a new diagnostic tool for the genotypic determination of HIV-2 coreceptor usage

M. Döring¹, P. Borrego², J. Büch¹, A. Martins², G. Friedrich¹, R. J. Camacho³, J. Eberle⁴, R. Kaiser⁵, T. Lengauer¹, N. Taveira²,⁶, N. Pfeifer¹

¹ Department for Computational Biology & Applied Algorithmics, Max Planck Institute for Informatics, Saarbrücken, Germany.
² Research Institute for Medicines (imed.ULisboa), Faculty of Pharmacy, University of Lisbon, Lisbon, Portugal.
³ Department of Microbiology & Immunology, Rega Institute for Medical Research, KU Leuven, Belgium.
⁴ Department of Virology, Max von Pettenkofer-Institut, Ludwig-Maximilians-University, Munich, Germany.
⁵ Institute for Virology, University of Cologne, Cologne, Germany.
⁶ Instituto Superior de Ciências da Saúde Egas Moniz (ISCSEM), Monte de Caparica, Portugal.

Relevance of HIV-2 coreceptor usage

- The selection of HIV-2 variants using the CXCR4 coreceptor (X4-capable) should be prevented because X4-capable variants are harder to neutralize than viruses using only CCR5 (R5) [1].
- Before prescribing CCR5-coreceptor antagonists to patients infected with HIV-2, clinicians should rule out the existence of X4-capable variants.
- Goal: differentiate R5 and X4-capable HIV-2 variants based on the amino acid sequence of the V3 loop.

Results

- A linear SVM (AUC=0.95) outperformed other models and was used in all subsequent analyses.
- For a set of 126 V3 sequences, the 10-fold nested CV sensitivity was 76.9% and the specificity was 97.3%.
- All samples from a set of nine, newly phenotyped V3 sequences were classified correctly by the SVM.
- We validated existing markers for X4-capability [2] and identified new, significant features ($p \leq 0.05$): variants 27K, 15G, and 85.

Materials and methods

Support vector machines (SVMs) were trained on a data set of 73 R5 and 52 X4-capable samples to classify binary-encoded V3 amino acid sequences as either R5 or X4-capable. Classifier performance was evaluated using 10-fold nested cross validation (CV). The predicted probabilities indicating whether a sequence originates from an X4-capable variant were transformed to false positive rates (FPRs).

We developed a visual representation of position-specific classifier weights to indicate amino acids associated with R5 and X4-capable variants (see Fig. 2). We evaluated established discriminatory sequence features from a rules-based approach by Visseaux et al. [2] and novel features detected by the SVM using Fisher’s exact test with multiple testing correction (Benjamini and Hochberg).

Highlights of the tool

- Accuracy: high sensitivity and specificity
- Interpretability: visualization of sequence-specific weights and output of FPRs
- Availability: an online web service is available at coreceptor-hiv2.geno2pheno.org
- Opportunities: enables large-scale epidemiological studies on HIV-2 coreceptor usage

Visualization of model weights

Figure 2: SVM weights for the V3 loop of a ROD10 isolate.

References