Arevir: A Secure Platform for Designing Personalized Antiretroviral Therapies Against HIV

Kirsten Roomp[1], Niko Beerenwinkel[2], Tobias Sing[1], Eugen Schülter[3], Joachim Büch[1], Saleta Sierra-Aragon[4], Martin Däumer[4], Daniel Hoffmann[3], Rolf Kaiser[4], Thomas Lengauer[1], and Joachim Selbig[5]

[1] Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany
{roomp, tobias.sing, buech, lengauer}@mpi-sb.mpg.de
[2] Department of Mathematics, University of California, Berkeley, USA
niko@math.berkeley.edu
[3] Center of Advanced European Studies and Research (caesar), Friedensplatz 16, 53175 Bonn, Germany
{eugen.schuelter, daniel.hoffmann}@caesar.de
[4] Institute of Virology, University of Cologne, Fürst-Pückler-Str. 56, 50935 Köln, Germany
saleta@arcor.de, martin.daeumer@medizin.uni-koeln.de, rolf.kaiser@uk-koeln.de
[5] University of Potsdam and Max Planck Institute for Molecular Plant Physiology, Am Mühlenberg 1, 14476 Golm-Potsdam, Germany
selbig@mpimp-golm.mpg.de

Abstract. Despite the availability of antiretroviral combination therapies, success in drug treatment of HIV-infected patients is limited. One reason for therapy failure is the development of drug-resistant genetic variants. In principle, the viral genomic sequence provides resistance information and could thus guide the selection of an optimal drug combination. In practice however, the benefit of this procedure is impaired by (1) the difficulty in inferring the clinically relevant information from the genotype of the virus and (2) the restricted availability of this information. We have developed a secure platform for collaborative research aimed at optimizing anti-HIV therapies, called *Arevir*. A relational database schema was designed and implemented together with a web-based user interface. Our system provides a basis for monitoring patients, decision-support, and computational analyses. Thus, it merges clinical, diagnostic and bioinformatics efforts to exploit genomic and patient therapy data in clinical practice.

1 Introduction

1.1 Antiretroviral Therapy

There are currently 25 licensed, antiretroviral agents (including 5 fixed-dose combinations) available in industrialized countries for the treatment of HIV-infected patients. The majority of these drugs targets one of the two viral enzymes, the protease or the reverse transcriptase (RT). Additionally, a new class of drugs called *entry inhibitors* is

U. Leser, F. Naumann, and B. Eckman (Eds.): DILS 2006, LNBI 4075, pp. 185 – 194, 2006.
© Springer-Verlag Berlin Heidelberg 2006

under development, with a subclass called fusion inhibitors. One such drug has been approved so far, which targets the gp41 protein located on the surface of HIV [1]. However, despite the introduction of combination therapy (called HAART – highly active antiretroviral therapy, usually consisting of three or more drugs) eradication of the virus from the patient's body cannot be achieved by current regimens [2]. Therefore, treatment strategies aim at maximal suppression of the viral load, i.e., the number of free virus particles per mL of the patient's blood serum. Besides the strong side effects of the inhibitors [3], the long-term effectiveness of HAART is also limited by the development of drug-resistant genetic variants [4]. Consequently, HIV resistance testing becomes increasingly important in the management of infected patients.

Resistance testing can be performed either by measuring viral activity in the presence or absence of a drug (phenotypic resistance testing [5]), or by scanning the viral genome for resistance-associated mutations (genotypic resistance testing [6]). It has been shown that patients can benefit from both genotypic and phenotypic testing [7]. Genotyping is faster and cheaper, whereas phenotypic results are easier to interpret. Direct sequencing produces genomic data of about 1200 base pairs of the HIV *pol* gene, which codes for protease and RT. This sequence carries the information about susceptibility or resistance of the patient's virus to each of the available drugs. However, it is challenging to infer resistance from the sequence and the optimal way of interpreting the genotype with respect to clinical outcome is not known.

1.2 Public Databases and Related Work

The HIV Sequence Database in Los Alamos provides a public repository for annotated HIV sequence data and is centered on sequence analysis. The HIV Drug Resistance Database in Stanford, formerly called the HIV RT/Protease Sequence Database, collects and analyzes sequences associated with the development of viral resistance. It is focused on sequences coding for the molecular targets of anti-HIV therapy. It includes drug susceptibility data and clinical histories [8]. The public website contains also a tool, *HIVdb*, for the interpretation of genotypic resistance. This system predicts resistance to a drug by scoring observed mutations in the drug's target protein. Mutation scores are derived manually by human experts based on reviewing links between mutations and resistance phenotypes described in the literature.

Another approach to interpreting genotypic resistance tests that avoids this bias lies in the systematic analysis of large sets of matched genotype-phenotype pairs. Statistical and machine learning methods have been applied successfully to derive models that predict phenotypic resistance from the genotype [9-13]. However, the predicted phenotype is only a first step towards understanding the clinical impact of resistance mutations. Selecting an optimal drug combination has become more difficult, because resistance testing adds complex genomic information to the decision-making process. Furthermore, the growing number of available drugs implies an exponentially growing number of possible drug combinations. Extending the data mining approach to related sequences, therapy histories and clinical outcomes promises to identify the determinants of the clinical resistance phenotype. However, this approach presumes large sets of curated and structured data.

Further related work can be found at the HIV Resistance Response Database Initiative (www.hivrdi.org), the Forum for Collaborative HIV Research (www.hivforum.org) and EuResist (www.euresist.org).