clinical samples that mirrored likely participant samples for the MARCH, would have been a better approach. As participant laboratories were, for the most part, already experienced in this technology, we chose to use only clinical samples in the wet panel testing (phase 2). However, the high variability of the results from phase 2 led to a further round of wet panel testing using both clonal and clinical samples (phase 2/3 onwards) in order to better understand where discrepancies between the laboratories had occurred.

One of the major challenges of the EQA program was defining “truth,” i.e., the correct tropism of the proviral DNA clinical samples. From the three separate results that were obtained for each sample, if any sequence was identified as X4, it was reported as X4. To overcome the partial DNA variability, a consensus-based approach for tropism determination of a sample was used. The final correct tropism was defined by what the majority of laboratories



**FIG 4** Phylogenetic tree of EQA phase 2 sequences for the 14 laboratories. Each sample is represented by a different color. In most cases, the sequences generated from the same sample by different laboratories clustered near each other. There was no evidence of sample mix-ups according to phylogenetic analysis. There was evidence of contamination (i.e., a sequence from one sample clustering with another) for 7 laboratories (laboratories A, B, D, F, H, I, and N), with most cases being restricted to a maximum of 1 sequence in 1 sample. There were 10 possible contamination events identified out of 685 sequences obtained (1.5%).

had called the sample. As a result, the consensus value for each sample was set as the median of the lowest FPR from each laboratory, and if this fell below the cutoff, then the consensus classification for that sample was X4.

Furthermore, to establish the factors that are associated with divergent results across laboratories, the EQA clinical samples underwent repeated amplifications and deep sequencing. Further analysis revealed a heterogeneous viral population within the clinical samples tested, i.e., a mix of R5 and X4 variants. The biological intrasample variability therefore explains the discordance observed in tropism results across laboratories and within the replicates tested. In addition, the vast majority of sequences clustered most closely with other sequences from the same specimen, which in most cases excludes the possibility of contamination. Thus, the majority of discrepant results likely arose from the random amplification of minority quasi-species and not from methodological deficiencies by the laboratory. Furthermore, this analysis highlights the reality that even triplicate testing is not able to sufficiently capture all the diversity that is present in a clinical specimen.

Recent studies support the use of triplicate testing but have been associated with an increased detection of X4 variants, and hence, the number of patients eligible for MVC treatment is reduced (13, 18, 19). Symons et al. (18) showed that with the addition of a third replicate, there was a 4% increase in X4 prediction in the study population, and this resulted in reclassification from R5 to X4 tropism. Duplicate testing has also been investigated using plasma HIV RNA and showed improved detection of X4 viruses compared to single testing (20). For the MARCH study, given the inherent variability of DNA and the clinical outcome being unknown, we prefer to be conservative and perform triplicate testing at an FPR of 10%. The detection of X4 sequences in any triplicate amplification from a sample should preclude that patient from switching to MVC.

Originally, the FPR threshold was set at 20%, which was a conservative cutoff. However, considering that the EQA was set up for implementation in a large-scale clinical study and potentially in routine clinical practice, it was critical to be conservative in terms of safety, i.e., to exclude those patients with X4 virus. The consequence of such a conservative FPR was that the majority of viruses

had a predicted tropism of X4. According to the European Consensus Group guidelines (9), it is recommended that proviral DNA tropism testing be performed in triplicate with an FPR of 10%, and the revised cutoff in the MARCH is consistent with that recommendation. A recent study by Svicher et al. (21) detected more R5-tropic viruses when an FPR of 10% was used than when an FPR of 20% was used. To support the use of a lower threshold and also to improve the detection of CCR5 tropism, an FPR of 20% was compared to FPRs of 2%, 5.75%, and 10% using data available from a subset of patients in the MOTIVATE and A4001029 trials (17). The comparison demonstrated that an FPR of 10% provided a clear discrimination between responders and nonresponders to MVC-containing regimens, and applying this to MARCH EQA produced fewer incorrect X4 classifications. Therefore, the FPR cutoff was revised to 10% for phase 2 onwards.

Several studies have also evaluated the use of lower FPRs at 10%, 5.75%, and 5% to assess the performance of genotypic tropism testing in proviral DNA with phenotypic tropism testing (10, 22). There was high concordance between the genotypic and phenotypic testing at FPRs of 10% and 5.75% of 96.3% and 100%, respectively (22), and 85% concordance at FPRs of 10% and 5% (10), which suggests that an FPR of 5.75% might be a good predictor for a positive response to MVC therapy (7). However, it is important to consider that these studies examined patients with detectable viral loads of >500 copies/ml, and they used HIV RNA instead of DNA as their starting material. For the MARCH, given that the patients are virologically suppressed, there is the potential to use a lower threshold to predict the clinical efficacy of MVC as a switch drug, but this would require further validation and assessment of clinical outcomes at lower thresholds, and potentially another intensive EQA. Once clinical data are available from the main MARCH, we will be able to perform a retrospective analysis to determine if a lower FPR would have a similar clinical efficacy of MVC at an FPR of 10% and, hence, optimize the use of CCR5 antagonists in a clinical setting.

Intensive quality assessment and quality control (QA/QC) of all laboratories should be implemented before embarking on multinational clinical studies in order to ensure that a common standard is used prior to testing participant samples (23–25). Recently, the DNA Tropism Italian Validation Concerted Action (DIVA) study (21) was set up to standardize genotypic tropism testing using proviral DNA for the routine clinical diagnostic laboratory in Italy. There were two phases; phase 1 involved the cross-validation of the methodology of V3 sequencing on proviral DNA among the 5 reference centers, and phase 2 was aimed at validating and sharing the methodology with 12 local virological centers involved in the project. Patient samples were used, and the majority had an undetectable viral load of <50 copies/ml. A single PCR was performed, and a geno2pheno FPR of 20% was applied. The study demonstrated high concordance in the genotypic tropism prediction using proviral DNA across the different laboratories in Italy (21). Interestingly, there were only 2 discordant samples between the local virological centers and the reference center, while in the current study, the high levels of DNA variation complicated the tropism prediction. The main difference between the two studies is that the DIVA study was performed singly versus in triplicate, and this perhaps resulted in less variability between the laboratories. Other factors that might have contributed to the differences observed were the sample size tested, the extraction procedure by which proviral DNA was obtained, the sequencing method used,

the replicates performed, and the FPR cutoff that was applied. Overall, an EQA program is a useful tool for helping to standardize a methodology in laboratories that do not routinely perform the testing. Panels should also include patient-derived samples to best reflect the patient population being tested.

Therefore, the MARCH EQA program of genotypic tropism testing using proviral DNA revealed the inherent variability of viral DNA in virally suppressed patients, and if only single or duplicate replicates had been performed, some of the variability and random amplification of a minority quasi-species that were revealed by triplicate testing would have been missed. All laboratories involved in the EQA program are now serving as tropism laboratories testing MARCH participant samples. The MARCH is the first randomized multicenter international study of the use of MVC as a switch drug in virologically suppressed HIV-1-infected patients. It will provide the largest clinical data set for the efficacy and safety of this approach, utilizing a genotypic assessment of tropism on proviral DNA to determine the sensitivity to MVC as an antiviral agent. As such, a robust determination of viral tropism is essential to the integrity of this study. The EQA program established for the MARCH plays a critical role in ensuring that the outcome of the study is reliable and applicable for the assessment of tropism in a clinical setting.

#### ACKNOWLEDGMENTS

The MARCH is an academic collaboration between UNSW and Pfizer/ViiV Healthcare. Three members of the MARCH study protocol steering committee (Simon Portsmouth,\* Fraser Drummond,\* and Eric LeFevre\*) are employees of Pfizer or ViiV Healthcare.

The study was funded by the Australian Government Department of Health and Ageing. The Kirby Institute is affiliated with the Faculty of Medicine, University of New South Wales.

The views expressed in this publication do not necessarily represent the positions of the Australian Government.

The MARCH Tropism Laboratory group comprises the following members: Horacio Salomon, Sergio Mazzini, and Andrea E. Rubio (Argentina); Anthony Kelleher,\* Kat Marks, Kazuo Suzuki, Nick Rismanto, Chris Birch, and Doris Chibo (Australia); Richard Harrigan, Luke Swenson, and Dennison Chan (Canada); William Acevedo, Patricio Manque, and Victor Polanco (Chile); Thomas Berg, Martin Obermeier, Rolf Kaiser,\* and Nadine Luebke (Germany); Suzie Coughlan and Jonathan Dean (Ireland); Orna Mor and Marina Wax (Israel); Wataru Sugiura\* and Yasumasa Iwatani (Japan); Gustavo Reyes Teran and Santiago Avila (Mexico); Tomas Pumarola (Spain); Kiat Ruxrungtham,\* Sunee Sirivichayakul, May Naphassanant, and Sasiwimol Ubolyam (Thailand); Steve Kaye (United Kingdom); and Sally Land, Joe Vincini, and Sue Best (NRL, Melbourne, Australia). The MARCH Study Group is composed of the following site and/or Coordinating Centre personnel: Marcelo Losso, Guillermo Vilorio, Angel Parlante, Marisa Sanchez, Mariana De Paz Sierra, Emiliano Bissio, Pablo Luchetti, Sergio Lupo, Luciana Peroni, Eduardo Warley, Lic Ines Vieni, Norma Porteiro, and Cecilia Vilas (Argentina); Mark Kelly, Abby Gibson, Mark Bloch, Teo Franic, William Donohue, Jill Thompson, Julian Elliot, Peta Hamill, Michelle Boglis, David Sowden, Suzanne Murray, Jennifer Hehir, Colleen Johnston, Sarah L. Pett,\* David Cooper,\* Karen MacRae, Brett Sinclair, Dominic Dwyer, Delene Assam, Ian Woolley,\* Ainsley Gillies, David Orth, and David Youds (Australia); Roger LeBlanc, François Lantaigne, Sharon Walmsley, Warmond Chan, John Gill,\* Brenda Beckthold, Graham Smith, Roberta Halpenny, and Jennifer Robinette Hills (Canada); Marcelo Wolff\* and Gladys Allende (Chile); Thierry Prazuck\* and Barbara de-dieuleveult (France); Christoph Stephan, Franziska Ebeling, Claudia Wengenroth, Björn Jensen, Cecilie Feind, Dirk Meyer-Olson, Kirsten Hoepfer, Renata Beider, Gerd Faetkenheuer, Eleonore Rund, Patrick Ingiliz, Andreas Wienbreyer, Juergen

Rockstroh,\* Angelika Engelhardt, Christoph Boesecke, Brigitta Spaeth, Heiko Jessen, Carmen Zedlack, and Hans-Jurgen Stellbrink\* (Germany); Paddy Mallon\* and Sibongile Simelane (Ireland); Wataru Sugiura\* and Mayumi Imahashi (Japan); Marcelo Losso, Waldo Belloso,\* Daniel Guelman, and Mariana Valdovinos (Latin America Coordinating Centre); Jaime Andrade-Villanueva, Karina Montes, Juan Sierra Madero,\* Jesus Eduardo Sanchez Hernandez, Eduardo Ruiz Ballesteros, Luis Mosqueda-Gomez, Monica Lopez-Segura (Mexico); José Gatell\*, Ana González-Cordón, Roger Paredes, Jordi Puig, Bonaventura Clotet, Marcial Delgado-Fernandez, Jose Arribas, and Juan Miguel Castro (Spain); Joan Albert Arnaiz, Helena Beleta, and Judit Pich (Spain Coordinating Centre); Kiat Ruxrungtham,\* Jintana Intasan, and Wirach Maek-a-nantawat (Thailand); Alan Winston, Scott Mullaney, Martin Fisher,\* Amanda Clarke, Celia Richardson, and Nicky Perry (United Kingdom); David Cooper,\* Sean Emery,\* Sarah L. Pett,\* Nisha Berthon-Jones, David Silk, Elise Tu, Kymme Courtney-Vega, Noorul Absar, Janaki Amin,\* and Hila Haskelberg (Kirby-MARCH team); and Anthony Kelleher,\* Kate Merlin, Julie Yeung, and Bertha Fsadni (Kirby-MARCH central laboratory group (AMR)). Members of the MARCH Protocol Steering Committee are those above with asterisks next to their names.

## REFERENCES

- Cormier EG, Dragic T. 2002. The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor. *J. Virol.* 76:8953–8957.
- Gulick RM, Lalezari J, Goodrich J, Clumeck N, DeJesus E, Horban A, Nadler J, Clotet B, Karlsson A, Wohlfeiler M, Montana JB, McHale M, Sullivan J, Ridgway C, Felstead S, Dunne MW, van der Ryst E, Mayer H, MOTIVATE Study Teams. 2008. Maraviroc for previously treated patients with R5 HIV-1 infection. *N. Engl. J. Med.* 359:1429–1441.
- Wilkin TJ, Goetz MB, Leduc R, Skowron G, Su Z, Chan ES, Heera J, Chapman D, Spritzler J, Reeves JD, Gulick RM, Coakley E. 2011. Reanalysis of coreceptor tropism in HIV-1-infected adults using a phenotypic assay with enhanced sensitivity. *Clin. Infect. Dis.* 52:925–928.
- Cooper DA, Heera J, Goodrich J, Tawadrous M, Saag M, DeJesus E, Clumeck N, Walmsley S, Ting N, Coakley E, Reeves JD, Reyes-Teran G, Westby M, Van Der Ryst E, Ive P, Mohapi L, Mingrone H, Horban A, Hackman F, Sullivan J, Mayer H. 2010. Maraviroc versus efavirenz, both in combination with zidovudine-lamivudine, for the treatment of antiretroviral-naïve subjects with CCR5-tropic HIV-1 infection. *J. Infect. Dis.* 201:803–813.
- Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R. 2007. Bioinformatics prediction of HIV coreceptor usage. *Nat. Biotechnol.* 25:1407–1410.
- Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, McLaughlin S, Shankarappa R, Margolick JB, Mullins JI. 2003. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J. Virol.* 77:13376–13388.
- McGovern RA, Thielen A, Portsmouth S, Mo T, Dong W, Woods CK, Zhong X, Brumme CJ, Chapman D, Lewis M, James I, Heera J, Valdez H, Harrigan PR. 2012. Population-based sequencing of the V3-loop can predict the virological response to maraviroc in treatment-naïve patients of the MERIT trial. *J. Acquir. Immune Defic. Syndr.* 61:279–286.
- McGovern RA, Thielen A, Mo T, Dong W, Woods CK, Chapman D, Lewis M, James I, Heera J, Valdez H, Harrigan PR. 2010. Population-based V3 genotypic tropism assay: a retrospective analysis using screening samples from the A4001029 and MOTIVATE studies. *AIDS* 24:2517–2525.
- Vandekerckhove LP, Wensing AMJ, Kaiser R, Brun-Vezinet F, Clotet B, De Luca A, Dressler S, Garcia F, Geretti AM, Klimkait T, Korn K, Masquelier B, Perno CF, Schapiro JM, Soriano V, Sonnerborg A, Vandamme AM, Verhofstede C, Walter H, Zazzi M, Boucher CA, the European Consensus Group on Clinical Management of Tropism Testing. 2011. European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect. Dis.* 11:394–407.
- Verhofstede C, Brudney D, Reynaerts J, Vaira D, Fransen K, De Bel A, Seguin-Devaux C, De Wit S, Vandekerckhove L, Geretti AM. 2011. Concordance between HIV-1 genotypic coreceptor tropism predictions based on plasma RNA and proviral DNA. *HIV Med.* 12:544–552.
- Prosperi MC, Bracciale L, Fabbiani M, Di Giambenedetto S, Razzolini F, Meini G, Colafigli M, Marzocchetti A, Cauda R, Zazzi M, De Luca A. 2010. Comparative determination of HIV-1 co-receptor tropism by Enhanced Sensitivity Trofile, gp120 V3-loop RNA and DNA genotyping. *Retrovirology* 7:56. doi:10.1186/1742-4690-7-56.
- Soulié C, Fourati S, Lambert-Niclot S, Malet I, Wirlden M, Tubiana R, Valantin MA, Katlama C, Calvez V, Marcelin AG. 2010. Factors associated with proviral DNA HIV-1 tropism in antiretroviral therapy-treated patients with fully suppressed plasma HIV viral load: implications for the clinical use of CCR5 antagonists. *J. Antimicrob. Chemother.* 65:749–751.
- Swenson LC, Moores A, Low AJ, Thielen A, Dong W, Woods C, Jensen MA, Wynhoven B, Chan D, Glascock C, Harrigan PR. 2010. Improved detection of CXCR4-using HIV by V3 genotyping: application of population-based and “deep” sequencing to plasma RNA and proviral DNA. *J. Acquir. Immune Defic. Syndr.* 54:506–510.
- Verhofstede C, Vandekerckhove L, Eygen VV, Demecheleer E, Vandembroucke I, Winters B, Plum J, Vogelaers D, Stuyver L. 2009. CXCR4-using HIV type 1 variants are more commonly found in peripheral blood mononuclear cell DNA than in plasma RNA. *J. Acquir. Immune Defic. Syndr.* 50:126–136.
- Woods CK, Brumme CJ, Liu TF, Chui CK, Chu AL, Wynhoven B, Hall TA, Trevino C, Shafer RW, Harrigan PR. 2012. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. *J. Clin. Microbiol.* 50:1936–1942.
- Hayden RT, Yan X, Wick MT, Rodriguez AB, Xiong X, Ginocchio CC, Mitchell MJ, Caliendo AM, College of American Pathologists Microbiology Resource Committee. 2012. Factors contributing to variability of quantitative viral PCR results in proficiency testing samples: a multivariate analysis. *J. Clin. Microbiol.* 50:337–345.
- Swenson LC, Dong WY, Mo T, Demarest J, Chapman D, Ellery S, Heera J, Valdez H, Poon AF, Harrigan PR. 2013. Use of cellular HIV DNA to predict virologic response to maraviroc: performance of population-based and deep sequencing. *Clin. Infect. Dis.* 56:1659–1666.
- Symons J, Vandekerckhove L, Paredes R, Verhofstede C, Bellido R, Demecheleer E, van Ham PM, van Lelyveld SF, Stam AJ, van Versendaal D, Nijhuis M, Wensing AM. 2012. Impact of triplicate testing on HIV genotypic tropism prediction in routine clinical practice. *Clin. Microbiol. Infect.* 18:606–612.
- Kagan RM, Johnson EP, Siaw M, Biswas P, Chapman DS, Su Z, Platt JL, Pesano RL. 2012. A genotypic test for HIV-1 tropism combining sanger sequencing with ultra-deep sequencing predicts virologic response in treatment-experienced patients. *PLoS One* 7:e46334. doi:10.1371/journal.pone.0046334.
- Raymond S, Recordon-Pinson P, Saliou A, Delobel P, Nicot F, Descamps D, Marcelin AG, Flandre P, Calvez V, Masquelier B, Izopet J, ANRS AC11 Resistance Study Group. 2011. Improved V3 genotyping with duplicate PCR amplification for determining HIV-1 tropism. *J. Antimicrob. Chemother.* 66:1972–1975.
- Svicher V, Alteri C, Montano M, D'Arrigo R, Andreoni M, Angarano G, Antinori A, Antonelli G, Alice T, Bagnarelli P, Baldanti F, Bertoli A, Borderi M, Boeri E, Bon I, Bruzzone B, Callegaro AP, Capobianchi MR, Carosi G, Cauda R, Ceccherini-Silberstein F, Clementi M, Chirianni A, Colafigli M, D'Arminio Monforte A, De Luca A, Di Biagio A, Di Nicuolo G, Di Perri G, Di Pietro M, Di Santo F, Fabeni L, Fadda G, Galli M, Gennari W, Ghisetti V, Giacometti A, Gori C, Gori A, Gulminetti R, Leoncini F, Maffongelli G, Maggiolo F, Manca G, Gargiulo F, Martinelli C, Maserati R, Mazzotta F, Meini G, Micheli V, et al. 2012. Performance of genotypic tropism testing on proviral DNA in clinical practice: results from the DIVA study group. *New Microbiol.* 35:17–25.
- Raymond S, Delobel P, Mavigner M, Cazabat M, Encinas S, Souyris C, Bruel P, Sandres-Sauné K, Marchou B, Massip P, Izopet J. 2010. CXCR4-using viruses in plasma and peripheral blood mononuclear cells during primary HIV-1 infection and impact on disease progression. *AIDS* 24:2305–2312.
- Hayden RT, Hokanson KM, Pounds SB, Bankowski MJ, Belzer SW, Carr J, Diorio D, Forman MS, Joshi Y, Hillyard D, Hodinka RL, Nikiforova MN, Romain CA, Stevenson J, Valsamakis A, Balfour HH,



- Jr, and U.S. EBV Working Group. 2008. Multicenter comparison of different real-time PCR assays for quantitative detection of Epstein-Barr virus. *J. Clin. Microbiol.* **46**:157–163.
24. Pang XL, Fox JD, Fenton JM, Miller GG, Caliendo AM, Preiksaitis JK, American Society of Transplantation Infectious Diseases Community of Practice, Canadian Society of Transplantation. 2009. Interlaboratory comparison of cytomegalovirus viral load assays. *Am. J. Transplant.* **9**:258–268.
25. Preiksaitis JK, Pang XL, Fox JD, Fenton JM, Caliendo AM, Miller GG, American Society of Transplantation Infectious Diseases Community of Practice. 2009. Interlaboratory comparison of Epstein-Barr virus viral load assays. *Am. J. Transplant.* **9**:269–279.