2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

Evolved Neural Networks for HIV-1 Co-receptor Identification

Gary B. Fogel, Enoch S. Liu, Marco Salemi, Susanna L. Lamers, Michael S. McGrath

***Abstract*—HIV-1 infects a variety of cell types such as macrophages, T-cells and dendritic cells by expressing different chemokine receptors. R5 HIV-1 viruses use the CCR5 co-receptor for entry, X4 viruses use the CXCR4 co-receptor, and several viral strains make use of both co-receptors (a so-called "dual tropic" or R5X4 virus). Both X4 and R5X4 viruses are associated with late stage rapid progression to AIDS. It remains difficult to identify viral co-receptor type in advance of treatment, especially the R5X4 variety. In this paper we extended previous work to classify HIV-1 tropism using evolved neural networks and a larger set of HIV-1 sequences and features to improve overall classification accuracy.**

I. Introduction

Approximately 34 million people worldwide are infected with HIV-1 and more than 1,100,000 people live with HIV infection in the United States [1]. While combined antiretroviral therapy (cART) has been effective in increasing the lifespan of those infected with HIV-1, it does not cure or clear individuals from viral infection [2-3]. Patients can become resistant to cART, and side effects such as metabolic disease or neurological disorders, can be fatal [4-6]. HIV-1-infected patients require close monitoring of viral load, along with tests that identify specific viral genetic mutations associated with drug resistance [7]. One class of drugs, called entry inhibitors (EIs), does not target the virus directly; instead they target receptors on the cell surface of specific immune cells (Figure 1). The chemokine co-receptor CCR5 (R5) is the major co-receptor for macrophage-tropic virus strains, and plays a crucial role in the sexual transmission of HIV-1 [8]. T-cell tropic viruses use the co-receptor CXCR4 (X4) to enter target cells, and usually, but not always, emerge late in disease progression. Some primary HIV-1 isolates are dual-tropic (R5X4) and can use both co-receptors [8]. Many studies have discussed the importance of viral tropism in disease progression for HIV-infected individuals [9-12] and extensive reviews of different viral phenotypes are available in the literature [13-16].

Artificial neural networks (ANNs) have been applied to HIV co-receptor prediction with limited success [17-21]. Most of these applications have used backpropagation for neural network training. For example Resch et al. [17] predicted HIV-1 co-receptor usage for X4 viruses from envelope V3 loop sequences using neural networks trained via Bayesian regulation modification of backpropagation. These efforts demonstrated the merit of the approach, however the mean reliability of correctly classifying X4 viruses was ~69% and considered insufficient for use in clinical settings despite 80% sensitivity and 89% specificity of the best ANN. Further, no classifier had been developed at that time to identify R5X4 strains from genotypic information alone.

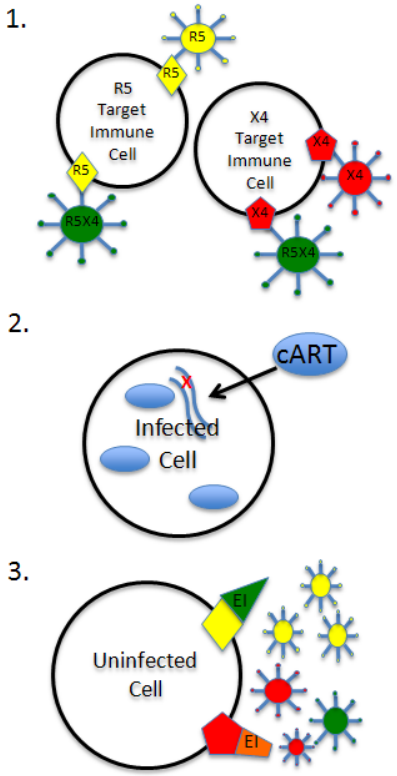

Fig. 1. Co-receptors, virus phenotypes and entry inhibitors. 1) HIV enters immune cells by binding co-receptors R5 or X4 on the cell surface. Three viral phenotypes of HIV are known: those that bind R5 receptors, those that bind X4 receptors and dual tropic (R5X4) viruses can bind both receptors. Complex genetic sequences on the virus surface determine which cellular co-receptor a virus will bind. 2) Once a cell is infected, combined antiretroviral therapy (cART) can reduce competent replication of virus; however, it can also interfere with other cellular functions. 3) EIs block virus binding to specific receptors and do not penetrate cells. Thus, this is an attractive alternative/ complement to cART; however, in order to use entry inhibitors, an individual must be assayed to determine which co-receptors their virus populations are using because of the risk of altering viral fitness to a more aggressive phenotype (see discussion below).

G.B. Fogel and E. Liu are with Natural Selection, Inc., San Diego, CA 92121 USA (phone: 858-455-6449; fax: 858 455-1560; email: gfogel@natural-selection.com).

M. Salemi is with the University of Florida, Department of Pathology and Laboratory Medicine, Gainesville, FL 32610 USA (email: salemi@pathology.ufl.edu)

S.L. Lamers is with BioInfoExperts, LLC, Thibodaux, LA 70302 USA (email: susannagene@bellsouth.net).

M.S. McGrath is with the University of California at San Francisco, Department of Laboratory Medicine, Pathology, and Medicine and the AIDS and Cancer Specimen Resource, San Francisco, CA 94143 USA (email: mmcgrath@hemeonc.ucsf.edu)

This work was supported by the U.S. National Institutes of Health under Grant R01 MH100984.

Although backpropagation is a common strategy for ANN optimization, convergence is only guaranteed to a locally optimal solution. A different approach to ANN optimization makes use of evolutionary computation to discover weight assignments or evolve the ANN architecture itself. Natural evolution provides inspiration for an approach that mimics random variation and selection to search the space of possible classifiers in a global sense. Evolved neural networks (ENNs) [22-25] have been applied on a wide variety of biochemical data mining problems [26-29].

Previously we presented the first use of ENNs to classify HIV-I co-receptor use [29]. That research was based on a small set of 149 HIV-1 V3 loop sequences (77 R5, 31 R5X4, and 41 X4 sequences) from a variety of HIV subtypes. Using multiple sequence alignment and removal of invariant positions, 9 features were calculated for each of 35 sequence positions (a total of 317 features) plus an additional 2 features V3-domain-level features were calculated. The feature space was reduced further to 248 to eliminate uninformative features. Fully connected feed-forward ENNs were used to map the input vectors for each sequence to co-receptor usage classification using increasingly largely feature sets as input. Given the small sample sizes, leave-one-out cross validation was used to minimize the resulting mean squared error of classification. This process was repeated for various inputs from 2 to 30, using two hidden nodes, and 1 output node in all cases. This process allowed us not only to evolve useful classifiers but identify feature combinations that were important for the classification process. ENNs were trained to classify R5 sequences from X4 sequences, and additional ENNs were trained to classify R5X4 sequences from either R5 or X4 sequences. The results of this approach led to a mean classification accuracy of 88.9% for R5 vs. X4 and mean classification accuracy of 75.5% for R5X4 vs. R5 or X4. These results for the first time provided the ability to classify dual-tropic viruses with reasonable accuracy.

In this paper we extended this approach using ENNs on a far larger sequence database, with additional features, to derive classifiers for four separate decisions (R5 vs. X4, R5 vs. R5X4, X4 vs. R5X4, and R5 vs. R5X4 vs. X4). Section II below reviews the methods used to develop the data, Section III outlines the methods used for ENN training and evaluation, Section IV reviews the results of the approach, and Section V provides some conclusions and future work.

II. Data Preparation

A. V3 loop sequences

HIV subtype B sequences (3,452 R5, 197 X4, and 545 R5X4) were downloaded from the HIV database at Los Alamos (http://www.hiv.lanl.gov/content/index) and translated into amino acid sequences. These sequences were then aligned computationally using ClustalW within the MEGA sequence analysis package [30]. The alignment was then manually curated to correct obvious alignment errors. Duplicate sequences were removed. A representative alignment of just three of these sequences (one for each tropism) is shown in Figure 2.

```
      1   *    *    *    *    *    *    *   40
R5    CERPNNNTR-RS-IQI---GPGRAWFEAEDIIGDIRKAHC
X4    CTRPNNNTR-KR-IRIQ-RGPGRAFVTIGK-IGNMRQAHC
R5X4  CIRPNNNTR-RS-IPI---GPGRAFYATGDIIGDIRQAYC
```

Fig. 2. Representative alignment. All 4194 sequences were aligned. Gaps were inserted as required to maximize the alignment. In the above excerpt, asterisks indicate every fifth position in the alignment. Dashes represent gapped positions.

B. Feature generation

In light of the sequence dataset resulting from II.A above, for each of the 40 alignment positions, features for model development (Table I) were calculated. While some of these were position dependent (e.g., glycosylation at specific positions), the remainder of the features were calculated for all positions and separately again for just the 5' end of the alignment (positions 9-14), for the 3' end of the alignment (positions 22-28) and lastly for a second region at the 3' end of the alignment (positions 31-37). These regional calculations were with respect to known associations of particular regions with HIV tropism in the literature.

The features in Table I were selected using the available

TABLE I
FEATURES FOR MODEL DEVELOPMENT

Molecular weight	Beta Levitt	Beta Chou and Fasman
Bulkiness	Antiparallel beta	Beta sheet
Polarity	AA Comp	Coil
Recognition factors	Relative mutability	Beta-sheet Levett
Hydrophobicity		
Sweet	Number of codons	Beta strand
Kyte and Doolittle	Polarity	Parallel beta
		AA Comp Swiss
Abraham and Leo	Refractivity	Prot
Bull and Breese	Eisenberg	Volume
Guy	Hopp and Woods	Charge
Miyazawa	Manavalan	HP Scale
Roseman	Black and Mould	Surface Area
		pKa alpha
Wolfenden	Fauchere	carboxylate
Wilson	Janin	pKa - amine
Cowan	Rao and Argos	pl at 25C
Aboderin	Tanford	Exchange
HPLC/TFA	Welling	Charge polarity
		Hydrophobicity
Meek	Parker	Membership Class
Mol fraction of	Cowan and	Mass Membership
buried res	Whittaker	class
		Surface Exposure
Chothia	Browne	Membership Class
		2D propensity
Grantham	Transmembrane	Membership Class
		AA Comp Swiss
Average Flex	Retention	Prot
	%accessible	Charge Conversion
Chou and Fasman	residues	Table
		Glycosylation
Alpha helix	Rose	Position 6 through 8
		Glycosylation
Beta Sheet	Avg Area buried	Position 5,7,9
	Alpha Chou and	Charge at position
Alpha Levitt	Fasman	12
Charge at position	Total sequence	Total sequence
30	glycosylation	charge

literature about tropism as a guide and also through resources such as those found via ExPASy ProtParam (www.expasy.org/protparam) [31]. This process resulted in ~3000 possible feature-positions that could be provided as input to a model for classification.

C. Feature processing

Using all features, linear regression was used to remove uncorrelated features and determine which features are most useful in separating the tropism classes independently for each of four decisions (R5 vs. R5X4, R5 vs. X4, R5X4 vs. X4, and R5 vs. R5X4 vs. X4). The top features identified over all tropism decisions were: overall V3 loop charge, charge at position 12, charge conversion table, Janin, Eisenberg, Volume, Refractivity, Tanford, Wolfenden, Cowen, Abraham and Leo, average area buried, bulkiness, Chothia, beta sheet). Although specific tropism decisions had slightly different orderings of the top features, these 15 were common across most tropism decisions and therefore considered useful for rapid model development.

III. Evolved Neural Network Training and Evaluation

For classifier development, feed-forward, fully-connected, ANNs with 15 inputs, 3 hidden nodes, and 1 output node, were evolved using a population of 100 parents and 100 offspring ANNs. Tournament selection with 4 opponent ANNs was used for selection. All hidden nodes used a sigmoid activation function, with initial sigma 0.1, initial weights 0.0 with inputs normalized to [0.1,0.9]. For the purpose of evolving ANNs, each ANN was coded as a real-valued vector of the weights and biases associated with the ANN in accordance with prior work [22-24]. Fitness was measured by taking the mean squared error (MSE) of the ANN prediction relative to the actual value as a measure of predictive accuracy for each sample using the equation:

$$MSE = \frac{1}{N\sum_{k=1}^{N}(P_k - O_k)^2} \qquad (1)$$

where P is the predicted activity for the kth sample, O is the observed activity for the kth sample, and N is the number of patterns in the training set. MSE was minimized using evolutionary computation. Optimization proceeded on the training data for 10,000 generations, monitoring both training and testing MSE to determine the number of generations that minimized both training and testing MSE without increased MSE on the testing samples. Once identified, ANNs were re-evolved for that number of generations, with the best ENN used to process the remaining held-out validation set for final evaluation. Convergence plots of the learning over the training examples provided a means to determine the most appropriate number of generations of evolution without overtraining. Each best evolved neural network was then processed using a threshold to determine tropism class above or below the threshold. For R5 vs. R5X4 this threshold was 0.3, for R5 vs. X4 the threshold was 0.25, for R5X4 vs. X4 the threshold was 0.52, and for R5 vs. R5X4 vs. X4 the thresholds were 0.7 and 1.10 respectively.

IV. Results

Convergence plots showing neural network optimization over generations of simulated evolution are shown in Figures 3 through 6 as average MSE on training and testing samples for the three shuffles of the data for each tropism class. In each case MSE decreases on the training data asymptotically as expected, while testing shows a similar decrease and then plateaus or even increases again in later generations indicating possible overtraining. Using this data, the number of generations with lowest MSE for both training and testing was obtained, and neural networks were then re-evolved for that number of generations, with the best-evolved neural network used to process the held-out validation data.

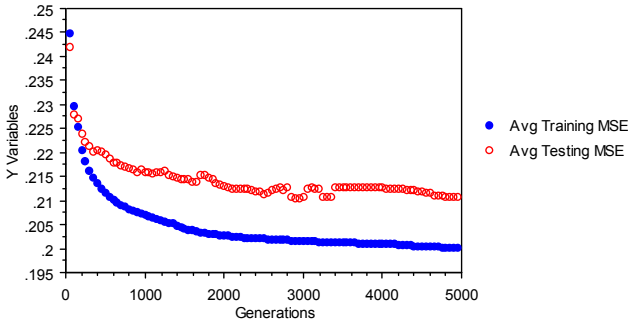

Fig. 3. Mean convergence for training and testing MSE over the three random shuffles of R5 vs. R5X4.

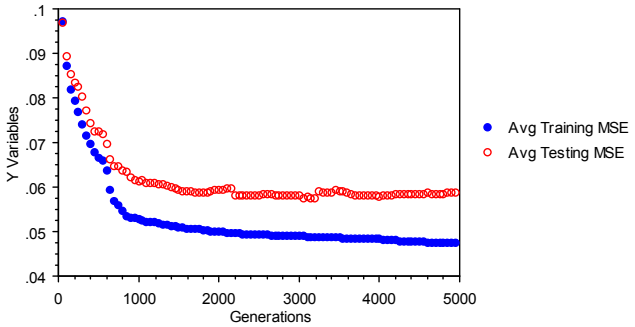

Fig. 4. Mean convergence for training and testing MSE over the three random shuffles of R5 vs. X4.

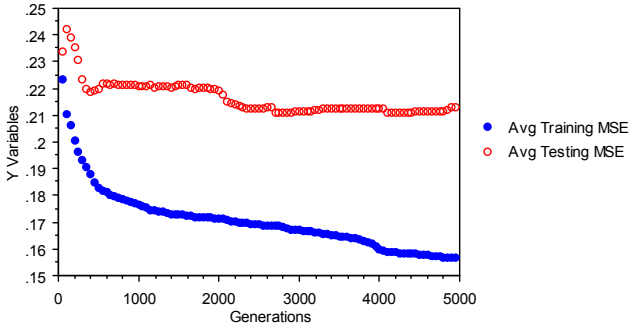

Fig. 5. Mean convergence for training and testing MSE over the three random shuffles of R5X4 vs. X4.

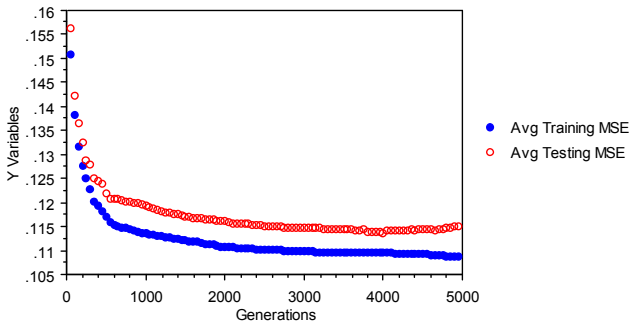

Fig. 6. Mean convergence for training and testing MSE over the three random shuffles of R5 vs. R5X4 vs. X4.

Results for each classification decision (R5 vs. R5X4, R5 vs. X4, X4 vs. R5X4, R5 vs. R5X4 vs. X4) are provided in Tables II through V respectively. Each table provides mean performance over all three random shuffles of the training, testing, and validation data. Performance on the decision of R5 vs. X4 was superior, and improved upon previous research, as did the accuracies of R5 vs. R5X4 and X4 vs. R5X4. The three-class decision of R5 vs. R5X4 vs. X4 was much harder for the neural networks to learn. Curiously the off diagonal errors were non-symmetrical. For instance when the actual tropism was R5, the neural networks almost never misclassified the sequence as being X4. Rather they misclassified as being R5X4. Similarly when the actual tropism was X4, the neural networks almost never misclassified the sequence as R5. Rather, they misclassified as R5X4. However, when the actual sequence was dual tropic (R5X4), the neural networks were twice as likely to misclassify the sequence as X4 than R5.

TABLE II
R5 VS. R5X4

Mean Training Performance (*n*=2398)		**Actual** R5 (*n*=2060)	**Actual** R5X4 (*n*=338)
Predicted	R5	1662/2060=80.7%	63/338=18.8%
	R5X4	399/2060=19.3%	274/338=81.2%

Mean Testing Performance (*n*=1199)		**Actual** R5 (*n*=1039)	**Actual** R5X4 (*n*=160)
Predicted	R5	809/1039=77.9%	34/160=21.2%
	R5X4	229/1039=22.1%	126/160=78.8%

Mean Validation Performance (*n*=400)		**Actual** R5 (*n*=353)	**Actual** R5X4 (*n*=47)
Predicted	R5	283/353=80.2%	11/47=24.1%
	R5X4	70/353=19.8%	36/47=75.9%

V. Discussion

The accurate assessment of co-receptor usage is important in several aspects of HIV research including studies about viral transmission, evolution, adaptation, viral reservoirs in specific tissues, and other *in vitro* and *in vivo* studies. Statistical and computational approaches to distinguish these tropism classes have their own history, generally with the realization that dual-tropic R5X4 viruses are most difficult to classify. However R5X4 viruses are important to monitor in light of their association with rapid disease progression.

TABLE III
R5 VS. X4

Mean Training Performance (*n*=2189)		**Actual** R5 (*n*=2070)	**Actual** X4 (*n*=119)
Predicted	R5	2033/2070=98.2%	8/119=6.4%
	X4	37/2070=1.8%	112/119=94.4%

Mean Testing Performance (*n*=1095)		**Actual** R5 (*n*=1040)	**Actual** X4 (*n*=56)
Predicted	R5	1003/1040=96.4%	6/56=10.1%
	X4	36/1040=3.5%	50/56=89.9%

Mean Validation Performance (*n*=365)		**Actual** R5 (*n*=343)	**Actual** X4 (*n*=22)
Predicted	R5	337/343=98.3%	3/22=15.2%
	X4	6/343=1.7%	19/22=84.8%

TABLE IV
R5X4 VS. X4

Mean Training Performance (*n*=445)		**Actual** R5X4 (*n*=329)	**Actual** X4 (*n*=116)
Predicted	R5X4	236/329=71.8%	31/116=27.0%
	X4	93/329=28.2%	83/116=71.6%

Mean Testing Performance (*n*=223)		**Actual** R5X4 (*n*=164)	**Actual** X4 (*n*=60)
Predicted	R5X4	115/164=70.0%	18/60=30.0%
	X4	49/164=30.0%	42/60=70.0%

Mean Validation Performance (*n*=74)		**Actual** R5X4 (*n*=52)	**Actual** X4 (*n*=22)
Predicted	R5X4	43/52=82.1%	5/22=24.2%
	X4	9/52=17.9%	17/22=75.8%

TABLE V
R5 vs. X4 vs. R5X4

Mean Training Performance (*n*=2517)		Actual		
		R5 (*n*=2071)	R5X4 (*n*=333)	X4 (*n*=113)
Predicted	R5	1401/2071= 67.6%	48/333= 14.5%	2/113= 1.5%
	R5X4	661/2071= 31.9%	167/333= 50.1%	31/113= 27.2%
	X4	10/2071= 0.5%	118/333= 35.4%	80/113= 71.3%

Mean Testing Performance (*n*=1258)		Actual		
		R5 (*n*=1036)	R5X4 (*n*=160)	X4 (*n*=62)
Predicted	R5	701/1036= 67.7%	24/160= 15.2%	1/62= 1.1%
	R5X4	331/1036= 31.9%	89/160= 55.9%	16/62= 25.1%
	X4	4/1036= 0.4%	46/160= 28.8%	46/62= 73.8%

Mean Validation Performance (*n*=419)		Actual		
		R5 (*n*=345)	R5X4 (*n*=52)	X4 (*n*=22)
Predicted	R5	244/345= 70.7%	11/52= 20.5%	1/22= 6.1%
	R5X4	09/345= 28.3%	31/52= 59.6%	4/22= 19.7%
	X4	3/345= 1.0%	10/52= 19.9%	16/22= 74.2%

As previously described, entry inhibitors bind specific chemokine receptors on cell surfaces and reduce the entry of viruses into required immune cells necessary for HIV replication. Many entry inhibitors targeting different cell receptors are under evaluation [32]. For example, Maraviroc targets the R5 co-receptor [33], AMD070 targets the X4 co-receptor [34] and Ibalizumab binds both X4 and R5 co-receptors [35]. Because R5 viruses are commonly transmitted and present during prolonged antiretroviral treatment [36], R5 inhibitors may be beneficial towards keeping viral loads low [38]. On the other hand, X4 viruses are associated with progression to AIDS; therefore, treatment with an R5 inhibitor while X4 viruses are present allows for the unwanted opportunity for growth and evolution of more aggressive viral populations in the infected individual [38]. For this reason, virus populations must be carefully monitored for co-receptor usage during treatment with an R5 antagonist. Alternatively, treatment with an X4 antagonist would be useless if the viral population was predominantly R5; however, such a drug may be useful in prolonging life in a patient with a mixed (R5X4) viral population. Ibalizumab binds both receptors and, like many other entry inhibitors is still under evaluation [35,38].

Signature pattern analysis, clustering, and phylogenetic analysis are commonly used to define viral subpopulations, however these approaches only assess nucleic or amino acid variation within a properly aligned set of sequences without considering other biological features present within the underlying data. Early viral tropism studies identified two charged amino acids positions in the envelope V3 domain that have been used to estimate tropism [39]; however, this method is unreliable for X4 viruses and largely ineffective in identifying dual-tropic (R5X4) viruses. Several bioinformatics prediction systems have been developed and are available via the Internet: WetCat [40], WebPSSM [41], geno2pheno [42]. These algorithms are used primarily by academia due to the high cost of biological assays. The algorithms are trained on genotypic information and known corresponding phenotype. So far, two cohorts have been analyzed with such genotypic approaches, resulting in frequencies of R5 virus strains that are within the range of those reported with biological assays [43]. However, all three of these systems are less reliable for X4 sequence identification and lack the ability to identify R5X4 with reasonable accuracy.

In this study, 15 of the features from Table 1 were ultimately used for classification in the ENNs. As in previous studies, various charged positions along the alignment were found important in tropism decisions. Importantly, while the system can help predict viral phenotypes, it can also provide insight to viral features that could be used in further studies aimed at reducing co-receptor binding. Seven interesting hydrophobicity scales were important for viral phenotype decisions, for example, the Janin scale [44] measures the free energy of transfer from the inside to the outside of globular proteins and the Tanford scale [45] measures the contribution of hydrophobic interactions to the stability of protein confirmation. Structural scales included the normalized frequency for beta-sheet formation, bulkiness and average area buried.

Frequently, HIV-associated studies are aimed at identifying subsets of viral sequences associated particular pathologies, for example, the identification of a brain-specific or lymphoma-specific virus has been researched [46-51]. The HIV Nef protein has been implicated in various HIV disease pathologies [52-53]. To support these efforts, An ENN analysis of Nef sequences isolated from tissue biopsies, such as brain and tumor sequences, could resolve if specific Nef proteins are not only associated with these tissues, but also pinpoint detailed features of Nef that contribute to the specific pathology.

VI. Conclusions

In this paper we have extended our previous efforts using evolved neural networks for tropism classification using biological and positional features of HIV-1 sequences to classify co-receptor phenotype. The current effort was derived from a substantial increase in the amount of sequence information available for modeling, the number of subtype-B HIV sequences with known co-receptor usage, and new

scales that describe amino acid physico-chemical properties. The result was an improved accuracy for all decisions using validation data. A similar approach can be used to sample other sequence populations associated with disease pathogenesis.

In the future, we plan on additional experiments to evolve the input layer to allow evolution to accept any of the ~3000 possible features as input, sub-selecting to smaller feature sets that may increase overall accuracy. Further, we intend to allow the entire architecture to evolve, including the number of hidden nodes and connections as these were chosen to be 3 and fully-connected arbitrarily, simply as a first pass on this dataset to see if performance could be improved over previous effort. In addition it should be recognized that the data set is largely unbalanced, with R5 sequences representing a large percentage of the available data. This imbalance may affect neural network performance, and future research will focus on repeating this analysis using balanced data sets for comparison.

REFERENCES

[1] Hall HI, Song R, Rhodes P, et al. "Estimation of HIV incidence in the United States." *JAMA.* Aug 6 2008;300(5):520-529.

[2] Alexaki A, Liu Y, Wigdahl B. "Cellular reservoirs of HIV-1 and their role in viral persistence." *Curr HIV Res.* Sep 2008;6(5):388-400.

[3] Aquaro S, Calio R, Balestra E, et al. "Clinical implications of HIV dynamics and drug resistance in macrophages." *J Biol Regul Homeost Agents.* 1998;12(1-2 Suppl):23-27.

[4] Anthony IC, Ramage SN, Carnie FW, Simmonds P, Bell JE. "Influence of HAART on HIV-related CNS disease and neuroinflammation." *J Neuropathol Exp Neurol.* Jun 2005;64(6):529-536.

[5] Estrada V, Portilla J. "Dyslipidemia related to antiretroviral therapy." *AIDS Rev.* Jan-Mar 2011;13(1):49-56.

[6] Nachega JB, Marconi VC, van Zyl GU, et al. "HIV Treatment Adherence, Drug Resistance, Virologic failure: Evolving concepts." *Infect Disord Drug Targets.* Mar 15 2011.

[7] Cortez KJ, Maldarelli F. "Clinical management of HIV drug resistance." *Viruses.* Apr 2011;3(4):347-378.

[8] Gorry PR, Ancuta P. "Coreceptors and HIV-1 pathogenesis." *Curr HIV/AIDS Rep.* Dec 29 2010.

[9] Berger EA, Murphy PM, Farber JM. "Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease." *0732-0582.* 1999;17:657-700.

[10] Broder CC, Collman RG. "Chemokine receptors and HIV." *0741-5400.* 1997;62(1):20-29.

[11] Murakami T, Yamamoto N. "Roles of chemokines and chemokine receptors in HIV-1 infection." *0925-5710.* 2000;72(4):412-417.

[12] Weber J, Piontkivska H, Quinones-Mateu ME. "HIV type 1 tropism and inhibitors of viral entry: clinical implications." *AIDS Reviews.* 2006;8(2):60-77.

[13] Clapham PR, McKnight A. "Cell surface receptors, virus entry and tropism of primate lentiviruses." *0022-1317.* 2002;83(8):1809-1829.

[14] Gorry PR, Churchill M, Crowe SM, Cunningham AL, Gabuzda D. "Pathogenesis of macrophage tropic HIV-1." *1570-162X.* 2005;3(1):53-60.

[15] Moore JP, Kitchen SG, Pugach P, Zack JA. "The CCR5 and CXCR4 coreceptors-- central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection." *0889-2229.* 2004;20(1):111-126.

[16] Weiss RA. "HIV receptors and cellular tropism." *IUBMB Life.* 2002;53(4-5):201-205.

[17] W. Resch, N. Hoffman, and R. Swanstrom, "Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks," *J. Virology*, vol. 76, 2001, pp. 3852-3864.

[18] J.A. Loannidis, T.A. Trikalinos, and M. Law, "HIV Lipodystrophy Case Definition Using Artificial Neural Network Modeling," *Antiviral Therapy*, vol. 8, 2003, pp. 435-441.

[19] D. Wang and B. Larder, "Enhanced Prediction of Lopinavir Resistance from Genotype by Use of Artificial Neural Networks," *J. Infectious Diseases*, vol. 188, pp. 653-660, 2003

[20] Z.L. Brumme, W.W.Y. Dong, B. Yip, B. Wynhoven, N.G. Hoffman, R. Swanstrom, M.A. Jensen, J.I. Mullins, R.S. Hogg, J.S.G. Montaner, and P.R. Harrigan, "Clinical and Immunological Impact of HIV Envelope V3 Sequence Variation after Starting Initial Triple Antiretroviral Therapy," *AIDS*, vol. 18, 2004, pp. F1-F9.

[21] L. Milich, B. Margolin, and R. Swanstrom, "V3 Loop of the Human Immunodeficiency Virus Type 1 Env Protein: Interpreting Sequence Variability," *J. Virology*, vol. 67, no. 9, 1993, pp. 5623-5634.

[22] D.B. Fogel, L.J. Fogel, and V.W. Porto, "Evolving Neural Networks," *Biological Cybernetics*, vol. 63, no. 6, 1990, pp. 487-493.

[23] V.W. Porto, D.B. Fogel, and L.J. Fogel, "Alternative Neural Network Training Methods," *IEEE Expert*, vol. 10, no. 3, 1995, pp. 16-22.

[24] X. Yao, "Evolving Artificial Neural Networks," *Proc. IEEE*, vol. 87, no. 9, 1999, pp. 1423-1447.

[25] N. Kohl and R. Miikkulainen, "An Integrated Neuroevolutionary Approach to Reactive Control and High-level Strategy," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 4, 2011, pp. 472-488.

[26] G.B. Fogel, "Computational Intelligence Approaches for Pattern Discovery in Biological Systems," *Briefings in Bioinformatics*, Vol. 9, 2008, pp. 307-316.

[27] D. Hecht, M. Cheung, and G.B. Fogel, "QSAR Using Evolved Neural Networks for the Inhibition of Mutant *Pf*DHFR by Pyrimethamine Derivatives," *BioSystems*, Vol. 92, 2008, pp. 10-15.

[28] G.B. Fogel and D. Hecht D, "Evolved Neural Networks for High Throughput Ligand Screening," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. Vol. 4(3), 2007, pp. 476-484.

[29] S. Lamers, M. Salemi, M. McGrath M, G.B. Fogel, "Prediction of R5, X4, and R5X4 HIV-1 Co-receptor Usage with Evolved Neural Networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics.* Vol. 5(2), 2008, pp. 291-300.

[30] Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0." *Molecular Biology and Evolution* 30: 2725-2729.

[31] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A "Protein Identification and Analysis Tools on the ExPASy Server*" in* John M. Walker (ed.) The Proteomics Protocols Handbook, Humana Press, 2005., pp 571-607.

[32] Gopi H, Chaiken I, Gopi H, Chaiken IGopi H, Chaiken Is; Drexel University, assignee. "Dual inhibitors of HIV-1 GP-120 interactions." US patent US7655464, 2009.

[33] Adamson CS, Freed EO. "Novel approaches to inhibiting HIV-1 replication." *Antiviral Res.* Jan 2010; 85(1):119-141.

[34] Wilkinson RA, Pincus SH, Shepard JB, et al. "Novel compounds containing multiple guanide groups that bind the HIV coreceptor CXCR4." *Antimicrob Agents Chemother.* Jan 2011;55(1):255-263.

[35] Caffrey M. "HIV envelope: challenges and opportunities for development of entry inhibitors." *Trends Microbiol.* Apr 2011; 19(4):191-197.

[36] Wang YM, Wang B, Dyer WB, Lachireddy K, Peng NK, Saksena NK. "Evidence for predominance of CCR5-using HIV-1 strains during highly active antiretroviral therapy." *Curr HIV Res.* Mar 2007; 5(2):221-234.

[37] Wilkin TJ, Gulick RM. "CCR5 antagonism in HIV infection: current concepts and future opportunities." *Annu Rev Med.* 2012; 63:81-93.

[38] Jacobson JM, Kuritzkes DR, Godofsky E, et al. "Safety, pharmacokinetics, and antiretroviral activity of multiple doses of ibalizumab (formerly TNX-355), an anti-CD4 monoclonal antibody, in human immunodeficiency virus type 1-infected adults." *Antimicrob Agents Chemother.* Feb 2009; 53(2):450-457.

[39] Hoffman NG, Seillier-Moiseiwitsch F, Ahn J, Walker JM, Swanstrom R. "Variability in the human immunodeficiency virus type 1 gp120 Env protein linked to phenotype-associated changes in the V3 loop." *J Virol.* Apr 2002; 76(8):3852-3864.

[40] Pillai S, Good B, Richman D, Corbeil J. "A new perspective on V3 phenotype prediction." *AIDS Res Hum Retroviruses.* Feb 2003; 19(2):145-149.

[41] Jensen MA, Coetzer M, van 't Wout AB, Morris L, Mullins JI. "A reliable phenotype predictor for human immunodeficiency virus type 1

subtype C based on envelope V3 sequences." *J Virol.* May 2006; 80(10):4698-4704.
[42] Beerenwinkel N, Daumer M, Oette M, et al. "Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes." *Nucleic Acids Res.* Jul 1 2003;31(13):3850-3855.
[43] Dybowski JN, Heider D, Hoffmann D. "Prediction of co-receptor usage of HIV-1 from genotype." *PLoS Comput Biol.* Apr 2010; 6(4):e1000743.
[44] Janin J "Surface and inside volumes in globular proteins." *Nature*, 1979;277:491-492.
[45] Tanford C "Contribution of hydrophobic interactions to the stability of the globular conformation of proteins." *J. Am. Chem. Soc.* 1962 84:4240-4274.
[46] Lamers SL, Poon AF, McGrath MS. "HIV-1 nef protein structures associated with brain infection and dementia pathogenesis." *PLoS One.* 2011 Feb 9;6(2):e16659.
[47] Lamers SL, Gray RR, Salemi M, Huysentruyt LC, McGrath MS. "HIV-1 phylogenetic analysis shows HIV-1 transits through the meninges to brain and peripheral tissues." *Infect Genet Evol.* 2011 Jan;11(1):31-7.
[48] Lamers SL, Salemi M, Galligan DC, Morris A, Gray R, Fogel G, Zhao L, McGrath MS. "Human immunodeficiency virus-1 evolutionary patterns associated with pathogenic processes in the brain." *J Neurovirol.* 2010 May;16(3):230-41.
[49] Salemi M, Lamers SL, Huysentruyt LC, Galligan D, Gray RR, Morris A, McGrath MS. "Distinct patterns of HIV-1 evolution within metastatic tissues in patients with non-Hodgkins lymphoma." P*LoS One.* 2009 Dec 3;4(12):e8153.
[50] Zhao L, Galligan DC, Lamers SL, Yu S, Shagrun L, Salemi M, McGrath MS. "High level HIV-1 DNA concentrations in brain tissues differentiate patients with post-HAART AIDS dementia complex or cardiovascular disease from those with AIDS." *Sci China C Life Sci*. 2009 Jul;52(7):651-6.
[51] Salemi M, Lamers SL, Yu S, de Oliveira T, Fitch WM, McGrath MS. "Phylodynamic analysis of human immunodeficiency virus type 1 in distinct brain compartments provides a model for the neuropathogenesis of AIDS." *J Virol.* 2005 Sep;79(17):11343-52.
[52] Lamers SL, Fogel GB, Singer EJ, Salemi M, Nolan DJ, Huysentruyt LC, McGrath MS. "HIV-1 Nef in macrophage-mediated disease pathogenesis." *Int Rev Immunol.* 2012 Dec;31(6):432-50.
[53] Lamers SL, Fogel GB, Huysentruyt LC, McGrath MS. "HIV-1 nef protein visits B-cells via macrophage nanotubes: a mechanism for AIDS-related lymphoma pathogenesis?" *Curr HIV Res*. 2010 Dec;8(8):638-40.