Genotypic Prediction of HIV Coreceptor Tropism: Application to Next-Generation Sequencing

Dr. J. Nikolaj Dybowski
PD Dr. Dominik Heider
Prof. Dr. Daniel Hoffmann

UNIVERSITÄT DUISBURG ESSEN
ZMB
ZENTRUM FÜR MEDIZINISCHE BIOTECHNOLOGIE

Background

:: Usage/Tropisms: R5, X4, and R5X4 determined by V3 loop

:: Potent R5-blockers (Maraviroc) available

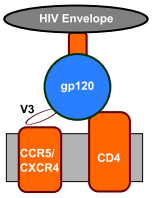
:: Ineffective against X4

:: Determine tropism prior to treatment

:: Computational alternatives to time-consuming, expensive cell-based assays

:: Prediction via charge rule [1] implicates importance of electrostatic potential

:: Preferential interaction of molecules with complementary electrostatic potential



Method

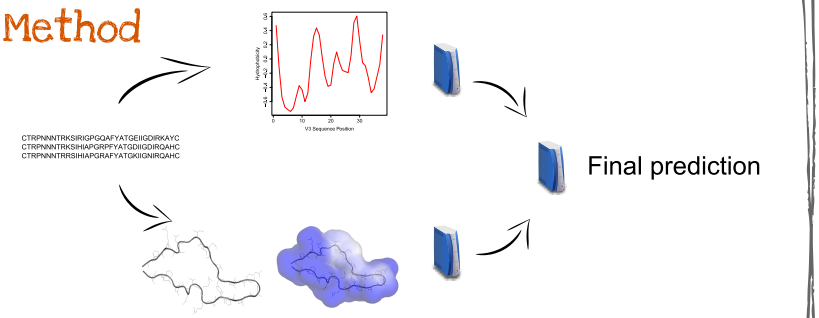


Fig. 1 **Schematic overview of the prediction method.**
Bottom: Construction of homology models based on crystal structure by Huang *et al.* [4], followed by calculation of electrostatic hull on which the first prediction is made using a Random Forest [5] machine learning model.
Top: Aminoacids of V3 sequences are translated to values modeling their hydrophobicity. Prediction again made by a Random Forest model.
At the second level a consensus prediction based on the independent first-level predictions is made to reach a final

Performance

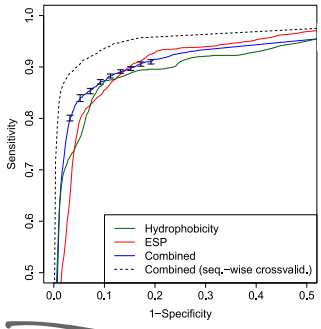


Fig. 2 **Method Performance.**
Performance measured by Receiver Operator Characteristic (ROC) curve. Displayed is the ability to correctly identify true positives (Sensitivity) against the false positive rate (1-Specificity).

Since some sequences originated from the same patient, crossvalidation was performed on a per-patient-level. The effect of this bias can be seen by the significant performance increase of the sequence-wise crossvalidation scheme.

Patient progression

:: Data from a study by Tsibris et al., 2009 [6]

:: Four intially R5-tropic patients (at week 0)

:: Treatment with R5 antagonist Vicriviroc (VVC)

:: For each patient and timepoint:

:: Sequencing of V3 region of quasispecies

:: Phenotypic testing (Monogram Trofile™)

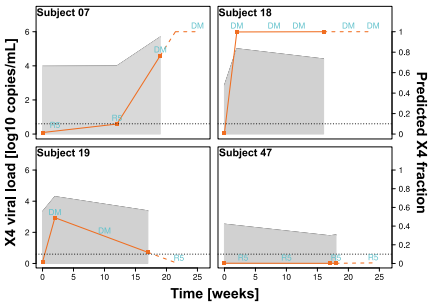


Fig. 3 **Predicting X4 load from quasipecies.**
Points represent relative amount of X4 using viruses in the quasispecies of patient. Labels show results of Monogram Trofile™ phenotypic tropism assay: R5 and DM, meaning R5-using and dual/mixed, respectively. Dotted, horizontal line represents Standard Trofile™ X4 detection cutoff.

Quasispecies evolution

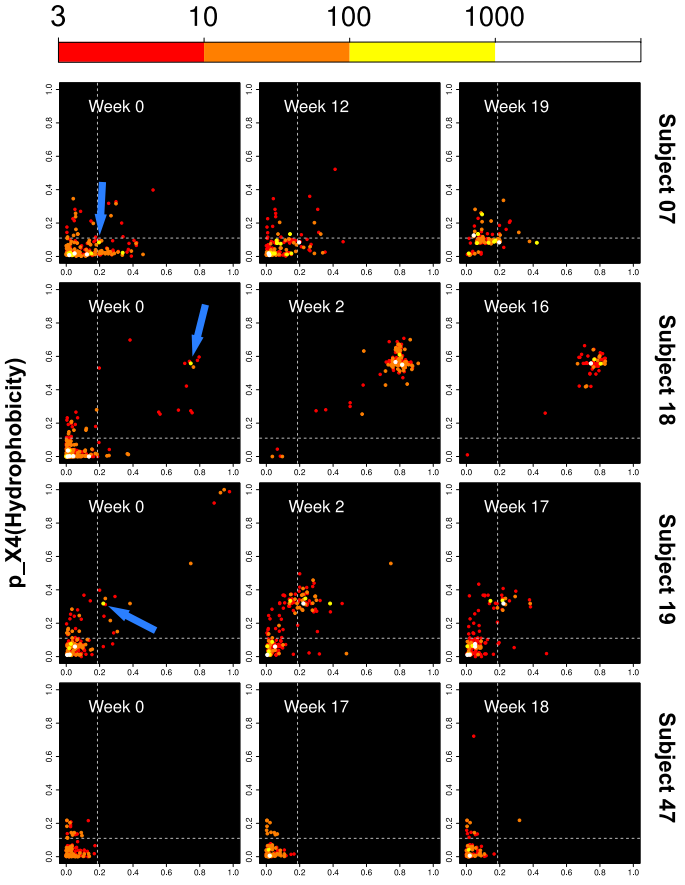


Fig. 4 **Emergence of X4 minorities under drug pressure.**
Virus population progression of patients 07, 18, and 19 (top to bottom). Axes show results of independent first-level predictions, based on electrostatic potential and hydrophobicity. Dashed lines represent cutoff levels between R5 and X4 classes. Cutoffs were chosen at 90% specificity in the training set. Green arrows indicate most prominent X4 minority variants at week 0. Development of patient tropism was driven almost excusively by these variants, each initially present at less than 1% of the quasispecies.

Conclusions

:: Predictions made on patient quasispecies reproduces phenotypic test results

:: Performs better at detection of minority variants

:: Minority variants (<1%) can be responsible for patient tropism switching

nikolaj.dybowski@uni-due.de

[1] Cardozo, T, Kimura, T, Philpott, S, Weiser, B, Burger, H, & Zolla-Pazner, S. (2007) AIDS Res. Hum. Retroviruses 23, 415–26.
[2] Kyte, J & Doolittle, R. (1982) J. Mol. Biol. 157, 105–132.
[3] Dybowski, J. N., Heider, D., and Hoffmann, D. (2010). PLoS Computational Biology, 6(4), e1000743.
[4] Huang, C. C, Tang, M, Zhang, M. Y, Majeed, S, Montabana, E, Stanfield, R. L, Dimitrov, D. S, Korber, B, Sodroski, J, Wilson, I. A, Wyatt, R, & Kwong, P. D. (2005) Science 310, 1025–8.
[5] Breiman, L. (2001) Machine Learning 45, 5–32.
[6] Tsibris, A & Korber, B & Arnaout, R & Russ, C & Lo, C, et al. (2009) PLoS ONE 4, e5683.