# Summary

* SVM performed slightly better than glmnet (for both training and testing error)
* Final model used 11 CpG predictors and was built with glmnet with a AUC of 0.981 and 0.977+-0.024 for training and testing error respectively (α = 0.75, λ = 0.25).
* The classifier predicted all of the unlabeled test set to Caucasian, which we doubt is the true case.
* We suspect the test set is too ‘different’ from the training data set for the classifier to perform accurately on the test set.

# Future Directions

* Normalizing and QCing the test and training datasets together may be necessary for DNA methylation classifiers to perform well
* Using MDS ethnicity coordinates from population stratification meta-analyses may provide ‘labels’ to assess classifier performance or improve model building. (self-reported ethnicity can be unreliable)