Genotype Interpretation Algorithms: The Key to Resistance Testing for HIV-1

Andrew Hwang

Biochemistry 218

Antiretroviral drug therapy has been one of the few promising areas in the fight against HIV, contributing to a significant reduction in mortality and morbidity due to HIV in the developed world. However, successful use of antiretroviral drugs is plagued by the phenomenon of drug resistance. A crucial challenge facing antiretroviral therapy is the development of accurate, convenient methods of resistance testing (6). This paper will review the role of resistance testing in the treatment of HIV. The paper will focus on genotype testing, a particularly promising method of resistance testing that uses computational analysis to determine drug susceptibility from sequence information. The role of phenotype information in improving genotype interpretation will be analyzed by comparing predictions made by genotype interpretation systems from three sources, Agence Nationale de Recherches sur le Sida (ANRS), HIV RT and Protease Sequence Database (HIVdb), and Rega Institute (Rega), to phenotypic data.

Background

Drug resistance

The three classes of antiretroviral drugs, nucleoside analog reverse transcriptase inhibitors (NRTI), non-nucleoside analog reverse transcriptase inhibitors (NNRTI), and protease inhibitors (PI) all target either HIV's reverse transcriptase enzyme or its protease. Drug resistance occurs when mutations in these targets make a drug no longer able to inhibit viral replication. Unfortunately, it has been found that many mutations also confer cross-resistance, where resistance to one drug additionally confers resistance to other drugs within that drug class.

Several factors contribute to the rapid development of drug resistance observed in HIV. The error-prone reverse transcriptase enzyme, which lacks the ability to proofread

during replication, along with the high rate of HIV replication, produce enormous genetic variation. The HIV population in any one person is a "quasi-species," consisting of many related but genetically distinct variants. In an untreated, infected individual, every single possible point mutation is estimated to occur between 10^4 and 10^5 times per day (18). During drug therapy, mutations conferring resistance are selected for, and recombination allows the virus to even more rapidly develop resistance by exchanging drug resistant mutations (13).

Drug resistance is an enormous and growing problem. In the United States and Europe, 10% of new infections are with strains of HIV that already are resistant to at least one class of anti-HIV drugs. This clearly poses major challenges to providing antiretroviral therapy since most effective therapy requires using drugs from at least two of the three drug classes (19).

Resistance Testing

Resistance testing has the potential to be a valuable tool in dealing with drug resistance. Retrospective studies have shown that resistance testing can predict clinical outcomes (3). Several large prospective studies where resistance testing is used as a guide for therapy have also demonstrated that resistance testing can produce better clinical outcomes than standard care procedures (22, 10, 6). While results have varied, it appears that resistance testing can be beneficially integrated into treatment. In fact, guidelines by the International AIDS Society-USA and the EuroGuidelines group recommend considering resistance testing for a variety of situations, including primary HIV-1 infection and treatment failure (18).

However, despite the benefits, resistance testing is a long way from becoming routine procedure in the treatment of HIV. This is true for a variety of factors, including the hazards of handling HIV in the workplace and the availability of alternative treatments if resistance is found. Other important challenges for resistance testing are improving reliability in detecting even tiny populations of drug resistance virus and accurately interpreting test results (7).

The two main types of resistance testing are phenotype testing and genotype testing. Phenotype testing looks at drug resistance *in vitro*, measuring the extent to which specific drugs or drug combinations inhibit viral replication in cultured cells. Genotype testing analyzes the sequence information of the virus, inferring drug resistance from mutations associated with resistance.

Phenotype testing

In this test, a recombinant virus is created with a reference strain of the virus along with the protease and reverse transcriptase genes of a patient's virus pool. This virus is subjected to increasing concentrations of the antiretroviral drug in order to determine the amount of drug required to inhibit replication by 50%, termed the IC_{50}. This IC_{50} value is compared to the value for a wild-type virus culture, and an increase in IC_{50} above a certain level indicates reduced susceptibility (23).

Although phenotype testing provides a direct, quantitative measurement of drug resistance, several practical difficulties have obstructed its adoption. Phenotype assays are very expensive, costing $700 to $900 per sample. In addition, the testing procedure is complex and only offered by a few companies. Furthermore, processing time can take as

long as 8 weeks. Unless these issues can be addressed, there is little chance that phenotype testing can possibly be adopted on a widespread basis (23)

Genotype testing

In genotype testing, the reverse transcriptase and protease samples are sequenced. The genotype assay returns either the sequence data or a list of the differences from the consensus sequence. This data is then analyzed by an interpretation system that infers drug resistance based on the mutations present (18).

Genotype testing has enjoyed much wider usage because the procedure is much simpler, the costs are lower, and the processing time is shorter. Genotype assays typically cost around $400 and take 2 weeks for processing (23). Reliable sequencing is readily available through a number of different methods, including dideoxynucleotide sequencing, clonal sequencing, and population-based sequencing. While the current genotyping kits have been primarily developed for subtype B strains, a recent study found that they were effective in sequencing non-subtype B strains as well (1). A recent cost-benefit analysis supported genotypic testing after treatment failure (5). These factors suggest that genotype testing will continue to be the more common method and, if resistance testing does become widespread in usage, it will probably be through genotype testing.

Genotype Interpretation Systems

The main challenge facing genotype testing is not obtaining the sequence data, but how to interpret it. Interpreting a given sequence requires analyzing over 100 amino acids across 40 codons that have been observed to confer some degree of resistance. These mutations can interact in a variety of different ways. For instance, a mutation that

confers resistance to one drug can suppress resistance in another. Cross-resistance, where a mutation selected for by one drug also confers resistance to other drugs in its class, further complicates interpretation. Systems must take into account all these complex interactions between different combinations of mutations. Then, the system must make a qualitative judgment of susceptibility based on a huge amount of clinical and laboratory data. As new results become available and new drugs are approved, the system must be updated to incorporate the new information (14).

Currently, there are over 20 interpretation systems available. Several studies have been performed analyzing the concordance of some of the most popular interpretation systems and found significant discrepancies. One study comparing three widely used algorithms, TruGene, Stanford HIV RT and Protease Sequence Database (HIVdb), and VirtualPhenotype found concordance in only 13.7% of samples (8). The largest such study found a better concordance of 66.4% when comparing 30,000 interpretations by ANRS, Rega, and Visible Genetics (VGI-6) (14). Nevertheless, it is clear that genotype interpretations vary widely and that there is no clear consensus on which interpretation is best.

These comparative studies have highlighted certain trouble spots for interpretation algorithms. Algorithms tend to disagree most often when interpreting resistance to nucleoside reverse transcriptase inhibitors (NRTIs). The drugs that were most commonly cited as having discordant interpretations were abacavir, didanosine, stavudine, amprenavir, and lopinavir. (8, 9, 17, 24). .

Genotype vs. Phenotype testing

Genotype and phenotype tests have been observed to produce discordant results. There are several possible explanations for this. If the resistance is caused by a rare mutation or rare combination of mutations, the genotype interpretation system may be unable to identify the sequence as resistant. In addition, if there is insufficient existing literature describing a drug, the interpretation system may not be informed of all the causes of resistance (11). However, in cases where there exist tiny populations of drug resistant viruses, genotype testing tends to be more sensitive than phenotype testing. Phenotype tests may be unable to observe the effects of this minor population, whereas a genotype test could recognize the drug resistant mutations from the sequence data. Another major factor contributing to discordant results is the complex interaction between mutations. For instance, one mutation may have a re-sensitization effect, conferring resistance to one drug but re-sensitizing the virus to another drug. In phenotype tests, the "re-sensitized" mutation would be masked (14).

Which test, then produces the more clinically relevant results? The fact that phenotype testing provides a direct test while genotype testing is based on inference suggests that phenotype testing may be better. However, phenotype tests have performed poorly in studies where it has been used as a guide. In contrast, clinical outcomes have generally been better in studies where genotype testing was used (18). In addition, two large trials that directly compared the two testing types obtained better clinical outcomes for genotype testing (10, 22).

Though genotype testing appears to be the more practical as well as the more effective of the two, the importance of phenotype testing should not be discounted. Many improvements still remain to be made in genotype interpretation systems, and phenotype

results play a crucial role in that process. As discussed above, phenotype tests can produce valuable, complementary information to genotype interpretations (11). An excellent example of this is a study by Parkin et al where phenotype information was used to improve predictions for the drug lopinavir by GeneSeq, a proprietary algorithm. Genotype and phenotype results were compared for concordance. Mutations that were over-represented in discordances were identified. These mutations were analyzed to determine mutations and variants that had not been included in the original algorithm (12).

I chose to compare phenotype results with genotype interpretations from three widely-used, non-proprietary algorithms on a large set of data. I was interested in describing the differences between genotype and phenotype results and discussing whether greater incorporation of phenotype results may be able to similarly improve these three algorithms.

Comparison of Genotype and Phenotype Tests

Material and Methods

Sequences and Phenotypic data

A total of 8079 sequences with available phenotypic data were analyzed. 4365 of these were RT sequences and 3714 were protease sequences. Sequences were taken from the Stanford HIV RT Protease Sequence Database.

Interpretation Systems

Sequence information was interpreted by three algorithms, HIVdb (20), ANRS (10), and Rega (24). All three algorithms are designed for clinical use and make interpretations with rule-based algorithms, where each rule is conditioned upon the

presence of certain mutations and assigns a level of inferred resistance to certain drugs. All three derive their rules from literature and clinical data.

ANRS and Rega output reports three levels of resistance, susceptible (S), intermediate (I), and resistant (R). HIVdb reports five levels, susceptible, potential low-level resistance, low-level resistance, intermediate resistance, and high-level resistance. For the purposes of comparison, susceptible and potential low-level resistance was considered susceptible (S), low-level resistance and intermediate resistance was considered intermediate (I), and high-level resistance was considered resistant (R).

Analysis

The sequences were analyzed through HIValg: Resistance Algorithm Comparison on the Stanford HIV RT Sequence Database (http://hivdb.stanford.edu). HIValg contains implementations of each of the interpretation systems using the Algorithm Specific Interface (2).

Results were compiled and analyzed in two Excel spreadsheets, one for protease data and the other for reverse transcriptase data. The number and percent of concordances and discordances was calculated. In addition, the number and percent of major discordances (S/R, R/S) and minor discordances (S/I, R/I, I/S, I/R) was also calculated. These values were also found for each drug and each drug class. This analysis was then repeated for each algorithm individually.

Numbers for each combination of interpretations (S/S, S/I, S/R, I/S, I/I, I/R, R/S, R/I, R/R) were calculated to find the weighted Kappa statistic (4). This analysis was performed to find the Kappa statistic for each algorithm individually as well as for only the protease data and only the reverse transcriptase data.

Results

Agreement of interpretations with phenotype (Table 1)

The Rega algorithm was found to have the best agreement among all sequences as well as the best agreement among reverse transcriptase sequences, with a weighted kappa value of .620 and .712, respectively. However, HIVdb was found to have the best agreement among just protease sequences, with a weighted kappa value of .564

Table 1

	ANRS	HIVdb	Rega
PR	.530	.564	**.**534
RT	.598	.620	.712
All	.562	.592	.620

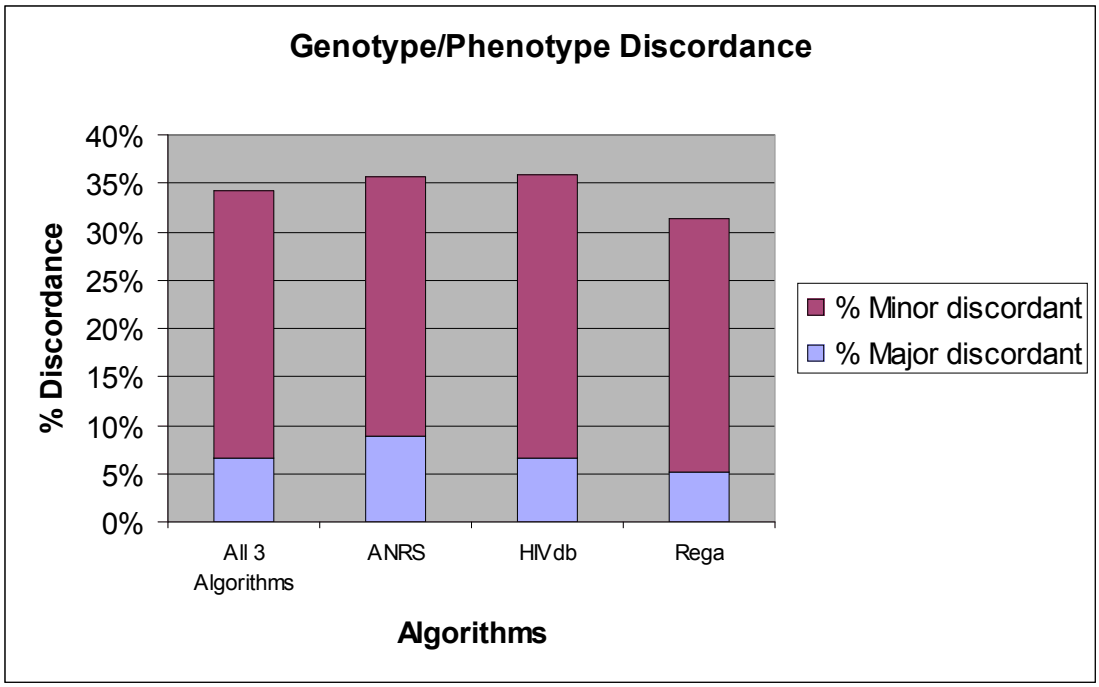
Discordance between genotype and phenotype (Figure 1)

Analysis of the data revealed significant discordance between genotype interpretation and phenotype overall. Of the interpretations made by the three algorithms, 34.22% of them were discordant with phenotypic data. However, further inspection reveals that most of these are minor discordances. 27.57% of the interpretations are minor discordances while only 6.65% are major discordances.

Looking at the algorithms individually, Rega interpretations were found to have the lowest prevalence of discordance overall, with a discordance of 31.39%. Rega also produced the lowest individual rates for both major discordance and minor discordance,

with values of 5.10% and 26.29%, respectively. ANRS interpretations had the second lowest prevalence of discordance overall, with a discordance of 8.78%. Rates for major and minor discordance were 8.78% and 35.07%, respectively. HIVdb interpretations had a discordance of 35.88%, the worst overall rate of discordance. However, the bulk of this consists of minor discordances. While its rate of minor discordances is 29.38%, its rate of major discordances is only 6.5%, significantly lower than the rate for ANRS.

Figure 1

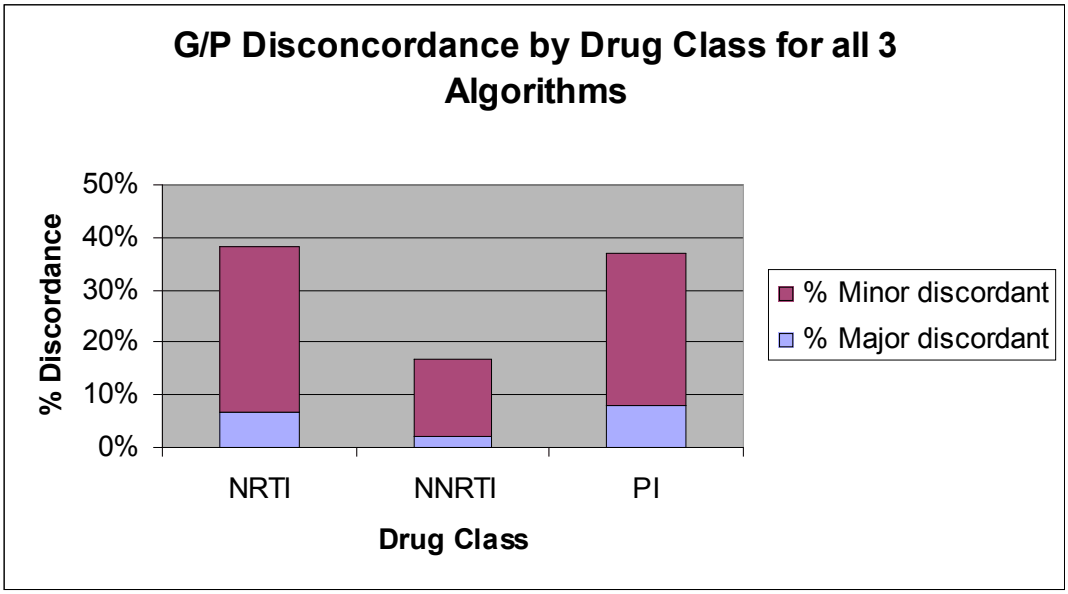


Discordance between genotype and phenotype by drug class (Figure 2, Figure 3)

Analysis of discordances by drug class revealed that interpretations for NNRTI drugs were considerably less discordant than interpretations for either NRTI or PI. The major discordance rate was 2.23% and the minor rate was 14.47%, adding up to a total discordance of 16.7%. All of these values are less than half the corresponding values for either NRTI's or PI's.

The next lowest rate of discordance was for PI interpretations, with a rate of 36.89% compared to 38.03% for NNRTI interpretations. However, although overall rate of discordance was lower, PI interpretations had a higher rate than NRTI of major discordance, with a rate of 8.08% compared to 6.63%.

Figure 2



A closer look at the discordance by drug class for each algorithm individually revealed some interesting trends (Figure 3). The algorithms generally had similar discordance rates for PI drugs, but discordance rates varied widely for NRTI drugs.

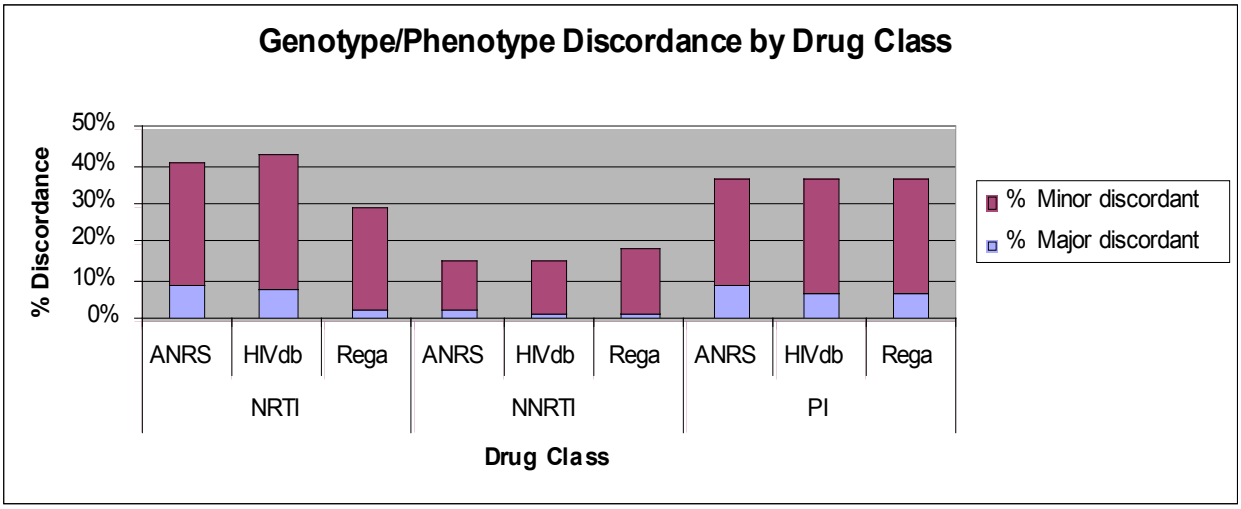
Although the Rega algorithm had the lowest rate of discordance overall, its performance varied significantly by drug class. For NNRTI drugs, where all algorithms had lower rates of discordance, the Rega algorithm actually had higher discordance than the other two algorithms. Rega interpretations had 1.89% major discordance and 17.04% minor discordance for an overall rate of 18.93%. However, for NRTI drugs, where the three algorithms in general had higher discordance, Rega interpretations had much lower rates of discordance. Its major rate of 3.28%, minor rate of 26.02%, and overall rate of

29.30% are all significantly lower than the rates of the other two algorithms. In general, the Rega algorithm was found to have a more level rate of discordance across drug classes.

In contrast, the HIVdb algorithm varied widely in its rate of discordance by drug class. For NNRTI drugs, where all algorithms had lower rates of discordance, HIVdb interpretations had an even lower rate. Major and minor discordance rates were 1.79% and 13.52%, respectively. However, for NRTI drugs, where the algorithms generally had higher discordance, HIVdb performed particularly poorly. The rate of major discordance and minor discordance was 7.73% and 36.03%, respectively, for a very high total rate of 43.76%.

Analysis also revealed that while the ANRS algorithm varied in its overall rate of discordance, it had the highest rate of major discordance across all drug classes. ANRS interpretations had major discordance rates of 3.41% for NNRTI drugs, 9.17% for NRTI drugs, and 9.82% for PI drugs.

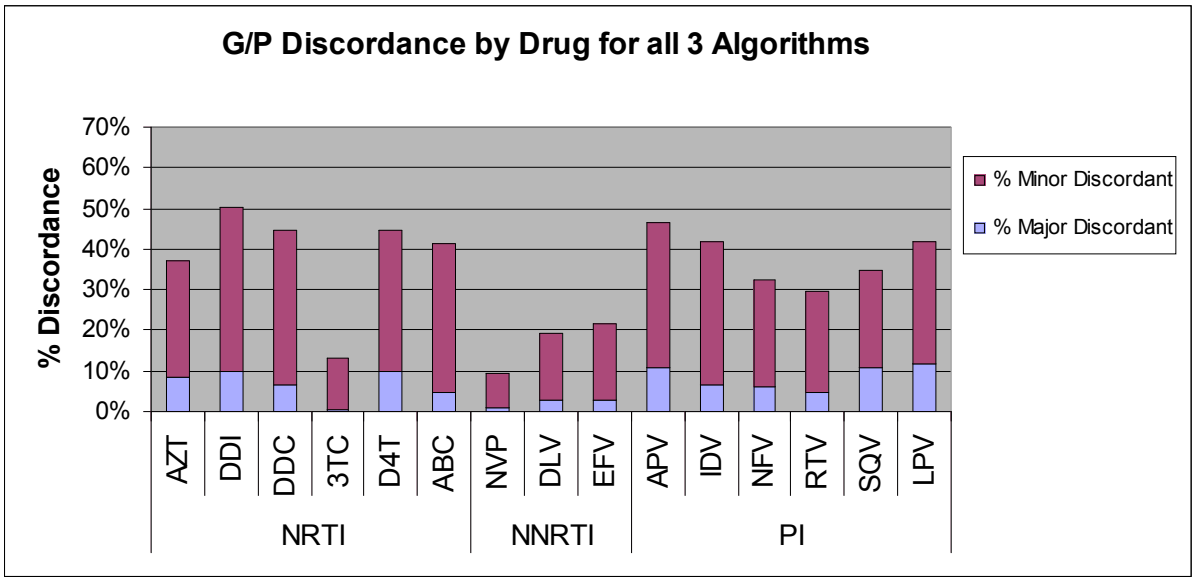
Figure 3



Discordance between Genotype and Phenotype by drug (Figure 4, Figure 5)

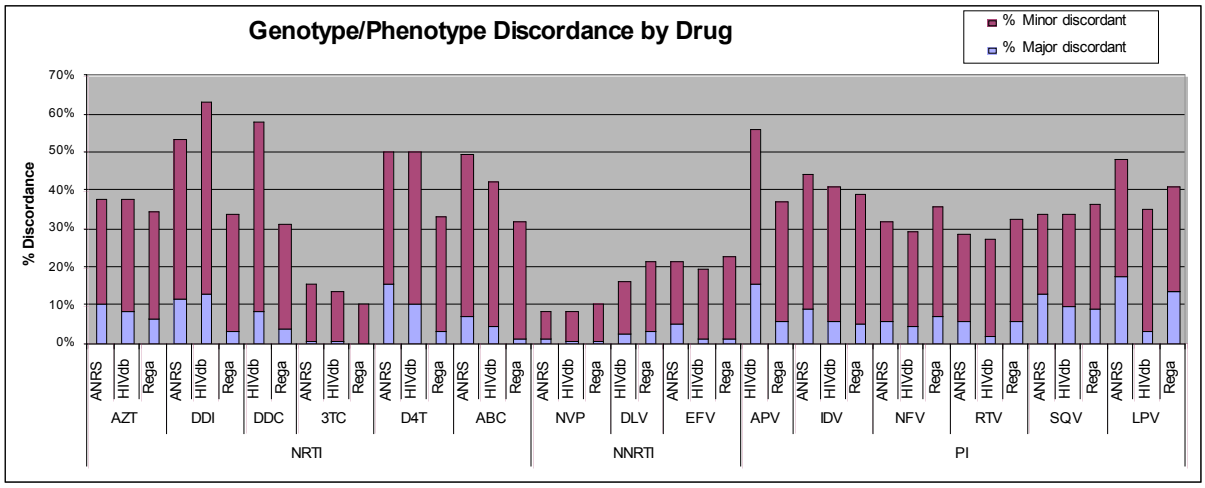
Inspection of genotype/phenotype correlation by drug is especially revealing, as the concordance varies by drug and interpretations for certain drugs stand out as having particularly poor concordance. The highest rate of discordance was seen for didsanosine (DDI) at 50.4%, zalcitabine (DDC) at 44.71%, stavudine (D4T) at 44.53%, abacavir (ABC) at 41.18%, amprenavir (APV) at 46.71%, indinavir (IDV) at 41.6%, and lopinavir (LPV) at 41.59%.

Figure 4



Analyzing the discordance by drug for each algorithm individually provides more concrete information on how well each algorithm performs for each drug. One particularly noticeable characteristic is that for HIVdb, the highest rates of discordance are found in a few specific drugs. HIVdb had much higher discordance than the other two algorithms for didanosine (DDI), zalcitabine (DDC), and amprenavir (APV).

Figure 5



Discussion

Although there was significant discordance between genotype/phenotype results throughout, this analysis highlighted several areas where discordance was significantly more prevalent. Interestingly, these corresponded well with areas where studies comparing genotype/genotype results between interpretation systems also found high discordance. For instance, similar to studies comparing interpretations between algorithms, this analysis found the highest rate of discordance for NRTI inhibitors. This analysis was also similar to those studies in finding high rates of discordance for abacavir, didanosine, stavudine, amprenavir, and lopinavir. This suggests that these may represent areas where genotype tests have large room for improvement and where phenotype testing may be a valuable source of information.

The analysis also indicated several areas that might warrant re-evaluation for each algorithm. In particular, the HIVdb algorithm stands out in that its interpretations for didanosine and amprenavir contribute disproportionately to its rate of discordance. This

suggests that it may be beneficial to re-inspect the rules for these two drugs to make sure they are updated and take into account relevant phenotypic information.

Various factors may contribute to the discordances observed. NRTI drugs may have especially high rates of discordance between genotype and phenotype results as well as between genotype algorithms because this drug class has demonstrated the most cross-resistance. Genotype algorithms may therefore have trouble interpreting the various mutation interactions (11). Abacavir and lopinavir may have high rates of discordance because these are the two drugs with the least data available, so algorithms may not take into account all mutations or combinations of mutations (22).

This analysis suggests that phenotype tests provide a great deal of information that could be used to improve all three of these algorithms. Further analysis is recommended examining specific mutation patterns contributing to the discordance. These mutation patterns may shed light on how these algorithms can be improved to better interpret sequence data.

References

1. Beddows S, Galpin S, Kazmi S, et al. "Performance of Two Commercially Available Sequence-Based HIV-1 Genotyping Systems for the Detection of Drug Resistance Against HIV Type 1 Group M Subtypes." *Journal of Medical Virology*. 70:337-342, 2003

2. Betts B and R Shafer. "Algorithm Specification Interface for Human Immunodeficiency Virus Type 1 Genotypic Interpretation." *Journal of Clinical Microbiology*, 2003;41,6:2792-2794.

3. Brun-Vézinet F, Descampes D, Ruffault A, et al. "Clinically relevant interpretation of genotype for resistance to abacavir. *AIDS*. 2003, 17:1795-1802.

4. C. C. Berry. "The kappa statistic." Journal of the American Medical Association, 268(18):2513, 1992

5. Corzillius M, Muhlberger N, Srozynski G, et al. "Cost effectiveness analysis of routine use of genotypic antiretroviral resistance testing after failure of antiretroviral treatment for HIV." *Antiviral Therapy* 2004; 9:27-36.

6. Hammer S, Pedneault L. "Antiretroviral resistance testing comes of age." *Antiviral Therapy*. 2000;5:1.

7. Kijak GH, Rubio AE, Pampuro SE, et al. Disrepant results in the interpretation of HIV-1 drug-resistance genotypic data among widely used algorithms. *HIV Med.* 2003;4:72-78.

8. Kijak GH, Rubio AE, Pampuro SE et al. "Discrepant results in the interpretation of HIV-1 drug resistance genotypic data among widely used algorithms." *HIV Medicine*. (2003) 4,1:72.

9. Luca A, Perno CF. "Impact of different HIV resistance interpretations by distinct systems on clinical utility of resistance testing." *Current Opinion in Infectious Diseases*. 2003,16:573-580.

10. Meynard, JL, Muriel V, Morand-Joubert L. "Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial." *AIDS* 2002, 16:727-736.

11. Parkin N, Chappey C, Maroldo L, et al. "Phenotypic and Genotypic HIV-1 Drug Resistance Assays Provide Complementary Information." 2002;128-136.

12. Parkin N, Chappey C, Petropoulos CJ. "Improving lopinavir genotype algorithm through phenotype correlations: novel mutation patterns and amprenavir cross-resistance." *AIDS*. 2003, 17:955-961.

13. Rambaut A, Posada D, Crandall K, Holmes E. "The Causes and Consequences of HIV Evolution." *Nature*. 2004;5:52-60.

14. Ravela J, Bradley B, Brun-Vézinet F, et al. "HIV-1 Protease and Reverse Transcriptase Mutation Patterns Responsible for Discordances Between Genotypic Drug Resistance Interpretation Algorithms." *Journal of Acquired Immune Deficiency Syndromes*. 33:8-14. 2003.

15. Reid C, Bassett R, Day S, et al. A dynamic rules-based interpretation system derived by an expert panel is predictive of virological failure. *Antiviral Ther.* 2001;6:21-39.

16. Schmidt B, Walter H, Schwingel E, et al. Comparison of different interpretation systems for genotypic HIV-1 drug resistance data. *Antiviral Ther.* 2001;6:102.

17. Schmidt B, Walter H, Zeitler N, Korn K. "Genotypic Drug Resistance Interpretation Systems – The Cutting Edge of Antiretroviral Therapy." *AIDS Rev*. 2002; 4:148-156.

18. Shafer, RW. "Genotypic Testing for Human Immunodeficiency Virus Type 1 Drug Resistance." *Clinical Microbiology Reviews*. 2002;15,2:247-277..

19. Shafer RW. "Stanford HIV Drug Resistance Database: Primer." http:/hivdb.stanford.edu/pages/documentPage/rationale.html.

20. Shafer RW, Stevenson D, Chan B. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* 1999;27:348-352.

21. Rhee S, Gonzales M, Kantor R. "Human immunodeficiency virus reverse transcriptase and protease sequence database." *Nucleic Acids Research*. 2003:31,1:298-303

22. Torti C, Quiros-Roldan E, Wilco K, et al. "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." *Journal of Infectious Diseases*. 2003;188:194-201.

23. Youree B, D'Aquila R. "Antiretroviral Resistance Testing for Clinical Management." *AIDS Rev.* 2002;4:3-12.

24. Van Laethem K, De Luca A, Antinori A, et al. A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1-infected patients. *Antiviral Ther.* 2002;7:123-129.

25. Zazzi M, Romano L, Venturi G, et al. "Comparative evaluation of three computerized algorithms for prediction of antiretroviral susceptibility from HIV type 1 genotype." *Journal of Antimicrobial Chemotherapy*. (2004)53, 356-360.