



***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

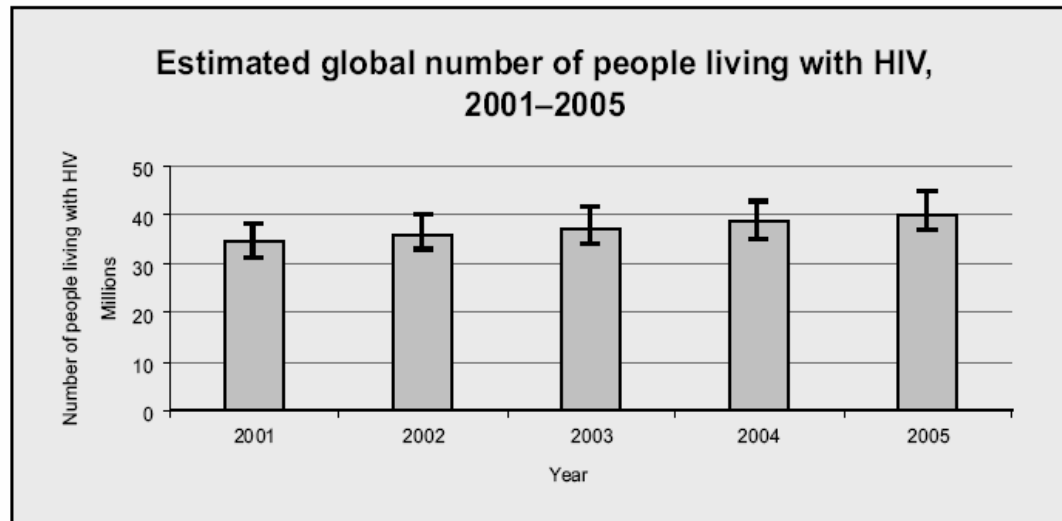
Overview

- 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

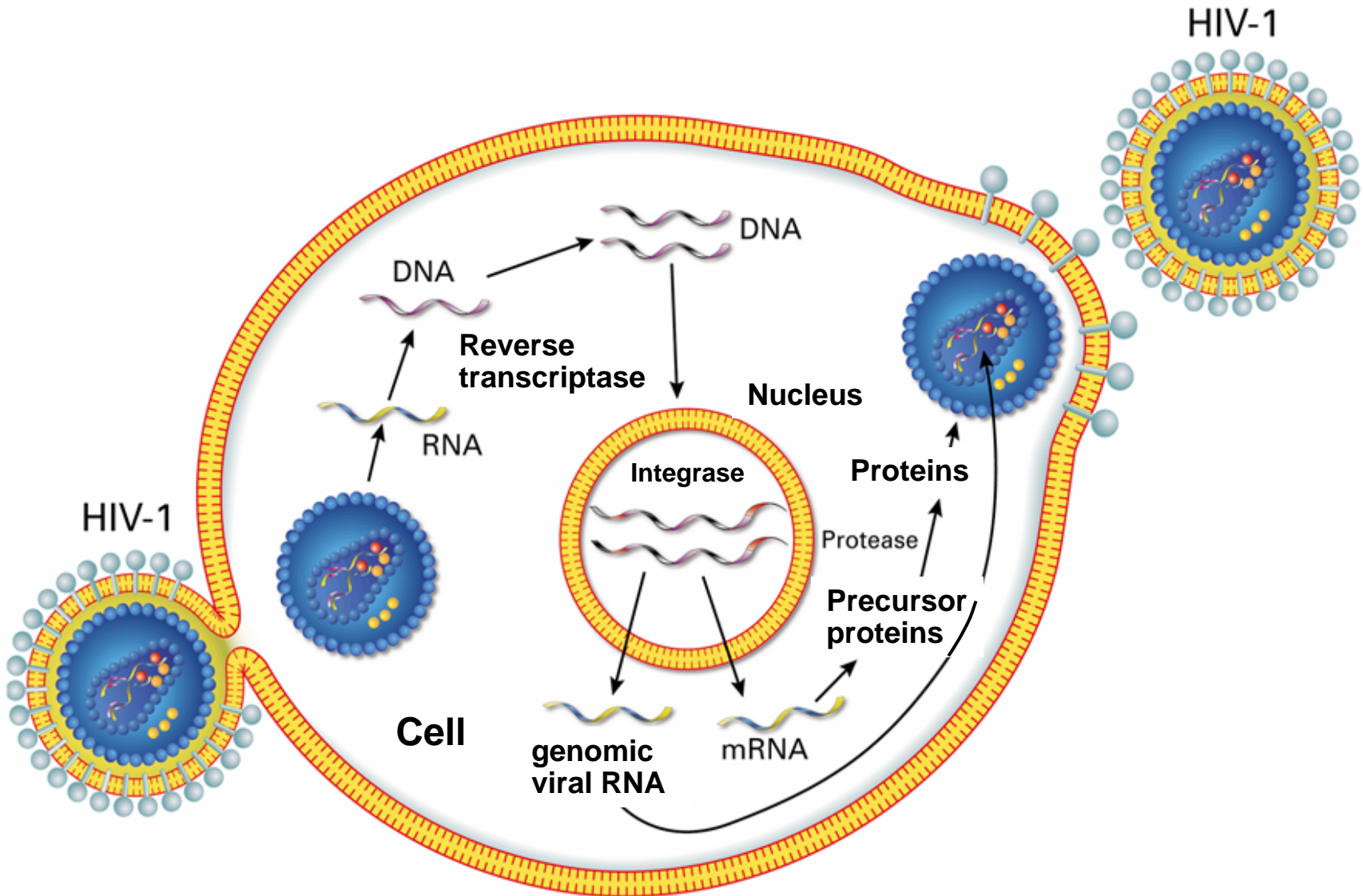
- World estimates of the HIV & AIDS epidemics in December 2005
 - Number of people living with HIV/AIDS: 40.3 million
 - People newly infected with HIV in 2005: 4.9 million
 - AIDS deaths in 2005: 3.1 million
- Regional statistics
 - 80% of these cases in sub-Saharan Africa
 - More than half a million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



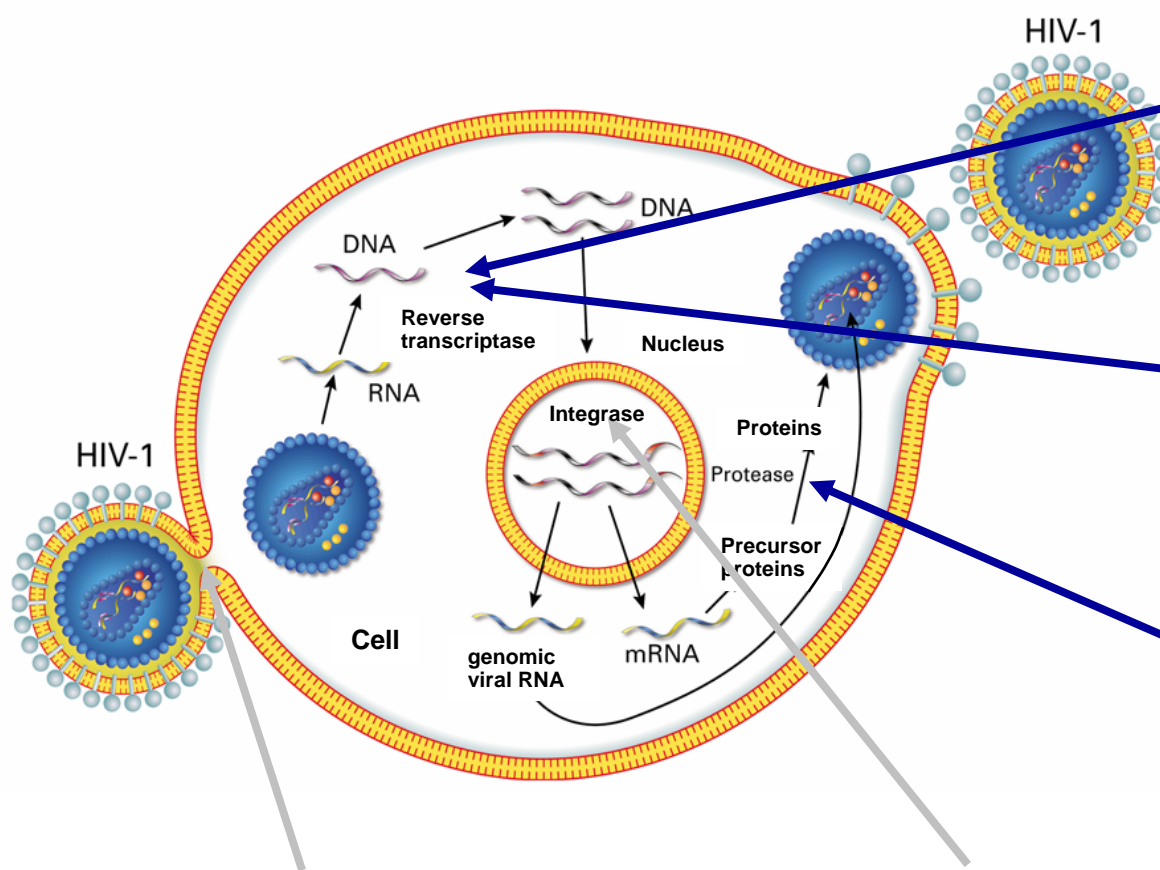
UNAIDS/WHO Report 2005



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

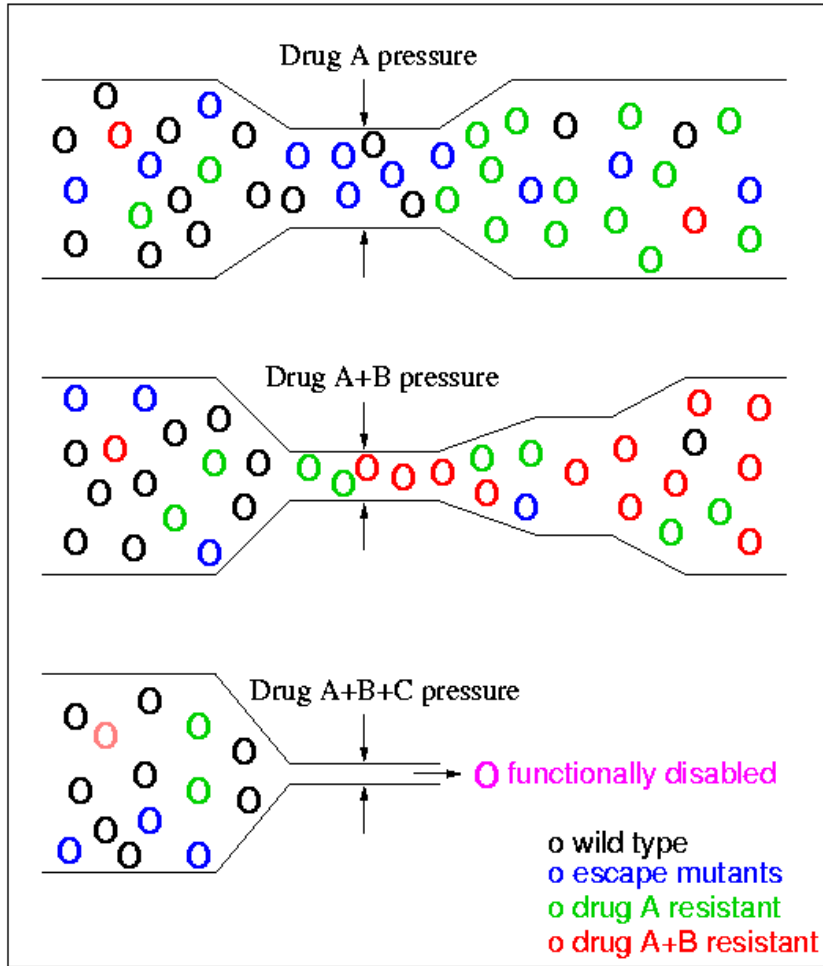
- Entry inhibitors (EI)
T20, T1249 Sch-C, T22, T134, ALX40-4, AMD3100

- Integrase Inhibitors (II)
TAK-799



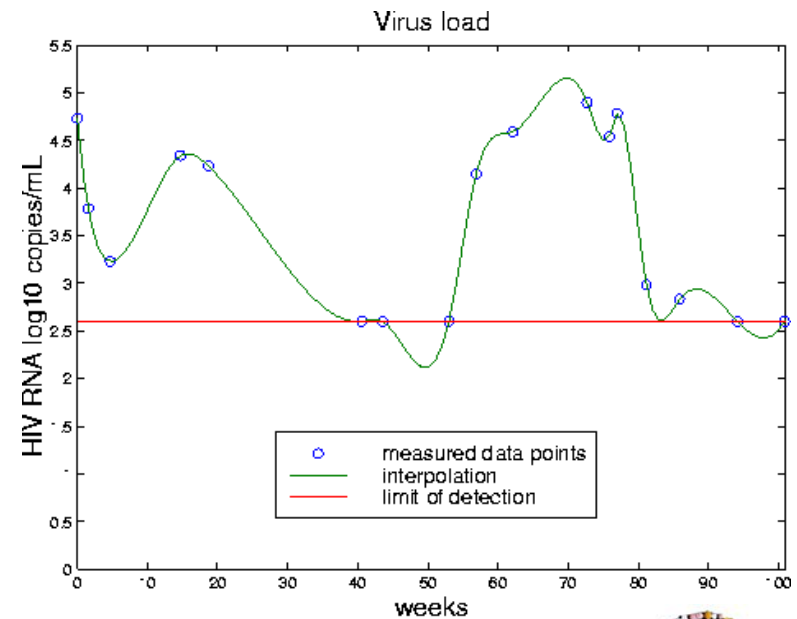
Drug resistance and treatment failure

- HIV is a “moving target”



- combination therapy
HAART*:

- ≥ 3 drugs, ≥ 2 drug classes



* Highly active anti-retroviral therapy

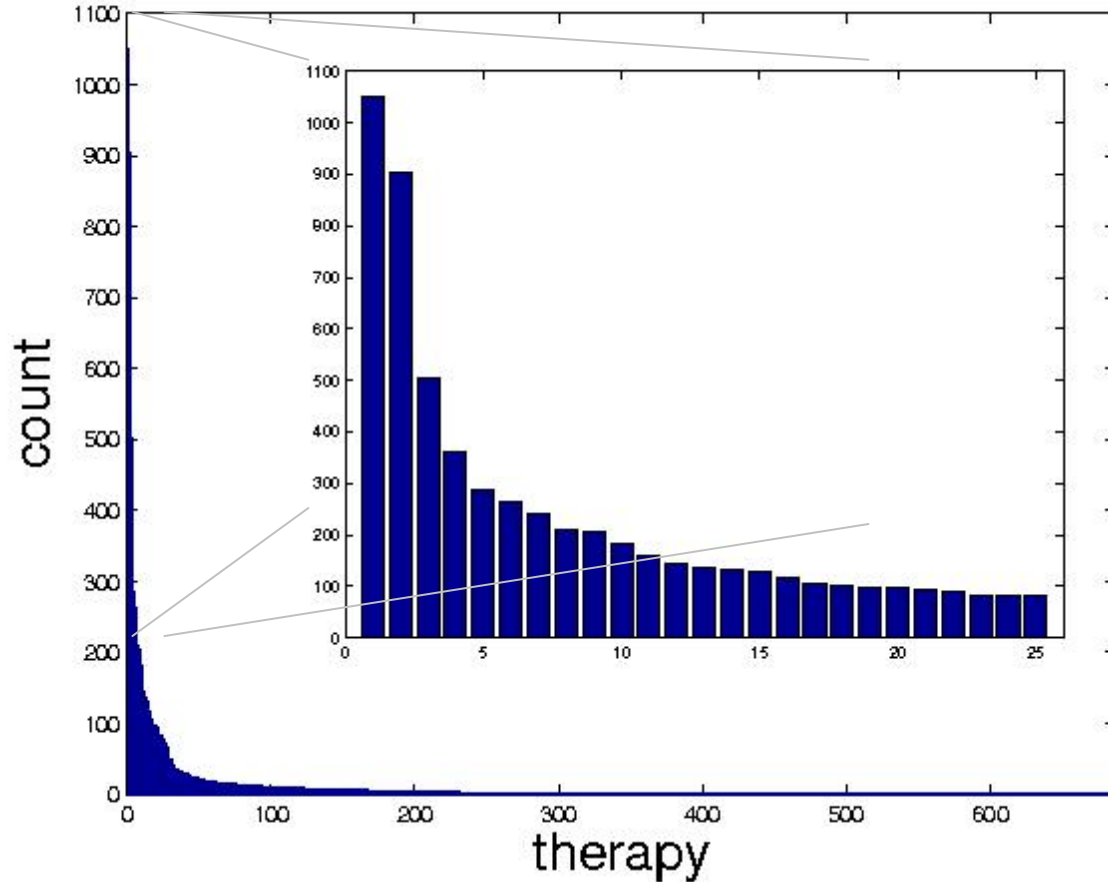


Resistance testing

- 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



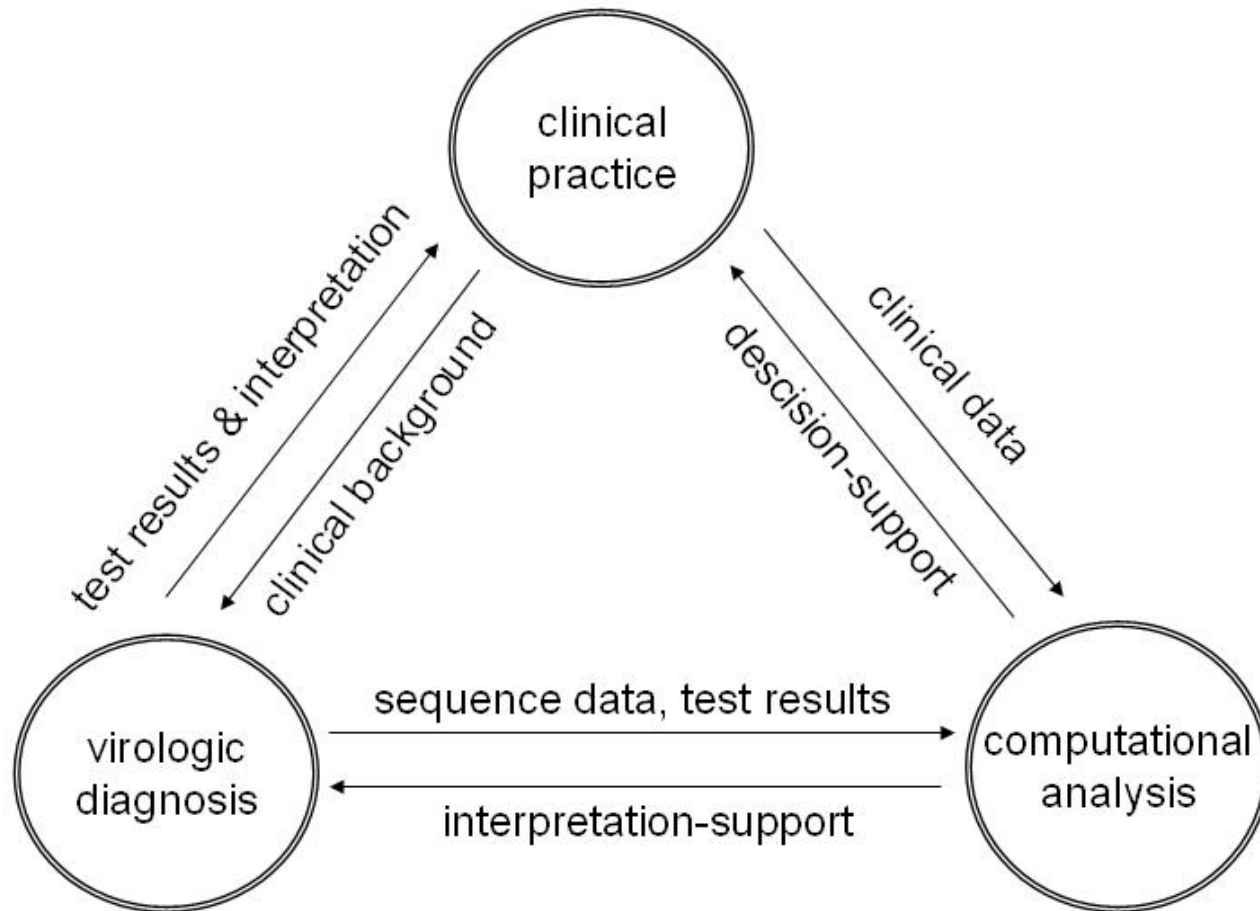
Combination Therapies



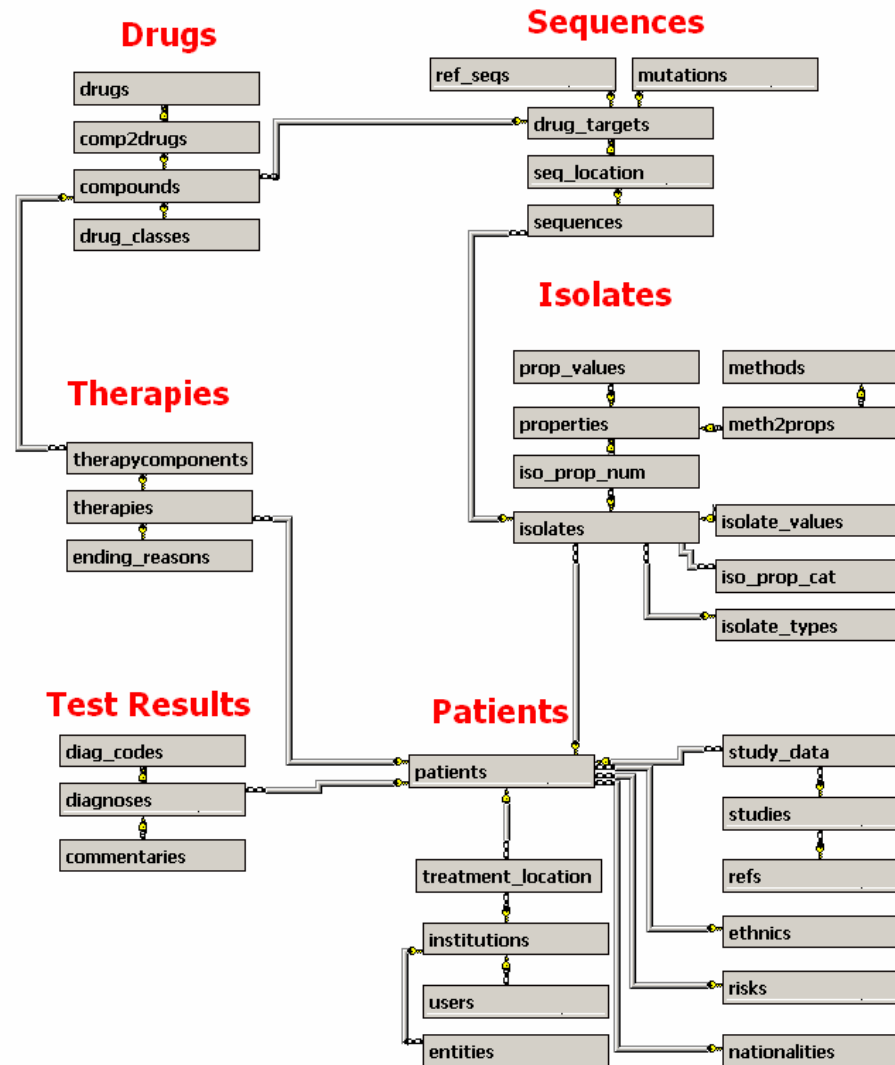
- There are 3,000 – 10,000 reasonable combination therapies
- In clinical practice, only 25 combinations are generally used



Supported information flow in *Arevir*



ER Diagram

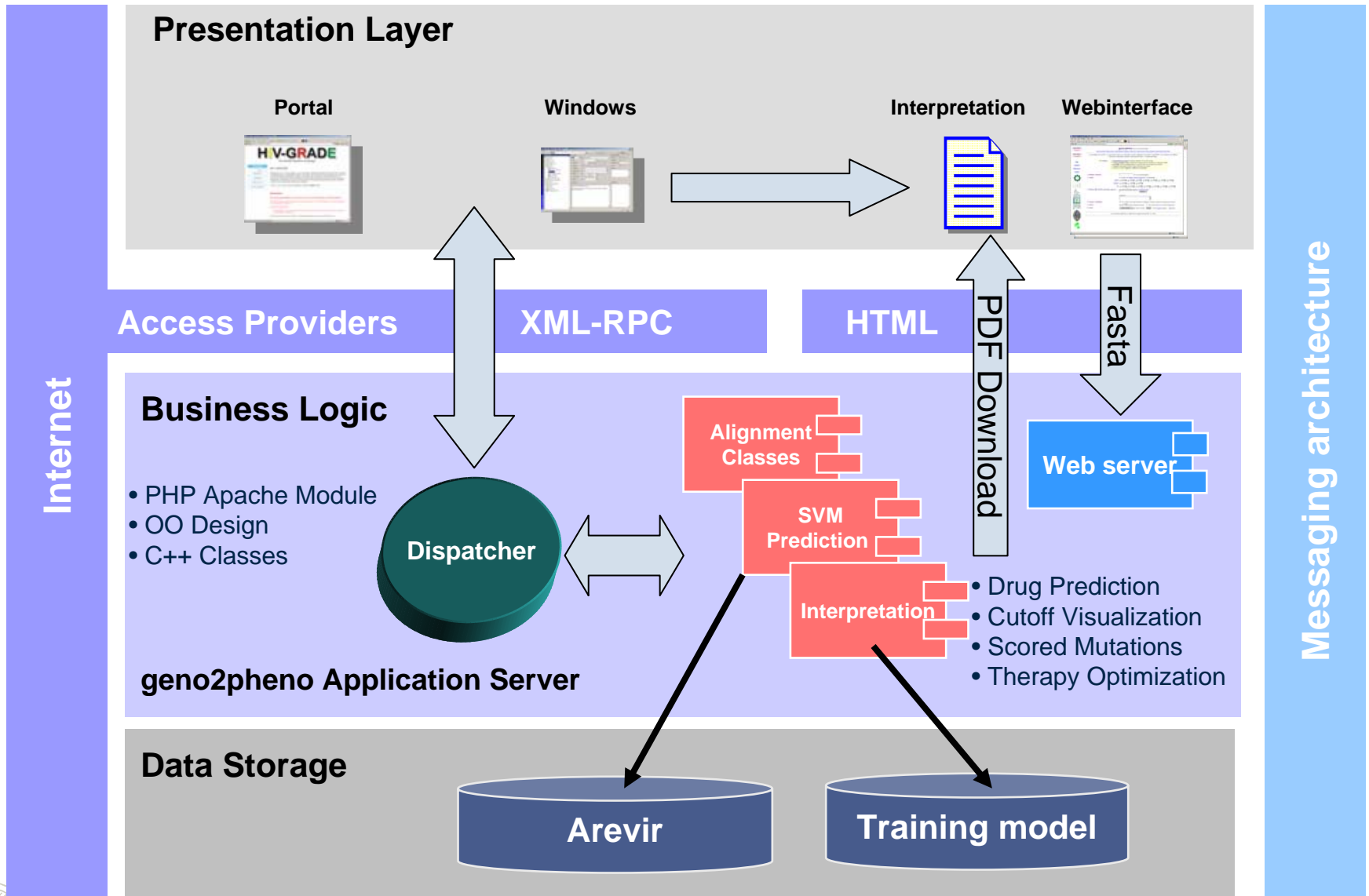


Arevir DB Content

- 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

Arevir database - Mozilla Firefox 127.0.0.1:5902

Datei Bearbeiten Ansicht Gehe

by Name **by Patient ID** **by Isolate** **by Resi ID**

Patient ID: Eugen1 Institute: Uni Köln, Virologie SEARCH

Queries **Personal Data** Data Map Diagnoses Therapies Isolates Genotype Import Data

Arevir ID: 9039

Local ID: **Institute:** Uni Köln, Virologie

From: **To:**

Year of Birth: **Sex:** female male indistinct unknown

Date of Death: **CCR5-Genotype:** wt/wt wt/d32 d32/d32 unknown

Nationality:

Ethnic Group: unknown afro asian caucasian hispanic

Risk Group: unknown none some

bisexual blood transfusion haemophiliac heterosexual homosexual

IVDA pattern II perinatal prenatal vocational

[Role: Admin] Name: Rolf Kaiser
Location: Uni Köln, Virologie

BACK RESET SEND



Clinician's Interface cont.

Arevir database - Mozilla Firefox 127.0.0.1:5902

by Name **by Patient ID** by Isolate by Resi ID

Patient ID: 3353 Institute: Use Arevir ID for Search SEARCH

Queries Personal Data Data Map Diagnoses Therapies Isolates **Genotype** Import Data

New isolate

Isolate name: 561

Arrival at Institution (date): Uni Köln, Virologie (2001-01-31)

Sampling date	Sequence (first 40 characters)
2001-01-31	CcTCAAATCACTCTTTGGCAACGACCCAtCGTCACAATAA

[align & predict sequence in geno2pheno](#) [Store sequence](#)

Isolate name: 432

Arrival at Institution (date): Uni Köln, Virologie (2000-10-14)

Sampling date	Sequence (first 40 characters)
---------------	--------------------------------

[Role: Admin] Name: Rolf Kaiser
Location: Uni Köln, Virologie



Clinician's Interface cont.

A screenshot of a web browser displaying a clinical interface for the Arevir database. The browser window title is "Arevir database - Mozilla Firefox" and the address bar shows "127.0.0.1:5902".

The interface has a green header with navigation tabs: "by Name", "by Patient ID" (selected), "by Isolate", and "by Resi ID". Below these are input fields for "Patient ID: 3353" and "Institute: Use Arevir ID for Search", along with a "SEARCH" button.

The main content area has several tabs: "Queries", "Personal Data", "Data Map" (selected), "Diagnoses", "Therapies", "Isolates", "Genotype", and "Import Data".

The "Data Map" tab displays a grid of data points. The vertical axis is labeled 0 to 6. The horizontal axis is labeled with letters a through v. Data points are represented by letters in various colors (red, green, blue, purple) and sizes. A dashed orange line is drawn across the grid at the level of 0. A vertical dashed line is at the level of 2. The text "today: Mar 3" is visible in the bottom right corner of the grid area.

Below the Data Map, there are sections for "Therapies:" and "Virus load [copies/ml]:".

Therapies:

- 1) Nov 8, 1995 - Nov 29, 1995 (3 weeks) ZDV
- 2) Nov 29, 1995 - Aug 15, 1996 (37 weeks) ZDV + 3TC
- 3) Aug 15, 1996 - Dec 11, 1996 (17 weeks) ZDV + 3TC + SQV
- 4) Dec 11, 1996 - Dec 15, 1996 (1 weeks) ZDV + 3TC + SQV + IDV
- 5) Dec 15, 1996 - Dec 15, 1998 (104 weeks) ZDV + 3TC + IDV
- 6) Dec 15, 1998 - Feb 4, 1999 (7 weeks) ddI + d4T + IDV + RTV
- 7) Feb 4, 1999 - Feb 24, 1999 (3 weeks) ddI + IDV + RTV
- 8) Feb 24, 1999 - Mar 18, 1999 (3 weeks) ddI + RTV
- 9) Mar 18, 1999 - Oct 14, 1999 (30 weeks) ddI + ABC + EFV + RTV
- 10) Oct 14, 1999 - Jan 7, 2000 (12 weeks) d4T + 3TC + ABC + RTV + APV
- 11) Jan 7, 2000 - (? weeks) d4T + ABC + LPV

Virus load [copies/ml]:

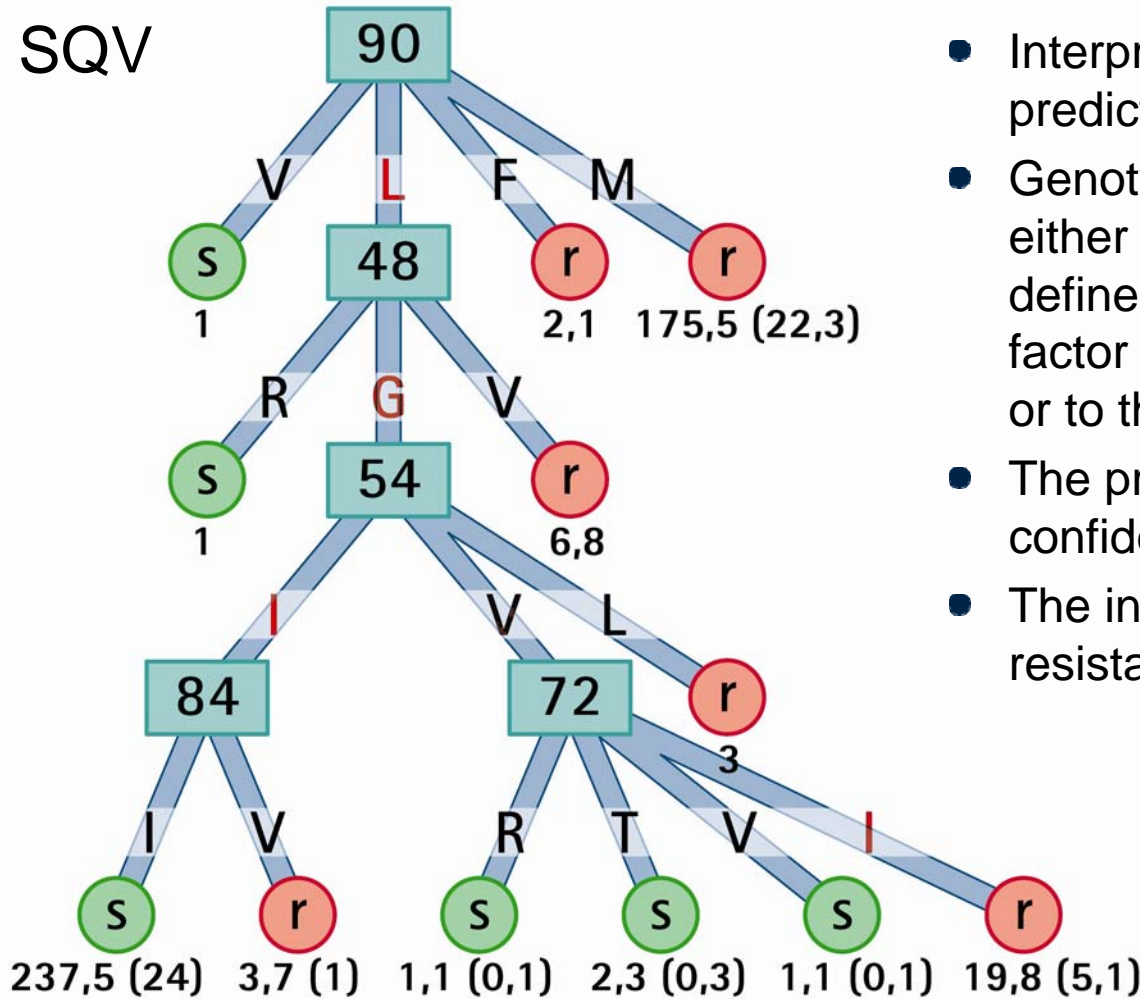
A) Mar 9, 1998	84000
B) Apr 8, 1998	9500
C) Apr 23, 1998	2600
D) May 20, 1998	33000
E) Jun 16, 1998	12000
F) Jul 10, 1998	2500
G) Aug 6, 1998	3200
H) Nov 30, 1998	8900
I) Jan 25, 1999	3100

At the bottom of the interface, there is a status bar with the text: "[Role: Admin] Name: Rolf Kaiser" and "Location: Uni Köln, Virologie".



Decision Trees

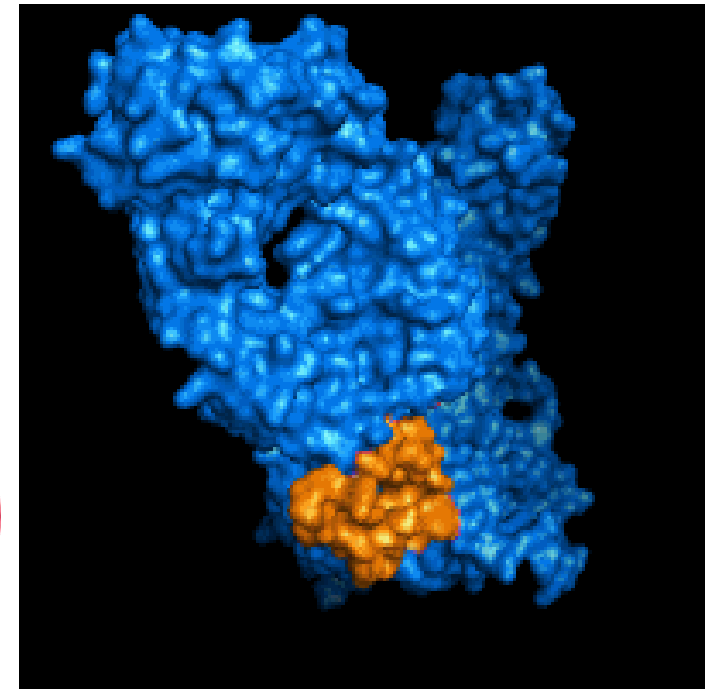
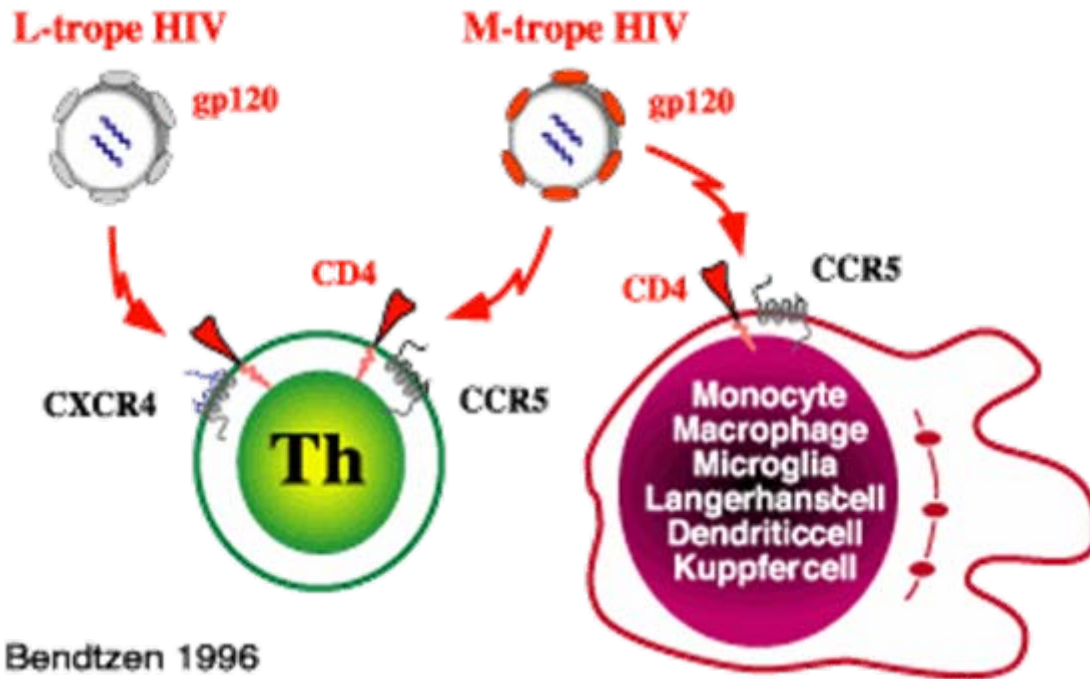
SQV



- 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**IHIGPGRAF**Y**ATG**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*)

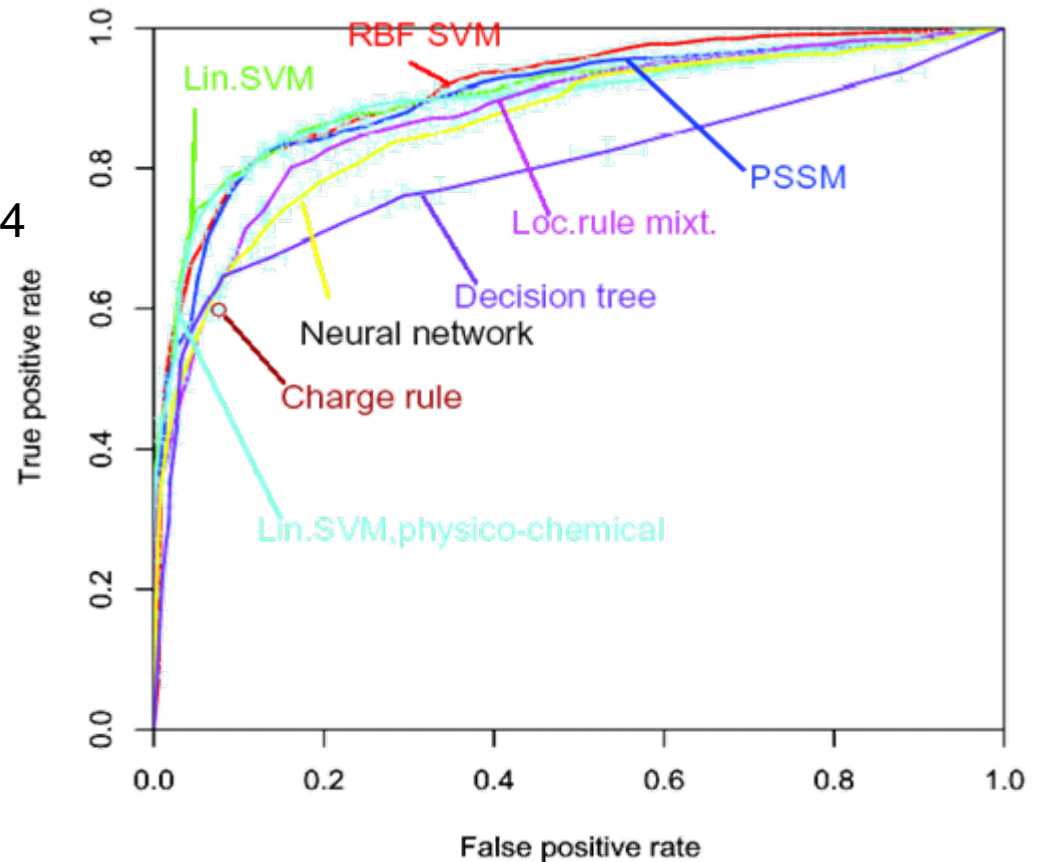
These results together suggest that for many V3 backgrounds, basic changes at 11 or 25 are neither necessary nor sufficient for a phenotype switch.

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

Disclaimer: All sequence interpretations are for research use only, not for diagnostic or clinical purposes!

1. General information

Sequence identifier: sample_sequence
FASTA header: AB002834.1(7338)-*-9188549
Date: Mar 21, 2006

2. Aligned V3 region

Multiple alignment to a V3 reference alignment (only *ncz* shown here). Position numbering follows [Kober et al.](#)

```
7110                               7187
ref. nt: tctacagaccaccac---aaccaatac-----ggaasagatccgatccag---aga
ref. aa: C T R P H - N N T - - R R I R I Q - R
query aa: C T R P H - N N T - - R K G I H I - -
query nt: tctacagaccaccac---aaccaatac-----ggaaggtatcacata-----

7188                               7193
ref. nt: ggcaccaggaga-----gcctctctcaata-----ggaaaa---ata
ref. aa: G P G R - - A F V T T - - G K - T
query aa: G F G R - - A L F Y A T - - - K I I
query nt: ggcaccaggaga-----gcattatttatgcaca-----aaataata

7194                               7193
ref. nt: ggaatatg-----agacagcacattgt
ref. aa: G N K - - R Q A H C
query aa: G D I - - R Q A H C
query nt: ggaatata-----agacagcacattgt
```

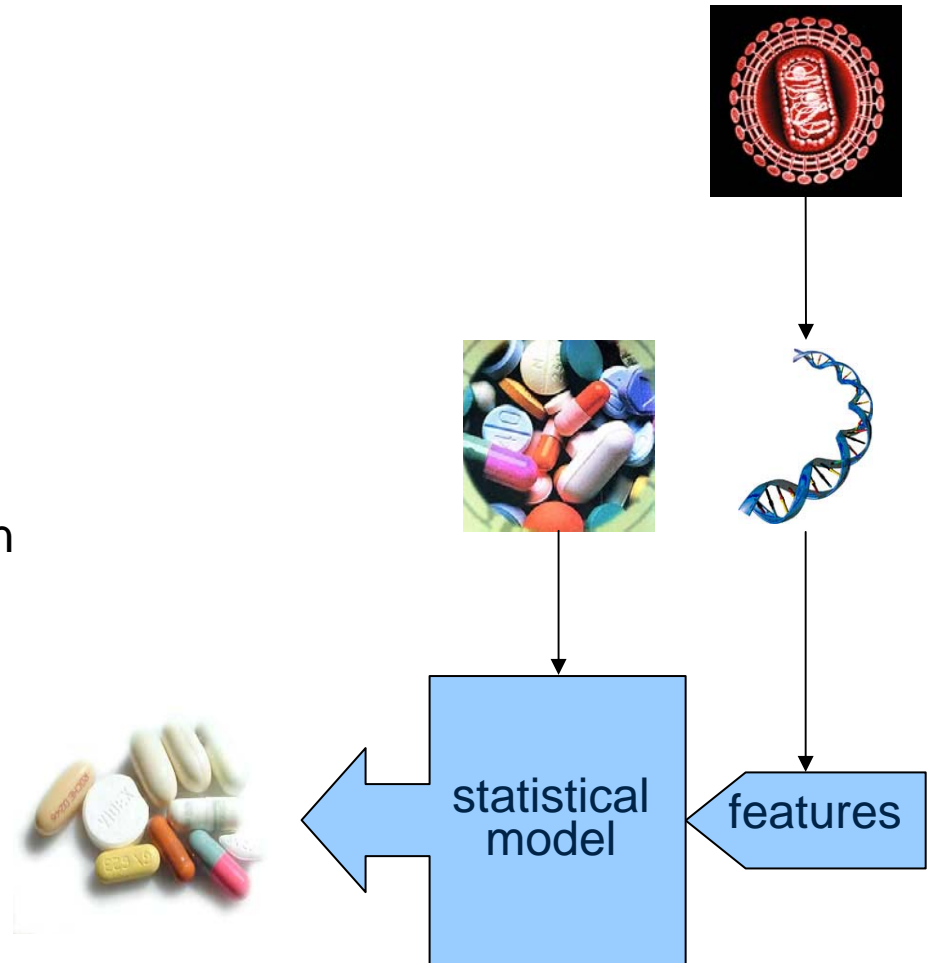
3. Predicted phenotype

Coreceptor	Prediction	Your chosen significance level
CCR5	cannot be used ($p = 0.141$)	False pos. rate: 0.10 True pos. rate: 0.58
CXCR4	can be used ($p = 0.085$)	False pos. rate: 0.10 True pos. rate: 0.80



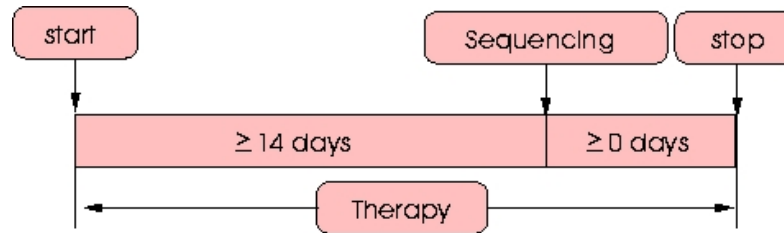
THEO

- 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

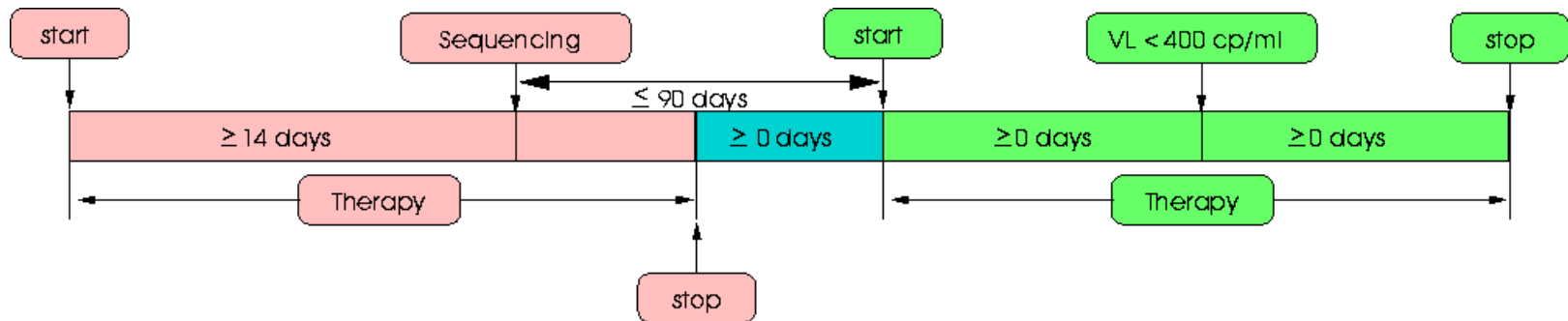


THEO cont.

- Definition of therapy failure and success
 - failure

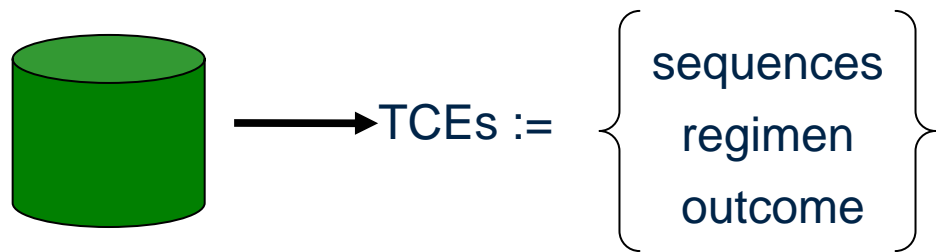


- success

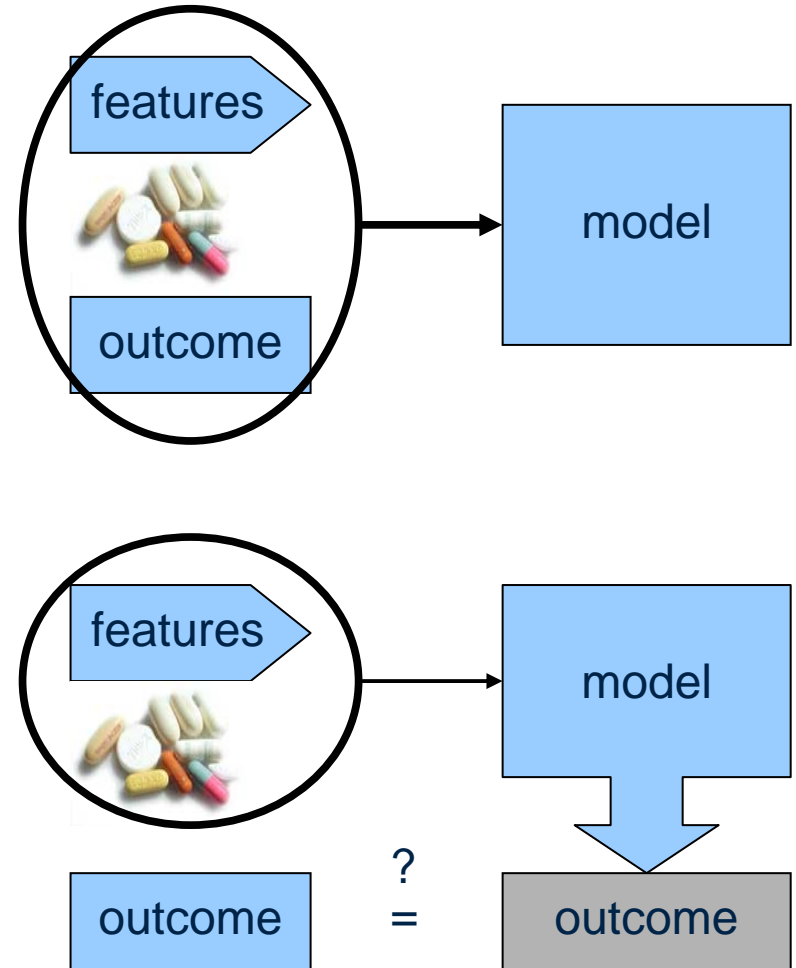


THEO cont.

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

No. of drugs <= - [] No. of pills per day <= - []

NRTIs: NNRTIs: Pls:

>= [] <= [] >= [] <= [] >= [] <= []

ZDV= [] NVP= [] IDV= []
 ddC= [] DLV= [] RTV= []
 ddI= [] EFV= [] SQV= []
 d4T= [] NFV= []
 3TC= [] ABC= [] LPV= []
 TDF= [] ATV= []

Reset

Compute

Selected drug combinations:

Success*	Regimen	Pills	Comment
0.86	d4T ABC NVP	5	d4T(2) ABC(2) NVP(1)
0.85	ddI ABC NVP	4	ddI(1) ABC(2) NVP(1)
0.85	ZDV ABC NVP	5	ZDV(2) ABC(2) NVP(1)
0.83	ddI d4T NVP	4	ddI(1) d4T(2) NVP(1)

*) predicted probability of virological success

Histogramm of all (all selected) therapies

Probability of virological success over 24+ weeks



THEO Applet cont.

• THERapy Optimization

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

The screenshot shows the THEO Applet interface with several red circles highlighting key settings: 'No. of drugs <= -', 'No. of pills per day <= -', 'NRTIs: <= 1', and 'ABC= exclude'. Below the settings is a table of 'Selected drug combinations' and a histogram titled 'Histogramm of all (all selected) therapies'.

Success*	Regimen	Pills	Comment
0.83	ddl d4T NVP	4	ddl(1) d4T(2) NVP(1)
0.83	ZDV ddl NVP	4	ZDV(2) ddl(1) NVP(1)
0.79	d4T TDF NVP	4	d4T(2) TDF(1) NVP(1)
0.78	ZDV TDF NVP	4	ZDV(2) TDF(1) NVP(1)

*) predicted probability of virological success

Histogramm of all (all selected) therapies

Probability of virological success over 24+ weeks



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



 **Arevir** www.geno2pheno.org

