

Predicting Response to Antiretroviral Treatment by Machine Learning: The EuResist Project

Maurizio Zazzi^a Francesca Incardona^b Michal Rosen-Zvi^c Mattia Prosperi^b
Thomas Lengauer^d Andre Altmann^d Anders Sonnerborg^f Tamar Lavee^c
Eugen Schülter^e Rolf Kaiser^e

^aDepartment of Biotechnology, University of Siena, Siena, and ^bInforma s.r.l., Rome, Italy; ^cIBM Haifa, Israel; ^dComputational Biology and Applied Algorithmics, Max Planck Institute of Information Technology, Saarbrücken, and ^eInstitute of Virology, University of Cologne, Cologne, Germany; ^fKarolinska Institute, Stockholm, Sweden

Key Words

HIV resistance · EuResist · Antiretroviral therapy · Interpretation system · Machine learning

Abstract

For a long time, the clinical management of antiretroviral drug resistance was based on sequence analysis of the HIV genome followed by estimating drug susceptibility from the mutational pattern that was detected. The large number of anti-HIV drugs and HIV drug resistance mutations has prompted the development of computer-aided genotype interpretation systems, typically comprising rules handcrafted by experts via careful examination of in vitro and in vivo resistance data. More recently, machine learning approaches have been applied to establish data-driven engines able to indicate the most effective treatments for any patient and virus combination. Systems of this kind, currently including the Resistance Response Database Initiative and the EuResist engine, must learn from the large data sets of patient histories and can provide an objective and accurate estimate of the virological response to different antiretroviral regimens. The EuResist engine was developed by a European consortium of HIV and bioinformatics experts and compares favor-

ably with the most commonly used genotype interpretation systems and HIV drug resistance experts. Next-generation treatment response prediction engines may valuably assist the HIV specialist in the challenging task of establishing effective regimens for patients harboring drug-resistant virus strains. The extensive collection and accurate processing of increasingly large patient data sets are eagerly awaited to further train and translate these systems from prototype engines into real-life treatment decision support tools.

Copyright © 2012 S. Karger AG, Basel

Background

The impressive advances in antiretroviral therapy notwithstanding, handling HIV drug resistance issues remains a challenge in the clinical management of many HIV-positive patients. While increasingly potent drug regimens can minimize or defer the development of drug resistance to first-line therapy, this occurrence cannot be completely avoided and lots of patients presently harbor drug-resistant HIV variants as a result of previous treatment failures. The large number of drug resistance mutations in the HIV genome and the possibility of combining

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2012 S. Karger AG, Basel
0300-5526/12/0552-0123\$38.00/0

Accessible online at:
www.karger.com/int

Maurizio Zazzi
Department of Biotechnology
University of Siena
IT-53100 Siena (Italy)
Tel. +39 057 723 3863, E-Mail zazzi@unisi.it

more than 20 antiretroviral drugs in different 3- or 4-drug regimens often render the choice of therapy puzzling. Indeed, several tools have been developed that help to infer the susceptibility of a specific virus variant to each of the individual drugs based on the mutation(s) detected by HIV genotyping [1]. Such tools include a variety of approaches ranging from simple mutation lists to rules-based algorithms, whereby each drug is considered fully, partially or not active depending on defined mutations and mutation sets. Regardless of their more or less complex nature, all the genotype interpretation systems (GIS) must be updated regularly to incorporate new knowledge in the field of HIV drug resistance. Such knowledge is being derived from the correlation between an HIV genotype and (1) *in vitro* drug susceptibility (phenotype), (2) patient treatment history, and (3) *in vivo* response to therapy. Along with the increased availability of genotype response to treatment correlation data, most GIS are now being tailored to predict drug efficacy *in vivo*.

The process of creating or updating a GIS has usually been done by a panel of experts via review and discussion of the pertinent literature followed by a vote to define or revise the algorithm. More recently, machine learning techniques have begun to be explored as methods to generate objective and clinically relevant interpretation of drug resistance data. Historically, the Virco proprietary VirtualPhenotype™ [2] and the freely available geno2pheno [3] systems were the first successful attempts to build data-driven GIS, although they were originally aimed at predicting the *in vitro* phenotype rather than the *in vivo* response to treatment. The current version of the Virco system (vircoTYPE) adds lower and upper clinical cutoffs for almost all drugs in order to translate the predicted phenotype into a clinically relevant estimate of efficacy [4]. The latest version of geno2pheno has evolved even more into a therapy optimizer prototype (called THEO) aimed at selecting treatment regimens based on an HIV genotype and some additional choices made by the user [5].

The first attempt to build a large international repository of genotype response data and exploit sophisticated data mining techniques to predict treatment outcome has been made by the HIV Resistance Response Database Initiative (RDI; www.hivrdi.org), a nonprofit organization registered in the UK. The RDI has been using artificial neural networks and random forest models with the ambitious task of predicting the absolute viral load change, or simply treatment success (undetectable viral load), following treatment switch for a given HIV genotype [6].

The system has recently been launched as an online tool for registered users (free of charge) and a prospective open pilot study of the system in clinical practice is ongoing [7]. Funded by the European Commission under the 6th Framework Programme, the EuResist project (www.euresist.org) has likewise collected genotype response data from different European countries and has explored several machine learning techniques to develop a treatment response prediction engine [8]. Both these initiatives share the need for a large amount of training data and the aim of developing freely available Web services. In addition, both have been exploring the inclusion of supplementary clinical data to improve the accuracy of the prediction, as well as prototype systems where treatment outcome is modeled in the absence of HIV genotype information [9, 10]. Although many issues are still to be resolved, these methods show promise and will provide novel treatment decision support tools in the near future.

The EuResist Approach

EuResist has collected data from multicentric HIV clinical databases in Germany, Italy, Luxembourg, Belgium, Portugal, Spain and Sweden. The source databases are regularly updated and the physically integrated EuResist database is periodically refreshed. The integration effort has also led to a definition of Health Level Seven (HL7) standards for the storage of HIV resistance data, currently under approval. Collected data include the demographics, viral loads, CD4 counts, treatment history and HIV genotypes of about 49,000 patients. From these, instances of the so-called treatment change episode (TCE, an acronym first created by the RDI team) are derived based on the TCE definition itself. In its original form, a TCE is made of a baseline viral load and a genotype, a new treatment and a follow-up viral load obtained while still on that treatment. Time constraints apply to all the TCE data and have been tentatively proposed by the Forum for HIV Research (www.hivforum.org/uploads/Resistance/DataAnalysisPlanRev1.pdf). The time intervals in the EuResist TCE are compliant with the Forum definition and currently focus on short-term 8-week responses. Data mining techniques have discovered several additional features and derived features that could have an impact on response to treatment and possibly improve the accuracy of the prediction engine. A partially overlapping but distinct set of potentially relevant features has been selected using different methods by the 3 machine learning groups participating in the EuResist project (the

Max-Planck Institute for Bioinformatics in Saarbrücken, Germany, the IBM Research Laboratories in Haifa, Israel and the Department of Automation, Roma Tre University together with the company Informa in Rome, Italy). The supplementary parameters considered include several indicators of treatment history, previously identified resistance mutations, route of infection, ethnicity, age, CD4 counts, consensus B local similarity and viral subtype. In contrast to the first version of the RDI system, EuResist aims at predicting a dichotomous outcome, i.e. treatment success or failure, rather than the change in viral load. This fits well to the current view of antiretroviral therapy, whereby treatment success is denoted by complete suppression of viral load [11]. Since high-level viremia may not decrease to undetectable levels in 8 weeks even with an effective treatment, the definition of success includes both an undetectable viral load and a decrease of at least 2 log₁₀ in viral load. The expanded set of attributes listed above together with the TCE parameters and the label of success or failure make up the definition of the EuResist standard datum. A set of >7,300 standard datum instances derived from the EuResist database is being used for training multiple machine learning approaches. Each learning strategy results in an engine that generates a prediction of short-term outcome for a patient given (at least) a particular drug combination and a genotype. Different engines are based on different subsets of attributes and these attributes may be differently represented in the engines. Next, these engines are combined in a final engine working as a free Web service to assist an infectious-diseases specialist in the establishment of the best treatment regimen (fig. 1).

Current Status

At the end of the EuResist project life cycle (June 2008), 3 engines had been developed and found to perform similarly in validation tests. Data have been continuously collected and updated and the engines retrained accordingly. The latest release of engines and data from November 2010 is currently in use by the Web service. Notwithstanding the extensive exploration of sophisticated machine learning techniques such as Support Vector Machines, Fuzzy Logic, Case-Based Reasoning and Random Forests, all 3 engines have ended up using more popular logistic regression models for the classification of the therapies as successes or failures. However, the engines employ different approaches to derive useful extra information from the EuResist database that is not directly

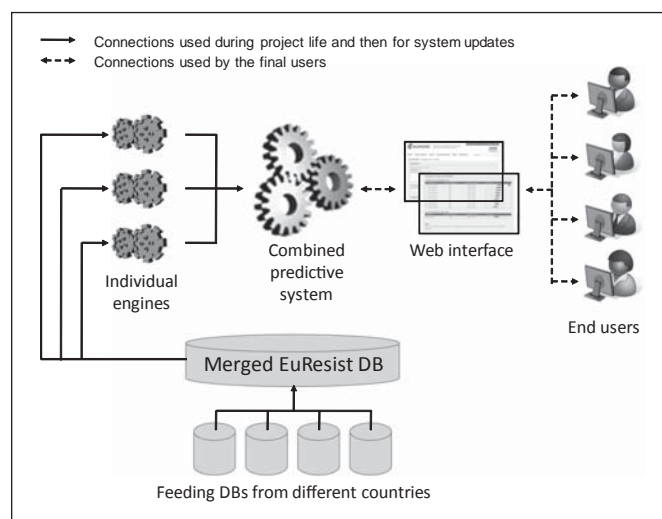


Fig. 1. The EuResist system integrates the different databases from the partners into an integrated database. This is used for training different individual machine learning procedures (engines). The different engines are combined in a joint predictive system which can be used free of charge and without registration via Internet. Via a Web interface, the end users (clinical virologists, clinicians, IT technicians/experts) receive the prediction of the most likely drug combination by submitting a patient's HIV sequence and other optional clinical data. They receive ranked suggestions of different drug combinations. DB = Database.

contained in the standard datum. The generative discriminative engine uses a Bayesian network to derive the probability of therapy success on the basis of clinical markers only (without genotype). On the other hand, the evolutionary engine focuses on a model of viral evolution under the selective pressure exerted by a specific drug, deriving a measure of the genetic barrier to resistance [12]. Finally, the mixed-effects engine considers a number of 2nd- and 3rd-order interactions among variables (drug × drug, drug × drug × drug and drug × mutation), thus accounting for composite effects [13].

A thorough analysis of the agreement and disagreement between the engines was carried out on the first release. It was found that all 3 have a similar performance, but are not identical and in fact disagree in 18.3% of cases. Notably, agreement of all 3 classifiers on the wrong label occurs far more frequently in instances labeled 'failure' than in instances labeled 'success' (37 vs. 4%), possibly due to issues of patients' adherence to therapy in a proportion of the training data. This problem and if and how it can be inferred from the data is now being explored. Several sophisticated fusion methods have been studied and compared to deliver one single

combined prediction derived from the 3 individual engines [14]. At present, even a simple mean combiner shows a better learning curve compared to the individual prediction systems. It generates more reliable predictions with fewer training data and reduces the standard deviation on multiple test repetitions, suggesting greater reliability. Based on validation studies on historical cases, the final engine currently predicts the correct outcome in about 76% of the cases. Interestingly, an investigation of the wrong predictions detected a relevant proportion of cases where the predicted success was achieved but later than at the target 8-week follow-up. This suggests that the performance of the system could be improved by redesigning the standard datum definition to match a clinically more relevant scenario. In line with this, a recent analysis also indicated that the accuracy of the prediction of the 24-week outcome equals that of the 8-week outcome even without retraining the system with 24-week follow-up data.

As for any novel diagnostic system, the EuResist engine is being compared with state-of-the-art tools that are available for assisting the choice of antiretroviral therapy in clinical practice. Different methods have different outputs and this complicates the comparison. For example, most GIS predict the activity of single drugs while EuResist predicts the success probability of treatment regimens. Moreover, there is no clear indication or consensus among GIS in terms of follow-up time and none allows the input of additional patient data. In order to enable reasonably fair comparisons among the systems, the sum of the GIS scores gained for single drugs can be analyzed by a logistic regression model and taken as an overall indicator for the predicted activity of the combination therapy. Initial results favor the use of the EuResist engine as an improved treatment decision support tool [8, 13]. In addition to this comparative analysis of methods, the EuResist team also provided a random selection of 25 case histories to a select panel of 10 HIV resistance experts, asking them to predict whether the indicated treatment would be successful or not. The experts could access all the patient records and use any of the available GIS to make their prediction. The goal of the study was the comparison between the EuResist engine and the expert opinion of the experts in a scenario which simulated clinical practice as much as possible. One out of the ten experts failed only 6 predictions as the EuResist system had, while all the others failed larger numbers of predictions [15]. These comparative studies will contribute to defining the role of a fully automated and freely available system in the clinical management of HIV patients.

The first public version of the EuResist engine was made available on the Web in September 2008. The user can input just the HIV genotype or add other attributes summarizing the patient history and baseline status. If these additional parameters are provided, the system uses them to refine the genotype-based prediction. The output is a list of the top ten best treatment regimens each with its probability of success. The probability is denoted as the range of the scores determined by the 3 individual engines as an indication of the agreement among them. Options for the inclusion or exclusion of specific drugs are provided. In this case, a second top ten list is provided based on user choices. The second version of the system has been online since November 2010. Apart from minor adjustments in the user interface and an expanded set of user-defined options, most importantly this new release incorporates the ability to model the activity of some of the newest drugs, such as darunavir, tipranavir and etravirine.

Open Issues and Future Perspectives

EuResist and similar initiatives introduce a new paradigm to the variety of antiretroviral treatment decision support tools. The output is a ranking of the best combinations of antiretroviral drugs with an estimate of the probability to achieve suppression of viral load. This is closer to the practical needs of the HIV specialist when compared to the usual list of predicted degrees of susceptibility of the virus to the individual drugs provided by the standard rules-based GIS. While this is certainly welcome, the advantage of the full-featured engine comes at the cost of inputting several additional data, particularly data summarizing the previous exposure to therapy and possibly also previous genotypes. Some physicians could be tempted to skip these extra requests and provide only the HIV genotype as they are used to doing and so they do not benefit from the optimal use of the engine. In some cases, the extra information could be simply not available to the user. However, electronic medical recording systems are also improving and expanding at a notable pace; thus, protocols for the automatic and secure delivery of patient data to an online prediction system are immediately within reach.

An implicit drawback of all machine learning systems is the need to have a reasonable amount of data for training. This raises at least three issues. First, data must be collected from multiple sources to obtain the requested critical amount. As with the EuResist database, the num-

ber of training examples obtained is far less than expected, even from large numbers of patients, because of the constraints dictated by the standard datum definition. Every effort to gather clinical information from many sources must deal with a data import from multiple formats, comply with differing ethical issues across countries and deal with a highly variable propensity to data sharing. Second, the quality of the training data is crucial, but effective quality assurance protocols are difficult to implement. Getting poor data cleaned up by the contributing centers is not always possible and labeling data as suspicious would be somewhat arbitrary. Third, there is a critical need for frequent updates and particularly for the uploading of data about novel drugs. If this requirement is not met in time, the system will probably be best suited to predict the outcome of old or even obsolete treatment regimens, but its performance will decrease for

those that include novel compounds, i.e. those where the help of the system would perhaps be the most welcome. In this context, agreements with regulatory boards and drug companies could be very helpful in updating the system to facilitate its contribution to the best use of novel regimens. In order to continue the work done and to contribute to further improvements in this perspective, the EuResist network was launched in 2009 to follow up the EuResist project mission and to further develop and deploy the services initiated by the original project.

The way to the clinical use of intelligent systems lies open and, notwithstanding the issues still to be solved, we believe that the HIV specialist will benefit from these continuously improved tools. An expanding consensus and the increased availability of high-quality training data are key factors in moving the current system prototypes into clinical practice.

References

- 1 Van Laethem K, Vandamme AM: Interpreting resistance data for HIV-1 therapy management – know the limitations. *AIDS Rev* 2006;8:37–43.
- 2 Vermeiren H, Van Craenenbroeck E, Alen P, Bacheler L, Picchio G, Lecocq P, Virco Clinical Response Collaborative Team: Prediction of HIV-1 drug susceptibility phenotype from the viral genotype using linear regression modeling. *J Virol Methods* 2007;145:47–55.
- 3 Beerenwinkel N, Däumer M, Oette M, Korn K, Hoffmann D, Kaiser R, Lengauer T, Selbig J, Walter H: Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic Acids Res* 2003;31:3850–3855.
- 4 Winters B, Montaner J, Harrigan PR, Gazzard B, Pozniak A, Miller MD, Emery S, van Leth F, Robinson P, Baxter JD, Perez-Elias M, Castor D, Hammer S, Rinehart A, Vermeiren H, Van Craenenbroeck E, Bacheler L: Determination of clinically relevant cutoffs for HIV-1 phenotypic resistance estimates through a combined analysis of clinical trial and cohort data. *J Acquir Immune Defic Syndr* 2008;48:26–34.
- 5 Altmann A, Beerenwinkel N, Sing T, Savenkov I, Doumer M, Kaiser R, Rhee SY, Fessel WJ, Shafer RW, Lengauer T: Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance. *Antivir Ther* 2007;12:169–178.
- 6 Larder B, Wang D, Revell A, Montaner J, Harrigan R, De Wolf F, Lange J, Wegner S, Ruiz L, Pérez-Eliás MJ, Emery S, Gatell J, Monforte AD, Torti C, Zazzi M, Lane C: The development of artificial neural networks to predict virological response to combination HIV therapy. *Antivir Ther* 2007;12:15–24.
- 7 Larder BA, Revell A, Mican JM, Agan BK, Harris M, Torti C, Izzo I, Metcalf JA, Rivera-Goba M, Marconi VC, Wang D, Coe D, Gazzard B, Montaner J, Lane HC: Clinical evaluation of the potential utility of computational modeling as an HIV treatment selection tool by physicians with considerable HIV experience. *AIDS Patient Care STDS* 2011;25:29–36.
- 8 Rosen-Zvi M, Altmann A, Prosperi M, Aharoni E, Neuvirth H, Sonnerborg A, Schuelter E, Struck D, Peres Y, Incardona F, Kaiser R, Zazzi M, Lengauer T: Selecting anti-HIV therapies based on a variety of genomic and clinical factors. *Bioinformatics* 2008;24:i399–i406.
- 9 Prosperi MC, Rosen-Zvi M, Altmann A, Zazzi M, Di Giambenedetto S, Kaiser R, Schülter E, Struck D, Sloot P, van de Vijver DA, Vandamme AM, Sonnerborg A, EuResist Study Group, Virolab Study Group: Antiretroviral therapy optimisation without genotype resistance testing: a perspective on treatment history based models. *PLoS One* 2010;5:e13753.
- 10 Revell AD, Wang D, Harrigan R, Hamers RL, Wensing AM, Dewolf F, Nelson M, Geretti AM, Larder BA: Modelling response to HIV therapy without a genotype: an argument for viral load monitoring in resource-limited settings. *J Antimicrob Chemother* 2010;65:605–607.
- 11 Hill A, Miralles D, Vangeneugden T, Lefebvre E: Should we now adopt the HIV-RNA < 50 copy endpoint for clinical trials of antiretroviral-experienced as well as naive patients? *AIDS* 2007;21:1651–1653.
- 12 Beerenwinkel N, Däumer M, Sing T, Rahn-enfuhrer J, Lengauer T, Selbig J, Hoffmann D, Kaiser R: Estimating HIV evolutionary pathways and the genetic barrier to drug resistance. *J Infect Dis* 2005;191:1953–1960.
- 13 Prosperi MC, Altmann A, Rosen-Zvi M, Aharoni E, Borgulya G, Bazso F, Sonnerborg A, Schülter E, Struck D, Ulivi G, Vandamme AM, Vercauteren J, Zazzi M, EuResist and Virolab Study Groups: Investigation of expert rule bases, logistic regression, and non-linear machine learning techniques for predicting response to antiretroviral treatment. *Antivir Ther* 2009;14:433–442.
- 14 Altmann A, Rosen-Zvi M, Prosperi M, Aharoni E, Neuvirth H, Schülter E, Büch J, Struck D, Peres Y, Incardona F, Sonnerborg A, Kaiser R, Zazzi M, Lengauer T: Comparison of classifier fusion methods for predicting response to anti HIV-1 therapy. *PLoS One* 2008;3:e3470.
- 15 Zazzi M, Kaiser R, Sonnerborg A, Struck D, Altmann A, Prosperi M, Rosen-Zvi M, Petroczi A, Peres Y, Schülter E, Boucher CA, Brun-Vezinet F, Harrigan PR, Morris L, Obermeier M, Perno CF, Phanuphak P, Pillay D, Shafer RW, Vandamme AM, van Laethem K, Wensing AM, Lengauer T, Incardona F: Prediction of response to antiretroviral therapy by human experts and by the EuResist data-driven expert system (the EVE study). *HIV Med* 2011;12:211–218.