max planck institut informatik

Do not blame the fitness!

Only slight impact of predicted replicative capacity for therapy response prediction

André Altmann

Department of Computational Biology and applied Algorithmics

Max Planck Institute for Informatics

D-66123 Saarbrücken Germany

7th European HIV Drug Resistance Workshop

25-27.03.09, Stockholm, Sweden

EUROPEAN HIV DRUG RESISTANCE WORKSHOP

max planck institut informatik

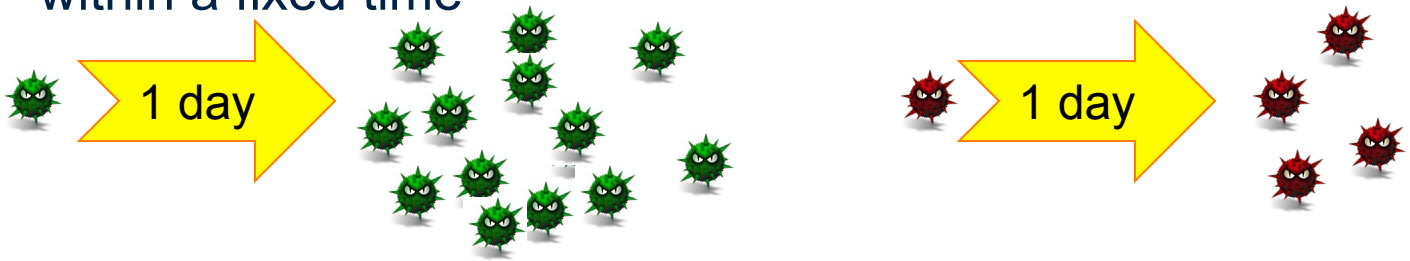
Outline

1. Prediction of RC from genotype (RT and Pro)
2. Validation of predicted RC (pRC)
3. Relevance of pRC for inferring response to ART
4. Conclusion

max planck institut informatik

Viral fitness

- Here: replication capacity w/o presence of drugs
  - Measure number of newly assembled infectious particles within a fixed time



  - Only Protease and RT were studied, not the complete virus
- Is viral fitness useful for predicting response to ART?

max planck institut informatik

- To study effect of RC during ART
  - First build a model that predicts RC from genotype on datasets comprising genotype-RC pairs

…KVIGSVLIGPTPANIIGRSLM… 1.9
…KVVGTVLVGPTPANIIGRNLM… 13.0
…KVIGSVLVGHTPSNIIGRNMM… 88.7
…KVVGTVLIGPTPVNVIGRNLM… 2.5

  - For this study two datasets were available

1. Monogram (Mark R Segal et *al.* Stat. Appl. Genet. Mol. Biol. 2004)
2. Erlangen (Hauke Walter et *al.*)
   317 and 253 samples

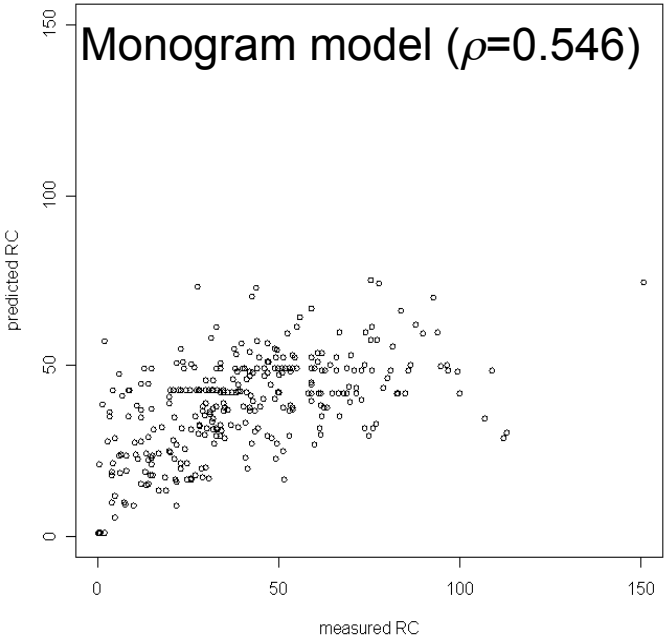
Prediction of RC from genotype

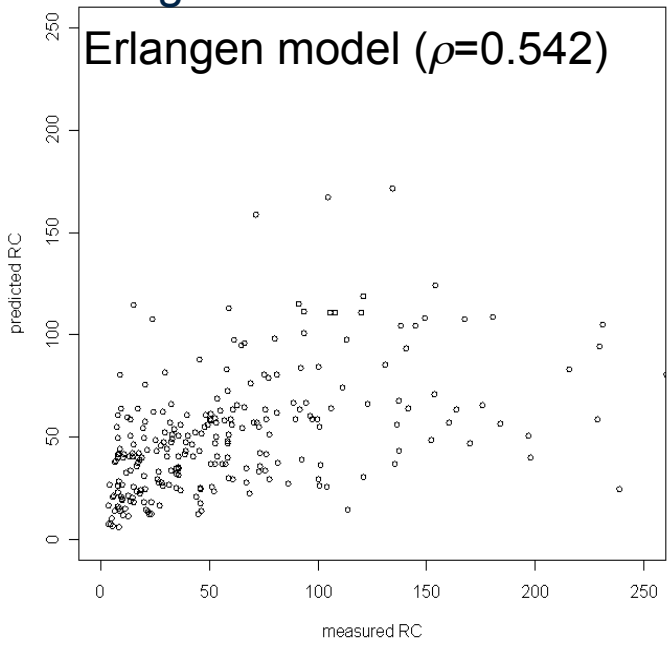
max planck institut informatik

Training of a support vector machine (SVM) for each dataset

1. Linear SVM for selecting important mutations
2. Polynomial SVM for modeling synergetic effects between mutations

- Estimation of model performance (spearman correlation)
  - Leave one out cross-validation on training data





max planck institut informatik

Important mutations in the SVM models

rank	Monogram mutation	influence	Erlangen rank
1	RT M184V		19
2	PR K43T		568
3	RT A158S		126
4	PR Q92R		401
5	PR I64L		886
6	PR K55R		602
7	PR E34K		483
8	PR I47V		366
9	PR V32I		131
10	PR P39S		141

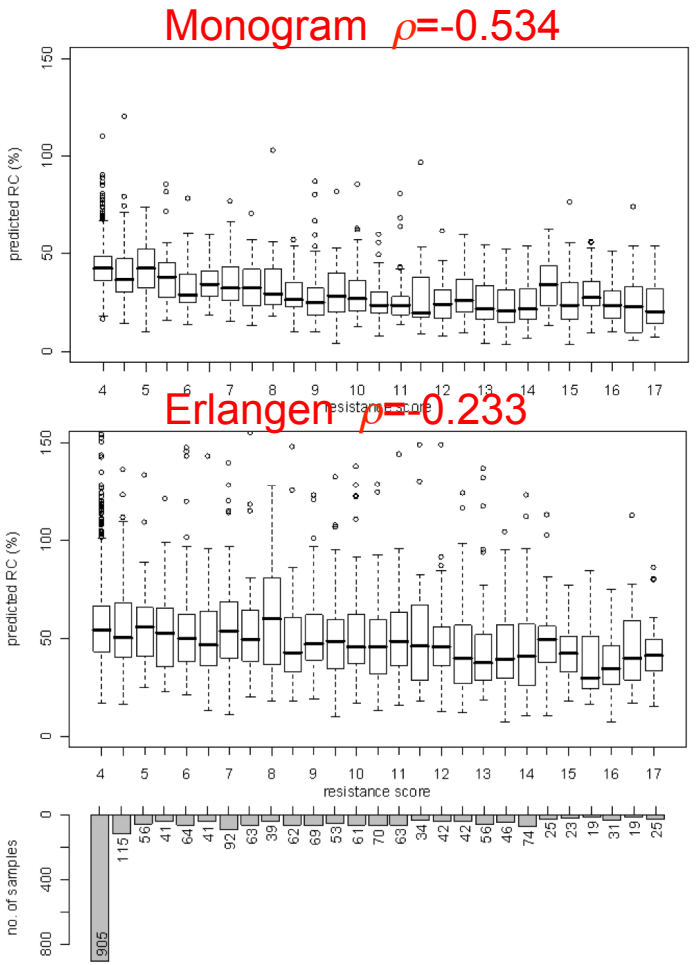
Erlangen mutation	influence	Monogram rank
RT Q207E		240
PR V82A		127
RT Y181C		150
RT T215Y		18
RT K20I		49
PR I13V		132
RT E122K		–
RT L74V		141
RT S162C		255
RT T39E		267

Validation of predicted RC

max planck institut informatik

Correlation of pRC with drug resistance

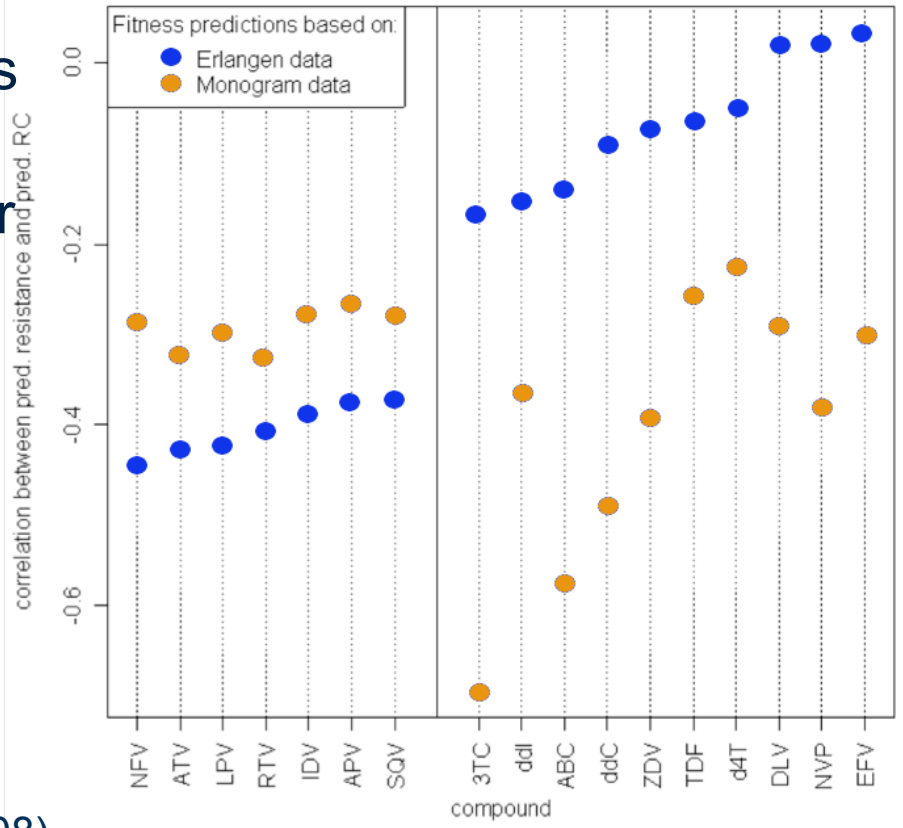
- 2,913 sequences (RT and Pro) from EuResist Integrated Database
- Resistance against 17 antiretroviral drugs was computed with geno2pheno
- Continuous value of predicted Fold Change (FC) was discretized using clinical cutoffs of geno2pheno
- Resistance against drugs was added: cumulative resistance score (CRS)



max planck institut informatik

Correlation of pRC with drug resistance

- Same dataset, but single drugs
- For the Erlangen model a clear separation of drug classes is visible
- Pro sequences of Erlangen data were highly mutated
  - 61% > 0 mutations
  - 25% > 4 mutations
  - For Monogram: 28% and 3%!
    (Robert W Shafer et al. AIDS Rev.2008)

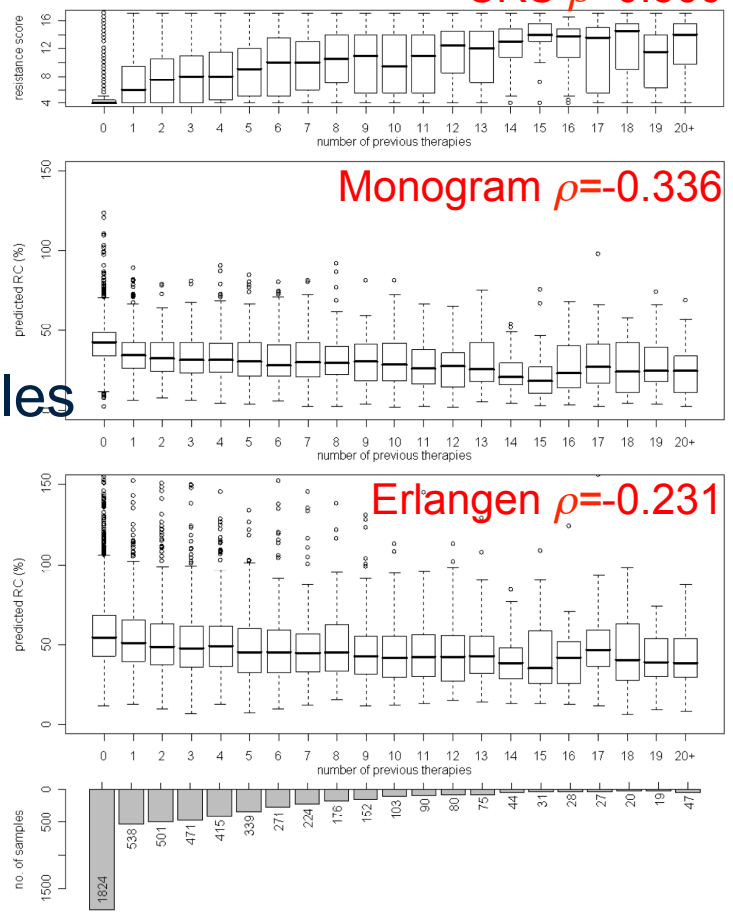


Validation of predicted RC

max planck institut informatik

Correlation of pRC with treatment experience

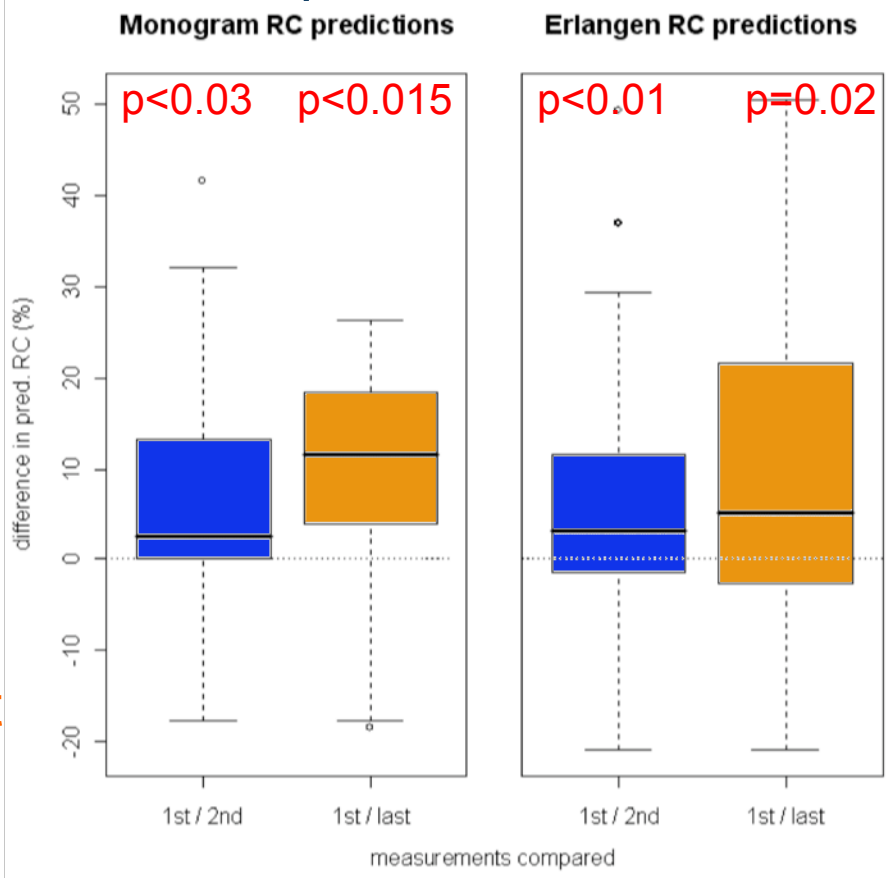
- 5,475 Pro and RT sequences from 3,869 patients extracted from the EuResist DB
- CRS and pRC computed for all samples
- Correlated to number of treatment changes before genotyping
- Patients with >19 treatment changes formed one group
- Naïve patients formed largest group



max planck institut informatik

Development of pRC during treatment interruptions

- 162 sequences from 57 patients undergoing a treatment interruption
- Sequences were obtained at end of treatment and at max. 4 time points during the break
- Difference in pRC between baseline and first (n=56) or last (n=30) measure during the break

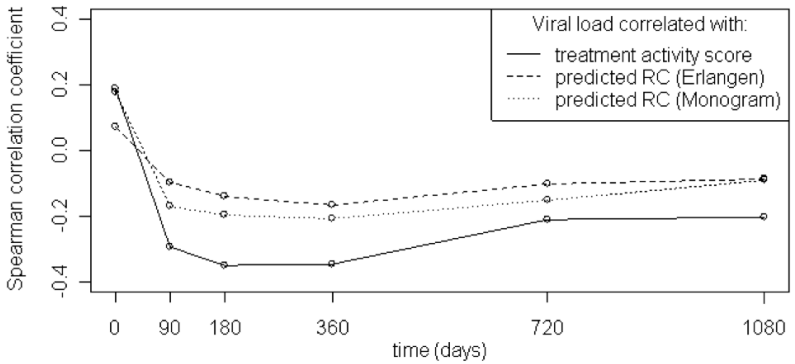


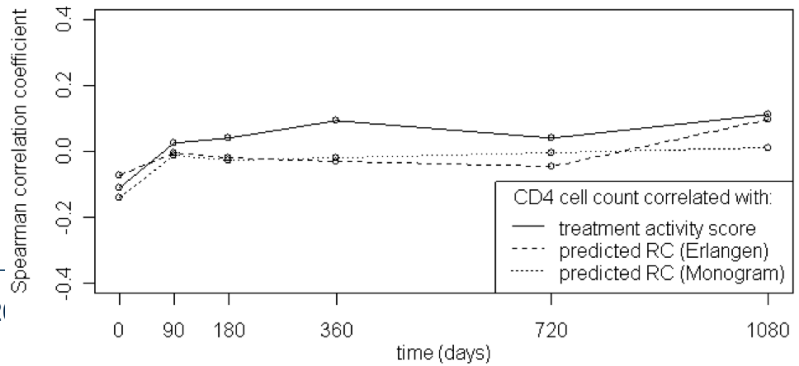
max planck institut informatik

Extraction of TCEs from the EuResist database

measure \ time	baseline	90 days follow-up	180 days follow-up	360 days follow-up	720 days follow-up	1080 days follow-up
viral load	2913	2031	2047	1457	675	333
$CD4^+$ T-cell count	2376	1621	1613	1154	526	252

- Baseline genotype, VL, and CD4 within 90 days before treatment start
- Follow-up measures at different time points
- Correlation of treatment activity score (TAS; PSS) and pRC with measurements at different time points



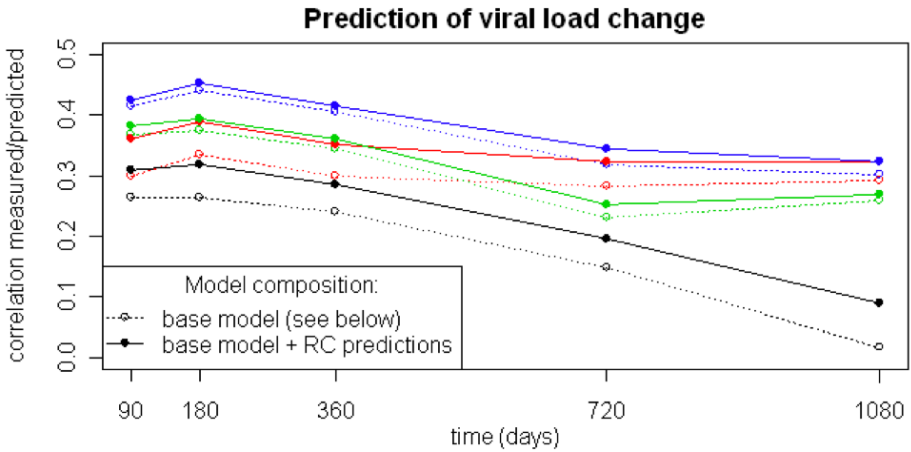


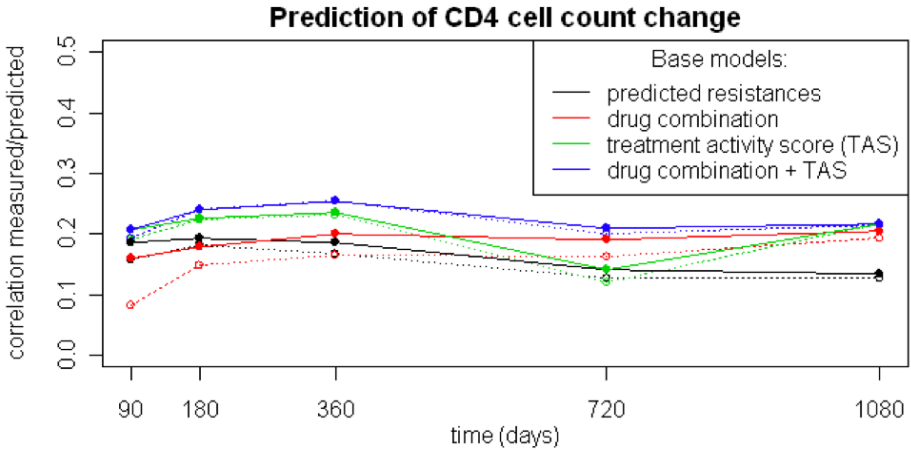
Relevance of pRC for inferring response to ART

max planck institut informatik

Predicting change in VL and CD4

- Linear regression model was trained using
  1. Predicted resistance to applied drugs
  2. TAS
  3. Drug combination
  4. Drug combination and TAS
- Models were built with and without pRC as covariate
- Performance was computed by 5x 5-fold cross-validation





Conclusions

max planck institut informatik

- Replication Capacity was predictable from genotype at moderate rates
  - Performance might be improvable with more more genotype-phenotype pairs
  - The pRC models selected different important mutations

- Pre-existing notions about RC were confirmed
  1. Inverse relation with drug resistance
  2. Relation to treatment experience
  3. Increase of RC during treatment interruptions

- Inclusion of pRC improved performance only moderately
  - No significant improvement over the best method

Resistance against applied antiretroviral drugs was dominant information for inferring response to ART

Do not blame the fitness!

Acknowledgements

EuResist
Integration of viral genomics with clinical data to predict response to anti-HIV treatment

Arevir

max planck institut informatik

Thomas Lengauer

Joachim Büch

Hendrik Weisser

Eugen Schülter

Melanie Balduin

Saleta Sierra Aragorn

Universität zu Köln

Rolf Kaiser

Klaus Korn

Hauke Walter

Monika Tschochner

KLINISCHE und MOLEKULARE VIROLOGIE
INSTITUT für
UNIVERSITÄT ERLANGEN-NÜRNBERG
Nationales Referenzzentrum für Retroviren

Anders Sönnerborg

Karolinska Institutet

Francesca Incardona

INFORMA CRO
LA RICERCA DEL FUTURO

Maurizio Zazzi

Thanks to Mark Segal for providing the Monogram dataset!