Antivirogram or PhenoSense: a comparison of their reproducibility and an analysis of their correlation

*Kai Wang, Ram Samudrala and John E Mittler**

Department of Microbiology, University of Washington, Wash., USA

*Corresponding author: Tel: +1 206 732 6160; Fax: +1 206 732 6055; E-mail: jmittler@u.washington.edu

The Antivirogram and PhenoSense assays are widely used phenotypic tests for HIV drug resistance. There are limited data on the reproducibility of each assay, and little is known about the correlation between the two. Using data from the Stanford HIV drug resistance database, we performed a comprehensive analysis of the reproducibility of each assay, and calculated the correlation and concordance of the two assays using both general IC_{50} fold change cutoff values and drug-specific cutoff values. Although the within-assay correlations were high (rank correlation coefficients r=0.94 and r=0.95 for the Antivirogram and PhenoSense assays, respectively), the between-assay correlation was considerably lower (r=0.36). Using drug-specific cutoff values for viruses classified as resistant by the Antivirogram or PhenoSense assays, respectively, only 71.4% [95% confidence intervals (95% CI): 58.7–82.1%] and 57.0% (95% CI: 45.3–68.1%) of the samples were classified as resistant using the other assay. The poor agreement between the assays was primarily due to the extremely poor correlation between these assays for samples with low resistance values (r=0.02 and r=0.61 for samples with the Antivirogram measurements lower or higher than 2.0, respectively). Since the cutoff values for both assays are relatively low, our analysis suggests that one should be very careful when interpreting measurements that are near the cutoff values for drug resistance.

Introduction

In recent years, phenotypic and genotypic resistance tests have become important tools in optimizing combination therapies for treating HIV-infected individuals [1]. Genotypic assays determine the presence of mutations known to confer decreased drug susceptibility, while phenotypic assays measure the susceptibility of the virus by determining the concentration of drug that inhibits viral replication in tissue culture [2]. The Antivirogram [3] and PhenoSense [4] assays, developed by Virco (Mechelen, Belgium) and Virologic (South San Francisco, Calif., USA), respectively, are both automated and commercialized recombinant virus-based phenotypic assays for HIV drug resistance. During the past few years, there have been a number of studies designed to evaluate the clinical benefit of drug resistance testing [5,6]. In most studies, clinical decision-making guided by genotypic data provided both virological and immunological benefits. However, phenotypic assays have been evaluated in only a few clinical trials and have not demonstrated such clear, favourable virological outcomes [5]. In fact, phenotypic testing has been shown to be clinically useful in just one of four prospective randomized clinical trials [7]. This has prompted us to carry out detailed analyses on the reported resistance values of these two phenotypic assays.

Previous studies conducted to evaluate the reproducibility of these two widely used phenotypic assays have been reviewed in [8]. Internal evaluations of Virco's Antivirogram assay showed a 1.2–2.5-fold variation in 50% inhibitory concentration (IC_{50}) values of 16 samples evaluated in 10 independent experiments [9]. Internal evaluation of Virologic's PhenoSense assay showed less than 2.5-fold variation in 99% of the 107 drug-susceptibility determinations for nine samples [4]. Another study on 39 samples showed 92% concordance on duplicated measurements 30 days apart by the PhenoSense assay, using 2.5 as the IC_{50} fold change cutoff value [10]. However, because different research groups used different sample collection and preparation protocols, these tests may underestimate the real-world variability in resistance testing. Since every clinic is different, cross-group reproducibility is more relevant to clinical settings than the reproducibility of multiple assays performed on the same samples and reported by the same research group.

In addition, because these two assays differ in various technical aspects, little is known about how

they correlate and concord with each other [8]. Although a comparative analysis of 50 plasma specimens demonstrated a 91.5% concordance between the Antivirogram and PhenoSense assays [11], this high concordance is misleading because of the high fraction of drug-susceptible virus in the study. Recently, we reanalysed these data and showed that the poor sampling method in this study prohibited drawing any definitive conclusions concerning the correlation or concordance of these two assays [12].

The Stanford HIV drug resistance database [13] contains a compilation of nearly all published, as well as some unpublished, HIV protease (PR) and reverse transcriptase (RT) sequences and, for many of the sequences, the corresponding IC_{50} fold change values measured by either the Antivirogram or the PhenoSense assay. Since the database collects data from many different sources, there are dozens of cases in which different patients with the same PR or RT sequence were measured by the same phenotypic assay in two different publications. Such data comprised paired records of repetitive measurements on the same sequence reported by different research groups, and we performed a comprehensive analysis of the reproducibility of the two assays using these paired records. There are also many cases in which the same sequence was analysed by two different phenotypic resistance assays, and such data comprised paired records of Antivirogram–PhenoSense measurements. We collected these paired records and performed a comprehensive analysis of the correlation and concordance of the two assays.

Methods

Data source

The raw dataset (v 1.2) used in our study was retrieved from the Stanford HIV drug resistance database (http://hivdb.stanford.edu/cgi-bin/GenoPhenoDS.cgi). The dataset contained genotypic sequence data and the corresponding phenotypic data for HIV-infected patients collected from various publications or unpublished studies. The phenotypic data were represented by the IC_{50} fold change over the wild-type virus with subtype B consensus sequence, and were determined by either Virologic's PhenoSense assay or Virco's Antivirogram assay. We discarded those data entries that did not contain exact quantitative IC_{50} fold change values.

The manufacturer-recommended drug-specific cutoff values were retrieved from Virologic's (http://www.phenosense.com) and Virco's (http://www.vircolab.com) websites in January 2004. Clinical cutoffs were used for five drugs [abacavir (ABC), didanosine (ddI), stavudine (d4T), tenofovir (TFV) and lopinavir (LPV)] for the PhenoSense assay, while all other cutoffs were biological cutoffs. Neither company gave a cutoff value for the relatively new drug atazanavir (ATV), so we arbitrarily set it to be 2.5 for both assays. Only seven ATV measurements are used in the reproducibility study and no ATV measurements are used in the correlation study.

Construction of paired records

For the reproducibility study, we first retrieved sequence data and the corresponding phenotypic data for each assay and each drug from the downloaded dataset. When two samples had the same PR or RT sequence and were measured by the same assay against the same drug, we constructed a paired record consisting of the IC_{50} fold change values for the two samples. In a few cases where three or more samples with the same sequence were measured by the same assay against the same drug, we constructed paired records between the first sample and all other samples.

For the correlation study, we constructed paired records of IC_{50} fold change values for samples that had the same PR or RT sequence and were phenotyped by both the Antivirogram and PhenoSense assays against the same drug.

Genotypic interpretation

Genotypic interpretation was performed for one sequence to detect possible data entry errors in the dataset we used. The HIVdb tool [21], available at http://hivdb.stanford.edu, gives a qualitative drug resistance interpretation for a given PR or RT sequence. The HIVlr tool [15], available at http://protinfo.compbio.washington.edu/pirspred, gives a quantitative prediction of IC_{50} fold change values.

Data analysis

All data analysis was performed using the statistics software STATA version 7 (Stata Corp, College Station, Tex., USA).

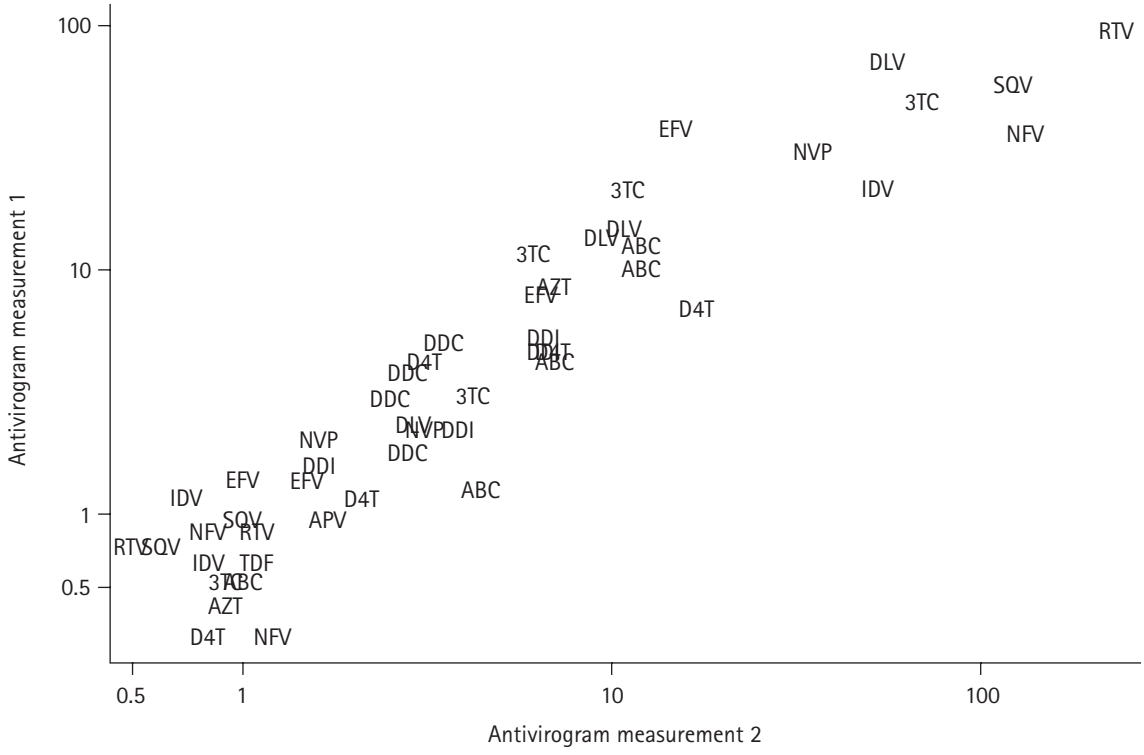
Results

Reproducibility of the individual phenotypic assays

We downloaded the most recent drug resistance dataset from the Stanford HIV drug resistance database, and collected paired records of repeated measurements by the same phenotypic assay on the same sequence. Based on the assumption that a phenotypic assay has comparable reproducibility for different anti-HIV drugs, we combined paired records for all drugs together in our analysis.

For the Antivirogram assay, we found 50 paired records of repeated measurements. The Spearman rank correlation coefficient between two measurements for these records was 0.94, indicating good correspondence between repeated measurements (Figure 1). The fold variation between each pair (calculated as the

Figure 1. Scatter plot of repeated IC_{50} fold change measurements on the same sequence using the Antivirogram assay



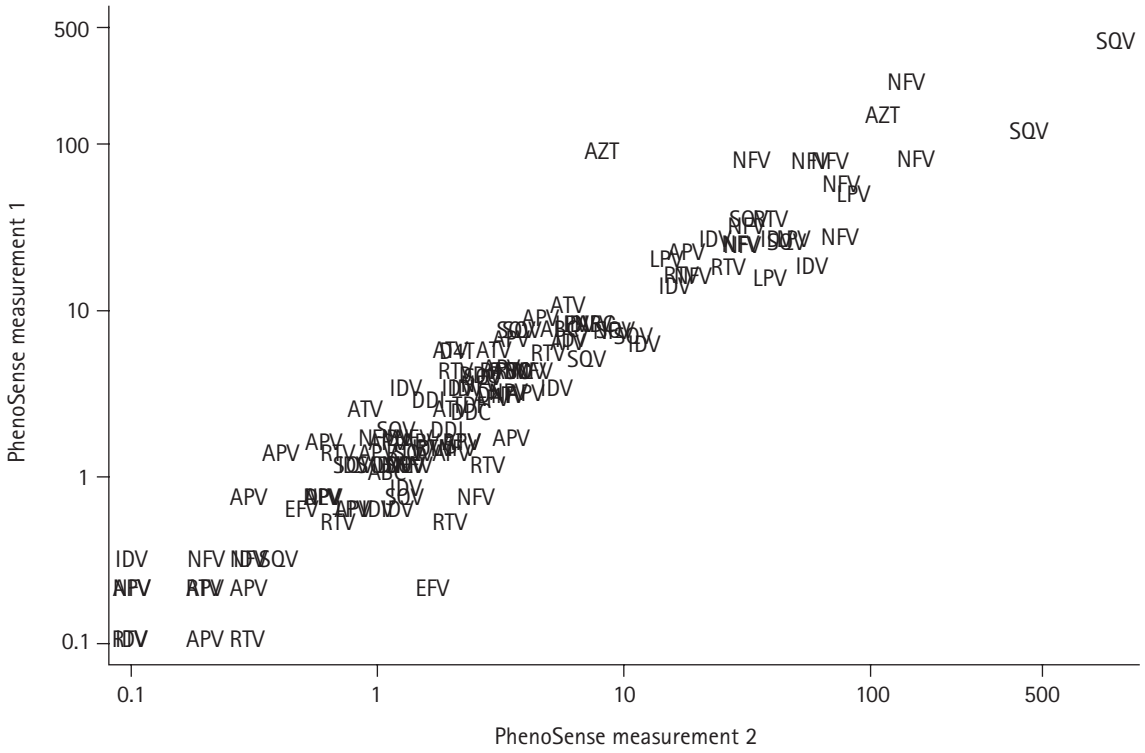
The X-axis and Y-axis represent IC_{50} fold change values on a log scale. Each point in the figure is represented by the name of the drug that the assay is based on. In general, the Antivirogram assay has good reproducibility, with good correlation between repeated measurements. ABC, abacavir; APV, amprenavir; ATV, atazanavir; AZT, zidovudine; d4T, stavudine; ddC, zalcitabine; ddI, didanosine; DLV, delavirdine; EFV, efavirenz; IDV, indinavir; LPV, lopinavir; NFV, nelfinavir; NVP, nevirapine; RTV, ritonavir; SQV, saquinavir; TDF, tenofovir; 3TC, lamivudine.

larger measurement over the smaller measurement) ranged from 1.0–4.0-fold with a median of 1.5-fold. Overall, 86.0% (43/50) of the paired records had less than 2.5-fold variation. Variability in lower measurements may be more likely to produce a higher fold variation. For those paired records that had both IC_{50} fold change measurements higher than 4.0, 81.0% (17/21) had less than 2.5-fold variation.

For the PhenoSense assay, we found 141 paired records using the same method. Since one sequence in the repeated PhenoSense measurement dataset had a very high fold variation (68.5-fold, due to reported IC_{50} fold changes of 0.4 and 27.4), we took a closer look at this sequence, which was measured against the drug delavirdine (DLV) and contained the following mutations: 7PT, 20R, 41L, 44D, 60I, 67N, 75M, 123E, 142V, 162Y, 184V, 196E, 197K, 203D, 208F, 210W, 211K, 215Y, 219N, 251I, 272P and 293V. The first measurement was collected from a published paper [14], while the second measurement was collected from unpublished work by database curators. Genotypic interpretation by the HIVdb tool [13] gave a prediction of 'susceptible' to DLV for this genotype, and another genotypic interpretation by the HIVlr tool [15] gave a quantitative prediction for the IC_{50} fold change of 1.06. Therefore, we concluded that there might be a data entry error for the second measurement and deleted this paired record from our analysis. The Spearman rank correlation coefficient between two measurements for the remaining records was 0.95, indicating good correspondence between repeated measurements (Figure 2). The variation between the paired records ranged from 1.0–10.3-fold with a median of 1.3-fold. Overall, 87.9% (123/140) of the paired records had less than 2.5-fold variation. For those paired records that had both IC_{50} fold change measurements higher than 2.5, 87.0% (67/77) had less than 2.5-fold variation. The high within-assay correlations and low fold-variation values for these phenotypic tests support our assumption that both assays have comparable reproducibility for different anti-HIV drugs.

To compare the reproducibility of these two assays, we plotted the cumulative distribution of fold variation for both assays (Figure 3), and failed to reject the null

Figure 2. Scatter plot of repeated IC_{50} fold change measurements on the same sequence using the PhenoSense assay



The X-axis and Y-axis represent IC_{50} fold change values on a log scale. Each point in the figure is represented by the name of the drug that the assay is based on. In general, the PhenoSense assay has good reproducibility, with good correlation between repeated measurements. ABC, abacavir; APV, amprenavir; ATV, atazanavir; AZT, zidovudine; d4T, stavudine; ddC, zalcitabine; ddI, didanosine; DLV, delavirdine; EFV, efavirenz; IDV, indinavir; LPV, lopinavir; NFV, nelfinavir; NVP, nevirapine; RTV, ritonavir; SQV, saquinavir; TDF, tenofovir; 3TC, lamivudine.

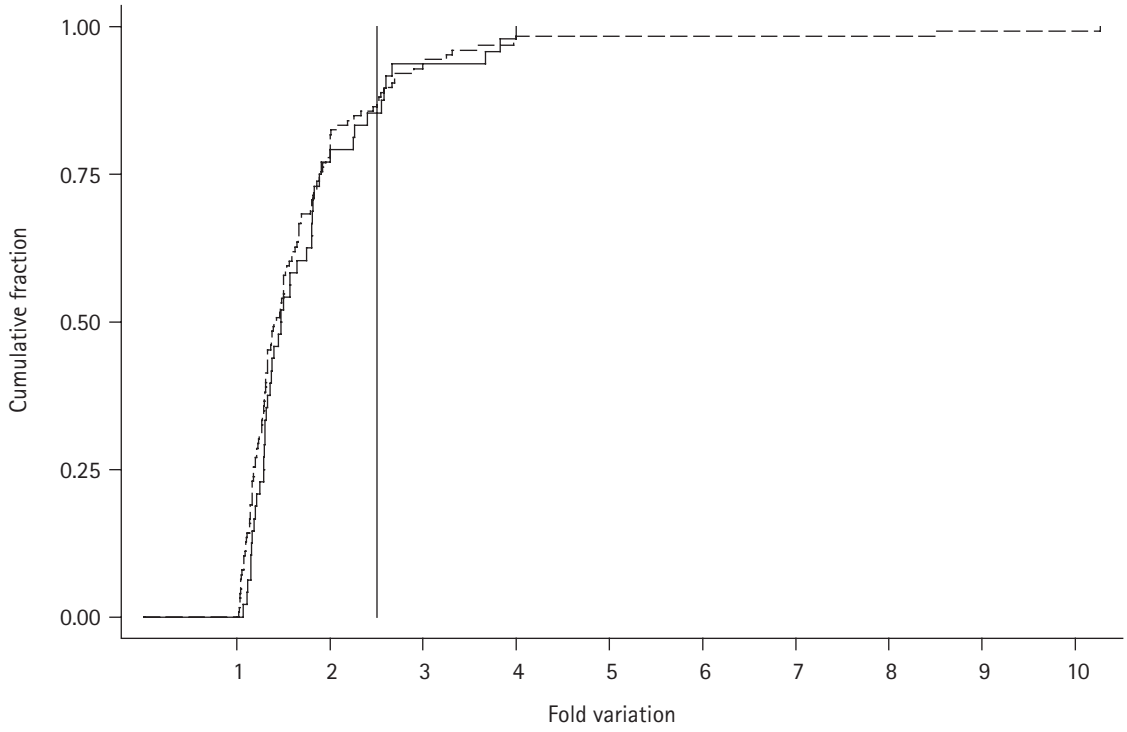
hypothesis that the two assays had same distribution of fold variation using the Wilcoxon rank sum test (P=0.13).

Two different kinds of IC_{50} fold change cutoff values are usually used for defining whether a given sequence is susceptible or resistant to anti-HIV drugs. The first kind is selected on the basis of assay performance characteristics, and adopts the same values for all drugs. These general cutoffs have been set to 4.0 for the Antivirogram assay and 2.5 for the PhenoSense assay [8]. For the Antivirogram assay, using the general IC_{50} fold change cutoff values of 4.0 for binary classification of susceptible versus resistant, 92.0% (46/50, 95% CI: 80.8–97.8%) of the paired measurements were concordant. For the PhenoSense assay, using the general IC_{50} fold change cutoff values of 2.5 for binary classification of susceptible versus resistant, 92.1% (129/140, 95% CI: 86.4–96.0%) of the paired measurements were concordant. The second kind of cutoff is a biological or clinical cutoff that adopts different values for different drugs (Table 1). Biological cutoffs reflect the normal variation in susceptibility observed for wild-type viruses, while clinical cutoffs indicate the values beyond which the probability of a clinical response begins to decline. These drug-specific cutoffs are derived from statistical analysis of clinical isolates and are considered to be more clinically relevant [16,17]. For the Antivirogram assay, 92.0% (46/50, 95% CI: 80.8–97.8%) of the paired measurements were concordant using these drug-specific cutoffs. For the PhenoSense assay, 87.1% (122/140, 95% CI: 80.4–92.2%) of the paired measurements were concordant using these drug-specific cutoffs.

Correlation and concordance between the two phenotypic assays

Besides the reproducibility of individual phenotypic assays, it is interesting to know how the Antivirogram and the PhenoSense assays correlate with each other. We collected 516 Antivirogram–PhenoSense paired records from the Stanford HIV drug resistance database. Overall, the Spearman rank correlation coefficient

Figure 3. Comparison of fold variation for repeated measurements between the Antivirogram and PhenoSense assays



The value on the Y-axis represents the cumulative fraction of paired records that have a fold variation less than the threshold value indicated on the X-axis. A vertical line is drawn to represent the 2.5-fold variation for both assays. The solid line and dashed line represent Antivirogram and PhenoSense assays, respectively. The distribution of fold variations for the two assays are very similar.

between the two assays for these records was 0.36. For PR and RT inhibitors, the rank correlation coefficients between the two assays were 0.35 and 0.37, respectively. The low correlation between the two assays was consistent with our previous findings [12]. We plotted the PhenoSense measurements against Antivirogram measurements (Figure 4), and superimposed it with a LOWESS (locally weighted scatter plot smoothing) [18] smooth curve. Judging from the smooth curve, a clear linear relationship between the two assays can only be observed when Antivirogram measurements were higher than 2.0. In fact, the Spearman rank correlation coefficient between the two assays was 0.02 for samples that had Antivirogram measurements lower than 2.0 and 0.61 for samples that had Antivirogram measurements higher than 2.0. For paired records that were classified as resistant by both assays (Antivirogram and PhenoSense measurements higher than 4.0 and 2.5, respectively), the rank correlation coefficient between the two assays was 0.83. For paired records that were classified as sensitive by both assays, the rank correlation coefficient was 0.12. Thus the low overall correlation between the two assays was mainly due to the presence of many samples with low resistance values.

Next, we investigated the concordance of the two assays by comparing the binary prediction of whether a virus was susceptible or resistant to a drug. Using the general IC_{50} fold change cutoff values of 4.0 and 2.5 for Antivirogram and PhenoSense assays, respectively, the overall concordance between the two assays was 90.5% (467/516); however, this was mainly due to the large fraction of drug-susceptible samples (Figure 4). In clinical practice it is more important to have high concordance when the samples are resistant to drugs. Of the 67 measurements that were defined as resistant by the Antivirogram assay, 65.7% (44/67, 95% CI: 53.1–76.8%) were defined as resistant by the PhenoSense assay. Of the 70 measurements that were defined as resistant by the PhenoSense assay, 62.9% (44/70, 95% CI: 50.5–74.1%) were defined as resistant by the Antivirogram assay. This yielded an average concordance of 64.3% when one of the paired samples was defined as resistant using the other assay. When we

Table 1. Summary of the concordance between the Antivirogram and PhenoSense assays using manufacturer-provided drug-specific cutoff values

Drug class	Drug name	Number of paired records	Number of concordant paired records	Antivirogram cutoff	Both resistant/ Antivirogram-resistant samples	PhenoSense cutoff	Both resistant/ PhenoSense-resistant samples
PI	APV	23	23	2.5	2/2	2.5	2/2
	ATV	NA	NA	2.5	NA	NA	NA
	IDV	41	39	3.0	3/5	2.5	3/3
	LPV	2	1	2.5	0/1	10.0	0/0
	NFV	41	36	4.0	6/8	2.5	6/9
	RTV	41	35	3.5	4/6	2.5	4/8
	SQV	40	34	2.5	2/7	2.5	2/3
NNRTI	DLV	41	37	10.0	6/6	2.5	6/10
	EFV	44	40	6.0	7/7	2.5	7/11
	NVP	40	35	8.0	5/6	2.5	5/9
NRTI	3TC	30	26	4.5	2/2	1.5	2/6
	ABC	35	31	3.0	0/2	4.5	0/2
	AZT	34	33	4.0	5/5	2.2	5/6
	d4T	35	32	3.0	1/1	1.7	1/4
	ddC	35	31	3.5	2/4	1.7	2/4
	ddI	34	31	3.5	0/1	1.7	0/2
	TDF	NA	NA	3.0	NA	1.4	NA
Total		516	464		45/63		45/79

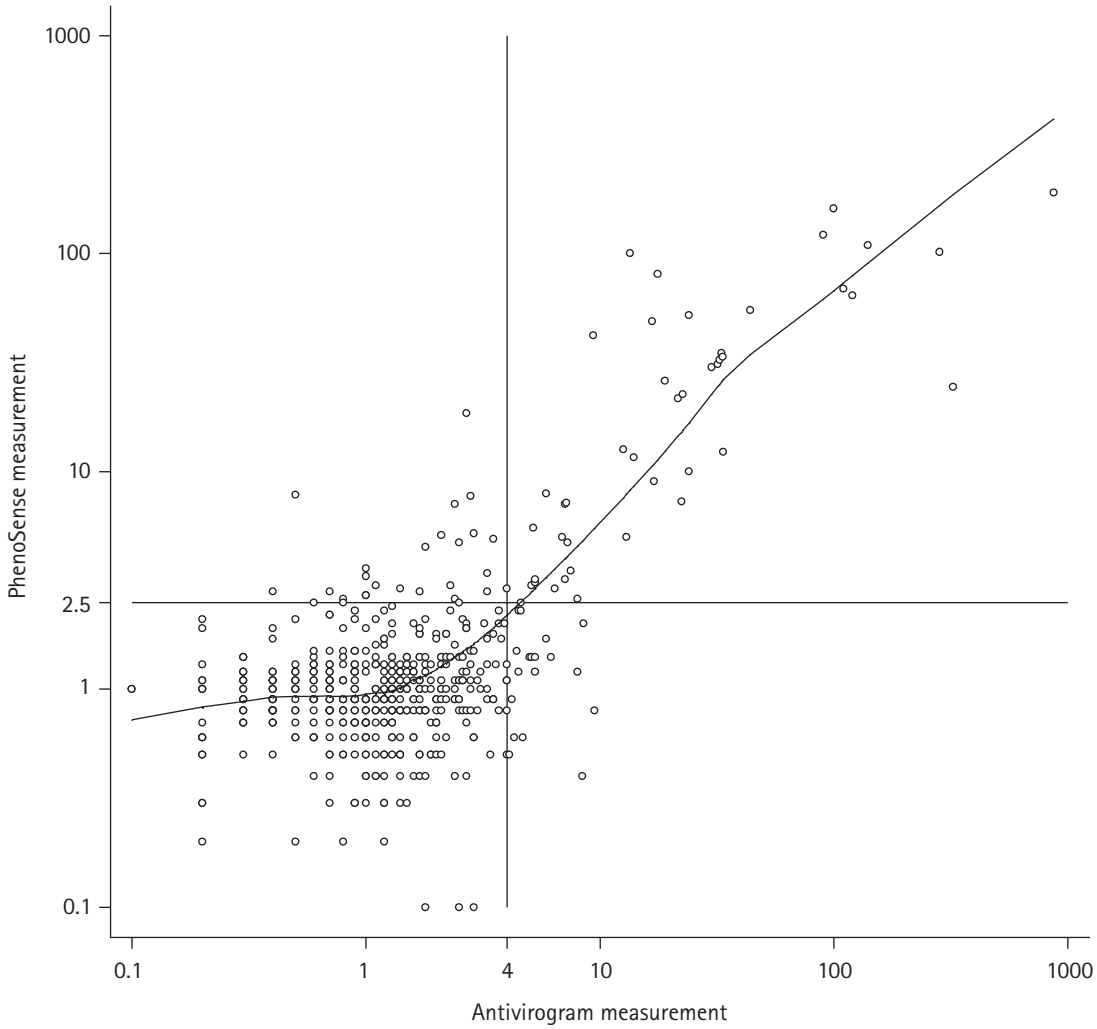
'Both resistant' means that the samples were defined as resistant by both assays; 'Antivirogram-resistant' or 'PhenoSense-resistant' means that the samples were defined as resistant by the corresponding assay. For samples that were classified as resistant by either assay, there is limited concordance between the Antivirogram assay and the PhenoSense assay. ABC, abacavir; APV, amprenavir; ATV, atazanavir; AZT, zidovudine; d4T, stavudine; ddC, zalcitabine; ddI, didanosine; DLV, delavirdine; EFV, efavirenz; IDV, indinavir; LPV, lopinavir; NFV, nelfinavir; NVP, nevirapine; RTV, ritonavir; SQV, saquinavir; TDF, tenofovir; 3TC, lamivudine.

did the equivalent calculations for the within-assay comparisons, we obtained average concordances of 91.3% and 91.7% for the Antivirogram–Antivirogram and PhenoSense–PhenoSense comparisons, respectively, for cases in which one of the paired samples was defined as resistant using the same assay.

Our analysis above demonstrated that the Antivirogram and PhenoSense assays had low concordance with each other when samples were defined as resistant by either assay. Although the cutoffs are based on biological data, one may suspect that the low concordance could be due to the fact that these assays use different cutoffs. To exclude this possibility, we compared the probability of samples being defined as resistant by the two assays. Using the general cutoff values of 4.0 and 2.5 for Antivirogram and PhenoSense assays, respectively, the odds ratio of being defined as resistant by Antivirogram assay versus PhenoSense assay was 0.88 (95% CI: 0.48–1.61), and we failed to reject the null hypothesis that samples had equal probability of being defined as resistant by two assays (McNemar test, P=0.78). To further address this 'difference in cutoff' issue, we redid the concordance analysis using a uniform cutoff value of 3.25 (average of 2.5 and 4.0) for both assays. With this cutoff, the overall concordance was 89.1%, although this, too, was mainly due to the presence of a large fraction of susceptible samples. Of the 84 measurements that were defined as resistant by the Antivirogram assay, only 45.2% (38/84, 95% CI: 34.3–56.5%) were defined as resistant by the PhenoSense assay. Of the 48 measurements that were defined as resistant by the PhenoSense assay, 79.2% (38/48, 95% CI: 65.0–89.5%) were defined as resistant by the Antivirogram assay. Overall, this analysis yielded an average concordance of 62.2% for cases in which one of the paired samples was defined as resistant using the other assay – a value that was, if anything, slightly worse than that obtained using the general assay-specific cutoffs.

To address the possibility that these concordances were low because the general cutoff values ignored differences between drugs, we also analysed the concordance between the two assays using manufacturer-provided drug-specific cutoff values (Table 1). The overall concordance between the two assays with these cutoffs, which may be the most clinically relevant since these are what are provided to the clinician, was 89.9% (464/516), a percentage which, once again, was mainly due to a large fraction of susceptible samples. Of the 63 measurements that were defined as resistant by the Antivirogram assay, 71.4% (45/63, 95% CI: 58.7–82.1%) were defined as resistant by the PhenoSense assay. Of the 79 measurements

Figure 4. Scatter plot of IC_{50} fold change values obtained from the PhenoSense assay versus those obtained from the Antivirogram assay on the same sequence



The X-axis and Y-axis both represent IC_{50} fold change values in log scale. A LOWESS smooth curve generated by robust locally weighted regression is superimposed on the scatter plot. A horizontal line is drawn to represent the general IC_{50} fold change cutoff value of 2.5 for the PhenoSense assay, and a vertical line is drawn to represent the general IC_{50} fold change cutoff value of 4.0 for the Antivirogram assay. The two phenotypic assays have limited correlation with each other, though a linear relationship can be observed for samples that have IC_{50} fold change values higher than 2.0 by the Antivirogram assay. For sequences that are classified as resistant by either assay, the two assays have low concordance.

that were defined as resistant by the PhenoSense assay, 57.0% (45/79, 95% CI: 45.3–68.1%) were defined as resistant by the Antivirogram assay. This yielded an average concordance of 64.2% when one of the paired samples was defined as resistant using the other assay, a value very similar to that obtained using the general drug-independent cutoffs.

We also analysed the relative contribution of different drug classes to the discordance between these two phenotypic assays. Using general cutoff values for nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleotide reverse transcriptase inhibitors (NNRTIs) and protease inhibitors (PIs), of the 67 measurements that were defined as resistant by the Antivirogram assay, 53.8% (7/13), 64.7% (22/34) and 75.0% (15/20) were defined as resistant by the PhenoSense assay, respectively. Using the same criteria, of the 70 measurements that were defined as resistant by the PhenoSense assay, 50.0% (7/14), 73.3% (22/30) and 57.7% (15/26) were defined as resistant by the

Antivirogram assay. Because of the relatively small sample size, the differences in accuracy between drug classes are not statistically significant ($P>0.017$ for all pairwise comparisons) using Fisher's exact test and Bonferroni adjustments.

We further investigated the possible cause of the observed low concordance between these two phenotypic assays. Most of the records studied here came from a single publication [11], which applied these phenotypic assays to samples from drug-naive patients and patients who were the sources of occupational exposures to HIV-1. In other words, most of these records consisted of viral samples that were not necessarily expected to be resistant to drugs. As a result, our paired records contained a very large fraction of susceptible samples (Figure 4). In fact, using general IC_{50} fold change cutoff values, 87.0% and 86.4% of the paired records were defined as susceptible by the Antivirogram and PhenoSense assays, respectively. Because many of these samples fell within the region of low Antivirogram–PhenoSense correlation and because the cutoff values also fell within this region, the low concordance observed in our study was not surprising. In clinical practice, the correlation and concordance between these assays will depend on the distribution of resistance levels in the target population. For patients with multiple therapy failures, many of the samples will be highly resistant to drugs, resulting in higher concordances and correlations between these assays. For therapy-naive or recently infected patients, however, where the prevalence of drug resistance is comparable with that of our paired records [19,20], these phenotypic assays will give correlations and concordances similar to those reported here.

Discussion

Based on our analysis of publicly available data, we conclude that both phenotypic assays have high cross-group reproducibility and that there is no evidence that one assay is more reproducible than the other. However, for samples in the Stanford HIV drug resistance database that had been phenotyped by both assays, the two assays were poorly correlated with each other and had limited concordance for samples that are defined as resistant by either assay. The low between-assay correlation and concordances are explained primarily by the poor correlation between the assays for susceptible samples. In clinical practice, where these assays are likely to be performed on patients in whom drug resistance is suspected, the correlation and concordance between the assays may be higher than those reported here. However, our analyses clearly demonstrate that one should be cautious when making decisions based on phenotypic measurements that are near the cutoff values for these assays.

Possible reasons for the poor agreement between the Antivirogram and PhenoSense assays for susceptible samples include variation in individual assay results, the possibility that testing methodologies may have changed over time, the possibility that the database contains incomplete or incorrect genotype or phenotype records and the fact that these assays use different testing methodologies. Variation in the individual assays cannot explain the poor agreement between these assays since both have good reproducibility. Although changes in testing methodology (and the reported resistance values) are a theoretical possibility, we have no evidence that either of these assays has changed significantly over time. The good reproducibility of both assays also argues against changes in testing methodology as an explanation for the poor concordance between the assays near the cutoff values. The Stanford HIV drug resistance database has been well maintained and updated over the years [13,21–23], and we expect the database errors to be kept to a minimum. Thus, we believe that the most likely explanation is the presence of numerous technical differences between the two assays. When the measurements are dependent on multiple variables (or technical aspects) and the relationships between measurements and each variable are divergent across assays, the reported numeric resistance values will not correlate well. In fact, the two assays vary in their approach to viral replication, the type of cell culture used for viral growth, the methods of transfection, the number of replication cycles, the means of measuring viral growth and even the viral genome region used for analysis.

Although we have performed analysis on almost all publicly available data, the database size that we have used is still not big enough, and the distribution of the resistance values in the database may not resemble that in the clinical settings. This is especially important for the Antivirogram–PhenoSense correlation study, since in a previous study we demonstrated poor correlation between these two assays using limited data from a single publication [12]. In the current study we tried to calculate the correlation and concordance of these two methods for samples that were defined as resistant by either assay, but we still did not have many additional data compared with our previous study. Since the Stanford HIV drug resistance database continually collects data from various sources, the same analysis can be repeated in the future once the size of the database increases.

We have previously demonstrated that both assays have poor concordance with each other using general IC_{50} fold change cutoff values [12], and one may suspect that this might have been due to the use of

non-optimal cutoff values. In this report, we have shown that concordances remain low even if we use the most recent drug-specific cutoff values (some of which are clinical cutoffs) provided at the manufacturer's websites. This suggests a problem with one or, perhaps, both of these assays. Resolution of these discrepancies could lead to further improvements in clinical outcomes in cases where physicians have used phenotypic tests to guide selection of combination therapy regimens. Furthermore, in recent years some HIV genotypic interpretation algorithms have been developed based solely on resistance values measured by a certain phenotypic assay [15,24–26]. Construction of genotypic models that are built upon each phenotypic assay may lead to more reliable consensus genotypic predictions of HIV drug resistance.

In the absence of multicentre prospective data that compares and supports the use of a phenotypic or genotypic assay over the other [6,27], we propose that phenotypic assay results that are near the cutoff values be treated with caution. One possibility for further research would be to compare the efficacy of current genotypic and phenotypic tests with tests that combine genotypic and phenotypic information. It has been shown that phenotypic and genotypic assays give complementary results; their combined use may provide additional clinically relevant information to guide the choice of combination regimens [28–31]. In fact, Virologic now offers dual genotypic and phenotypic testing on a single patient sample. This assay, called PhenoSense GT, combines PhenoSense and GeneSeq testing to identify all relevant resistant-associated mutations and potential viral mixtures while assessing the phenotypic effect of the combination of these mutations on individual drugs [5]. Virco also offers a genotypic resistance test called *Virtual*Phenotype, which provides a probabilistic estimate of the phenotype by matching the genotype with a large proprietary database of samples that have been both genotyped and phenotyped [17]. The assays needed to conduct more comprehensive tests of drug resistance assays are therefore already available for both of these commercial assays. When phenotypic assay results are available, the supplemented use of genotypic assays may lead to more comprehensive and more accurate predictions of HIV drug resistance.

Acknowledgements

This work was supported by NIH grant R21-AI52063-01 (JEM) and a Searle Scholar's Award (RS). We thank the authors of the published HIV resistance datasets for making their data publicly available. We thank the curators of the Stanford HIV drug resistance database for collecting and compiling the drug resistance dataset used in our study.

References

1. Aslanzadeh J. HIV resistance testing: an update. *Annals of Clinical & Laboratory Science* 2002; **32:**406–413.
2. Hanna GJ & D'Aquila RT. Clinical use of genotypic and phenotypic drug resistance testing to monitor antiretroviral chemotherapy. *Clinical Infectious Diseases* 2001; **32:**774–782.
3. Hertogs K, de Bethune MP, Miller V, Ivens T, Schel P, Van Cauwenberge A, Van Den Eynde C, Van Gerwen V, Azijn H, Van Houtte M, Peeters F, Staszewski S, Conant M, Bloor S, Kemp S, Larder B & Pauwels R. A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency virus type 1 isolates from patients treated with antiretroviral drugs. *Antimicrobial Agents & Chemotherapy* 1998; **42:**269–276.
4. Petropoulos CJ, Parkin NT, Limoli KL, Lie YS, Wrin T, Huang W, Tian H, Smith D, Winslow GA, Capon DJ & Whitcomb JM. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrobial Agents & Chemotherapy* 2000; **44:**920–928.
5. Wilson JW. Update on antiretroviral drug resistance testing: combining laboratory technology with patient care. *AIDS Reader* 2003; **13:**25–30, 35–38.
6. Hirsch MS, Brun-Vézinet F, Clotet B, Conway B, Kuritzkes DR, D'Aquila RT, Demeter LM, Hammer SM, Johnson VA, Loveday C, Mellors JW, Jacobsen DM & Richman DD. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an International AIDS Society-USA Panel. *Clinical Infectious Diseases* 2003; **37:**113–128.
7. Shafer RW. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clinical Microbiology Reviews* 2002; **15:**247–277.
8. Demeter L & Haubrich R. Phenotypic and genotypic resistance assays: methodology, reliability, and interpretation. *Journal of Acquired Immune Deficiency Syndromes* 2001; **26:**S3–S9.
9. Vingerhoets J, Bloor S, Michiels L, Scholliers A, Derhertogh P, Van Den Eynde C, De Vroey V, Van Cauwenberge A, de Koning I, Alcorn T, Larder B & Hertogs K. The accuracy and reproducibility of high throughput genotypic and phenotypic HIV-1 resistance testing under EN45001 and CLIA accreditation labels. *3rd International Workshop on Drug Resistance*. 23–26 June 1999, San Diego, Calif., USA. Abstract 77.
10. Haubrich R, Kemper C, Witt M, Keiser P, Dube M, Forthal D, Currier J, Seefried E, Hwang J, McCutchan JA, Hellmann N, Heilek G, Kie Y & Richman D. Reproducibility of an HIV phenotype resistance assay in clinical practice. *39th Interscience Conference on Antimicrobial Agents & Chemotherapy*. 26–29 September 1999, San Francisco, Calif., USA. Abstract 417.
11. Qari SH, Respess R, Weinstock H, Beltrami EM, Hertogs K, Larder BA, Petropoulos CJ, Hellmann N & Heneine W. Comparative analysis of two commercial phenotypic assays for drug susceptibility testing of human immunodeficiency virus type 1. *Journal of Clinical Microbiology* 2002; **40:**31–35.
12. Wang K, Samudrala R & Mittler JE. Weak agreement between predictions of 'reduced susceptibility' from Antivirogram and Phenosense assays. *Journal of Clinical Microbiology* 2004; **42:**2353–2354.
13. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J & Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research* 2003; **31:**298–303.
14. Gonzales MJ, Machekano RN & Shafer RW. Human immunodeficiency virus type 1 reverse-transcriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a northern California clinic-based population. *Journal of Infectious Diseases* 2001; **184:**998–1006.
15. Wang K, Jenwitheesuk E, Samudrala R & Mittler JE. Simple linear model provides highly accurate genotypic predictions of HIV-1 drug resistance. *Antiviral Therapy* 2004; **9:**343–352.

16. Miller V. Interpretation of resistance assay results. *Antiviral Therapy* 2001; **6:**1–9.
17. Harrigan PR, Montaner JS, Wegner SA, Verbiest W, Miller V, Wood R & Larder BA. World-wide variation in HIV-1 phenotypic susceptibility in untreated individuals: biologically relevant values for resistance testing. *AIDS* 2001; **15:**1671–1677.
18. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 1979; **74:**829–836.
19. Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC, Koup RA, Mellors JW, Connick E, Conway B, Kilby M, Wang L, Whitcomb JM, Hellmann NS & Richman DD. Antiretroviral-drug resistance among patients recently infected with HIV. *New England Journal of Medicine* 2002; **347:**385–394.
20. Duwe S, Brunn M, Altmann D, Hamouda O, Schmidt B, Walter H, Pauli G & Kucherer C. Frequency of genotypic and phenotypic drug-resistant HIV-1 among therapy-naive patients of the German Seroconverter Study. *Journal of Acquired Immune Deficiency Syndromes* 2001; **26:**266–273.
21. Kantor R, Machekano R, Gonzales MJ, Dupnik K, Shapiro JM & Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database: an expanded data model integrating natural language text and sequence analysis programs. *Nucleic Acids Research* 2001; **29:**296–299.
22. Shafer RW, Stevenson D & Chan B. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research* 1999; **27:**348–352.
23. Shafer RW, Jung DR, Betts BJ, Xi Y & Gonzales MJ. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research* 2000; **28:**346–348.
24. Beerenwinkel N, Schmidt B, Walter H, Kaiser R, Lengauer T, Hoffmann D, Korn K & Selbig J. Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. *Proceedings of the National Academy of Sciences, USA* 2002; **99:**8271–8276.
25. Wang D & Larder B. Enhanced prediction of lopinavir resistance from genotype by use of artificial neural networks. *Journal of Infectious Diseases* 2003; **188:**653–660.
26. Sevin AD, DeGruttola V, Nijhuis M, Schapiro JM, Foulkes AS, Para MF & Boucher CAB. Methods for investigation of the relationship between drug-susceptibility phenotype and human immunodeficiency virus type 1 genotype with applications to AIDS clinical trials group 333. *Journal of Infectious Diseases* 2000; **182:**59–67.
27. Idemyor V. The evolving role of antiretroviral drug resistance testing in HIV-infected individuals. *HIV Clinical Trials* 2002; **3:**413–417.
28. Parkin N, Chappey C, Maroldo L, Bates M, Hellmann NS & Petropoulos CJ. Phenotypic and genotypic HIV-1 drug resistance assays provide complementary information. *Journal of Acquired Immune Deficiency Syndromes* 2002; **31:**128–136.
29. Parkin NT, Chappey C & Petropoulos CJ. Improving lopinavir genotype algorithm through phenotype correlations: novel mutation patterns and amprenavir cross-resistance. *AIDS* 2003; **17:**955–961.
30. Scudeller L, Torti C, Quiros-Roldan E, Patroni A, Lo Caputo S, Moretti F, Mazzotta F, Donati E, Vivarelli A & Carosi G. HIV susceptibility to amprenavir: phenotype-based versus rules-based interpretations. *Journal of Antimicrobial Chemotherapy* 2003; **52:**776–781.
31. Vandamme AM, Houyez F, Banhegyi D, Clotet B, De Schrijver G, De Smet KA, Hall WW, Harrigan R, Hellmann N, Hertogs K, Holtzer C, Larder B, Pillay D, Race E, Schmit JC, Schuurman R, Schulse E, Sonnerborg A & Miller V. Laboratory guidelines for the practical use of HIV drug resistance tests in patient follow-up. *Antiviral Therapy* 2001; **6:**21–39.

Received 14 February 2004, accepted 6 April 2004