

# Classifying Modes of Adaptation in Ancient Human DNA with Deep Learning



BRENDAN AERIA<sup>1</sup>, Mariana Harris<sup>2,3</sup>, Maya Weissman<sup>3</sup>, Nandita Garud<sup>3,4</sup>

<sup>1</sup> BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA  
<sup>2</sup> Department of Computational Medicine, UCLA  
<sup>3</sup> Department of Ecology and Evolutionary Biology, UCLA  
<sup>4</sup> Department of Human Genetics, UCLA



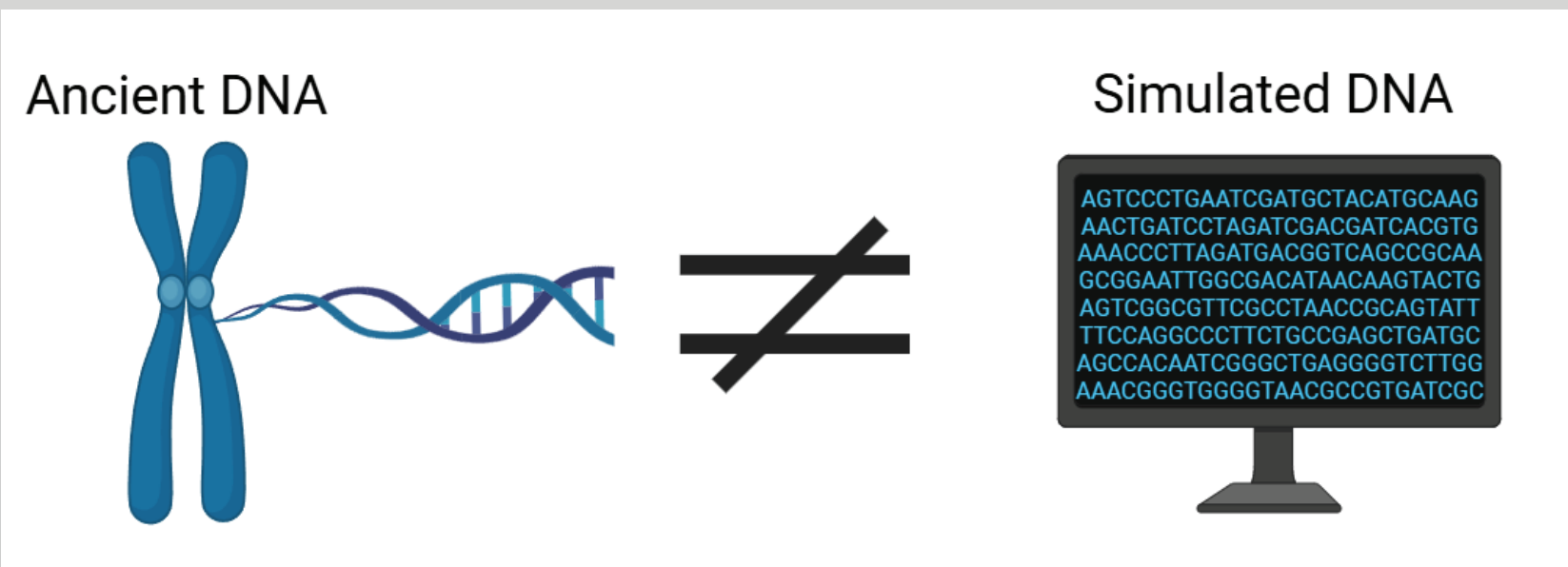
## Introduction

- **Selective sweeps:** the process by which beneficial mutations rise to high frequency, driving adaptation
  - **Hard Sweeps:** The beneficial mutation arises from a single new source.
  - **Soft Sweeps:** The beneficial mutation was already present on multiple genetic backgrounds.
- Distinguishing between hard and soft sweeps in ancient DNA (aDNA) is challenging due to high rates of missing data and complex population history.
- Current deep learning methods often overlook a powerful signal: the distinct evolutionary patterns of **synonymous** (silent) versus **non-synonymous** (protein-changing) mutations.

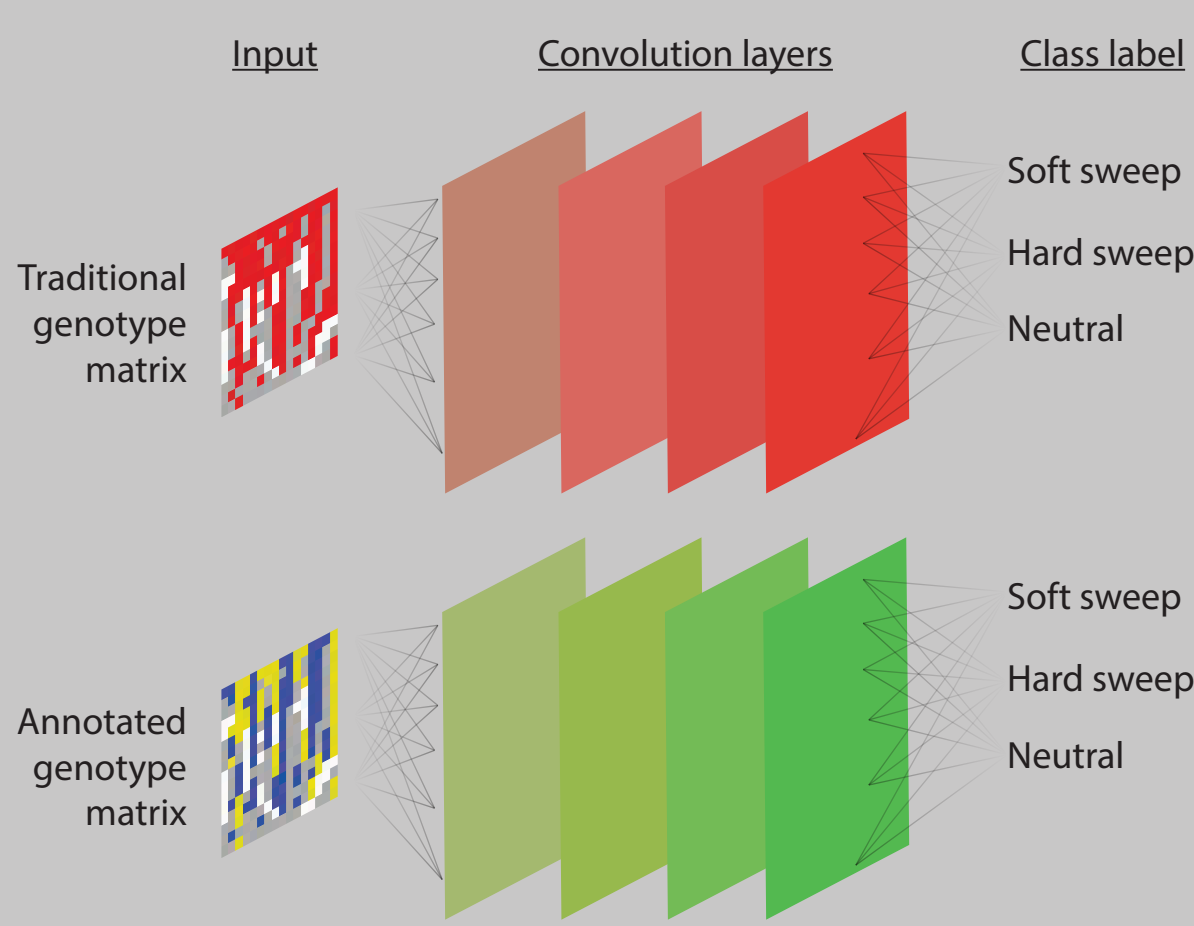
Our Goal: To construct a neural network that accurately classifies selective sweeps

## Hypothesis

- 1) Learning the similarities between simulated and ancient DNA will improve model performance.

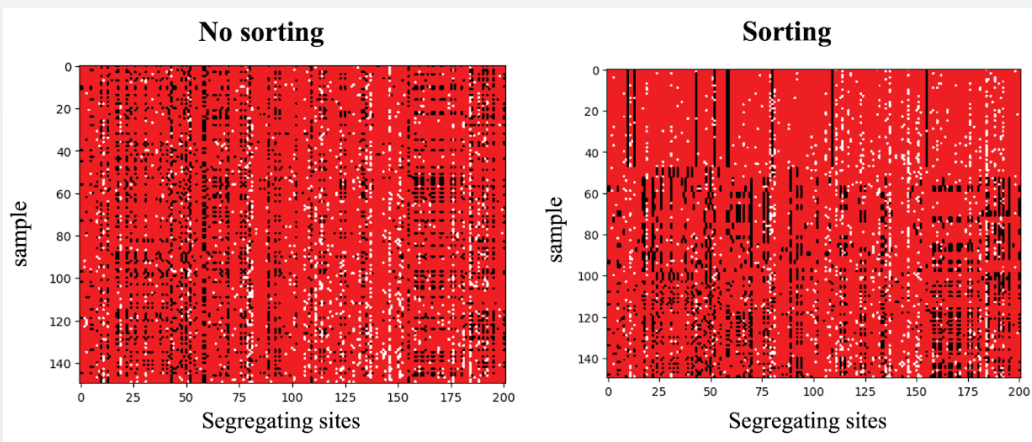
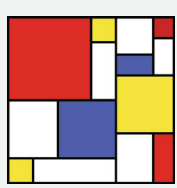


- 2) A genotype matrix with functional annotations will yield a more accurate selective sweep classification model.

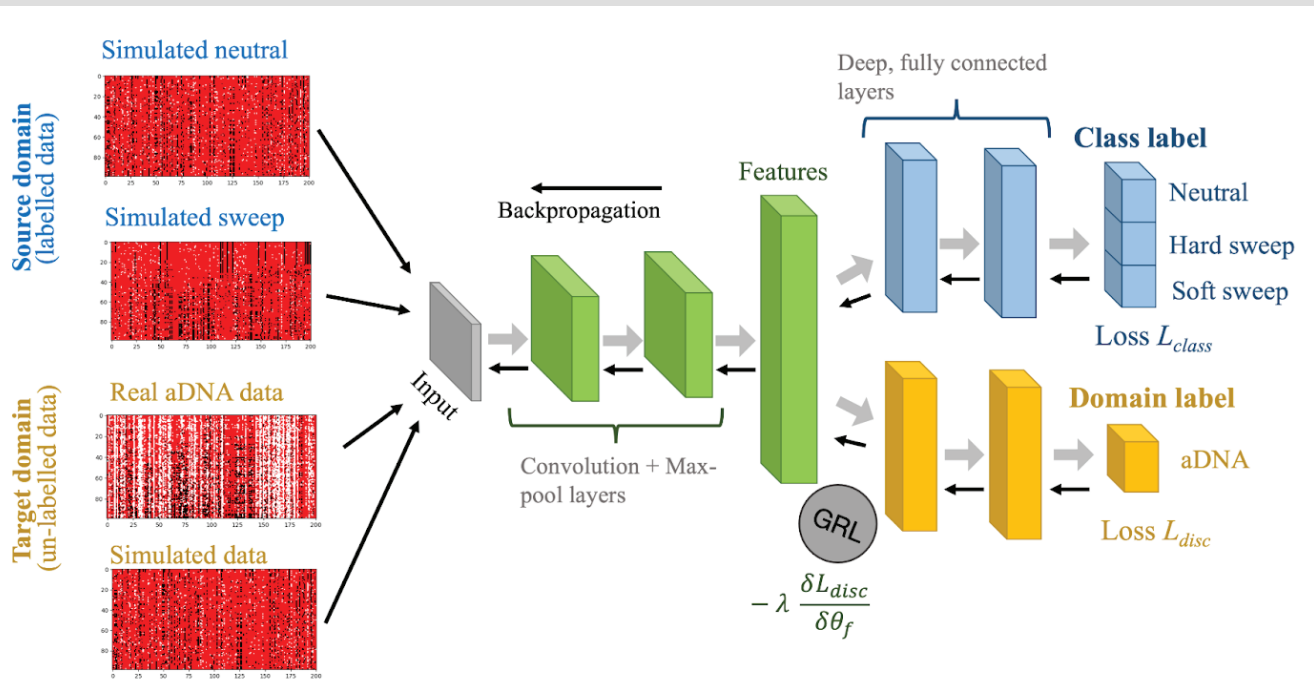


## Methods

- 1). Simulate haplotypes using SLiM
- 2). Annotate the function change of mutations
- 3). Align and sort haplotypes
- 4). Train a CNN using aDNA as the target domain and the simulated DNA as the source domain
- 5). Evaluate the model performance by plotting receiver operating characteristic curves

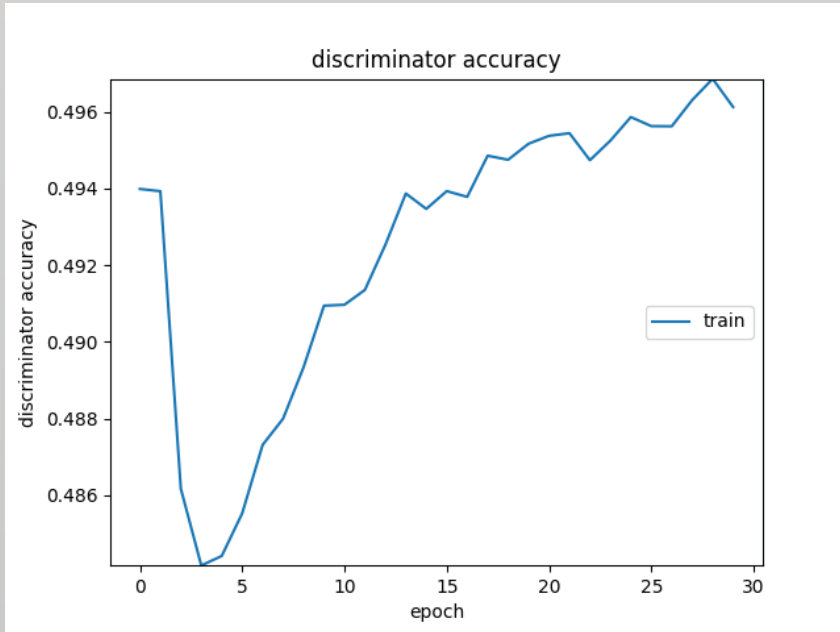


### Domain Adaptive Neural Network (DANN) Architecture Diagram



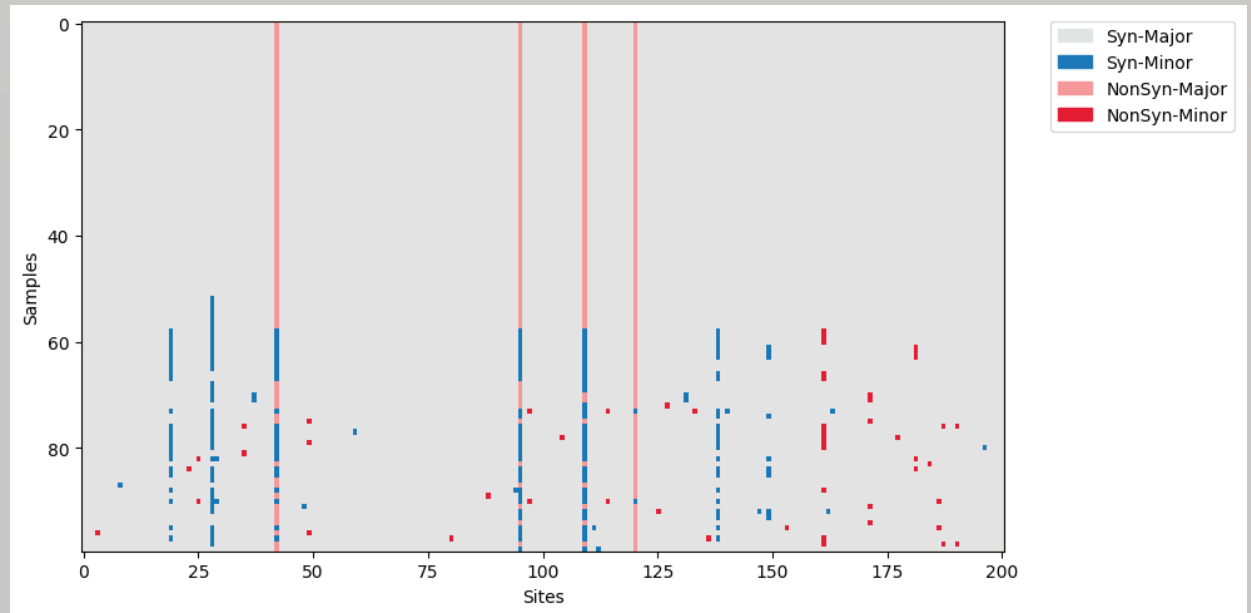
By training a DANN, we test to see if learning the similarities between simulated and ancient DNA will improve model performance.

### Discriminator Accuracy Demonstrates Gradient Reversal

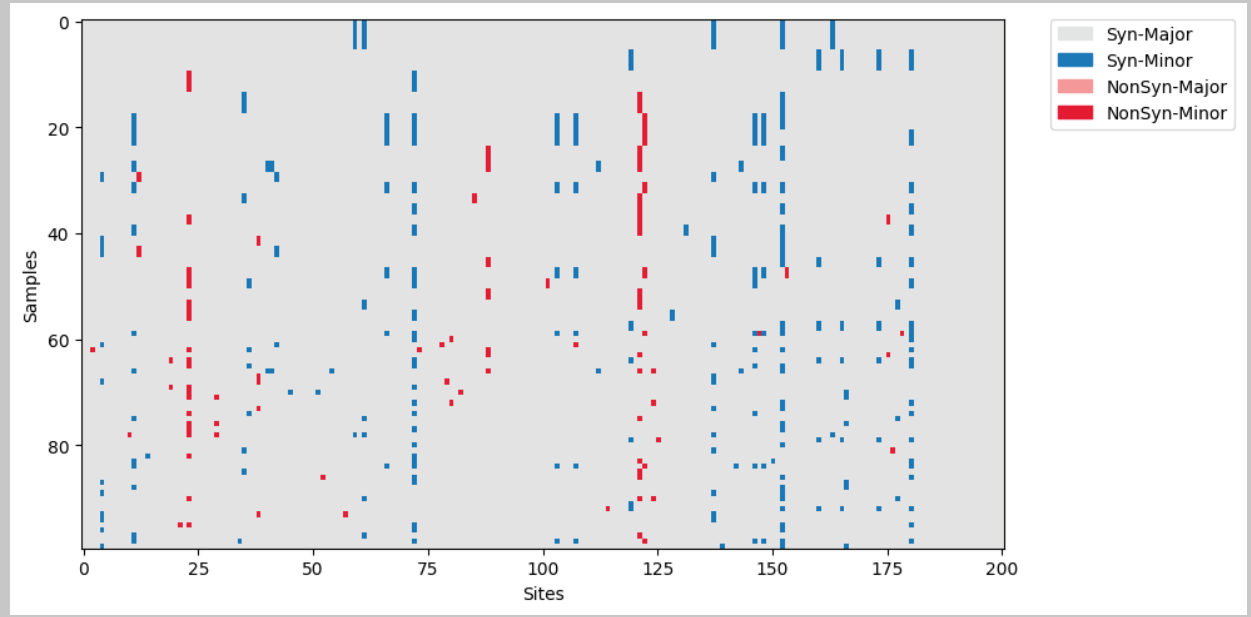


After epoch 4 in training the DANN, the Gradient Reversal Layer is activated, indicated by the rapid increase in discriminator accuracy.

### Functional annotation matrix of a sweep



### Functional annotation matrix of neutrality

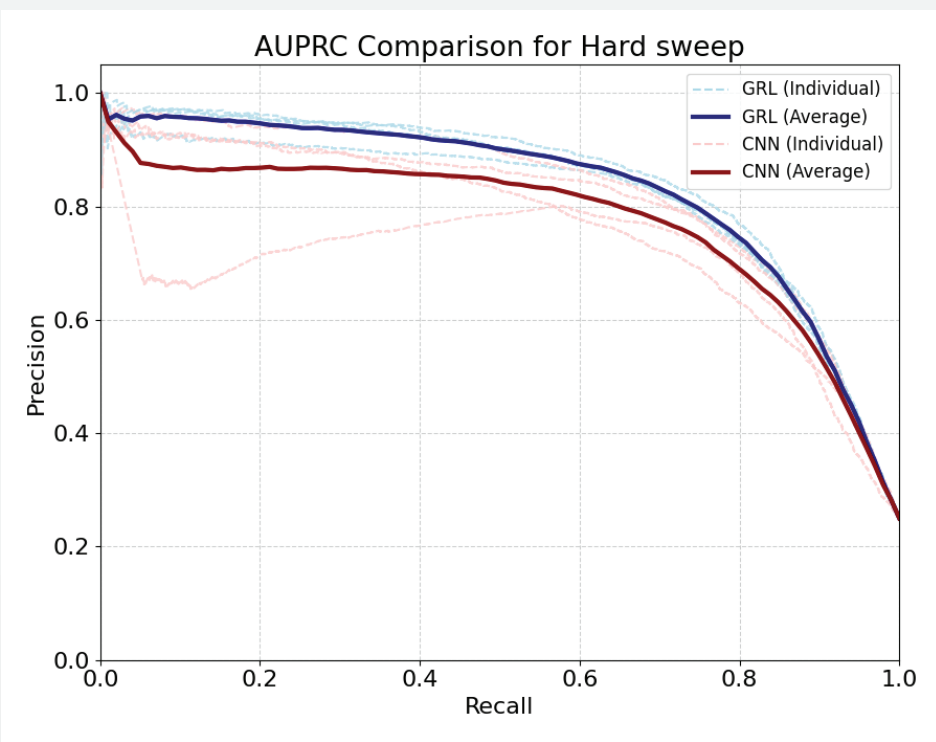


The long stretch of identical haplotypes in samples 1 to 50 occurs as a result of a beneficial mutation sweeping across the population.

Contrary to the matrix with a sweep occurring, there is no clear repetition of the same haplotype. This indicates that no selection has swept across the population.

## Results

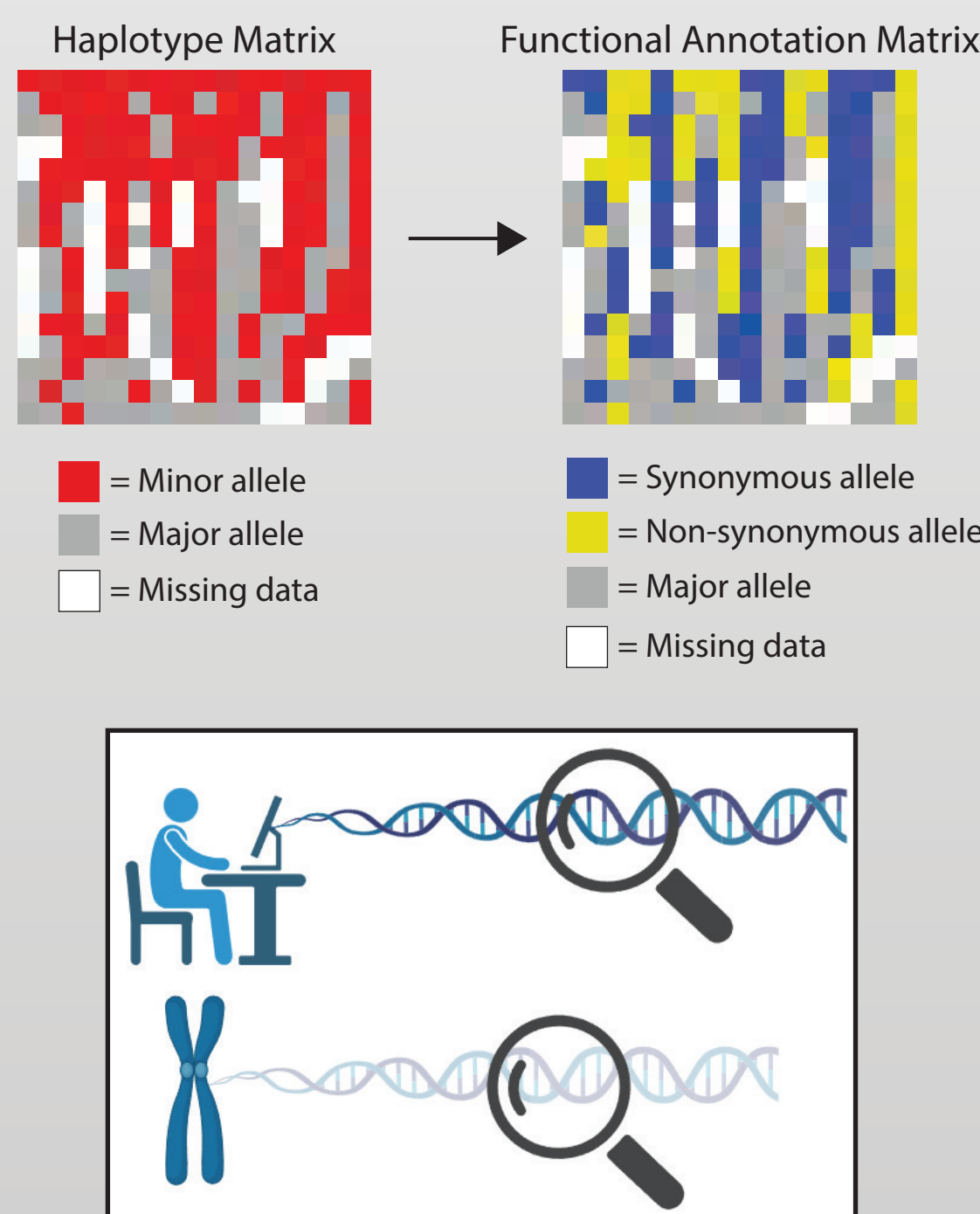
The DANN consistently outperforms the standard CNN when it comes to classifying hard sweeps.



Similar results were found for classifying neutrality. However, detecting soft sweeps was nearly at guess proficiency for both CNN and DANN.

## Next Steps

- Annotate more halpotypes with synonymous and non-synonymous alleles
- Train our existing neural nets on this annotated data
- Use these newly trained models to detect sweeps in ancient human populations



## Conclusions

- A gradient reversal layer is a powerful technique that allows for generalization between source and target domain.
- A convolutional neural network struggles to classify soft sweeps in combination with high missing data rate.
- Using functional information about a mutation is a promising direction for future sweep classification.

## References:

