Statistical Applications in Genetics and Molecular Biology

Volume 10, *Issue* 1 2011 *Article* 6

Learning from Past Treatments and Their Outcome Improves Prediction of *In Vivo* Response to Anti-HIV Therapy

Hiroto Saigo[*] Andre Altmann[†] Jasmina Bogojeska[‡]
Fabian Müller[**] Sebastian Nowozin[††] Thomas Lengauer[‡‡]

[*] Max Planck Institute for Informatics, hiroto.saigo@mpi-inf.mpg.de
[†] Max Planck Institute for Informatics, altmann@mpi-inf.mpg.de
[‡] Max Planck Institute for Informatics, jasmina@mpi-inf.mpg.de
[**] Max Planck Institute for Informatics, fmueller@mpi-inf.mpg.de
[††] Max Planck Institute for Biological Cybernetics, sebastian.nowozin@tuebingen.mpg.de
[‡‡] Max Planck Institute for Informatics, lengauer@mpi-inf.mpg.de

Learning from Past Treatments and Their Outcome Improves Prediction of *In Vivo* Response to Anti-HIV Therapy[*]

Hiroto Saigo, Andre Altmann, Jasmina Bogojeska, Fabian Müller, Sebastian Nowozin, and Thomas Lengauer

Abstract

Infections with the human immunodeficiency virus type 1 (HIV-1) are treated with combinations of drugs. Unfortunately, HIV responds to the treatment by developing resistance mutations. Consequently, the genome of the viral target proteins is sequenced and inspected for resistance mutations as part of routine diagnostic procedures for ensuring an effective treatment. For predicting response to a combination therapy, currently available computer-based methods rely on the genotype of the virus and the composition of the regimen as input. However, no available tool takes full advantage of the knowledge about the order of and the response to previously prescribed regimens. The resulting high-dimensional feature space makes existing methods difficult to apply in a straightforward fashion. The machine learning system proposed in this work, sequence boosting, is tailored to exploiting such high-dimensional information, i.e. the extraction of longitudinal features, by utilizing the recent advancements in data mining and boosting.

When applied to predicting the latest treatment outcome for 3,759 treatment-experienced patients from the EuResist integrated database, sequence boosting achieved superior performance compared to SVMs with RBF kernels. Moreover, sequence boosting allows an easy access to the discriminative treatment information.

Analysis of feature importance values provided by our model confirmed known facts regarding HIV treatment. For instance, application of potent and recently licensed drugs was beneficial for patients, and, conversely, the patient group that was subject to NRTI mono-therapies in the past had poor treatment perspectives today. Furthermore, our model revealed novel biological insights. More precisely, the combination of previously used drugs with their in vivo response is more in-

[*] Hiroto Saigo and Andre Altmann are joint first authors. Please send correspondence to Hiroto Saigo, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka, Japan; email: hiroto@mpi-inf.mpg.de, or Andre Altmann, Max Planck Institute of Psychiatry, 80804 Munich, Germany; email altmann@mpi-inf.de. The authors thank the EuResist project (contract number EU-STREP IST-2004-027173) and the EuResist coordinator Maurizio Zazzi for providing the data. The authors also thank Rolf Kaiser for comments on the medical interpretation.

formative than the information of previously used drugs alone. Using this information improves the performance of systems for predicting therapy outcome.

KEYWORDS: data mining, discriminative sequence features, boosting, HIV, clinical, optimization

1 Introduction

The human immunodeficiency virus (HIV) was discovered in the early 80's (Barré-Sinoussi et al., 1983). Until now, it has claimed the lives of more than 25 million people, and currently more than 33 million people are reported to be infected with HIV[1]. HIV is a retrovirus, i.e. its genome is coded in RNA, which first has to be reversely transcribed to DNA for exploiting the replication machinery of the host cell (Fields et al., 2007). The process of reverse transcription is carried out by the viral protein reverse transcriptase (RT). Current HIV therapy is limited to suppressing the viral load (i.e. number of copies of viral RNA in one ml of blood serum) and therefore delaying disease progression to AIDS and death. The high mutation rate of HIV (Gao et al., 2004) is due to RT lacking a proof-reading mechanism. This poses a challenge to antiretroviral treatment, since it is only a matter of time until mutations are generated that allow the virus to replicate in the presence of a drug. Due to the replicative advantage, these drug resistance mutations are selected evolutionarily and cause the failure of the ongoing regimen. In order to delay resistance development, modern anti HIV therapies comprise multiple drugs attacking the virus at multiple stages of the replication cycle (Clavel and Hance, 2004). Fusion inhibitors (FIs) prevent the entry of HIV into its host cells. Nucleoside and non-nucleoside reverse transcriptase inhibitors (abbreviated NRTIs and NNRTIs, respectively) inhibit the viral RT. Protease inhibitors (PIs) bind to the active site of the viral protease that cleaves precursor proteins into functionally active units (for further details see for instance Clavel and Hance (2004)).

Eventually also these highly active antiretroviral therapies (HAARTs) fail and the treating clinician has to find a new combination of active antiretroviral drugs. This task is complicated by the phenomenon of cross-resistance, which means that resistance mutations selected by one drug also confer resistance against drugs with same mode of action targeting the same viral protein. To arrive at a beneficial selection of drugs to administer to the patient, the sequence of the genetic regions coding for the viral target proteins is obtained from the patient's virus. This sequence is then inspected for resistance mutations. This process is state-of-the-art, but one major obstacle remains: as soon as drug resistance mutations do no longer present a replicative advantage, they may disappear from the currently predominant viral variant. This can happen if a treatment is altered or paused. Unfortunately, the patient harbors previous viral variants in the form of proviral DNA in several infected tissues (Fields et al., 2007). This constitutes a memory of resistance mutations provoked by previous treatments. As a consequence, recycling of drugs typi-

[1] http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/2008_Global_report.asp

cally leads to a rapid reselection of previously existing resistance mutations, which are not prevalent in the predominant viral variant in the patient's blood and consequently are not detectable by conventional sequencing methods. For avoiding such a short-term viral rebound, the treating clinician considers the patient's treatment history, i.e. the previously administered drugs, in addition to the viral genotype when selecting a new regimen. Treatment history has long been recognized as clinically relevant (Bratt et al., 1998). More recently it was shown that taking all available (past and present) genotypes of the patient into account improves the prediction of treatment response in heavily pretreated patients (Zaccarelli et al., 2009)

For assisting the interpretation of genotypic sequences statistical learning methods were used to assess resistance against single drugs (Beerenwinkel et al., 2002). This concept was recently extended to predict *in vivo* response to combination treatments (Altmann et al., 2007, 2009; Larder, 2007). However, so far computer-based methods make use of the patient's treatment history only by using binary indicators of previous exposure to a drug as additional features (Bickel et al., 2008; Larder, 2007; Rosen-Zvi et al., 2008). While this representation perfectly summarizes previous drug applications, it may miss important and informative cause-effect relationships, such as: the drug efavirenz (EFV) selects mutation RT103N, which leads to the administration of a new drug combination including lopinavir (LPV), but not any drug from the same class as EFV (see Figure 2 for an example). Here, the notation RT103N indicates that the wild-type amino acid at position 103 in the RT was replaced with Asparagine. This RT mutation alone is sufficient to confer complete resistance to the NNRTIs EFV and NVP (Antinori et al., 2002). Therefore, the failure of the latest NNRTI containing regimen may be attributed to the occurrence of mutation RT103N as response to the previous use of EFV. It is worth noting, however, that in clinical practice viral genotypes are not always measured, and consequently important mutations such as RT103N go unnoticed. In that case the treating clinician has to consider the possibility of accumulation of NNRTI resistance mutations from the previously administered regimens.

A further indicator for accumulated resistance mutations is the number of treatment changes (#TC), which is defined as the number of times the patient's treatment has been changed or interrupted. A higher #TC increases the risk of the patient experiencing a treatment failure (see Figure 1), pointing towards the increasing difficulty of treating patients that experienced many treatment changes. This is a direct consequence of the accumulated mutations in the currently dominant viral variant and in the viral reservoirs. Information on the treatment history is useful for predicting response to antiretroviral therapy in treatment-experienced patients, since mutations in the viral reservoirs are not detectable by standard genotyping. Therefore, the validation of our method, called *sequence boosting* focuses on

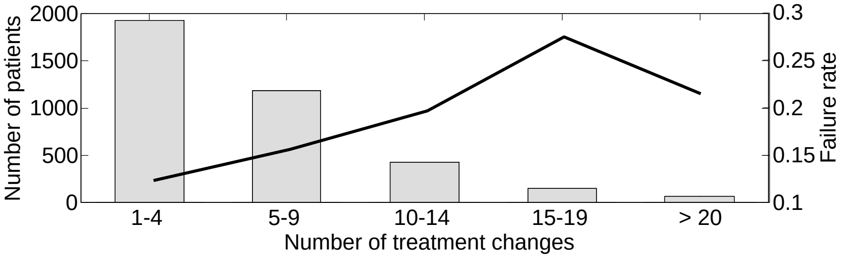


Figure 1: Ratio of patients with long treatment lines. The number of treatment changes is defined as the number of times the patient's treatment has been changed or interrupted. The figure depicts the increase of the treatment failure rate as the number of treatment changes grows, pointing to the increasing difficulty of treating patients that experienced many treatment changes.

treatment-experienced patients. In the following we present further related work. In Section 2 we introduce our source for HIV treatment data, the EuResist Integrated database. Furthermore, we explain how patient data obtained from the database is converted into *sequence features* that can be used by our learning method *sequence boosting*. Briefly, *sequence boosting* classifies the response to the current therapy based on all events in the patient's treatment record (viral genotypes, previously administered therapies, and response to these therapies) under consideration of their order. In essence, the resulting method is a linear classifier using non-linear features. In Sections 3 and 4 we present the results and discussion, respectively, of our computational experiments. And finally we conclude with Section 5.

1.1 Related work

The prediction of *in vivo* response to antiretroviral therapy has been approached in previous works. For instance, Rosen-Zvi et al. (2008) used logistic regression models with different sets of features for predicting the outcome of antiretroviral combination therapies. Among these features were up to three-way interaction terms between indicator variables for drugs, previously administered drugs, and mutations. The *sequence boosting* method presented here, considered up to N-way interaction features, where N is the number of treatment events in the longest treatment record in the training data. Other approaches, such as Bayesian networks (Deforche et al., 2006) or transfer learning (Bickel et al., 2008) were applied to the same treatment outcome prediction problem, but none of them employed such a large number of features.

Although the goal is different, a closely related and well-studied topic is the prediction of the phenotypic drug resistance. This topic is similar in a sense that it uses genotypic information as features and the target is the prediction of binary or real-valued response values. Various statistical learning methods have been applied in this area, including linear regression (Rabinowitz et al., 2006; Rhee et al., 2006; Saigo et al., 2007), decision trees (Beerenwinkel et al., 2002), support vector machines (SVMs) (Beerenwinkel et al., 2003; Sing et al., 2005; Sing and Beerenwinkel, 2007), artificial neural networks (Wang and Larder, 2003; Larder, 2007), bayesian networks (Deforche et al., 2008) and Markov models (Foulkes and DeGruttola, 2002, 2003).

2 Materials and Methods

2.1 The EuResist integrated database

The EuResist integrated database (release November 2007), which is the source of data for the computational experiments, comprises data from four different countries: Germany, Italy, Luxembourg, and Sweden. The database contains 61,831 different treatments from 18,467 patients collected in the years 1987 through 2007. For each patient the viral load (VL) measurements, therapies (sets of administered drugs), and genotypes are recorded.

For instance, Figure 2 shows an excerpt of a patient's treatment record, covering the last two treatment switches. This patient has started with FTC+TDF+LPV+RTVb and was switched to 3TC+AZT+EFV after an increase in viral load. Viral suppression was not maintained for a long period of time, thereafter FTC+TDF+NVP was selected. The viral load did not decrease in response to this treatment and the therapy was therefore considered a treatment failure. At the end of the second treatment term, the mutation 103N was observed in the RT coding region.

By definition, baseline VL and genotype are only assigned to a treatment, if they were obtained at most 90 days before treatment start. A follow-up VL is attributed to a treatment only if it is available in a specific time-interval (here: between 28 and 84 days) after onset of the therapy. Following the guidelines of the EuResist consortium (Rosen-Zvi et al., 2008), treatment response is dichotomized to success and failure, with a treatment success defined by a drop of the follow-up VL either below the limit of detection, i.e. 400 copies/ml, or by two orders of magnitude compared to the baseline VL. If no follow-up VL is available in that time frame, the corresponding treatment receives no label. Of note, the time frame of one to three months for the follow-up VL in the definition is considered a short-term response, as opposed to medium-term response of about 6 months and long-term response with even more distant end-points.

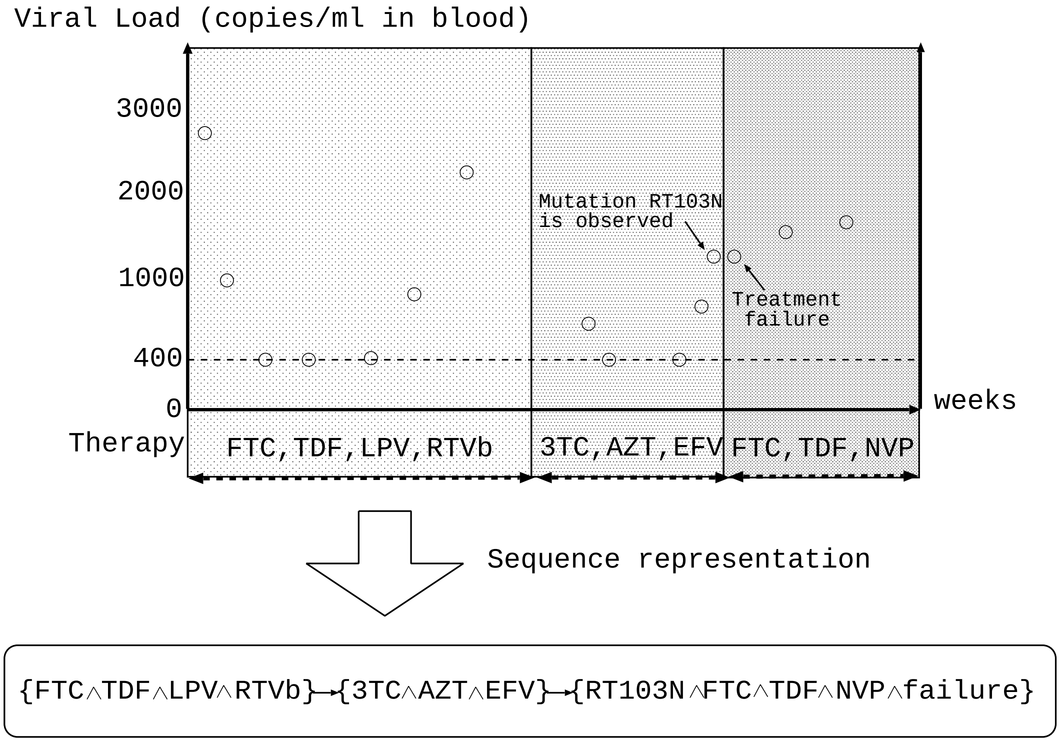


Figure 2: Example of a treatment record (top) and its sequence representation (bottom). (top) The chart shows the treatment record of a patient over the last two treatment switches. This patient has started treatment with FTC+TDF+LPV+RTVb, and was switched to 3TC+AZT+EFV after an increase in viral load. Suppression of viral load by the second regimen did not last long, thereafter FTC+TDF+NVP was selected. Here, the viral load did not decrease and thus the regimen is considered a treatment failure. At the end of the second treatment term, the mutation 103N was observed in the RT coding region (RT103N). (bottom) Sequence representation of the treatment history corresponding to the chart above.

2.2 Sequence representation of treatment history

The treatment records stored in the EuResist Integrated database have to be converted for making it a suitable input for *sequence boosting*. Each record in our training data takes the form (x_i, y_i) where $x_i \in \{0,1\}^{T \times F}$ and $y_i \in \{0,1\}$ denote the treatment sequence of the i-th patient and the outcome of the latest therapy, respectively. We use the definition of treatment outcome as presented in the previous paragraph. F is the largest number of treatment events in each treatment, i.e., “the number of drugs (25)” + “the number of mutations (108)” + “the outcome of the therapy (1)” = 134. The list of all the treatment events is summarized in the Appendix (Table A1). T is the length of the longest treatment sequence in training data, which, in our case, is 38. The list of 108 mutations is based on the list maintained by the International AIDS Society (Johnson et al., 2008). The binary vector x_i is equivalently represented as a set X_i, containing all indices t and f such that $x_{i,t,f} = 1$, where $\{t|1 \leq t \leq T\}$ and $\{f|1 \leq f \leq F\}$ are indices for the treatment and the additional information (such as observed mutations before the onset of the therapy, applied drugs and their short-time response) in each treatment, respectively. For instance, the appearance of events f_1 and f_2 in treatment t followed by the appearance of events f_3 and f_4, is represented as $x_{t,f_1} \wedge x_{t,f_2} \wedge x_{t+1,f_3} \wedge x_{t+1,f_4}$. In the following, we represent the corresponding sequence feature as $\{x_1 \wedge x_2\} \rightarrow \{x_3 \wedge x_4\}$, where the $\rightarrow$ operator means that the treatment on the right hand side follows the treatment on the left hand side. This representation is a generalization of the treatment change episodes (TCE) presented by Altmann et al. (2007). If drugs, mutations, and therapy outcome occur simultaneously in the treatment, then the order of the events is i) the observed mutations, ii) prescribed drugs, followed by iii) therapy outcome. This ordering originates from the observation that a resistance induced treatment change is usually preceded by the sequencing of the viral population in the patient. Likewise, the therapy outcome has to be preceded by a change in the drug combination.

Example 1 *A sequence* {FTC∧TDF∧LPV∧RTVb}→{3TC∧AZT∧EFV} → {RT103N ∧FTC∧TDF∧NVP∧*failure*} *represents a treatment including* FTC, TDF, LPV *and* RTVb *followed by the combination* 3TC+AZT+EFV. *At the end the second treatment, a mutation* RT103N *is observed and* FTC+TDF+NVP *is administered, which turned out to be a treatment failure.*

The terms *sequence* and *sequence feature* are formally defined as follows:

Definition 1 *(Sequence). A sequence $s = (s_1, s_2, \ldots s_T)$ is defined as an ordered list of elements s_t. Each element s_t is a finite set of integers corresponding to the indicators for treatment events in the t-th treatment.*

Example 2 *In Example 1, s_1 has treatment events* FTC, TDF, LPV, RTVb, *s_2 has treatment events* 3TC, AZT, EFV, *and s_3 has treatment events* RT103N, FTC, TDF, NVP *and failure.*

Definition 2 *(Sequence feature). A sequence s^1 is a sequence feature of s^2 if there exists a strictly increasing element mapping such that each element of s^1 is a subset of its corresponding element of s^2.*

Example 3 {EFV}→{NVP} *is a sequence feature of the sequence in Example 1, since* {}⊆{FTC, TDF, LPV, RTVb}, {EFV}⊆{3TC, AZT, EFV} *and* {NVP}⊆ {RT103N, FTC, TDF, NVP, failure}, *where* {} *stands for the empty set.*

Given two sets of sequences corresponding to complete treatment records from two groups of patients (one experiencing a treatment success in the latest treatment and the other a treatment failure), we want to identify sequence features (treatment change episodes) that are observed frequently in one set, but infrequently in the other set.

2.3 Sequence boosting

Sequence boosting classifies the response to the current therapy based on all events in the patient's treatment record (viral genotypes, previously administered therapies, and short-term response to these therapies) under consideration of their order. In essence, the resulting method is a linear classifier using non-linear features. To this end, the *sequence boosting* method constructs a feature space progressively by adding a sequence feature in each iteration. We follow the LPBoost (Demiriz et al., 2002) approach in which the parameter vector is regularized w.r.t. the L1-norm (LASSO) resulting in most sequence features having zero weights. This is particularly useful in our case, since even if the whole sequence feature space is expensive to construct, we can disregard and skip adding the sequence features with zero-weights to the feature space. Let X be the sequence representation of the patient's treatment record, and s be an arbitrary sequence feature. We represent the presence or absence of s in X by an indicator function $I(X)$ that returns 1 if $s \in X$, and 0 otherwise. Our classifier is a linear combination of sequence features:

$$f(X) = \mathrm{sgn}\left(\sum_{s \in \mathscr{S}} \alpha_s I_s(X) \right) \tag{1}$$

where s is an instance of the complete sequence feature space $\mathscr{S}$ derived from the training set, and α_s is the corresponding weight to be learned.
In order to illustrate the function of the indicator function further, consider following example of a treatment history[2]:

$$\begin{aligned} X_1 \quad = \quad & \{\mathrm{FTC} \wedge \mathrm{TDF} \wedge \mathrm{LPV} \wedge \mathrm{RTVb}\} \rightarrow \{\mathrm{3TC} \wedge \mathrm{AZT} \wedge \mathrm{EFV}\} \\ & \rightarrow \{\mathrm{RT103N} \wedge \mathrm{FTC} \wedge \mathrm{TDF} \wedge \mathrm{NVP} \wedge \mathit{failure}\}, \end{aligned}$$

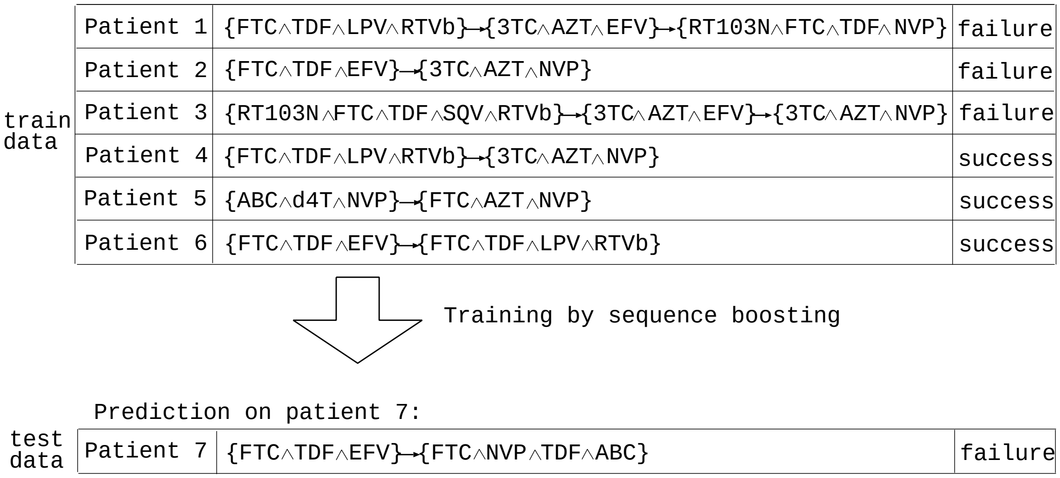
and the following examples for sequence features:

$$s_1 = \{\mathrm{EFV} \wedge \mathit{failure}\}, s_2 = \{\mathrm{NVP} \wedge \mathit{failure}\}, s_3 = \{\mathrm{EFV}\} \rightarrow \{\mathrm{NVP}\}.$$

The sequence features s_2 and s_3 occur in the treatment history X_1, but s_1 does not. Therefore $I_{s_2}(X_1) = I_{s_3}(X_1) = 1$ and $I_{s_1}(X_1) = 0$. Unlike in the conventional

[2] Figure 2 illustrates derivation of the sequence representation from the actual treatment history.

representation of features as fixed size vectors, this sequence representation allows us to compare treatment histories of different length. Thereby we circumvent the missing data problem. For instance, the treatment response to a past regimen is undoubtfully useful, but unfortunately not always measured and recorded. Figure 3 illustrates the training and evaluation procedure of *sequence boosting*.



train data			
	Patient 1	{FTC∧TDF∧LPV∧RTVb}→{3TC∧AZT∧EFV}→{RT103N∧FTC∧TDF∧NVP}	failure
	Patient 2	{FTC∧TDF∧EFV}→{3TC∧AZT∧NVP}	failure
	Patient 3	{RT103N∧FTC∧TDF∧SQV∧RTVb}→{3TC∧AZT∧EFV}→{3TC∧AZT∧NVP}	failure
	Patient 4	{FTC∧TDF∧LPV∧RTVb}→{3TC∧AZT∧NVP}	success
	Patient 5	{ABC∧d4T∧NVP}→{FTC∧AZT∧NVP}	success
	Patient 6	{FTC∧TDF∧EFV}→{FTC∧TDF∧LPV∧RTVb}	success

Prediction on patient 7:

test data			
	Patient 7	{FTC∧TDF∧EFV}→{FTC∧NVP∧TDF∧ABC}	failure

Figure 3: Schematic figure of training and evaluation of *sequence boosting*. We use patients whose past treatment records and latest treatment outcomes are available. In this example, we have two groups of patients whose recent treatment outcome are either failure (1,2,3) or success (4,5,6). While the data from these six patients are used for training the system, the information on patient 7 is reserved only for evaluation of the system. In this example, only one sequence feature EFV→ NVP is obtained during training, since presence or absence of this feature in treatment history perfectly discriminates one group from the other. Using this rule, the system predicts the recent treatment outcome of the patient 7 as failure since this patient also has experienced the treatment of EFV followed by NVP: EFV→NVP.

2.3.1 Problem formulation

Our objective function is the same as the one of LPBoost (Demiriz et al., 2002) (also known as the Linear Programming Machine (Schölkopf and Smola, 2002)), i.e. it maximizes the margin ρ:

$$
\begin{aligned}
max_{\alpha,\xi,\rho} \quad & \rho - D\sum_{i=1}^{n}\xi_i, && (2)\\
s.t. \quad & \sum_{s\in\mathscr{S}} y_i\alpha_s I_s(X) + \xi_i \geq \rho \quad i = 1\ldots n, && (3)\\
& \sum_{s\in\mathscr{S}} \alpha_s = 1, \quad \alpha \geq 0, \quad \xi \geq 0, && (4)
\end{aligned}
$$

where we set $D = \frac{1}{n\nu}$ according to Demiriz et al. (2002). This problem is hard to solve, due to the large number of sequence features and corresponding weights α. Therefore we consider a restricted problem on a subset of variables α: we start with an empty set of variables, and a new variable is added in each iteration. By the duality of linear programming, adding a variable in the primal problem is equivalent to adding a constraint in the dual problem. The former approach is often termed column generation, and the latter is referred to as constraint generation, or cutting-plane method. The dual problem of the above linear programming problem is:

$$
\begin{aligned}
min_{\lambda,\gamma} \quad & \gamma, && (5)\\
s.t. \quad & \sum_{i=1}^{n} y_i\lambda_i I_s(X_i) \leq \gamma \quad s \in \mathscr{S}, && (6)\\
& \sum_{\imath=1}^{n} \lambda_i = 1, \quad 0 \leq \lambda \leq D. && (7)
\end{aligned}
$$

From the dual point of view, only a subset of constraints is used to determine the current solution, and a new constraint that most strongly violates the constraint (6) is added to the solution set in each iteration. In our case, finding the most strongly violated constraint (constraint generation subproblem) is equivalent to finding the sequence feature s which maximizes the *gain function*:

$$
g(s) = \sum_{i=1}^{n} y_i\lambda_i I_s(X_i). \qquad (8)
$$

In the following subsection, we give a branch-and-bound algorithm called *Discriminative PrefixSpan* to solve this constraint generation subproblem. The *sequence boosting* algorithm terminates if the newly obtained constraint satisfies $g(s) \leq \hat{\gamma} + \varepsilon$, where ε is a parameter for early stopping, and $\hat{\gamma}$ is the value of the dual solution. As soon as the dual problem is feasible by taking $(\lambda, \hat{\gamma} + \varepsilon)$, then the weak duality of linear programming tells us that the value of the primal solution $\gamma^* = \rho - D\sum_{i=1}^{n}\xi_i$ is bounded from above by $\gamma^* < \hat{\gamma} + \varepsilon$. This suggests that even if we were to include all the sequence features, the value of the primal solution can be increased at most by ε. In all the experiments ε is set to 0.01. The pseudocode of *sequence boosting* is given in Algorithm 1.

Algorithm 1 *sequence boosting*

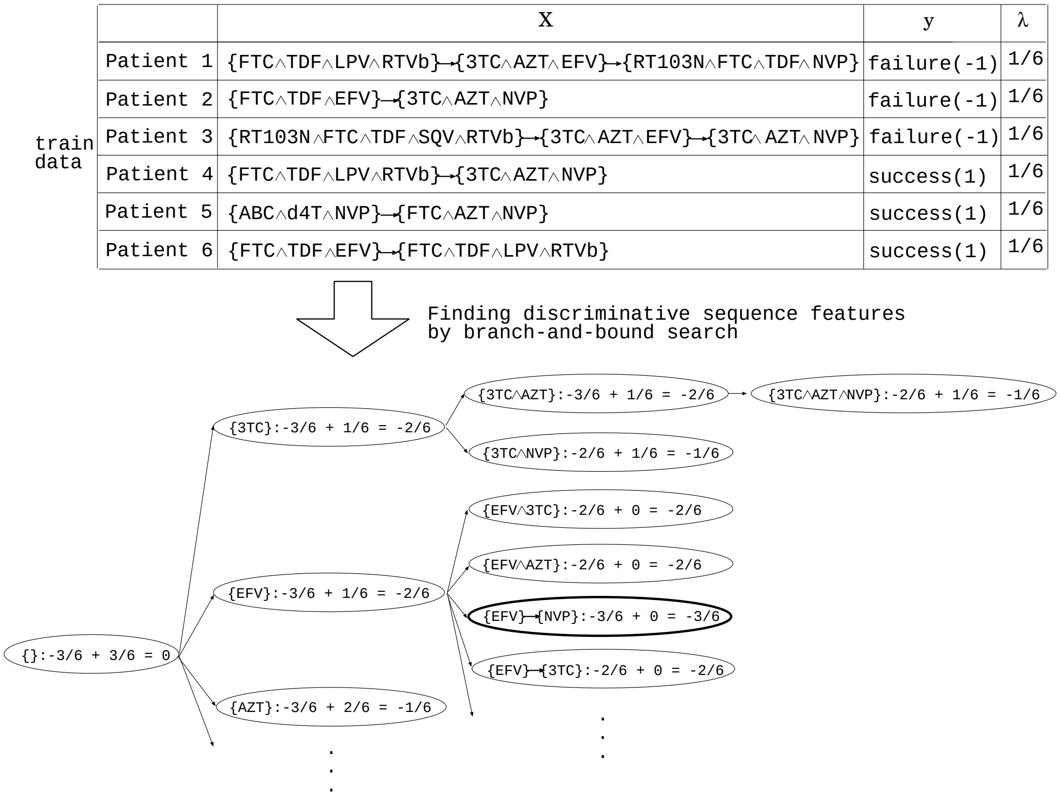
```
Require: $X, y, \nu, \varepsilon$
Ensure: $\alpha, H$
procedure sequence boosting
    $H^{(0)} = \emptyset$, $\lambda_i^{(0)} = 1/n$, level = 1
    loop
        $s^* \leftarrow \mathrm{Discriminative\,PrefixSpan}(\lambda, X, y)$
        if $\sum_{i=1}^{n} y_i \lambda_i 1_{s^*}(X_i) \leq \gamma + \varepsilon$ then
            break                                    ▷ No more sequence features
        end if
        $H \leftarrow H \cup \{s^*\}$                ▷ Update the sequence feature set
        $(\lambda, \alpha) \leftarrow$ Solve the dual problem (5),(6),(7)
    end loop
    return $(\alpha)$
end procedure
```

2.3.2 Searching for sequence features

A straightforward approach to finding the sequence feature which maximizes the gain function (8) out of all possible sequence features is to first enumerate all sequence features using PrefixSpan (Pei et al., 2004), then compute the gain (8) for each of them, and at the end take the maximum. In the PrefixSpan approach, the sequence features are organized in a search tree. The root node of the tree is an empty set, and each node is extended by adding one item to the sequence of its parent node. At each node, the frequency (number of occurrences in the dataset: support) of the sequence feature is counted, and used for tree pruning.

In our case, instead of the frequency, the gain function is computed at each node, since we are interested in discriminative sequence features rather than frequent ones (Figure 4). This tree is traversed by one of the basic search algorithms such as depth first search (DFS), breath first search (BFS), or the A^* algorithm. Here, we used a variant of A^*, and deepened the tree depth gradually (Nowozin et al., 2008). This strategy makes pruning effective when there are short high-gain sequence features.

For increasing efficacy, the size of the search tree has to be limited: suppose that one has traversed the tree up to the node with a sequence feature s, and the maximum gain found so far is g_{cur}. If there are no sequence features among the supersets s' of the feature s which exceed the gain g_{cur}, then one can quit the traversal at this point and prune the downstream part of the tree. In order to judge whether the tree can be pruned at s, we use the following theorem (Morishita, 2001; Kudo et al., 2005):



train data

	X	y	λ
Patient 1	{FTC∧TDF∧LPV∧RTVb}→{3TC∧AZT∧EFV}→{RT103N∧FTC∧TDF∧NVP}	failure(-1)	1/6
Patient 2	{FTC∧TDF∧EFV}→{3TC∧AZT∧NVP}	failure(-1)	1/6
Patient 3	{RT103N∧FTC∧TDF∧SQV∧RTVb}→{3TC∧AZT∧EFV}→{3TC∧AZT∧NVP}	failure(-1)	1/6
Patient 4	{FTC∧TDF∧LPV∧RTVb}→{3TC∧AZT∧NVP}	success(1)	1/6
Patient 5	{ABC∧d4T∧NVP}→{FTC∧AZT∧NVP}	success(1)	1/6
Patient 6	{FTC∧TDF∧EFV}→{FTC∧TDF∧LPV∧RTVb}	success(1)	1/6

Finding discriminative sequence features
by branch-and-bound search
{}:-3/6 + 3/6 = 0
{3TC}:-3/6 + 1/6 = -2/6
{EFV}:-3/6 + 1/6 = -2/6
{AZT}:-3/6 + 2/6 = -1/6
{3TC∧AZT}:-3/6 + 1/6 = -2/6
{3TC∧AZT∧NVP}:-2/6 + 1/6 = -1/6
{3TC∧NVP}:-2/6 + 1/6 = -1/6
{EFV∧3TC}:-2/6 + 0 = -2/6
{EFV∧AZT}:-2/6 + 0 = -2/6
{EFV}→{NVP}:-3/6 + 0 = -3/6
{EFV}→{3TC}:-2/6 + 0 = -2/6

Figure 4: A treatment history in sequence representation (top) and its corresponding search tree (bottom). In each oval of the search tree, a sequence feature s and the derivation of its corresponding gain ($g(s) = \sum_{i=1}^{6} y_i \lambda_i I_s(X_i)$) are shown. This example shows the exploration of the search tree at the first iteration of *sequence boosting* where the λs are initialized as uniform values. In this case the search tree explores sequence features which appear frequently in patients 1 to 3 (whose treatments turned out to be a failure), but infrequently in patients 4 to 6 (whose treatments turned out to be successful), or vice versa. The most discriminative sequence feature $\{\text{EFV} \to \text{NVP}\}$ is marked in the bold oval.

Theorem 1 *For any sequence feature s' such that $s \subseteq s'$, $g(s') < g_{cur}$, if*

$$\max\left\{ \sum_{\{i|s\subseteq x_i, y_i=0\}} \lambda_i, \sum_{\{i|s\subseteq x_i, y_i=1\}} \lambda_i \right\} < g_{cur}. \qquad (9)$$

In terms of time complexity *sequence boosting* is NP-hard due to the NP-hardness of its subproblem: PrefixSpan. The scalability of our method depends on the size of the dataset and its density. If the data are dense, then the output length of the PrefixSpan increases and the total speed decreases quickly. However, our data representation is sparse, i.e., most of the sequence features appear much less frequently than the size of the dataset *n*. Thus intersection operations on a search tree can efficiently prune candidates, and we do not need to search for very long sequence features. This sparseness helps us exploring a space which at first glance seems intractably large.

2.4 Computational experiments

The first computational experiment focused on the comparison of *sequence boosting* with logistic regression and SVMs using linear and nonlinear kernels to investigate the discriminative power of the sequence features. Since the space of sequence features is too large to be utilized as feature vector for SVMs and logistic regression, the patient's treatment record has to be encoded in a different way[3]. In the second experiment, we compared which one of the treatment information - therapy, genotype, or therapy outcome - is important for the prediction. Furthermore, *sequence boosting* allows easy access to discriminative treatment information. Thus, in the final experiment, we filtered and interpreted selected sequence features retrieved by *sequence boosting*.

For improving interpretability, we encoded treatment information of the current therapy with different integers than treatment information from past therapies. Additionally, we used three special indicators representing the absence of all PIs, all NRTIs, or all NNRTIs denoted as NoPIs, NoNRTIs, or NoNNRTIs, respectively. Furthermore, it is desirable that each treatment contains more than one treatment event, since a larger set with more treatment events such as $\{\text{TDF} \wedge \textit{success}\}$ is more informative than just $\{\textit{success}\}$. However, it could be disadvantageous for the classification performance to simply set the minimum number of events in a treatment too large. The reason is that a treatment with more events are less frequent in the training set, and can represent only smaller fraction of the data. Therefore, in an initial experiment, we varied the parameter of the model, which controls the minimum number of events in a treatment and observed its influence on the prediction accuracy. The optimal value identified in this initial experiment was used for all remaining experiments.

2.4.1 Comparison with other methods

In this experiment we compare *sequence boosting* with SVMs and logistic regression. For SVMs and logistic regression, we provided a feature vector of fixed length

[3] SVMs with non-linear kernels can make use of the same amount of information. Extracting discriminative rules from non-linear kernels, however, is challenging. Thus, we do not pursue this interesting direction for engineering a suitable kernel.

that has i) binary indicators for drugs and mutations of the current treatment, ii) integer values counting the frequencies of the drugs in past treatments, iii) integers representing the number of times a mutation was observed in previous genotypes, and iv) three additional numerical variables, which indicate the number of successes and failures in the past as well as the total number of treatment changes, respectively. This feature representation extends the ones used in Bickel et al. (2008) and Rosen-Zvi et al. (2008). Additionally, since it is a more common representation, we used binary indicators instead of frequencies for previously used drugs and observed mutations. Thus, the representation provides the same quantitative information as the sequence features but without their order, i.e. the sequence information. SVMs with polynomial kernels of degree 2 and 3 correspond to the interaction features in Rosen-Zvi et al. (2008).

We prepared three datasets depending on the number of treatment changes: i) patients with #TC ≥ 10, ii) patients with #TC ≥ 5, and iii) patients with #TC ≥ 1. The baseline accuracy, i.e. frequency of successful therapies of these settings is 0.783, 0.832 and 0.851, respectively. There were 3,759 patients in total, 1,830 with #TC≥ 5 and 646 with #TC≥ 10.

For evaluation, we performed 10-fold cross-validation where, in each fold, 80% of the data were used for training, 10% were used for adjusting the regularization parameter, and the other 10% were used solely for the performance assessment. For SVM and *sequence boosting*, the regularization parameter ν, which controls the balance between overfitting and underfitting, was chosen from $\{0.1, 0.2, \ldots, 0.6\}$. For estimating the baseline performance in terms of accuracy, we employ two different predictors: one that performs random guesses according to the prevalence of the classes (guess), and one that always predicts the majority class (majority). For the former the performance is computed as $x^2 + (1-x)^2$, with x being the frequency of successful therapies. This baseline with background probabilities has previously been employed by Baldi et al. (2000).

Performance of the classifiers is assessed in both, accuracy and area under the receiver operating characteristics (ROC) curve (AUC). In order to test for significance of the observed improvement of *sequence boosting* over the other methods, we employed a Wilcoxon rank-sum test with a 5% significance level.

2.4.2 Comparison of different treatment information

In principle, there are three types of treatment information: genotype (GT), therapy (TH) and therapy outcome (TO). For addressing the question which treatment information or which interactions of treatment information are most informative for the classification, the training data were restricted to the following settings: i) only genotype information (GT), ii) only therapy information (TH), iii) only therapy outcome information (TO), and their two-way combinations, i.e. iv) therapy outcome and genotype, v) therapy and therapy outcome, and vi) therapy and genotype. We compared all of these combinations using *sequence boosting* via 10-fold cross-validation (using the same setup as described in the previous section). In order to ensure a fair comparison of the different treatment information with the baseline method, i.e. predictions based on the genotype preceding the therapy, we selected a subset of patients with a genotype attributed to the current treatment. After this restriction 446 patients remained and the corresponding baseline accuracy (frequency of successful therapies) was 0.774.

2.4.3 Identifying discriminative sequence features

In general, the set of features selected by a given feature selection method depends on the training set. The LASSO feature selection employed by *sequence boosting* poses no exception of this fact (Zou, 2006). For ensuring the robustness of the selected sequence features, we employed bootstrapping on the dataset comprising patients with 10 or more past treatments (646 patients). A *sequence boosting* model was trained for each of 1000 bootstrap replicates (with replacement) of the original dataset. We filtered out those sequence features with i) frequency ≤ 100, ii) importance ≤ 100 and iii) p-value ≥ 0.1. Here, frequency denotes the number of times that the sequence feature was selected from a bootstrap sample. Importance of a sequence feature was computed by dividing its mean weight by the variance of its weights computed in the bootstrap repetitions. The p-values were computed using a Fisher's exact test based on a contingency table (sequence feature vs. therapy outcome).

Furthermore, for computing an interaction p-value between the treatment outcome (TO) and treatment and genotype information (TH and GT), we use the likelihood ratio test. Briefly, the likelihood ratio test compares two logistic regression models: one with and the other without an interaction term. P-values are computed in such a way that under the null hypothesis of no difference, the deviance (log-likelihood difference between the two models) follows a χ^2-distribution (Bickel and Doksum, 2002).

3 Results

Table 1 shows that the classification performance decreased as we increased the minimum number of events in a treatment, even though the difference was not significant at a 5% level by the Wilcoxon rank-sum test after Bonferroni correction.

Thus, the minimum number of events in a treatment was set to 3 for all computational experiments, since a therapy consisting of only one drug has at least 3 feature components due to the introduction of absence indicators: NoPIs, NoNRTIs and NoNNRTIs. Of note, the performance values obtained with a minimum sequence features size of 3 showed the least variance.

Table 1: Difference in performance (ACC and AUC) with respect to the minimum size of the sequence feature. The parameter that controls the minimal sequence feature size was varied. Performance was assessed in 10-fold cross-validation on the dataset where patients have 10 or more treatments (#TC $\geq$ 10).

minimum sequence feature size	ACC	AUC
1	0.808 ± 0.0414	0.748 ± 0.0520
2	0.788 ± 0.0493	0.741 ± 0.0677
3	0.786 ± 0.0196	0.741 ± 0.0411
4	0.781 ± 0.0368	0.703 ± 0.0961
5	0.768 ± 0.0497	0.700 ± 0.1010

3.1 Comparison with other methods

Tables 2 and 3 list the prediction performance (measured in terms of accuracy (ACC) and area under the ROC curve (AUC)) of various methods in the three different settings. It can be observed that *sequence boosting* constantly outperforms other classifiers. Results that exhibit a statistically significant difference are marked with •. Performance achieved with the binary encoding for historic drugs and mutations was in general inferior to the integer encoding (see Tables A2 and A3 in the Appendix). Improvements over the baseline in terms of ACC are quite pronounced for the predictor that performs random guesses; compared to the majority predictor the improvements (if at all) are for all machine learning approaches only negligible. The rightmost column of Tables 2 and 3 indicates that non-linear methods (*sequence boosting* and SVM with polynomial and RBF kernels) perform better than linear methods (logistic regression and SVM with linear kernel), which motivates us to identify non-linear sequence features.

Table 2: Comparison of different classification methods in terms of accuracy. Results where *sequence boosting* outperforms other methods with statistically significant difference are marked with ●.

Method	#TC ≥10	#TC ≥5	#TC ≥1	mean
baseline (guess)	0.660	0.720	0.746	-
baseline (majority)	0.783	0.832	0.851	-
sequence boosting	0.808 ± 0.0414	0.822 ± 0.0140	0.851 ± 0.0046	0.827 ± 0.0219
SVM poly. ($d = 2$)	●0.768 ± 0.0479	0.811 ± 0.0301	●0.832 ± 0.0188	0.804 ± 0.0326
SVM poly. ($d = 3$)	●0.760 ± 0.0520	0.808 ± 0.0276	●0.836 ± 0.0136	0.801 ± 0.0384
SVM RBF	●0.783 ± 0.0064	0.824 ± 0.0048	●0.847 ± 0.0052	0.818 ± 0.0324
SVM linear	0.780 ± 0.0421	●0.805 ± 0.0203	●0.772 ± 0.0210	0.786 ± 0.0172
Logistic regression	●0.700 ± 0.0556	●0.780 ± 0.0302	●0.838 ± 0.0118	0.773 ± 0.0693
mean	0.758 ± 0.0408	0.806 ± 0.0226	0.825 ± 0.0141	-

Table 3: Comparison of different classification methods in terms of the area under the ROC curve (AUC). Results where *sequence boosting* outperforms other methods with statistically significant difference are marked with ●.

Method	#TC ≥10	#TC ≥5	#TC ≥1	mean
baseline	0.500	0.500	0.500	-
sequence boosting	0.748 ± 0.0520	0.713 ± 0.0620	0.686 ± 0.0323	0.716 ± 0.0311
SVM poly. ($d = 2$)	0.702 ± 0.0863	0.699 ± 0.0444	●0.653 ± 0.0340	0.685 ± 0.0275
SVM poly. ($d = 3$)	●0.684 ± 0.0824	0.711 ± 0.0440	0.667 ± 0.0373	0.687 ± 0.0222
SVM RBF	0.729 ± 0.0107	0.684 ± 0.0458	●0.660 ± 0.0233	0.681 ± 0.0501
SVM linear	0.731 ± 0.0941	0.670 ± 0.0766	●0.601 ± 0.0523	0.667 ± 0.0680
Logistic regression	●0.567 ± 0.0673	●0.616 ± 0.0530	0.676 ± 0.0490	0.620 ± 0.0546
mean	0.682 ± 0.0682	0.676 ± 0.0528	0.651 ± 0.0392	-

3.2 Comparison of different treatment information

Table 4 shows the results on the comparison of different treatment information. In terms of accuracy no difference between the different treatment information is visible. From the AUC column, however, we can observe that a single type of treatment information (i.e. only TO, only GT or only TH) failed to rank patients correctly. For achieving a reasonable AUC performance, it was necessary to use the therapy information (TH) in combination with either the therapy outcome information (TO) or the genotypic information (GT).

Table 4: Comparison of different features in terms of accuracy (ACC) and AUC. TO (Therapy Outcome) indicates the success/failure of a previous treatment, GT (Genotype) indicates the presence/absence of the 108 mutations, TH (Therapy) indicates the drug compounds.

Feature set	ACC	AUC
TO	0.778 ± 0.0004	0.520 ± 0.0078
GT	0.756 ± 0.0016	0.549 ± 0.0059
TH	0.733 ± 0.0001	0.617 ± 0.0039
TO + GT	0.760 ± 0.0009	0.533 ± 0.0079
TH + TO	0.759 ± 0.0024	0.752 ± 0.0074
TH + GT	0.776 ± 0.0022	0.756 ± 0.0042
TH + GT + TO	0.783 ± 0.0013	0.737 ± 0.0051

Table 5: A list of interesting sequence features. Sequence Features in the upper rows are associated with success of the current treatment. Sequence Features in the lower rows are associated with failure of the current treatment. Bold font indicates features from the current regimen. Interaction p-values (fifth column) were only computed for sequence features comprising *success* or *failure*.

importance	frequency	p-value	sequence feature	$\{$TH, GT$\} \times$ TO
293	117	3.47e-06	$\{$d4T $\wedge$ NFV $\wedge$ success$\}$	1.33e-3
290	128	3.47e-05	$\{$LPV $\wedge$ RTVb $\wedge$ success$\}$	0.10
163	415	6.75e-02	$\{$**3TC** $\wedge$ **DRV** $\wedge$ **NoNNRTIs**$\}$	-
161	150	1.83e-02	$\{$**3TC** $\wedge$ **RTVb** $\wedge$ **DRV**$\}$	-
-128	231	1.89e-08	$\{$RT210W $\wedge$ RT215Y $\wedge$ failure$\}$	5.36e-4
-193	402	2.64e-06	$\{$d4T $\wedge$ NFV $\wedge$ failure$\}$	0.01
-209	411	3.61e-09	$\{$LPV $\wedge$ RTVb $\wedge$ failure$\}$	2.10e-4
-278	374	8.75e-05	$\{$AZT $\wedge$ NoNNRTIs $\wedge$ NoPIs$\}$ $\rightarrow$ $\{$ddI $\wedge$ NoNNRTIs $\wedge$ NoPIs$\}$	-

3.3 Identifying discriminative sequence features

In total we obtained 8215 unique sequence features from the 1000 bootstrap replicates. After filtering, however, only 171 sequence features remained (listed in Table A4 in the Appendix). Table 5 depicts the subset of all discovered sequence features that are discussed in the next section. Sequence features in bold font denote the inclusion of treatment information from the current therapy, i.e. the one whose response has to be inferred. Sequence features that include drugs from the current therapy are especially interesting, since the outcome of a therapy is a direct response to the drugs in the prescribed regimen. It is likely that the past treatment has an effect on the current treatment. The leftmost column shows the importance of the variables, the second column the frequency of the sequence features in the 1000 bootstrap repetitions, and the third column shows the p-value computed by a

Fisher's exact test. The fourth column shows the obtained sequence features. The rightmost column shows the interaction p-values between the therapy and genotype information (TH and GT) and the therapy outcome information (TO) calculated by a likelihood ratio test. Of note, in many of the sequence features in which TH and GT appear together with TO in the same treatment, interaction p-values were smaller than the 5% significance threshold.

4 Discussion

The comparison with other methods showed that *sequence boosting* achieved superior performance compared to both, logistic regression and SVMs (with linear and non-linear kernels). Furthermore, focusing on the mean performance of the three settings, we can observe that the mean accuracy increases as we select patients with fewer treatment changes. This is simply due to the fact that patients, who are in an early stage of the disease, can be treated with a higher success rate. This fact is also reflected by the success rate (baseline and guess) in the three datasets and visualized in Figure 1. Thus, for the remaining two computational experiments, we focused on the more challenging setting: predicting the treatment outcomes of patients who have experienced many treatment changes (#TC $\geq$ 10).

When comparing different treatment information, it turned out that for achieving optimal performance short-term therapy outcome (TO) has to be provided either with genotypic information (GT) or therapy information (TH). Interestingly, the combination TH+TO performed as well as the combination TH+GT. This suggests that information on the therapy history (TH) is dependent on the previous treatment outcomes or the mutations in past and present genotypes. It is likely that interactions between these treatment information are crucial for the performance. Further evidence for this hypothesis is provided by analysis of important sequence features. More precisely, the observed small interaction p-values between treatment outcome and remaining therapy information (TH and GT) may explain the sharp increase of the classifier performance when TH is combined with TO or GT. For example, the sequence feature $\{\text{d4T} \wedge \text{NFV} \wedge \textit{success}\}$ with the interaction p-value $1.33{\times}10^{-3}$ and importance 293 suggests that the usage of drug d4T with NFV in the past itself is not predictive. However, once this treatment has experienced a short-term success in the past, then this indicates a high probability of a successful current treatment. Such interactions are investigated in more detail in the following.

4.1 Interpretation of sequence features

The *sequence boosting* method generates a linear classifier using non-linear features. This facilitates interpretation of features and provides an understanding of how the classifier works. In this section we closely look at the interesting sequence features listed in Table 5.

Sequence boosting identified usage of DRV (darunavir) in the current regimen (indicated by bold font) as a supporting factor for a successful regimen ($\{\mathbf{3TC}\wedge\mathbf{DRV}\wedge\mathbf{NoNNRTIs}\},\{\mathbf{RTVb}\wedge\mathbf{DRV}\wedge\mathbf{NoNNRTIs}\}$). In fact, DRV has been recently approved by the US Food and Drug Administration (FDA) and is recommended for treating heavily treatment-experienced patients. Combination therapies using DRV are among the most potent regimens currently available. A sequence feature that is frequently selected and linked to treatment failure comprises mutations RT210W and RT215Y in one of the available patient's genotypes $\{\mathrm{RT210W}\wedge\mathrm{RT215Y}\wedge\textit{failure}\}$. These two mutations are thymidine analog mutations and confer resistance against all NRTIs. Since NRTIs are part of almost every combination therapy the presence of these mutations seriously limits the effectiveness of the current regimen. An example for a sequence feature representing a treatment switch is $\{\mathrm{AZT}\wedge\mathrm{NoNNRTIs}\wedge\mathrm{NoPIs}\}\rightarrow\{\mathrm{ddI}\wedge\mathrm{NoNNRTIs}\wedge\mathrm{NoPIs}\}$. This sequence feature is related to failure of the current regimen as indicated by the negative importance value. A switch from an NRTI-only regimen to another NRTI-only regimen was only common in the pre-HAART area. Therefore, this sequence feature is associated with patients that had a high number of treatment switches (#TC) and are therefore less likely to be successfully treated.

Sequence boosting cannot only recover well-known facts. A large number of important sequence features include the short-term treatment outcome of previous regimens. The fact that response to a previous regimen affects the current regimen is not obvious at first. However, if a patient failed a regimen comprising a potent drug (e.g. LPV) after only a short time, then there were enough mutations in the predominant viral variant or viral reservoirs to cause failure of the drug. In contrast, if the treatment was successful, then the patient's virus did not acquire enough resistance mutations to impair the drug (yet). As an example we compared the resistance mutations in patients that have the sequence feature $\{\mathrm{LPV}\wedge\mathrm{RTVb}\wedge\textit{success}\}$ (73 patients) with patients with the sequence feature $\{\mathrm{LPV}\wedge\mathrm{RTVb}\wedge\textit{failure}\}$ (66 patients). Protease mutations were more prevalent in the *failure* group than in the *success* group. The observed enrichment is significant according to a paired Wilcoxon rank-sum test on frequencies of single mutations (p-value=4.49×10^{-10}). Moreover, LPV associated resistance mutations (Johnson et al., 2008) are significantly more enriched than other protease mutations (p-value=1.01×10^{-3} using a one-sided Wilcoxon rank sum test). Figure 5 shows the corresponding detailed histogram of frequencies for all mutations.

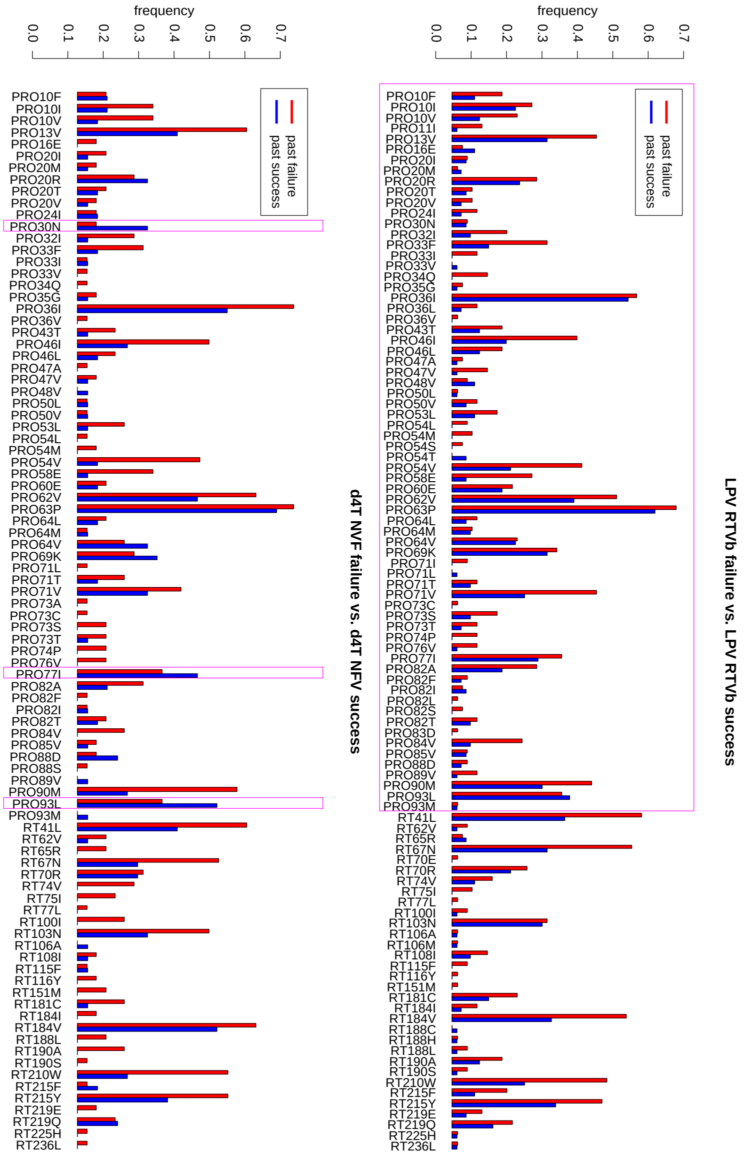


Figure 5: Mutation frequencies in the most recent genotype after therapies including {d4T∧NFV}(left) or {LPV∧RTVb}(right). (left) Patients that experience a treatment failure in general had protease mutations that confer resistance against many PIs including NFV. Interestingly, in patients that were successfully treated with NFV, the mutation PRO30N (highlighted in a box) was enriched. This mutation confers high-level resistance only against NFV, and it can therefore be expected to have developed during the course of the successful treatment. The same holds true for the NFV related mutations PRO77I and PRO93L. (right) Among patients who have experienced a treatment with boosted LPV, protease mutations (highlighted in a box) were more prevalent in the *failure* group than in the *success* group. This explains the higher risk of failing the recent treatment for the failure group, since resistance mutations to PIs have been already stored in the reservoir during or before the failure of the past treatment with boosted LPV.

Another example of this phenomenon is provided by patients that were treated in the past with d4T and NFV and either had a successful (n=33) $\{$d4T$\wedge$NFV$\wedge$*success*$\}$ or failing (n=35) therapy $\{$d4T$\wedge$NFV$\wedge$*failure*$\}$. Patients that experience a treatment failure, in general, had protease mutations that confer resistance against many PIs including NFV. Interestingly, in patients that were successfully treated with NFV, the mutation PRO30N was enriched. This mutation confers high-level resistance only against NFV, and it can therefore be expected to have developed during the course of the successful treatment. The same holds for mutations PR77I and PR93L, which are mainly associated with NFV resistance.

The application of *sequence boosting* to the problem revealed that previous exposure to a drug is not the crucial information for predicting response to a new regimen. It is more important to know whether drugs were part of a successful or failing past regimen to draw conclusions about drug resistance especially in the absence of genotypic information.

5 Conclusion

Sequence boosting combines an optimization technique with a sequence mining method. In contrast to SVMs with non-linear kernels, the generated models are interpretable, which enables clinicians to reason about the obtained discriminative sequence features. In order to improve confidence in the classifier's decision it is valuable to open the "black box" and analyze on what evidence the classifier's decision is based. In our computational experiments we found that sequence features based on information on the treatment history perform well especially for patients with many treatment changes. By studying the significance of interactions, our approach revealed that information on past treatments paired with their short term outcome is as valuable as the treatment information paired with genotypic information. The successful results on a large HIV data encourages us to apply the same tool to a broader range of clinical time series problems.

Appendix

Table A1: Overview of considered treatment information. “Mutations” column shows the list of mutations used in this work. In the “Drugs” column, the drugs in the same group share the same mode of action against the same molecular target. Drug groups are Protein Inhibitors (PIs), Nucleotide Reverse Transcriptase Inhibitors (NRTIs), Non-Nucleotide Reverse Transcriptase Inhibitors (NNRTIs), and Fusion Inhibitors (FIs).

Mutations	Drugs	Treatment Outcome
· Reverse transcriptase mutations	· NRTIs	· Success
41L, 62V, 65R, 67N, 69i, 70E, 70R, 74V, 75I, 77L, 100I, 103N, 106A,106M, 108I, 115F, 116Y, 151M, 181C, 181I, 184I, 184V, 188C, 188H, 188L, 190A, 190S, 210W, 215F, 215Y, 219E, 219Q, 225H, 236L	Lamivudine (3TC), Abacavir (ABC), Zidovudine (AZT), Stavudine (d4T), Zalcitabine (ddC), Didanosine (ddI), Tenofovir (TDF), Emtricitabine (FTC)	If viral load drops i) below 400 copies/ml, or ii) two magnitude from the treatment start.
	· NNRTIs	· Failure
· Protease mutations	Delavirdine (DLV), Efavirenz (EFV), Nevirapine (NVP) Etravirine (TMC125)	If no success.
10C, 10F, 10I, 10R, 10V, 11I, 13V, 16E, 20I, 20M, 20R, 20T, 20V, 24I, 30N, 32I, 33F, 33I, 33V, 34Q, 35G, 36I, 36L, 36V, 43T, 46I, 46L, 47A, 47V, 48V, 50L, 50V, 53L, 53Y, 54A, 54L, 54M, 54S, 54T, 54V, 58E, 60E, 62V, 63P, 64L, 64M, 64V, 69K, 71I, 71L, 71T, 71V, 73A, 73C, 73S, 73T, 74P, 76V, 77I, 82A, 82F, 82I, 82L, 82S, 82T, 83D, 84V, 85V, 88D, 88S, 89V, 90M, 93L, 93M	· PIs Amprenavir (APV), Atazanavir (ATV), Indinavir (IDV) Lopinavir (LPV), Nelfinavir (NFV), Ritonavir (RTV), boosted dose Ritonavir (RTVb), Saquinavir (SQV), Fosamprenavir (FPV), Tipranavir (TPV), Darunavir (DRV)	
	· FIs	
	Enfuvirtide (T20)	

Table A2: Comparison of different classification methods (in binary encoding) in terms of accuracy. As a feature for comparing methods, we used binary indicators of drugs and mutations in the past and present. Results where *sequence boosting* outperforms with statistically significant difference are marked with •.

Method	#TC ≥10	#TC ≥5	#TC ≥1	mean
baseline (guess)	0.660	0.720	0.746	-
baseline (majority)	0.783	0.832	0.851	-
sequence boosting	0.808 ± 0.0414	0.822 ± 0.0140	0.851 ± 0.0046	0.827 ± 0.0219
SVM poly. ($d = 2$)	•0.779 ± 0.0317	0.816 ± 0.0205	•0.835 ± 0.0179	0.810 ± 0.0234
SVM poly. ($d = 3$)	0.782 ± 0.045	0.811 ± 0.0179	•0.838 ± 0.0166	0.810 ± 0.0265
SVM RBF	•0.782 ± 0.0077	0.823 ± 0.0055	•0.844 ± 0.0047	0.816 ± 0.0597
SVM linear	•0.777 ± 0.0375	•0.802 ± 0.0162	•0.704 ± 0.0194	0.761 ± 0.0244
Logistic regression	•0.678 ± 0.0762	•0.750 ± 0.0144	•0.835 ± 0.0144	0.754 ± 0.035
mean	0.760 ± 0.0396	0.800 ± 0.0149	0.811 ± 0.0146	-

Table A3: Comparison of different classification methods (in binary encoding) in terms of the area under the ROC curve (AUC). As a feature for comparing methods, we used binary indicators of drugs and mutations in the past and present. Results where *sequence boosting* outperforms with statistically significant difference are marked with •.

Method	#TC ≥10	#TC ≥5	#TC ≥1	mean
baseline	0.500	0.500	0.500	-
sequence boosting	•0.748 ± 0.0520	•0.713 ± 0.0620	•0.686 ± 0.0323	0.716 ± 0.0311
SVM poly. ($d = 2$)	•0.674 ± 0.0709	•0.665 ± 0.0528	•0.637 ± 0.0259	0.659 ± 0.0499
SVM poly. ($d = 3$)	•0.651 ± 0.0848	•0.651 ± 0.0595	•0.642 ± 0.0331	0.648 ± 0.0591
SVM RBF	•0.704 ± 0.0705	•0.672 ± 0.0519	0.646 ± 0.0261	0.674 ± 0.0495
SVM linear	•0.674 ± 0.0628	•0.617 ± 0.0668	•0.555 ± 0.0615	0.615 ± 0.0637
Logistic regression	•0.484 ± 0.1178	•0.65 ± 0.0543	•0.65 ± 0.0543	0.595 ± 0.0755
mean	0.638 ± 0.0814	0.651 ± 0.0571	0.626 ± 0.0402	-

Table A4: A list of patterns with high importance obtained by bootstrapping. Patterns in the upper rows are associated with success of the current treatment. Patterns in the lower rows are associated with failure of the current treatment. IMP stands for importance, FREQ stands for frequency. The last column shows the interaction p-values by likelihood ratio test between the therapy feature (TH) or genotype (GT) and the therapy outcome feature (TO). Bold font indicates features from the current regimen. Interaction p-values (fifth column) were only computed for patterns comprising *success* or *failure*.

IMP	FREQ	p-value	pattern	{TH, GT} × TO
-362	133	5.16e-04	{AZT ∧ IDV ∧ NoNNRTIs}	-
-349	199	1.18e-02	{3TC ∧ d4T ∧ NoNNRTIs}	-
-339	134	7.60e-02	{AZT ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ ddI ∧ NoNNRTIs ∧ NoPIs}	-
-325	159	8.59e-04	{ddI ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ ddI ∧ NoNNRTIs}	-
-305	189	9.61e-03	{AZT ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ NoNNRTIs ∧ NoPIs}	-
-303	194	4.21e-03	{IDV ∧ NoNNRTIs ∧ failure}	0.62
-299	220	2.92e-02	{**TDF** ∧ **RTVb** ∧ **NoNNRTIs**}	-
-288	120	1.58e-07	{TDF ∧ ATV ∧ NoNNRTIs}	-
-278	374	8.75e-05	{AZT ∧ NoNNRTIs ∧ NoPIs} → {ddI ∧ NoNNRTIs ∧ NoPIs}	-
-277	192	5.92e-04	{**ddI** ∧ **RTVb** ∧ **NoNNRTIs**}	-
-276	145	4.91e-05	{3TC ∧ LPV ∧ RTVb ∧ NoNNRTIs}	-
-276	281	5.87e-02	{3TC ∧ d4T ∧ NoPIs}	-
-265	326	7.18e-04	{ddI ∧ RTVb ∧ NoNNRTIs}	-
-263	142	1.84e-02	{TDF ∧ ATV ∧ RTVb}	-
-263	174	2.44e-05	{AZT ∧ NoNNRTIs ∧ NoPIs} → {ddI ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ NoNNRTIs ∧ NoPIs}	-
-262	366	2.57e-06	{3TC ∧ NoNNRTIs ∧ failure}	0.06
-260	116	6.87e-02	{IDV ∧ NoNNRTIs ∧ success} → {3TC ∧ IDV ∧ NoNNRTIs}	0.58
-260	152	7.52e-04	{d4T ∧ ddI ∧ NoNNRTIs} → {3TC ∧ d4T ∧ NoNNRTIs}	-
-258	158	8.27e-02	{AZT ∧ LPV ∧ RTVb}	-
-257	158	7.50e-10	{3TC ∧ RTVb ∧ failure}	7.97e-5
-252	226	3.74e-03	{ddI ∧ TDF ∧ failure}	0.87
-252	425	1.32e-04	{ddI ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ NoNNRTIs ∧ NoPIs}	-
-251	438	3.90e-06	{LPV ∧ RTVb ∧ NoNNRTIs}	-
-248	831	2.07e-06	{ddI ∧ NoNNRTIs ∧ failure}	0.01
-248	118	1.23e-02	{3TC ∧ NoNNRTIs ∧ failure} → {LPV ∧ RTVb ∧ NoNNRTIs}	4.15e-3
-247	504	4.90e-02	{3TC ∧ NVP ∧ NoPIs}	-
-246	180	5.13e-02	{AZT ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ SQV ∧ NoNNRTIs}	-
-242	581	1.67e-05	{3TC ∧ RTVb ∧ NoNNRTIs}	-
-241	138	2.55e-05	{3TC ∧ TDF ∧ NoNNRTIs} → {**TDF** ∧ **RTVb** ∧ **NoNNRTIs**}	-
-241	423	2.10e-02	{ABC ∧ d4T ∧ NoNNRTIs}	-
-239	597	2.01e-05	{d4T ∧ SQV ∧ NoNNRTIs}	-
-239	371	1.36e-05	{IDV ∧ RTVb ∧ NoNNRTIs}	-
-235	780	1.99e-02	{ddC ∧ NoNNRTIs ∧ NoPIs}	-
-234	184	1.78e-03	{3TC ∧ d4T ∧ NFV}	-
-231	116	6.91e-02	{AZT ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ ddI ∧ NoNNRTIs}	-
-230	147	3.44e-02	{**TDF** ∧ **FTC** ∧ **RTVb**}	-
-230	133	4.81e-06	{AZT ∧ ddC ∧ NoNNRTIs}	-
-227	215	3.54e-02	{3TC ∧ ABC ∧ failure}	0.02
-227	496	1.03e-03	{3TC ∧ ABC ∧ NoNNRTIs}	-
-225	566	1.13e-05	{d4T ∧ RTVb ∧ NoNNRTIs}	-
-224	148	3.05e-03	{3TC ∧ RTVb ∧ NoNNRTIs} → {**3TC** ∧ **RTVb** ∧ **NoNNRTIs**}	-
-217	281	6.91e-03	{LPV ∧ RTVb ∧ NoNNRTIs} → {**3TC** ∧ **RTVb** ∧ **NoNNRTIs**}	-
-216	153	5.40e-02	{3TC ∧ AZT ∧ NoPIs} → {3TC ∧ RTVb ∧ NoNNRTIs}	-
-216	123	2.13e-06	{d4T ∧ NoNNRTIs ∧ failure} → {LPV ∧ RTVb ∧ NoNNRTIs}	0.01
-214	152	6.57e-02	{3TC ∧ ddI ∧ NoNNRTIs} → {3TC ∧ RTVb ∧ NoNNRTIs}	-
-214	288	9.09e-11	{RTVb ∧ NoNNRTIs ∧ failure}	1.24e-5

DOI: 10.2202/1544-6115.1604

Table A4, continued.

IMP	FREQ	p-value	pattern	{TH, GT} × TO
-212	101	6.09e-03	{ABC ∧ d4T ∧ NoNNRTIs ∧ NoPIs}	-
-211	277	1.36e-02	{AZT ∧ NoNNRTIs ∧ NoPIs} → {3TC ∧ d4T ∧ NoPIs}	-
-210	121	4.45e-03	{ddI ∧ TDF ∧ RTVb}	-
-209	411	3.61e-09	{LPV ∧ RTVb ∧ failure}	2.10e-4
-208	697	3.91e-03	{AZT ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ NoNNRTIs ∧ NoPIs} → {3TC ∧ d4T ∧ NoNNRTIs}	-
-206	581	4.73e-04	{3TC ∧ AZT ∧ failure}	0.12
-203	184	6.96e-06	{ddI ∧ RTVb ∧ NoNNRTIs ∧ failure}	0.36
-199	101	1.22e-02	{d4T ∧ RTV ∧ NoNNRTIs}	-
-194	502	7.49e-07	{3TC ∧ RTVb ∧ NoNNRTIs} → {**TDF** ∧ **RTVb** ∧ **NoNNRTIs**}	-
-193	234	5.81e-06	{AZT ∧ NoNNRTIs ∧ NoPIs} → {3TC ∧ d4T ∧ NoNNRTIs}	-
-193	402	2.64e-06	{d4T ∧ NFV ∧ failure}	0.01
-193	188	1.03e-02	{3TC ∧ d4T ∧ NoNNRTIs} → {d4T ∧ RTV ∧ SQV}	-
-193	124	7.32e-13	{d4T ∧ NVP ∧ NFV}	-
-192	246	9.59e-07	{3TC ∧ NFV ∧ NoNNRTIs} → {3TC ∧ ABC ∧ NoNNRTIs}	-
-192	126	2.95e-04	{3TC ∧ NoPIs ∧ RT67N}	-
-191	104	4.79e-08	{PRO63P ∧ RT67N ∧ RT70R}	-
-186	368	2.96e-02	{ddI ∧ NoNNRTIs ∧ NoPIs} → {ddC ∧ NoNNRTIs ∧ NoPIs}	-
-185	135	5.07e-02	{TDF ∧ SQV ∧ NoNNRTIs}	-
-185	172	1.31e-03	{3TC ∧ SQV ∧ NoNNRTIs} → {d4T ∧ NoNNRTIs ∧ failure}	0.47
-185	175	3.93e-04	{d4T ∧ LPV ∧ NoNNRTIs}	-
-180	180	1.15e-02	{AZT ∧ RTVb ∧ NoNNRTIs} → {**AZT** ∧ **RTVb** ∧ **NoNNRTIs**}	-
-176	176	1.62e-03	{SQV ∧ NoNNRTIs ∧ failure} → {d4T ∧ SQV ∧ NoNNRTIs}	0.04
-175	125	1.57e-02	{3TC ∧ IDV ∧ NoNNRTIs} → {d4T ∧ SQV ∧ NoNNRTIs}	-
-174	179	3.58e-02	{RTVb ∧ PRO63P ∧ RT67N}	-
-174	273	3.14e-04	{AZT ∧ NoNNRTIs ∧ NoPIs} → {3TC ∧ SQV ∧ NoNNRTIs}	-
-173	109	1.46e-10	{d4T ∧ ddI ∧ RTVb} → {d4T ∧ RTVb ∧ NoNNRTIs}	-
-172	127	1.12e-02	{d4T ∧ SQV ∧ NoNNRTIs} → {d4T ∧ NoNNRTIs ∧ failure}	0.13
-172	218	1.04e-02	{TDF ∧ ATV ∧ failure}	4.90e-5
-171	166	3.80e-05	{d4T ∧ NoNNRTIs ∧ failure} → {d4T ∧ SQV ∧ NoNNRTIs}	-
-171	117	3.98e-05	{ddI ∧ APV ∧ NoNNRTIs}	-
-170	145	2.51e-04	{3TC ∧ AZT ∧ PRO13V}	-
-166	174	9.56e-05	{d4T ∧ NoNNRTIs ∧ PRO13V}	-
-166	277	2.12e-12	{3TC ∧ RTVb ∧ FPV}	-
-165	107	1.34e-05	{PRO10I ∧ PRO63P ∧ RT70R}	-
-164	183	2.24e-05	{RTVb ∧ FPV ∧ NoNNRTIs}	-
-162	129	2.44e-07	{3TC ∧ RT215Y ∧ failure}	0.02
-160	278	4.86e-07	{APV ∧ RTVb ∧ NoNNRTIs}	-
-158	360	3.54e-04	{ABC ∧ d4T ∧ failure}	0.01
-155	207	7.13e-08	{RTVb ∧ SQV ∧ NoNNRTIs} → {d4T ∧ RTVb ∧ NoNNRTIs}	-
-148	340	5.47e-05	{d4T ∧ ddI ∧ NoNNRTIs} → {d4T ∧ SQV ∧ NoNNRTIs}	-
-147	112	1.87e-02	{AZT ∧ NoNNRTIs ∧ NoPIs} → {ddI ∧ NoNNRTIs ∧ NoPIs} → {ddC ∧ NoNNRTIs ∧ NoPIs}	-
-139	124	3.01e-03	{3TC ∧ PRO20R ∧ PRO36I}	-
-133	101	2.71e-09	{d4T ∧ EFV ∧ RTVb} → {d4T ∧ LPV ∧ NoNNRTIs}	-
-132	119	4.72e-07	{PRO63P ∧ RT184V ∧ RT215Y}	-
-128	231	1.89e-08	{RT210W ∧ RT215Y ∧ failure}	5.36e-4
-118	102	1.74e-04	{d4T ∧ ddI ∧ NoNNRTIs}	-

Table A4, continued.

IMP	FREQ	p-value	pattern	{TH, GT} × TO
152	179	6.47e-06	{3TC ∧ d4T ∧ NoNNRTIs} → {RTVb ∧ NoNNRTIs ∧ failure}	0.43
161	150	1.83e-02	{**3TC** ∧ **RTVb** ∧ **DRV**}	-
163	415	6.75e-02	{**3TC** ∧ **DRV** ∧ **NoNNRTIs**}	-
168	111	2.40e-02	{3TC ∧ NoNNRTIs ∧ NoPIs}	-
171	455	6.96e-03	{3TC ∧ d4T ∧ NoNNRTIs} → {ddI ∧ RTVb ∧ NoNNRTIs}	-
177	304	3.71e-08	{3TC ∧ NoNNRTIs ∧ failure} → {ddI ∧ NoNNRTIs ∧ failure}	0.64
183	268	5.76e-04	{NoNNRTIs ∧ PRO93L ∧ RT184V}	-
184	134	9.80e-02	{RTVb ∧ T20 ∧ success}	0.03
197	134	5.06e-03	{3TC ∧ d4T ∧ NFV} → {3TC ∧ d4T ∧ NFV}	-
204	132	1.02e-02	{TDF ∧ RTVb ∧ NoNNRTIs} → {RTVb ∧ NoNNRTIs ∧ success}	5.19e-7
208	239	1.61e-05	{d4T ∧ RTVb ∧ NoNNRTIs} → {d4T ∧ RTVb ∧ failure}	0.03
210	140	6.85e-06	{d4T ∧ ddI ∧ NoPIs} → {LPV ∧ RTVb ∧ NoNNRTIs}	-
211	294	6.14e-04	{RTVb ∧ NoNNRTIs ∧ PRO93L}	-
213	168	3.10e-11	{d4T ∧ EFV ∧ NoPIs} → {3TC ∧ RTVb ∧ NoNNRTIs}	-
214	569	5.96e-02	{3TC ∧ NoNNRTIs ∧ NoPIs} → {3TC ∧ IDV ∧ NoNNRTIs}	-
214	218	1.84e-07	{3TC ∧ AZT ∧ NoPIs} → {3TC ∧ IDV ∧ NoNNRTIs}	-
214	250	3.87e-03	{TDF ∧ RTVb ∧ NoNNRTIs ∧ success}	3.38e-3
215	330	8.33e-02	{LPV ∧ RTVb ∧ NoNNRTIs} → {3TC ∧ NoNNRTIs ∧ NoPIs}	-
216	166	2.53e-02	{d4T ∧ ddI ∧ NoNNRTIs} → {ddI ∧ NoNNRTIs ∧ NoPIs}	-
217	117	3.96e-04	{AZT ∧ ddI ∧ NoNNRTIs} → {ddI ∧ NoNNRTIs ∧ NoPIs}	-
217	207	1.12e-03	{3TC ∧ ddC ∧ NoPIs}	-
220	415	1.53e-03	{3TC ∧ AZT ∧ NoNNRTIs} → {3TC ∧ d4T ∧ IDV}	-
220	167	1.40e-02	{3TC ∧ RTVb ∧ NoNNRTIs} → {3TC ∧ NoNNRTIs ∧ NoPIs}	-
220	243	3.75e-04	{ddI ∧ EFV ∧ success}	0.09
224	214	5.92e-07	{NoNNRTIs ∧ PRO63P ∧ PRO93L}	-
226	842	9.17e-05	{3TC ∧ NoPIs ∧ success}	1.61e-6
226	386	1.17e-06	{ddI ∧ NoNNRTIs ∧ success}	0.05
226	102	3.05e-02	{3TC ∧ d4T ∧ NoNNRTIs} → {d4T ∧ RTVb ∧ NoNNRTIs}	-
229	342	5.87e-05	{d4T ∧ NVP ∧ success}	1.48e-4
230	264	3.32e-02	{3TC ∧ ddI ∧ NoNNRTIs} → {3TC ∧ ddI ∧ NoNNRTIs}	-
230	168	3.94e-04	{d4T ∧ NVP ∧ NoPIs}	-
231	334	4.16e-02	{AZT ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ IDV ∧ NoNNRTIs}	-
233	250	3.72e-02	{ddI ∧ RTVb ∧ NoNNRTIs} → {RTVb ∧ NoNNRTIs ∧ success}	0.01
233	491	5.08e-04	{3TC ∧ RTVb ∧ success}	0.03
234	240	3.54e-08	{d4T ∧ ddI ∧ NoPIs} → {3TC ∧ RTVb ∧ NoNNRTIs}	-

Table A4, continued.

IMP	FREQ	p-value	pattern	{TH, GT} × TO
237	287	7.17e-04	{ABC ∧ ddI ∧ NoPIs}	-
237	223	7.69e-02	{ABC ∧ TDF ∧ NoNNRTIs}	-
239	109	6.51e-02	{3TC ∧ ABC ∧ NoPIs}	-
240	155	1.06e-06	{3TC ∧ ddI ∧ NoNNRTIs} → {3TC ∧ TDF ∧ NoNNRTIs}	-
241	189	9.75e-02	{d4T ∧ NFV ∧ NoNNRTIs} → {d4T ∧ NFV ∧ NoNNRTIs}	-
242	149	1.21e-03	{ddI ∧ NoNNRTIs ∧ failure} → {ddI ∧ NoNNRTIs ∧ NoPIs}	0.04
242	106	7.71e-06	{3TC ∧ LPV ∧ NoNNRTIs} → {3TC ∧ RTVb ∧ NoNNRTIs}	-
243	137	1.08e-06	{EFV ∧ NoPIs ∧ success}	0.01
244	224	1.18e-03	{3TC ∧ RTVb ∧ NoNNRTIs} → {3TC ∧ TDF ∧ NoNNRTIs}	-
244	102	1.73e-03	{AZT ∧ NoNNRTIs ∧ failure} → {3TC ∧ ddI ∧ NoNNRTIs}	0.01
246	116	7.30e-02	{3TC ∧ AZT ∧ TDF ∧ NoNNRTIs}	-
246	159	3.11e-02	{AZT ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ NoNNRTIs ∧ NoPIs} → {3TC ∧ IDV ∧ NoNNRTIs}	-
246	132	3.00e-02	{3TC ∧ ddC ∧ NoNNRTIs}	-
247	104	9.23e-03	{RTVb ∧ NoNNRTIs ∧ success} → {LPV ∧ RTVb ∧ NoNNRTIs}	0.02
248	896	3.09e-11	{d4T ∧ NoNNRTIs ∧ success}	8.15e-5
248	195	2.49e-02	{d4T ∧ EFV ∧ success}	0.09
254	170	2.42e-02	{ABC ∧ AZT ∧ NoNNRTIs}	-
256	373	3.60e-04	{NVP ∧ NoPIs ∧ success}	0.01
256	360	1.65e-02	{ABC ∧ EFV ∧ NoPIs}	-
258	115	1.04e-03	{ABC ∧ AZT ∧ NoPIs}	-
259	141	5.45e-02	{3TC ∧ d4T ∧ NoNNRTIs} → {3TC ∧ d4T ∧ NoNNRTIs} → {3TC ∧ d4T ∧ NoNNRTIs}	-
261	584	8.62e-04	{d4T ∧ ddI ∧ success}	0.01
261	311	1.49e-04	{3TC ∧ ABC ∧ NoNNRTIs} → {3TC ∧ ABC ∧ NoPIs}	-
262	102	7.43e-02	{ddI ∧ NoPIs ∧ success}	0.01
267	174	4.33e-02	{AZT ∧ ddI ∧ NoNNRTIs} → {AZT ∧ ddI ∧ NoNNRTIs}	-
267	160	4.05e-05	{ddI ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ NoNNRTIs ∧ NoPIs} → {AZT ∧ NoNNRTIs ∧ NoPIs}	-
268	528	7.11e-02	{3TC ∧ ddI ∧ NoNNRTIs}	-
272	143	7.26e-02	{AZT ∧ NoNNRTIs ∧ NoPIs} → {3TC ∧ NoNNRTIs ∧ NoPIs}	-
273	192	2.65e-03	{d4T ∧ ddI ∧ NoNNRTIs]} → {d4T ∧ ddI ∧ NoPIs}	-
274	277	2.45e-02	{3TC ∧ d4T ∧ NoNNRTIs} → {3TC ∧ d4T ∧ NFV}	-
275	184	9.68e-06	{ddI ∧ NFV ∧ success}	1.21e-4
284	497	1.76e-04	{IDV ∧ NoNNRTIs ∧ success}	1.53e-3
287	364	7.60e-04	{ABC ∧ EFV ∧ success}	0.21
290	128	3.47e-05	{LPV ∧ RTVb ∧ success}	0.10
293	117	3.47e-06	{d4T ∧ NFV ∧ success}	1.33e-3
298	163	7.93e-06	{d4T ∧ NoPIs ∧ success}	4.95e-21
299	109	2.70e-03	{ddI ∧ NoNNRTIs ∧ NoPIs}	-
300	167	4.64e-02	{ddI ∧ NoNNRTIs ∧ PRO63P}	-
302	268	1.27e-02	{3TC ∧ ddI ∧ NoPIs}	-
304	172	5.96e-02	{3TC ∧ ABC ∧ AZT ∧ NoNNRTIs}	-
308	101	7.72e-05	{AZT ∧ ddI ∧ NoPIs} → {ddI ∧ NoNNRTIs ∧ NoPIs}	-
310	108	3.61e-03	{3TC ∧ NoNNRTIs ∧ success} → {3TC ∧ AZT ∧ NoPIs}	0.01
311	183	7.08e-05	{ddI ∧ NoNNRTIs ∧ NoPIs} → {3TC ∧ NoNNRTIs ∧ NoPIs}	-
348	132	2.35e-04	{3TC ∧ ABC ∧ AZT} → {ABC ∧ AZT ∧ NoPIs}	-
374	155	2.52e-10	{NFV ∧ NoNNRTIs ∧ success}	7.28e-4

References

Altmann, A., N. Beerenwinkel, T. Sing, I. Savenkov, M. Däumer, R. Kaiser, S. Y. Rhee, W. J. Fessel, R. W. Shafer, and T. Lengauer (2007): “Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance,” *Antiviral Therapy*, 12, 169–178.

Altmann, A., M. Däumer, N. Beerenwinkel, Y. Peres, E. Schülter, J. Büch, S. Rhee, A. Sönnerborg, W. Fessel, R. Shafer, M. Zazzi, R. Kaiser, and T. Lengauer (2009): “Predicting response to combination antiretroviral therapy: retrospective validation of geno2pheno-theo on a large clinical database,” *Journal of Infectious Diseases*, 199(7), 999–1006.

Antinori, A., M. Zaccarelli, A. Cingolani, F. Forbici, M. G. Rizzo, M. P. T. otta, S. D. Giambenedetto, P. Narciso, A. Ammassari, E. Girardi, A. D. Luca, and C. F. Perno (2002): “Cross-resistance among nonnucleoside reverse transcriptase inhibitors limits recycling efavirenz after nevirapine failure,” *AIDS Res Hum Retroviruses*, 18, 835–8.

Baldi, P., S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen (2000): “Assessing the accuracy of prediction algorithms for classification: an overview,” *Bioinformatics (Oxford, England)*, 16, 412–24.

Barré-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vézinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier (1983): “Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS),” *Science*, 220, 868–71, URL `http://www.sciencemag.org/cgi/reprint/220/4599/868`

Beerenwinkel, N., M. Däumer, M. Oette, K. Korn, D. Hoffman, R. Kaiser, T. Lengauer, J. Selbig, and H. Walter (2003): “Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes,” *NAR*, 31(13), 3850–3855.

Beerenwinkel, N., B. Schmidt, H. Walter, R. Kaiser, T. Lengauer, D. Hoffman, K. Korn, and J. Selbig (2002): “Diversity and complexity of HIV-1 drug resistance: A bioinformatics approach to predicting phenotype from genotype,” *PNAS*, 99(12), 8271–8276.

Bickel, P. and K. Doksum (2002): *Mathematical Statistics: Basic Ideas and Selected Topics*, Prentice-Hall.

Bickel, S., J. Bogojeska, T. Lengauer, and T. Scheffer (2008): “Multi-task learning for HIV therapy screening,” in *25th International Conference on Machine Learning*, 56–63.

Bratt, G., A. Karlsson, A. C. Leandersson, J. Albert, B. Wahren, and E. Sandström (1998): "Treatment history and baseline viral load, but not viral tropism or CCR-5 genotype, influence prolonged antiviral efficacy of highly active antiretroviral treatment," *AIDS*, 12, 2193–202, URL `http://meta.wkhealth.com/pt/pt-core/template-journal/lwwgateway/media/landingpage.htm?issn=0269-9370&volume=12&issue=16&spage=2193`

Clavel, F. and A. J. Hance (2004): "HIV drug resistance," *N Engl J Med*, 350, 1023–35, URL `http://content.nejm.org/cgi/content/extract/350/10/1023`.

Deforche, K., A. Cozzi-Lepri, K. Theys, B. Clotet, R. J. Camacho, J. Kjaer, K. V. Laethem, A. Phillips, Y. Moreau, J. D. Lundgren, A.-M. Vandamme, and E. S. Group (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.

Deforche, K., T. Silander, R. Camacho, Z. Grossman, M. Soares, K. Laethem, R. Kantor, Y. Moreau, and A.-M. Vandamme (2006): "Analysis of HIV-1 pol sequences using bayesian networks: implications for drug resistance," *Bioinformatics*, 22, 2975–2979.

Demiriz, A., K. Bennet, and J. Shawe-Taylor (2002): "Linear programming boosting via column generation," *Machine Learning*, 46(1-3), 225–254.

Fields, B. N., D. M. Knipe, and P. M. Howley (2007): "Fields virology," *Lippincott Williams & Wilkins*, URL `http://books.google.com/books?id=5OOsomr0w18C&printsec=frontcover`.

Foulkes, A. S. and V. DeGruttola (2002): "Characterizing the relationship between HIV-1 genotype and phenotype: Prediction-based classification," *Biometrics*, 58, 146–156.

Foulkes, A. S. and V. DeGruttola (2003): "Characterizing the progression of viral mutations over time," *J. Am. Stat. Assoc*, 98, 859–867.

Gao, F., Y. Chen, D. N. Levy, J. A. Conway, T. B. Kepler, and H. Hui (2004): "Unselected mutations in the human immunodeficiency virus type 1 genome are mostly nonsynonymous and often deleterious." *Journal of Virology*, 78, 2426–2433.

Johnson, V. A., F. Brun-Vèzinet, B. Clotest, H. F. Günthard, D. R. Kuritzkes, D. P. Pillay, J. M. Schapiro, and D. D. Richman (2008): "Update of the drug resistance mutations in HIV-1: Spring 2008," *Topics in HIV Medicine*, 16(1), 62–68.

Kudo, T., E. Maeda, and Y. Matsumoto (2005): "An application of boosting to graph classification," in *Advances in Neural Information Processing Systems 17*, MIT Press, 729–736.

Larder, B. (2007): "The development of artificial neural networks to predict virological response to combination hiv therapy," *Antivir Ther*, 12(1), 15–24.

Morishita, S. (2001): "Computing optimal hypotheses efficiently for boosting," in *Discovery Science*, 471–481.

Nowozin, S., G. Bakir, and K. Tsuda (2008): "Discriminative subsequence mining for action classification," in *International Conference on Computer Vision*, 1919–1923.

Pei, J., J. Han, B. Mortazavi-asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M. Hsu (2004): "Mining sequential patterns by pattern-growth: The prefixspan approach," *IEEE Transactions on Knowledge and Data Engineering*, 16(11), 1424–1440.

Rabinowitz, M., L. Myers, M. Banjevic, A. Chan, J. Sweetkind-Singer, J. Haberer, K. McCann, and R. Wolkowicz (2006): "Accurate prediction of HIV-1 drug response from the reverse transciptase and protease amino acid sequences using sparse models created by convex optimization," *Bioinformatics*, 22, 541–549.

Rhee, S. Y., J. Taylor, G. Wadhera, A. Ben-Hur, D. L. Brutlag, and R. W. Shafer (2006): "Genotypic predictors of human immunodeficiency virus type 1 drug resistance," *PNAS*, 103, 17355–17360.

Rosen-Zvi, M., A. Altmann, M. Prosperi, E. Aharoni, H. Neuvirth, A. Sönnenborg, E. Schülter, D. Struck, Y. Peres, F. Incardona, R. Kaiser, M. Zazzi, and T. Lengauer (2008): "Selecting anti-HIV therapies based on a variety of genomic and clinical factors," *Bioinformatics*, 24, i399–i406.

Saigo, H., T. Uno, and K. Tsuda (2007): "Mining complex genotypic features for predicting HIV-1 drug resistance," *Bioinformatics*, 23(18), 2455–2462.

Schölkopf, B. and A. J. Smola (2002): *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, URL `http://www.learning-with-kernels.org`

Sing, T. and N. Beerenwinkel (2007): "Mutagenetic tree fisher kernel improves prediction of HIV drug resistance from viral genotype," in B. Schölkopf, J. Platt, and T. Hoffman, eds., *Advances in Neural Information Processing Systems 19*, Cambridge, MA: MIT Press, 1297–1304.

Sing, T., V. Svicher, N. Beerenwinkel, F. Ceccherini-Silberstein, M. Däumer, R. Kaiser, H. Walter, K. Korn, D. Hoffmann, M. Oette, J. K. Rockstoh, G. Fätkenheuer, C.-F. Perno, and T. Lengauer (2005): "Characterization of novel HIV drug resistance mutations using clustering, multidimensional scaling and SVM-based feature ranking," in *ninth European Conference on Principles and Practice of Knowledge Discovery in Databases*, 285–296.

Wang, D. and B. Larder (2003): "Enhanced prediction of lopinavir resistance from genotype by use of artificial neural networks," *Journal of Infectious Diseases*, 188, 653–660.

Zaccarelli, M., P. Lorenzini, F. Ceccherini-Silberstein, V. Tozzi, F. Forbici, C. Gori, M. P. Trotta, E. Boumis, P. Narciso, C. F. Perno, and A. Antinori (2009): "Historical resistance profile helps to predict salvage failure," *Antivir Ther (Lond)*, 14, 285–91, URL `http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&Cmd=Retrieve&list_uids=19430103&dopt=abstractplus`.

Zou, H. (2006): "The adaptive LASSO and its oracle properties." *J. Am. Stat. Ass.*, 101, 1418–1429.