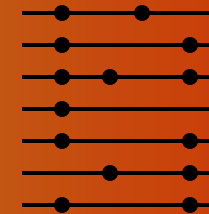
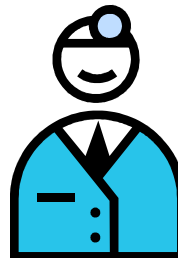
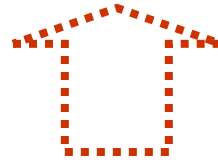
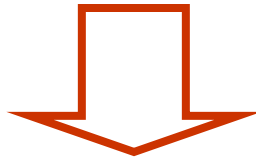
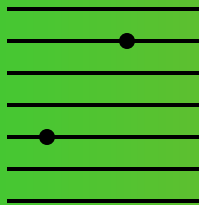
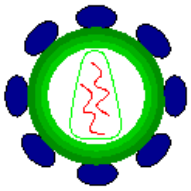


# **The evolutionary potential of HIV predicts response to antiretroviral therapy**

**Niko Beerenwinkel**

Department of Mathematics  
University of California at Berkeley

<http://math.berkeley.edu/~niko/>



Therapy failure



Drug resistance profile:

ZDV	D4T	DDI	3TC	ABC	...
EFV	DLV	NVP			
SQV	IDV	NFV	APV	LPV	

- **Predict mutations from treatment**
- Predict drug resistance from mutations
- **Predict response to new drug combinations**
- Find optimal combination

# Mutagenetic tree model

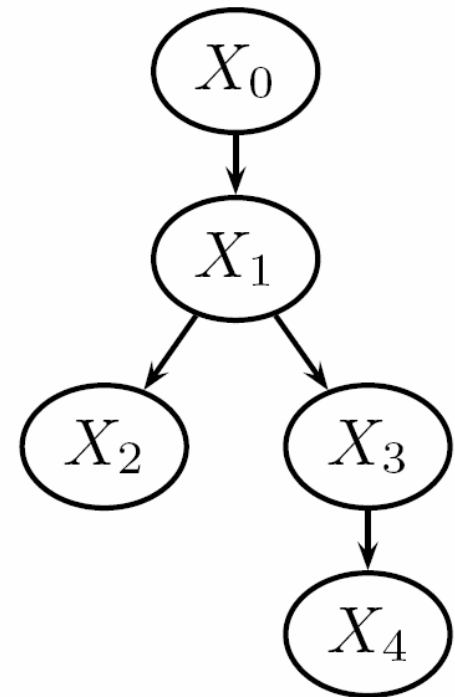
---

- Let  $X = (X_1, \dots, X_M)$  be binary random variables indicating the occurrence of resistance-associated mutations.
- A **mutagenetic tree** is a graphical model (Bayesian network) given by a tree and the *constraint* probability matrices

$$\vartheta^m = \begin{matrix} & 0 & 1 \\ 0 & \begin{pmatrix} 1 & 0 \\ 1 - \vartheta_{11}^m & \vartheta_{11}^m \end{pmatrix} \\ 1 & \end{matrix}$$

such that

$$\text{Prob}(X = x) = \prod_{m \in [M]} \vartheta_{x_{\text{pa}(m)}, x_m}^m$$

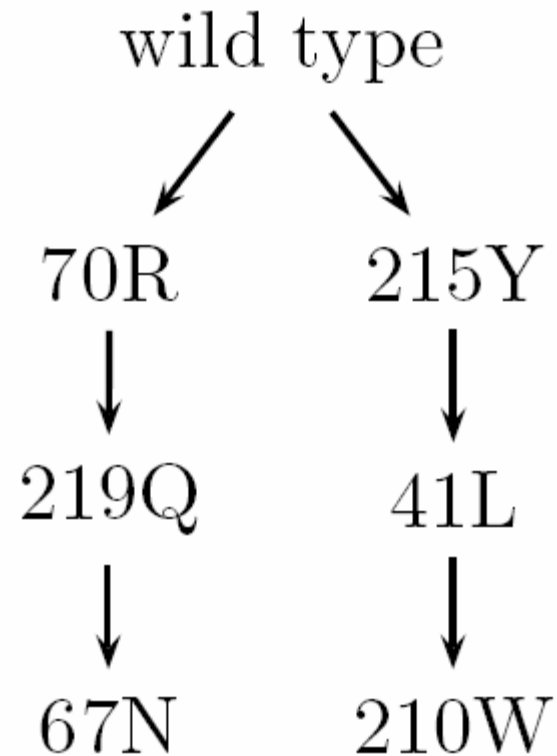


# Example: Accumulation of TAMs under ZDV

---

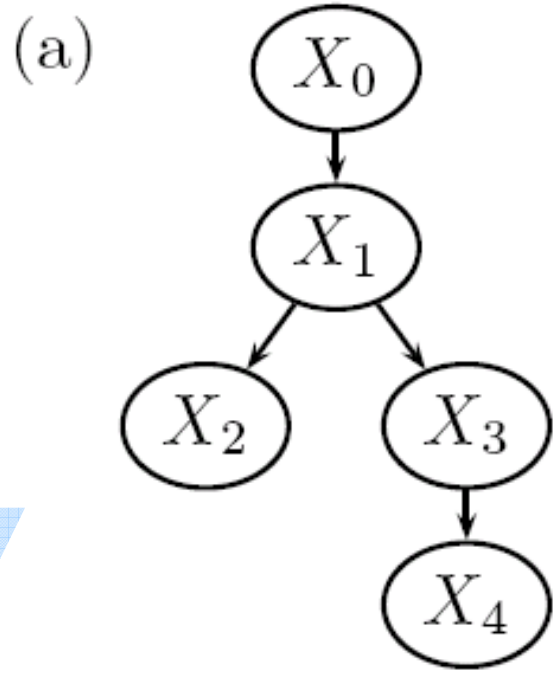
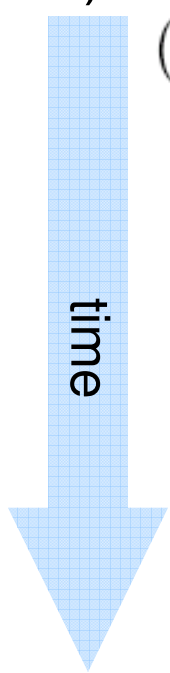
## ■ Mutagenetic tree models

- can be learned efficiently from cross-sectional data.
- can be interpreted (and estimated) as a process in time, i.e, as a **model of evolution**.

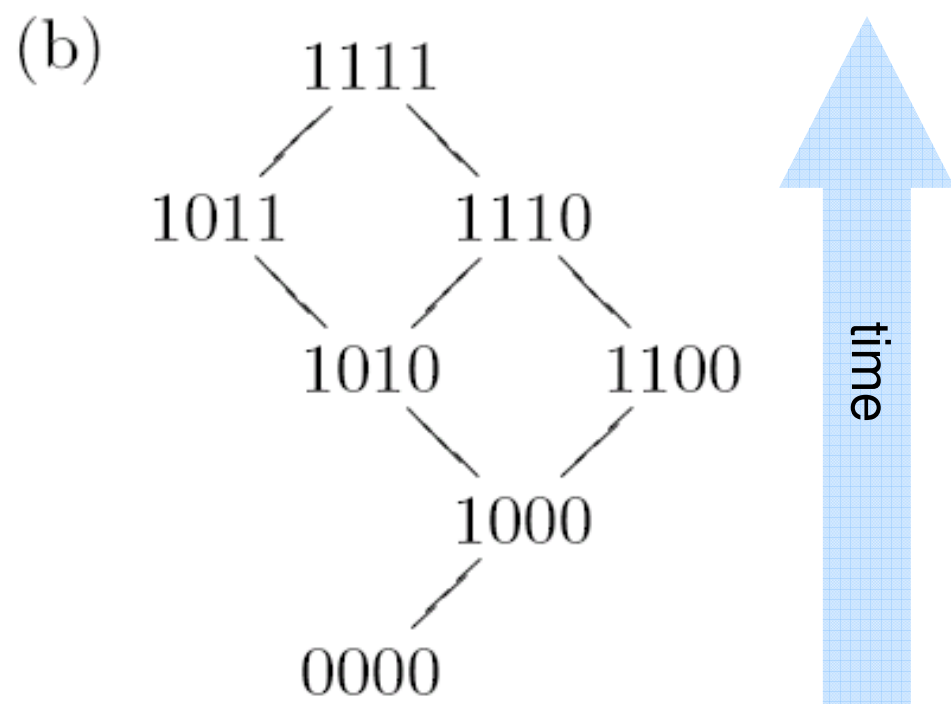


# Lattice of compatible states

■ )

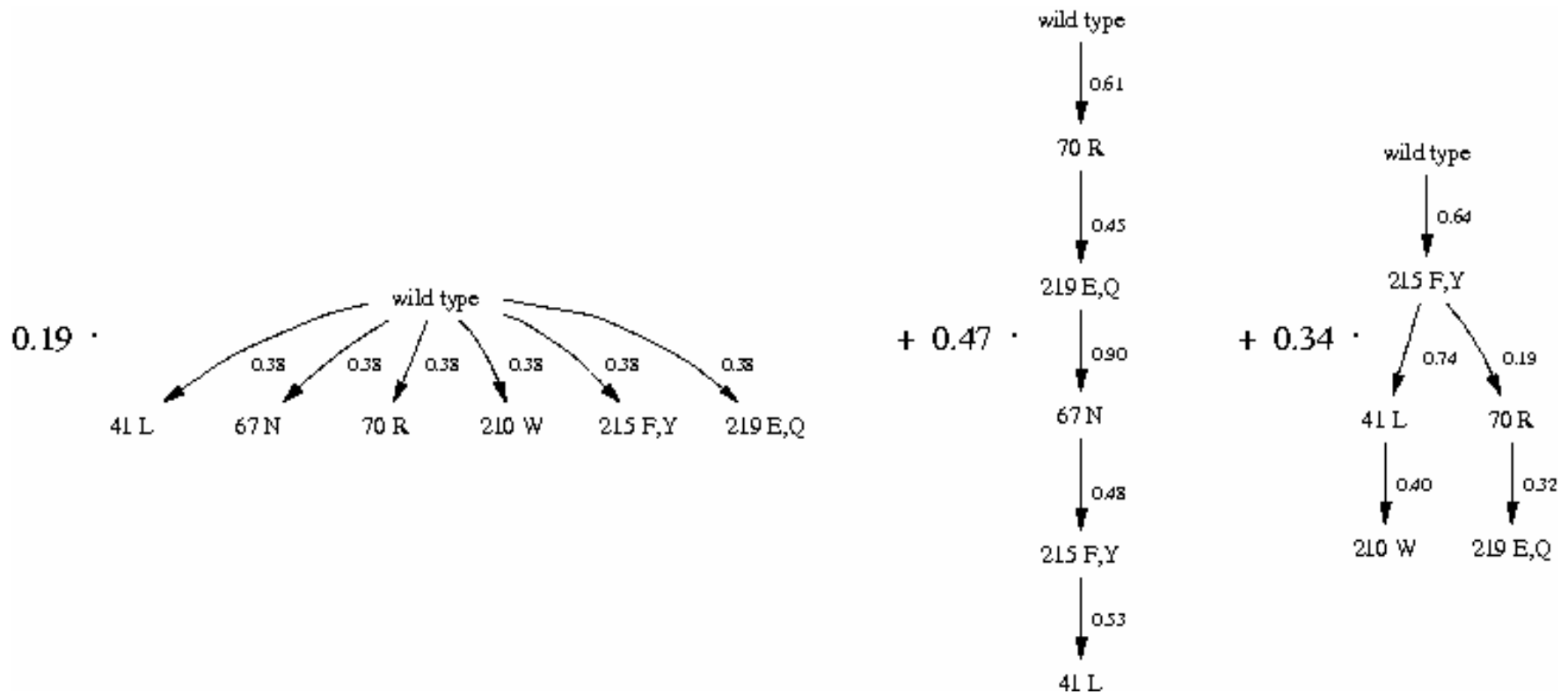


Mutagenetic tree



State space  
(mutational patterns)

# Mutagenetic trees mixture model



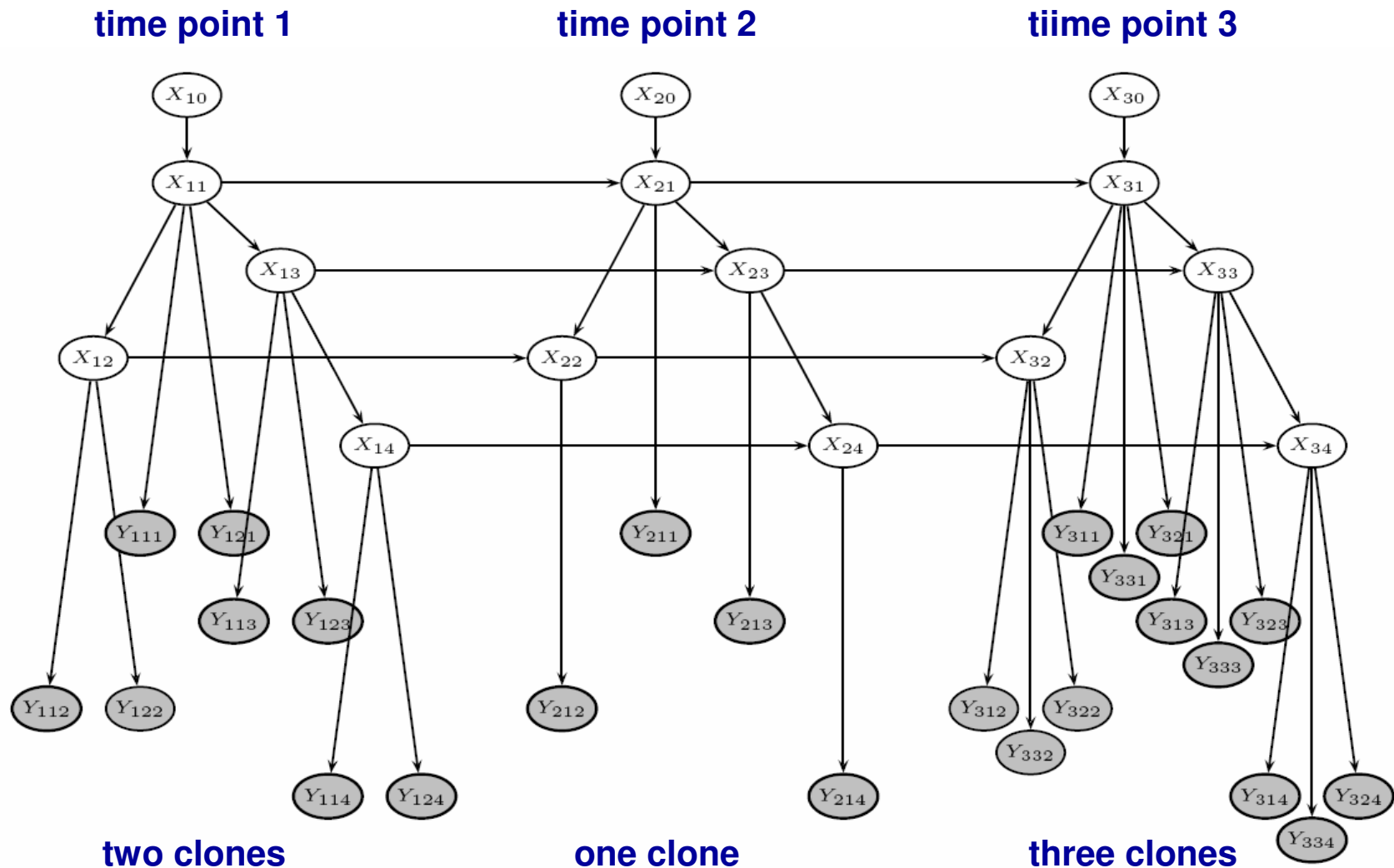
“noise component”

70-219  
pathway

215-41  
pathway

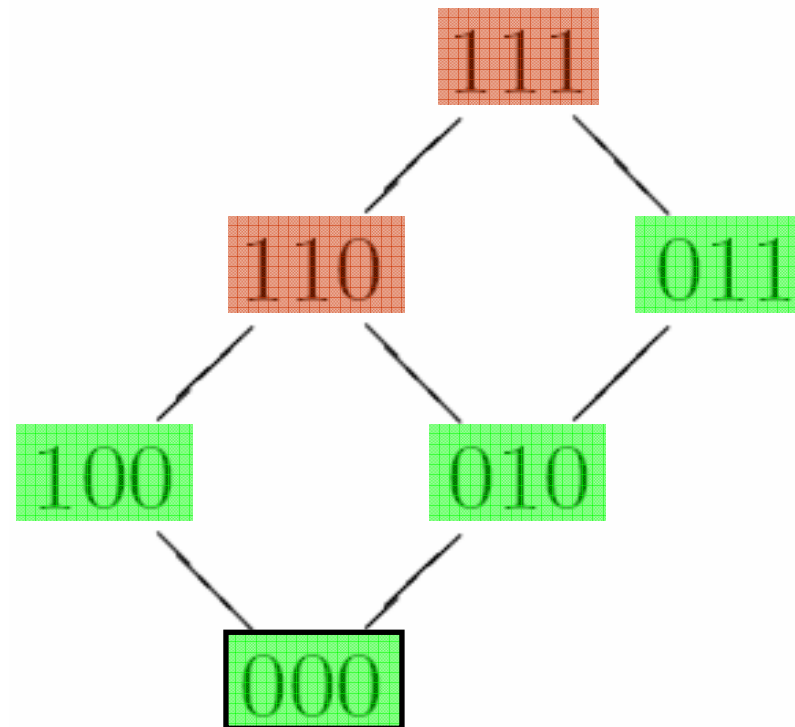
# Mutagenetic tree hidden Markov model

- Estimated from longitudinal clonal data.



# Genetic barrier

---



**resistant**  
**susceptible**

(estimated  
from *in vitro*  
phenotypes)

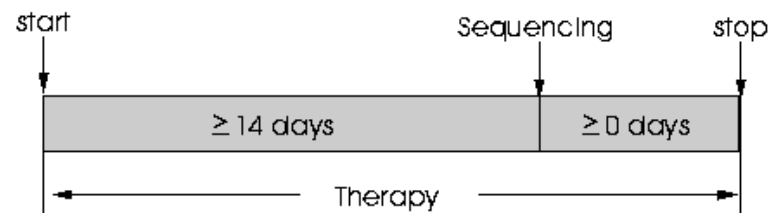
- Definition: The *genetic barrier* is the probability of not reaching any resistant escape state (= 1 – “risk of escape”).



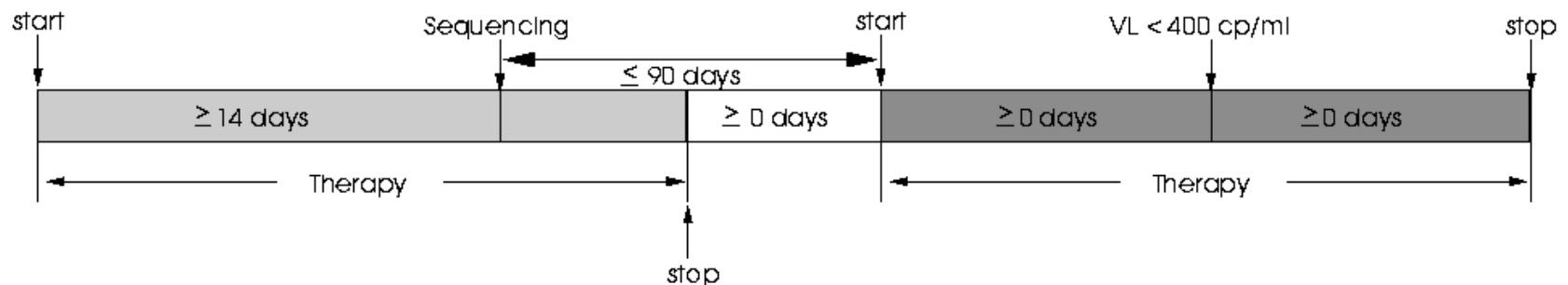
# Predicting therapy outcome

- 5,875 treatment change episodes (TCEs)
- 4,584 failures and 1,291 successes:

*failure*



*success*



- Compare different classifiers and feature encodings

# Features

---

- **Indicator**: All mutations
- **Phenotype\***: Predicted fold-change in susceptibility
- **Activity score\***: Exhaustive sequence space search, report “worst case” mutants
- **Genetic barrier\***
- **Genetic progression score (GPS)**: Expected waiting time of the mutational pattern in the timed mutagenetic tree

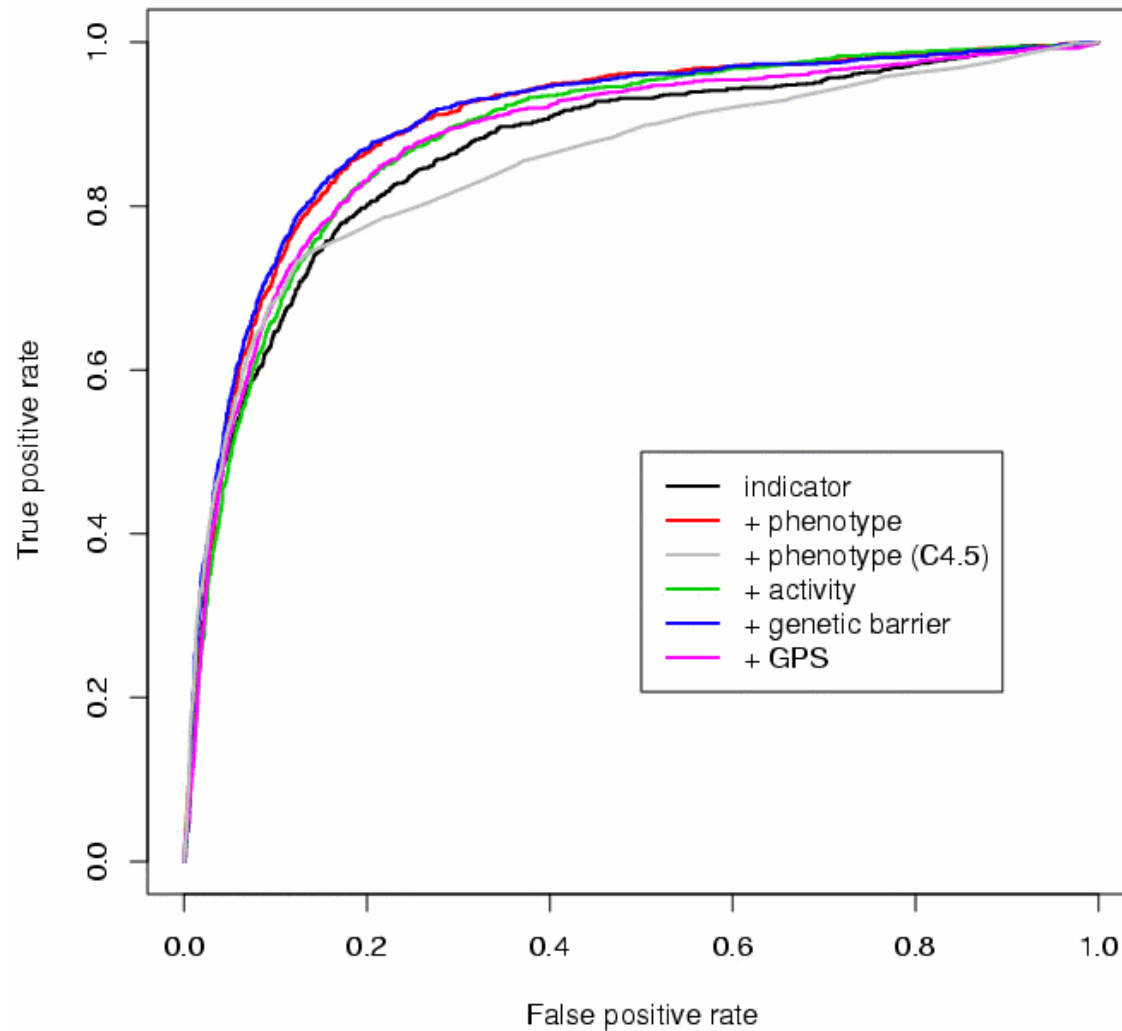
\* uses *in vitro* phenotypic data

# Performance: area under the ROC curve

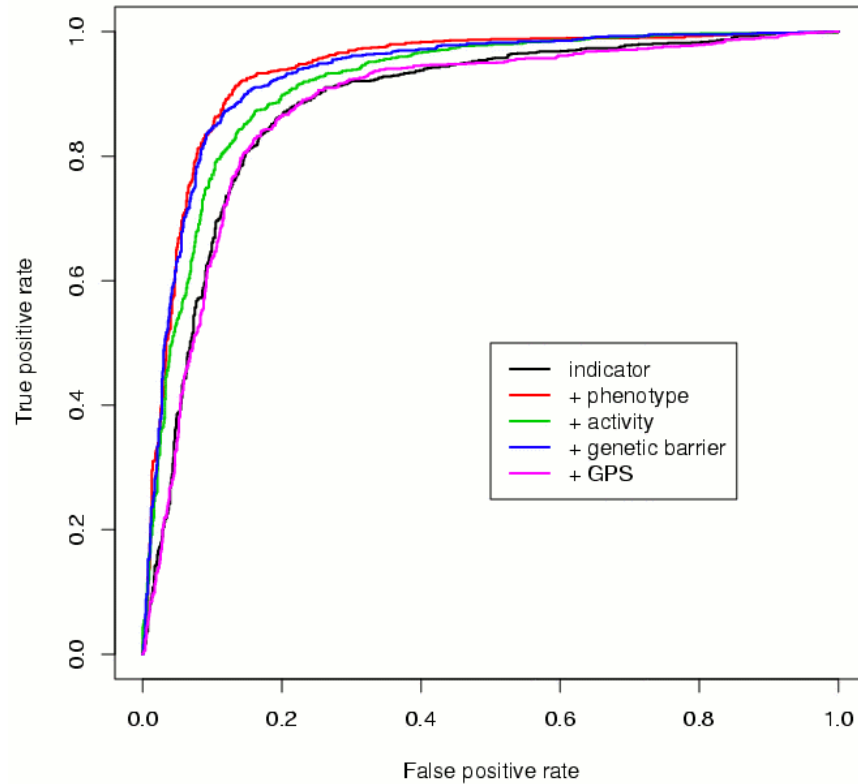
Learning method →	LDA	LSR	SVM	C4.5	LOGR	LMT	<i>average</i>
↓ Feature encoding							
<b>Indicator</b>	0.826 ±0.00 6	0.825 ± 0.007	0.809 ± 0.009	0.835 ± 0.007	0.823 ± 0.007	0.867 ± 0.007	0.831 ± 0.008
<b>+ phenotype*</b>	0.904 ± 0.004	0.903 ± 0.004	0.902 ± 0.004	0.847 ± 0.005	0.903 ± 0.004	0.902 ± 0.005	<b>0.894 ±</b> 0.009
<b>+ activity*</b>	0.865 ± 0.006	0.863 ± 0.006	0.857 ± 0.006	0.851 ± 0.008	0.865 ± 0.006	0.884 ± 0.008	0.864 ± 0.005
<b>+ genetic barrier*</b>	0.883 ± 0.003	0.884 ± 0.006	0.871 ± 0.009	0.857 ± 0.011	0.880 ± 0.005	0.901 ± 0.003	<b>0.879 ±</b> 0.006
<b>+ GPS</b>	0.853 ± 0.008	0.854 ± 0.006	0.855 ± 0.007	0.856 ± 0.006	0.862 ± 0.005	0.882 ± 0.006	0.860 ± 0.005
<b><i>average</i></b>	0.866 ± 0.013	0.866 ± 0.013	0.859 ± 0.015	0.849 ± 0.004	0.867 ± 0.013	0.887 ± 0.007	

# ROC curves for logistic model trees

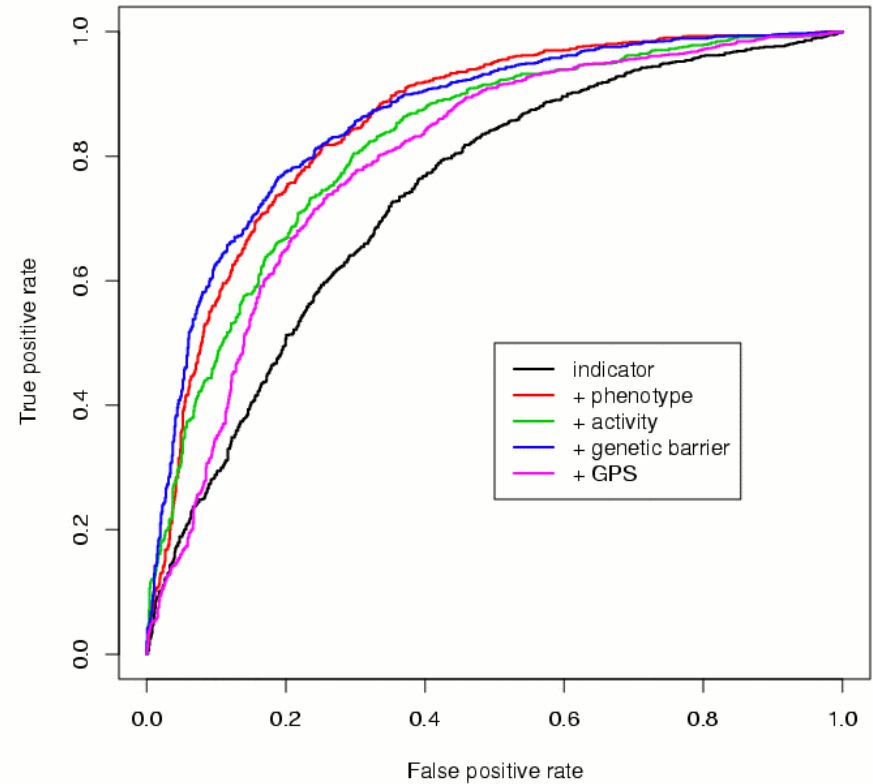
---



# ROC curves for logistic model trees

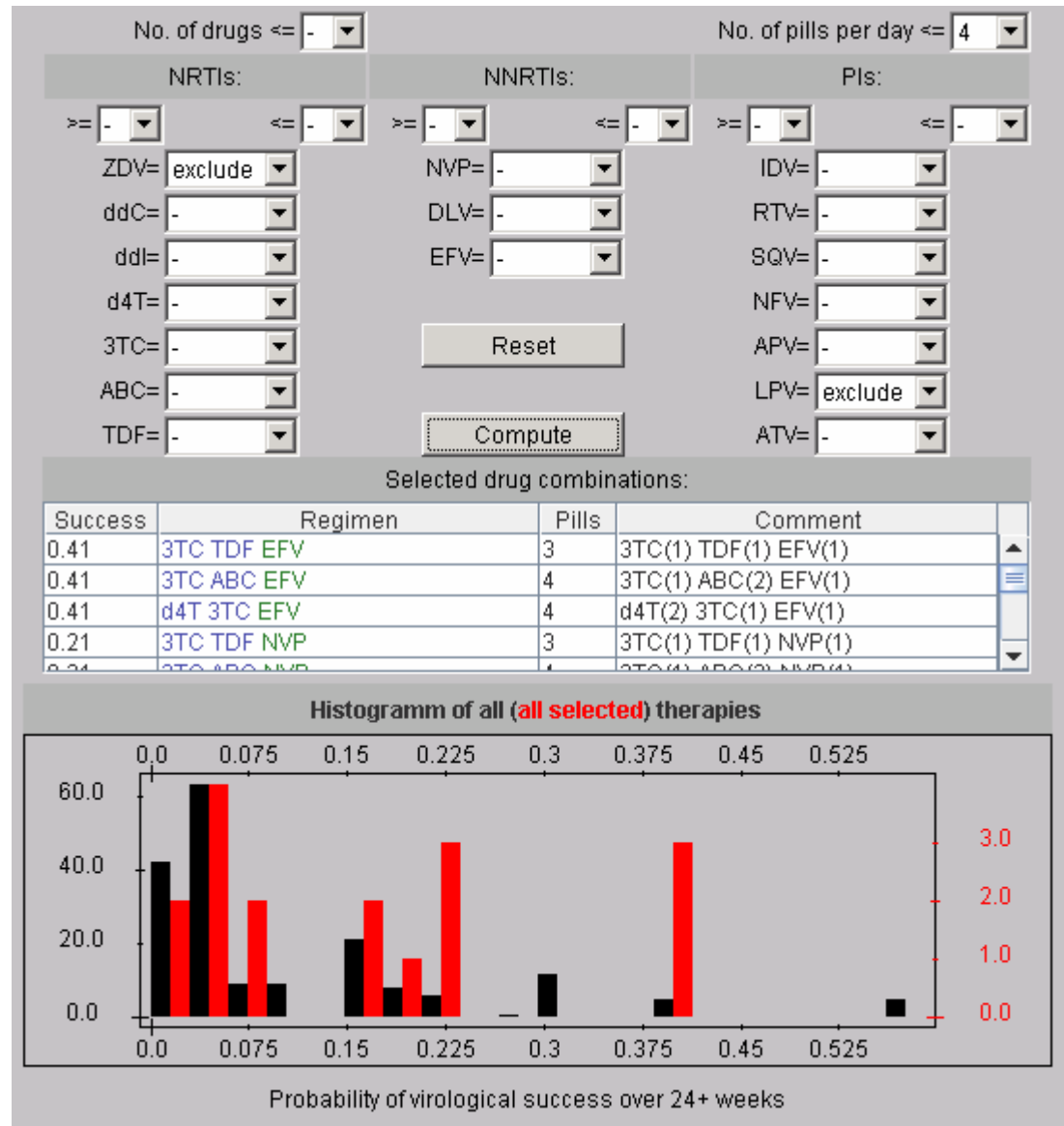


1,097 balanced sequences



1,022 balanced TCEs

## THEO – Therapy Optimizer



# Acknowledgements

---

- Andre Altmann, Tobias Sing  
Max Planck Institute for Informatics, Saarbruecken
- Mathias Drton  
The University of Chicago, Dept. of Statistics
- Nick Eriksson, Lior Pachter, Bernd Sturmfels  
UC Berkeley, Dept. of Mathematics
  
- DFG – Deutsche Forschungsgemeinschaft
- Forum for Collaborative HIV Research
  
- References: <http://math.berkeley.edu/~niko/>