JAC antiviral

Journal of Antimicrobial Chemotherapy (2004) **53**, 356–360
DOI: 10.1093/jac/dkh021
Advance Access publication 19 December 2003

Comparative evaluation of three computerized algorithms for prediction of antiretroviral susceptibility from HIV type 1 genotype

Maurizio Zazzi[1]*, Laura Romano[1], Giulietta Venturi[1], Robert W. Shafer[2], Caroline Reid[3], Federico Dal Bello[4], Cristina Parolin[4], Giorgio Palù[4] and Pier E. Valensin[1]

[1]Department of Molecular Biology, University of Siena, Siena, Italy; [2]Division of Infectious Diseases, Stanford University, Stanford, CA; [3]Bayer HealthCare, Diagnostics Division, Toronto, Canada; [4]Department of Histology, Microbiology and Medical Biotechnologies,University of Padova, Padova, Italy

Received 29 April 2003; returned 24 June 2003; revised 23 July 2003; accepted 13 October 2003

***Objectives*: To compare three methods for using HIV-1 genotype to predict antiretroviral drug susceptibility.**

***Methods*: We applied three genotypic interpretation algorithms to 478 reverse transcriptase (RT) and 410 protease sequences for which phenotypic data were available. Sequences were obtained from clinical practice and from published sequences in the Stanford HIV-1 RT and Protease Sequence Database. The genotypic interpretation algorithms included: Stanford HIVdb program (HIVdb), the Visible Genetics/Bayer Diagnostics Guidelines 6.0 (VGI) and a genotypic interpretation program (AntiRetroScan, ARS) developed at the University of Siena, Italy. Genotypic interpretations were normalized to a three-level output: susceptible, intermediate and resistant. Discordances were defined as differences between genotype and phenotype for the same virus isolate. Discordances for which an isolate was considered susceptible by one test but resistant by another test were considered major discordances.**

***Results*: The frequency of major discordances between genotype and phenotype was 10.6, 13.7 and 15.7% for ARS, VGI and HIVdb, respectively ($P < 0.0001$ for ARS versus HIVdb and for ARS versus VGI; $P = 0.002$ for VGI versus HIVdb). The correlation between genotype and phenotype was highest for non-nucleoside RT inhibitors and lowest for nucleoside RT inhibitors. Half of the major discordances involved stavudine, didanosine and zalcitabine. The concordance among the three genotypic algorithms was high, with weighted Kappa values ranging between 0.76 and 0.84 for the pairwise comparisons between each of the algorithms.**

***Conclusions*: Genotype interpretation algorithms correctly predict phenotype in 85–90% of cases, but the rate of concordance is not uniformly distributed among different drugs. These data provide insight into the potential additional benefit derived from phenotyping.**

Keywords: antiretroviral drug resistance, genotype, phenotype, virtual phenotype

Introduction

Although combination therapy with viral reverse transcriptase (RT) and protease (PR) inhibitors limits HIV type 1 (HIV-1) replication, inability to eliminate infection completely makes antiretroviral treatment a chronic, possibly lifelong therapy. Such prolonged treatment is characterized by toxicity and inadequate adherence, leading to suboptimal drug exposure and development of drug resistance in a majority of patients. Thus, drug resistance testing has become an integral part of the clinical management of infected patients.[1]

In vitro drug susceptibility testing requires complex procedures that are difficult to standardize, and is presently provided by very few specialized companies at high cost (e.g. the Virco N. V. Antivirogram and the ViroLogic Inc. Phenosense assay).[2,3] However, detailed and continuously updated knowledge of the HIV-1 mutations responsible for drug resistance has made it possible to infer the drug susceptibility profile of a given HIV-1 isolate from its *pol* gene sequence.[4] More than 100 amino acid substitutions at over 40 codons have been described that can confer some degree of resistance to one or more of the currently available nucleoside (NRTIs) and non-nucleoside RT inhibitors (NNRTIs), or protease inhibitors (PIs).[5] The number of resistance mutations and the complexity of the interactions between these mutations have stimulated the development of a number of

*Corresponding author. Tel: +39-0577-233850; Fax: +39-0577-233870; E-mail: zazzi@unisi.it

genotype interpretation systems, several of which are now run by Web-based or stand-alone computer programs.[6]

Most genotypic interpretation algorithms use a set of rules to assign a level of susceptibility (e.g. susceptible, possible or intermediate resistance, resistance) to each drug. Each rule is conditioned upon the presence of one or more specific mutations and assigns a level of inferred resistance to one or more different drugs. The Visible Genetics/Bayer Diagnostics genotype interpretation rules, an integral part of the software included in the widely used TRUGENE HIV-1 genotyping test and OpenGene sequencing system, are an example of this approach. An alternative method is being used at Stanford University, where each mutation is given a score or 'penalty' for each drug and the susceptibility level is determined by the sum of the values derived from the mutations found (http://hivdb.stanford.edu/).

We have developed a genotype interpretation system (AntiRetroScan) combining the 'penalty' and the 'rules-based' approaches. The performances of this combined method, the penalty-based Stanford system and the Visible Genetics/Bayer Diagnostics rules-based method were evaluated by comparing predicted phenotypes for 478 RT and 410 PR sequences with phenotypic data.

Materials and methods

Sequences and phenotypic data

Phenotype data from the Stanford HIV RT and Protease Sequence Database (http://hivdb.stanford.edu/) and the HIV Monitoring Service database at the University of Siena (SHMS), Italy, were used. To standardize the analysis, only phenotype results obtained by the Virco Antivirogram and the ViroLogic Phenosense phenotyping systems were selected. All RT and PR sequences had at least one resistance mutation. In all, 433 RT and 365 PR sequences with real phenotype results were obtained from the Stanford database. Forty-five RT and PR sequences with matched Antivirogram results were obtained from SHMS. Each sequence had susceptibility data available for a variable number of drugs depending on the time when the phenotyping had been conducted. Drug-specific cut-off values as indicated in the current version of the Antivirogram (2.5-fold for saquinavir and amprenavir, 3-fold for stavudine, abacavir and indinavir, 3.5-fold for didanosine, zalcitabine and ritonavir, 4-fold for zidovudine, tenofovir and nelfinavir, 4.5-fold for lamivudine, 6-fold for efavirenz, 8-fold for nevirapine, 10-fold for delavirdine and lopinavir) and Phenosense (1.4-fold for tenofovir, 1.7-fold for didanosine, stavudine and zalcitabine, 2.2-fold for zidovudine, 2.5-fold for lamivudine, nevirapine, delavirdine, efavirenz, indinavir, ritonavir, nelfinavir, saquinavir and amprenavir, 4.5-fold for abacavir, 10-fold for lopinavir) assays were used to assign susceptibility or resistance. The total number of drug susceptibility predictions was 6053.

Genotype interpretation systems

The Stanford HIVdb program (HIVdb) scans a sequence for known resistance mutations and calculates a final score for each drug based on the sum of the corresponding penalties retrieved from a reference table in which the resistance mutations have definite values for the different drugs. Depending on this final score, one of five susceptibility levels (susceptible, potential low-level resistance, low-level resistance, intermediate resistance, high-level resistance) is then assigned to each drug.

The software included in the Visible Genetics/Bayer Diagnostics TRUGENE HIV-1 genotyping test and OpenGene sequencing system (VGI) extracts known resistance mutations from the output sequence. The presence of a mutation combination defined in any of the 85 rules triggers an intermediate resistance or resistance call for the drug(s) defined by the rule. The highest degree of resistance (no evidence of resistance, possible resistance or resistance) assigned to each drug is reported as the final susceptibility result.

AntiRetroScan 1.0 (ARS) is a Windows desktop program developed at the University of Siena in April 2001. It was initially derived from the Stanford HIVdb program with the aim of using the same algorithm off-line and automatically storing the output in a pre-existing local Microsoft Visual Foxpro database. The program uses the same penalty approach as HIVdb but mutation–drug pair values have been updated independently from the HIVdb. In addition, a set of 12 rules has been incorporated to adjust the final scores based on predicted interactions among mutations. Eight rules remodel the scores for NRTIs, including proportional increases in zidovudine, stavudine and tenofovir scores depending on the total number of thymidine analogue mutations (TAMs), abacavir score in the presence of TAMs and M184V, abacavir score in the presence of TAMs and V118I, lamivudine score in the presence of TAMs and E44A/D or V118I, and a decrease of the zidovudine, stavudine and tenofovir resensitization by M184V in the presence of insertions at RT codon 69. Four rules remodel the scores to PIs including an increase in amprenavir score in the presence of specific mutation pairs (V32I + I47V, L33F + V82A, V82F + L90M, M46I + V82F), an increase in saquinavir, indinavir, ritonavir, nelfinavir and amprenavir scores in the presence of more than eight total PI resistance mutations, a slight decrease in indinavir score in the presence of less than six total PI resistance mutations and a progressive increase in lopinavir score depending on the number of specific mutations (L10F/I/R/V, K20M, L24I, M46I/L, I47V, G48V, I50V, I54M/V, A71T/V, V82A/F/T/S, I84V, L90M). Similar to HIVdb, the system defines five drug susceptibility levels, although the result is formally expressed as probability of resistance (none, negligible, modest, substantial, high), rather than as resistance levels.

Data analysis

In order to allow inter-algorithm comparisons, the five-category output yielded by HIVdb and ARS was collapsed to the three-level output yielded by VGI. To achieve this, the bottom and top two categories were scored as susceptibility and resistance, respectively, while the middle category was considered as intermediate. Pairwise concordance among the genotypic systems was analysed using a weighted Kappa statistic. Major discordances between genotype and phenotype were recorded when an isolate was considered susceptible by one test but resistant by another test. Minor discordances were considered for isolates ranked as intermediate by genotype and susceptible or resistant by phenotype. The difference in the frequencies of discordances with phenotype between each genotypic interpretation system was computed by χ^2 analysis.

Results

The median percentage of isolates that were resistant by phenotype was 31% (range 17% for didanosine to 70% for nelfinavir). The percentage distribution of predicted phenotype categories for these *in vitro* susceptibility data is shown in Figure 1. For each of the three genotypic interpretation systems, the highest rate of major and minor discordance occurred for the NRTIs and the lowest for the NNRTIs. The frequency of major discordances between genotype and phenotype was 10.6, 13.7 and 15.7% for ARS, VGI and HIVdb, respectively ($P < 0.0001$ for ARS versus HIVdb and for ARS versus VGI; $P = 0.002$ for VGI versus HIVdb). The frequency of total (major plus minor) discordances was also significantly lower for ARS (24.7%) with respect to VGI (30.1%) and HIVdb (28%) ($P < 0.0001$ for both comparisons), but HIVdb outperformed VGI in this analysis ($P = 0.012$).

Didanosine, zalcitabine and stavudine generated 40–45% of the total major discordances (Figure 2). In nearly all of these cases, the

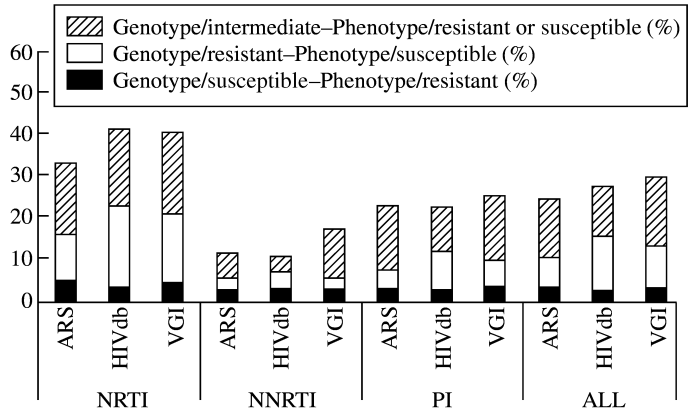

Figure 1. Prevalence of major (genotype susceptible/phenotype resistant; genotype resistant/phenotype susceptible) and minor (genotype intermediate/phenotype susceptible or resistant) discordance between each genotype interpretation system (ARS, AntiRetroScan; HIVdb, Stanford HIVdb; VGI, Visible Genetics) and phenotype.

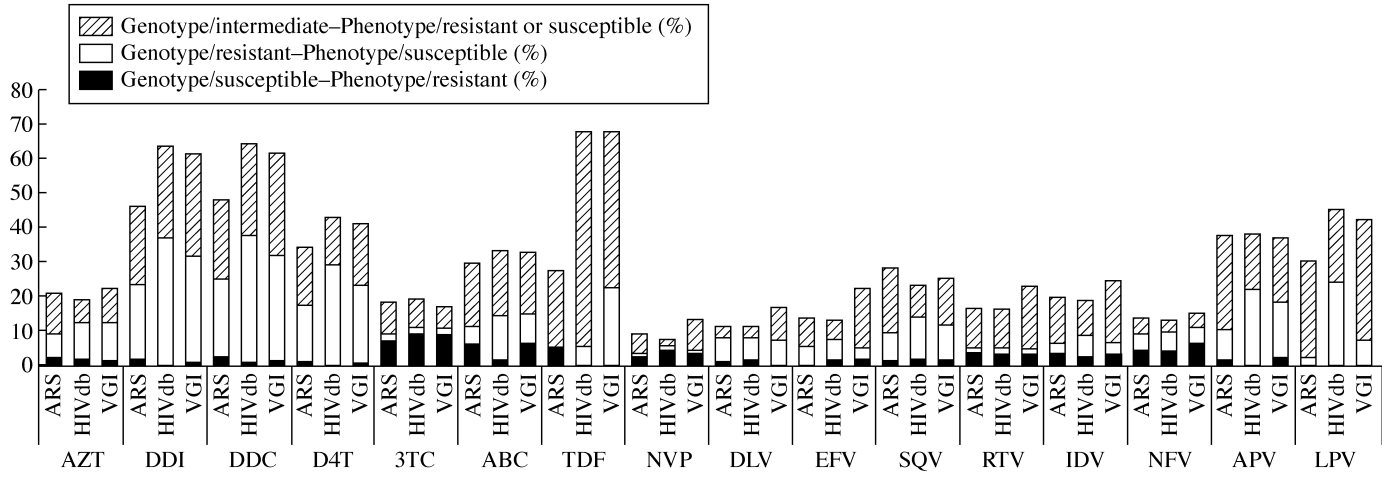

Figure 2. Distribution of major and minor genotype–phenotype discordance per drug for the three genotype interpretation systems (ARS, AntiRetroScan; HIVdb, Stanford HIVdb; VGI, Visible Genetics) with respect to phenotype data. AZT, zidovudine; DDI, didanosine; DDC, zalcitabine; D4T, stavudine; 3TC, lamivudine; ABC, abacavir; TDF, tenofovir; NVP, nevirapine; DLV, delavirdine; EFV, efavirenz; SQV, saquinavir; RTV, ritonavir; IDV, indinavir; NFV, nelfinavir; APV, amprenavir; LPV, lopinavir.

genotypic interpretation was resistant, whereas the phenotypic interpretation was susceptible. Indeed, when these cases were excluded from the analysis, major discordances for ARS, VGI and HIVdb were 6.4, 7.9 and 8.7%, respectively. In contrast, lamivudine was the only NRTI for which all the systems underestimated resistance. A similar trend was observed with the NNRTI nevirapine and with the PI ritonavir. Phenotypic resistance to the PI saquinavir, amprenavir and lopinavir was more often overestimated than underestimated, especially by HIVdb and VGI.

Pair-wise analysis among the three genotype interpretation systems revealed that <1% of isolates were interpreted as susceptible by one system and resistant by another. The rates of minor discordance between algorithms were 16.4% for ARS compared with HIVdb, 14% for ARS compared with VGI and 12% for HIVdb compared with VGI. The overall weighted Kappa values for comparisons among the three genotypic interpretations ranged from 0.76 to 0.84. However, the rate of agreement was not uniformly distributed among different drugs (Figure 3).

Discussion

We have used a large phenotype dataset to test the ability to predict *in vitro* HIV drug susceptibility from genotype by widely used online (HIVdb) and commercially available genotyping assay-associated (VGI) algorithms, as well as by a newly developed HIV genotype interpretation system (ARS).

Overall, the three genotype interpretation systems exhibited major discordances with phenotyping in 10.6–15.7% (mean 13.3%) of cases. While phenotypic tests currently use a fold-resistance cut-off to yield a dichotomous output (resistance or susceptibility), drug resistance is continuous rather than dichotomous. Therefore, some of the major discordances actually derived from intermediate genotypes matching susceptible or resistant phenotypes that were close to the phenotypic cut-off. For example, when phenotypic results that were within a fold range that was 20% higher or lower than the manufacturers cut-off were considered indicative of intermediate susceptibility, the major discordance rate for ARS, VGI and HIVdb

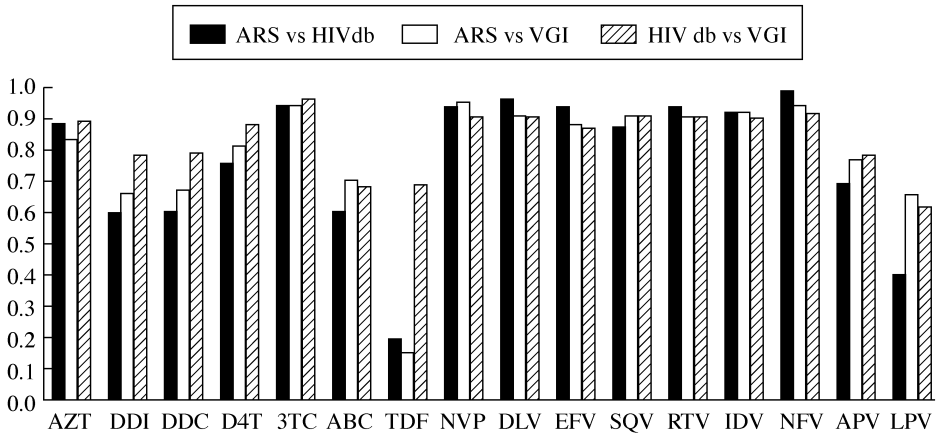

Figure 3. Drug-specific weighted Kappa values for the agreement between each genotype interpretation system pair (ARS, AntiRetroScan; HIVdb, Stanford HIVdb; VGI, Visible Genetics).

decreased to 8, 10.9 and 12.7%. However, it is presently not known what cut-off values can be associated to an intermediate response to different drugs *in vivo*.

Nearly half of major discordances were for stavudine, didanosine and zalcitabine, and derived from a resistant genotype with a susceptible phenotype in the presence of TAMs. For incompletely understood reasons, phenotypic resistance to these drugs is hard to measure by currently available technology and their cut-off values are close to or within assay variability. There is *in vivo* evidence that stavudine and didanosine select for TAMs[7,8] and that virological response to these drugs may decrease in these cases even in the presence of fold-resistance levels that are below the established biological cut-off values.[9,10] Should such evidence be definitely confirmed, then nearly half of the major discordances between genotype and phenotype found in this work could be attributed to phenotype limitations rather than to actual misinterpretation of genotype.

On the other hand, underestimation of real *in vitro* phenotypic resistance to lamivudine by all the three genotypic systems was mostly due to modest increases in fold-resistance levels in the presence of TAMs (data not shown), as also reported in a previous analysis.[11] Although it has been shown that HIV-1 RT with TAMs has decreased susceptibility to lamivudine,[12] a clinical role for this low-level resistance remains uncertain, particularly in light of the very large dynamic susceptibility range typical of this drug. Resistance patterns involving combinations of nucleotide excision mutations (NEMs) and M184V were also associated with a higher rate of inter-algorithm discordance for tenofovir, didanosine, zalcitabine and abacavir, as reported in other comparative studies.[13,14] Among the PIs, lopinavir remains the drug with the lowest inter-algorithm agreement, probably because of the continuous updating of its complex resistance profile in the absence of typical signature mutations.[15]

The analysis described in this study used algorithms that were updated during a similar period (February, August and October 2002 for VGI, HIVdb and ARS, respectively), and all recognize known correlations between genotype and phenotype, genotype and virological outcome, and genotype and treatment history as a common data source for genotype interpretation. Thus, it is not surprising that the three systems showed a high degree of correlation. The analysis presented here does not prove that using penalty- and rules-based approaches together was responsible for the increased reliability of prediction of phenotype by ARS over the other two methods. Most importantly, any genotype interpretation system must be evaluated as a tool to predict virological response to treatment *in vivo*, rather than phenotype *in vitro*. While this approach is being pursued worldwide for many different systems, this work suggests that updated genotype interpretation systems correctly predict drug susceptibility for most resistance profiles. Preliminary results also indicate that the discordance between each of the three systems considered and the VirtualPhenotype[16] is even lower than that found with respect to real phenotype (9.5% versus 13.3%; data not shown). Since the VirtualPhenotype has been shown to be equal to or better than real phenotype at predicting virological response,[17,18] the potential benefits derived from phenotyping with respect to genotyping may be presently limited to a low proportion of cases in clinical practice.

Acknowledgements

This work was supported by Istituto Superiore di Sanità, Ministero della Sanità, Rome (grants 30D.82, 30B.15) and Fondazione Monte dei Paschi di Siena, Siena (grant 'Diagnostica microbiologica diretta mediante tecniche biomolecolari'), Italy.

References

1. Yeni, P. G., Hammer, S. M., Carpenter, C. C. *et al.* (2002). Antiretroviral treatment for adult HIV infection in 2002: updated recommendations of the International AIDS Society—USA Panel. *Journal of the American Medical Association* **288**, 222–35.

2. Hertogs, K., de Bethune, M. P., Miller, V. *et al.* (1998). 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 and Chemotherapy* **42**, 269–76.

3. Petropoulos, C. J., Parkin, N. T., Limoli, K. L. *et al.* (2000). A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrobial Agents and Chemotherapy* **44**, 920–8.

4. Demeter, L. & Haubrich, R. (2001). International perspectives on antiretroviral resistance. Phenotypic and genotypic resistance assays: methodology, reliability, and interpretations. *Journal of Acquired Immune Deficiency Syndromes* **26**, S3–9.

5. Shafer, R. W., Jung, D. R., Betts, B. J. *et al.* (2000). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research* **28**, 346–8.

6. Schmidt, B., Walter, H., Zeitler, N. *et al*. (2002). Genotypic drug resistance interpretation systems – the cutting edge of antiretroviral therapy. *AIDS Reviews* **4**, 148–56.

7. Pellegrin, I., Izopet, J., Reynes, J. *et al*. (1999). Emergence of zidovudine and multidrug-resistance mutations in the HIV-1 reverse transcriptase gene in therapy-naive patients receiving stavudine plus didanosine combination therapy. STADI Group. *AIDS* **13**, 1705–9.

8. Coakley, E. P., Gillis, J. M. & Hammer, S. M. (2000). Phenotypic and genotypic resistance patterns of HIV-1 isolates derived from individuals treated with didanosine and stavudine. *AIDS* **14**, F9–15.

9. Shulman, N. S., Hughes, M. D., Winters, M. A. *et al*. (2002). Subtle decreases in stavudine phenotypic susceptibility predict poor virologic response to stavudine monotherapy in zidovudine-experienced patients. *Journal of Acquired Immune Deficiency Syndromes* **31**, 121–7.

10. Calvez, V., Costagliola, D., Descamps, D. *et al*. (2002) Impact of stavudine phenotype and thymidine analogues mutations on viral response to stavudine plus lamivudine in ALTIS 2 ANRS trial. *Antiviral Therapy* **7**, 211–8.

11. Parkin, N. T., Chappey, C., Maroldo, L. *et al*. (2002). Incidence and nature of phenotype-genotype discordance: maximizing the utility of resistance testing. In *Program and Abstracts of the Ninth Conference on Retroviruses and Opportunistic Infections, Seattle, WA, 2002*. Abstract 580-T, p. 267. Foundation for Retrovirology and Human Health, Alexandria, VA, USA.

12. Duan, C., Poticha, D., Stoeckli, T. C. *et al*. (2001). Inhibition of purified recombinant reverse transcriptase from wild-type and zidovudine-resistant clinical isolates of human immunodeficiency virus type 1 by zidovudine, stavudine, and lamivudine triphosphates. *Journal of Infectious Diseases* **184**, 1336–40.

13. Kijak, G. H., Rubio, A. E., Pampuro, S. E. *et al*. (2003). Discrepant results in the interpretation of HIV-1 drug-resistance genotypic data among widely used algorithms. *HIV Medicine* **4**, 72–8.

14. Puchhammer-Stockl, E., Steininger, C., Geringer, E. *et al*. (2002). Comparison of virtual phenotype and HIV-SEQ program (Stanford) interpretation for predicting drug resistance of HIV strains. *HIV Medicine* **3**, 200–6.

15. Kempf, D. J., Isaacson, J. D., King, M. S. *et al*. (2002). Analysis of the virological response with respect to baseline viral phenotype and genotype in protease inhibitor-experienced HIV-1-infected patients receiving lopinavir/ritonavir therapy. *Antiviral Therapy* **7**, 165–74.

16. Larder, B. A., Kemp, S. D. & Hertogs, K. (2000). Quantitative prediction of HIV-1 phenotypic drug resistance from genotypes: the virtual phenotype (VirtualPhenotype). *Antiviral Therapy* **5**, *Suppl. 3*, 49 (abstract 63).

17. Mazzotta, F., Lo Caputo, S., Torti, C. *et al*. (2003). Real versus virtual phenotype to guide treatment in heavily pretreated patients: 48-week follow-up of the Genotipo-Fenotipo di Resistenza (GenPheRex) trial. *Journal of Acquired Immune Deficiency Syndromes* **32**, 268–80.

18. Perez-Elias, M. J., Garcia-Arata, I., Munoz, V. *et al*. (2002). A randomized, prospective study of Phenotype (P) versus Virtual Phenotype (VirtualP) testing for patients failing antiretroviral therapy (ART). In *Program and Abstracts of the Ninth Conference on Retroviruses and Opportunistic Infections, Seattle, WA, 2002*. Abstract 586, p. 269. Foundation for Retrovirology and Human Health, Alexandria, VA, USA.