We fitted a logistic regression to the first three principal components. We then applied leave-one-out cross-validation to assess the ability of the principal components to predict patient/control. We found the model predicted the disease well, with 3/20 prediction errors (two false positives and one false negative)

Figure .. shows the predicted probabilities for all 20 subjects.

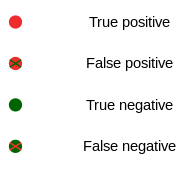
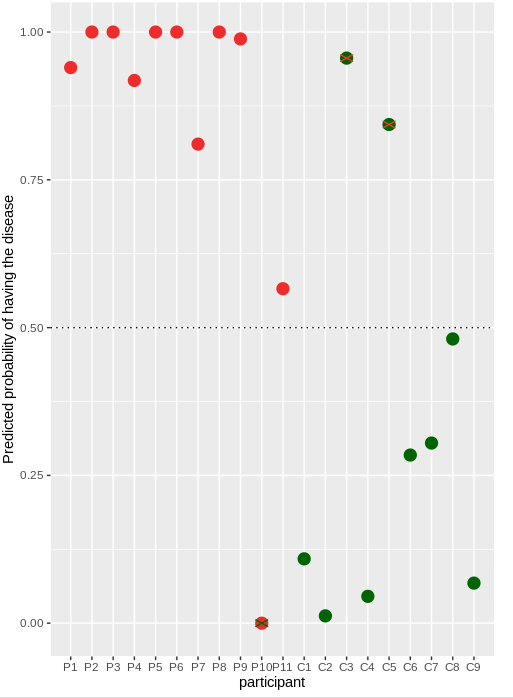


Figure ..: Results of Principal Component logistic regression after leave-one-out cross-validation. One patient (P10) and two controls (C3 and C5) were misclassified.

We see that P10,C3 and C5 are misclassified, as would be expected from the results of the PCA (figure.. ) A permutation test shows that the probability of achieving a prediction error of 3/20 or more extreme, is 0.04, so we may say it is unlikely that the observed predictive capacity of the principal components has been caused by chance.

We fitted 2,3 and 4-means clustering model on the Pareto-scaled data, and on the first two and three principal components, but this did not provide a satisfactory clustering, in the sense that this did not separate patients from controls.

There are 147 mitochondrian proteins among the 2970 under investigation. We tested the hypothesis that these are specifically implied in predicting the disease with the above regression model. This turned out not to be the case: Simulation shows that a random sample of 147 other proteins does no worse on average than the 147 mitochondrian proteins.