max planck institut informatik

Arevir: a secure platform for designing personalized antiretroviral therapies against HIV

Kirsten Roomp

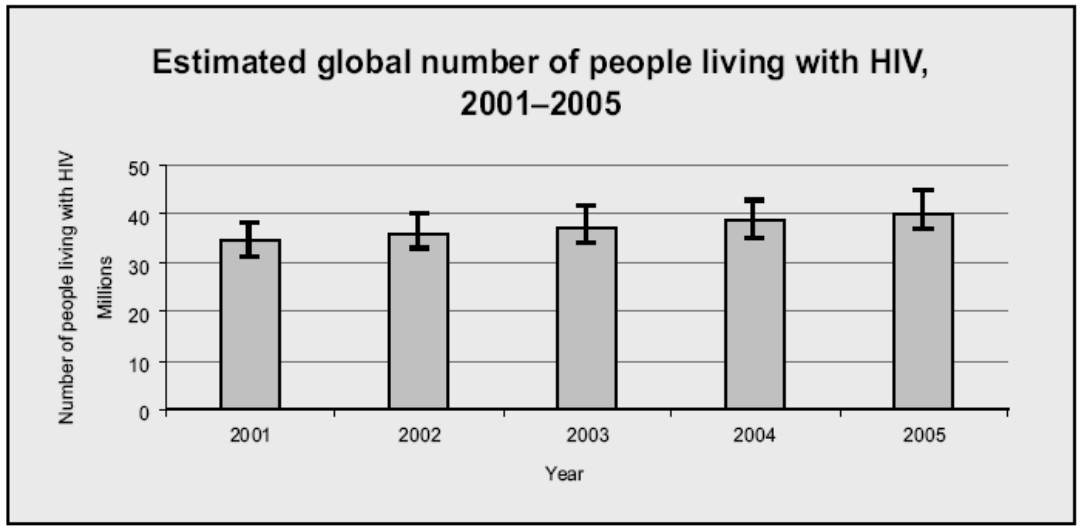
Computational Biology and Applied Algorithmics
Max Planck Institute for Informatics
66123 Saarbrücken
Germany

Overview

- Introduction to HIV therapy
- Arevir
- geno2pheno
- Patient consent and patient identifiers
- Web interfaces
  - Clinician’s interface
  - geno2pheno[resistance]
  - geno2pheno[coreceptor]
  - THEO
- Conclusions

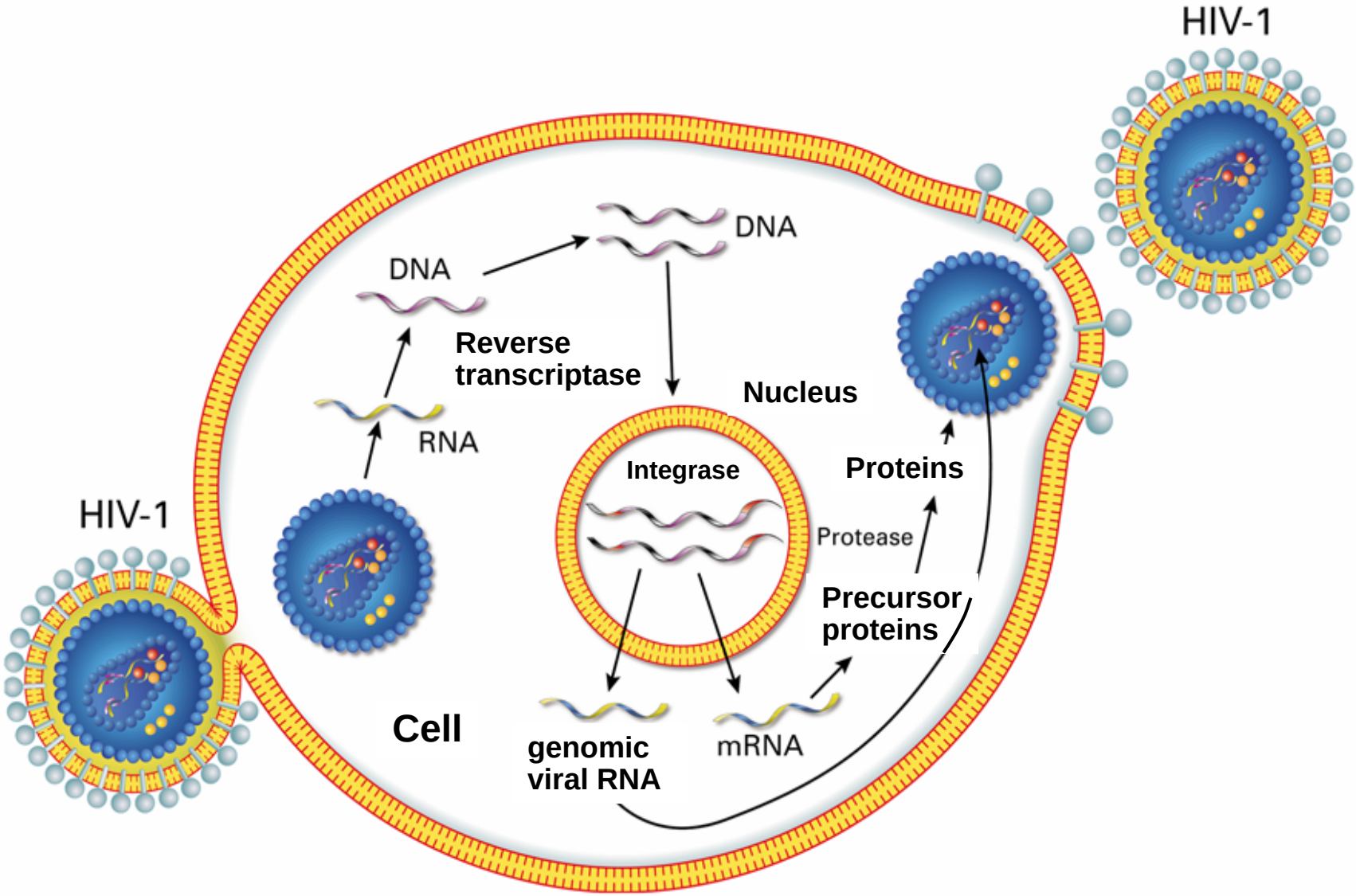
HIV and AIDS Statistics

- World estimates of the HIV & AIDS epidemics in December 2005
  - Number of people living with HIV/AIDS: 40.3 million
  - People newly infected with HIV in 2005: 4.9 million
  - AIDS deaths in 2005: 3.1 million
- Regional statistics
  - 80% of these cases in sub-Saharan Africa
  - More than half a million people are living with HIV in Western Europe

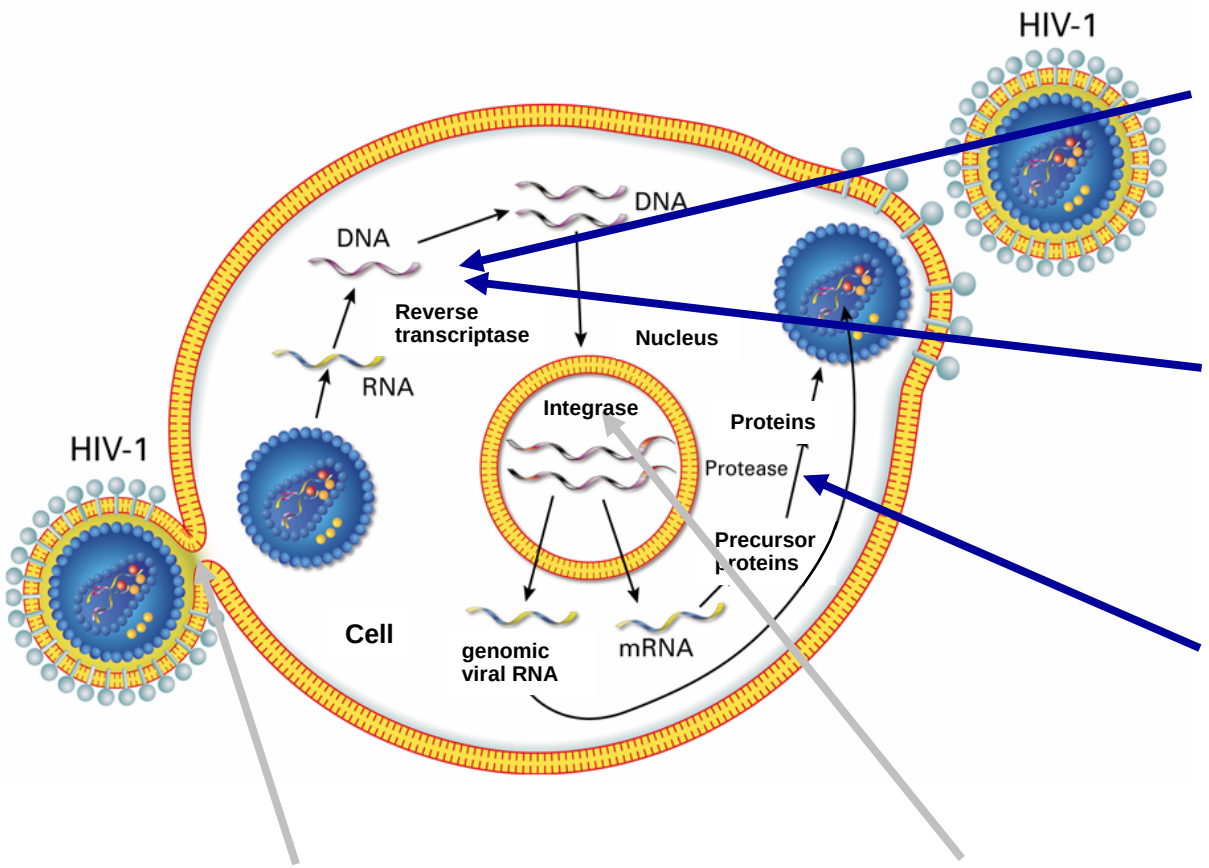


UNAIDS/WHO Report 2005

HIV replication cycle



HIV replication cycle and drug targets



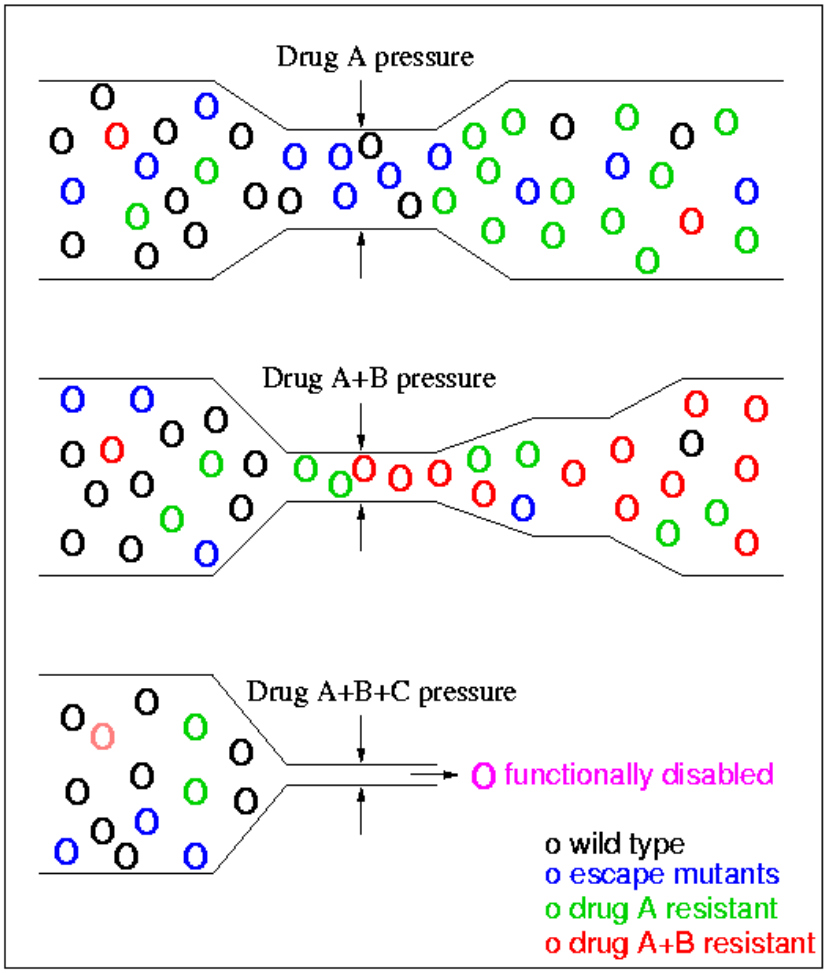
Drug classes

- Nucleoside reverse transcriptase inhibitors (NRTI)
  ZDV, ddI, ddC, d4T, 3TC, ABC, TDF
- Non-nucleoside reverse transcriptase inhibitors (NNRTI)
  EFV, NVP, DLV
- Protease inhibitors (PI)
  SQV, IDV, RTV, NFV, APV, LPV, ATV

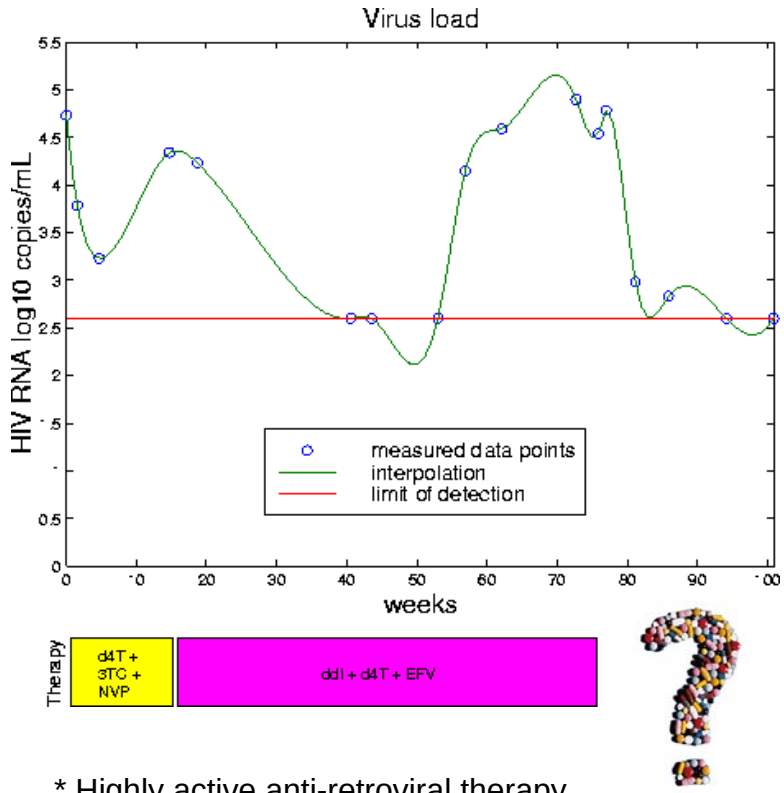
- Entry inhibitors (EI)
  T20, T1249 Sch-C, T22, T134, ALX40-4, AMD3100
- Integrase Inhibitors (II)
  TAK-799

Drug resistance and treatment failure

- HIV is a “moving target”



- combination therapy HAART*:
  - ≥ 3 drugs, ≥ 2 drug classes



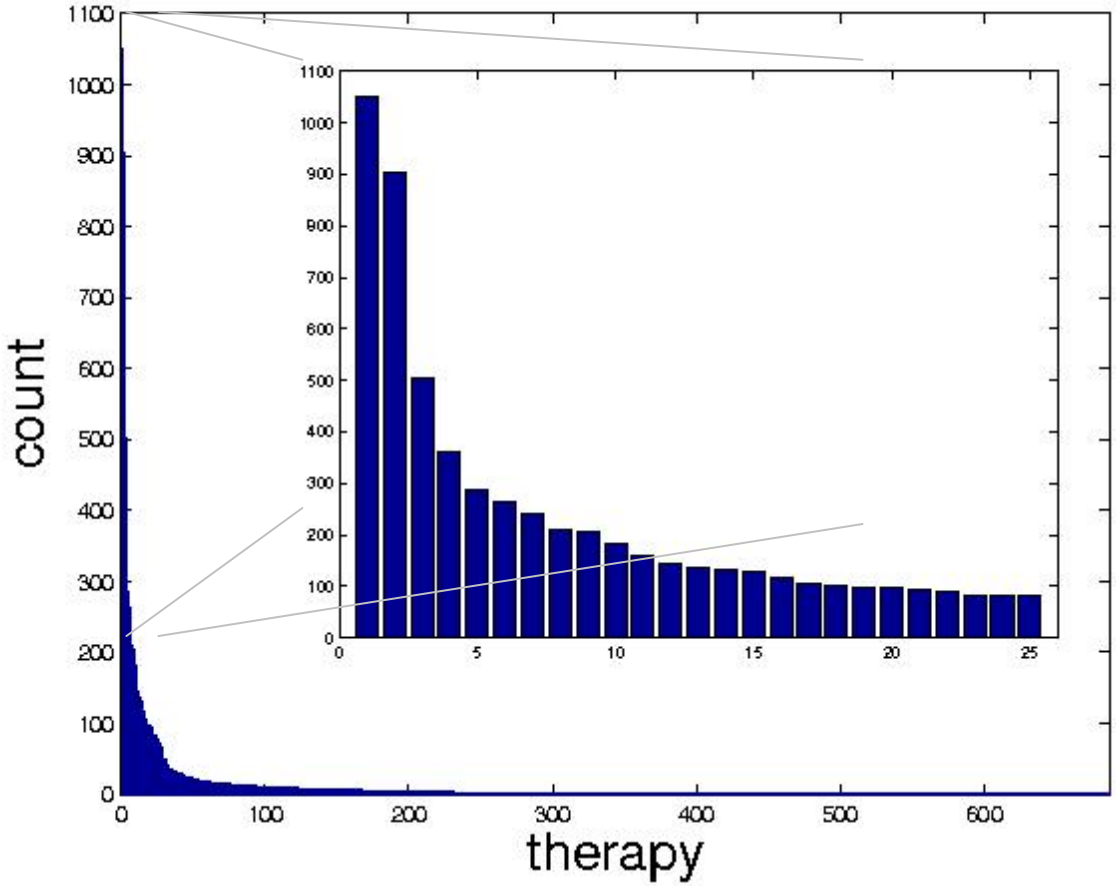
* Highly active anti-retroviral therapy

Resistance testing

- Phenotypic Resistance Testing
  - *in vitro*
  - recombinant assay for pol gene
  - Labour intensive
  - Restricted to specialized labs
  - Takes 4-8 weeks
  - Costs ~1500 US$
  - Output is a single number per drug: easy to interpret

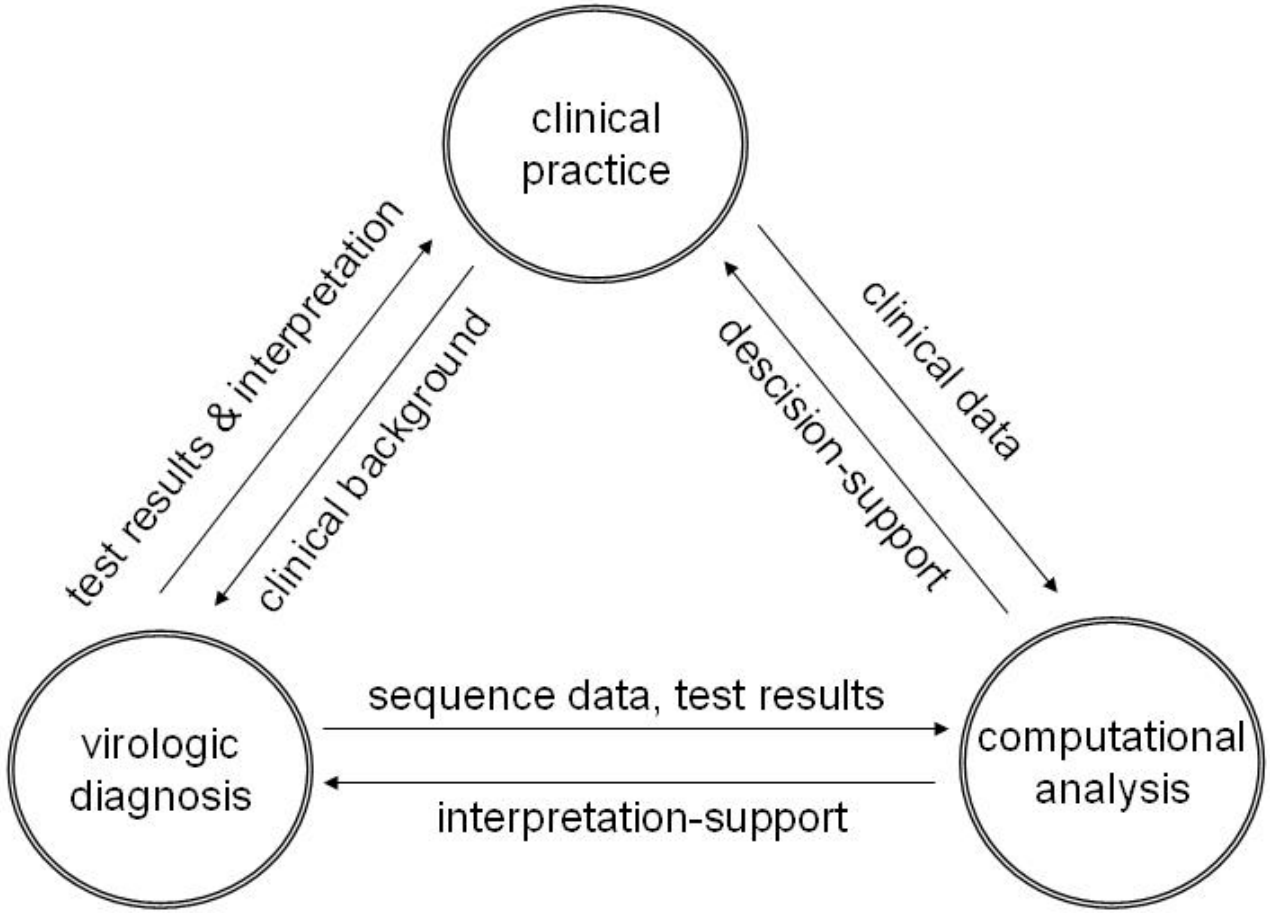
- Genotypic Resistance Testing
  - Sequencing of drug targets in virus from patient's blood serum
  - Standardized kits
  - No infectious virus needed
  - Takes only a few days
  - Cheaper: ~300 US$
  - Output is the DNA sequence of the viral pol gene: interpretation challenging

Combination Therapies

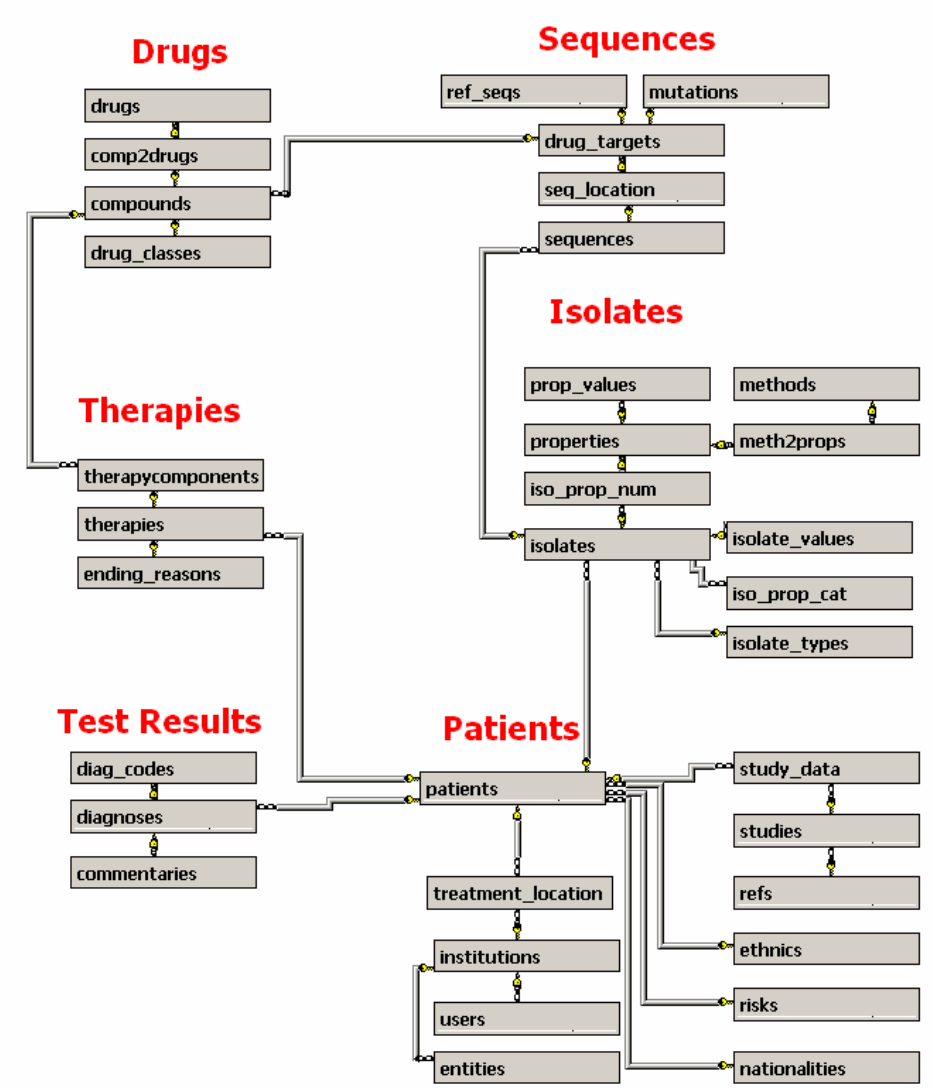


- There are 3,000 – 10,000 reasonable combination therapies
- In clinical practice,only 25 combinations are generally used

Supported information flow in *Arevir*



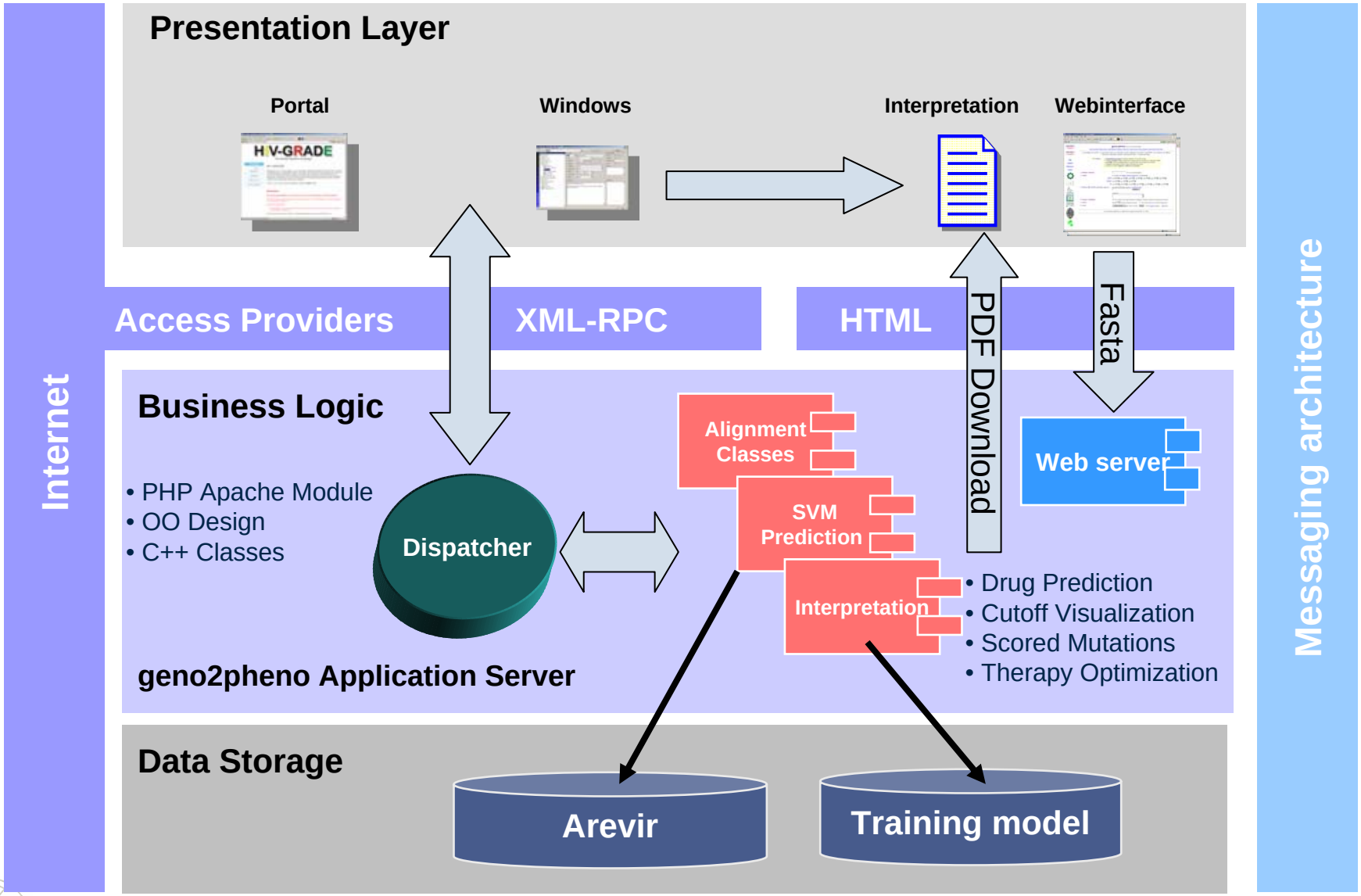
ER Diagram


Drugs
drugs
comp2drugs
compounds
drug_classes
Sequences
ref_seqs
mutations
drug_targets
seq_location
sequences
Isolates
prop_values
methods
properties
meth2props
iso_prop_num
isolates
isolate_values
iso_prop_cat
isolate_types
Therapies
therapycomponents
therapies
ending_reasons
Test Results
diag_codes
diagnoses
commentaries
Patients
patients
study_data
studies
refs
ethnics
risks
nationalities
treatment_location
institutions
users
entities

Arevir DB Content

- Current implementation was intended for use on a national level within Germany
- Collaborators from 17 clinical centers, 3 virologic labs and 3 information technology institutes
- July 2006, the database contains
  - 5,720 patients
  - 9,685 therapies
  - 5,365 viral genomic sequences and
  - 48,502 clinical test results
- Virtually all components of the system are scalable to larger settings
- However, since data quality is a key factor and has been identified as a major challenge, emphasis lies on well-defined data sets and close cooperation

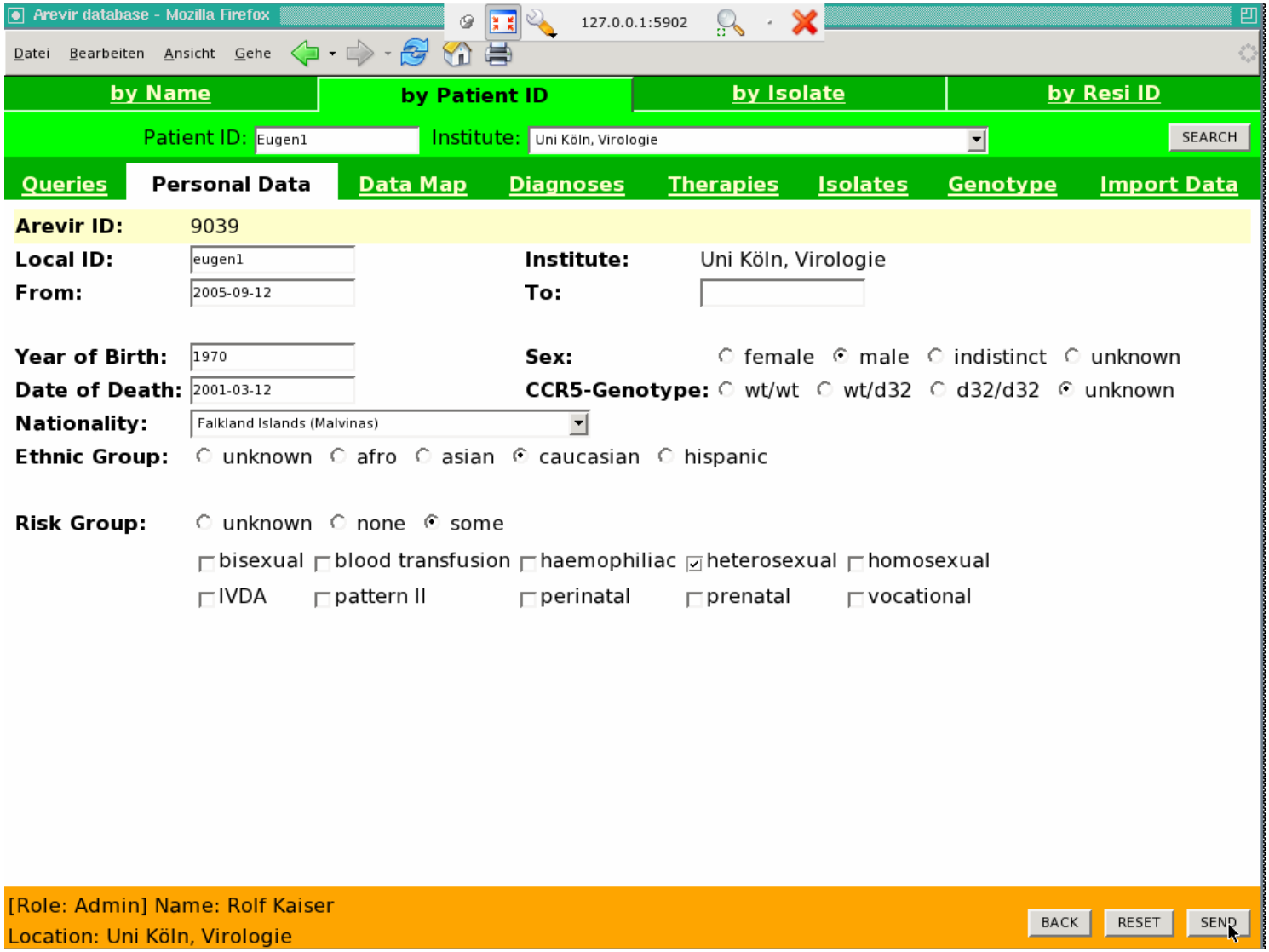
Arevir and geno2pheno



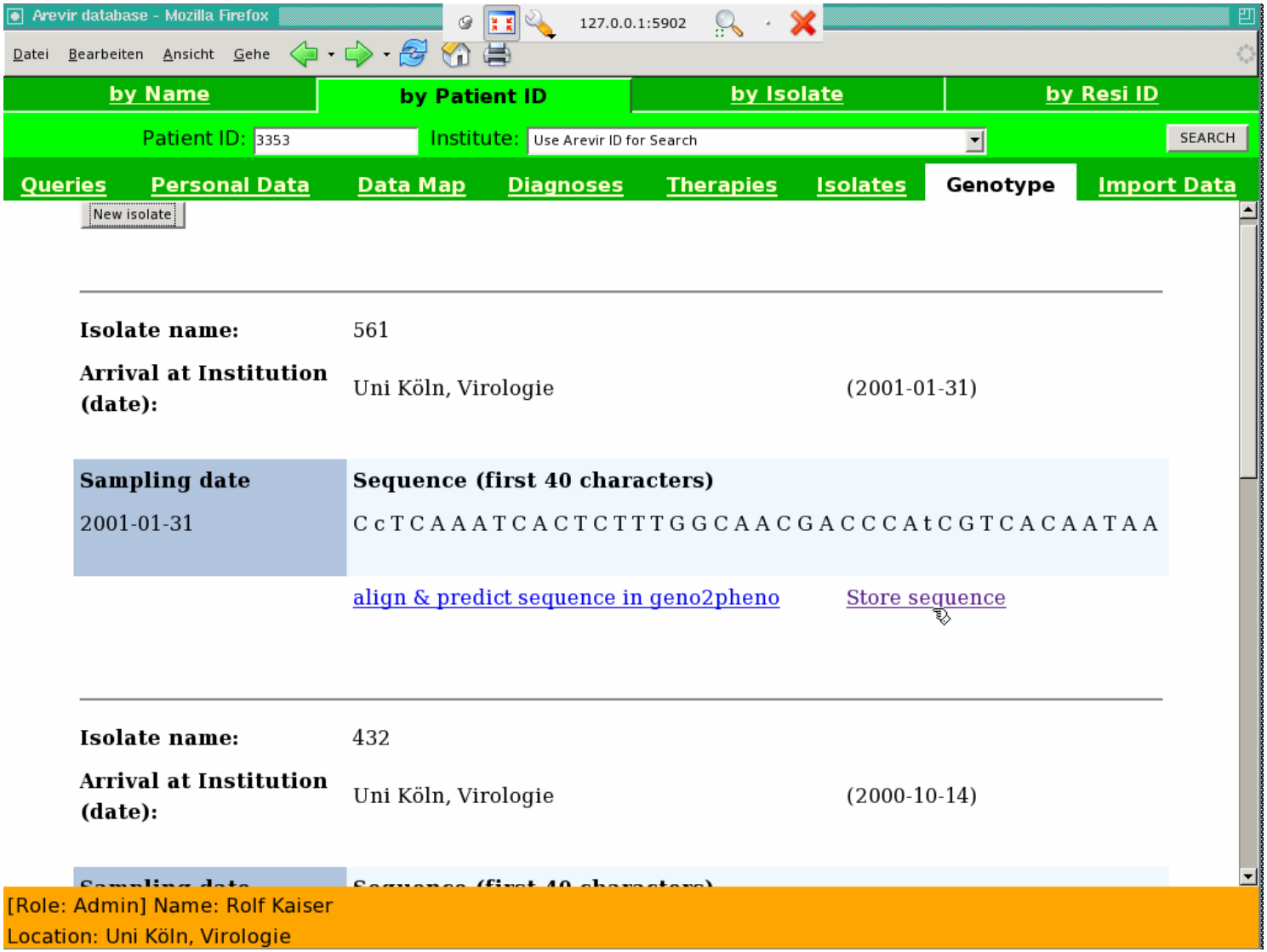
Patient Consent and Patient Identifiers

- Patient Consent
  - For enrollment in *Arevir*, patients need to consent explicitly to providing their data and can revoke their agreement at any time
  - They are informed in detail about project goals and technical realizations
- Patient Identifiers
  - Strict security measures allow the data to be accessed over the web by identifying a patient by its name and date of birth
  - Unlike using anonymous patient identifiers, this method assures usability in clinics and promotes data integrity
  - But the restrictive system architecture entails some limitations on speed and ease of use, notably on printing web contents
  - Patient names are not stored in plaintext in the database - we use a one-way hash function to generate pseudonyms. The Secure Hash Algorithm (SHA-1) is applied to patient name and date of birth, producing a 160-bit hash code.
  - Storing pseudonyms instead of plaintext patient names implicates that given the hash function only comparisons between requested patients and the database contents are possible. This procedure minimizes the risk of the database being abused for uncovering HIV-infections.
  - Finally, computational analyses on patient data are performed only on anonymous data by dropping the pseudonyms table prior to further processing.

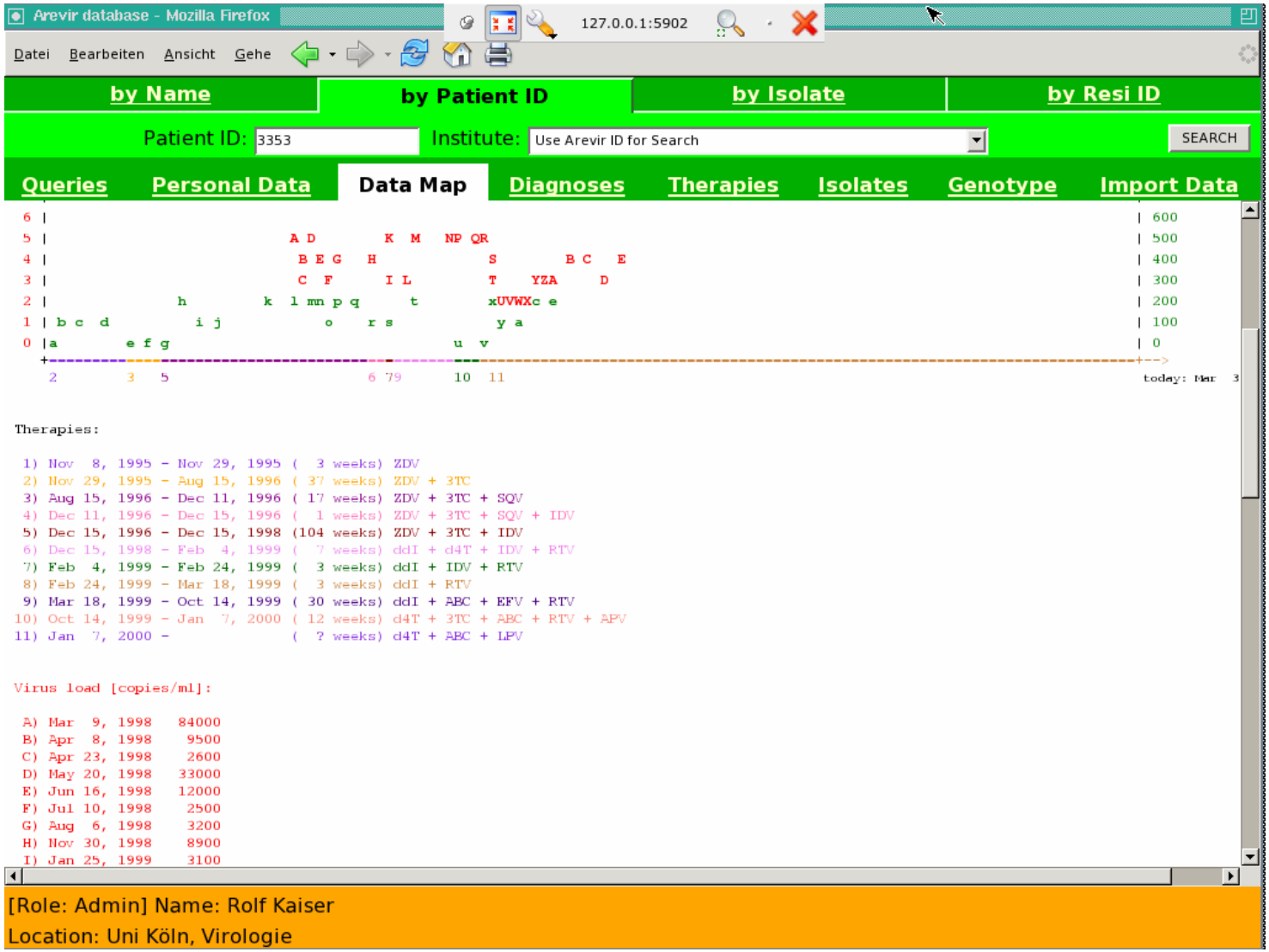
Clinician's Interface

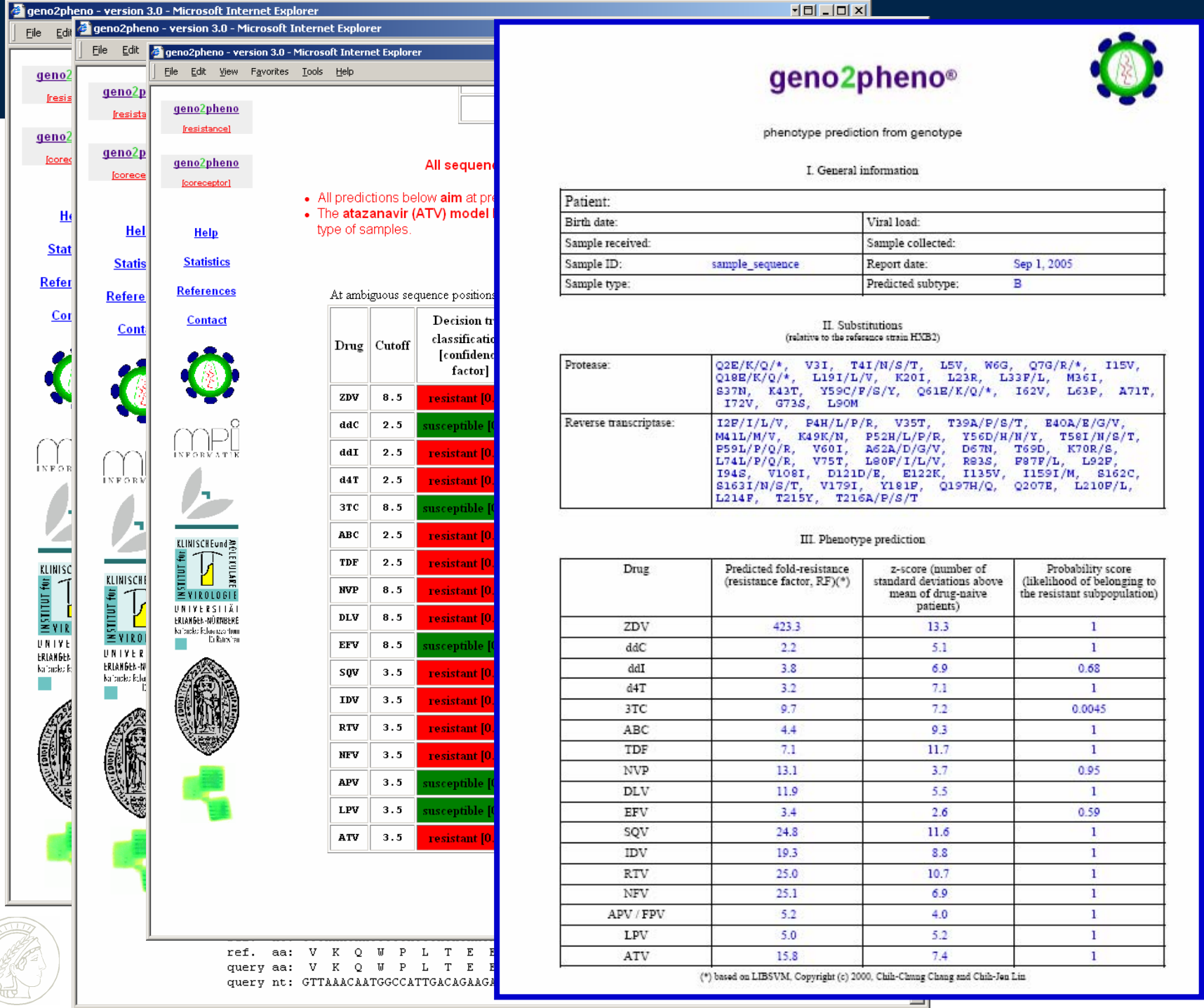


Clinician's Interface cont.



Clinician's Interface cont.





geno2pheno®

phenotype prediction from genotype

I. General information

Patient:			
Birth date:		Viral load:	
Sample received:		Sample collected:	
Sample ID:	sample_sequence	Report date:	Sep 1, 2005
Sample type:		Predicted subtype:	B

II. Substitutions

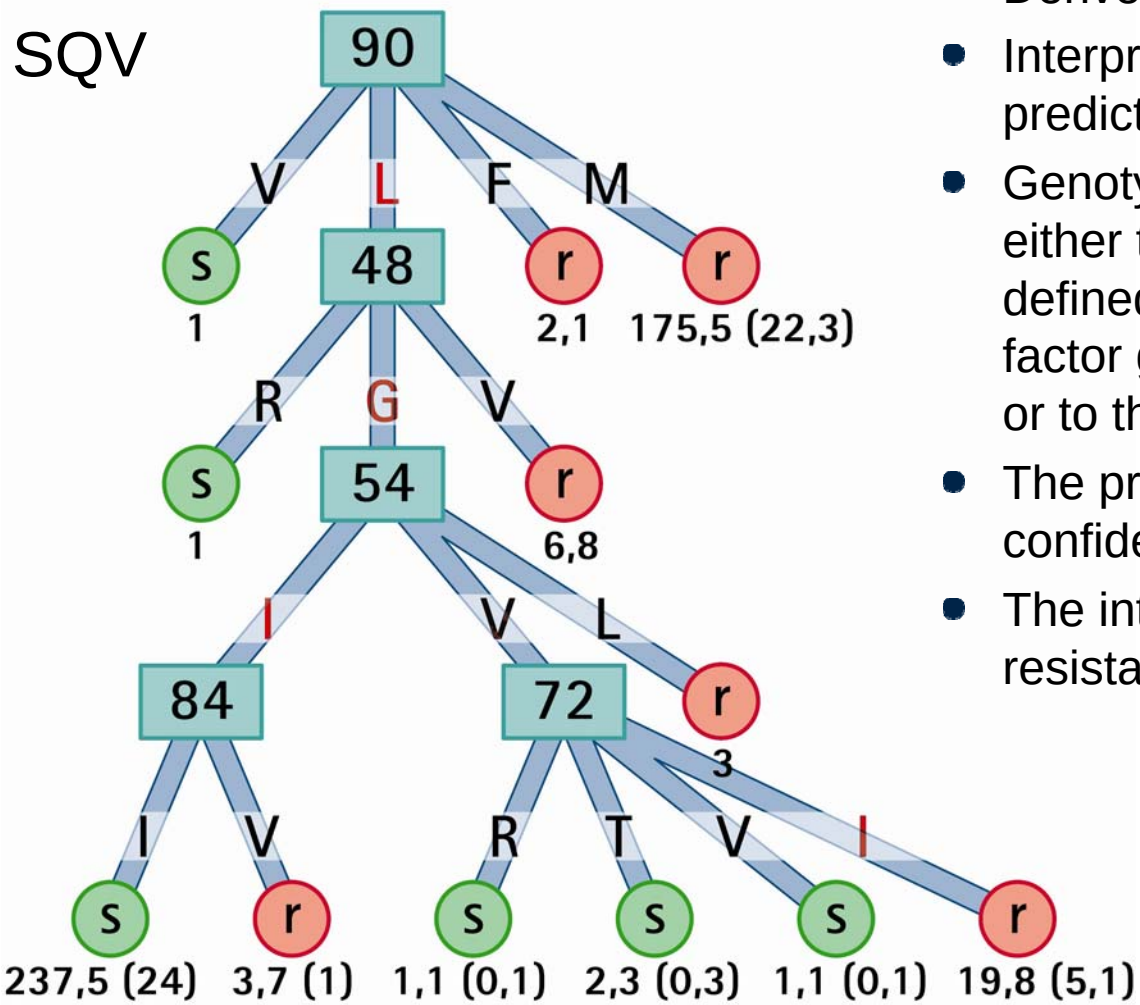
(relative to the reference strain HXB2)

Protease:	Q2E/K/Q/*, V3I, T4I/N/S/T, L5V, W6G, Q7G/R/*, I15V, Q18E/K/Q/*, L19I/L/V, K20I, L23R, L33P/L, M36I, S37N, K43T, Y59C/F/S/Y, Q61E/K/Q/*, I62V, L63P, A71T, I72V, G73S, L90M
Reverse transcriptase:	I2P/I/L/V, P4H/L/P/R, V35T, T39A/P/S/T, E40A/E/G/V, M41L/M/V, K49K/N, P52H/L/P/R, Y56D/H/N/Y, T58I/N/S/T, P59L/P/Q/R, V60I, A62A/D/G/V, D67N, T69D, K70R/S, L74L/P/Q/R, V75T, L80F/I/L/V, R83S, P87F/L, L92P, I94S, V108I, D121D/E, E122K, I135V, I159I/M, S162C, S163I/N/S/T, V179I, Y181F, Q197H/Q, Q207E, L210P/L, L214F, T215Y, T216A/P/S/T

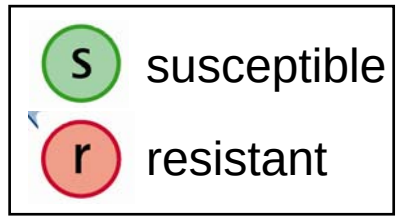
III. Phenotype prediction

Drug	Predicted fold-resistance (resistance factor, RF)(*)	z-score (number of standard deviations above mean of drug-naive patients)	Probability score (likelihood of belonging to the resistant subpopulation)
ZDV	423.3	13.3	1
ddC	2.2	5.1	1
ddI	3.8	6.9	0.68
d4T	3.2	7.1	1
3TC	9.7	7.2	0.0045
ABC	4.4	9.3	1
TDF	7.1	11.7	1
NVP	13.1	3.7	0.95
DLV	11.9	5.5	1
EFV	3.4	2.6	0.59
SQV	24.8	11.6	1
IDV	19.3	8.8	1
RTV	25.0	10.7	1
NFV	25.1	6.9	1
APV / FPV	5.2	4.0	1
LPV	5.0	5.2	1
ATV	15.8	7.4	1

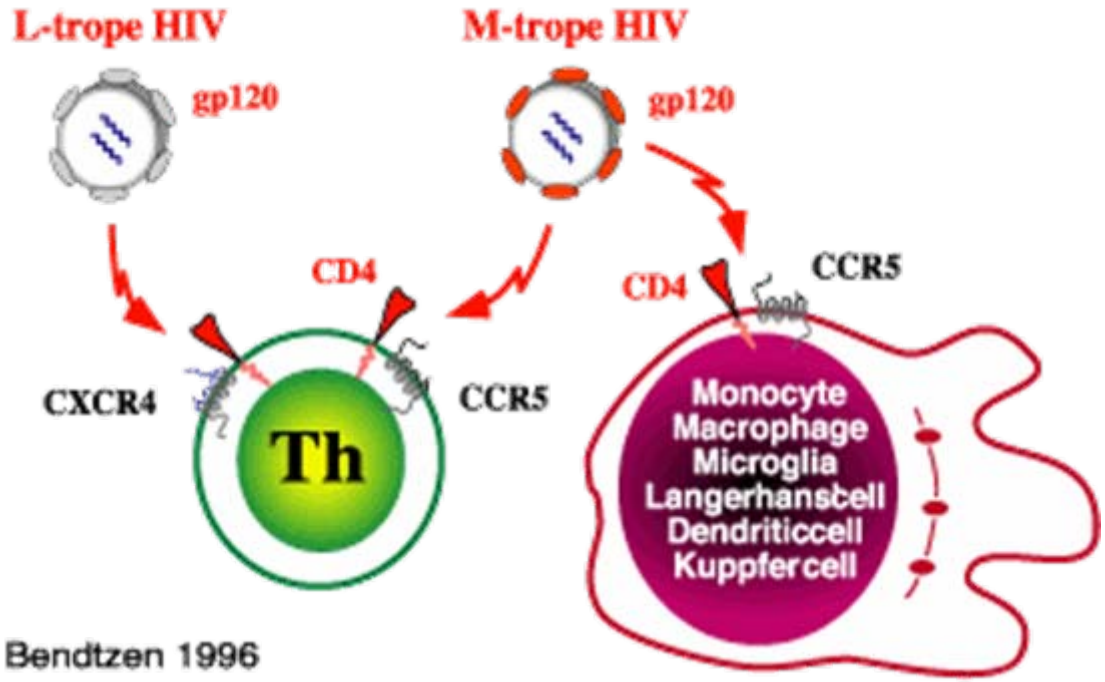
Decision Trees

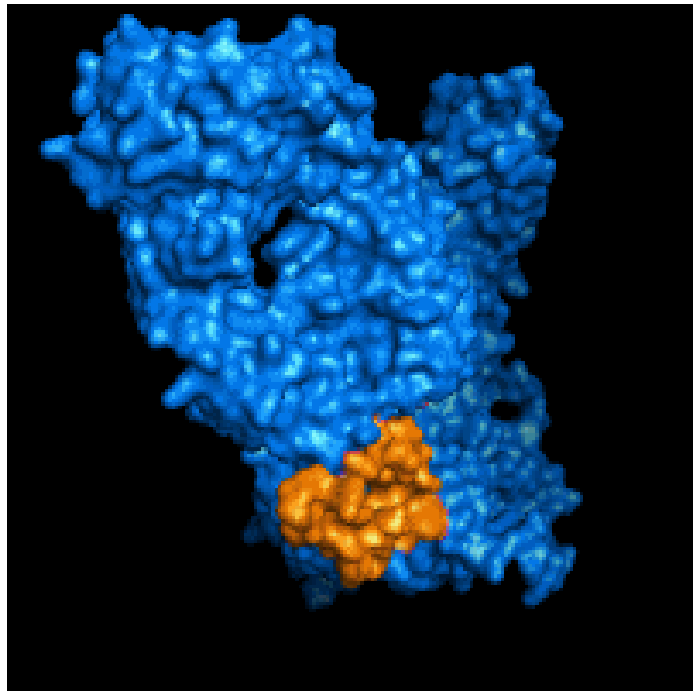


- Derived from phenotypic datasets
- Interpretable statistical model: class prediction by decision trees
- Genotypes are predicted to belong either to the resistant group (red; defined by attaining a resistance factor greater or equal to the Cutoff) or to the susceptible group (green)
- The prediction is accompanied by a confidence factor
- The interdependence of different resistance mutations is represented



Coreceptor Usage





V3 region; 11/25 rule

CTRPNNNTRK**S**IHIGPGRAFYATG**E**IIGDIRQAHC

Fouchier'92 (*J Virol)*

The overall reliability of all sequence motif-based methods for phenotype inference, especially for coreceptor usage prediction, was limited.

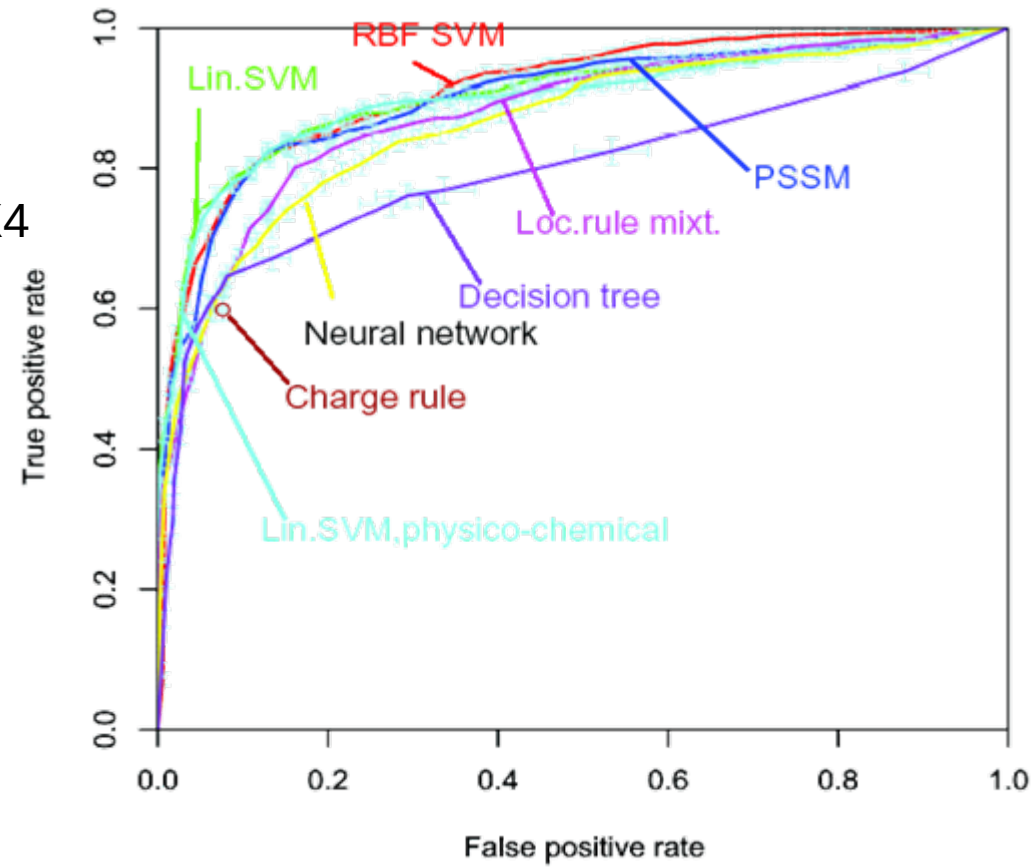
Resch'01 (*Virology)*

These results together suggest that for many V3 backgrounds, basic changes at 11 or 25 are neither necessary nor sufficient for a phenotype switch.

Jensen'03 (*J Virol)*

Predicting R5/X4

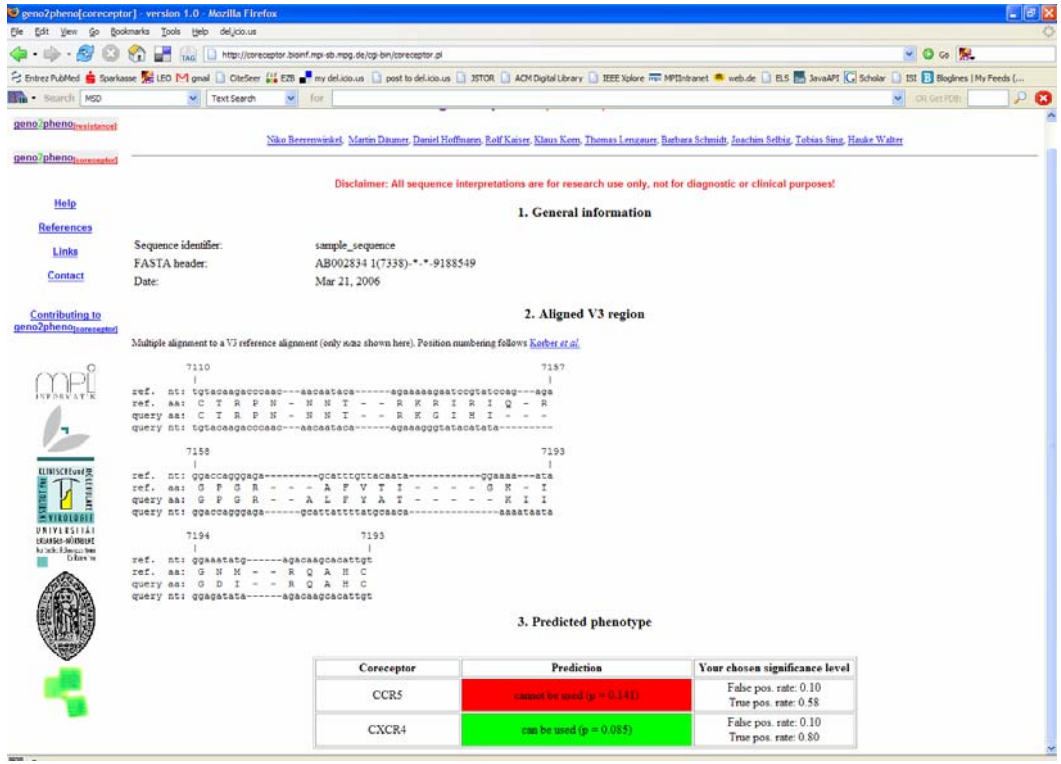
- Method comparison
- 1,110 clonal g/p pairs
  - 332 patients
  - 769 R5, 131 R5/X4, 210 X4
- Setup:
  - "-": R5, "+": R5/X4+X4
  - at most 1 seq./pat.
  - 10x10-fold cross-val.
- Result:
  - SVM vs. 11/25: +16.9%



Briggs'00, Resch'01, Pillai'03, Jensen'03, Sing'04

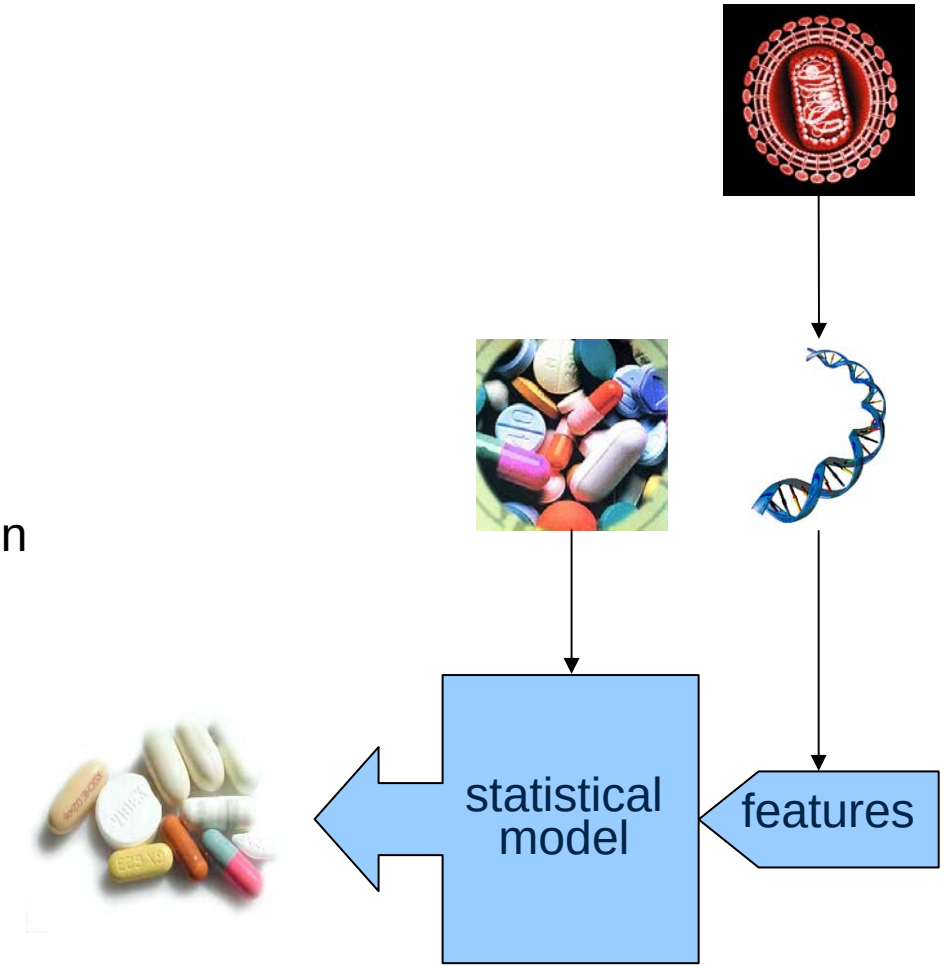
geno2pheno[coreceptor]

- Performance: clinical << clonal data.
- Improvement by combining different markers
- Alternative model to 11/25: many sites contribute
- Structure-based descriptors look promising
- Next: sequence 900 *env* bp of all clinical samples



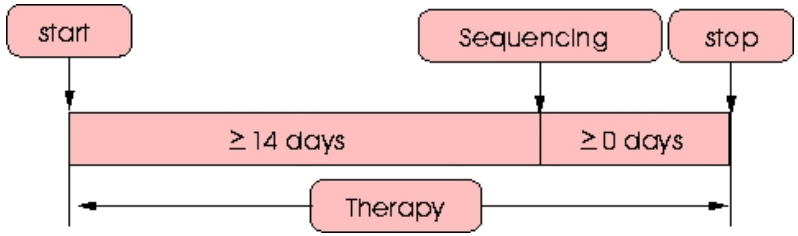
THEO

- Optimize therapy outcome
- given
  - sequences of RT and PRO
  - set of therapies
- “optimal”
  - therapy success
- additional knowledge
  - application pattern of a regimen
  - include/exclude certain drugs

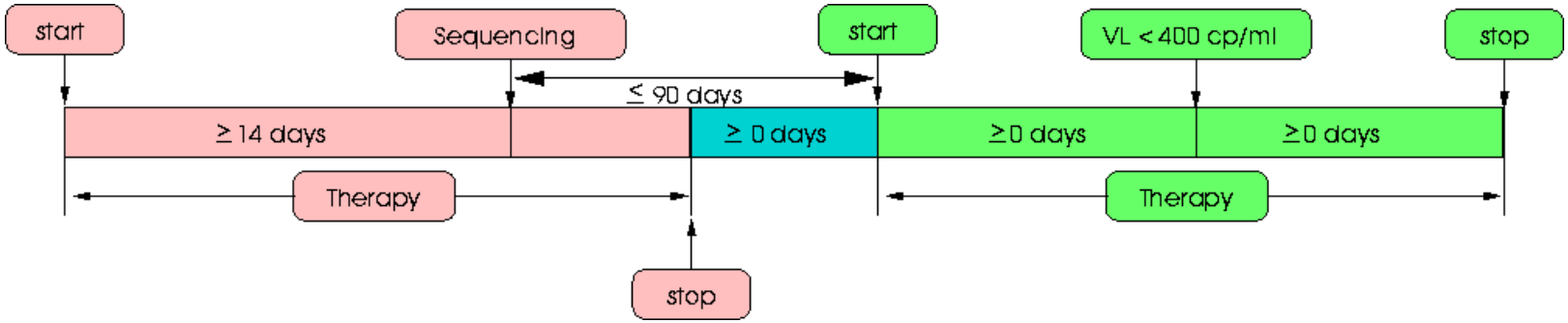


THEO cont.

- Definition of therapy failure and success
  - failure



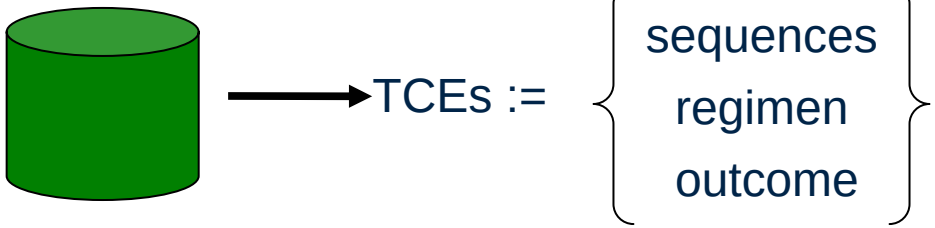
  - success

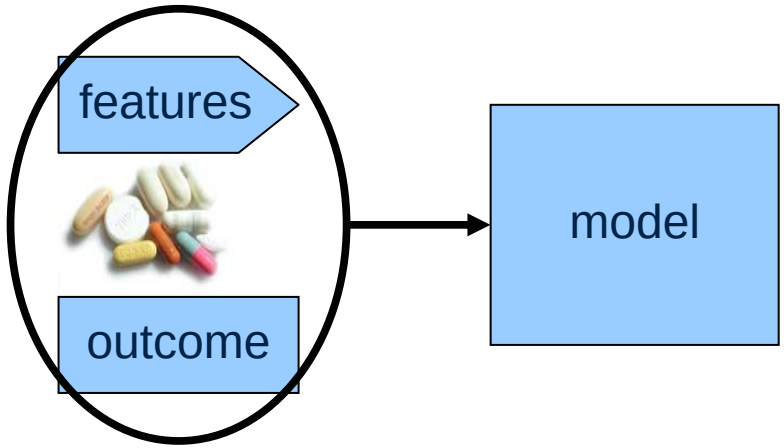


THEO cont.

- Predict therapy outcome
  - given
    - sequences of RT and PRO
    - compounds of the regimen

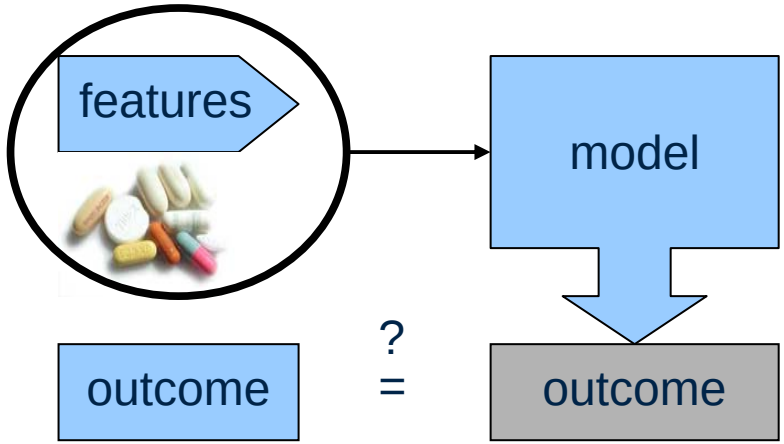
  $\Rightarrow$ binary classification problem





- Method for model training

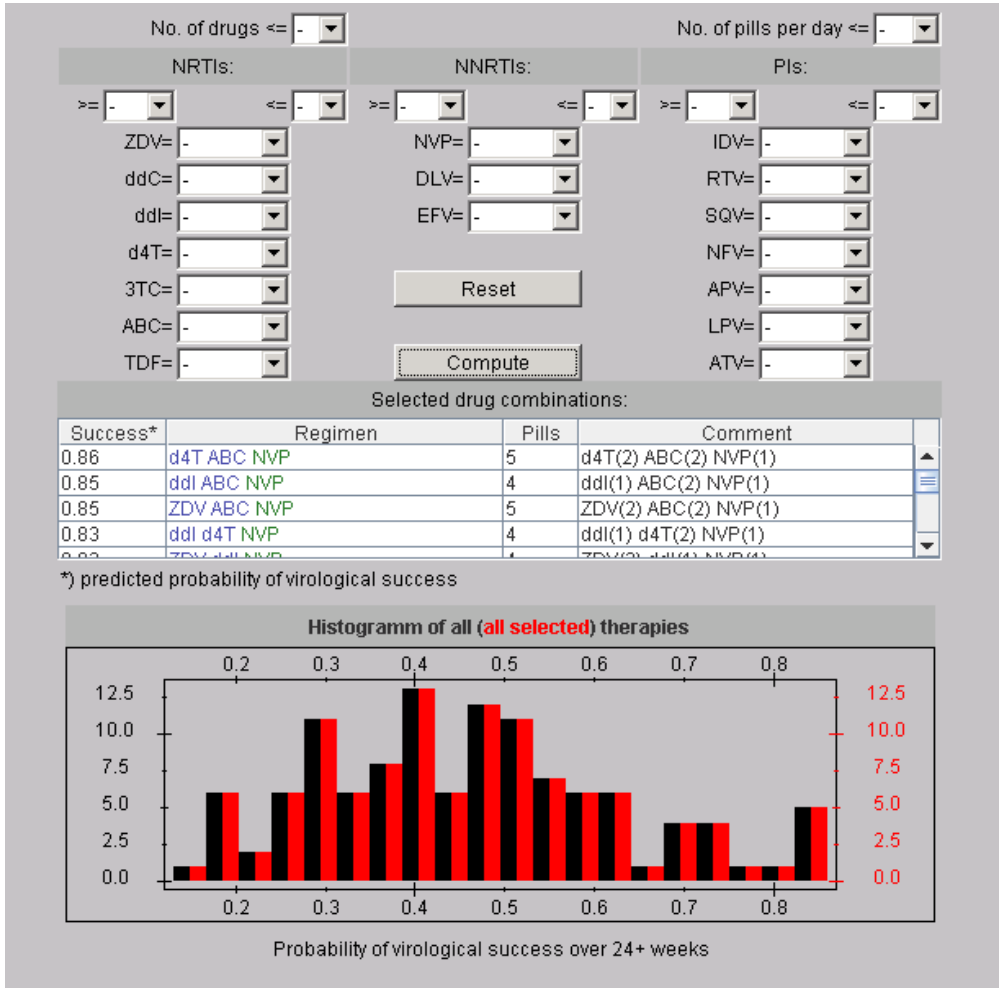
- Validation of
  - model
  - features



THEO Applet

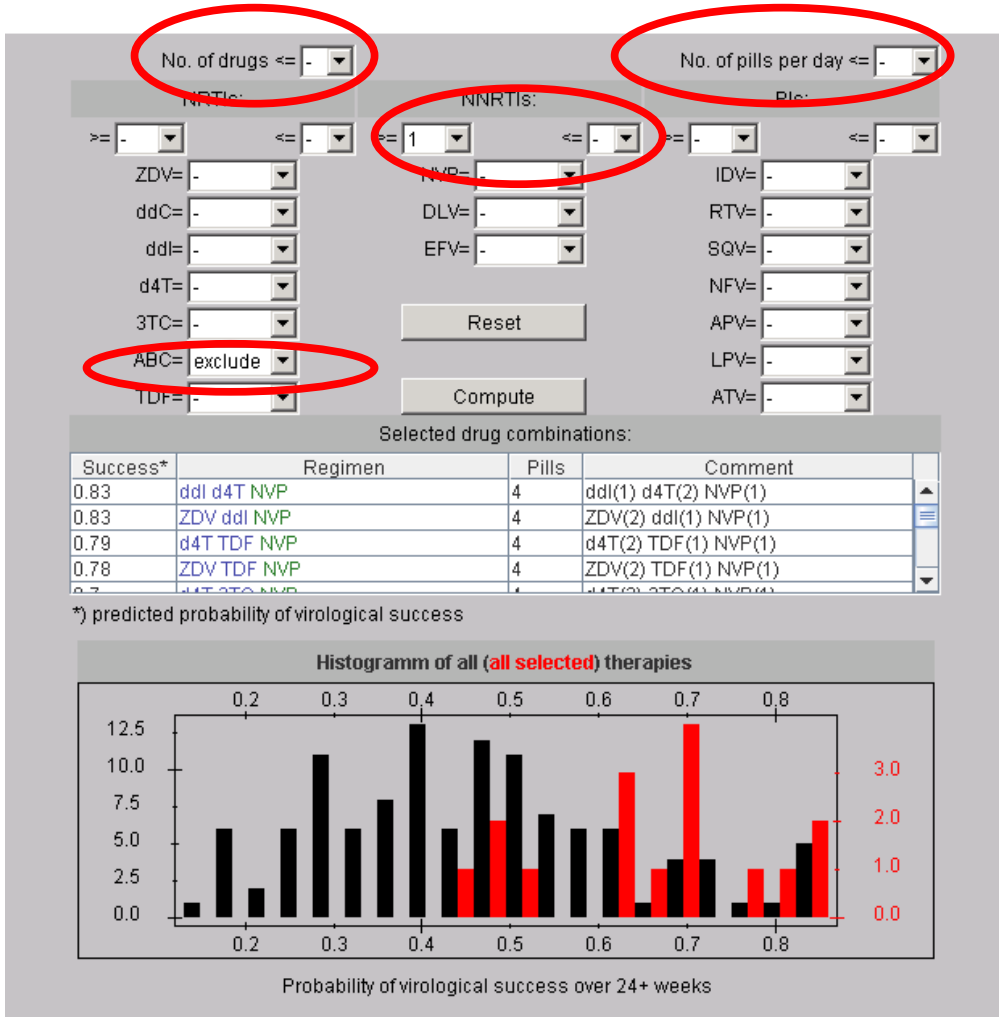
THErapy Optimization

Drug	Predicted fold-resistance (resistance factor, RF)(*)	z-score (number of standard deviations above mean of drug-naive patients)
ZDV	54.7	8.2
ddC	2.4	5.6
ddI	1.6	1.6
d4T	1.6	1.9
3TC	138.4	17.5
ABC	5.0	10.2
TDF	2.1	4.2
NVP	2.6	0.7
DLV	0.6	-2.6
EFV	0.8	-0.9
SQV	4.4	5.0
IDV	29.2	10.2
RTV	54.4	13.6
NFV	19.6	6.3
APV / FPV	8.5	5.4
LPV	26.0	10.9
ATV	10.4	6.1



THEO Applet cont.

- THErapy Optimization
  - limit no. of drugs
  - limit daily burden
  - include/exclude drugs
  - set number of drugs per class



Conclusions

- We have presented a web-based data management system for collaborative research on HIV of direct clinical relevance
- The system has the goal of optimizing antiretroviral therapies in view of viral sequence data
- Our focus is on providing a basis for patient management, evidence-based decision-support and research at the same time
- These seemingly diverse tasks can be unified in a natural way into one system on the basis of a common data model
- This approach may be seen as a real-life example of incorporating bioinformatics methods into clinical practice
- The presented data model proves its flexibility in admitting new clinical parameters, and new drugs with new target molecules

Acknowledgments

Thomas Lengauer	MPI for Informatics, Saarbrücken
Tobias Sing	
Andre Altmann	
Jörg Rahnenführer	
Niko Beerenwinkel	Berkley, USA
Daniel Hoffmann	Caesar, Bonn
Eugen Schülter	
Joachim Selbig	MPI für Pflanzenphysiologie, Golm
Rolf Kaiser	Virologisches Institut, Universität zu Köln
Martin Däumer	
Saleta Sierra-Aragon	
Barbara Schmidt	Institut für klinische und molekulare Virologie, Universität Erlangen-Nürnberg
Hauke Walter	
Klaus Korn	
Jürgen Klein	Fraunhofer Institut für Algorithmen Wissenschaftliches Rechnen, Sankt Augustin
Eberhard Schrüfer	
Marc Oette	Klinik für Gastroenterologie, Universität Düsseldorf
Gerd Fätkenheuer	Klinik für Innere Medizin I, Universität zu Köln
Jürgen Rockstroh	Klinik für Innere Medizin I, Universität Bonn
Ulrich Spengler	
Benedikt Weissbricht	Virologisches Institut, Universität Würzburg
Thomas Berg	Medizinisches Labor Berg, Berlin
Patrick Braun	PZB, Aachen
Valentina Svicher	Klinik für Experimentelle Medizin, Università di RomaTor Vergata, Italy
Francesca Ceccherini-Silberstein	
Richard Harrigan	BC Centre for Excellence in HIV, Vancouver, Canada

Arevir **www.geno2pheno.org**