Are non-B subtypes less susceptible to antiretroviral drugs?
– a bioinformatical approach to prediction of non-B subtype susceptibility.

J Kjær1, L Høj1, A Cozzi-Lepri2, Z Fox3, JD Lundgren1,3

1Copenhagen HIV Programme, University of Copenhagen, Faculty of Health Sciences
2Panum Institute/Building 21.1
3Centre for Viral Diseases, Righospitalet, Copenhagen, Denmark

BACKGROUND

The first generation of antiretrovirals that have been developed and used to treat primarily subtype B infected patients are finally becoming available in African countries. The subtype B is however not the dominant subtype in Africa and question remains if these subtypes are less susceptible to these drugs. We will here try to answer this question by use of artificial neural networks (ANNs).

METHOD

We have previously showed that, based on the physicochemical properties of the amino acids, ANNs are able to extrapolate predictions to non-B subtypes with high accuracy (see box below) despite being trained only with subtype B sequence data and a matching IC50 fold change (FC). The ANNs use the predicted logIC50 FC susceptibility for protease or reverse transcriptase for a given drug based on the difference between physicochemical properties of the amino acids in HXB2 sequence (logIC50 FC=0) and any protease or reverse transcriptase presented to the ANNs (Figure 3). We applied the use of these ANNs to further predict the logIC50 FC susceptibility for the most dominant subtypes in Africa: A, C, G and CRF_AG to the drugs: abacavir, amprenavir, atazanavir, didanosine, efavirenz, indinavir, lamivudine, lopinavir, nevirapine, ritonavir, saquinavir, stavudine, tenofovir, zalcitabine, zidovudine. We obtained reference sequences for these subtypes from the Los Alamos HIV sequence database. These subtypes were used for input to the ANNs to predict the logIC50 FC for each of the above drugs. We used an unpaired t-test to identify significant differences in mean of the logIC50 FC values between B subtype and non-B subtypes.

RESULTS

From Los Alamos we extracted 6 sequences for subtype A, 3 sequences for subtype B and 3 sequences for each of the subtypes C, G and CRF_AG resulting in 96 logIC50 FC predictions for subtype B, 16 for subtype D and 64 for each of the subtypes C, G and CRF_AG.

The predicted mean IC50 FC across the 16 drugs for each subtype was:

<table>
<thead>
<tr>
<th>Subtype</th>
<th>mean logIC50 FC</th>
<th>mean IC50 FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.16</td>
<td>1.18</td>
</tr>
<tr>
<td>B</td>
<td>0.13</td>
<td>1.26</td>
</tr>
<tr>
<td>C</td>
<td>-0.30</td>
<td>0.74</td>
</tr>
<tr>
<td>G</td>
<td>0.27</td>
<td>1.57</td>
</tr>
<tr>
<td>CRF_AG</td>
<td>0.14</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Plots of the mean predictions for the individual drugs and for each subtype is shown in the Figure 1. Please note that these values have been transformed from the originally predicted logIC50 FC and each data point represents the mean per drug and subtype.

We found no significant difference in logIC50 FC for subtypes C or CRF_AG compared to the subtype B logIC50 FC mean. But there was a significant [p<0.05] difference between B and A (p=0.03) and B and G (p=0.02) predictions.

Detailed logIC50 FC predictions are listed in Table 1. Drugs with a predicted IC50 FC > 2 fold are:

- Abacavir
- Didanosine
- Lamivudine
- Stavudine
- Zalcitabine
- Lamivudine
- Stavudine
- Zalcitabine
- Nevirapine
- Amprenavir
- Atazanavir
- Indinavir
- Lopinavir
- Nevirapine
- Saquinavir

Excluding these drugs from the analysis resulted in only subtype C (mean IC50 FC=0.20) to be significantly different in susceptibility compared to subtype B (p=0.04).

LIMITATIONS

The extracted reference sequences from the Los Alamos data base constitutes a small and selected dataset. A few unusual polymorphisms in some of these sequences can easily have become a bias in the results presented here. Of note the logIC50 FC values for subtype B which were expected closer to 0 for all drugs and for the subtypes C, G and CRF_AG hyper-susceptibility to the NNRTIs are being predicted. But instead of using HXB2 (IC50 FC=- for all drugs given the design of the ANNs) for comparison we decided to use the subtype B reference sequences from Los Alamos to allow for the same level of bias to the subtype B predictions as we would have to accept in regards to subtype A, C, G and CRF_AG. Only a larger dataset including larger subtype variation would allow to adjust for this and to investigate specific polymorphisms that cause significant change in IC50 FC values.

CONCLUSION

Predictions showed significant difference between the predicted mean IC50 FC values for the both A and G HIV-1 subtypes compared to IC50 FC for subtype B. This is in part driven by reduced susceptibility to amprenavir, atazanavir, nevirapin and zidovudine and does also affect susceptibility for C and CRF_AG subtypes. To the majority of the drugs we did not find any reduced susceptibility to the drugs for the subtypes and C subtypes may even have an increased susceptibility to these drugs.


"If sequence data was not available for non-B subtypes therefore we could not validate the use of ANNs for either NNRTIs or NNRTIs. For PR we found 3 subtype C sequences for atazanavir, 15 subtype C and one subtype B sequences for amprenavir, indinavir, nevirapine, ritonavir and saquinavir; and 16 subtype C and one subtype B sequence for laminidine. This allowed for a total of 200 IC50 FC values to be predicted for primarily C subtypes by ANNs trained with unique subtype sequences. The IC50 FC values were mainly in the susceptible range with a mean IC50 FC value of 1.0±0.9 (range 0.2–5.8) and with 76% observed and 75% predicted above IC50 FC=1.5 fold. The overall correlation coefficient between the observed and the predicted IC50 FC values was found to be 0.89 (p<0.05 for nevirapine, 0.85 for indinavir and 0.56 for atazanavir, ritonavir and saquinavir, respectively). The overall PPV for all 200 predictions was 0.86."

Download poster at: www.cphiv.dk