Antimicrobial Agents and Chemotherapy

Standardized Comparison of the Relative Impacts of HIV-1 Reverse Transcriptase (RT) Mutations on Nucleoside RT Inhibitor Susceptibility

George L. Melikian, Soo-Yon Rhee, Jonathan Taylor, W. Jeffrey Fessel, David Kaufman, William Towner, Paolo V. Troia-Cancio, Andrew Zolopa, Gregory K. Robbins, Ron Kagan, Dennis Israelski and Robert W. Shafer

Antimicrob. Agents Chemother. **2012, 56(5):2305. DOI: 10.1128/AAC.05487-11.**
Published Ahead of Print 13 February 2012.

AAC Journals.ASM.org

Standardized Comparison of the Relative Impacts of HIV-1 Reverse Transcriptase (RT) Mutations on Nucleoside RT Inhibitor Susceptibility

George L. Melikian,[a] Soo-Yon Rhee,[a] Jonathan Taylor,[b] W. Jeffrey Fessel,[c] David Kaufman,[d] William Towner,[e] Paolo V. Troia-Cancio,[f] Andrew Zolopa,[a] Gregory K. Robbins,[g] Ron Kagan,[h] Dennis Israelski,[a] and Robert W. Shafer[a]

Division of Infectious Diseases, Department of Medicine, Stanford University, Stanford, California, USA[a]; Department of Statistics, Stanford University, Stanford, California, USA[b]; Kaiser Permanente Medical Care Program—Northern California, San Francisco, California, USA[c]; Mount Sinai Medical Center, New York, New York, USA[d]; Department of Infectious Diseases, Kaiser Permanente Los Angeles, Los Angeles, California, USA[e]; Division of Infectious and Immunologic Diseases, UC Davis Medical Center, Davis, California, USA[f]; Division of Infectious Diseases, Harvard University, Boston, Massachusetts, USA[g]; and Quest Diagnostics Incorporated, San Juan Capistrano, California, USA[h]

Determining the phenotypic impacts of reverse transcriptase (RT) mutations on individual nucleoside RT inhibitors (NRTIs) has remained a statistical challenge because clinical NRTI-resistant HIV-1 isolates usually contain multiple mutations, often in complex patterns, complicating the task of determining the relative contribution of each mutation to HIV drug resistance. Furthermore, the NRTIs have highly variable dynamic susceptibility ranges, making it difficult to determine the relative effect of an RT mutation on susceptibility to different NRTIs. In this study, we analyzed 1,273 genotyped HIV-1 isolates for which phenotypic results were obtained using the PhenoSense assay (Monogram, South San Francisco, CA). We used a parsimonious feature selection algorithm, LASSO, to assess the possible contributions of 177 mutations that occurred in 10 or more isolates in our data set. We then used least-squares regression to quantify the impact of each LASSO-selected mutation on each NRTI. Our study provides a comprehensive view of the most common NRTI resistance mutations. Because our results were standardized, the study provides the first analysis that quantifies the relative phenotypic effects of NRTI resistance mutations on each of the NRTIs. In addition, the study contains new findings on the relative impacts of thymidine analog mutations (TAMs) on susceptibility to abacavir and tenofovir; the impacts of several known but incompletely characterized mutations, including E40F, V75T, Y115F, and K219R; and a tentative role in reduced NRTI susceptibility for K64H, a novel NRTI resistance mutation.

Nucleoside/nucleotide reverse transcriptase (RT) inhibitors (NRTIs) are the backbone of antiretroviral (ARV) therapy. Each of the initial treatment regimens recommended by the Department of Health and Human Services (34) and the World Health Organization (38) include two complementary NRTIs and an ARV belonging to a second drug class.

In a previous study, we applied several data-mining approaches to quantify associations between NRTI-associated HIV-1 drug resistance mutations and *in vitro* susceptibility data (24). About 630 susceptibility test results were available for abacavir (ABC), didanosine (ddI), lamivudine (3TC), stavudine (d4T), and zidovudine (AZT), and 350 were available for tenofovir (TDF). In that study, we used a predefined list of nonpolymorphic NRTI-selected mutations to reduce the number of independent variables influencing NRTI susceptibility. Here we analyze a data set that is about twice as large and uses two regression methods in tandem: one to identify genotypic predictors of NRTI susceptibility from the many RT mutations present in the data set (rather than relying on a predefined list of mutations, as we did previously) and one to quantify the impact of RT mutations on NRTI susceptibility. In addition, we used several approaches to determine whether models that included statistical interactions among NRTI resistance mutations improved the prediction of reductions in NRTI susceptibility.

MATERIALS AND METHODS

HIV-1 isolates. We analyzed HIV-1 isolates in the HIV Drug Resistance Database (HIVDB) (22) for which *in vitro* NRTI susceptibility testing had been performed by the PhenoSense (Monogram, South San Francisco, CA) assay (20). About 35% of the test results were from studies published previously by other laboratories; 65% were from studies by our research group or from data recently contributed by one of several collaborating clinics. About 425 genotype-phenotype correlations have not appeared in the published literature previously (for a copy of the data set, see the supplemental material). The Stanford University Human Subjects Committee approved this study.

Drug susceptibility results were expressed as the fold change in susceptibility, defined as the ratio of the 50% effective concentration (EC_{50}) for a tested isolate to that for a standard wild-type control isolate. EC_{50} results for 3TC and emtricitabine (FTC) with a fold change in susceptibility of >200 were censored (i.e., reported as >200) by the PhenoSense assay. In such cases, we assigned a fold change of 200 for these two NRTIs, as well as for AZT, for samples which had fold change results of >200.

The subtype of each isolate either was determined by using the REGA subtyping algorithm (5) and the NCBI viral genotyping resource (26) or was identified directly from the phenotype report. Mutations were de-

Received 9 August 2011 Returned for modification 16 September 2011
Accepted 3 February 2012
Published ahead of print 13 February 2012
Address correspondence to George L. Melikian, gmelikia@stanford.edu.
Supplemental material for this article may be found at http://aac.asm.org/.

doi:10.1128/AAC.05487-11
The authors have paid a fee to allow immediate free access to this article.

fined as differences from the consensus subtype B amino acid RT sequence (available at http://hivdb.stanford.edu/pages/documentPage/consensus_amino_acid_sequences.html). Nonpolymorphic mutations were defined as mutations that occur at a prevalence of ≤0.5% in the absence of ARV selective pressure (1).

To minimize bias, we excluded susceptibility results obtained when more than one virus from the same individual contained the same mutations at the following influential NRTI resistance positions: 65, 74, 115, 151, 184, and 215. Because the presence of mixtures may confound genotype-phenotype correlations, we also excluded viruses with sequences containing electrophoretic mixtures at these positions.

Identification of mutations associated with decreased NRTI susceptibility. To identify mutations that decrease susceptibility to one or more NRTIs, we used the LASSO (least absolute shrinkage and selection operator) procedure to examine all mutations occurring in 10 or more virus samples. LASSO is particularly useful for selecting a subset of predictors when the set of possible predictors is large (7). LASSO constructs a model by fitting a least-squares solution with the added constraint that $\Sigma\ |\beta_i|_1$ (the L^1 norm of the parameter vector) be $\leq s$, where s is a regularization parameter determined by cross-validation. During cross-validation, LASSO used four-fifths of the data for selecting a model and one-fifth for validating the selected model using the GLMNET package in R (21). Each 5-fold cross-validation was repeated 10 times to estimate the variance in the predicted model. The validation set (one-fifth of entire data) was used to decide when to stop adding variables to the model. The regularization parameter—the LASSO penalty used to identify the optimal number of explanatory features—was chosen as the smallest parameter whose mean cross-validation error was less than or equal to the minimum cross-validation error plus 1 standard deviation of the cross-validation error at the minimum. The dependent variable was the $\log_{10}$ fold change in HIV susceptibility. Each of the regression coefficients represented an HIV-1 amino acid mutation. LASSO coefficient means that were more than 3 standard deviations above or below zero after 10 repeated runs of 5-fold cross-validation were considered statistically significant predictors of susceptibility to NRTIs.

To quantify the effect of the LASSO-selected mutations on NRTI susceptibility, we used least-squares regression (LSR). For this regression analysis, we also used 5-fold cross validation and 10-fold repetition to estimate the variance among the fitted coefficients. Seven LSR models—one for each NRTI—were created. In these models, each of the selected mutations was an explanatory variable and the log of the fold change in susceptibility was the response variable. For each 5-fold cross-validation, 80% of data was used for learning regression coefficients and 20% was used for testing. LSR coefficients (each corresponding to an HIV-1 amino acid mutation) that were more than 3 standard deviations above or below zero in the 10 repeated runs of 5-fold cross-validation were considered statistically significant predictors of susceptibility to NRTIs.

Regression analyses (for both the LASSO and LSR models) were standardized by scaling the log fold distributions for each of seven NRTIs to a distribution with a standard deviation of 1. Standardizing the regression coefficients made it possible to compare the magnitude of a coefficient for an RT mutation across NRTIs despite the highly variable dynamic susceptibility ranges among the NRTIs. Consequently, the regression coefficients reflect the standard deviation change in log fold associated with each specific mutation (rather than the actual log fold difference).

Contribution of NRTI mutations to decreased susceptibility. Prediction accuracy was evaluated using continuous and categorical approaches. The continuous approach involved calculating the mean squared error (MSE) between the actual and predicted standardized log fold change in susceptibility. The categorical approach involved determining how often the predicted phenotype correlated with one of three predefined susceptibility categories: susceptibility, low/intermediate resistance, and high-level resistance. The predefined susceptibility categories for each NRTI were identical to those used in our previous publication (24). They were chosen to approximate the geometric mean of the published estimated clinical cutoffs provided with the PhenoSense reports. For AZT, 3TC, and FTC, an isolate with <3-fold-decreased susceptibility was considered susceptible; an isolate with 3- to 25-fold-decreased susceptibility was considered to exhibit low/intermediate resistance; and an isolate with >25-fold-decreased susceptibility was considered highly resistant. For ddI, d4T, and TDF, a fold resistance of <1.5 was considered to indicate susceptibility; 1.5- to 3.0-fold resistance was considered low/intermediate resistance; and >3.0-fold resistance was considered a high level of resistance. For ABC, <2-fold resistance was considered to indicate susceptibility; 2- to 6-fold resistance was considered low/intermediate resistance; and >6-fold resistance was considered a high level of resistance.

Mutational interactions. We used four approaches to investigate whether models with interactions improved the prediction of *in vitro* susceptibility. (i) The deletion/substitution/addition (DSA) algorithm explored interactions among the mutations identified by LASSO (31). (ii) Multivariate adaptive regression splines (MARS) progressively tune the maximum allowed interaction constraint parameter *mi* from 1 to 3 (8). (iii) We extended our LSR by including the stepwise addition of interactions to the input matrix of mutation pairs that had previously been identified as significantly covarying in a previous study (23). (iv) We conducted an exhaustive search of all potential two-way interactions among the LASSO-identified mutations by constructing a variable interaction matrix that included all possible two-way interactions in addition to each individual LASSO-identified mutation. We next used LASSO to fit a drug-specific regression model using this larger interaction matrix. Cross-validation was used in both stages to minimize overfitting.

RESULTS

Summary of NRTI susceptibility analysis results. Phenotypic susceptibility results were available for 1,739 HIV-1 isolates from 1,478 individuals. These included 1,687 clinical isolates and 52 laboratory clones or site-directed mutants. To reduce bias resulting from individuals who had more than one virus tested, we excluded from our analysis 228 viruses from individuals with more than one virus having the same mutations at each of the following NRTI resistance positions: 65, 74, 115, 151, 184, and 215. To reduce the confounding effect of virus populations containing mixtures of two or more residues at the same position, we excluded 256 isolates with electrophoretic mixtures at the same positions.

Among the 1,273 isolates included in our analysis, more than 1,100 susceptibility results were available for 3TC, ABC, AZT, d4T, and ddI, 952 for TDF, and 577 for FTC. Overall, 45% of results met the predefined criteria for susceptibility; 28% met those for low/intermediate resistance; and 26% met those for high-level resistance. Table 1 shows the numbers of isolates within each susceptibility category for each of the seven NRTIs. Of the 1,273 isolates, 98.2% belonged to subtype B. Isolates were obtained between 1995 and 2011 (median year: 2003; interquartile range, 2000 to 2007).

Figure 1 shows the extent of cross-resistance between each pair of NRTIs by showing the correlation of the standardized log fold change in susceptibility for each pair of NRTIs. The two cytidine analogs, 3TC and FTC, had the highest correlation ($r = 0.99$). The second and third highest correlations were those between the two thymidine analogs AZT and d4T ($r = 0.83$) and between AZT and TDF ($r = 0.83$). Extremely low correlations were present between the standardized log fold susceptibilities to TDF and 3TC (0.02), TDF and FTC (0.04), AZT and 3TC (0.11), and AZT and FTC (0.22).

NRTI resistance mutations and their effects on specific NRTIs. Among the 177 mutations at 90 positions that occurred 10

TABLE 1 Numbers of HIV-1 isolates with genotype-phenotype correlations for each of the seven NRTIs by predefined resistance category

NRTI[a]	No. (%) of isolates[b]: Susceptible	With low/intermediate resistance	With high-level resistance	Total no. of isolates
3TC	419 (34.5)	204 (16.8)	588 (48.4)	1,211
ABC	373 (32.3)	443 (38.3)	335 (29.0)	1,151
AZT	644 (52.3)	281 (22.8)	324 (26.3)	1,249
d4T	631 (51.0)	363 (29.3)	238 (19.2)	1,232
ddI	568 (45.9)	555 (44.8)	110 (8.9)	1,233
TDF	613 (65.6)	214 (22.9)	125 (13.4)	952
FTC	212 (36.4)	95 (16.3)	270 (46.4)	577
All	3,460 (45.5)	2,155 (28.3)	1,990 (26.1)	7,605

[a] NRTI, nucleoside reverse transcriptase inhibitor; 3TC, lamivudine; ABC, abacavir; AZT, zidovudine; d4T, stavudine; ddI, didanosine; TDF, tenofovir; FTC, emtricitabine.
[b] Susceptible isolates were defined as having susceptibility decreased <1.5-fold for d4T, ddI, and TDF, <2-fold for ABC, and <3-fold for AZT, 3TC, and FTC. Those with low/intermediate resistance were defined as having susceptibility decreased 1.5- to 2.9-fold for d4T, ddI, and TDF, 2- to 5.9-fold for ABC, and 3.0- to 24.9-fold for AZT, 3TC, and FTC. Those with high-level resistance were defined as having susceptibility decreased ≥3.0-fold for d4T, ddI, and TDF, ≥6-fold for ABC, and ≥25-fold for AZT, 3TC, and FTC.

or more times in our data set, LASSO identified 28 mutations at 26 positions as significant predictors of decreased susceptibility to one or more NRTIs. These mutations included V35I, E40F, M41L, K43E, K64H, K65R, D67N, T69ins, K70R, L74V, V75T, F77L, R83K, A98G, K102Q, Y115F, V118I, I135T, Q151M, M184V/I, E203D, H208Y, L210W, T215F/Y/D, D218E, and K219R. To quantify the contribution of the LASSO-identified mutations to reduced susceptibility, we created an NRTI-specific LSR model for each of the seven NRTIs. M184I, which was present in 16 patient samples, was combined with M184V in our analysis. T69ins includes a variety of different double amino acid insertions at this position—most commonly two serines. Figure 2 shows the regression coefficients of the LASSO-identified mutations that were significantly associated with reduced susceptibility to at least one NRTI in the LSR model. The complete list of regression coefficients for each mutation in each of the seven NRTI models is given in Table S1 in the supplemental material.

The median number of LASSO-identified mutations per sample was 3 (interquartile range, 1 to 6). The prevalence of each of the LASSO-identified mutations in the drug susceptibility data set was highly correlated with the prevalence of these mutations in sequences from the approximately 13,000 NRTI-treated individuals in the Stanford HIV Drug Resistance Database (Pearson's *r*, 0.99; *P*, <0.001) (22) (Fig. 3).

The highest regression coefficients (in one or more of the seven LSR models) were those for K65R, T69ins, Y115F, Q151M, and M184V/I (significantly more than 1.0) and those for E40F, K64H, K70R, L74V, V75T, F77L, and T215F/Y (between 0.5 and 1.0). K64H, which was present in only 16 and 13 isolates undergoing d4T and TDF susceptibility testing, had standardized regression coefficients for these two drugs of 0.63 (95% confidence interval [95% CI], 0.629 to 0.631) and 1.17 (95% CI, 1.164 to 1.176), respectively.

The T69 insertion had a coefficient of >0.5 for all seven NRTIs; K65R had a coefficient of >0.5 for six NRTIs; E40F, F77L, Q151M, M184V/I, and T215F/Y each had coefficients of >0.5 for four NRTIs. M184V and M184I were associated with increased susceptibility to TDF, AZT, and d4T; L74V was associated with increased susceptibility to TDF and AZT; and K65R was associated with increased susceptibility to AZT.

Four of the 28 mutations associated with decreased NRTI susceptibility were polymorphic in one or more group M subtypes, including K43E, V118I, I135T, and E203D.

Least-squares regression prediction performance. Table 2 summarizes the categorical and continuous prediction performance of LSR by using each of the mutations identified by LASSO. The categorical performance, or classification accuracy, was the proportion of isolates for which the regression model correctly predicted whether the phenotype was within the bounds of one of the three predefined susceptibility categories: susceptible, exhibiting low/intermediate resistance, or highly resistant. The classification accuracies ranged from 0.77 for ddI, 0.78 to 0.82 for ABC, AZT, TDF, and d4T, and 0.92 to 0.94 for 3TC and FTC. The predictions and actual results were completely discordant (i.e., susceptible versus highly resistant) for about 0.5% of tests (range, 0.26% for ABC to 0.96% for TDF) and partially discordant (i.e., intermediate versus susceptible or intermediate versus highly resistant), on average, for 13% of tests (range, 5.3% for FTC to 22.5% for ddI) (see Table S2 in the supplemental material).

The standardized log fold MSE of 50 trials (5-fold cross-validation performed 10 times) per NRTI ranged from 0.08 (FTC) to 0.35 (TDF), with a standard deviation range of 0.03 (FTC) to 0.08 (TDF) (Table 2).

NRTI mutation interactions. None of the four approaches that incorporated mutational interactions (that is, evaluation for nonlinear effects, such as synergy or antagonism, in NRTI resistance for pairs of mutations)—the deletion/substitution/addition (DSA) partitioning algorithm, multivariate adaptive regression splines (MARS), extension of LASSO to include subsets of previously identified covarying mutations, and extension of LASSO to include all pairwise interactions—improved the accuracy of prediction of reductions in NRTI susceptibility over that with their respective noninteraction versions. Although several models identified pairs of mutations (e.g., T69ins plus T215Y, F77L plus Q151M, and K65R plus Q151M) that interacted synergistically to reduce NRTI susceptibility, these isolated effects did not result in an overall improvement in prediction accuracy and therefore did not justify the use of a complex interaction model.

DISCUSSION

NRTI resistance mutations include those that inhibit NRTI incorporation into the HIV-1 primer DNA strand and those that promote the excision of chain-terminating NRTIs via ATP-mediated pyrophosphorolysis. K65R, K70E, L74V, F115Y, M184V/I, and Q151M plus the Q151M-associated mutations (A62V, V75I, F77L, and F116Y) inhibit NRTI incorporation; whereas M41L, D67N, K70R, L210W, T215Y/F, K219Q/E, and the amino acid T69ins promote NRTI excision. M41L, D67N, K70R, L210W, T215Y/F, and K219Q/E are called thymidine analog mutations (TAMs) because they are selected primarily by the thymidine analogs AZT and d4T. The TAMs have been subclassified into two overlapping clusters: type I (M41L, L210W, and T215Y) and type II (D67N, K70R, T215F, and K219Q/E) TAMs. The mechanisms of action of two additional mutations, T69D and V75T, which were reported in the 1990s to reduce susceptibility to ddC and d4T, respectively (6, 14, 29), have been less well characterized.

With the analysis of increasingly large databases, many addi-

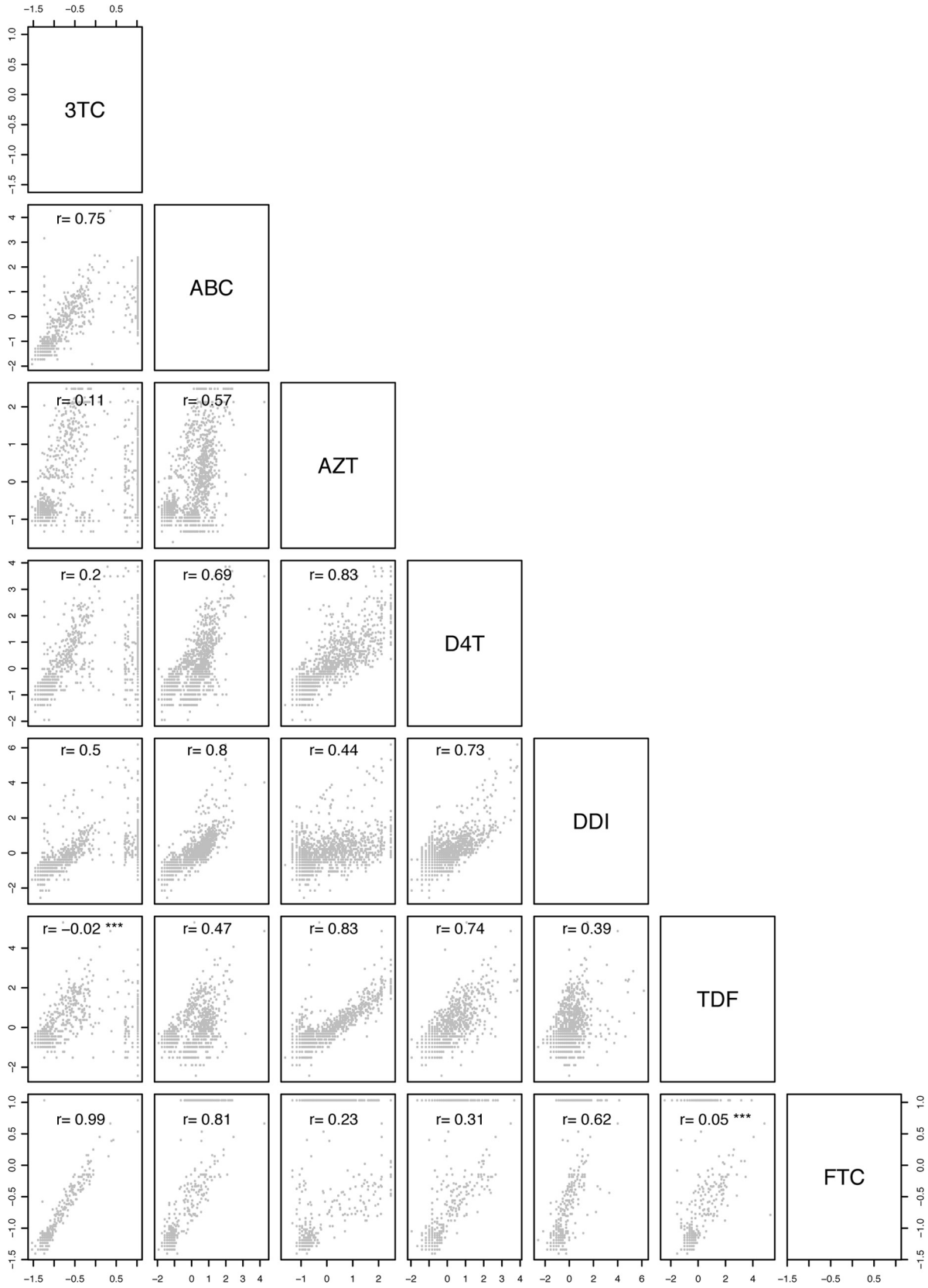

FIG 1 Phenotypic correlation matrix showing standardized HIV-1 log fold cross-resistance between each pair of the seven NRTIs. The Pearson correlation coefficients (*r*) for each of the 21 NRTI pairs are shown. ***, $P > 0.3$; in all other cases, $P < 0.0001$. Drug abbreviations: 3TC, lamivudine; ABC, abacavir; AZT, zidovudine; D4T, stavudine; DDI, didanosine; TDF, tenofovir; FTC, emtricitabine.

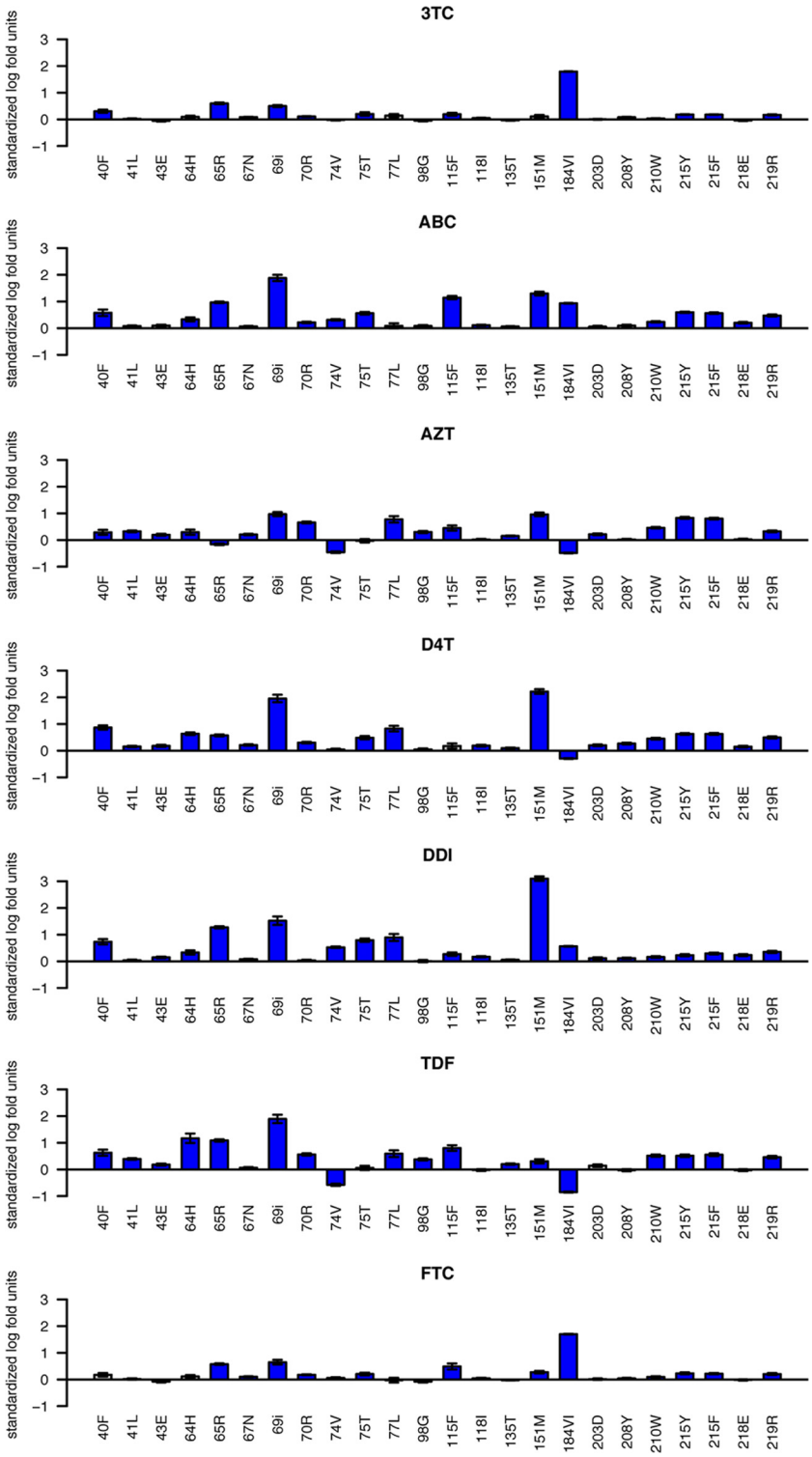

FIG 2 Regression coefficients of the RT mutations found to be significantly associated with decreased susceptibility to at least one NRTI in the least-squares regression models. The mutations shown occurred at least 10 times in the data set. Positive coefficients represent mutations that decrease drug susceptibility; negative coefficients represent mutations that increase drug susceptibility. The *y* axis reflects the mean additive log fold change in HIV-1 susceptibility (expressed in standard deviation units) for the log fold distribution of the respective NRTI. The error bars indicate the standard deviation of the mean generalized error, determined 50 times (10 repetitions of 5-fold cross-validation). Bars representing coefficients whose cross-validated means (as absolute values) are ≥3 standard deviations from zero are blue; other coefficient bars are gray, indicating a lack of statistical significance after cross-validation. Drug abbreviations: 3TC, lamivudine; ABC, abacavir; AZT, zidovudine; D4T, stavudine; DDI, didanosine; TDF, tenofovir; FTC, emtricitabine.

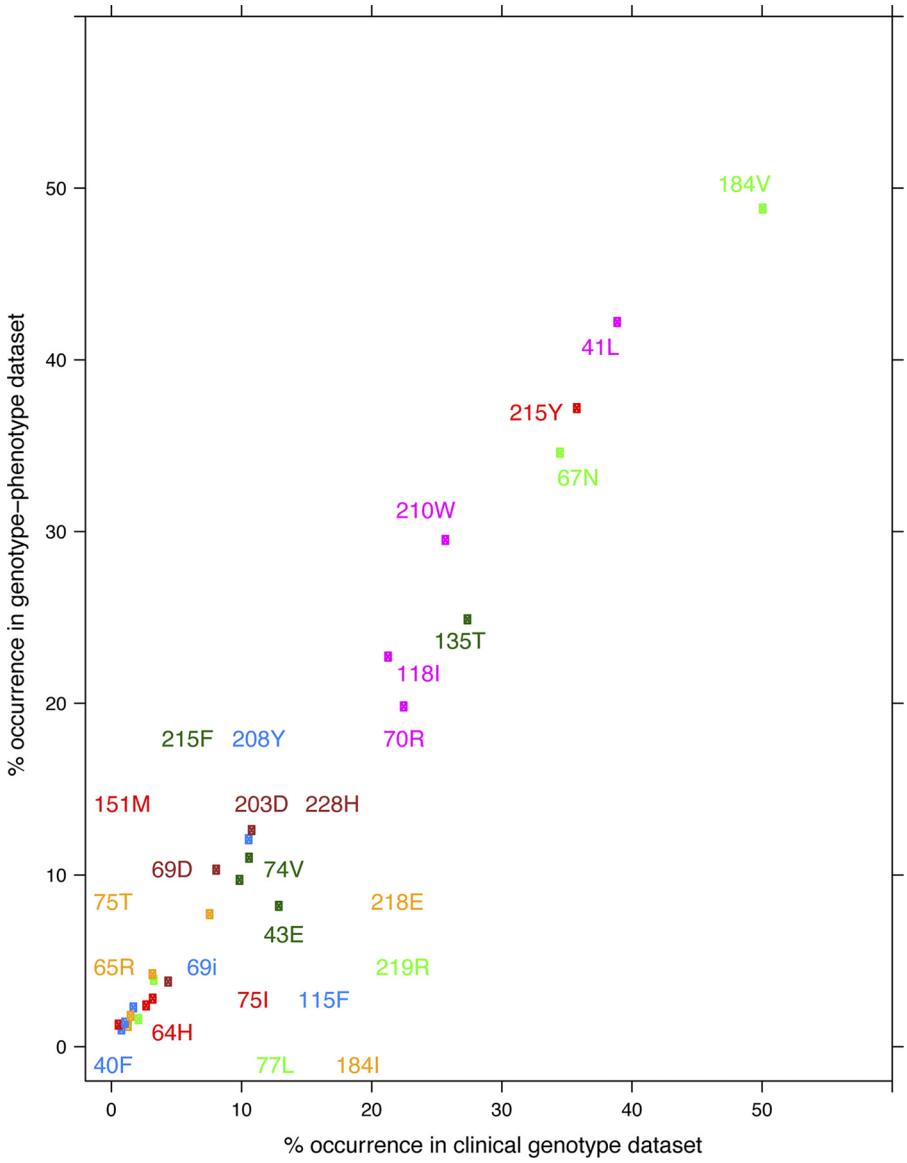

FIG 3 Correlation between the frequency of each of the 28 NRTI resistance mutations in the Stanford HIV Drug Resistance Database (the clinical genotype data set) and that in the genotype-phenotype data set.

TABLE 2 Predictive accuracy and standardized MSE of LSR models[a]

Drug[a]	Classification accuracy[b]	Standardized MSE[c]
3TC	0.92 (0.02)	0.10 (0.03)
ABC	0.82 (0.02)	0.19 (0.04)
AZT	0.79 (0.03)	0.23 (0.04)
d4T	0.81 (0.02)	0.21 (0.04)
ddI	0.77 (0.02)	0.24 (0.04)
TDF	0.78 (0.03)	0.35 (0.08)
FTC	0.94 (0.02)	0.08 (0.03)

[a] 3TC, lamivudine; ABC, abacavir; AZT, zidovudine; d4T, stavudine; ddI, didanosine; TDF, tenofovir; FTC, emtricitabine.
[b] Proportion of isolates for which the regression model correctly predicted whether the phenotype was within the bounds of one of the three predefined NRTI susceptibility categories: susceptible, with low/intermediate resistance, or highly resistant. Values in parentheses denote standard deviations.
[c] Mean squared error between actual and predicted phenotypes. Phenotypes have been standardized to zero mean and unit variance, such that predicted values reflect standard deviation units. Values represent means (with standard deviations in parentheses) derived from 10 repeated and independent runs of 5-fold cross-validation.

tional NRTI-selected mutations have been identified and in some cases have been shown to decrease NRTI susceptibility. Several of these mutations occur at known NRTI resistance positions: K65N, D67G/E/S/T, K70Q/N/G/S/T, L74I, V75M/A/S, and K219N/R/W/D/H (2, 4, 11, 25, 30, 37). Others are at novel positions in the 5′ polymerase coding domain: E40F, K43E/Q/N, E44D/A, V118I, E203K, H208Y, D218E, K223Q/E, and L228H/R (9, 12, 33) (13, 30). Finally, several mutations 3′ to the polymerase coding domain facilitate nucleotide excision, presumably by slowing enzymatic translocation, allowing more time for nucleoside reverse transcriptase inhibitor (NRTI) excision (19). The most important of these mutations, N348I (10, 39), was not evaluated in our study, because it lies outside the RT region that is tested by the PhenoSense assay.

Methodological innovations and prediction accuracy. It has been difficult to determine the phenotypic impact of RT muta-

tions on individual NRTIs, because clinical NRTI-resistant HIV-1 isolates usually contain multiple mutations, often in complex patterns. Moreover, the NRTIs have highly variable *in vitro* dynamic susceptibility ranges (i.e., the fold difference in EC_{50} between highly drug resistant and wild-type viruses). The EC_{50}s of AZT, 3TC, and FTC for highly resistant viruses are usually more than 100 times higher than those for wild-type viruses. In contrast, the EC_{50}s of d4T, ddI, and TDF for highly resistant viruses are rarely more than 5 times higher than those for wild-type viruses. Nonetheless, reductions in susceptibility with EC_{50}s as low as 1.5 times higher than that of the wild type are clinically significant for d4T, ddI, and TDF. The dynamic range for ABC is slightly higher than that for d4T, ddI, and TDF.

To facilitate the comparability of a mutation's effect on different NRTIs despite their different dynamic ranges, we standardized the coefficients for each mutation by dividing the dependent variable (log fold change in HIV susceptibility) by its variance. This provides the ability to assess the relative influences of mutations on decreased susceptibility even for those NRTIs with narrow dynamic ranges. We also chose to study only those phenotypes performed by PhenoSense because of the greater reproducibility of this assay for NRTIs with narrow dynamic ranges (40).

The overall classification accuracy for 3TC, ABC, AZT, d4T, ddI, and TDF was 81.5%, compared with 80.0% in our previous 2006 analysis (24). The classification accuracy improved by $\geq$3.0% for 3TC and ABC and by about 1.0% for the remaining NRTIs. The standardized MSE for these six NRTIs also improved compared with that in our previous analysis, with a decrease from 0.24 to 0.20 over all NRTIs. The rather modest improvement in prediction accuracy despite the increase in the number of genotype-phenotype correlations in this study compared with our previous 2006 study most likely resulted from the ways in which the independent variables were selected in the two studies. In the 2006 study (24), we used external knowledge to choose the independent variables by including nonpolymorphic mutations that had previously been shown to be selected by NRTI therapy. In this study, we made no prior assumptions about the mutations and used the LASSO algorithm—which is particularly useful for selecting a subset of predictors when the set of possible predictors is large—to analyze all 177 mutations that occurred in viruses from 10 or more individuals.

Although the LASSO algorithm is parsimonious, 18 mutations—particularly those with the greatest regression coefficients—were significantly associated with decreased susceptibility to one or more NRTIs in the current and 2006 studies: M41L, K43E, K65R, D67N, T69ins, K70R, L74V, V75T, Y115F, Q151M, M184V/I, H208Y, L210W, T215F/Y, D218E, and K219R. In contrast, K43N/Q, V75M, F116Y, E203K, and L228H were significantly associated with decreased susceptibility only in the 2006 study, whereas E40F, K64H, F77L, A98G, V118I, I135T, and E203D were significantly associated with decreased susceptibility only in the current study.

The fact that regression models containing interaction terms did not significantly improve prediction accuracy suggests that most interactions among NRTI resistance mutations are additive rather than multiplicative. Although a small number of mutational effects may be multiplicative (e.g., T69 insertion and T215Y, F77L, and Q151M), we did not test models that used only preselected mutation pairs. Models that include interactions may not improve prediction accuracy for two additional reasons. Although highly correlated mutations may have multiplicative effects, the numbers of samples in which each of the two mutations occurs alone may be insufficient to demonstrate an interaction. Interactions may also be difficult to observe if some of the independent variables in a model are surrogates for a multiplicative interaction. For example, as noted in the following section, several additional mutations frequently occurred in combination with M41L, L210W, and T215Y (see Fig. S1 in the supplemental material). The inclusion of these additional mutations, therefore, may have made it difficult to identify multiplicative effects among the three type I TAMs.

New insights into NRTI mutations and reduced susceptibility. (i) Known NRTI resistance associations. Our results are consistent with much of the published literature on NRTI susceptibility, including two large *in vitro* studies (35, 36), three intensification or salvage therapy trials that reported associations between preexisting NRTI mutations and the virological response to a new NRTI (15, 17, 18), and numerous studies of individual NRTI resistance mutations. We showed that M184V decreases susceptibility (in descending order) to 3TC or FTC, ABC, and ddI and increases susceptibility (in descending order) to TDF, AZT, and d4T. We showed that D67N and K219Q/E are the TAMs with the least effect on NRTI susceptibility. Indeed, K219Q was not even selected by the LASSO algorithm, while K219E yielded small regression coefficient values. In contrast, the type II TAMs K70R and T215F were found to have statistically significant coefficients for TDF (K70R and T215F) and ABC (T215F).

Y115F, a mutation discovered for its contribution to ABC resistance, was also found to decrease susceptibility to TDF significantly—a finding that has been reported previously (32, 35) but has not garnered much attention. The original study that reported that V75T reduced susceptibility to d4T noted that V75T reduced susceptibility to ddI (14). However, this association has not generally been cited. In contrast, our results indicate that V75T appears to contribute as much to reduced susceptibility to ddI as it does to reduced susceptibility to d4T.

Despite the finding that most mutations were associated with decreased susceptibility to multiple NRTIs, the correlations in the levels of resistance between AZT and 3TC, AZT and FTC, TDF and 3TC, and TDF and FTC were strikingly low. This observation, which was reported previously by Whitcomb et al. (36), results from the fact that the most common NRTI resistance mutation, M184V, which causes reduced susceptibility to 3TC and FTC, increases susceptibility to AZT and TDF. This mutational interaction likely explains the clinical efficacy of NRTI backbones containing AZT or TDF in combination with a cytidine analog such as 3TC or FTC. However, not all efficacious dual NRTI backbones benefit from this interaction. The combination of ABC and 3TC is highly effective under most circumstances despite the fact that M184V decreases susceptibility to both NRTIs. The effectiveness of this combination may result from the fact that ABC has the greatest antiviral activity except for the cytidine analogs (27). Nonetheless, the NRTI backbone of ABC and 3TC was found to be less effective than that of TDF and FTC for patients with high viral loads in a recent large clinical trial (28).

(ii) Novel NRTI resistance associations. E40F and K219R, two previously reported but poorly characterized NRTI-associated mutations, were associated with significantly decreased susceptibility to six and seven NRTIs, respectively. This association appears to be the result of each mutation's strong correlation with type I TAMs. Among the 13 patients with viruses containing E40F,

11 (84%) also had M41L, L210W, and T215Y. Among the 49 patients with viruses containing K219R, 41 (84%) also had the same three type I TAMs. In contrast, 26% of all viruses in the study had each of the three type I TAMs.

K64H, K64N, and K64Y are nonpolymorphic mutations that are strongly selected by NRTI therapy (22, 30). Each of these K64 variants was recently reported to occur in <0.1% of 12,730 ARV-naïve patients compared with 0.5% to 1.1% of 4,598 patients with a history of receiving NRTIs but not NNRTIs (30). In the current study, K64H was significantly associated with decreased susceptibility to d4T (16 patients; regression coefficient, 0.63) and TDF (13 patients; regression coefficient, 1.2). K64H occurred in combination with ≥3 type II TAMs in 12 patients and in combination with M41L, L210W, and T215Y in 4 patients. To further define the effect of mutations at position 64, we performed site-directed mutagenesis to back mutate clones with K64H from four isolates and clones with the less frequently detected mutations K64N and K64Y from one isolate each. Susceptibility testing of the six isogenic pairs of clones showed that K64H induced a median 1.4-fold (range, 1.3- to 1.6-fold) and 1.3-fold (range, 1.2- to 1.8-fold) decreased susceptibility to d4T and TDF, respectively. K64N induced 2.4-fold and 1.4-fold decreased susceptibility to d4T and TDF, respectively. K64Y induced 2.1-fold and 1.8-fold decreased susceptibility to d4T and TDF, respectively. Further studies of viruses with these mutations are ongoing.

A98G was first reported to reduce susceptibility to several NNRTIs in the early 1990s (3). However, we recently reported that A98G was selected by NRTIs as well as NNRTIs, because it occurred in 25 (0.2%) of 12,370 ARV-naïve patients, 97 (2.1%) of 4,598 patients treated with NRTIs but not NNRTIs, and 711 (8.5%) of 8,367 patients treated with NNRTIs (usually in combination with NRTIs) (30). The most likely explanation for the association with slightly decreased susceptibility to AZT and TDF was that 42/68 (62%) of viruses with A98G also had M41L, L210W, and T215Y. The only other NNRTI mutation shown to influence NRTI susceptibility was Y181C, which, as previously reported, modestly increased susceptibility to AZT and TDF (16).

Conclusion. Initial and salvage ARV therapies have become increasingly effective in well-resourced countries. Potent ARVs from five mechanistic classes are now routinely used in combination with NRTIs. It has therefore become increasingly difficult to assess the impact of baseline NRTI resistance mutations on the response to an NRTI used as part of a salvage therapy regimen. Therefore, correlations between RT mutations and *in vitro* NRTI susceptibility are increasingly important for quantifying the effects of NRTI mutations on susceptibility to NRTIs.

Our study provides a comprehensive yet fine-grained view of the most common NRTI resistance mutations. Because our results were standardized by the variance in the log fold resistance levels for each NRTI, we provide the first analysis that quantifies the relative phenotypic effect of each mutation across each of the NRTIs. Despite the use of a feature selection approach designed to assess the potential roles of many different RT mutations, the NRTI resistance mutations we identified with the greatest effect on NRTI susceptibility were for the most part known nonpolymorphic treatment-selected mutations. Although one of these mutations, K64H, was not previously reported to decrease susceptibility to NRTIs, it was recently reported to be under strong NRTI selection pressure (30), and site-directed mutagenesis experiments were consistent with our regression model. For several other mutations, novel associations with decreased susceptibility to specific NRTIs were identified and in some cases explained by their association with other, more common NRTI resistance mutations.

ACKNOWLEDGMENTS

The study was supported by funding from two grants: AI068581 (NIH) and CHRP D08-ST-033.

REFERENCES

1. **Bennett DE, et al.** 2009. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PLoS One **4**:e4724.
2. **Bradshaw D, et al.** 2007. Novel drug resistance pattern associated with the mutations K70G and M184V in human immunodeficiency virus type 1 reverse transcriptase. Antimicrob. Agents Chemother. **51**:4489–4491.
3. **Byrnes VW, et al.** 1993. Comprehensive mutant enzyme and viral variant assessment of human immunodeficiency virus type 1 reverse transcriptase resistance to nonnucleoside inhibitors. Antimicrob. Agents Chemother. **37**:1576–1579.
4. **Delaugerre C, et al.** 2005. Selection of a rare resistance profile in an HIV-1-infected patient exhibiting a failure to an antiretroviral regimen including tenofovir DF. J. Clin. Virol. **32**:241–244.
5. **de Oliveira T, et al.** 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics **21**:3797–3800.
6. **Fitzgibbon JE, et al.** 1992. Human immunodeficiency virus type 1 *pol* gene mutations which cause decreased susceptibility to 2′,3′-dideoxycytidine. Antimicrob. Agents Chemother. **36**:153–157.
7. **Friedman J, Hastie T, Tibshirani R.** 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. **33**:1–22.
8. **Friedman JH, Roosen CB.** 1995. An introduction to multivariate adaptive regression splines. Stat. Methods Med. Res. **4**:197–217.
9. **Gonzales MJ, et al.** 2003. Extended spectrum of HIV-1 reverse transcriptase mutations in patients receiving multiple nucleoside analog inhibitors. AIDS **17**:791–799.
10. **Hachiya A, et al.** 2008. Amino acid mutation N348I in the connection subdomain of human immunodeficiency virus type 1 reverse transcriptase confers multiclass resistance to nucleoside and nonnucleoside reverse transcriptase inhibitors. J. Virol. **82**:3261–3270.
11. **Hachiya A, et al.** 2011. K70Q adds high-level tenofovir resistance to "Q151M complex" HIV reverse transcriptase through the enhanced discrimination mechanism. PLoS One **6**:e16242.
12. **Hertogs K, et al.** 2000. A novel human immunodeficiency virus type 1 reverse transcriptase mutational pattern confers phenotypic lamivudine resistance in the absence of mutation 184V. Antimicrob. Agents Chemother. **44**:568–573.
13. **Huigen MC, et al.** 2008. Identification of a novel resistance (E40F) and compensatory (K43E) substitution in HIV-1 reverse transcriptase. Retrovirology **5**:20.
14. **Lacey SF, Larder BA.** 1994. Novel mutation (V75T) in human immunodeficiency virus type 1 reverse transcriptase confers resistance to 2′,3′-didehydro-2′,3′-dideoxythymidine in cell culture. Antimicrob. Agents Chemother. **38**:1428–1432.
15. **Lanier ER, et al.** 2004. Antiviral efficacy of abacavir in antiretroviral therapy-experienced adults harbouring HIV-1 with specific patterns of resistance to nucleoside reverse transcriptase inhibitors. Antivir. Ther. **9**:37–45.
16. **Larder BA.** 1992. 3′-Azido-3′-deoxythymidine resistance suppressed by a mutation conferring human immunodeficiency virus type 1 resistance to nonnucleoside reverse transcriptase inhibitors. Antimicrob. Agents Chemother. **36**:2664–2669.
17. **Miller M, et al.** 2004. Baseline genotypic analysis of treatment-naive patients taking tenofovir DF (TDF) or stavudine (d4T) in combination with lamivudine (3TC) and efavirenz (EFV). Antivir. Ther. **9**:S173.
18. **Molina JM, et al.** 2005. Didanosine in HIV-1-infected patients experiencing failure of antiretroviral therapy: a randomized placebo-controlled trial. J. Infect. Dis. **191**:840–847.
19. **Nikolenko GN, et al.** 2005. Mechanism for nucleoside analog-mediated abrogation of HIV-1 replication: balance between RNase H activity and nucleotide excision. Proc. Natl. Acad. Sci. U. S. A. **102**:2093–2098.
20. **Petropoulos CJ, et al.** 2000. A novel phenotypic drug susceptibility assay

for human immunodeficiency virus type 1. Antimicrob. Agents Chemother. **44**:920–928.
21. **R Development Core Team.** 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
22. **Rhee SY, et al.** 2003. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. **31**:298–303.
23. **Rhee SY, Liu TF, Holmes SP, Shafer RW.** 2007. HIV-1 subtype B protease and reverse transcriptase amino acid covariation. PLoS Comput. Biol. **3**:e87.
24. **Rhee SY, et al.** 2006. Genotypic predictors of human immunodeficiency virus type 1 drug resistance. Proc. Natl. Acad. Sci. U. S. A. **103**:17355–17360.
25. **Ross LL, et al.** 2006. A rare HIV reverse transcriptase mutation, K65N, confers reduced susceptibility to tenofovir, lamivudine and didanosine. AIDS **20**:787–789.
26. **Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusova T.** 2004. A web-based genotyping resource for viral sequences. Nucleic Acids Res. **32**:W654–W659.
27. **Saag MS, et al.** 1998. Antiretroviral effect and safety of abacavir alone and in combination with zidovudine in HIV-infected adults. Abacavir Phase 2 Clinical Team. AIDS **12**:F203–F209.
28. **Sax PE, et al.** 2009. Abacavir-lamivudine versus tenofovir-emtricitabine for initial HIV-1 therapy. N. Engl. J. Med. **361**:2230–2240.
29. **Selmi B, et al.** 2001. The valine-to-threonine 75 substitution in human immunodeficiency virus type 1 reverse transcriptase and its relation with stavudine resistance. J. Biol. Chem. **276**:13965–13974.
30. **Shahriar R, et al.** 2009. Nonpolymorphic human immunodeficiency virus type 1 protease and reverse transcriptase treatment-selected mutations. Antimicrob. Agents Chemother. **53**:4869–4878.
31. **Sinisi SE, van der Laan MJ.** 2004. Deletion/substitution/addition algorithm in learning with applications in genomics. Stat. Appl. Genet. Mol. Biol. **3**:Article18.
32. **Stone C, Ait-Khaled M, Craig C, Griffin P, Tisdale M.** 2004. Human immunodeficiency virus type 1 reverse transcriptase mutation selection during in vitro exposure to tenofovir alone or combined with abacavir or lamivudine. Antimicrob. Agents Chemother. **48**:1413–1415.
33. **Svicher V, et al.** 2006. Involvement of novel human immunodeficiency virus type 1 reverse transcriptase mutations in the regulation of resistance to nucleoside inhibitors. J. Virol. **80**:7186–7198.
34. **U.S. Department of Health and Human Services Panel on Clinical Practices for Treatment of HIV Infection. January** 2011. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf.
35. **Vermeiren H, et al.** 2007. Prediction of HIV-1 drug susceptibility phenotype from the viral genotype using linear regression modeling. J. Virol. Methods **145**:47–55.
36. **Whitcomb JM, Parkin NT, Chappey C, Hellmann NS, Petropoulos CJ.** 2003. Broad nucleoside reverse-transcriptase inhibitor cross-resistance in human immunodeficiency virus type 1 clinical isolates. J. Infect. Dis. **188**:992–1000.
37. **Wirden M, et al.** 2009. Antiretroviral combinations implicated in emergence of the L74I and L74V resistance mutations in HIV-1-infected patients. AIDS **23**:95–99.
38. **World Health Organization HIV/AIDS Programme.** 2010. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach, 2010 revision. World Health Organization, Geneva, Switzerland.
39. **Yap SH, et al.** 2007. N348I in the connection domain of HIV-1 reverse transcriptase confers zidovudine and nevirapine resistance. PLoS Med. **4**:e335.
40. **Zhang J, Rhee SY, Taylor J, Shafer RW.** 2005. Comparison of the precision and sensitivity of the Antivirogram and PhenoSense HIV drug susceptibility assays. J. Acquir. Immune Defic. Syndr. **38**:439–444.