IMPROVING CLASSIFICATION FOR RECENT INFECTION USING TOP SCORING PAIRS

#0975

Athena Chen[1], Oliver Laeyendecker[2,3], Charles S. Morrison[4], Susan H. Eshleman[2], Ingo Ruczinski[1]

[1]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, [2]Johns Hopkins University School of Medicine, Baltimore, MD, USA, [3] NIH, Bethesda, MD, USA, [4]FHI 360, Durham, NC, USA

BACKGROUND

- Many methods for cross-sectional HIV incidence testing classify individuals as recently-infected using characteristics of the antibody (Ab) response to HIV infection, such as titer, class, and avidity.
- These Ab responses can be affected by viral suppression, resulting in false-recent misclassification and overestimation of HIV incidence.
- Based on previous ART initiation guidelines, low viral load (VL) is often used as a surrogate for non-recent infection.
- As availability and use of ART increases worldwide, algorithms that do not include HIV VL and are not impacted by viral suppression are needed for cross-sectional HIV incidence estimation.
- We developed a classifier based on top scoring pairs (TSP) that identifies recent HIV infections without use of VL data.

METHODS

- A massively parallel antibody profiling system (VirScan) was used to quantify Ab binding to 3,384 peptides spanning the HIV genome. The sample set included 258 samples from 57 women in Zimbabwe and Uganda with HIV subtype C infection and known duration of infection (Table 1).
- Samples were randomly split into training and testing sets by individual. Optimal classification cutoffs and the number of voting pairs were determined using the training data and were then used to analyze the testing data. The performance of the final classifier was validated using an independent data set.
- The 4TSP classifier used the relative abundance of Abs to four peptide pairs to classify samples as recent (2-6 months) or non-recent (≥18 months). Overall sample classification was determined by a majority voting system (Figure 1).
- We compared the performance of the TSP algorithm to a widely used algorithm: the Limiting Antigen (LAg) Avidity assay (ODn ≤1.5) and VL >1,000 copies/mL (LAg+VL). Data were compared using receiver operating curves (ROC), classification accuracies, and probability of appearing recently infected as a function of duration of infection. Logistic regression was used to model probability of appearing recently infected.

A top scoring pairs (TSP) classifier of four peptide pairs performed better than the widely used LAg+VL algorithm for identifying recent infections without relying on viral load information.

The mean window period of the 4TSP classifier is twice as long as the mean window period for the LAg+VL algorithm, which may make it possible to obtain accurate incidence estimates from smaller cross-sectional surveys.

- **4TSP: 217 days (95% CI: 183 – 257 days)**
- **LAg + VL: 108 days (95% CI: 76 – 146 days)**

Table 1: Characteristics of samples in the training, test, and validation sets.

	Training	Testing	Validation
# of persons	38	19	19
# of samples	176	82	26
Recent infection	63	27	1
Non-recent infection	113	55	25
Median # of samples/participant (range)	5 (2 – 7)	4 (2 – 7)	1 (1 – 3)
Median duration of infection (range; years)	2.9 (0.2 – 8.7)	3.1 (0.2 – 8.3)	5.1 (1.6 – 8.8)
Recent infection	0.3 (0.2 – 0.6)	0.3 (0.2 – 0.4)	0.3 (0.3 – 0.3)
Non-recent infection	4.5 (1.5 – 8.7)	4.9 (1.5 – 8.3)	5.1 (1.6 – 8.8)
Median age of participant (range)	29 (21 – 43)	27 (19 – 44)	33 (21 – 45)
Median CD4 cell count (range; cells/mm^3)	464 (59 – 1113)	494 (71 – 1216)	127 (54 - 945)
Median $\log_{10}$ viral load (range; copies/mL)	4.0 (2.1 – 5.8)	4.3 (2.2 – 5.9)	4.4 (2.0 – 6.1)

Figure 2: ROC curves for the four peptide pairs and the LAg-Avidity assay.

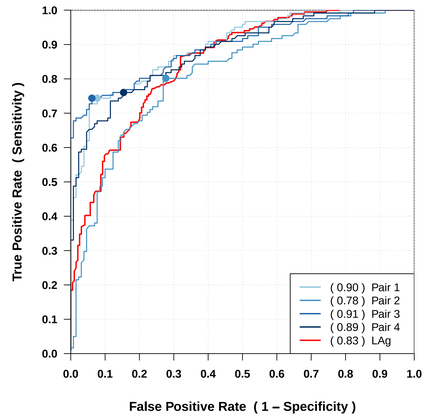


Figure 1: 4TSP Classifier for recent (2-6 months) and non-recent (≥18 months) infection

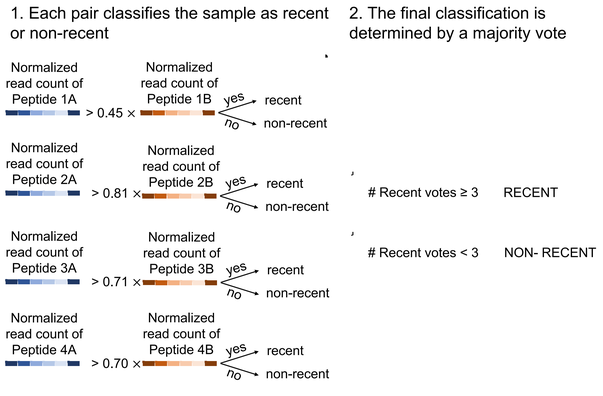
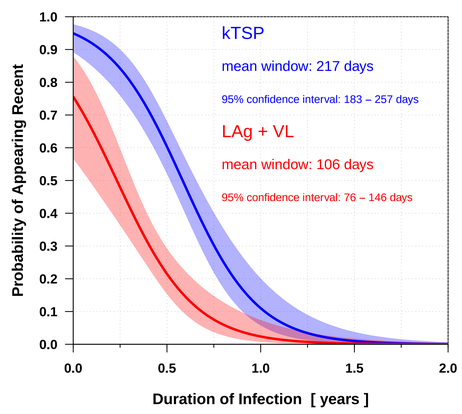


Figure 3: Probability of being classified as recent as a function of duration of infection for the 4TSP classifier and the LAg + VL algorithm.



RESULTS

- At the same classification accuracy for non-recent samples, each of the four peptide pairs correctly classified more of the recent samples than the LAg-Avidity assay alone (Figure 2).
- The 4TSP classifier had a larger mean window period than the LAg+VL algorithm (217 days vs. 106 days; Figure 3).
- In the training and testing sets, the 4TSP classifier correctly classified a higher proportion of recent samples than the LAg+VL algorithm (Table 2). Because the number of recent samples in the validation set was low (N=1), further studies are needed to validate the 4TSP classifier.

Table 2: Proportion of samples correctly classified by the 4TSP classifier and the LAg+VL algorithm. Some samples were missing VL data and could not be classified as recent or non-recent by the current LAg+VL algorithm.

	4TSP Classifier			LAg + VL		
	Correct	Total	Percent	Correct	Total	Percent
Training	162	176	92%	139	169	82%
Recent	49	63	78%	27	57	47%
Non-recent	113	113	100%	112	112	100%
Test	77	82	94%	61	77	62%
Recent	22	27	81%	8	24	33%
Non-recent	55	55	100%	53	53	100%
Validation	24	26	92%	25	26	96%
Recent	0	1	0%	0	1	0%
Non-recent	24	25	96%	25	25	100%

SUMMARY AND CONCLUSIONS

- We identified four peptide pairs that outperformed the standard LAg+VL algorithm for identifying samples from individuals with recent HIV infection. These peptides can be incorporated into a simple assay for field use.
- With a larger mean window period, the TSP classifier may be likely to provide accurate estimates of incidence from smaller cross-sectional surveys, and has the advantage that it does not require use of VL data. Further validation is needed to evaluate the performance of the classifier for estimating incidence.
- The TSP approach can be easily applied to other populations and viral subtypes to identify novel peptide signatures for cross-sectional incidence estimation.

Johns Hopkins Bloomberg School of Public Health · Johns Hopkins School of Medicine · NIH National Institute of Allergy and Infectious Diseases · Center for AIDS Research, Johns Hopkins University · fhi 360