**Methods**

The second part of the challenge required the prediction of the emvarHit and LogSkew.comb parameters.

For predicting the probability of emvarHit, we used the combined k-mers from the reference and altered sequences and counted them as one feature. We first count all possible2-mers in the reference and alternate sequences and then for the same position in the two sequences we combine the couple of 2-mers into a 4-mer. This new feature space is then fed to the SVM binary classifier to estimate the emvar probability.

Fig-1 shows the AUCs when doing the 5 fold cross-validation on the training data provided in the challenge.



**Fig 1: ROC for emvar hit prediction**



Predicted logfold change

Actual logfold change

**Fig 2: Predicted vs actual log fold change for the reference sequences**