www.geno2pheno.org

MAX-PLANCK-GESELLSCHAFT

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

EuResist
Integration of viral genomics with clinical data to predict response to anti-HIV treatment
www.euresist.org

115

Do not blame the fitness! – Only slight impact of predicted replicative capacity for therapy response prediction

Andre Altmann[1], Hendrik Weisser[1], Francesca Incardona[2], Anders Sönnerborg[3], Maurizio Zazzi[4], Rolf Kaiser[5], Thomas Lengauer[1], Hauke Walter[6]

1 Max Planck Institute for Informatics, Saarbrücken, Germany; 2 Informa srl, Rome, Italy; 3 Division of Infectious Diseases, Department of Medicine, Karolinksa Institute, Stockholm, Sweden; 4 Department of Molecular Biology, University of Siena, Siena, Italy;
5 Institute of Virology, University of Cologne, Cologne, Germany; 6 Institute of Clinical and Molecular Virology, University of Erlangen, Germany

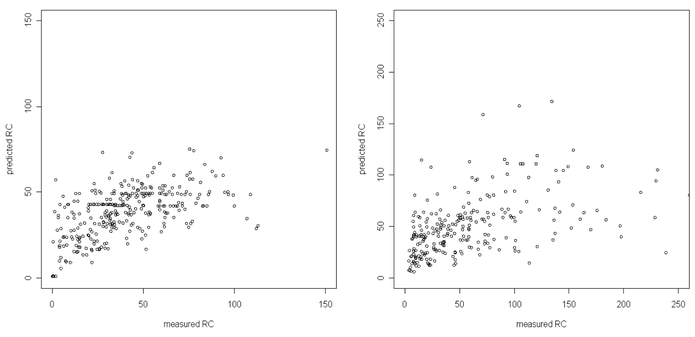
Background

Replication capacity (RC) of specific HIV isolates is occasionally blamed for unexpected treatment responses. However, the role of viral RC in response to antiretroviral therapy (ART) is not yet fully understood. We developed a method for predicting RC from genotype and studied the impact of predicted viral RC (pRC) on the change of viral load (VL) and CD4$^+$ T-cell count (CD4) during the course of therapy.

Methods & Results

Predicting replication capacity from genotype

Two data sets comprising genotype-RC pairs were used to train support vector machine (SVM) models. One SVM model using a polynomial kernel (degree 3) was trained for every data set. The model trained on the data set originating from *Erlangen* (253 genotype-RC pairs) achieved a Spearman correlation (ρ) of 0.542 (right scatter plot) in Leave-One-Out-Cross-Validation. The model trained on the *Monogram* data (n=317) [1] reached ρ=0.546 (left scatter plot).

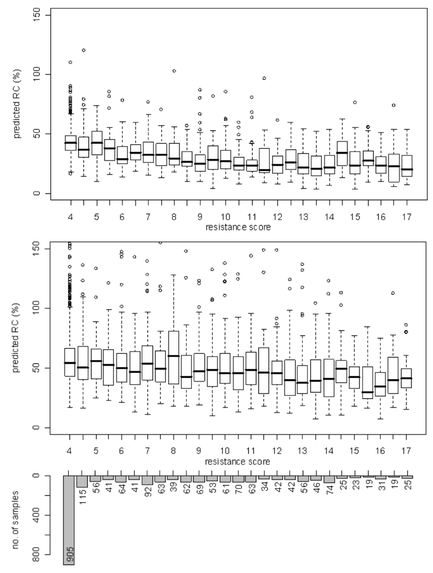


The contribution of mutations to the predicted replication capacity differed among data sets / SVM models (see table). However, protease sequences in the Erlangen data set were highly mutated (61% had one or more and 25% had 5 and more mutations of [2]) compared to protease sequences of the Monogram data set (28% had one or more and 3% had 5 and more mutations).

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

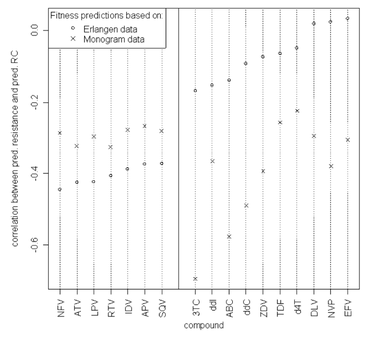
Erlangen mutation	influence	Monogram rank
RT Q207E	inc.	240
PR V82A	dec.	127
RT Y181C	inc.	150
RT T215Y	dec.	18
RT K20I	inc.	49
PR I13V	dec.	132
RT E122K	inc.	–
RT L74V	inc.	141
RT S162C	inc.	255
RT T39E	dec.	267

Correlation of pRC with resistance

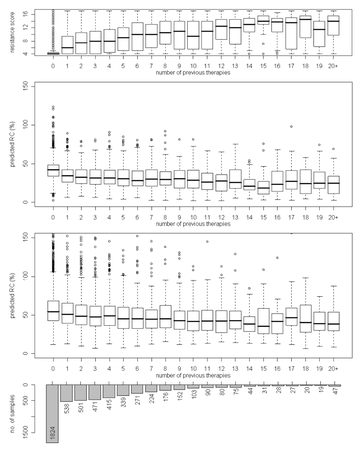


For 2,913 protease and reverse transcriptase (RT) sequences extracted from the EuResist database resistance against 17 antiretroviral drugs was computed with geno2pheno$_{[resistance]}$. The continuous values were discretized using the geno2pheno clinical cut-offs: Susceptible (0.0), Intermediate (0.5), Resistant (1.0). For every sequence resistance against all drugs was summed, resulting in a *cumulative resistance score (CRS)* that ranges between 0 and 17. The CRS was plotted against the pRC of the Monogram model (top; ρ=-0.534) and of the Erlangen model (bottom; ρ= -0.233).

The pRC of both models was also correlated to the resistance of single drugs (right figure). For the Erlangen model a clear separation of PIs, NRTIs, and NNRTIs was visible. PIs were most (inversely) correlated to pRC, whereas NNRTIs were not correlated at all.



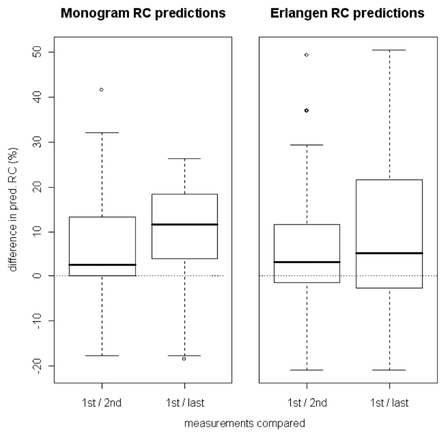
Relation of pRC and treatment experience



For 5,475 protease and RT sequences extracted from 3,869 patients of the EuResist database the CRS and pRC was computed and correlated to the number of treatments prior to the sequencing. Treatment naïve patients formed the largest group. CRS was clearly positively correlated to the number of treatments (upper box plot; ρ=0.560), and pRC computed with the Monogram model (middle box plot; ρ=-0.336) and with the Erlangen model (lower box plot; ρ=-0.231) was negatively correlated with treatment experience.

pRC during treatment interruptions

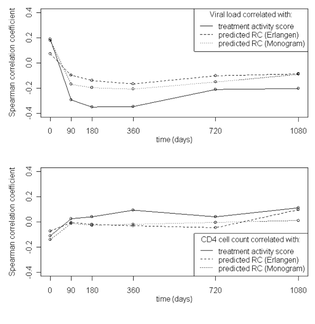
pRC was computed for 162 sequences of 57 patients undergoing a treatment interruption. One sequences was obtained at end of treatment and at up to four different time points during the break. The first measure during the break was about two months after end of treatment. The last measure during the break was at varying time points. The box plots on the right display the difference in predicted RC between the baseline measure and the first (n=56) and last (n=30) measure during the break, respectively.



Clinical relevance of pRC

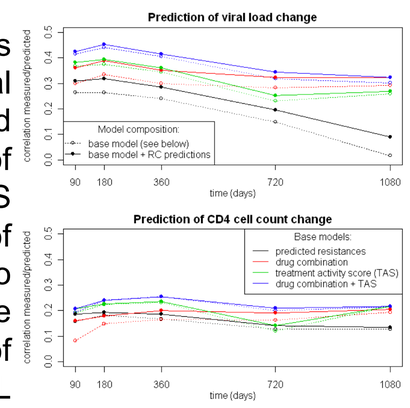
Treatment change episodes (TCEs) were extracted from the EuResist integrated database. Baseline measures were taken up two 90 days prior to treatment start. Follow-up measurements were taken at different time points.

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



The figure on the left shows the Spearman correlations between clinical markers (VL and CD4) and predicted RC or the *treatment activity score* (TAS) at baseline and at different time points during the treatment. The TAS is equivalent to a phenotypic susceptibility score computed with geno2pheno. Correlations with VL are in general better. In addition, TAS is better correlated to VL and CD4 than pRC.

The figure on the right shows the correlations between actual and predicted change of the clinical markers (VL and CD4). Linear regression was used for the predictions with predicted resistance of applied drugs (I), applied drug combination (II), TAS (III), or drug combination and TAS (IV). Each of these base models was used with or without the two pRC values from both models. Prediction of change in VL (CD4) was best at 180 (360) days. Inclusion of pRC improved the prediction of change in VL significantly for models I-III.



Conclusions

Viral RC, as measured by two different phenotypic tests, could be predicted from genotype with moderate accuracy. Pre-existing notions about RC were confirmed, e.g. increase of pRC during treatment interruption, relation of RC with treatment experience, expected direction of correlation of pRC with baseline measurements. Indeed, pRC could slightly improve prediction of virological treatment response. In general, pRC was significantly correlated with drug resistance. In summary pRC does not appear to provide substantial information over drug resistance, since the latter remains the dominant factor in predicting response to ART.

Acknowledgements

The work was supported by the Eu*Resist* project (IST -4- 027173-STP). We thank Mark Segal for making the Monogram RC dataset available to us. We also thank Melanie Balduin for providing the treatment interruptions dataset.

References

1. Mark R Segal, Jason D Barbour, and Robert M Grant. Relating HIV-1 sequence variation to replication capacity via trees and forests. *Stat. Appl. Genet. Mol. Biol.*, 3:Article2; discussion article 7, article 9, 2004.
2. Robert W Shafer and Jonathan M Schapiro. HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev.*, 10(2):67–84, 2008.