CoRSeq\textsuperscript{23}: Novel Coreceptor Usage Prediction Algorithms for HIV-1 Subtypes B, C, D and CRF01\_AE
Kieran Cashin\textsuperscript{1,2}, Lachlan R. Gray\textsuperscript{1,2}, Katherine Harvey\textsuperscript{1,2}, Danielle Perez-Bercoff\textsuperscript{1}, Guinevere Q. Lee\textsuperscript{5}, Jasmina Sterjovski\textsuperscript{1}, Michael Roche\textsuperscript{1}, P. Richard Harrigan\textsuperscript{5}, Melissa J. Churchill\textsuperscript{1,6}, Paul R. Gorry\textsuperscript{1,2}

\textsuperscript{1}Center for Biomedical Research, Burnet Institute, Melbourne, Australia, \textsuperscript{2}Department of Microbiology and Immunology, University of Melbourne, Melbourne, Australia, \textsuperscript{3}Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia, \textsuperscript{4}Center Recherche Public de la Santé, Luxembourg, Luxembourg, \textsuperscript{5}BC Center for Excellence in HIV/AIDS, University of British Columbia, Vancouver, BC, Canada, \textsuperscript{6}Department of Microbiology and Immunology, Monash University, Melbourne, Australia

**Background:** Maraviroc (MVC) is a CCR5 antagonist used in HIV-1 antiretroviral therapy that can only be prescribed to patients with exclusively CCR5-using virus populations. Thus reliable determination of HIV-1 coreceptor usage in patients is clinically important for prescription of MVC. Genotypic coreceptor usage prediction algorithms, based on HIV-1 V3 sequence, have been developed as inexpensive and rapid alternatives to traditional phenotypic assays that, due to their cost and lengthy turn-around time, have limited prescription of MVC. Most algorithms are designed against and predict coreceptor usage of HIV-1 subtype B (B-HIV), but fail to reliably predict coreceptor usage of other HIV-1 subtypes, which constitute >85% of infections worldwide. We have developed a suite of V3 sequence-based coreceptor usage prediction algorithms (CoRSeq\textsubscript{3}) that are highly sensitive and specific for determining coreceptor usage of HIV-1 subtypes B, C (C-HIV), D (D-HIV) and CRF01\_AE (AE-HIV).

**Methodology:** CoRSeq\textsubscript{3} B-HIV, C-HIV, D-HIV and AE-HIV algorithms were designed and tested using every respective phenotypically characterised patient-derived V3 sequence in the Los Alamos Database, and from our recently published studies. Unique sequences were randomly assigned to “design” and “test” sets for each subtype and analysed for phenotype specific characteristics.

**Results:** Analysis of “design” sets revealed differing V3 characteristics for CCR5- and CXCR4-using viruses, including charge, length, glycosylation and amino acid mutations, which informed CoRSeq\textsubscript{3} prediction criteria. CoRSeq\textsubscript{3} predictive accuracy was then tested against independent “test” sequences. We found that CoRSeq\textsubscript{3} algorithms have more favorable sensitivity and specificity profiles for predicting HIV-1 coreceptor usage than any alternative, including geno2pheno (Table).

**Conclusions:** Once clinically validated in future studies our highly sensitive and specific B-HIV, C-HIV, D-HIV and AE-HIV algorithms may enhance access to MVC and future coreceptor antagonists for countries where non-B HIV-1 strains predominate, and which have expanding economies and improving health care systems, such as India and China where C-HIV is endemic and Thailand, Indonesia and Vietnam where AE-HIV predominates. Developed nations with increasing AE-HIV prevalence such as Japan may also benefit from CoRSeq\textsubscript{3}. CoRSeq\textsubscript{3} has been developed as an online platform that will soon be freely available at www.burnet.edu.au/coreceptor.

<table>
<thead>
<tr>
<th>Coreceptor Usage Prediction Technique</th>
<th>B-HIV</th>
<th>C-HIV</th>
<th>D-HIV</th>
<th>AE-HIV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>185 CXCR4-using and 526 R5</td>
<td>80 CXCR4-using and 429 R5</td>
<td>57 CXCR4-using and 80 R5</td>
<td>50 CXCR4-using and 126 R5</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coreceptor Usage Prediction Technique</th>
<th>CoRSeq\textsubscript{3}_B</th>
<th>CoRSeq\textsubscript{3}_C</th>
<th>CoRSeq\textsubscript{3}_D</th>
<th>CoRSeq\textsubscript{3}_AE</th>
<th>geno2pheno 1% FPR</th>
<th>geno2pheno 2.5% FPR</th>
<th>geno2pheno 5% FPR</th>
<th>geno2pheno 10% FPR</th>
<th>geno2pheno 15% FPR</th>
<th>geno2pheno 20% FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-HIV</td>
<td>81.1</td>
<td>79.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C-HIV</td>
<td>-</td>
<td>-</td>
<td>88.8</td>
<td>90</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-HIV</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>86</td>
<td>83.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AE-HIV</td>
<td>34.6</td>
<td>99.3</td>
<td>52.5</td>
<td>99.8</td>
<td>49.1</td>
<td>97.5</td>
<td>60</td>
<td>99.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>geno2pheno 1% FPR</td>
<td>58.4</td>
<td>97.2</td>
<td>67.5</td>
<td>99.8</td>
<td>70.2</td>
<td>87.5</td>
<td>78</td>
<td>92.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>geno2pheno 2.5% FPR</td>
<td>68.1</td>
<td>93.7</td>
<td>72.5</td>
<td>97.2</td>
<td>86</td>
<td>75</td>
<td>82</td>
<td>82.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>geno2pheno 5% FPR</td>
<td>74.1</td>
<td>85.3</td>
<td>80</td>
<td>93.7</td>
<td>89.5</td>
<td>53.8</td>
<td>88</td>
<td>57.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>geno2pheno 10% FPR</td>
<td>74.6</td>
<td>80.1</td>
<td>83.8</td>
<td>91.4</td>
<td>89.5</td>
<td>52.5</td>
<td>88</td>
<td>51.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>geno2pheno 15% FPR</td>
<td>77.8</td>
<td>69.8</td>
<td>85</td>
<td>85.8</td>
<td>89.5</td>
<td>48.8</td>
<td>90</td>
<td>40.5</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>