PLOS | COMPUTATIONAL BIOLOGY

The Individualized Genetic Barrier Predicts Treatment Response in a Large Cohort of HIV-1 Infected Patients

Niko Beerenwinkel[1,2]*, Hesam Montazeri[1,2], Heike Schuhmacher[1,2], Patrick Knupfer[1], Viktor von Wyl[3], Hansjakob Furrer[4], Manuel Battegay[5], Bernard Hirschel[6], Matthias Cavassini[7], Pietro Vernazza[8], Enos Bernasconi[9], Sabine Yerly[10], Jürg Böni[11], Thomas Klimkait[12], Cristina Cellerai[13], Huldrych F. Günthard[3], The Swiss HIV Cohort Study[¶]

1 Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland, 2 SIB Swiss Institute of Bioinformatics, Basel, Switzerland, 3 Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland, 4 Clinic for Infectious Diseases, Bern University Hospital, Bern, Switzerland, 5 Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel, Switzerland, 6 Division of Infectious Diseases, Geneva University Hospital, Geneva, Switzerland, 7 Division of Infectious Diseases, University Hospital Lausanne, Lausanne, Switzerland, 8 Division of Infectious Diseases, Cantonal Hospital St. Gallen, St. Gallen, Switzerland, 9 Division of Infectious Diseases, Regional Hospital Lugano, Lugano, Switzerland, 10 Laboratory of Virology, University Hospital Geneva, Geneva, Switzerland, 11 Swiss National Center for Retroviruses, Institute of Medical Virology, University of Zurich, Zurich, Switzerland, 12 Institute for Medical Microbiology, University of Basel, Basel, Switzerland, 13 Division of Immunology and Allergy, Centre Hospitalier Universitaire Vadois, Lausanne, Switzerland

Abstract

The success of combination antiretroviral therapy is limited by the evolutionary escape dynamics of HIV-1. We used Isotonic Conjunctive Bayesian Networks (I-CBNs), a class of probabilistic graphical models, to describe this process. We employed partial order constraints among viral resistance mutations, which give rise to a limited set of mutational pathways, and we modeled phenotypic drug resistance as monotonically increasing along any escape pathway. Using this model, the individualized genetic barrier (IGB) to each drug is derived as the probability of the virus not acquiring additional mutations that confer resistance. Drug-specific IGBs were combined to obtain the IGB to an entire regimen, which quantifies the virus' genetic potential for developing drug resistance under combination therapy. The IGB was tested as a predictor of therapeutic outcome using between 2,185 and 2,631 treatment change episodes of subtype B infected patients from the Swiss HIV Cohort Study Database, a large observational cohort. Using logistic regression, significant univariate predictors included most of the 18 drugs and single-drug IGBs, the IGB to the entire regimen, the expert rules-based genotypic susceptibility score (GSS), several individual mutations, and the peak viral load before treatment change. In the multivariate analysis, the only genotype-derived variables that remained significantly associated with virological success were GSS and, with 10-fold stronger association, IGB to regimen. When predicting suppression of viral load below 400 cps/ml, IGB outperformed GSS and also improved GSS-containing predictors significantly, but the difference was not significant for suppression below 50 cps/ml. Thus, the IGB to regimen is a novel data-derived predictor of treatment outcome that has potential to improve the interpretation of genotypic drug resistance tests.

Author Summary

Drug resistance remains a challenge in the management of HIV-infected patients. The accumulation of mutations during ongoing viral replication is the origin of drug resistance development. Understanding this evolutionary process in a quantitative manner is an important prerequisite for minimizing the risk of resistance development and for the optimal selection of drug combinations for each individual patient. We present probabilistic graphical models for describing the evolution of drug resistance, and we derive the individualized genetic barrier (IGB), a single quantity summarizing the genetic potential of the virus for evolutionary escape from selective drug pressure. The predictive power of the IGB is demonstrated on a large well characterized clinical cohort of HIV patients and compared to classical predictors.

Introduction

Despite an increasing arsenal and improved potency of antiretroviral drugs, the optimal use of combination antiretroviral therapy against HIV-1 infection remains challenging [1]. Complicating factors include drug interactions and toxicities, adherence to therapy, and development of drug resistance [2]. Because genotypic drug resistance testing is performed on a routine basis today and because mutational patterns are unique for each patient, treatment choices are, in principle, highly personalized. In practice, however, it can be difficult to identify an optimal drug combination for each individual patient due to the combinatorial complexity of both the set of feasible drug combinations and of viral mutational patterns.

In addition to controlled clinical trials, analyzing data from large observational cohort studies is a promising way to identify predictors of treatment outcome, even if the availability of drugs and therapeutic strategies change over time [3]. This approach can be based on modeling the risk of acquiring additional mutations [4], on estimating future drug options [5], on predicting the time to virological failure [6,7], or on classifying the regimens of treatment change episodes (TCEs) as successful versus failing, depending on the patient's response to therapy. A TCE consists of predictor variables including the applied drug combination, viral genotype, treatment history, demographic and clinical parameters, and a response variable such as the change in viral load.

HIV-1 genotype has been shown to be a strong predictor of therapeutic success in retrospective and prospective studies [8–14], but the large number of mutations complicates prediction. TCE classification is a noisy, high-dimensional prediction problem with unobserved confounding factors and sparse data. It has been addressed by several statistical learning methods [15–25]. Comparative studies have emphasized the importance of selection and representation of features, especially of the viral genotype, over the choice of the learning algorithm [26–28]. In order to directly correlate genotype with clinical response, rules-based approaches, such as the genotypic susceptibility score (GSS) [29–34] and statistical models [23,26,28] have been proposed, often outcompeting human experts [35].

Drug resistance development is driven by viral evolution and thus models of viral evolutionary escape from drug pressure have been proposed to improve therapy response prediction [16,22,36]. Specifically, the individualized genetic barrier (IGB) to drug resistance has been suggested as a predictor of treatment outcome. The IGB is defined as the probability of the virus not to become resistant to a certain drug [37–39]. A high IGB means that viral

evolutionary escape from the selective pressure of the drug is unlikely. Related quantities are the average number of mutations and the average time to reach drug resistance derived from simulated HIV-1 evolutionary trajectories on an estimated fitness landscape [36,40,41]. This approach has been explored for treatment with zidovudine plus lamivudine and with nelfinavir [42], but it does not scale to the variety of combination therapies observed in clinical databases, because sufficient data for estimating fitness landscapes is available only for a few drug combinations. Earlier, the term 'calculated genetic barrier' has been used to assess the number of mutations necessary to acquire specific drug resistance-associated mutations, which were found to be similar among HIV-1 subtypes [43].

In the present study, we apply a simplified definition of the IGB which can be computed efficiently for any drug combination based on a statistical model that captures the order and the dynamics of accumulating mutations and the associated levels of phenotypic drug resistance [44]. The IGB to resistance to a certain drug is the probability that the virus will not accumulate additional mutations leading to a resistant strain. This drug-specific IGB has been demonstrated to be a strong predictor of virological response in two large observational cohort studies [26,28]. Here, we derive a novel predictor, the IGB to the entire drug combination which measures the genetic potential for evolutionary escape of the virus from the selective pressure of combination therapy.

In order to assess the performance of the IGB as a predictor of treatment outcome, we analyzed TCE data from the Swiss HIV Cohort Study (SHCS) database, a large, long-term observational, multi-center, clinical database with integrated results of genotypic drug resistance tests [45,46]. We identified risk factors of therapeutic failure and constructed models of treatment outcome considering as predictors the applied regimen, treatment history, viral genotype, GSS, drug-specific IGBs, IGB to regimen, and demographic and clinical variables including patient adherence. Overall, we found the IGB to the entire regimen to be the strongest and most significant predictor. Our results demonstrate that the viral genotype is represented efficiently by the IGB to regimen, a single, interpretable probability summarizing the predicted dynamics of viral evolutionary escape.

Results

For each drug, viral evolutionary escape from its selective pressure was modeled using Isotonic Conjunctive Bayesian Networks (I-CBNs). In these probabilistic graphical models, dependencies among mutations are described by a partial order, which defines the genotype lattice, i.e., the set of genotypes compatible with the order constraints, and hence the set of possible mutational escape pathways (Figure 1). To each genotype, its level of phenotypic drug resistance is associated using isotonic regression, such that drug resistance is monotonically non-decreasing along any mutational pathway from the wild type towards the genotype carrying all mutations. Using cross-sectional matched genotype-phenotype pairs from the Stanford HIV Drug Resistance Database, I-CBN models were learned for a total of 18 antiretroviral drugs (Supporting Figures S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, Supporting Table S2). Each model includes up to eleven pre-selected mutations (see Methods).

From the I-CBN models, transition probabilities among genotypes were derived and the individualized genetic barrier (IGB) to resistance development to each drug was computed as the probability of the observed genotype not acquiring additional mutations that would transform it into a genotypic state predicted
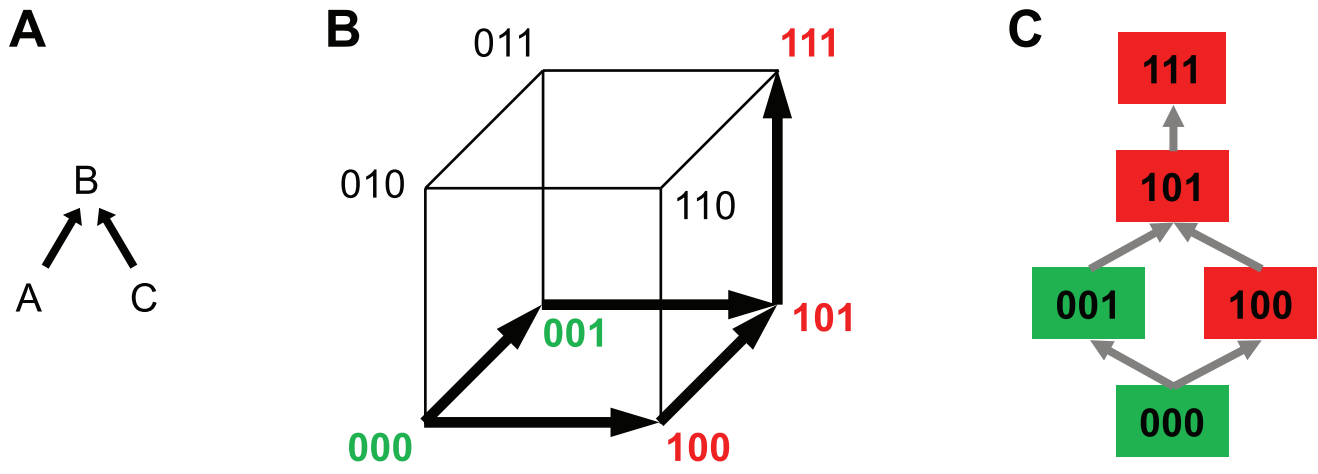
Figure 1. Schematic illustration of I-CBN model and individualized genetic barrier (IGB). (A) A partially ordered set of three mutations, A, B, and C, is considered with the two relations $A \prec B$ and $C \prec B$, resulting in two possible escape pathways of the virus, namely $A \rightarrow C \rightarrow B$ or $C \rightarrow A \rightarrow B$. (B) The partial order constraints give rise to the genotype lattice consisting of genotypes 000, 001, 100, 101, and 111 indicated with bold arrows, where genotypes are encoded as binary strings such that 000 is the wild type $\varnothing$ (no mutations), 100 is defined by mutation A and identified with $\{A\}$, 101 with $\{A,C\}$, etc. The genotype lattice $\mathcal{G}$ is shown inside the embedding hypercube $\{0,1\}^3 \simeq 2^{\{A,B,C\}}$. For each antiretroviral drug, genotypes are labeled as either susceptible (green) or resistant (red). (C) Genotype lattice isolated from the embedding hypercube. The IGB is the probability of the virus not reaching a resistant state.
doi:10.1371/journal.pcbi.1003203.g001

to be resistant. For a drug combination, the IGB was obtained as the sum over all drugs of the regimen of the drug-specific IGBs. Thus, the IGB to regimen can be regarded as the expected number of active components in the drug cocktail taking viral evolutionary escape mechanisms into account. To assess the predictive power of the IGB in a clinical setting, we analyzed a large cohort of HIV-1-infected patients and compared the IGB to several known predictors of therapy response (Figure 2), including the GSS, obtained from the Stanford HIV Drug Resistance Database website (HIVdb 6.2.0).

TCEs from the time period 1988–2010 were derived from the SHCS database (Table 1 and 2) and labeled as either failure or success (see Methods). Therapy success was defined as viral load reduction below 50 cps/ml (400 cps/ml) during treatment. We obtained 2185 (2631) genotype-therapy pairs, including 73% (63%) failures. The usage of individual drugs and the 30 most frequent drug combinations are shown in Supporting Figures S2

and S3, respectively. The historical development of drug usage patterns is reported in Supporting Table S3, where the regimens are annotated as either being recommended as first-line or alternative regimens according to current treatment guidelines [47], or as past first-line or second-line recommended regimens that are still in use in developing countries or occasionally used if drug resistant virus is present at baseline or as salvage regimens, or as regimens that are not in use anymore as first-line regimens but were before, including those still used under special circumstances, such as unusual tolerability.

In order to predict the outcome (failure versus success) of each therapy, we considered applied drugs, demographic and clinical variables, viral genotype, IGBs to received drugs, and IGB to regimen (Figure 2, Table S1). Univariate logistic regression resulted in a total of 50 (44) features that were significantly associated with therapy outcome (Figure S22). Among the predictive drugs, the use of ZDV, d4T, 3TC, and NFV were
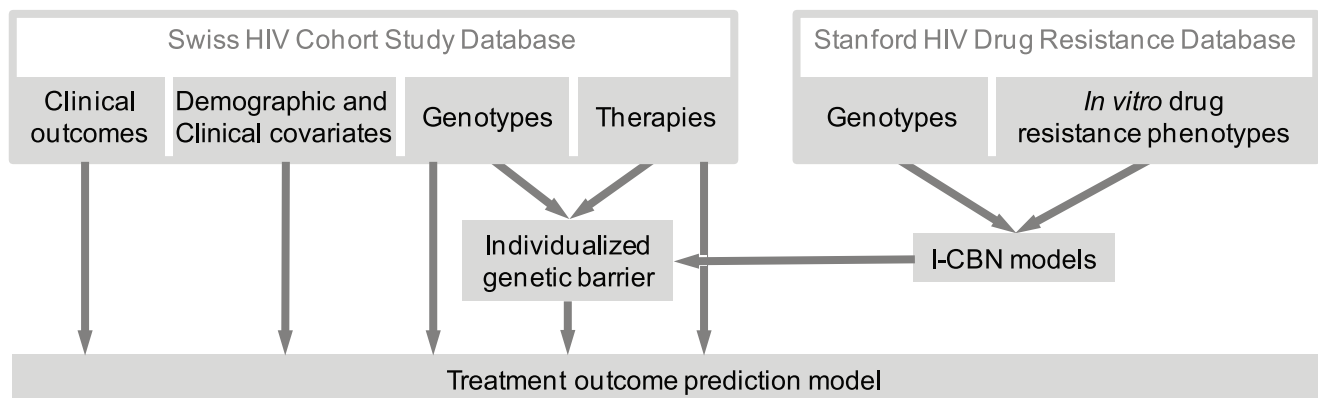
Figure 2. Data flow. Matched pairs of viral genotype and drug resistance phenotype from the Stanford HIV Drug Resistance Database (top right) were used to learn I-CBN models for all drugs separately. The drug-specific individualized genetic barriers (IGBs) are derived from these models. The IGB to regimen is computed for each genotype-therapy pair in the Swiss HIV Cohort Database and its predictive power is assessed in prediction models that also account for classical demographic, clinical, and genetic covariates.
doi:10.1371/journal.pcbi.1003203.g002

Table 1. Characteristics of the numerical predictors in the SHCS database.

Numerical variables	50 cps/ml		400 cps/ml	
	median	(IQR)	median	(IQR)
Age	40	(35–46)	40	(35–46)
Minimum CD4 T cell count (cells/mm^3)	108	(40–200)	110	(40–206)
Maximum viral load (log10 copies/ml)	5.17	(4.72–5.63)	5.15	(4.66–5.61)

doi:10.1371/journal.pcbi.1003203.t001

associated with increased risk of therapeutic failure, while ABC, TDF, FTC, EFV, RTV, LPV/r, ATV, and ATV/r increased the odds of therapeutic success. Most of the significant amino acid changes in the viral protease (PR) gene (10I, 30N, 33F, 46I, 54V, 71V, 82A, 84V, 90M) and reverse transcriptase (RT) gene (39A, 41L, 44D, 67N, 74V, 103N, 118I, 123S, 210W, 215Y, 297R) have been associated with resistance to multiple PR inhibitors (PIs) and RT inhibitors (RTIs), respectively, and all except PR 30N and RT 123S increased the risk of treatment failure. A higher IGB to any of 15 (16) individual drugs increased the chance of successful virological response. The IGB to the entire drug combination and the GSS were also significant predictors.

In the multivariate analysis, only 12 (14) variables were significant, nine (ten) of which are indicating the inclusion of individual drugs in the regimen (Figure 3). The usage of the nucleoside RTIs (NRTIs) ZDV, ddI, d4T, and 3TC, and of the PIs APV and SQV, were associated with negative treatment outcome, whereas the four boosted PIs (i.e., given together with low-dose RTV to improve their bioavailability) SQV/r, IDV/r, LPV/r, and ATV/r had positive predictive power. Among the

many genotype-derived predictors, only GSS and IGB to regimen reached statistical significance at the 1% level in the multivariate model. For the 50 cps/ml success definition, the odds ratio (OR) of therapeutic success was ten-fold higher for the IGB (OR 23.6, 95% confidence interval [CI] 12.21–45.4, $p < 10^{-19}$) as compared to the GSS (OR 2.1, 95% CI 1.6–2.7, $p < 10^{-7}$), and similarly for 400 cps/ml (IGB OR 25.0, 95% CI 14.7–42.5, $p < 10^{-30}$ versus GSS OR 1.8, 95% CI 1.5–2.2, $p < 10^{-7}$), indicating that the IGB provides an effective summary of the risk of treatment failure due to viral genetic changes. In addition, increased overall maximum (peak) viral load before treatment remained a significant predictor of therapy outcome in the multivariate logistic regression model.

For optimal treatment outcome prediction, we also explored the use of regularized logistic regression models. Specifically, the elastic net, which combines L_1 and L_2 regularization was applied to identify sparse classifiers of therapy outcome. Classifier performance was evaluated in ROC curves summarized by the area under the ROC curve (AUC), and analyzed according to the historical drug usage patterns (Table S3). The competitive models (high AUC) are only those using all clinical and demographic

Table 2. Characteristics of the categorical predictors in the SHCS database.

Categorical variables		50 cps/ml		400 cps/ml	
		frequency	(%)	frequency	(%)
Gender	female	562	(25.72%)	705	(26.8%)
	male	1623	(74.28%)	1926	(73.2%)
AIDS	no	1461	(66.86%)	1775	(67.46%)
	yes	724	(33.14%)	856	(32.54%)
Transmission group	blood	16	(0.73%)	27	(1.03%)
	heterosexual	719	(32.91%)	881	(33.49%)
	IDU	491	(22.47%)	598	(22.73%)
	male homosexual	879	(40.23%)	1033	(39.26%)
	mother-to-child	12	(0.55%)	16	(0.61%)
	others/unknown	68	(3.12%)	76	(2.88%)
Ethnic Group		9	(0.41%)	10	(0.38%)
	asian	41	(1.88%)	58	(2.2%)
	black	281	(12.86%)	347	(13.19%)
	hispano american	34	(1.56%)	46	(1.75%)
	white	1743	(79.77%)	2080	(79.06%)
	unknown	77	(3.52%)	90	(3.42%)
Adherence to treatment	low	496	(22.7%)	610	(23.19%)
	high	1586	(72.59%)	1899	(72.18%)
	others/unknown	103	(4.71%)	122	(4.64%)

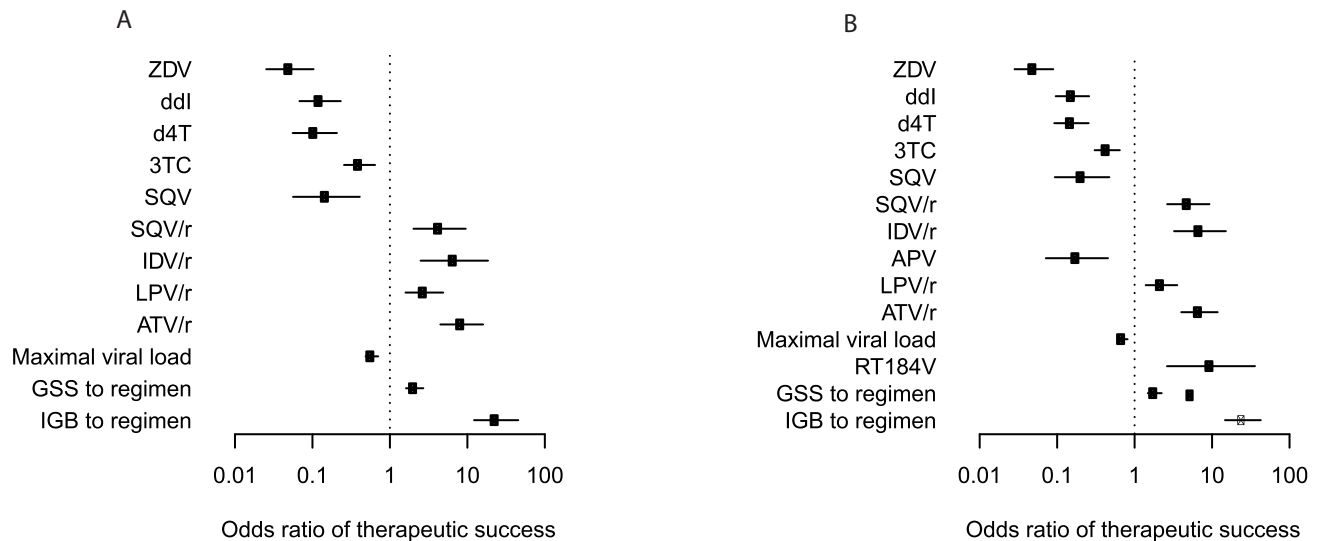doi:10.1371/journal.pcbi.1003203.t002

Figure 3. Multivariate analysis of predictors of response to antiretroviral combination therapy in the SHCS database. Associations have been tested using a logistic regression model and odds ratios of therapeutic success, defined as viral load reduction below 50 cps/ml (A) and 400 cps/ml (B), are reported together with their 95% confidence intervals on a logarithmic scale. Benjamini-Hochberg-corrected p-values are represented as black ($p<0.001$) and grey ($p<0.01$) symbols. Only predictors with a p-value smaller than 0.01 are included.
doi:10.1371/journal.pcbi.1003203.g003

variables, mutations, and drugs (Tables S5, S6, Figure 4). When comparing IGB to GSS as predictors in this setting, we found a significant advantage of the IGB for 400 cps/ml if all other features are included in the models ($p=0.01$, Wilcoxon rank sum test). Furthermore, the IGB also improves treatment outcome prediction if added to models that already contain the GSS ($p=0.0002$). For 50 cps/ml, we did not find significant differences

in AUC between IGB and GSS when used in prediction models that included all other covariates, nor did the GSS-containing model improve upon adding IGB. The significant increase for the larger dataset with the 400 cps/ml success definition demonstrates the predictive power of the IGB and indicates that GSS and IGB, although correlated, contain some orthogonal information, which, if combined, can further improve treatment outcome prediction.
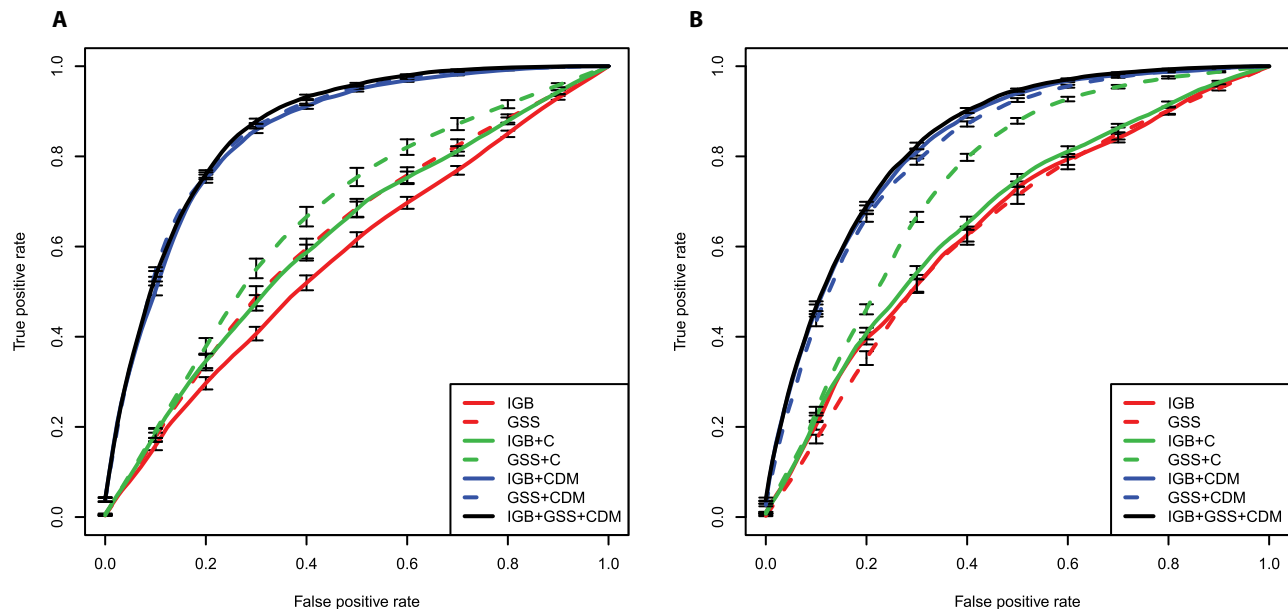
Figure 4. ROC curves quantifying the performance of elastic net regularized logistic regression models in predicting treatment outcome, defined as a reduction of viral load below 50 cps/ml (A) and 400 cps/ml (B). The areas under the ROC curves (AUC values) are reported in Table S5 and Table S6. Prediction models are encoded by the sets of predictors used, where C refers to the demographic and clinical variables, D refers to drugs, and M to mutations. For example, the model IGB+CDM includes as predictors IGB to regimen, clinical and demographic predictors, applied drugs, and mutations. The models with all predictors perform significantly better than all other models.
doi:10.1371/journal.pcbi.1003203.g004

Discussion

We have comprehensively analyzed factors of therapy outcome in the SHCS database using univariate, multivariate, and regularized multivariate logistic regression models. As predictors of therapeutic success we identified the applied drugs, the GSS, and as the strongest predictor the IGB to regimen, a novel predictor derived from viral genotype.

Including genotype information into treatment outcome prediction is challenging because of the large number of observed mutations and the complexity of the genotype-phenotype relationship. Here, we have explored the IGB to drug resistance as a summary measure of the escape dynamics of the virus under treatment. The underlying idea of this modeling approach is that the IGB captures how difficult it is for the virus to escape from the selective pressure of individual drugs or from the entire drug combination. This piece of information is different from assessing the current genotypic or phenotypic drug resistance state of the virus, as intended, for example, by the GSS. The IGB makes a prediction about the expected escape dynamics of the virus population given its current genetic state.

The computation of the IGB involves an evolutionary model of genetic progression under selective drug pressure along multiple mutational pathways and a notion of evolutionary escape, which was based here on the predicted level of phenotypic drug resistance. We applied I-CBN models for jointly describing genetic progression and associated phenotypic change of the virus. In particular, phenotype predictions are non-linear in the mutations, which allows for capturing epistatic effects, i.e., the same mutation can have different effects on the resistance phenotype depending on the genetic background of the virus (Figure 1). The I-CBN models were estimated from independent genotype-phenotype data. Using these models, the complex, high-dimensional, genotypic data of each virus can be summarized efficiently by the IGB to resistance to each drug. Thus, rather than modeling interactions between drugs and individual mutations, the IGB provides a comprehensive model of drug-genotype interaction.

In the present study, we have extended the concept of the IGB to the entire regimen in a fashion that allows for computing this quantity for any drug combination and hence for large clinical datasets. The IGB to regimen can be regarded as the expected number of active drugs in the regimen. Assuming independent effects among drugs, we compute the regimen IGB from the drug IGBs. These simplifying assumptions are made for computational feasibility. They present a conceptual limitation of the approach and more elaborate models are conceivable. In addition, other variables not included in this study might be important, for example, pharmacological properties of drug combinations and host genetic factors. Here, the IGB, a single interpretable quantity, was found to be the strongest genotype-derived predictor of virological response and hence the most efficient representation of the viral genotype with respect to therapy outcome.

We have used throughout two definitions of virological success of treatment, namely reduction of viral load below 50 cps/ml and below 400 cps/ml. The latter less stringent cutoff was included because in the past it represented the limit of detection of viral load assays. Today viral load values of 50 cps/ml and lower can be measured and reduction below 50 cps/ml (or below the limit of detection) is an accepted therapeutic goal. We generally found very similar results for the two datasets, but the advantage of using IGB over GSS (the de facto standard genotype interpretation tool) reached statistical significance only for 400 cps/ml, but not for 50 cps/ml. This finding may, in part, be due to the

larger dataset and hence increased statistical power for 400 cps/ml as compared to 50 cps/ml. In the future, larger datasets will be required to further evaluate the IGB and its potential to predict treatment outcome without the need for expert rules. This property of the IGB is particularly appealing for new drugs, for which reliable rules are not readily available before evidence has accumulated in published studies. Larger datasets and more elaborate statistical variable importance methods [48] will also increase the power to detect other factors of therapeutic outcome, but the general consistency between the 50 cps/ml and 400 cps/ml success definitions suggests that a sizable fraction of important variables have been identified. In addition, larger TCE databases will allow for analyzing alternative endpoints, such as time to virological failure or virological response after a fixed period of time.

In the univariate analysis, most drugs had a positive effect on treatment outcome, with the exception of ZDV, d4T, 3TC, and NFV. The negative associations might be due to the prominent use of the drug combinations (ZDV or d4T) +3TC+ (IDV or NFV), 90% of which were failures. The four drugs were among the first to be approved for antiretroviral therapy and used in early suboptimal regimens. Moreover, they were poorly tolerated and therefore one can expect a general lower adherence to treatment. A similar observation was made in the multivariate analysis, where ZDV, ddI, SQV, 3TC and d4T were significant predictors decreasing the odds of therapeutic success. This effect might also be due to the common early use of these drugs in mono therapy and their later use in salvage regimens, even if multiple resistance mutations had already accumulated [49]. Among PIs, a pronounced trend was that boosting with RTV increased the odds of successful treatment. The fraction of PI boosting in the dataset is reported in Supporting Table S4.

A few variables did not show significant association with therapy outcome although they might have been expected to. For example, adherence is a well-known predictor of treatment success [50,51], but it failed to reach significance in the multivariate model, most likely due to lack of adherence data for about 45% of the patients. The missing data resulted from collecting adherence data within the SHCS only since January 2003. Indeed, in a multivariate analysis restricted to the subset of 1183 TCEs with observed adherence a more pronounced effect can be observed. We have not included a set of variables in this study that are known to be predictors because of the construction of the dataset. The definition of the dataset of genotype-therapy pairs allows for including several sequential TCEs from the same patient. Most TCEs are actually derived from unique patients, but some patients occur multiple times. Each TCE gives rise to two therapy cases, a failure, which had given rise to the switch, followed by a salvage regimen, which can be a failure or a success. Therefore, we did not include variables that are affected by the sequential ordering of therapies, such as the total time a patient was under therapy with a certain drug or the calendar year of treatment.

In summary, the IGB to regimen is a new predictor of treatment outcome that captures, in a single quantity, the virus' genetic potential for developing drug resistance under the selective pressure of the combination therapy. The IGB can be computed efficiently for any viral genotype and any drug combination. It may thus contribute to improved interpretation of genotypic drug resistance tests and to the rational design of individualized therapies. Future prospective studies are required to apply these results to other patient populations and to eventually integrate them into clinical practice.

Methods

Swiss HIV Cohort Study (SHCS) database

Founded in 1988, the SHCS is a nationwide, prospective, multicenter, clinic-based cohort with continuous enrolment and semi-annual study visits representing approximately 50% of all HIV-infected and 75% of all treated patients in Switzerland [46]. The SHCS has been approved by ethical committees of all participating institutions, and written informed consent has been obtained from all participants. The SHCS drug resistance database contains the results of 13,201 genotypic resistance tests from 9,231 patients, stored in a central database [45]. Resistance data stem from routine clinical testing (60%) and from tests performed retrospectively from frozen repository plasma samples (40%) (Table 1 and 2).

The SHCS has been approved by the following ethical committees of all participating institutions: Kantonale Ethikkommission Bern; Ethikkommission beider Basel; comité d'éthique du département de médecine de Hôpitaux Universitaires de Genéve; commission d'éthique de la recherche clinique, Lausanne; comitato etico cantonale, Bellinzona; Ethikkommission des Kanton St.Gallens; and Ethik-Kommission Zürich, all Switzerland. Written informed consent has been obtained from all participants [46].

Treatment change Episode (TCE) data

TCEs were obtained from the SHCS database as follows. Each TCE consists of a failing therapy followed by a salvage therapy (Supporting Figure S1). We required that the failing therapy was at least four month long and that the genotype was measured no more than 90 days before and no more than 30 days after onset of the uninterrupted salvage therapy [26]. In order to restrict to failing regimens due to viral rebound and to exclude convenience treatment changes or single determinations of low-level viremias (blips), a failing therapy was defined by either two consecutive viral load measurements above 500 cps/ml, or a single viral rebound followed by therapy switch, or single rebound after 180 days and lack of viral suppression below the limit of detection.

Therapies were labeled 'success' versus 'failure' as follows. Any failing therapy was considered a failure. Salvage therapies were considered successful, if viral load dropped below 50 cps/ml at any time point during treatment, otherwise they were considered failures. Because viral load assays with a sensitivity of 50 cps/ml were not available for the whole observation period, we also considered an alternative definition of therapy success as a viral load reduction below 400 cps/ml. The TCE dataset spans the time period 1988–2010, but 75% of TCEs date from 2000 or later.

Isotonic Conjunctive Bayesian Network (I-CBN) models

Genetic progression of the virus under selective drug pressure and the resulting phenotypic drug resistance changes were modeled jointly using I-CBNs [44]. In this model, mutations occur subject to partial order constraints which define the genotype lattice, the set of genotypes compatible with the constraints, and drug resistance is non-decreasing along any mutational pathway (Figure 1). Formally (see [44] for details), let $(\mathcal{E}, \prec)$ be a partially ordered set of n mutations. Each genotype is identified with the subset $g \subseteq \mathcal{E}$ of mutations it carries. The genotype lattice $\mathcal{G}$ induced by $(\mathcal{E}, \prec)$ is the set of all genotypes g for which it holds that $e \in g$ implies $e' \in g$ whenever $e' \prec e$ in $\mathcal{E}$. We denote by $\mathrm{Exit}_{\prec}(g)$ the set of accessible mutations from genotype g under the given partial order constraints. The I-CBN is a statistical model for the random variables $X \in 2^{\mathcal{E}} \simeq \{0,1\}^n$, describing observed genotypes, and $Y \in \mathbb{R}$, describing associated

drug resistance phenotypes, both of which are observed from true hidden genotypes $Z \in \mathcal{G} \subseteq 2^{\mathcal{E}}$ subject to noise. The probability of an unobserved genotype Z is defined as

$$P(Z = g) = \prod_{e \in g} \theta_e \cdot \prod_{e \in \mathrm{Exit}_{\prec}(g)} (1 - \theta_e) \qquad (1)$$

where the parameters θ_e denote the conditional probabilities of mutation $e \in \mathcal{E}$ given that all of its predecessor mutations have occurred, $\theta_e = P(e \in Z | \forall e' \in Z, e' < e, e' \neq e)$. The observed random variables X and Y are independent given Z. The genotype observation error is modeled as

$$P(X|Z) = (1 - \varepsilon)^{n - \delta(X,Z)} \varepsilon^{\delta(X,Z)} \qquad (2)$$

where δ denotes the Hamming distance and errors are assumed to occur independently among sites at rate ε. The observed drug resistance phenotype Y is the log fold-change in susceptibility. For each genotype Z, it follows a normal distribution

$$P(Y|Z) = \mathrm{Norm}(\mu_Z, \sigma) \qquad (3)$$

subject to the monotonicity contraints $\mu_g \leq \mu_h$ for all genotypes $g \subseteq h$. The complete model for X and Y is then the marginalization

$$P(X, Y) = \sum_Z P(Z) P(X|Z) P(Y|Z) \qquad (4)$$

Parameter estimation for this model was performed using the EM algorithm described in [44].

The model was applied separately to 18 antiretroviral drugs, using between 280 and 2303 (median 1448) cross-sectional genotype-phenotype pairs, i.e., observations of (X, Y), obtained from the Stanford HIV Drug Resistance Database, restricted to subtype B sequences and to Phenosense or Antivirogram assays [52]. For each drug, we selected its resistance-associated mutations reported on the Stanford HIVdb website lumping together mutations occurring at the same site, or if unavailable, applied L_1-penalized (lasso) linear regression [53,54] to select from all PR or RT mutations occurring at least ten times a sparse set $\mathcal{E}$ of $n = 10$ predictor mutations. The performance of the models is reported as the Pearson correlation coefficient between true and predicted phenotypes, estimated from a separate, random subset of 20% of the data. Phenotypic cutoff values were derived from the distribution of fold-change values as described previously [15,26] and used to dichotomize resistance predictions (Supporting Table S2).

Individualized Genetic Barrier (IGB)

Given an I-CBN model, transition probabilities among genotypes $g, h \in \mathcal{G}$ can be computed as

$$P(g \rightarrow h) = \prod_{e \in h \setminus g} \theta_e \cdot \prod_{e \in \mathrm{Exit}_{\prec}(g) \setminus h} (1 - \theta_e) \qquad (5)$$

Using these transition probabilities and the predicted drug resistance phenotypes μ_g, we define the IGB of genotype $g \in \mathcal{G}$ to resistance to drug d as the probability of the virus not reaching any genotypic state predicted as resistant,

$$\mathrm{IGB}_d(g) = 1 - \sum_{h \in \mathcal{G}_d} P(g \rightarrow h) \qquad (6)$$

where $\mathcal{G}_d \subset \mathcal{G}$ is the subset of all genotypes g predicted to be resistant to drug d, i.e., for which μ_g is greater than the resistance cutoff (Supporting Table S2).

Genotypes outside the lattice $\mathcal{G}$ (not complying with the partial order constraints) are regarded as erroneous observations of the genotypes in the lattice. The IGB of such a genotype f is

$$\text{IGB}_d(f) = 1 - \sum_{g \in G} P(g|f)[1 - \text{IGB}_d(g)] \qquad (7)$$

where $P(g|f)$ is the probability of the actual genotype being g given that f has been observed. By Bayes' theorem,

$$P(g|f) = \frac{P(f|g)P(g)}{\sum_{h \in \mathcal{G}} P(f|h)P(h)} \qquad (8)$$

where $P(f|g) = (1 - \varepsilon)^{n - \delta(f,g)} \varepsilon^{\delta(f,g)}$ is modeled as in Eq. 2.

The genetic barrier to escape from a regimen R is defined as the sum of the drug-specific barriers over all drugs in the regimen

$$\text{IGB}_R(g) = \sum_{d \in R} \text{IGB}_d(g) \qquad (9)$$

Because the IGB to each drug can be regarded as an estimate of the activity of the drug (the probability of not escaping), the IGB to a regimen may be interpreted as the expected number of active drugs in the regimen. Note that $0 \leq \text{IGB}_R(g) \leq |R|$, that $\text{IGB}_R(g) \approx 0$ means that evolutionary escape is almost certain, and that adding a drug to a regimen can only increase the genetic barrier to the regimen.

Statistical analysis

For classifying therapies as failures versus successes, univariate, multivariate, and regularized multivariate logistic regression was used. For a set of precitors $x_1, \ldots, x_m$, the therapeutic success probability p is modeled by the regression

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m \qquad (10)$$

where β_j are the regression coefficients. The odds ratio of therapeutic success associated with a one-unit increase in predictor j is e^{β_j}. P-values for the predictors are corrected for multiple testing using the Benjamini-Hochberg procedure. For regularization, we applied the elastic net [55], which combines an L_1 (lasso) penalty encouraging sparse solutions with an L_2 (ridge) penalty that tends to average across correlated features. Classifier performance was evaluated using ROC curves and is reported as the area under the ROC curve (AUC). The data was ten times randomly split into 40% for estimation of the two hyperparameters (one for the degree of each type of regularization) and 60% for model fitting and testing, which was done by 10-fold cross-validation [56].

The R language for statistical computing (http://www.r-project.org/) was used for all analyses, including the R packages icbn, glmnet, and ROCR. An R script for computing the IGB is available at: http://www.cbg.ethz.ch/software/igb. The Stanford HIVDB Sierra web service was used for GSS computation.

Acknowledgments

Supporting Information

Figure S1 Treatment change episode (TCE). Each TCE consists of a failing therapy followed by a salvage therapy. The failing therapy gives rise to a failure, whereas the salvage therapy can be either a success or a failure, depending on whether viral load suppression below 50 cps/ml (400 cps/ml) was achieved during treatment or not (see Methods). Genotypes are measured prior to or at the beginning of the salvage regimen. Examples of successful salvage therapy and failing salvage therapy are given in part (A) and (B) of this figure, respectively.
(EPS)

Figure S2 Drug usage in the SHCS database. Drug frequencies among successful (green) and failing (red) regimens for the TCEs of the SHCS database. Successful treatment was defined as a reduction in viral load below 50 cps/ml (A) or 400 cps/ml (B).
(EPS)

Figure S3 Most abundant drug combinations in the SHCS database. Frequencies of the 30 most abundant drug combinations in the SHCS database. Successful treatment was defined as a reduction in viral load below 50 cps/ml (A) or 400 cps/ml (B).
(EPS)

Figure S4 I-CBN model for resistance development to ZDV. Partially ordered set of RT mutations 41L, 67N, 70R, 74I, 74V, 184V, 210W, 215F, 215Y, 219Q associated with resistance to ZDV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S5 I-CBN model for resistance development to DDI. Partially ordered set of RT mutations 41L, 65R, 69Ins, 74VI, 151M, 184VI, 210W, 215FY associated with resistance to DDI (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S6 I-CBN model for resistance development to DDC. Partially ordered set of RT mutations 41L, 65R, 67N, 75M, 75T, 116Y, 151M, 184V, 210W, 211N associated with resistance to DDC (A) and induced genotype lattice (B). Genotypes

are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S7 I-CBN model for resistance development to D4T. Partially ordered set of RT mutations 41L, 65R, 67N, 69Ins, 70R, 151M, 184VI, 210W, 215FY, 219QE associated with resistance to D4T (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S8 I-CBN model for resistance development to 3TC. Partially ordered set of RT mutations 41L, 67N, 70R, 181C, 184V, 190A, 210W, 215F, 215Y, 219Q associated with resistance to 3TC (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S9 I-CBN model for resistance development to ABC. Partially ordered set of RT mutations 41L, 65R, 69Ins, 74VI, 115F, 151M, 184VI, 210W, 215FY associated with resistance to ABC (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S10 I-CBN model for resistance development to TDF. Partially ordered set of RT mutations 41L, 65R, 69Ins, 70R, 74VI, 115F, 151M, 184VI, 210W, 215FY associated with resistance to TDF (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S11 I-CBN model for resistance development to FTC. Partially ordered set of RT mutations 65R, 69Ins, 151M, 184VI associated with resistance to FTC (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S12 I-CBN model for resistance development to EFV. Partially ordered set of RT mutations 100I, 101EP, 103NS, 106AM, 181CIV, 188LHC, 190ASE, 230L associated with resistance to EFV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S13 I-CBN model for resistance development to NVP. Partially ordered set of RT mutations 100I, 101EP, 103NS, 106AM, 181CIV, 188LHC, 190ASE, 230L associated with resistance to NVP (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S14 I-CBN model for resistance development to RTV. Partially ordered set of PR mutations 24I, 30N, 32I, 46I, 46L, 54V, 73S, 82A, 84V, 90M associated with resistance to RTV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S15 I-CBN model for resistance development to SQV. Partially ordered set of PR mutations 48VM, 54VTALM, 82AT, 84V, 88S, 90M associated with resistance to SQV (A) and

induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S16 I-CBN model for resistance development to IDV. Partially ordered set of PR mutations 32I, 46IL, 47V, 54VTALM, 76V, 82AFTS, 84V, 88S, 90M associated with resistance to IDV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S17 I-CBN model for resistance development to NFV. Partially ordered set of PR mutations 30N, 46IL, 47V, 48VM, 54VTALM, 82AFTS, 84V, 88DS, 90M associated with resistance to NFV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S18 I-CBN model for resistance development to LPV. Partially ordered set of PR mutations 32I, 46IL, 47VA, 48VM, 50V, 54VTALM, 76V, 82AFTS, 84V, 90M associated with resistance to LPV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S19 I-CBN model for resistance development to APV. Partially ordered set of PR mutations 24I, 32I, 46I, 46L, 48V, 53L, 54V, 82A, 84V, 90M associated with resistance to APV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S20 I-CBN model for resistance development to ATV. Partially ordered set of PR mutations 10I, 32I, 33F, 46I, 48V, 54V, 71V, 82A, 84V, 90M associated with resistance to ATV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S21 I-CBN model for resistance development to TPV. Partially ordered set of PR mutations 32I, 46IL, 47VA, 54VAM, 82TL, 84V associated with resistance to TPV (A) and induced genotype lattice (B). Genotypes are colored green if predicted susceptible and red if predicted resistant.
(EPS)

Figure S22 Univariate analysis of predictors of response to antiretroviral combination therapy in the SHCS database. Associations have been tested using logistic regression models and odds ratios of therapeutic success, defined as viral load reduction below 50cps/ml (A) and 400cps/ml (B), are reported together with their 95% confidence intervals on a logarithmic scale. Benjamini-Hochberg-corrected p-values are represented as black ($p < 0.001$) and grey ($p < 0.01$) symbols. Only predictors with a p-value smaller than 0.01 are included.
(EPS)

Table S1 Complete list of all variables analyzed with respect to treatment outcome. Groups NRTI, NNRTI, and PI consist of binary variables, one for each drug, indicating the presence of the respective drug in the regimen. For PIs, boosted (given together with low-dose RTV) and unboosted formulations are distinguished, except for LPV which is always applied boosted. The variable RTV refers to the use of ritonavir as the only PI in the regimen. Demographic and clinical variables include age and

gender of the patient, whether he or she had AIDS, the maximum viral load and the minimum CD4 T cell count measured anytime before treatment onset, transmission group (BLOOD, HET, IDU, MSM, or OTHER), and adherence. Patient adherence was assessed in questionnaires and measured as the percentage of missed dosages [50,51] for 1183 (45%) of the patients, and then dichotomized. For the multivariate analysis only, unobserved values of patient adherence were imputed by a logistic regression model (one for each dataset) from all remaining variables except the response (treatment outcome). For each drug, the individualized genetic barrier (IGB) is the probability of the virus not escaping from the selective pressure of the drug. The IGB to regimen is defined as the sum of the drug-specific IGBs over all drugs in the regimen. Mutations in the PR and RT of HIV-1 are denoted by the sequence position followed by the amino acid. Each variable is binary indicating the presence of the respective amino acid at the respective position in the protein. Only mutations that occurred in at least 5% of the samples are considered.
(PDF)

Table S2 Construction of I-CBN models. For each drug, is reported the number N of genotype-phenotype pairs the model has been learned from, the correlation coefficient R between predicted and true drug resistance phenotypes, the list of selected mutations, and the cutoff value C defining resistant versus susceptible viruses. The correlation coefficient has been estimated from an independent test set consisting of 20% of the data that was not used for training. For ZDV, DDI, D4T, 3TC, ABC, TDF, FTC, EFV, NVP, SQV, IDV, NFV, LPV, and TPV, the corresponding drug resistance-associated mutations reported on the Stanford HIV Drug Resistance Database website were used, while for DDC, RTV, APV, and ATV, we selected ten mutations using L1-penalized linear regression (lasso).
(PDF)

Table S3 Different categories of drug combinations in SHCS databse. The first category includes drug combinations currently recommended as first-line or alternative regimens according to the JAMA recommendations [42]. Category 2 includes regimens that were recommended as first-line or second-line regimens in the past, regimens that are still in use in developing countries or are used sometimes if drug resistant virus is present at baseline, or salvage regimens. Category 3 includes older regimens that are not in use anymore as first-line regimens but were before, regimens that are not corresponding to guidelines, including those that are sometimes used in special circumstances, such as unusual tolerability, etc. To evaluate the prediction performance (sensitivity and specificity) of each

category, leave-one-out cross-validation experiments were performed.
(PDF)

Table S4 PI usage and boosting fraction. Reported is the total number of regimens in the SHCS database that include the respective PI, and in parenthesis, the percentage that the PI is boosted, i.e., given together with low-dose ritonavir (RTV).
(PDF)

Table S5 Comparative performance in predicting treatment outcome, defined as a reduction of viral load below 50cps/ml, for different elastic net regularized logistic regression models. Comparative performance in predicting treatment outcome, defined as a reduction of viral load below 50cps/ml, for different elastic net regularized logistic regression models. In columns 3–8, the p-value of a two-sided Wilcoxon rank sum test for differences in the area under the ROC curve (AUC; column 2) is reported. Prediction models (column 1) are encoded by the sets of predictors used, where C refers to the demographic and clinical variables, D refers to drugs, and M to mutations. For example, the model IGB+CDM includes as predictors IGB to regimen, clinical and demographic predictors, applied drugs, and mutations.
(PDF)

Table S6 Comparative performance in predicting treatment outcome, defined as a reduction of viral load below 400cps/ml, for different elastic net regularized logistic regression models. Comparative performance in predicting treatment outcome, defined as a reduction of viral load below 400cps/ml, for different elastic net regularized logistic regression models. In columns 3–8, the p-value of a two-sided Wilcoxon rank sum test for differences in the area under the ROC curve (AUC; column 2) is reported. Prediction models (column 1) are encoded by the sets of predictors used, where C refers to the demographic and clinical variables, D refers to drugs, and M to mutations. For example, the model IGB+CDM includes as predictors IGB to regimen, clinical and demographic predictors, applied drugs, and mutations.
(PDF)

Author Contributions

Conceived and designed the experiments: NB PK VvW HFG. Performed the experiments: VvW HF MB BH MC PV EB SY JB TK CC HFG. Analyzed the data: NB HM HS PK VvW HFG. Contributed reagents/materials/analysis tools: VvW HF MB BH MC PV EB SY JB TK CC HFG. Wrote the paper: NB HM HS PK VvW HFG.

References

1. 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: 321–333.
2. Hirsch MS, Günthard HF, Schapiro JM, Brun-Vézinet F, Clotet B, et al. (2008) Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an international AIDS Society-USA panel. Clin Infect Dis 47: 266–285.
3. Saigo H, Altmann A, Bogojeska J, Mueller F, Nowozin S, et al. (2011) Learning from past treatments and their outcome improves prediction of in vivo response to anti-HIV therapy. Stat Appl Genet Mol Biol 10: Article 6.
4. Lawyer G, Altmann A, Thielen A, Zazzi M, Snnerborg A, et al. (2011) HIV-1 mutational pathways under multidrug therapy. AIDS Res Ther 8: 26.
5. Jiang H, Deeks SG, Kuritzkes DR, Lallemant M, Katzenstein D, et al. (2003) Assessing resistance costs of antiretroviral therapies via measures of future drug options. J Infect Dis 188: 1001–1008.
6. Fitzgerald AP, DeGruttola VG, Vaida F (2002) Modelling HIV viral rebound using non-linear mixed effects models. Stat Med 21: 2093–2108.
7. Prosperi MCF, Di Giambenedetto S, Fanti I, Meini G, Bruzzone B, et al. (2011) A prognostic model for estimating the time to virologic failure in HIV-1 infected

patients undergoing a new combination antiretroviral therapy regimen. BMC Med Inform Decis Mak 11: 40.
8. Torti C, Quiros-Roldan E, Keulen W, Scudeller L, Caputo SL, et al. (2003) Comparison between rules-based human immunodeficiency virus type 1 genotype interpretations and real or virtual phenotype: concordance analysis and correlation with clinical outcome in heavily treated patients. J Infect Dis 188: 194–201.
9. DeGruttola V, Dix L, D'Aquila R, Holder D, Phillips A, et al. (2000) The relation between baseline HIV drug resistance and response to antiretroviral therapy: re-analysis of retrospective and prospective studies using a standardized data analysis plan. Antivir Ther 5: 41–48.
10. Haupts S, Ledergerber B, Böni J, Schüpbach J, Kronenberg A, et al. (2003) Impact of genotypic resistance testing on selection of salvage regimen in clinical practice. Antivir Ther 8: 443–454.
11. Cingolani A, Antinori A, Rizzo MG, Murri R, Ammassari A, et al. (2002) Usefulness of monitoring HIV drug resistance and adherence in individuals failing highly active antiretroviral therapy: a randomized study (ARGENTA). AIDS 16: 369–379.

12. Tural C, Ruiz L, Holtzer C, Schapiro J, Viciana P, et al. (2002) Clinical utility of HIV-1 genotyping and expert advice: the Havana trial. AIDS 16: 209–218.

13. Mazzotta F, Caputo SL, Torti C, Tinelli C, Pierotti P, et al. (2003) Real versus virtual phenotype to guide treatment in heavily pretreated patients: 48-week follow-up of the genotipo-fenotipo di resistenza (GenPheRex) trial. J Acquir Immune Defic Syndr 32: 268–280.

14. Meynard JL, Vray M, Morand-Joubert L, Race E, Descamps D, et al. (2002) Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. AIDS 16: 727–736.

15. Beerenwinkel N, Lengauer T, Däumer M, Kaiser R, Walter H, et al. (2003) Methods for optimizing antiviral combination therapies. Bioinformatics 19 Suppl 1: i16–i25.

16. Beerenwinkel N, Sing T, Lengauer T, Rahnenführer J, Roomp K, et al. (2005) Computational methods for the design of effective therapies against drug resistant HIV strains. Bioinformatics 21: 3943–3950.

17. Lengauer T, Sing T (2006) Bioinformatics-assisted anti-HIV therapy. Nat Rev Microbiol 4: 790–797.

18. Sinisi SE, Polley EC, Petersen ML, Rhee SY, van der LaanMJ (2007) Super learning: an application to the prediction of HIV-1 drug resistance. Stat Appl Genet Mol Biol 6: Article7.

19. Larder B, Wang D, Revell A, Montaner J, Harrigan R, et al. (2007) The development of artificial neural networks to predict virological response to combination HIV therapy. Antivir Ther 12: 15–24.

20. Altmann A, Rosen-Zvi M, Prosperi M, Aharoni E, Neuvirth H, et al. (2008) Comparison of classifier fusion methods for predicting response to anti HIV-1 therapy. PLoS One 3: e3470.

21. Rosen-Zvi M, Altmann A, Prosperi M, Aharoni E, Neuvirth H, et al. (2008) Selecting anti-HIV therapies based on a variety of genomic and clinical factors. Bioinformatics 24: i399–i406.

22. Prosperi MCF, D'Autilia R, Incardona F, Luca AD, Zazzi M, et al. (2009) Stochastic modelling of genotypic drug-resistance for human immunodeficiency virus towards long-term combination therapy optimization. Bioinformatics 25: 1040–1047.

23. Prosperi MCF, Altmann A, Rosen-Zvi M, Aharoni E, Borgulya G, et al. (2009) Investigation of expert rule bases, logistic regression, and non-linear machine learning techniques for predicting response to antiretroviral treatment. Antivir Ther 14: 433–442.

24. Bogojeska J, Bickel S, Altmann A, Lengauer T (2010) Dealing with sparse data in predicting outcomes of HIV combination therapies. Bioinformatics 26: 2085–2092.

25. Bogojeska J, Lengauer T (2012) Hierarchical Bayes model for predicting effectiveness of HIV combination therapies. Stat Appl Genet Mol Biol 11: Article 11.

26. Altmann A, Beerenwinkel N, Sing T, Savenkov I, Däumer M, et al. (2007) Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance. Antivir Ther 12: 169–178.

27. Altmann A, Sing T, Vermeiren H, Winters B, Craenenbroeck EV, et al. (2009) Advantages of predicted phenotypes and statistical learning models in inferring virological response to antiretroviral therapy from HIV genotype. Antivir Ther 14: 273–283.

28. Altmann A, Däumer M, Beerenwinkel N, Peres Y, Schülter E, et al. (2009) Predicting the response to combination antiretroviral therapy: Retrospective validation of geno2pheno-THEO on a large clinical database. J Infect Dis 199: 999–1006.

29. Swanstrom R, Bosch RJ, Katzenstein D, Cheng H, Jiang H, et al. (2004) Weighted phenotypic susceptibility scores are predictive of the HIV-1 RNA response in protease inhibitor-experienced HIV-1-infected subjects. J Infect Dis 190: 886–893.

30. Rhee SY, Fessel WJ, Liu TF, Marlowe NM, Rowland CM, et al. (2009) Predictive value of HIV-1 genotypic resistance test interpretation algorithms. J Infect Dis 200: 453–463.

31. Frentz D, Boucher CAB, Assel M, Luca AD, Fabbiani M, et al. (2010) Comparison of HIV-1 genotypic resistance test interpretation systems in predicting virological outcomes over time. PLoS One 5: e11505.

32. Van Laethem K, De Luca A, Antinori A, Cingolani A, Perna CF, et al. (2002) A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1-infected patients. Antivir Ther 7: 123–129.

33. Zazzi M, Prosperi M, Vicenti I, Giambenedetto SD, Callegaro A, et al. (2009) Rules-based HIV-1 genotypic resistance interpretation systems predict 8 week and 24 week virological antiretroviral treatment outcome and benefit from drug potency weighting. J Antimicrob Chemother 64: 616–624.

34. De Luca A, Cingolani A, Giambenedetto SD, Trotta MP, Baldini F, et al. (2003) Variable prediction of antiretroviral treatment outcome by different systems for interpreting genotypic human immunodeficiency virus type 1 drug resistance. J Infect Dis 187: 1934–1943.

35. Zazzi M, Kaiser R, Snnerborg A, Struck D, Altmann A, et al. (2011) Prediction of response to antiretroviral therapy by human experts and by the EuResist data-driven expert system (the EVE study). HIV Med 12: 211–218.

36. Deforche K, Cozzi-Lepri A, Theys K, Clotet B, Camacho RJ, et al. (2008) Modelled in vivo HIV fitness under drug selective pressure and estimated genetic barrier towards resistance are predictive for virological response. Antivir Ther 13: 399–407.

37. Beerenwinkel N, Däumer M, Sing T, Rahnenführer J, Lengauer T, et al. (2005) Estimating HIV evolutionary pathways and the genetic barrier to drug resistance. J Infect Dis 191: 1953–1960.

38. Gish R, Jia JD, Locarnini S, Zoulim F (2012) Selection of chronic hepatitis b therapy with high barrier to resistance. Lancet Infect Dis 12: 341–353.

39. Götte M (2012) The distinct contributions of fitness and genetic barrier to the development of antiviral drug resistance. Curr Opin Virol 2: 644–650.

40. Theys K, Deforche K, Beheydt G, Moreau Y, van Laethem K, et al. (2010) Estimating the individualized HIV-1 genetic barrier to resistance using a nelfinavir fitness landscape. BMC Bioinformatics 11: 409.

41. Deforche K, Camacho R, Laethem KV, Lemey P, Rambaut A, et al. (2008) Estimation of an in vivo fitness landscape experienced by hiv-1 under drug selective pressure useful for prediction of drug resistance evolution during treatment. Bioinformatics 24: 34–41.

42. Theys K, Deforche K, Libin P, Camacho RJ, Laethem KV, et al. (2010) Resistance pathways of human immunodeficiency virus type 1 against the combination of zidovudine and lamivudine. J Gen Virol 91: 1898–1908.

43. van de Vijver DA, Wensing AMJ, Angarano G, Asjö B, Balotta C, et al. (2006) The calculated genetic barrier for antiretroviral drug resistance substitutions is largely similar for different hiv-1 subtypes. J Acquir Immune Defic Syndr 41: 352–360.

44. Beerenwinkel N, Knupfer P, Tresch A (2011) Learning monotonic genotype-phenotype maps. Stat Appl Genet Mol Biol 10: 3.

45. von Wyl V, Yerly S, Böni J, Brgisser P, Klimkait T, et al. (2007) Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. Arch Intern Med 167: 1782–1790.

46. Schoeni-Affolter F, Ledergerber B, Rickenbach M, Rudin C, Günthard HF, et al. (2010) Cohort profile: the Swiss HIV Cohort Study. Int J Epidemiol 39: 1179–1189.

47. Thompson MA, Aberg JA, Hoy JF, Telenti A, Benson C, et al. (2012) Antiretroviral treatment of adult HIV infection: 2012 recommendations of the international antiviral Society-USA panel. JAMA 308: 387–402.

48. van der Laan MJ, Rose S (2011) Targeted Learning. Springer.

49. Scherrer AU, von Wyl V, Böni J, Yerly S, Klimkait T, et al. (2011) Viral suppression rates in salvage treatment with raltegravir improved with the administration of genotypic partially active or inactive nucleoside/tide reverse transcriptase inhibitors. J Acquir Immune Defic Syndr 57(1):24–31.

50. Glass TR, Geest SD, Hirschel B, Battegay M, Furrer H, et al. (2008) Self-reported non-adherence to antiretroviral therapy repeatedly assessed by two questions predicts treatment failure in virologically suppressed patients. Antivir Ther 13: 77–85.

51. Glass TR, Battegay M, Cavassini M, Geest SD, Furrer H, et al. (2010) Longitudinal analysis of patterns and predictors of changes in self-reported adherence to antiretroviral therapy: Swiss HIV Cohort Study. J Acquir Immune Defic Syndr 54: 197–203.

52. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, et al. (2003) Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res 31: 298–303.

53. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Roy Stat Soc B 58: 267–288.

54. Rabinowitz M, Myers L, Banjevic M, Chan A, Sweetkind-Singer J, et al. (2006) Accurate prediction of HIV-1 drug response from the reverse transcriptase and protease amino acid sequences using sparse models created by convex optimization. Bioinformatics 22: 541–549.

55. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Statist Soc B 67, Part 2: 301320.

56. Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning, 2nd edition. Springer.