Transmitted Drug Resistance in the CFAR Network of Integrated Clinical Systems Cohort: Prevalence and Effects on Pre-Therapy CD4 and Viral Load

Art F. Y. Poon[1], Jeannette L. Aldous[2], W. Christopher Mathews[2], Mari Kitahata[3], James S. Kahn[4], Michael S. Saag[5], Benigno Rodríguez[6], Stephen L. Boswell[7], Simon D. W. Frost[8], Richard H. Haubrich[2]*

1 BC Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada, 2 Department of Medicine, University of California San Diego, San Diego, California, United States of America, 3 University of Washington, Seattle, Washington, United States of America, 4 University of California San Francisco, San Francisco, California, United States of America, 5 University of Alabama Birmingham, Birmingham, Alabama, United States of America, 6 Case Western Reserve University, Cleveland, Ohio, United States of America, 7 Fenway Community Health/Harvard Medical School, Boston, Massachusetts, United States of America, 8 University of Cambridge, Cambridge, United Kingdom

Abstract

Human immunodeficiency virus type 1 (HIV-1) genomes often carry one or more mutations associated with drug resistance upon transmission into a therapy-naïve individual. We assessed the prevalence and clinical significance of transmitted drug resistance (TDR) in chronically-infected therapy-naïve patients enrolled in a multi-center cohort in North America. Pre-therapy clinical significance was quantified by plasma viral load (pVL) and CD4+ cell count (CD4) at baseline. Naïve bulk sequences of HIV-1 protease and reverse transcriptase (RT) were screened for resistance mutations as defined by the World Health Organization surveillance list. The overall prevalence of TDR was 14.2%. We used a Bayesian network to identify co-transmission of TDR mutations in clusters associated with specific drugs or drug classes. Aggregate effects of mutations by drug class were estimated by fitting linear models of pVL and CD4 on weighted sums over TDR mutations according to the Stanford HIV Database algorithm. Transmitted resistance to both classes of reverse transcriptase inhibitors was significantly associated with lower CD4, but had opposing effects on pVL. In contrast, position-specific analyses of TDR mutations revealed substantial effects on CD4 and pVL at several residue positions that were being masked in the aggregate analyses, and significant interaction effects as well. Residue positions in RT with predominant effects on CD4 or pVL (D67 and M184) were re-evaluated in causal models using an inverse probability-weighting scheme to address the problem of confounding by other mutations and demographic or risk factors. We found that causal effect estimates of mutations M184V/I ($-1.7\log_{10}$ pVL) and D67N/G ($-2.1\sqrt[3]{\text{CD4}}$ and $+0.4\log_{10}$ pVL) were compensated by K103N/S and K219Q/E/N/R. As TDR becomes an increasing dilemma in this modern era of highly-active antiretroviral therapy, these results have immediate significance for the clinical management of HIV-1 infections and our understanding of the ongoing adaptation of HIV-1 to human populations.

Introduction

A substantial fraction ($\approx 5\%$–20%) of human immunodeficiency virus type 1 (HIV-1) variants that are transmitted and establish infections in individuals who have never been exposed to antiretroviral drugs (i.e., ARV-naïve) already carry one or more mutations that are associated with resistance to ARVs [1,2]. Resistance mutations are generally slow to revert in ARV-naïve hosts [3,4] and can persist at low frequencies in treatment-experienced hosts as well [5]. The pre-existence of resistance mutations can significantly diminish effectiveness of subsequent ARV therapy [6,7]; for example, transmitted K103N mutations are associated with increased regimen failure and poorer outcomes on non-nucleoside reverse transcriptase inhibitors (NNRTIs) [8]. On the other hand, resistance mutations can also incur a measurable cost with respect to virus replicative capacity in vitro [9–11] and competitive growth in vivo [12]. Within therapy-naïve hosts, the fitness costs of transmitted drug resistance (TDR) mutations may influence the subsequent evolution and population dynamics of HIV-1. With the exception of limiting the choice of initial ARV regimens, however, the direct clinical consequences of TDR mutations in therapy-naïve hosts remain inconclusive [13–17].

Our central hypothesis in this study is that the presence of TDR mutations in therapy-naïve hosts has a measurable effect on not only the subsequent rate of growth of the virus population within the host, but also on the rate of depletion of CD4+ T-lymphocytes

that are targeted by HIV-1. To evaluate this hypothesis, we assess the relationship between TDR mutations and prognostic clinical markers – CD4+ T-cell count (CD4) and plasma HIV-1 RNA load (pVL) – at baseline using data from the Centers for AIDS Research (CFAR) Network of Integrated Clinical Systems (CNICS), an observational cohort created from a data-sharing collaboration between multiple clinical sites in North America [18]. 'Baseline' refers to the measurements of clinical markers upon HIV diagnosis and/or entry into clinical care. It does not correspond to a consistent time point after infection and many patients will have already progressed into the chronic phase of infection. Nevertheless, these data can provide useful insight into the clinical implications of TDR in therapy-naïve hosts.

First, we quantify the effects of TDR mutations at both aggregate and position-specific levels. An aggregate approach sums across mutations occupying different positions in HIV-1 protease and reverse transcriptase, implicitly assuming that every mutation has a similar effect (in both direction and magnitude) on baseline CD4 or pVL. In contrast, a position-specific analysis recognizes the possibility that some TDR mutations will have a disproportionately large effects while others have negligible effects. Furthermore, it enables us to identify compensatory interactions, such that the effect of a specific mutation on CD4 or pVL depends on the presence or absence of other mutations; such interactions are an important aspect of the evolution of drug resistance in HIV-1 [11]. Second, we re-evaluate statistically significant associations of individual TDR mutations in a causal modeling framework, i.e., inverse probability-of-treatment weighting [19]. A statistical association between a treatment (a TDR mutation) and an outcome (baseline CD4 or pVL) in an observational study may be biased if one or more variables, such as other co-transmitted resistance mutations, are not only correlated with the treatment but also affect the outcome [19]. In the presence of such confounding factors, the significant association could be wrongly attributed to a causal effect of the treatment on the outcome. Using statistical procedures that can adjust for confounding by weighting observational data has become an important component of understanding the consequences of drug resistance in HIV [20].

Methods

Ethics statement

All patients in the CNICS database have provided informed written consent, as reviewed and approved by the institutional review boards of all participating sites – University of Alabama, Birmingham (UAB); University of California, San Diego (UCSD); University of California, San Francisco (UCSF); University of Washington (UW); Case Western Reserve University (CWRU); and Harvard University (FENWAY) – to have their clinical information used for the purposes of research. Patients were informed on the minimum information being collected, alternatives to participation, and possible uses of the data. In addition, CNICS has received a Certificate of Confidentiality from the National Institutes of Health (NIH).

Data collection

The study population comprised individuals from 6 CNICS sites: University of Alabama, Birmingham; University of California, San Diego; University of California, San Francisco; University of Washington; Case Western Reserve University; and Harvard University [18]. All patients in the CNICS database have provided informed written consent, as reviewed and approved by the institutional review boards of all participating sites, to have their

clinical information used for the purposes of research. HIV-1 *pol* sequences in the region encoding protease and/or part of reverse transcriptase (i.e., the first 246–400 codons of RT) were obtained at each site by conventional Sanger sequencing directly from RT-PCR amplification products from sample extractions of viral RNA. The exact bulk sequencing protocol varied among sites, as some sites used commercial laboratory services (viz., Monogram, Quest, Virco, LabCorp) whereas others used in-house assays based on commercial kits. At the time of this study, 4914 codon or amino acid sequences were available in the CNICS database but many were obtained when patients were on-therapy. Amino acid sequences were reconstituted from reports of residue polymorphisms relative to the HXB2 reference sequence. Sequences were linked to: unique anonymized patient identifiers; CNICS site; sampling date; age; ARV regimens; demographic factors (self-reported gender at baseline, race/ethnicity); HIV risk factors (men who have sex with men, MSM; injection drug use, IDU); baseline plasma HIV RNA titres (pVL, virus/mL plasma); and baseline CD4 cell counts (cells/mL). A patient record was included in this study if: (1) a sequence was sampled before therapy, and (2) CD4 and/or pVL were sampled within 120 days of the sequence sample date. To normalize the distributions of pVL and CD4, we used log_{10} and cube-root transformations [21], respectively. Race and ethnicity records were aggregated into the following groups: non-Hispanic white, non-Hispanic black, Hispanic, and other. Non-random associations between demographic and risk factors were identified using log-linear models.

Sequence analysis

Sequences were aligned pairwise against the NL4-3 *pol* reference sequence using an implementation of the Gotoh algorithm in HyPhy [22,23] under default settings. Surveillance drug resistance mutations (SDRMs) were tallied for each amino acid sequence according to the 2009 update of the World Health Organization list of surveillance drug resistance mutations (SDRMs) [24]. Amino acid polymorphisms (e.g., bulk sequence mixtures) were resolved to the resistant residue when applicable.

Statistical analysis

Resistance scores by drug class (protease inhibitors, PIs; nucleoside and non-nucleoside reverse transcriptase inhibitors, NRTIs/NNRTIs) were calculated according to the Stanford HIV Drug Resistance database algorithm [25]. The Stanford scoring system is a linear predictor of drug resistance phenotypes that works by assigning integer weights to a sequence for encoding resistance-associated amino acid residues at specific positions. Weights (or 'scores') for a given sequences were calculated for every drug in a given drug class. We used the highest score of any drug in a drug class as the class-specific score. These quantitative scores, denoted by S_{PI}, S_{NRTI} and S_{NNRTI} respectively, were used to analyze the aggregate effects of TDR on baseline CD4 and pVL. Aggregate or position-specific effects of TDR on baseline CD4 or pVL were evaluated alongside demographic and risk factors in linear models. In each case we used a bi-directional stepwise algorithm in R (stepAIC [26]) to select a 'best-fitting' linear model based on the Akaike information criterion (AIC), which penalizes model likelihood by the number of parameters. We seeded the algorithm with the full additive model (without interaction terms) and limited model complexity to second-order interactions. To avoid over-fitting the data from evaluating a large number of alternative models, we averaged parameter estimates and P values over an ensemble of models (best +5 next-best) using Akaike weights [27].

To estimate causal effects of individual SDRMs, we used inverse probability-of-treatment weight (IPTW) estimators [19]. The conditional probability $Pr(A|\Lambda)$ for a given 'treatment' variable (A) was computed by fitting a logistic model of A against the set of all potential confounding variables (Λ). To account for compensatory interactions among mutations, we combined the corresponding presence/absence variables into a single multinomial treatment as suggested to us by Miguel A. Hernán (pers. comm.). For example:

$$A_{M184} = \begin{cases} 0 & \text{if} \quad M184X=0 \text{ and } (K103X=0 \text{ and } K219X=0) \\ 1 & \text{if} \quad M184X=1 \text{ and } (K103X=0 \text{ and } K219X=0) \\ 2 & \text{if} \quad M184X=1 \text{ and } (K103X=1 \text{ or } K219X=1) \end{cases}$$

where X signifies any SDRM residue at that position. Conditional probabilities for multinomial treatments were obtained by fitting a multinomial logistic model with the *multinom* function in the R package *nnet*. Because weighting observations by the inverse of $Pr(A|\Lambda)$ can introduce excessive variability, we used stabilized weights that normalize $Pr(A|\Lambda)$ by the marginal probability $Pr(A)$, which was estimated by fitting a logistic model containing only the intercept term [19]. The stabilized weights were incorporated into a linear model of baseline CD4 or pVL on treatment A using the *svyglm* function in the R package *survey* [28], which uses an estimating equation-based approach to fitting generalized linear models with robust standard errors. Cases with missing data were omitted from these analyses.

A Bayesian network analysis to detect co-transmission of TDR mutations was carried out using a Markov chain Monte Carlo (MCMC) procedure in HyPhy [29]. A Bayesian network is a compact representation of the joint probability distribution of multiple variables (*i.e.*, the presence/absence of TDR mutations), where 'joint' implies that the probability distribution of a variable may be conditional on other variables in the network [30]. In other words , a Bayesian network provides an efficient framework for exploring the statistical interactions among variables. Each variable is represented by a 'node' in the network, and connections between nodes ('edges') are drawn to indicate that one variable is conditioned on the second. The MCMC procedure obtains a random sample of Bayesian networks from the posterior probability distribution defined by the data. The Markov chain was propagated for 10^6 steps, discarding the first half as burn-in and thinning the remaining sample to 100 steps at equal intervals. Edges in the network signifying co-transmission were accepted if they were present in 90% or more of the networks in the sample.

Results

Population

The study population was comprised of 14111 individuals with a median baseline CD4 of 300 cells/mL ($n=14011$) and a median baseline pVL of 24400 copies/mL ($n=13953$). Median year of birth was 1964. The majority of the population self-reported as 'white' ($n=8694$). 3725 individuals self-reported as 'black', 207 as 'Asian/Pacific Islander' and 835 as 'other'; there were 650 non-respondents. At the time of this study, the CNICS database contained 1585 sequence records that we classified as therapy-naïve. Out of these, 1575 naïve sequences were linked to baseline CD4 and pVL records. 1526 naïve sequences were associated with baseline CD4 records within 120 days of the sequence sampling date and were otherwise censored from subsequent analyses. Likewise, 1516 naïve sequences were associated with baseline pVL within 120 days of sampling and were otherwise censored. The

median baseline CD4 and pVL of patients with naïve sequences in this data set were 285 cells/mL and $10^{4.7}$ copies/mL, respectively. Composition of the naïve sample with respect to race was significantly associated with gender ($\chi^2=68.2$, $df=3$, $P=1.03\times10^{-14}$); *e.g.*, the sample included a disproportionately greater number of naïve sequences from white males (odds ratio, $OR=3.1$) and black females ($OR=3.0$). Sample composition by race was also significantly associated with MSM ($\chi^2=138.8$, $df=3$, $P=7.05\times10^{-30}$), largely because of a greater number of white MSMs than expected by marginal frequencies ($OR=2.85$).

Prevalence of transmitted SDRMs

Overall, 225 out of 1585 naïve sequences (14.2%) contained one or more SDRMs. Of these 225 sequences, 66 (4.2%) contained at least one PI-associated SDRM, 130 (8.2%) contained at least one NRTI-associated SDRM, and 131 (8.3%) contained at least one NNRTI-associated SDRM. In addition, 17 (1.1%) sequences contained SDRMs in all three drug classes, which is substantially greater than the number expected by chance (<0.5 sequences) if the joint transmission of mutations was independent of drug class. A decline from pre-2003 levels in the fraction of naïve sequences with at least 1 SDRM was mirrored by a similar decline in the mean number of SDRMs per naïve sequence (Figure 1). However, we observed an increase in the most recent sample year (2008) in the upper 95% quartile of the distribution in the number of SDRMs. Neither measure of prevalence was significantly associated with demographic or risk factors in generalized linear models ($P>0.05$). Counts of SDRMs by residue position are provided in Table 1. A Bayesian network analysis detected significant co-transmission of SDRMs in four groups (Figure 2). The largest group generally corresponded to NRTI resistance-associated
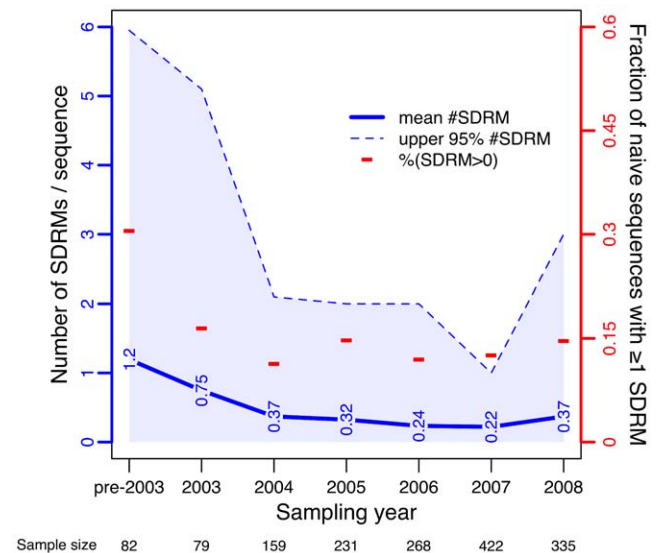
Figure 1. Prevalence of transmitted surveillance drug resistance mutations (SDRMs) over time. Sampling years 1998–2002 were grouped into a single level ('pre-2003') to adjust for small sample sizes ($n=3,5,18$, and 56, respectively). Sample sizes for the bins are displayed below the x-axis. Dashes (−) indicate the fraction of naïve sequences in a given sampling period that contain at least one SDRM (right y-axis). The trend in mean number of SDRMs per naïve sequence (left y-axis) over time is displayed as a solid line annotated with the actual values. The 95% quantile per time point is depicted by a lighter line bounding a shaded region to illustrate the trend in the distributions (all medians were zero).

doi:10.1371/journal.pone.0021189.g001

Table 1. Number and percentage prevalence of putatively transmitted surveillance drug resistance mutations (SDRMs) by residue position in antiretroviral (ARV)-naïve sequences.

Class	position	count	% prevalence	Class	position	count	% prevalence
PI	L23	1	0.06	NRTI(cont'd)	K70	9	0.57
	L24	1	0.06		L74	11	0.69
	D30	15	0.95		V75	2	0.13
	V32	2	0.13		F77	1	0.06
	M46	20	1.26		Y115	1	0.06
	I47	2	0.13		F116	1	0.06
	G48	0	0		Q151	0	0
	I50	2	0.13		M184	40	2.52
	F53	1	0.06		L210	29	1.83
	I54	16	1.01		T215	77	4.86
	G73	3	0.19		K219	30	1.89
	L76	1	0.06	NNRTI	L100	1	0.06
	V82	12	0.76		K101	7	0.44
	N83	0	0		K103	82	5.17
	I84	10	0.63		V106	2	0.13
	I85	2	0.13		V179	1	0.06
	N88	15	0.95		Y181	24	1.51
	L90	30	1.89		Y188	10	0.63
NRTI	M41	49	3.09		G190	24	1.51
	K65	3	0.19		P225	1	0.06
	D67	28	1.77		M230	0	0
	T69	9	0.57				

PI = protease inhibitor; NRTI = nucleoside reverse transcriptase inhibitor; NNRTI = non-nucleoside reverse transcriptase inhibitor.
doi:10.1371/journal.pone.0021189.t001

mutations (with the exception of K103, I54 and V82) and contained well-characterized mutational pathways, *e.g.*, D67/T69/K70/K219 and T215/M41/L210. Two smaller groups (M46/I84/I85 and D30/N88) comprised mutations associated with PI resistance, the latter group being associated with resistance to nelfinavir in particular. The remaining group (K101/G190) comprised mutations associated with resistance to NNRTIs. In sum, co-transmitted SDRMs tended to confer resistance to the same drug or drug class.

Aggregate SDRM predictors of baseline CD4 and HIV RNA

The summated Stanford scores for NRTI and NNRTI (S_{NRTI}, S_{NNRTI}) and race/ethnicity were significantly ($\alpha = 0.05$) associated with baseline CD4 (Table 2). For every 10 units of S_{NRTI} and S_{NNRTI}, the model predicted a decline in $\sqrt[3]{\text{CD4}}$ by about 0.07 and 0.04 units, respectively. For example, if a hypothetical patient whose CD4 count matched the study median (285 cells/mL) became infected by an HIV-1 variant carrying SDRMs such that $S_{\text{NRTI}} = 100$ and $S_{\text{NNRTI}} = 100$, we would predict a baseline CD4 count of $(\sqrt[3]{285} - 0.069 \times 10 - 0.044 \times 10)^3 \approx 162$ cells/mL. The joint effect of transmitted RTI resistance on CD4 is illustrated in Figure 3. Only about 5.8% of the variance in $\sqrt[3]{\text{CD4}}$ was explained by this model. Similarly, S_{NRTI}, S_{NNRTI}, gender and race/ethnicity were significantly associated with baseline pVL (Table 2). For every 10 units in S_{NRTI}, pVL was predicted to decline by about $0.03 \log_{10}$ units; in contrast, pVL was predicted to increase by about $0.02 \log_{10}$ units for every 10 units in S_{NNRTI}

(Figure 4). Only 2% of the variance in $\log_{10}$ pVL was explained by this model. S_{PI} was not significantly associated with either CD4 or pVL at baseline.

Position-specific SDRM predictors of baseline CD4 and HIV RNA

Although aggregate SDRM-based statistics such as Stanford scores are convenient and easy to interpret as proximate measures of drug class-specific resistance, they may mask the effects of individual mutations on clinical outcome. Consequently, we assessed the effects of position-specific SDRMs in linear models of baseline CD4 and pVL. To minimize the number of predictor variables and avoid over-fitting of the data, we limited our analysis to positions with a minimum SDRM frequency of 1% in naïve sequences (see Table 1). These positions were: M46, I54, and L90 in protease, and; M41, D67, M184, L210, T215, K219, K103, Y181, and G190 in reverse transcriptase. The presence or absence of SDRMs at each position was represented by a binary variable (herein denoted by appending an 'X' to the amino acid positional notation, *e.g.*, M46X = {0,1}).

D67X and K219X were each associated with significantly lower baseline CD4 ($P < 0.04$; Table 3); both sets of SDRMs are thymidine analog mutations (TAMs) conferring resistance to NRTIs. For example, the model would predict a hypothetical decline from 400 cells/mL to 146 and 235 cells/mL with the introduction of SDRMs at positions D67 and K219, respectively. In addition, there was a significant interaction effect between D67X and K219X associated with a compensatory increase by
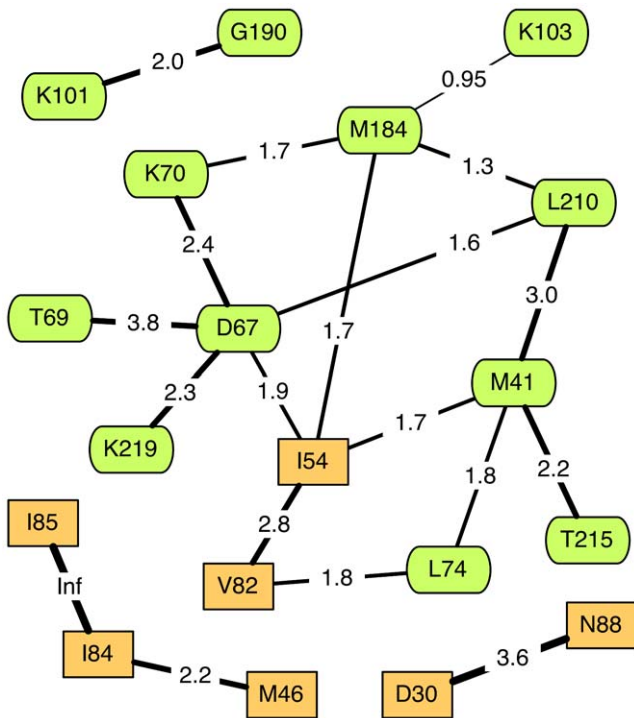
Figure 2. Bayesian network depicting conditional dependencies among surveillance drug resistance mutations (SDRMs) in ARV-naïve sequences. Each labeled node corresponds to a position in protease [rectangular] or RT (rounded). Nodes without dependencies (L23, L24, V32, I47, I50, F53, G73, L76, K65, V75, F77, Y115, F116, L100, V106, V179, Y181, and Y188) were omitted from the graph for clarity. Connections between nodes (edges) are labeled with the $\log_{10}$ odds ratio of the 2×2 contingency table for the presence/absence of SDRMs at the respective sites, *e.g.*, an SDRM at position PR-D30 is estimated to be $10^{3.6} \approx 4000$ times more likely to be present in an ARV-naïve sequence that contains an SDRM at position PR-N88. Line widths for edges are also drawn in proportion to the $\log_{10}$ odds ratios. Inf = infinity, *i.e.*, unable to estimate because of zero count(s) in the contingency table.

doi:10.1371/journal.pone.0021189.g002

Figure 3. CD4 cell count (raw) plotted against the sum of Stanford scores for nucleoside and non-nucleoside reverse transcriptase inhibitors (S_{NRTI} and S_{NNRTI}, respectively). The upper limit of CD4 (2073 cells/mL) was truncated to 1000 cells/mL (omitting 19 outliers) to emphasize the overall trend. Because the predicted baseline CD4 tended to decline with both S_{NRTI} and S_{NNRTI}, we combined the scores into a single ordinal variable to facilitate interpretation. The linear model prediction is displayed as a solid line (generated by fitting a smoothing spline to the predicted values with smoothing parameter $\alpha = 0.9$), with 95% confidence intervals displayed as dashed lines.

doi:10.1371/journal.pone.0021189.g003

about $+2.5$ units $\sqrt[3]{CD4}$ ($P = 0.006$; Figure 5). This model accounted for 6.8% of the variance in $\sqrt[3]{CD4}$.

M184X was significantly associated with an approximately ten-fold lower pVL at baseline ($P = 1.3 \times 10^{-7}$; Table 3). M46X and

L210X were also significantly associated with lower baseline pVL ($P < 0.01$) and D67X was associated with greater pVL ($P = 0.009$). We also observed several significant mutational interaction terms in the model. For instance, interactions between M184X and either K103X or K219X were associated with a compensatory increase in pVL ($P = 0.018$ and $P = 0.023$, respectively; Figure 5). Similarly, a decline in pVL associated with M46X was compensated by T215X ($P = 0.033$). An interaction term between D67 and K219, which was encountered in the preceding model on CD4, was also statistically significant in this model of pVL

Table 2. Coefficient estimates from Akaike Information Criterion (AIC)-selected linear models of baseline $\sqrt[3]{CD4}$ and plasma HIV RNA on Stanford scores by drug class and demographic and risk factors.

	$\sqrt[3]{CD4}$		$\log_{10}$ **plasma HIV RNA**	
	Estimate (95% CI)	*P* **value**	**Estimate (95% CI)**	*P* **value**
S_{NRTI}	$-6.9 \, (-12.2, \, -1.1) \times 10^{-3}$	0.02	$-3.2 \, (-0.8, \, -5.7) \times 10^{-3}$	0.009
S_{NNRTI}	$-4.4 \, (-8.4, \, -0.5) \times 10^{-3}$	0.035	$1.7 \, (0.04, \, 3.4) \times 10^{-3}$	0.045
Male	NS		0.19 (0.06, 0.32)	0.006
Race/ethnicity				
Black	$-0.35 \, (-0.61, \, -0.09)$	0.008	$-0.15 \, (-0.26, 0.04)$	0.007
Hispanic	$-0.57 \, (-0.86, \, -0.28)$	1.1×10^{-4}	NS	

All estimates were averaged over an ensemble of linear models using Akaike weights. Only statistically significant ($\alpha = 0.05$) terms after weighting are reported. Linear effects of factors were estimated relative to the model intercept, *i.e.*, a hypothetical female white individual. NS = not significant.
doi:10.1371/journal.pone.0021189.t002
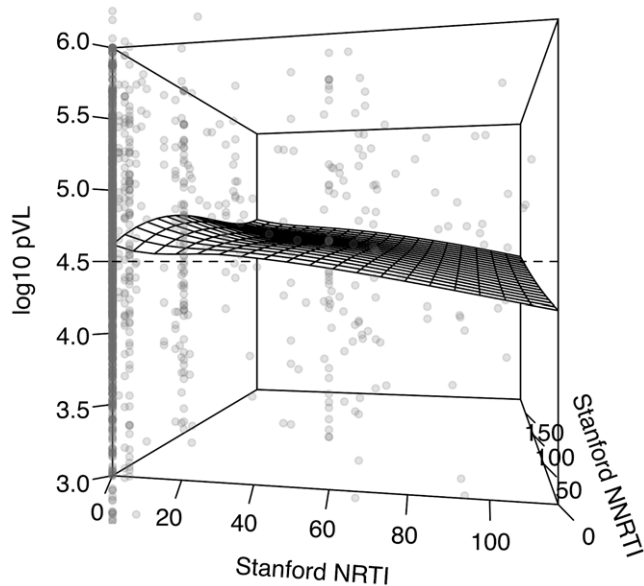
Figure 4. Three-dimensional scatterplot of $\log_{10}$-transformed plasma viral load (pVL) as a function of Stanford scores for NRTI and NNRTIs (S_{NRTI} and S_{NNRTI}). Overall, higher S_{NNRTI} was associated with higher pVL, while S_{NRTI} was inversely associated with pVL. The range of $\log_{10}$ pVL (1.58,6.96) was truncated to emphasize the overall trend (omitting 73 outliers). The linear model prediction is displayed as a wireframe surface (generated by local polynomial regression with smoothing parameter $\alpha = 0.8$).
doi:10.1371/journal.pone.0021189.g004

($P < 0.05$) with the presence of K219X diminishing the effect of D67X on pVL by about 0.6 $\log_{10}$ units. Approximately 6.7% of variation in pVL was explained in this model, over three times as much as the aggregate model.

Causal inference

We used an inverse probability weighting scheme (using stabilized weights) to disentangle the causal effects of SDRMs from confounding variables that influence both the probability of the 'treatment' (an SDRM at the position in question) and CD4 or pVL. Based on the results of the preceding section, we focussed on the compensated and uncompensated effects of M184X on baseline pVL, and of D67X on both pVL and $\sqrt[3]{CD4}$. K103X and K219X were assumed to be interchangeable with respect to their compensatory interactions with M184X. Consequently, the multinomial outcome A_{M184} comprised $n = 1445$, 19 and 15 cases for the respective levels 0, 1 and 2 (i.e., wildtype, uncompensated M184X, and M184X compensated by K103X and/or K219X). In an unweighted univariate linear model, M184X was significantly associated with a -1.1 decline in $\log_{10}$ pVL in the absence of K103X and/or K219X (Table 4). Stabilized weights adjusting for the conditional probability of carrying M184X ranged from 0.02 to 18.72 with a mean of 1.016. Adding stabilized weights to the model exposed a greater deleterious effect of M184X on $\log_{10}$ pVL (-1.69 units) when uncompensated by K103X and/or K219X. Similarly, re-weighting linear models resulted in larger estimated effects of uncompensated D67X on both baseline CD4 (-2.1 units $\sqrt[3]{CD4}$) and pVL (2.5-fold increase; Table 4). Stabilized weights adjusting for the conditional probability of carrying D67X ranged from 0.006 to 12.3 with a mean of 1.003. For both pVL and CD4, the effects of D67X compensated by K219X became statistically significant and reversed sign relative to the unadjusted models (Table 4).

Discussion

The overall prevalence of transmitted drug resistance (TDR) mutations in therapy-naïve patients has been reported in a large number of study populations around the world and ranges from 0 to 25% [1]. On the other hand, the transmission of a specific mutation is a relatively infrequent event whose prevalence requires very large samples (over 1000 individuals) to estimate. This requirement has generally been achieved by integrating study populations from similar geographic regions, such as the Combined Analysis of Resistance Transmission over Time of Chronically and Acute Infected HIV Patients (CATCH) study encompassing 27 countries in Europe [31], the Swiss HIV Cohort Study [32], the UK Collaborative HIV cohort study [14], and the Centers for AIDS Research (CFAR) Network of Integrated Clinical Systems (CNICS) cohort reported in this study. Here we briefly compare and contrast the prevalence of drug class-specific and site-specific TDR across these integrated cohorts. First, NRTI-associated SDRMs generally tend to be more prevalent than SDRMs associated with PIs or NNRTIs. Our sample deviates slightly from this trend in that the prevalence of naïve sequences with ≥ 1 NNRTI-associated SDRM is similar to sequences with ≥ 1 NRTI-associated SDRM, largely due to an atypically high frequency of transmitted mutations at RT-K103 (5.2%) in our sample. Second, SDRMs that are slow to revert in the absence of selection (e.g., M41L, K103N, and T215X in RT and L90M in protease), implying a negligible cost to fitness [1], tend to have relatively high prevalence in therapy-naïve sequences.

The foremost and best-documented clinical implication of TDR is that it impedes the virological response to subsequent drug therapy [6–8]. In contrast, the clinical implications of TDR in the absence of drug therapy are subtle and less well understood. These effects have been difficult to elucidate because the transmission of SDRMs remains an infrequent event at the scale of individual mutations. Investigators have attempted to overcome the limitations of insufficient sample sizes by evaluating the effects of aggregate statistics of TDR, such as an 'all-or-none' statistic that groups sequences carrying one or more SDRMs [14,16,17]. While this approach is convenient and easy to interpret, it is subject to confounding because different drug classes or mutations can have opposite effects on prognostic clinical markers. Our use of a large data set from a multi-center cohort enabled us to overcome some limitations of sample size, and to break down the effects of TDR on baseline CD4 and pVL by drug class or amino acid position. For example, we found that S_{NRTI} and S_{NNRTI} had opposing associations with pVL; because these quantities are correlated, their effects on pVL would have been confounded in an 'all-or-none' analysis. Similarly, we have observed several compensatory interactions among mutations that would otherwise mask the actual effects of individual mutations on CD4 or pVL (e.g., D67N/G and K219Q/E).

Because this study population was comprised of seroprevalent rather than seroincident individuals, each baseline measurement of CD4 and pVL was sampled at a point in time following an unknown date of transmission. Consequently, there is no guarantee that these baseline measurements were taken at a consistent stage of infection. In other words, the CD4 count or pVL at baseline is the result of dynamics over an unknown period of time, starting from an unknown value upon infection. These missing data would be required to calculate rates of change, which would be easier to generalize to specific cases in clinical practice where these quantities (time since infection and prognostic markers at infection) may be known. This is a common problem of observational cohort studies and may explain why a substantial

Table 3. Coefficient estimates from AIC-selected linear models of baseline $\sqrt[3]{CD4}$ and $\log_{10}$ plasma HIV RNA on position-specific SDRMs and demographic and risk factors.

	$\sqrt[3]{CD4}$		$\log_{10}$ plasma HIV RNA	
	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value
M41X	1.0 (0.05, 1.95)	0.04	NS	
M46X	NS		−0.98 (−1.6, −0.35)	0.003
D67X	−2.1 (−3.9, −0.4)	0.026	0.96 (0.25, 1.7)	0.009
M184X	NS		−1.06 (−1.44, −0.67)	1.3×10^{-7}
L210X	NS		−0.6 (−1.0, −0.17)	0.007
K219X	−1.2 (−2.4, −0.08)	0.037	NS	
D67X:K219X	2.5 (0.14, 4.8)	0.006	−0.59 (−1.6, 0.4)	0.047
K219X:G190X	−3.0 (−5.6, −0.4)	0.014	NS	
L90X:M41X	−1.7 (−4.4, 0.1)	0.007	NS	
I54X:L210X	2.0 (−0.59, 4.7)	0.04	NS	
M46X:T215X	NS		1.7 (0.88, 2.5)	0.033
M184X:K103X	NS		0.58 (−0.14, 1.3)	0.018
M184X:K219X	NS		0.87 (−0.07, 1.8)	0.023
K219X:Y181X	NS		1.43 (0.34, 2.5)	0.013
Male	−0.58 (−0.9, −0.24)	7.5×10^{-4}	NS	
Sample date	23.3 (7.4, 38.9) $\times 10^{-5}$	0.004	−20.0 (−38.4, 0.1) $\times 10^{-5}$	0.04
Race/ethnicity				
Black	−0.35 (−0.61, −0.10)	0.007	−0.16 (−0.26, −0.05)	0.003
Hispanic	−0.57 (−0.85, −0.29)	8.3×10^{-5}	NS	
other	−0.53 (−0.98, −0.08)	0.02	NS	

SDRMs are annotated by HXB2 reference residue and position (gene-specific numbering), followed by an 'X' to indicate a non-reference residue. Positions 46, 54, and 90 are in protease and all others are in reverse transcriptase. All estimates were averaged over an ensemble of linear models using Akaike weights. Only terms that were statistically significant ($\alpha = 0.05$) are reported here. NS = not significant.

doi:10.1371/journal.pone.0021189.t003
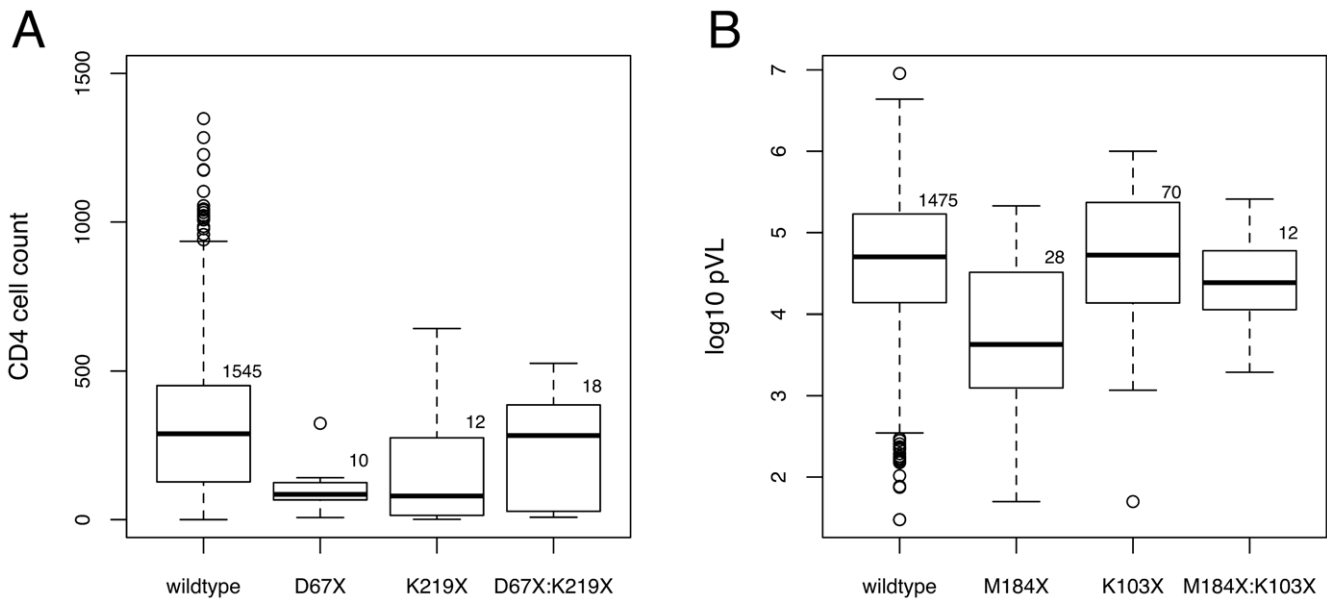
Figure 5. Box-and-whisker plots illustrating (A) the effects of SDRMs at positions D67 and K219 on CD4 cell count and (B) the effects of SDRMs at positions M184 and K103 on plasma viral load (pVL). 'Wildtype' denotes sequences lacking SDRMs at both positions, irrespective of whether any other SDRMs were present at other positions in protease or RT. Solid lines indicate the group median and open circles denote outliers that fall outside the region defined by 1.5 times the interquartile range. Plot (A) was rescaled with respect to untransformed CD4 counts to emphasize differences among groups, trimming 4 outliers from the 'wildtype' group (1602, 1626, 1838, and 2073 cells/mL) from the plot region. Sample sizes per group are annotated on each plot.

doi:10.1371/journal.pone.0021189.g005

Table 4. Unadjusted and stabilized weight-adjusted linear models of multinomial exposure variables A_{M184} (M184X and (K103X or K219X)) and A_{D67} (D67X and K219X) on $\log_{10}$ plasma HIV RNA and $\sqrt[3]{\text{CD4}}$, respectively.

Level	Unadjusted model		Adjusted model	
	Estimate (95% C.I.)	P	Estimate (95% C.I.)	P
$\log_{10}$ plasma HIV RNA $\sim A_{\text{M184}}$				
$A_{\text{M184}} = 0$	4.63 (4.58,4.67)			
$A_{\text{M184}} = 1$	$-$**1.10** ($-$1.57,$-$0.63)	4.4×10^{-6}	$-$**1.69** ($-$2.63,$-$0.74)	4.85×10^{-4}
$A_{\text{M184}} = 2$	$-$0.236 ($-$0.568,0.096)	0.16	0.207 ($-$0.35,0.763)	0.47
$\sqrt[3]{\text{CD4}} \sim A_{\text{D67}}$				
$A_{\text{D67}} = 0$	6.22 (6.12,6.33)			
$A_{\text{D67}} = 1$	$-$**1.77** ($-$2.8,$-$0.74)	7.97×10^{-4}	$-$**2.14** ($-$2.66,$-$1.63)	7.5×10^{-16}
$A_{\text{D67}} = 2$	$-$0.768 ($-$1.93,0.40)	0.196	**0.51** (0.07,0.89)	0.016
$\log_{10}$ plasma HIV RNA $\sim A_{\text{D67}}$				
$A_{\text{D67}} = 0$	4.61 (4.56,4.65)			
$A_{\text{D67}} = 1$	**0.41** (0.21,0.62)	9.0×10^{-5}	**0.40** (0.32,0.48)	$<2 \times 10^{-16}$
$A_{\text{D67}} = 2$	$-$0.21 ($-$0.64,0.22)	0.34	**0.23** (0.05,0.41)	0.011

The model intercepts are taken to be $A = 0$ by default, *i.e.*, none of the designated mutations are present; these estimates are unchanged in the adjusted model and are not repeated for clarity of presentation. Significant ($\alpha = 0.05$) effect estimates are bolded.

doi:10.1371/journal.pone.0021189.t004

fraction ($\approx 90\%$) of variation in baseline CD4 or pVL remained unaccounted for in the linear models. Furthermore, because the baseline samples were taken either upon entry into care or prior to initiating ARV therapy (ART), they may constitute a biased sample of the patient population with respect to CD4 and pVL. Recent guidelines, for example, indicate that patients should be initiated on ART when their CD4 cell count falls below a threshold of 500 cells per mL regardless of HIV RNA [33] However, treatment guidelines have evolved over time [34] and using a large sample size covering a decade of clinical practice can ameliorate some of these biases. As guidelines have pushed initiation of therapy to earlier time points, it could be argued that less fit virus, with lower HIV RNA and higher CD4, would not merit clinical significance since virtually every patient should start therapy soon after identification; in other words, an association of TDR with lower viral load would not result in a longer delay between detection of HIV and the need to start therapy. However, viral fitness and replication level may be very significant in the prevalence of resistance mutations in the circulating virus population, the pool of virus in a community that could potentially be transmitted. As the pool of patients with suppressed virus increases due to ARV treatment, the relative rates of transmitted resistance may paradoxically increase due to a diminishing pool of untreated wildtype virus and a greater prevalence of resistance in the circulating virus population (10%–20% of patients who are failing therapy [1]). In this scenario, viral fitness and its effect on HIV RNA could determine which viruses are transmitted. Mutations such as M184V may be transmitted less often due to the reduced overall viral load in those subjects, whereas mutations such as K103N may increase in prevalence. THe overall rates of transmitted resistance would be a complex interplay between many host and viral factors.

Some of these issues might be resolved by the analysis of longitudinal data that are available for this study population [18]. However, this approach would create even more problems than it could potentially address. First, we would have to accommodate the effects of antiretroviral therapy that generally follows soon after HIV genotyping at baseline. This is not a trivial task, not only

because it introduces a potentially large number of variables to the analysis (one for every drug, or a minimum of three for the predominant drug classes), but also because variation in adherence and confounding due to variation in prescribed regimens across clinical sites would severely complicate the interpretation of results. Second, we would have to accommodate large amounts of missing data due to differential follow-up among patients as well as a sparsity of HIV sequence data relative to CD4 and pVL measurements. In other words, HIV genotyping is performed much less frequently. Third, it is more difficult to adjust for confounding in longitudinal than cross-sectional data sets by inverse probability weighting. Previous studies have successfully employed a marginal structural modelling approach in similar but simplified contexts [20,35]. Marginal structural models are essentially a time-dependent generalization of the inverse probability weighting approach used here [19]. However, they are difficult to apply in highly-structured contexts (where pVL, CD4, drug regimen and HIV genotype all influence one another over time) and large numbers of missing data. In sum, the analysis of longitudinal data is a substantially different question outside the scope of this study on the effect of TDR on pre-therapy pVL and CD4 and must be left to future work.

Of all position-specific effects of SDRMs on baseline prognostic markers inferred from these data, one of the largest and most statistically significant was M184V/I. The effects of these substitutions on the kinetics and fidelity of HIV-1 reverse transcriptase and replication capacity *in vitro* are well-characterized [36–38]. Previous studies have also documented compensatory/epistatic interactions of other mutations acting on M184V/I, including K219Q [11] and N384I [39]. Furthermore, Paredes and colleagues have recently observed that *in vivo* fitness is reduced in viruses carrying the M184V mutation in the absence of lamivudine [12]. Even so, the effect of M184V/I on viral fitness *in vivo* is confounded by the effects of other SDRMs and demographic/risk factors that should be handled using a causal modeling approach [19]. Taking such an approach, we have found that uncompensated M184V/I mutations are causally associated with a 50-fold reduction ($10^{-1.7}$) in baseline pVL in therapy-naïve patients.

Previous studies have reported no greater than a 3-fold $(10^{-0.5})$ reduction in pVL due to M184V/I [38,40,41], while M184I itself has been estimated to reduce fitness by 23% relative to M184V [42]. This discrepancy may be due to failing to account for interactions with other mutations such as the compensatory interactions with K103 and/or K219, or confounding by other genetic and non-genetic factors affecting pVL. On the other hand, experimental studies of M184V/I-containing recombinant HIV-1 are potentially not subject to these confounding factors, and M184V has been reported to reduce fitness by as much as 16-fold *in vitro* [11]. However, it is difficult to estimate the fitness effect of a mutation *in vivo* from *in vitro* measures of its effect on replication efficiency as there are other components of RT functionality that can be affected [43].

In contrast, the clinical consequences of D67N/G and K219Q/E are not as well known. The effects of D67N/G and K219Q/E on CD4 in our study were mirrored by significant effects on pVL; *i.e.*, substitutions at D67 were causally associated with a 2.5-fold increase in pVL that was only partially compensated by

substitutions at K219. These residues potentially form a salt bridge and likely affect the formation of the RT 3′ pocket during polymerization [44]. Substitutions at these sites have previously been suggested to incur a moderate fitness cost *in vitro* [11,45], although these previous studies had not employed statistical methods to separate out the effects of these substitutions from confounding by other mutations and demographic or risk factors . Our results give the first evidence that the transmission of D67N/G may have deleterious virological and immunological consequences in therapy-naïve patients that may be compensated by K219Q/E.

Author Contributions

Conceived and designed the experiments: RHH SDWF AFYP JLA. Performed the experiments: AFYP. Analyzed the data: AFYP. Contributed reagents/materials/analysis tools: WCM MK JSK MSS BR SLB. Wrote the paper: AFYP.

References

1. Booth CL, Geretti AM (2007) Prevalence and determinants of transmitted antiretroviral drug resistance in HIV-1 infection. J Antimicrob Chemother 59: 1047–56.
2. Wheeler WH, Ziebell RA, Zabina H, Pieniazek D, Prejean J, et al. (2010) Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.-2006. AIDS 24: 1203–12.
3. Pao D, Andrady U, Clarke J, Dean G, Drake S, et al. (2004) Long-term persistence of primary genotypic resistance after HIV-1 seroconversion. J Acquir Immune Defic Syndr 37: 1570–3.
4. Little SJ, Frost SDW, Wong JK, Smith DM, Kosakovsky Pond SL, et al. (2008) Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. J Virol 82: 5510–5518.
5. Halvas EK, Wiegand A, Boltz VF, Kearney M, Nissley D, et al. (2010) Low frequency nonnucleoside reverse-transcriptase inhibitor-resistant variants contribute to failure of efavirenz-containing regimens in treatment- experienced patients. J Infect Dis 201: 672–80.
6. Johnson JA, Li JF, Wei X, Lipscomb J, Irlbeck D, et al. (2008) Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. PLoS Med 5: e158.
7. Simen BB, Simons JF, Hullsiek KH, Novak RM, Macarthur RD, et al. (2009) Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis 199: 693–701.
8. Lecossier D, Shulman NS, Morand-Joubert L, Shafer RW, Joly V, et al. (2005) Detection of minority populations of HIV-1 expressing the K103N resistance mutation in patients failing nevirapine. J Acquir Immune Defic Syndr 38: 37–42.
9. Mammano F, Trouplin V, Zennou V, Clavel F (2000) Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. J Virol 74: 8524–31.
10. Collins JA, Thompson MG, Paintsil E, Ricketts M, Gedzior J, et al. (2004) Competitive fitness of nevirapine-resistant human immunodeficiency virus type 1 mutants. J Virol 78: 603–11.
11. Cong M, Heneine W, García-Lerma JG (2007) The fitness cost of mutations associated with human immunodeficiency virus type 1 drug resistance is modulated by mutational interactions. J Virol 81.
12. Paredes R, Sagar M, Marconi VC, Hoh R, Martin JN, et al. (2009) In vivo fitness cost of the M184V mutation in multidrug-resistant human immunodeficiency virus type 1 in the absence of lamivudine. J Virol 83: 2038–43.
13. Balotta C, Berlusconi A, Pan A, Violin M, Riva C, et al. (2000) Prevalence of transmitted nucleoside analogue-resistant HIV-1 strains and pre-existing mutations in pol reverse transcriptase and protease region: outcome after treatment in recently-infected individuals. Antivir Ther 5: 7–14.
14. Grant RM, Hecht FM, Warmerdam M, Liu L, Liegler T, et al. (2002) Time trends in primary HIV-1 drug resistance among recently infected persons. JAMA 288: 181–8.
15. Shet A, Leslie B, Hiroshi M, Mehandru S, Chung C, et al. (2006) Tracking the prevalence of transmitted antiretroviral drug-resistant HIV-1: a decade of experience. JAIDS 41: 439–446.
16. Bezemer D, de Ronde A, Prins M, Porter K, Gifford R, et al. (2006) Evolution of transmitted HIV-1 with drug-resistance mutations in the absence of therapy: effects on CD4+ T-cell count and HIV-1 RNA load. Antivir Ther 11: 173–8.
17. Pillay D, Bhaskaran K, Jurriaans S, Prins M, Masquelier B, et al. (2006) The impact of transmitted drug resistance on the natural history of HIV infection and response to first-line therapy. AIDS 20: 21–8.
18. Kitahata MM, Rodriguez B, Haubrich R, Boswell S, Mathews WC, et al. (2008) Cohort profile: the Centers for AIDS Research Network of Integrated Clinical Systems. Int J Epidemiol 37: 948–55.
19. Robins JM, Hernán MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11: 550–60.
20. Hogg RS, Bangsberg DR, Lima VD, Alexander C, Bonner S, et al. (2006) Emergence of drug resistance is associated with an increased risk of death among patients first starting HAART. PLoS Med 3: e356.
21. Taylor JM, Wang Y, Ahdieh L, Chmiel JS, Detels R, et al. (2000) Causal pathways for CCR5 genotype and HIV progression. J Acquir Immune Defic Syndr 23: 160–71.
22. Gotoh O (1982) An improved algorithm for matching biological sequences. J Mol Biol 162: 705–708.
23. Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676–679.
24. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, et al. (2009) Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PLoS One 4: e4724.
25. Liu TF, Shafer RW (2006) Web resources for HIV type 1 genotypic-resistance test interpretation. Clin Infect Dis 42: 1608–18.
26. Venables WN, Ripley BD Modern Applied Statistics with S. Springer.
27. Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociol Method Res 33: 261–304.
28. Lumley T (2004) Analysis of Complex Survey Samples. Journal of Statistical Software 9: 1–19.
29. Poon AFY, Kosakovsky Pond SL, Richman DD, Frost SDW (2007) Mapping protease inhibitor resistance to human immunodeficiency virus type 1 sequence polymorphisms within patients. J Virol 81: 13598–607.
30. Pearl J (1988) Probabilistic reasoning in intelligent systems: Networks of plausible inference. San MateoCalif.: Morgan Kaufmann Publishers. 552 p.
31. Wensing AMJ, van de Vijver DA, Angarano G, Asjö B, Balotta C, et al. (2005) Prevalence of drugresistant HIV-1 variants in untreated individuals in Europe: implications for clinical management. J Infect Dis 192: 958–66.
32. Yerly S, von Wyl V, Ledergerber B, Böni J, Schüpbach J, et al. (2007) Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. AIDS 21: 2223–9.
33. Thompson MA, Aberg JA, Cahn P, Montaner JSG, Rizzardini G, et al. (2010) Antiretroviral treatment of adult HIV infection: 2010 recommendations of the International AIDS Society-USA panel. JAMA 304: 321–33.
34. Wilkin TJ, Gulick RM (2008) When to start antiretroviral therapy? Clin Infect Dis 47: 1580–6.
35. Hernán MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 11: 561–70.
36. Wainberg MA, Drosopoulos WC, Salomon H, Hsu M, Borkow G, et al. (1996) Enhanced fidelity of 3TC-selected mutant HIV-1 reverse transcriptase. Science 271: 1282–5.
37. Götte M, Arion D, Parniak MA, Wainberg MA (2000) The M184V mutation in the reverse transcriptase of human immunodeficiency virus type 1 impairs rescue of chain-terminated DNA synthesis. J Virol 74: 3579–85.
38. Devereux HL, Emery VC, Johnson MA, Loveday C (2001) Replicative fitness in vivo of HIV-1 variants with multiple drug resistance-associated mutations. J Med Virol 65: 218–24.
39. von Wyl V, Ehteshami M, Symons J, Bürgisser P, Nijhuis M, et al. (2010) Epidemiological and biological evidence for a compensatory effect of connection

domain mutation N348I on M184V in HIV-1 reverse transcriptase. J Infect Dis 201: 1054–62.

40. Castagna A, Danise A, Menzo S, Galli L, Gianotti N, et al. (2006) Lamivudine monotherapy in HIV-1-infected patients harbouring a lamivudine-resistant virus: a randomized pilot study (E-184V study). AIDS 20: 795–803.

41. Harrison L, Castro H, Cane P, Pillay D, Booth C, et al. (2010) The effect of transmitted HIV-1 drug resistance on pre-therapy viral load. AIDS 24: 1917–22.

42. Frost SD, Nijhuis M, Schuurman R, Boucher CA, Brown AJ (2000) Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. J Virol 74: 6262–8.

43. Wei X, Liang C, Götte M, Wainberg MA (2003) Negative effect of the M184V mutation in HIV-1 reverse transcriptase on initiation of viral DNA synthesis. Virology 311: 202–12.

44. Huang H, Chopra R, Verdine GL, Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science 282: 1669–75.

45. García-Lerma JG, MacInnes H, Bennett D, Weinstock H, Heneine W (2004) Transmitted human immunodeficiency virus type 1 carrying the D67N or K219Q/E mutation evolves rapidly to zidovudine resistance in vitro and shows a high replicative fitness in the presence of zidovudine. J Virol 78: 7545–52.