Comparison between Rules-Based Human Immunodeficiency Virus Type 1 Genotype Interpretations and Real or Virtual Phenotype: Concordance Analysis and Correlation with Clinical Outcome in Heavily Treated Patients

Carlo Torti,1 Eugenia Quiros-Roldan,1 Wilco Keulen,6,7 Luigia Scudeller,2 Sergio Lo Caputo,1 Charles Boucher,4 Francesco Castelli,1 Francesco Mazotta,2 Piera Pierotti,3 Anne Mieke Been-Tiktak,7∗ Giovanni Buccoliero,4 Michele De Gennaro,3 Giampiero Carosi,1 and Carmine Tinelli,1 for the GenPherex Study Group of the MaSTeR Cohort∗

1Institute of Infectious and Tropical Diseases, University of Brescia, Brescia, 2Biostatistics Unit, Istituto di Ricovero e Cura a Carattere Scientifico, Policlinico San Matteo, Pavia, 3Department of Infectious Diseases, S. M. Annunziata Hospital, Firenze, 4Division of Infectious Diseases, Santissima Annunziata Hospital, Taranto, and 5Division of Infectious Diseases, General Hospital, Lucca, Italy; 6University Medical Center, and 7Virology Networks, Utrecht, The Netherlands

We compared 2 rules-based genotype interpretation systems and real or virtual phenotype through a retrospective analysis of a prospective trial. Genotypes were determined with VircoGEN II (VIRCO) and were interpreted with either RetroGram 1.4 or TRUGENE HIV-1 (guidelines 3.0) or original virtual phenotype (Virtual Phenotype; VIRCO), as available in the year 2000. Among 188 human immunodeficiency virus (HIV) type 1 isolates, overall concordance (κ agreement) was observed for the 2 rules-based systems, whereas striking discordances were noted between them and real and virtual phenotype interpretations for stavudine, didanosine, zalcitabine, abacavir, and amprenavir (κ < 0.4). Clinical evaluation of a subset of 173 patients showed that both rules-based sensitivity scores were independently associated with HIV RNA loads <400 copies/mL at week 16 of during-treatment analysis (TRUGENE: odds ratio [OR], 2.90; 95% confidence interval [CI], 1.52–5.52; P = .001; RetroGram: OR, 2.34; 95% CI, 1.21–4.55; P = .012), whereas, in contrast to real or virtual phenotype, interpretations according to biological cut-offs were not (OR, 1.91; 95% CI, 0.77–4.76; P = .162).

The emergence of drug-resistant human immunodeficiency virus (HIV) is an important factor limiting the effectiveness of antiretroviral therapy for the treatment of HIV infection. Several large international studies have demonstrated that therapeutic changes guided by resistance testing increase the possibility of clinical success [1–4]. These results led to the recommendation of routine HIV-1 drug-resistance testing in a number of clinical settings [5–8].

It is currently uncertain whether resistance information based on genotype or phenotype is a more significant predictor of clinical outcome, although there is accumulating evidence that appears to favor genotype-based resistance information [1–4, 9–12]. Genotype tests have entered into clinical practice more widely because of lower cost, shorter turnaround time, and simpler procedure. However, genotype testing provides an indirect measurement of resistance and needs to be interpreted into expected drug activity [13].

In addition to lists of mutations and guidelines, there are currently 2 sophisticated approaches for interpreting genotypic patterns: first, software programs, which
run algorithms based on current knowledge both from literature data and expert opinion [13]; and, second, virtual phenotype (v-PHT), which is determined by matching the patient’s HIV genotype with genotypes with known phenotype profiles in a large proprietary database [14]. In the present study, we used the unique clinical data of 188 pretreated HIV-1–infected patients to compare different genotype resistance interpretation systems and recombinant real phenotype (r-PHT) in relation to the virological outcome from salvage therapy.

PATIENTS AND METHODS

Patient enrollment and characteristics. HIV-infected patients enrolled in the GenPhex (Genotypic versus Phenotypic Resistance) project were considered for inclusion in this study. The GenPhex project is a prospective, randomized trial that demonstrated that virological outcome is not significantly different in response to salvage therapy based on r-PHT (Antivirogram; VIRCO) or v-PHT (Virtual Phenotype; VIRCO) when results were evaluated, in conjunction with patients’ clinical history, by an independent panel of experts [15].

The present study was approved by the local ethics committee of each collaborating center. All patients gave written informed consent. Human experimentation guidelines have been followed according to the declaration of Helsinki.

The inclusion criteria were as follows: (1) virological failure (defined as plasma virus load $\geq 1000$ HIV-RNA copies/mL after $\geq 6$ months of continuous treatment) while receiving a combination antiretroviral regimen including $\geq 3$ drugs; (2) $\geq 2$ years of exposure to antiretrovirals; and (3) experience with $\geq 6$ antiretrovirals.

Patients were randomly assigned in a 1:1 ratio to the r-PHT ($n = 101$) or v-PHT ($n = 100$) study arm. HIV-resistance results were not available for 13 patients because of low virus load at the time of testing or because of technical reasons and have been excluded from the analysis. Virological, immunological, and clinical data were collected at study entry and during the study at weeks 4, 16, 32, and 48. Patients’ adherence to the prescribed therapy also was recorded by the treating physician at weeks 16, 32, and 48 by means of patient self-report as a percentage of missing doses (i.e., $\geq 50\%$, poor adherence; $50\%–20\%$, fair adherence; and $<20\%$, good adherence) during the week preceding consultation.

HIV-resistance testing and interpretation tools. Resistance tests were performed by VIRCO (r-PHT, Antivirogram; v-PHT, VircogenII [VIRCO]). The versions of the r-PHT and v-PHT tests used during this study in the year 2000 categorized drug resistance into 3 classes ($<4$, $4–10$, and $>10$ fold-resistance) for all drugs (technical cut-offs) and represent an earlier version of the currently available version that uses drug-specific biological cut-offs. Although technical cut-offs have been used for decision making during the study, results were reinterpreted with the new biological cut-offs for the purpose of these analyses [14]. Neither r-PHT nor v-PHT for lopinavir was considered, because this drug was not tested at the time of the study. All mutations reported were imported and reinterpreted using both RetroGram 1.4 (Virology Networks) and TRUGENE HIV-1 (guidelines 3.0; version 1.15; Visible Genetics) for the purposes of this study.

Comparison between systems. Concordance of both RetroGram and TRUGENE interpretations with either r-PHT or v-PHT results was investigated by 2 complementary analyses. First, the $\kappa$ measure of interrate agreement ranked into 5 classes was performed, with $\kappa \geq 0.4$ considered to be significant for agreement [16]. Genotype interpretation by RetroGram results in a 4-category suitability ranking as follows: class A, “can be used;” class B, “consider use if no class A drug available;” class C, “consider use if no class A or B drug available;” and class D, “consider use if no class A, B, or C drug available.” In contrast, the TRUGENE algorithm uses 3 categories to predict drug activity: class S, “no evidence of resistance;” class PR, “possible resistance;” and class R, “resistance.” To be able to compare both systems, 3 different output normalization systems were developed: first, $A = S$, $B = PR$, and $C + D = R$; second, $A + B = S + PR$ and $C + D = R$; and third, $A = S$ and $B + C + D = PR + R$. Because of the concordance found between RetroGram and TRUGENE, a composite output normalization between these 2 systems was created for the comparison with either r-PHT or v-PHT, whose interpretation output uses only 2 classes: 1 (“sensitive”) or 0 (“resistant”). Three further normalizations were used for comparison of rules-based with either r-PHT or v-PHT as follows: fourth, $A = S = 1$, $B + C + D$ and/or $PR + R = 0$; fifth, $A + B = S + PR = 1$, $C + D$, and/or $R = 0$; and sixth, $A + B + C = S + PR = 1$, $D$, and/or $R = 0$.

Second, discordant interpretations were calculated for each drug, ranked in 2 categories as follows: first, sensitive r-PHT or v-PHT result but any degree of drug resistance as measured by RetroGram and/or TRUGENE (P-S/G-R); and second, resistant r-PHT or v-PHT result but samples classed as sensitive according to either the RetroGram or TRUGENE system (P-R/G-S).

The number of samples considered for the concordance analysis did not correspond to the number of patients enrolled into the study, since genotypes and phenotypes were not available for all drugs, because of technical reasons. All concordance analyses were conducted primarily by a conservative approach, considering missing drug-resistance results as resistant. Analyses were repeated, excluding missing drug-resistance results, to provide confirmation.

Correlation with the virological outcome. Clinical success was defined as a plasma virus load $<400$ HIV-1 RNA copies/mL and was analyzed throughout the follow-up period until
week 48. During-treatment (DT) analysis was the primary analysis, including only enrolled patients with data regarding the original treatment combinations. Intent-to-treat (ITT) analysis also was performed using the last-observation-carried-forward (LOCF) method to impute the information for patients still receiving any antiretroviral therapy at any time points.

Drugs prescribed to >20% of patients were as follows: stavudine (73.6% in the r-PHT arm and 70.5% in the v-PHT arm), didanosine (57.5% in the r-PHT arm and 60.2% in the v-PHT arm), booster-dose ritonavir (49.4% in the r-PHT arm and 47.7% in the v-PHT arm), abacavir (35.6% in the r-PHT arm and 31.8% in the v-PHT arm), lamivudine (24.1% in the r-PHT arm and 22.7% in the v-PHT arm), amprenavir (24.1% in the r-PHT arm and 19.3% in the v-PHT arm), and saquinavir (20.7% in the r-PHT arm and 23.9% in the v-PHT arm). DT analysis gave the following rates of virological success without significant differences between r-PHT and v-PHT arms of the GenPhex study: 21 (14.9%) of 141 at week 4, 18 (17.1%) of 105 at week 16, 16 (19.0%) of 84 at week 32, and 13 (40.6%) of 32 at week 48.

To assess the correlation between different interpretation systems and virological outcome, a sensitivity score of 1 was assigned for each drug in the salvage regimen in case of complete sensitivity with any of the systems and zero for any degree of resistance. Sensitivity scores for the prescribed drugs were added up to obtain sensitivity scores for the treatment regimen and then correlated with the treatment response [17]. Ritonavir was not considered in the score because it was always prescribed as a boosting dose.

DT rates of virological success were correlated to the sensitivity scores at week 16 of the follow-up period as the primary study analysis. Week 16 was chosen because more patients were receiving the initial treatment combination that was selected on the basis of resistance results after inclusion into the study and because the length of treatment was considered to be reasonably long enough to judge its antiviral activity. Eighty-seven patients in the r-PHT arm and 86 in the v-PHT arm had started therapy and were considered for this analysis.

Next, several factors were tested by univariate logistic regression model for the association with the virological outcome at week 16: sex; age in years; risk factors for HIV acquisition (intravenous drug use vs. others); Centers for Disease Control and Prevention 1993 clinical stage [18] (class C vs. others); baseline CD4 cell count (>200 vs. <200 cells/mm3); baseline HIV RNA level (<4.7 vs. ≥4.7 log10 copies/mL); duration in years of previous exposure to antiviral therapy; number of previous treatment lines; treatment arm (r-PHT vs. v-PHT); patients’ adherence (poor and fair vs. good); and r-PHT, v-PHT, RetroGram, or TRUGENE sensitivity scores of the new regimens (0 vs. 1, 2, and 3–4).

Missing drug-resistance results have been considered as resistant (score 0). All analyses were repeated, excluding patients with any missing drug-resistance results for the prescribed drugs, to provide confirmation.

**Statistical methods.** The χ² or Fisher’s exact test was used for comparisons of proportions, as appropriate. Logistical analysis (univariate and multivariate models) was used to model the relationship between virological outcome and the sensitivity scores (the predictors or independent variables). The variables found to be significant (P < .2) in the univariate analysis were entered into multivariate models, to identify independently predictive factors. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated. Likelihood ratio tests were used to perform analysis between pairs of maximum-likelihood models; P < .05 was considered to be statistically significant. All tests were 2-sided. Analyses were performed with Statistica for Windows (2001 release; StatSoft) and with STATA (release 7.0; StataCorp).

**RESULTS**

**RetroGram and TRUGENE genotype interpretations.** In the present study, genotype data were used from the GenPhex clinical cohort study. These data were retrospectively analyzed by 2 widely used genotype interpretation systems: RetroGram and TRUGENE.

A high degree of κ agreement was found for most antiretrovirals with the first output normalization, except for abacavir (κ = 0.52) and lopinavir (κ = 0.54). Analysis with the second output normalization system showed no significant differences, with the exception of lopinavir (κ = 0.27). Analysis with the third output normalization, in which all intermediate and full resistance levels were grouped together, showed the best correlations for any of the drugs (figure 1A).

**Rules-based interpretation versus r-PHT or v-PHT interpretations.** κ agreements were not significantly different between the output normalization systems applied, with only modest improvement seen with the fifth and sixth output normalization systems. Results obtained with the fourth output normalization (which also was used for correlation with the virological outcome) are shown in figure 1B. In particular, κ agreement was not significant for stavudine, didanosine, zalcitabine, abacavir, and amprenavir. For most drugs, κ agreement was higher in the comparison with v-PHT, rather than with r-PHT, results. The highest difference between r-PHT and v-PHT was found for zidovudine (κ = 0.34 and κ = 0.79, respectively; figure 1B).

Concordance between the different systems was further investigated as the percentage of discordant interpretations. The percentage of P-R/G-S results (defined as sensitive with either RetroGram or TRUGENE among samples that were classified as resistant with either r-PHT or v-PHT) was <4% for all drugs (data not shown). Conversely, the percentage of P-S/G-R (de-
HIV-Resistance Interpretations

Figure 1. A, $\kappa$ agreement values between RetroGram and TRUGENE. The following output normalizations were used: RetroGram class A = TRUGENE class S, and RetroGram classes B + C + D = TRUGENE classes P + R. B, $\kappa$ agreement values between RetroGram/TRUGENE (composite score) and real or virtual phenotype according to biological cut-offs. $\kappa$ values obtained with the following output normalization are presented: RetroGram class A = TRUGENE class S = real or virtual phenotype interpretation class 1, and RetroGram classes B + C + D = TRUGENE classes P + R = real or virtual phenotype interpretation class 0. 3TC, lamivudine; ABC, abacavir; APV, amprenavir; D4T, stavudine; DDC, zalcitabine; DDI, didanosine; DLV, delavirdine; EFV, efavirenz; IDV, indinavir; LPV, lopinavir; NFV, nelfinavir; NVP, nevirapine; RTV, ritonavir; SQV, saquinavir; ZDV, zidovudine.

fined as resistant with RetroGram and/or TRUGENE among samples that were sensitive with either r-PHT or v-PHT) was >50% for most of the nucleoside reverse-transcriptase inhibitors (zidovudine, 63% and 24%; stavudine, 72% and 68%; didanosine, 54% and 51%; zalcitabine, 57% and 51%; and abacavir, 51% and 49%, for comparison of rules-based interpretations with r-PHT and v-PHT, respectively). Lower P-S/G-R results were noted for both protease inhibitors (except from amprenavir [50% and 38%]) and nonnucleoside reverse-transcriptase inhibitors.

Correlation with the virological outcome. Sensitivity scores were created for each system, as described above. Rules-based systems were more likely to interpret resistance of the prescribed drugs, as demonstrated by the fact that, among patients who started treatment, 41.6% received at least 2 drugs for which HIV-1 was sensitive as judged by TRUGENE, 15.8% by RetroGram, and 74.4% by r-PHT or v-PHT.

At week 16, r-PHT and v-PHT sensitivity scores of the prescribed regimens showed lower association with the virological suppression (overall $\chi^2 = 8.01; P = .005$; figure 2A). By contrast, sensitivity scores of both rules-based systems showed higher significant associations ($\chi^2 = 23.13$ for RetroGram; $\chi^2 = 22.57$ for TRUGENE; $P < .0001$; figure 2B and 2C).

Among factors tested in the univariate analysis, those associated with the virological success were as follows: baseline HIV-1 RNA level (OR, 3.05; 95% CI, 1–9.31; $P = .049$), adherence (OR, 0.33; 95% CI, 0.12–0.95; $P = .039$), and r-PHT or v-PHT (OR, 2.11; 95% CI, 1.01–4.39; $P = .045$), RetroGram (OR, 2.51; 95% CI, 1.43–4.43; $P = .001$), or TRUGENE (OR, 2.61; 95% CI, 1.53–4.45; $P < .001$) sensitivity scores. Results did not change significantly when patients with missing resistance results for any of the drugs ($n = 10$) were excluded from the analysis. In the multivariate analysis, only the rules-based sensitivity scores remained predictive of virological suppression at week 16 (table 1).

Univariate and multivariate analyses at later time points during the GenPherex study confirmed and strengthened the association between sensitivity scores and the virological outcome, both with an ITT-LOCF and a DT approach. For instance, with ITT-LOCF multivariate analysis at the 32–48-week composite time point, the associations observed between sensitivity scores and the virological outcome were as follows: for RetroGram, OR, 2.10; 95% CI, 1.32–3.34; and $P = .002$; for TRUGENE, OR, 2.37; 95% CI, 1.51–3.72; and $P < .0001$; and for r-PHT or v-PHT scores, OR, 1.99; 95% CI, 1.09–3.64; and $P = .025$.

DISCUSSION

Although several large international studies have demonstrated that therapeutic changes guided by resistance testing increase the possibility of clinical success [1–4, 9, 10], it is not clear
whether genotyping or phenotyping methods are a better instrument to choose therapy after virological failure [1–4, 9–12].

The present study was initiated to investigate potential differences between 2 different rules-based interpretation systems [20]. In addition, clinical relevance of differences between rules-based systems and r-PHT and v-PHT was evaluated using a large set of clinical data obtained from the GenPherex study, in which patients were randomly assigned on the basis of either r-PHT or v-PHT to guide treatment changes. Genotype data from this study were used in a retrospective comparative analysis of 2 widely used genotype interpretation systems: RetroGram and TRUGENE.

The concordance analysis presented here shows a good correlation between the 2 rules-based interpretation systems, RetroGram and TRUGENE, even in this cohort of heavily pretreated patients. Of interest, the lowest correlation was found for the less-experienced drugs, abacavir and lopinavir, which suggests that knowledge of their resistance patterns is still incomplete, precluding consensus of expert opinion on their clinical relevance.

The comparison between either r-PHT or v-PHT and rules-based interpretations showed striking discordances. Discordant results differed by drug, with the lowest agreement for stavudine, didanosine, and zalcitabine. Abacavir and amprenavir discordances also were higher in the triple comparison, rather than in the comparison between RetroGram and TRUGENE alone, which confirms that poor characterization of resistance patterns and scarce correlation with phenotypic resistance in the current databases make interpretation more complicated.

However, for most drugs, the degree of concordance between v-PHT results and the 2 rules-based interpretations was higher than that between r-PHT results and the same rules-based interpretations, in particular for zidovudine. Four complementary reasons can be suggested to explain better correlations between rules-based interpretations and v-PHT with respect to r-PHT results: (1) v-PHT considers only selected mutations for the matching of patients’ viruses in the database, and this criterion is more consistent to that of rules-based interpretations; (2) because v-PHT is based on genotypic sequencing, it can be more sensitive in detecting minority resistant subpopulations in the viral mixture [21]; (3) complex resistance patterns in such heavily pretreated patients might not necessarily correspond to phenotypic resistance because of complex interactions between mutations [13]; and (4) different cut-offs between r-PHT and v-PHT have been postulated [14]. Among these possible reasons, the first one is likely to be the best explanation for the better correlation found between v-PHT and rules-based genotype interpretations, compared with that between r-PHT results and rules-based genotype interpretations.

To compare rules-based HIV-1 genotype interpretations with either r-PHT or v-PHT in relation to the clinical outcome, clinical data from the GenPherex study were used. In this retrospective study, logistic regression analysis showed better independent associations between sensitivity scores obtained with rules-based interpretations and the virological response, in contrast to the r-PHT or v-PHT interpretation. One explanation for this finding could be that the biological cut-offs, defined by the 95% CI of the fold-resistance distribution in untreated
Table 1. Multivariate analyses of predictors of virological suppression (<400 human immunodeficiency virus [HIV] RNA copies/mL) in 3 separate models with real or virtual phenotype, RetroGram or TRUGENE sensitivity scores as covariates.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real or virtual phenotype¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study arm</td>
<td>0.55 (0.15–1.99)</td>
<td>.36</td>
</tr>
<tr>
<td>Clinical stage, CDC 1993 C³ vs. others</td>
<td>0.66 (1.17–2.53)</td>
<td>.54</td>
</tr>
<tr>
<td>Baseline CD4 cell count, &gt;200 vs. &lt;200 cells/mm³</td>
<td>2.70 (0.58–12.59)</td>
<td>.20</td>
</tr>
<tr>
<td>Baseline HIV RNA level, &lt;4.7 vs. &gt;4.7 log₁₀ copies/mL</td>
<td>1.85 (0.46–7.51)</td>
<td>.38</td>
</tr>
<tr>
<td>Adherence, poor and fair vs. good</td>
<td>0.19 (0.05–0.76)</td>
<td>.018</td>
</tr>
<tr>
<td>No. of treatment lines experienced</td>
<td>0.92 (0.67–1.28)</td>
<td>.64</td>
</tr>
<tr>
<td>VIRCO sensitivity score, 0 vs. 1, 2, and 3–4</td>
<td>1.91 (0.77–4.76)</td>
<td>.16</td>
</tr>
<tr>
<td>RetroGramc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study arm</td>
<td>0.68 (0.19–2.44)</td>
<td>.55</td>
</tr>
<tr>
<td>Clinical stage, CDC 1993 C³ vs. others</td>
<td>0.79 (0.20–3.19)</td>
<td>.74</td>
</tr>
<tr>
<td>Baseline CD4 cell count, &gt;200 vs. &lt;200 cells/mm³</td>
<td>3.58 (0.72–17.70)</td>
<td>.11</td>
</tr>
<tr>
<td>Baseline HIV RNA level, &lt;4.7 vs. &gt;4.7 log₁₀ copies/mL</td>
<td>1.87 (0.45–7.8)</td>
<td>.39</td>
</tr>
<tr>
<td>Adherence, poor and fair vs. good</td>
<td>0.23 (0.05–0.98)</td>
<td>.04</td>
</tr>
<tr>
<td>No. of treatment lines experienced</td>
<td>0.94 (0.67–1.32)</td>
<td>.71</td>
</tr>
<tr>
<td>Retrogram sensitivity score, 0 vs. 1, 2, and 3–4</td>
<td>2.34 (1.21–4.55)</td>
<td>.01</td>
</tr>
<tr>
<td>TRUGENEd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study arm</td>
<td>0.68 (0.17–2.72)</td>
<td>.58</td>
</tr>
<tr>
<td>Clinical stage, CDC 1993 C³ vs. others</td>
<td>2.55 (0.12–2.45)</td>
<td>.43</td>
</tr>
<tr>
<td>Baseline CD4 cell count, &gt;200 vs. &lt;200 cells/mm³</td>
<td>3.60 (0.66–19.74)</td>
<td>.14</td>
</tr>
<tr>
<td>Baseline HIV RNA level, &lt;4.7 vs. &gt;4.7 log₁₀ copies/mL</td>
<td>1.75 (0.38–7.97)</td>
<td>.47</td>
</tr>
<tr>
<td>Adherence, poor and fair vs. good</td>
<td>0.17 (0.04–0.71)</td>
<td>.01</td>
</tr>
<tr>
<td>No. of treatment lines experienced</td>
<td>0.91 (0.64–1.31)</td>
<td>.62</td>
</tr>
<tr>
<td>VGI sensitivity score, 0 vs. 1, 2, and 3–4</td>
<td>2.90 (1.52–5.52)</td>
<td>.001</td>
</tr>
</tbody>
</table>

NOTE. “Study arm” indicates randomization into the real- or virtual-phenotype arm of the GenPherex study. CDC, Centers for Disease Control and Prevention; CI, confidence interval; OR, odds ratio.

¹ Log likelihood, −33.97 (P = .0073).
² CDC 1993 guidelines were published elsewhere [18].
³ Log likelihood, −33.702 (P = .0008).

patients, do not invariably reflect the actual clinical cut-offs. It has been demonstrated recently that even subtle changes in resistance to drugs extensively used in the study cohort, such as stavudine and didanosine, can have an impact on the virological outcome [22, 23], which indicates that studies aimed at defining clinically relevant cut-offs are urgently needed. It is difficult, however, to infer from our data the actual responsibility of stavudine and didanosine in particular, for the discordant prediction of the virological response, because of their prescription into regimens including multiple drugs for which substantial rates of discordant interpretations also have been found, such as abacavir and amprenavir.

Moreover, additional factors besides the sensitivity scores had an impact on the virological outcome from the results of this analysis. In particular, patients’ adherence and the viroimmunological status at baseline emerged as independent predictors of the virological outcome, but they did not obscure the impact of the rules-based genotypic sensitivity scores at each point of follow-up and with any statistical analyses performed.

It is important, however, to discuss several limitations of the present study. First, in the GenPherex study, r-PHT and v-PHT were used to guide selection of the antiretroviral regimens used. Therefore, comparing these methods of measuring resistance to rules-based systems (which were not used in the study) may have introduced the possibility of bias in this retrospective analysis. In general, it would have been a stronger study if all systems were applied equally—that is, on a data set that was not influenced by the results of the tests being compared. However, in thinking this through, it is important to consider which direction the bias would affect the results. The bias should actually favor the r-PHT or v-PHT over the rules-based interpretations in this comparison because they should be more “overfit.” The fact that technical cut-offs were used in the prospective study while biological cut-offs were used in this analysis may have...
mitigated some of this problem of “overfitted” data, but this difference also may have negatively affected the predictive capacity of the r-PHT or v-PHT. To address this issue, bootstrap analysis was used to determine whether results are “overfit” however, results did not change significantly (data not shown).

Second, updated versions of resistance interpretation algorithms have been introduced into clinical practice after completion of the study. For instance, the 6.0 version of the Visible-Genetics rules are now in use for interpretation of the TRUGENE assay, making the 3.0 results somewhat out of date; newer versions of the rules might provide different or better results. Moreover, it is also important to state that phenotype tests other than those described here are available. More studies are necessary to assess performance of different phenotype methods, compared with rules-based genotype interpretations. In this regard, the prospective ANRS 088 study [11] has demonstrated virological benefit deriving from genotype-driven over phenotype-driven antiretroviral therapy, but only in patients experiencing a first protease inhibitor failure. In addition, regarding the possible use of r-PHT or v-PHT relative to rules-based interpretations, it might be advantageous to have results quantitating the relative resistance to different drugs, especially in heavily pretreated patients whose virus is resistant to most drugs, because it can be helpful to select the less “resistant” drugs for salvage regimens.

Third, it is possible that the better association between rules-based system interpretation and the virological outcome was influenced by the peculiarity of this cohort of patients who received extensive prescription of nucleoside reverse-transcriptase inhibitors, despite extensive pretreatment with this drug class.

In conclusion, in the present study, rules-based genotype interpretation showed better correlation with the virological outcome for a large group of heavily pretreated patients, in contrast to v-PHT or r-PHT interpreted through the biological cut-offs. Further prospective comparative studies are mandatory in patients with different degrees of antiretroviral experience, taking into account the broad spectrum of factors with a potential impact on the clinical outcome. Such a study (Resistance and Dosage Adapted Regimens [RADAR]) is currently under way.

**GENPHEREX STUDY GROUP MEMBERS**

GenPherex Study group members in Italy include F. Mazzotta, S. Lo Caputo, and P. Pierotti (Firenze); G. Carosi, F. Castelli, C. Torti, E. Quiros-Roldan, and L. Tomasoni (Brescia); C. Carnevale and A. Pan (Cremona); R. Maserati and L. Minoli (Pavia); A. Poggio and V. Mondino (Verbania); M. Totti and E. Donati (Grosseto); F. Alberici and M. Sisti (Piacenza); G. Cadeo and D. Vangi (Brescia); A. Chirianni and A. Loiacono (Napoli); A. Lazzarin and N. Gianotti (Milano Hospital San Raffaele); E. Leontini and M. Pozzi (Firenze); V. Vullo (Roma); G. Pastore and N. Ladisa (Bari); D. Dionisio and A. Vivarelli (Pistoia); A. Scasso and M. De Gennaro (Lucca); F. Resta and G. Buccoliero (Taranto); P. Delle Foglie (Trento); F. Ghinelli and L. Sighinolfi (Ferrara); G. Angarano (Foggia); and C. Tinelli and L. Scudeller (Statistical Unit, Pavia).

**References**


