

# UCSF

## UC San Francisco Previously Published Works

### Title

Identification of dual-tropic HIV-1 using evolved neural networks

### Permalink

<https://escholarship.org/uc/item/0pd6s7dn>

### Authors

Fogel, Gary B  
Lamers, Susanna L  
Liu, Enoch S  
[et al.](#)

### Publication Date

2015-11-01

### DOI

10.1016/j.biosystems.2015.09.007

Peer reviewed



# HHS Public Access

Author manuscript

*Biosystems*. Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

*Biosystems*. 2015 November ; 137: 12–19. doi:10.1016/j.biosystems.2015.09.007.

## Identification of Dual-Tropic HIV-1 Using Evolved Neural Networks

Gary B. Fogel<sup>a</sup>, Susanna L. Lamers<sup>b,\*</sup>, Enoch S. Liu<sup>a</sup>, Marco Salemi<sup>c</sup>, and Michael S. McGrath<sup>d</sup>

<sup>a</sup>Natural Selection, Inc., San Diego, CA 92121

<sup>b</sup>Bioinfoexperts, LLC, Thibodaux, LA 70302

<sup>c</sup>University of Florida, Department of Pathology and Laboratory Medicine, Gainesville, FL 32610

<sup>d</sup>University of California at San Francisco, Department of Laboratory Medicine and the AIDS and Cancer Specimen Resource, San Francisco, CA 94143

### Abstract

Blocking the binding of the envelope HIV-1 protein to immune cells is a popular concept for development of anti-HIV therapeutics. R5 HIV-1 binds CCR5, X4 HIV-1 binds CXCR4, and dual-tropic HIV-1 can bind either coreceptor for cellular entry. R5 viruses are associated with early infection and over time can evolve to X4 viruses that are associated with immune failure. Dual-tropic HIV-1 is less studied; however, it represents functional antigenic intermediates during the transition of R5 to X4 viruses. Viral tropism is linked partly to the HIV-1 envelope V3 domain, where the amino acid sequence helps dictate the receptor a particular virus will target; however, using V3 sequence information to identify dual-tropic HIV-1 isolates has remained difficult. Our goal in this study was to elucidate features of dual-tropic HIV-1 isolates that assist in the biological understanding of dual-tropism and develop an approach for their detection. Over 1559 HIV-1 subtype B sequences with known tropisms were analyzed. Each sequence was represented by 73 structural, biochemical and regional features. These features were provided to an evolved neural network classifier and evaluated using balanced and unbalanced data sets. The study resolved R5X4 viruses from R5 with an accuracy of 81.8% and from X4 with an accuracy of 78.8%. The approach also identified a set of V3 features (hydrophobicity, structural and polarity) that are associated with tropism transitions. The ability to distinguish R5X4 isolates will improve computational tropism decisions for R5 vs. X4 and assist in HIV-1 research and drug development efforts.

### Keywords

HIV-1 coreceptor; viral tropism; artificial neural network; evolutionary computation; HIV phenotype

\*Corresponding Author: Susanna L. Lamers, Bioinfoexperts, LLC; Tel.: 985-413-0455; Fax: 985-493-3487; susanna@bioinfo.com.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Introduction

Approximately 34 million people are infected with HIV-1 worldwide and more than 1.1 million live with HIV-1 infection in the United States [1]. While combined antiretroviral therapy (cART) has increased the lifespan of HIV-1-infected individuals, cART does not clear viral infection [2-3]. Its success, however, has supported the current mission to cure HIV-1 disease [4]. However, patients can acquire resistance to cART, which results from specific genetic mutations in the viral genome [5-6]. Resistance creates the need for patients to change medication with drugs that attack the virus in different ways [7]. HIV-1 variation, whether occurring naturally or in direct response to cART, also influences escape from immune surveillance [8], disease pathogenesis [9-10], development of viral reservoirs [10], and a wide spectrum of diseases associated with metabolic disorder [11-13], neurological disorder [14] and cancer [15-16]. In order to enter immune cells, HIV-1 first binds a CD4 cellular receptor and then another receptor, usually CCR5 (R5) or CXCR4 (X4) [17]. These co-receptors are present on the surface of both macrophages and T-cells [18]; however, the complex preference of HIV-1 for specific co-receptors varies under different conditions, such as passage in cell cultures [18] or the state of the immune system [19]. Still, as co-receptor binding is required for successful HIV-1 infection, exploiting the process for therapeutic intervention remains a popular concept. Entry inhibitors (EIs) are a class of drugs designed specifically to block coreceptors, thereby limiting the ability of HIV-1 to infect new cells. Currently, Maraviroc [20-21] blocks CCR5 while drug development continues for drugs that target CXCR4. Administering these drugs requires matching the therapy to the appropriate receptor in use by HIV-1. Furthermore, HIV-1 can evolve to acquire resistance to EIs [22] in a manner similar to that of other antiretroviral drugs. Therefore, the ability to monitor which receptor(s) are being used for viral entry is critical for appropriate treatment options based on the current and anticipated evolution of the virus.

R5 HIV-1 plays a crucial role in the transmission and establishment of HIV-1 [23]. Classically, the emergence of X4 viruses has been associated with the progression of acquired immune deficiency syndrome [24-25]. However, many patients never evolve X4 HIV-1 and instead evolve a highly macrophage-tropic HIV-1 with enhanced tropism for R5 [19] or “dual-tropic” HIV-1 (R5X4) [26], which can use both co-receptors for cellular entry [23-27-28-18]. Dual tropic HIV-1 is an interesting intermediate in that it is less efficient in binding either receptor than R5 or X4, but still allows for viral entry [26]. Furthermore, recent views suggest that X4 emergence is not associated with a highly pathogenic virus, but rather is the result of reduced host immune efficiency [26], which permits the accumulation of viral diversity [29], resulting in a wider spectrum of co-receptor usage. In this light, the ability to identify R5 and R5X4 viruses prior to X4 emergence with high accuracy would accelerate the study of HIV-1 evolution [30], the staging of disease progression [31-35-19-36-38], and the development of appropriate personalized therapies for those infected [22-39].

One of the principal determinants of HIV-1 interactions with R5 or X4 cellular coreceptors is the V3 domain of the HIV-1 envelope protein [40]. Co-receptor selection of viral isolates in this region is influenced by amino acid substitutions, insertions, and deletions [41]. HIV-1

V3 sequence data sets with known phenotypes combined with artificial neural networks (ANNs) or other machine learning strategies, primarily with a focus on backpropagation for training, have been applied to predict R5 vs. X4 co-receptor usage with reasonable success [40, 42–45]. However, the continued inability to identify dual-tropic R5X4, could account for decreased sensitivity and specificity of these methods if true R5X4 sequences are actually being misclassified as either R5 or X4 for model development and testing. Further, in the case of artificial neural networks (ANNs), 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 and/or evolve the ANN architecture itself. Evolved neural networks (ENNs) [46–50] have been applied with success to a wide variety of biochemical data mining problems [46, 51–53] and afford the researcher with the opportunity not only optimize a model for input-output mapping but to examine which feature combinations taken from a larger set of possible features are most relevant for high accuracy classification.

In a previous publication, we presented the first use of ENNs to classify HIV-1 co-receptor use [53]. That initial research was based on a small public set of 149 HIV-1 V3 loop sequences (77 R5, 31 R5X4, and 41 X4 sequences) from a variety of HIV-1 subtypes with known tropisms. 9 biochemical features for each of 35 amino acid positions and 2 additional V3-domain-level features were calculated. Fully connected feed-forward ENNs were used to map the features for each sequence to co-receptor usage classification using increasingly larger feature sets as inputs. The effort not only produced useful classifiers but also helped identify feature combinations that were important for classification. 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. This approach led to a mean classification accuracy of 88.9% for R5 vs. X4 and a mean classification accuracy of 75.5% for R5X4 vs. R5 or X4. This initial approach demonstrated strong potential for correctly classifying dual-tropic HIV-1 using an expanded set of sequences and features. In this paper, we used a larger database of over 1559 sequences and a broader assortment of 73 features to derive classifiers for four separate tropism decisions (R5 vs. X4, R5 vs. R5X4, X4 vs. R5X4, and R5 vs. R5X4 vs. X4). While many strategies for nonlinear machine learning could be applied to this problem such as support vector machines (SVMs), we specifically chose to use ENNs for this work in order to compare results to our previous effort as described above. Further, we evaluated the effects of balanced vs. unbalanced data sets on model performance. The results provide an indication of important features associated with tropism classification as well as improved detection of dual-tropic viruses.

## Methods

**V3 Loop Sequences.** Publically available V3 loop sequences (relative to HXB2 positions 7110–7217) for HIV-1 subtype B were downloaded from the HIV Database at the Los Alamos National Laboratory (<http://www.hiv.lanl.gov/content/index>) and translated into amino acid sequences. The search criteria limited each data set to either “only CCR5,” “only CXCR4,” or “only R5X4.” The resulting database consisted of 3452 R5 sequences, 197 X4 sequences, and 545 R5X4 sequences. The sequences were aligned using ClustalW within the MEGA5 sequence analysis package [75] and then manually edited to correct for any obvious

alignment errors. Identical sequences were removed. The final alignment used for feature generation below contained 1223 unique R5 sequences, 241 unique R5X4 sequences, and 95 unique X4 sequences. No limitation was put on the phenotype or culture method in order to preserve sequence diversity.

**Feature Generation and Processing.** While many studies have focused on characteristics at specific sites in the V3 loop relative to tropism [76–78], we took a different unbiased approach using both site-specific and regional characteristics. For each sequence, 73 features were calculated for each of the 40 alignment positions (Supplemental Table 1). These features were selected using the available tropism literature and also through resources such as ExPASy programs ProtScale and ProtParam ([www.expasy.org](http://www.expasy.org)) [79]. While some features were position-dependent (e.g., glycosylation at specific positions), the remaining features were calculated for all positions and, to determine if regional features were associated with tropism, the alignment was further reduced to smaller segments: the 5' end of the alignment (positions 9–14), 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). This process resulted in ~3,000 possible feature-positions that could be provided as input to a model for classification. Linear regression was then used to determine which features were 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) and to reduce the number of features to those with highest correlation to tropism class decision for each decision type.

**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 [46–49]. 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:

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

Where  $P$  is the predicted activity for the  $k$ th sample,  $O$  is the observed activity for the  $k$ th 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 5,000 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.

This process of learning was applied to the original data given its unbalanced tropism classes and also applied in separate experiments with artificially balanced classes to determine the extent to which data balance assisted neural network learning. Balancing was achieved by

first setting aside 10% of data from each class for a validation set. Three separate training/testing sets were then generated from the remaining data using 66% of the remaining data for training and 34% for testing. Class sizes in the training set were balanced by randomly duplicating sequences in the smaller classes until they matched the total size of the largest class. Convergence plots showing the minimization of mean squared error over the training examples were used 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.

## Results

The resulting alignment contained 1223 R5, 241 R5X4, and 95 X4 unique sequences spanning 40 positions, including gaps used to maximize the alignment. The length of the alignment allowed the inclusion of all sequences in the study, even those that were unusually long. Previous studies have relied on charges at positions 11 and 25 for tropism identification [54, 55]; in our developed alignment these charged positions were at positions 12 and 30 respectively. A representative alignment of one sequence from each tropism is shown in Fig. 1. Multiple sequences from a single individual were occasionally identified in the alignments, especially in the large R5 sequence alignment. It is debatable how the genetic background of these sequences could, if at all, influence the subsequent analysis; while a “subject” has a specific evolutionary history that results in specific evolutionary paths, “within-subject” evolution incorporates hundreds of environments (anatomical sites and many cell types) where HIV-1 can evolve [10, 56, 57]. Furthermore, patients are known to harbor mixtures of tropism phenotypes. For these reasons, we included unique sequences when isolated from the same subject.

While specific tropism decisions exhibited slightly different orderings of the top features, the features that provided the greatest correlation to all tropism decisions are listed in Table 1. These were considered useful for model development as a strategy to identify classifiers that used the same inputs and yet could classify different tropisms, and also served as input to the evolved neural networks. Convergence plots showing neural network optimization over generations of simulated evolution are shown in Fig. 2 as average MSE on training and testing samples for the three shuffles of the unbalanced and balanced data for each tropism class. In each case, MSE decreased asymptotically on the training data as expected, while testing showed a similar decrease and then plateaus or even increases in later generations indicating potential 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. In the case where no increase in testing MSE was identified, the best-evolved neural network was chosen from generation 5,000 as the representative model for validation data processing.

Results for each classification decision using unbalanced data are provided in detail as supplemental material with mean performance over all three random shuffles of the training, testing, and validation data (Supplemental Tables S2–S5) and in condensed form in Tables 2–3. These tables provide the number of samples used for training, testing, and validation,

and the resulting true positive, true negative, false positive, and false negative classification rates expressed as percentages. In each case, superior models maximize correct classification while minimizing off-diagonal false positives and false negatives. Such confusion matrices are a standard means to evaluate classification performance.

For the unbalanced 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 rarely misclassified the sequence as X4 (e.g., Supplementary Table S4), instead misclassifying it as R5X4. Similarly, when the actual tropism was X4, the neural networks almost never misclassified the sequence as R5. However R5 sequences were regularly misclassified as R5X4 (Supplementary Table S3 and S6). When the actual sequence was dual tropic (R5X4), the neural networks were twice as likely to misclassify the sequence as X4 rather than R5.

For the balanced data, performance on the decision of R5 vs. X4 was again superior, and an improvement upon previous research, as were 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. Off-diagonal errors remained largely non-symmetrical in validation cases however in the case of R5 vs. X4, the off-diagonal error was highest when the ENNs were misclassified R5 as X4, a result directly opposite to that of the unbalanced condition.

Our best models trained for the decision of R5 vs. R5X4 performed at a mean accuracy of 78.1% vs. 81.5% on the validation data, suggesting a slight improvement when using balanced data for this decision. This may be due to the large bias in R5 sequences ( $n=1,223$ ) relative to R5X4 sequences ( $n=241$ ) in the original unbalanced training data. With regards to R5 vs. X4, this bias of sequences in the unbalanced training data was even more pronounced ( $n=1,223$  vs.  $n=95$  respectively). Similarly, models from the balanced dataset outperformed models from the unbalanced dataset on the validation set data (mean of 97.3% vs. 91.6%), again suggesting the importance of working with balanced data for modeling. However, this trend did not hold when working with other tropism decisions. For instance in the decision of R5X4 vs. X4, the original unbalanced data had  $n=241$  and  $n=95$  sequences respectively. Despite the considerably higher accuracy of the balanced approach on training data (mean of 80.0% vs. 71.7%), accuracy on the validation set was nearly equal for the balanced and unbalanced approaches (mean of 76.1% vs. 78.0% respectively). Lastly, for the decision of R5 vs. R5X4 vs. X4, mean accuracy on the validation set was roughly equal between balanced and unbalanced (66.2% and 68.2% respectively). Curiously, in the balanced condition the accuracy of correctly classifying the dual-tropic condition decreased from 59.6% to 46.7%. This model confusion is likely the result of converting a two-class decision into a three-class decision where conflicting information exists in the 15 features being provided to the model.



## Discussion

The accurate assessment of co-receptor usage is of importance for studies regarding viral transmission, evolution, adaptation, reservoirs in specific tissues, and other *IN VITRO* and *IN VIVO* systems. Such assessment may also prove important for proper allocation of treatment regimens designed to help block specific cellular receptors. Biological assays for tropism assessment are currently available and use a recombinant-virus assay to determine HIV-1 coreceptor tropism [58], but these are time-consuming and expensive, and thus encourage continued assessing of tropism using computational methods. Early viral tropism studies identified that X4 HIV-1 can be more highly charged and specifically that two charged amino acids positions (11 and 25) in the envelope V3 domain are particularly useful in discriminating R5 from X4 HIV-1 [41]; however, this is largely understood as an ineffective stand-alone method for predicting HIV-1 tropism. Useful bioinformatics prediction systems for tropism detection of R5 or X4 HIV-1 have been developed and are available via the Internet: webpssm [59, 60] and **geno2pheno** [61]. These algorithms are used as a reasonable substitute for biological assays. WebPSSM has two predictive scoring matrices that were developed using sequences of known coreceptor phenotype, as assayed on indicator cells expressing CD4 and either CCR5 or CXCR4 or sequences of known syncytium-inducing phenotype on the MT2 cell line [60, 59]. Certainly, assessing co-receptor usage using highly controlled cell lines has its advantages, however not all patients and viruses in patients should be expected to behave precisely similar to the behavior of the MT2 cell line. Another advantage of WebPSSM is its ability to predict phenotypes for two of the twelve major HIV-1 subtypes. PSSM Sinsi method is calculated using sequences of known syncytium-inducing phenotype on the MT2 cell line, PSSM x4r5 method is calculated using sequences of known coreceptor phenotype, as assayed on indicator cells expressing exogenous CD4 and either CCR5 or CXCR4, Geno2pheno was developed with publically available phenotyped sequence data and utilizes a support vector machine (SVM) trained for the detection of X4 tropism. Geno2pheno also has the ability to incorporate clinical information into the prediction algorithm and both programs have developed convenient alignment procedures specialized to handle inherent V3 envelope length and sequence variation. PSSM will score sequences as R5 or X4-capable, whereas geno2pheno scores are either “X4” or predicted “non-X4” However, neither program attempts to predict dual-tropic HIV-1.

Utilizing our R5X4 sequence data set, geno2pheno calculated that 22% of the sequences were likely not X4 and WebPSSM predicted that 43% were R5 (Table 4). This indirect or direct misclassification of dual tropic HIV-1 as R5 is important considering a new wave of research focused on targeting macrophage infection [30, 62–64]. Moreover, with the ability to identify dual-tropic HIV-1, therapeutics could be developed to target HIV-1 at a stage that is already associated with low binding efficiency [26] and prior to the generation of X4 variants [65].

The top features used for classification in the ENNs provide insight to the biochemistry associated with tropism and receptor binding. This insight is the basis of a biological understanding of dual-tropism itself. The hydrophobic nature of V3 at the interface of receptor binding has been suggested as an important V3 feature that can be exploited to



develop new broad-spectrum viral entry inhibitors [65, 66]. In this study, 7 of the 15 features used were hydrophobicity scales such as the Janin scale [67], which measures the free energy of transfer from the inside to the outside of globular proteins, and the Tanford scale [68], which measures the contribution of hydrophobic interactions to the stability of protein confirmation. While many different hydrophobic scales exist, the current experiment narrows the field of scales pertinent to the V3 transition among phenotypes. Several structural scales were also found to be useful including the normalized frequency for beta-sheet formation, bulkiness, and average area buried. These features, especially the beta-sheet/beta-turn formation, have a relationship to the structural characteristics of the V3 loop associated with receptor binding [69–73]. Surprisingly, glycosylation patterns, which are associated with the R5 to X4 switch were not identified as useful for decisions across all three tropisms, indicating another structural feature that obscures correct classifications when only considering two outcomes. This was unusual given that glycosylation continues to be considered as an important feature for tropism dynamics in the literature. Similar to hydrophobicity, structural elements of V3 are important for the development of novel therapeutics [66].

It should also be noted that in all cases using balanced data, the performance of the models on testing data showed far more rapid improvement in reduced MSE than in the unbalanced condition. This suggests that balancing of the data improves the ability of simulated evolution to discover useful weight assignments on the neural networks. In most cases when using balanced datasets, ENNs performance on the testing data was still improving at 5,000 generations, placing an artificial cutoff on their performance optimization. However for the sake of comparison to previous runs, we kept the number of generations for unbalanced and balanced datasets. Given the large difference in the number of sequences for each tropism type, balancing the data helps to generate ENNs that have improved generalization on the validation set, and therefore likely improved utility as a biomarker. However, the features shared by the models resulting from both training approaches are themselves very interesting for biologists desiring an improved understanding of the dynamics of tropism.

The current effort was possible due to a substantial increase in the amount of publically available sequence information available for modeling, the number of subtype-B HIV-1 sequences with known co-receptor usage, and a wide range of scales describing amino acid physico-chemical properties. The result was improved accuracy for R5X4 identification, which presents the opportunity to also improve R5 and X4 predictions. For example, a cascade of ENN models will be developed that seek to first identify dual-tropic viruses from R5 or X4, then, once the possibility of dual-tropic viruses have been ruled out, a different ENN classifier will be used to determine if the HIV-1 is R5 or X4 with higher accuracy. HIV-associated studies are more often aimed at identifying subsets of viral sequences associated with microenvironments or clinical courses; for example, the identification of a brain-specific or lymphoma-specific virus has been researched [74, 56, 16, 10]. To support these efforts, we are developing an ENN pipeline that could be used to identify amino acid features in any protein associated with a known output. In future work we also hope to assay the merit of alternative classifier approaches on this same data to determine which representations and approaches provide improved accuracy.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors would like to acknowledge the HIV Database at the Los Alamos National Laboratory, whose careful HIV-1 sequence curating makes studies such as these possible and Sevan Ficici for his thoughts and assistance on dataset balancing. The authors would also like to thank the anonymous reviewers for their comments. This work was supported by National Institutes of Health grants R01 MH100984, R01 NS063897 and UM1CA181255.

## References

- Hall HI, Song R, Rhodes P, Prejean J, An Q, Lee LM, et al. Estimation of HIV incidence in the United States. *JAMA*. 2008; 300(5):520–9. DOI: 10.1001/jama.300.5.520 [PubMed: 18677024]
- Alexaki A, Liu Y, Wigdahl B. Cellular reservoirs of HIV-1 and their role in viral persistence. *Current HIV research*. 2008; 6(5):388–400. [PubMed: 18855649]
- Aquaro S, Calio R, Balestra E, Bagnarelli P, Cenci A, Bertoli A, et al. Clinical implications of HIV dynamics and drug resistance in macrophages. *Journal of biological regulators and homeostatic agents*. 1998; 12(1–2 Suppl):23–7. [PubMed: 9689575]
- Stevenson M. CROI 2014: basic science review. *Topics in antiviral medicine*. 2014; 22(2):574–8. [PubMed: 24901883]
- Barrie KA, Perez EE, Lamers SL, Farmerie WG, Dunn BM, Sleasman JW, et al. Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology*. 1996; 219(2):407–16. DOI: 10.1006/viro.1996.0266 [PubMed: 8638406]
- Gulnik SV, Suvorov LI, Liu B, Yu B, Anderson B, Mitsuya H, et al. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry*. 1995; 34(29):9282–7. [PubMed: 7626598]
- Cortez KJ, Maldarelli F. Clinical management of HIV drug resistance. *Viruses*. 2011; 3(4):347–78. DOI: 10.3390/v3040347 [PubMed: 21994737]
- Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*. 1995; 267(5197):483–9. [PubMed: 7824947]
- 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. *Journal of virology*. 2005; 79(17):11343–52. doi: 79/17/11343 [pii] 10.1128/JVI.79.17.11343-11352.2005. [PubMed: 16103186]
- Salemi M, Lamers SL, Huysentruyt LC, Galligan D, Gray RR, Morris A, et al. Distinct patterns of HIV-1 evolution within metastatic tissues in patients with non-Hodgkins lymphoma. *PloS one*. 2009; 4(12):e8153. doi: 10.1371/journal.pone.0008153 [PubMed: 19997510]
- Fitch KV, Srinivasa S, Abbara S, Burdo TH, Williams KC, Eneh P, et al. Noncalcified coronary atherosclerotic plaque and immune activation in HIV-infected women. *The Journal of infectious diseases*. 2013; 208(11):1737–46. DOI: 10.1093/infdis/jit508 [PubMed: 24041790]
- Bernstein LE, Berry J, Kim S, Canavan B, Grinspoon SK. Effects of etanercept in patients with the metabolic syndrome. *Archives of internal medicine*. 2006; 166(8):902–8. DOI: 10.1001/archinte.166.8.902 [PubMed: 16636217]
- Estrada V, Portilla J. Dyslipidemia related to antiretroviral therapy. *AIDS reviews*. 2011; 13(1):49–56. [PubMed: 21412389]
- Anthony IC, Ramage SN, Carnie FW, Simmonds P, Bell JE. Influence of HAART on HIV-related CNS disease and neuroinflammation. *Journal of neuropathology and experimental neurology*. 2005; 64(6):529–36. [PubMed: 15977645]
- Ng VL, McGrath MS. The immunology of AIDS-associated lymphomas. *Immunological reviews*. 1998; 162:293–8. [PubMed: 9602372]

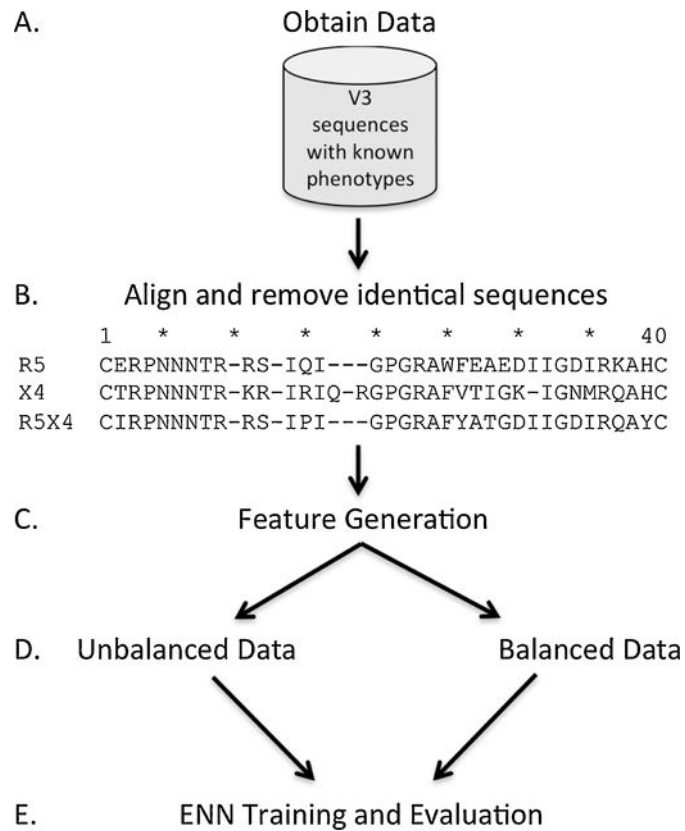
16. Lamers SL, Salemi M, Galligan DC, Morris A, Gray R, Fogel G, et al. Human immunodeficiency virus-1 evolutionary patterns associated with pathogenic processes in the brain. *J Neurovirol*. 2010; 16(3):230–41. DOI: 10.3109/13550281003735709 [PubMed: 20367240]
17. <http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5710a2.htm>.
18. Goodenow MM, Collman RG. HIV-1 coreceptor preference is distinct from target cell tropism: a dual-parameter nomenclature to define viral phenotypes. *Journal of leukocyte biology*. 2006; 80(5):965–72. DOI: 10.1189/jlb.0306148 [PubMed: 16923919]
19. Gorry PR, Churchill M, Crowe SM, Cunningham AL, Gabuzda D. Pathogenesis of macrophage tropic HIV-1. *Current HIV research*. 2005; 3(1):53–60. [PubMed: 15638723]
20. MacArthur RD, Novak RM. Reviews of anti-infective agents: maraviroc: the first of a new class of antiretroviral agents. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2008; 47(2):236–41. DOI: 10.1086/589289 [PubMed: 18532888]
21. Dorr P, Westby M, Dobbs S, Griffin P, Irvine B, Macartney M, et al. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrobial agents and chemotherapy*. 2005; 49(11):4721–32. DOI: 10.1128/AAC.49.11.4721-4732.2005 [PubMed: 16251317]
22. Starr-Spires LD, Collman RG. HIV-1 entry and entry inhibitors as therapeutic agents. *Clinics in laboratory medicine*. 2002; 22(3):681–701. [PubMed: 12244592]
23. Gorry PR, Ancuta P. Coreceptors and HIV-1 pathogenesis. *Current HIV/AIDS reports*. 2011; 8(1): 45–53. DOI: 10.1007/s11904-010-0069-x [PubMed: 21188555]
24. Roundtable for the Development of Drugs and Vaccines against AIDS (Institute of Medicine). Weiss, R.; Mazade, L. Surrogate endpoints in evaluating the effectiveness of drugs against HIV infection and AIDS : September 11–12, 1989: conference summary. Publication, vol no IOM-90-001. Washington, D.C.: National Academy Press; 1990.
25. Moyle GJ, Wildfire A, Mandalia S, Mayer H, Goodrich J, Whitcomb J, et al. Epidemiology and predictive factors for chemokine receptor use in HIV-1 infection. *The Journal of infectious diseases*. 2005; 191(6):866–72. DOI: 10.1086/428096 [PubMed: 15717260]
26. Tasca S, Ho SH, Cheng-Mayer C. R5X4 viruses are evolutionary, functional, and antigenic intermediates in the pathway of a simian-human immunodeficiency virus coreceptor switch. *Journal of virology*. 2008; 82(14):7089–99. DOI: 10.1128/JVI.00570-08 [PubMed: 18480460]
27. Loftin LM, Kienzle M, Yi Y, Collman RG. R5X4 HIV-1 coreceptor use in primary target cells: implications for coreceptor entry blocking strategies. *Journal of translational medicine*. 2011; 9(Suppl 1):S3.doi: 10.1186/1479-5876-9-S1-S3 [PubMed: 21284902]
28. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, et al. HIV-1 nomenclature proposal. *Science*. 2000; 288(5463):55–6. [PubMed: 10766634]
29. Mild M, Gray RR, Kvist A, Lemey P, Goodenow MM, Fenyó EM, et al. High inpatient HIV-1 evolutionary rate is associated with CCR5-to-CXCR4 coreceptor switch. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2013; 19:369–77. DOI: 10.1016/j.meegid.2013.05.004 [PubMed: 23672855]
30. Campbell JH, Hearps AC, Martin GE, Williams KC, Crowe SM. The importance of monocytes and macrophages in HIV pathogenesis, treatment, and cure. *AIDS*. 2014; doi: 10.1097/QAD.0000000000000408
31. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annual review of immunology*. 1999; 17:657–700. DOI: 10.1146/annurev.immunol.17.1.657
32. Broder CC, Collman RG. Chemokine receptors and HIV. *Journal of leukocyte biology*. 1997; 62(1):20–9. [PubMed: 9225988]
33. Murakami T, Yamamoto N. Roles of chemokines and chemokine receptors in HIV-1 infection. *International journal of hematology*. 2000; 72(4):412–7. [PubMed: 11197206]
34. 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. [PubMed: 16848274]
35. Clapham PR, McKnight A. Cell surface receptors, virus entry and tropism of primate lentiviruses. *The Journal of general virology*. 2002; 83(Pt 8):1809–29. [PubMed: 12124446]

36. 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. *AIDS research and human retroviruses*. 2004; 20(1):111–26. DOI: 10.1089/088922204322749567 [PubMed: 15000703]
37. Weiss RA. HIV receptors and cellular tropism. *IUBMB life*. 2002; 53(4–5):201–5. DOI: 10.1080/15216540212652 [PubMed: 12120995]
38. Doms RW, Trono D. The plasma membrane as a combat zone in the HIV battlefield. *Genes & development*. 2000; 14(21):2677–88. [PubMed: 11069884]
39. Katzenstein TL. Molecular biological assessment methods and understanding the course of the HIV infection. *APMIS Supplementum*. 2003; 114:1–37. [PubMed: 14626050]
40. Resch W, Hoffman N, Swanstrom R. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. *Virology*. 2001; 288(1):51–62. DOI: 10.1006/viro.2001.1087 [PubMed: 11543657]
41. 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. *Journal of virology*. 2002; 76(8):3852–64. [PubMed: 11907225]
42. Ioannidis JP, Trikalinos TA, Law M, Carr A. HIV lipodystrophy case definition using artificial neural network modelling. *Antiviral therapy*. 2003; 8(5):435–41. [PubMed: 14640391]
43. Wang D, Larder B. Enhanced prediction of lopinavir resistance from genotype by use of artificial neural networks. *The Journal of infectious diseases*. 2003; 188(5):653–60. DOI: 10.1086/377453 [PubMed: 12934180]
44. Brumme ZL, Dong WW, Yip B, Wynhoven B, Hoffman NG, Swanstrom R, et al. Clinical and immunological impact of HIV envelope V3 sequence variation after starting initial triple antiretroviral therapy. *AIDS*. 2004; 18(4):F1–9. [PubMed: 15090786]
45. Milich L, Margolin B, Swanstrom R. V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. *Journal of virology*. 1993; 67(9):5623–34. [PubMed: 8350415]
46. Fogel GB. Computational intelligence approaches for pattern discovery in biological systems. *Briefings in bioinformatics*. 2008; 9(4):307–16. DOI: 10.1093/bib/bbn021 [PubMed: 18460474]
47. Fogel DB, Fogel LJ, Porto VW. Evolving Neural Networks. *Biological Cybernetics*. 1990; 63(6):487–93.
48. Porto VW, Fogel DB, Fogel LB. Alternative Neural Network Training Methods. *IEEE Expert*. 1995; 10(3):16–22.
49. Yao X. Evolving Artificial Neural Networks. *Proc IEEE*. 1999; 87(9):1423–47.
50. Kohl N, Mikkulainen R. An Integrated Neuroevolutionary Approach to Reactive Control and High-level Strategy. *IEEE Transactions on Evolutionary Computation*. 2011; 16(4):472–88.
51. Hecht D, Cheung M, Fogel GB. QSAR using evolved neural networks for the inhibition of mutant PfDHFR by pyrimethamine derivatives. *Bio Systems*. 2008; 92(1):10–5. DOI: 10.1016/j.biosystems.2007.10.005
52. Hecht D, Fogel GB. High-throughput ligand screening via preclustering and evolved neural networks. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*. 2007; 4(3):476–84. DOI: 10.1109/tcbb.2007.1038
53. Lamers SL, Salemi M, McGrath MS, Fogel GB. Prediction of R5, X4, and R5X4 HIV-1 coreceptor usage with evolved neural networks. *IEEE/ACM Trans Comput Biol Bioinform*. 2008; 5(2):291–300. DOI: 10.1109/TCBB.2007.1074 [PubMed: 18451438]
54. Sierra S, Kaiser R, Thielen A, Lengauer T. Genotypic coreceptor analysis. *European journal of medical research*. 2007; 12(9):453–62. [PubMed: 17933727]
55. Poveda E, Briz V, Roulet V, Del Mar Gonzalez M, Faudon JL, Skrabal K, et al. Correlation between a phenotypic assay and three bioinformatic tools for determining HIV co-receptor use. *AIDS*. 2007; 21(11):1487–90. DOI: 10.1097/QAD.0b013e32826fb741 [PubMed: 17589199]
56. 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. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2011; 11(1):31–7. DOI: 10.1016/j.meegid.2010.10.016 [PubMed: 21055482]

57. Blackard JT. HIV compartmentalization: a review on a clinically important phenomenon. *Current HIV research*. 2012; 10(2):133–42. [PubMed: 22329519]
58. Whitcomb JM, Huang W, Fransen S, Limoli K, Toma J, Wrin T, et al. Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. *Antimicrobial agents and chemotherapy*. 2007; 51(2):566–75. DOI: 10.1128/AAC.00853-06 [PubMed: 17116663]
59. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *Journal of virology*. 2003; 77(24):13376–88. [PubMed: 14645592]
60. 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. *Journal of virology*. 2006; 80(10):4698–704. DOI: 10.1128/JVI.80.10.4698-4704.2006 [PubMed: 16641263]
61. Beerenwinkel N, Daumer M, Oette M, Korn K, Hoffmann D, Kaiser R, et al. Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic acids research*. 2003; 31(13):3850–5. [PubMed: 12824435]
62. Walker, J.; Burdo, T.; Miller, A.; Misgin, K.; Sulciner, M.; McGrath, M., et al. Macrophages in Hearts of SIV+ Rhesus Macaques with Cardiac Disease Are Decreased Using PA300. *Conference on Retroviruses and Opportunistic Infection*; Atlanta, GA. 2013.
63. Rappaport J. Editorial: The Monocyte/Macrophage in the Pathogenesis of AIDS: The Next Frontier for Therapeutic Intervention in the CNS and Beyond: Part I. *Current HIV research*. 2014; 12(2): 75–6. [PubMed: 25088181]
64. Flegler AJ, Cianci GC, Hope TJ. CCR5 conformations are dynamic and modulated by localization, trafficking and G protein association. *PloS one*. 2014; 9(2):e89056.doi: 10.1371/journal.pone.0089056 [PubMed: 24586501]
65. Badani H, Garry RF, Wimley WC. Peptide entry inhibitors of enveloped viruses: the importance of interfacial hydrophobicity. *Biochimica et biophysica acta*. 2014; 1838(9):2180–97. DOI: 10.1016/j.bbamem.2014.04.015 [PubMed: 24780375]
66. Cosset FL, Lavillette D. Cell entry of enveloped viruses. *Advances in genetics*. 2011; 73:121–83. DOI: 10.1016/B978-0-12-380860-8.00004-5 [PubMed: 21310296]
67. Janin J. Surface and inside volumes in globular proteins. *Nature*. 1979; 277(5696):491–2. [PubMed: 763335]
68. Tanford C. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J Am Chem Soc*. 1962; 84:4240–4274.
69. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*. 1998; 393(6686):648–59. DOI: 10.1038/31405 [PubMed: 9641677]
70. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, et al. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature*. 1998; 393(6686):705–11. DOI: 10.1038/31514 [PubMed: 9641684]
71. Rizzuto CD, Wyatt R, Hernandez-Ramos N, Sun Y, Kwong PD, Hendrickson WA, et al. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science*. 1998; 280(5371):1949–53. [PubMed: 9632396]
72. Wyatt R, Sodroski J. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science*. 1998; 280(5371):1884–8. [PubMed: 9632381]
73. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *The Journal of biological chemistry*. 1990; 265(18):10373–82. [PubMed: 2355006]
74. Lamers SL, Poon AF, McGrath MS. HIV-1 nef protein structures associated with brain infection and dementia pathogenesis. *PloS one*. 2011; 6(2):e16659.doi: 10.1371/journal.pone.0016659 [PubMed: 21347424]

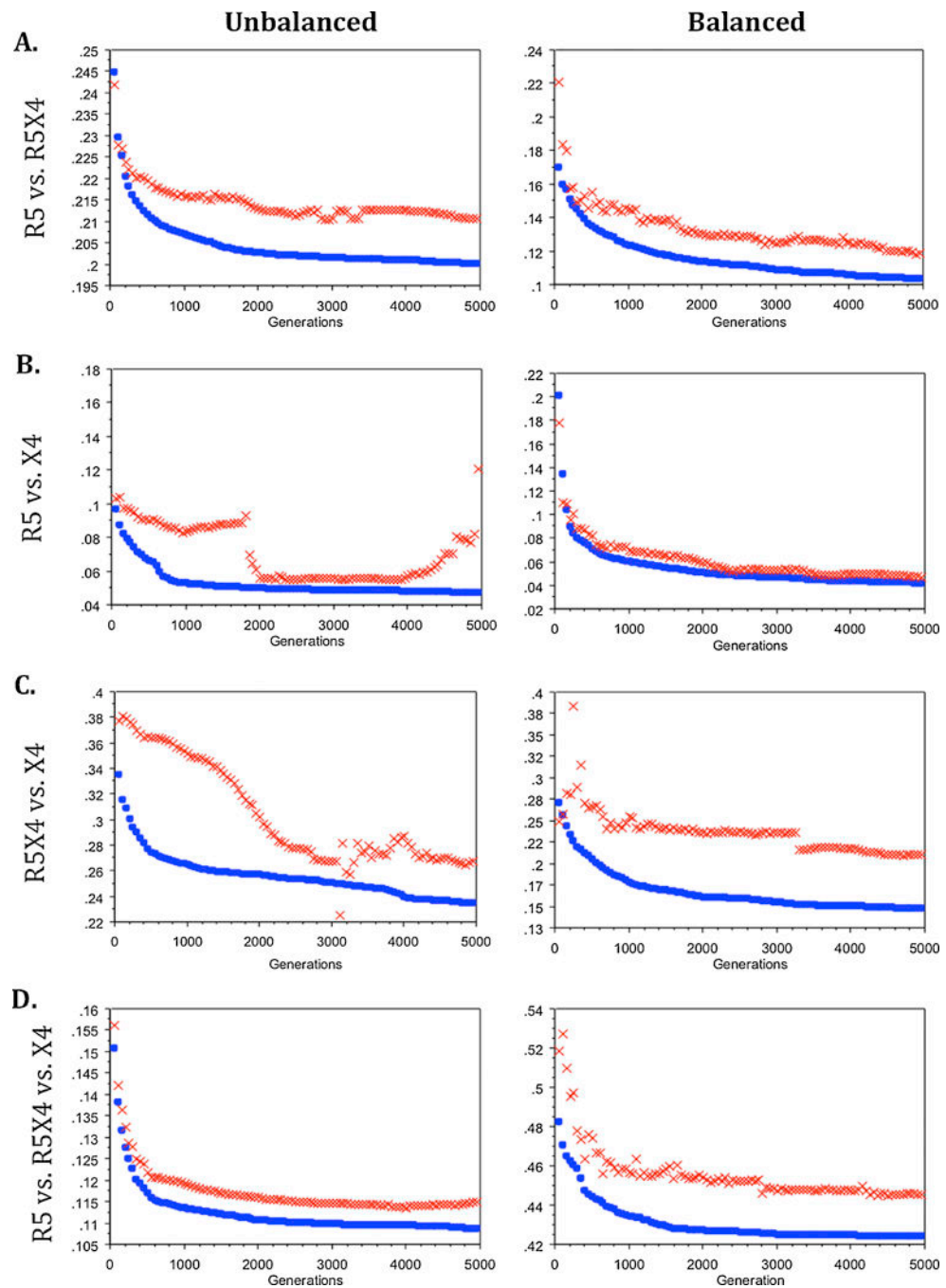


75. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*. 2011; 28(10):2731–9. DOI: 10.1093/molbev/msr121 [PubMed: 21546353]
76. Briggs DR, Tuttle DL, Sleasman JW, Goodenow MM. Envelope V3 amino acid sequence predicts HIV-1 phenotype (co-receptor usage and tropism for macrophages). *AIDS*. 2000; 14(18):2937–9. [PubMed: 11153675]
77. Wilkin TJ, Gulick RM. CCR5 antagonism in HIV infection: current concepts and future opportunities. *Annual review of medicine*. 2012; 63:81–93. DOI: 10.1146/annurev-med-052010-145454
78. Delgado E, Fernandez-Garcia A, Vega Y, Cuevas T, Pinilla M, Garcia V, et al. Evaluation of genotypic tropism prediction tests compared with in vitro co-receptor usage in HIV-1 primary isolates of diverse subtypes. *The Journal of antimicrobial chemotherapy*. 2012; 67(1):25–31. DOI: 10.1093/jac/dkr438 [PubMed: 22010208]
79. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods in molecular biology*. 1999; 112:531–52. [PubMed: 10027275]
80. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. Hydrophobicity of amino acid residues in globular proteins. *Science*. 1985; 229(4716):834–8. [PubMed: 4023714]
81. Zimmerman JM, Eliezer N, Simha R. The characterization of amino acid sequences in proteins by statistical methods. *Journal of theoretical biology*. 1968; 21(2):170–201. [PubMed: 5700434]
82. Darby, NJ.; Creighton, TE. *Protein structure: In focus*. Oxford University Press; 1993.
83. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in enzymology and related areas of molecular biology*. 1978; 47:45–148. [PubMed: 364941]
84. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, MR.; Appel, RD., et al. *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press; 2005.
85. Chothis C. The Nature of Accesible and Buried Surfaces in Proteins. *J Mol Biol*. 1976; 105:1–14. [PubMed: 994183]
86. Abraham DJ, Leo AJ. Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins*. 1987; 2(2):130–52. DOI: 10.1002/prot.340020207 [PubMed: 3447171]
87. Wolfenden R, Andersson L, Cullis PM, Southgate CC. Affinities of amino acid side chains for solvent water. *Biochemistry*. 1981; 20(4):849–55. [PubMed: 7213619]
88. Eisenberg D. Three-dimensional structure of membrane and surface proteins. *Annual review of biochemistry*. 1984; 53:595–623. DOI: 10.1146/annurev.bi.53.070184.003115
89. Cowan R, Whittaker RG. Hydrophobicity indices at ph 7.5 determined by HPLC. *Peptide Research*. 1990; 3:75–80. [PubMed: 2134053]
90. Jones DD. Amino acid properties and side-chain orientation in proteins: a cross correlation approach. *Journal of theoretical biology*. 1975; 50(1):167–83. [PubMed: 1127956]



**Fig. 1.** Steps Involved With Neural Network Development. A) 4,194 Envelope V3 sequences with known phenotypes were downloaded from the Los Alamos HIV database. B) Sequences were trimmed to the V3 loop domain and the amino acid translation was aligned within a 40 base pair window. The alignment was reduced to include only non-identical sequences. Shown is a representative alignment that contains one sequence for each tropism; asterisks indicate every fifth position in the alignment and dashes represent gapped positions. C) 73 features were calculated for each position in the alignment. D) The naturally unbalanced data was also represented as a balanced dataset with equal numbers of sequences for each tropism. E) Data was presented to a neural network as training, testing, and validation in triplicate using evolution to optimize the neural networks to minimize MSE on the decision of tropism class.





**Fig. 2.** Convergence during model optimization on training and testing data over three random shuffles of the data scored relative to minimized mean square error (MSE). Left Panels = Unbalanced Data, and Right Panels = Balanced Data. A) R5 vs. R5X4; B) R5 vs. X4; C) R5X4 vs. X4; D) R5 vs. R5X4 vs. X4.

**Table 1**

The fifteen features identified as having the highest average linear correlation to tropism across all four tropism decisions.

<b>Class</b>	<b>Features</b>	<b>Reference</b>
Size, Shape, Structure	Avg. area buried	[80]
	Bulkiness	[81]
	Volume	[82]
	Beta turn Chou and Fasman	[83]
Polarity Scales	Charge at Position 12	[45]
	Charge scale for all amino acids	[84]
	Overall Charge (K and R = +1, D and E = -1)	[84]
Hydrophobicity Scales	Chothia	[85]
	Abraham and Leo	[86]
	Wolfenden et al.	[87]
	Eisenberg et al.	[88]
	Janin	[67]
	Tanford	[68]
Other	Cowan and Whittaker	[89]
	Refractivity	[90]

**Table 2**

Mean validation performance for 2-tiered tropism comparisons using unbalanced and balanced training methods.

[0,1-4]Unbalanced		[0,5-8]Balanced			
[0,1-2]Comparison		[0,3-4]Actual		[0,7-8]Actual	
<b>R5 vs. R5X4 (n = 400)</b>	<b>R5 (n=353)</b>	<b>R5X4 (n=47)</b>	<b>R5 vs. R5X4 (n=400)</b>	<b>R5 (n=345)</b>	<b>R5X4 (n=55)</b>
Predicted	R5 283/353=80.2%	11/47=24.1%	Predicted	R5 280/345=81.3%	10/55=18.2%
	R5X4 70/353=19.8%	36/47=75.9%		R5X4 65/345=18.7%	45/55=81.8%
<b>R5 vs. X4 (n=365)</b>	<b>R5 (n=343)</b>	<b>X4 (n=22)</b>	<b>R5 vs. X4 (n=365)</b>	<b>R5 (n=345)</b>	<b>X4 (n=20)</b>
Predicted	R5 337/343=98.3%	3/22=15.2%	Predicted	R5 323/345=94.6%	0/20=0%
	X4 6/343=1.7%	19/22=84.8%		X4 19/345=5.4%	20/20=100%
<b>R5X4 vs. X4 (n=74)</b>	<b>R5X4 (n=52)</b>	<b>X4 (n=22)</b>	<b>RX45 vs. X4 (n=75)</b>	<b>R5X4 (n=55)</b>	<b>X4 (n=20)</b>
Predicted	R5X4 43/52=82.1%	5/22=24.2%	Predicted	R5X4 43/55=78.8%	5/20=26.7%
	X4 9/52=17.9%	17/22=75.8%		X4 12/55=21.1%	15/20=73.3%

**Table 3**  
 Mean validation performance for 3-tiered tropism comparisons using unbalanced and balanced training methods.

[0,1-5]Unbalanced		[0,6-10]Balanced	
[0,1-2]Mean Validation Performance (n=419)	[0,3-5]Actual	[0,6-7]Mean Validation Performance (n=420)	[0,8-10]Actual
Predicted	R5 (n=345) 244/345=70.70%	X4 (n=22) 1/22=6.10%	R5 (n=345) 265/345=76.80%
	R5X4 (n=52) 11/52=20.50%	X4 (n=22) 4/22=19.70%	R5X4 (n=55) 13/55=24.20%
	3/345=1.00%	16/22=74.20%	26/55=46.70%
			6/345=1.70%
			16/55=29.10%
			15/20=75.00%

**Table 4**

Prediction of coreceptor usage for dual tropic viruses using web-based prediction methods. Geno2Pheno FPR (false-positive rate) is the probability of classifying and R5-virus falsely as X4. Geno2Pheno has several FPR cutoff values to choose from; here we used the Recommendations from the European Consensus Group on Clinical Management on HIV-1 Tropism Testing FPR of 10%.

Prediction	PSSM (Sinsi)	PSSM (x4r5)	Geno2Pheno (10% FPR)
X4	62%	57%	78%
R5 <sup>4</sup>	38%	43%	22%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript