

# Ulcerative Colitis Metabolomics sPLSDA

Isabel Prado-Tucker

2023-07-12

Data source: (The Inflammatory Bowel Disease Multi'omics Database)[<https://ibdmdb.org/>]

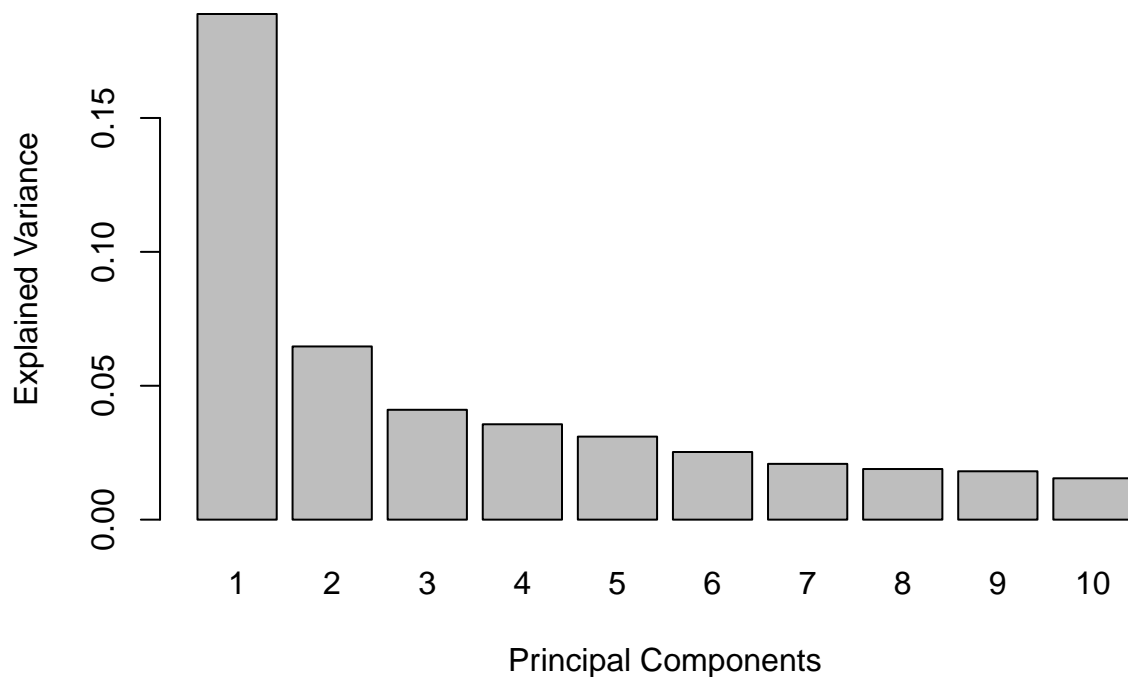
Data selection: n=120 (60 UC and 30 nonIBD samples selected with M/F split). Only C8-pos method data selected from metabolomic dataset.

Data preparation: Selected data was run through xmsPanda regression analysis to determine selected features. This analysis is performed on the regression selected features data.

---

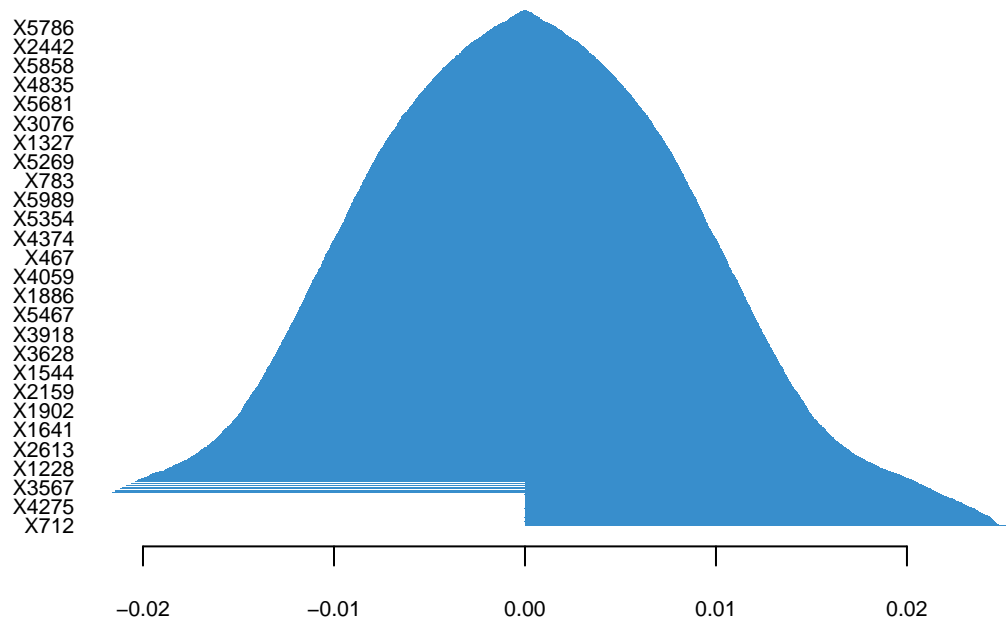
## Initial sPLS-DA model

FIGURE 1: Barplot of the variance each principal component explains of the UC metabolomic



data.

# Loadings on comp 1



## Loadings on comp 2

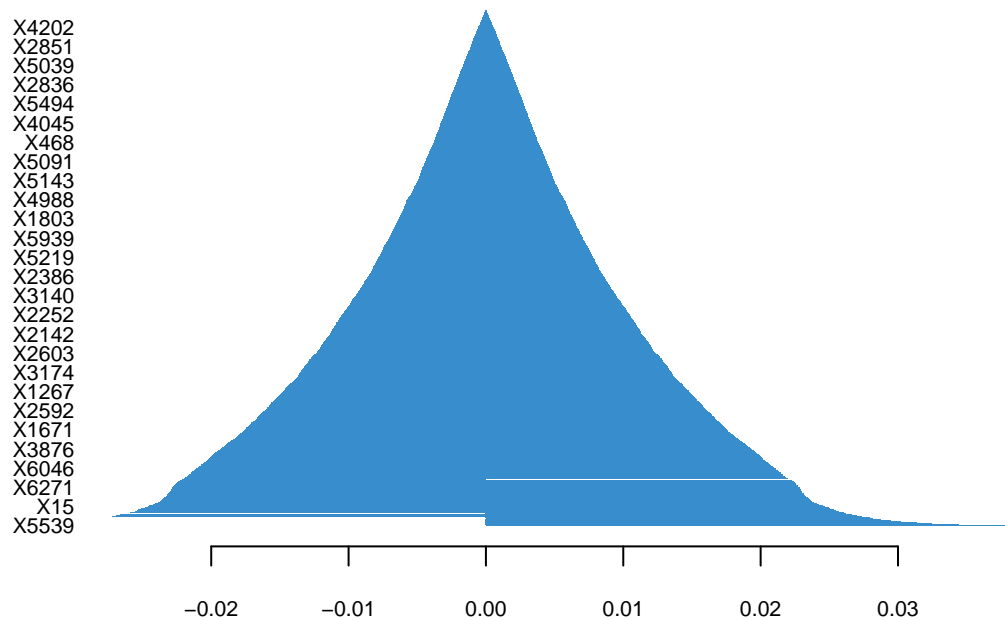
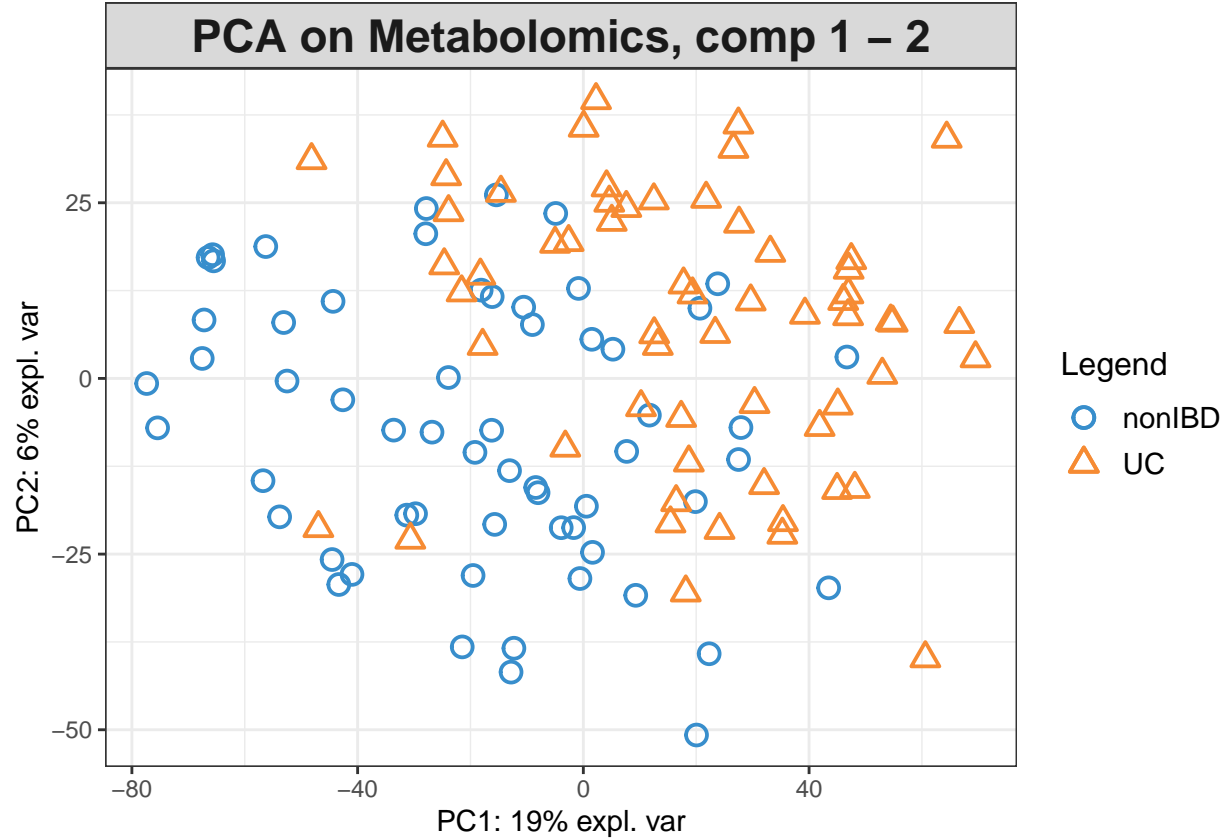


FIGURE 2: Preliminary (unsupervised) analysis with PCA on the UC metabolomic data



# Contribution on comp 1

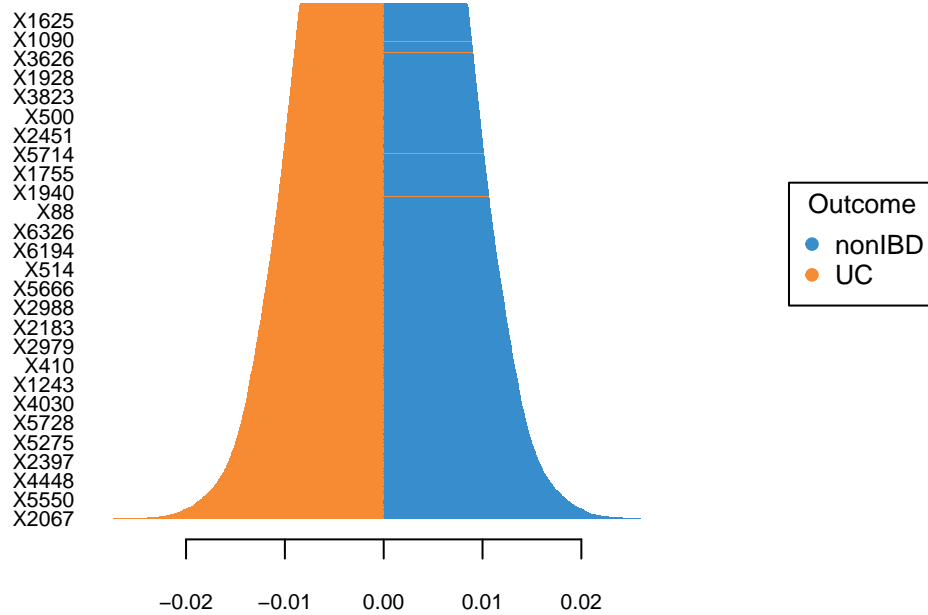
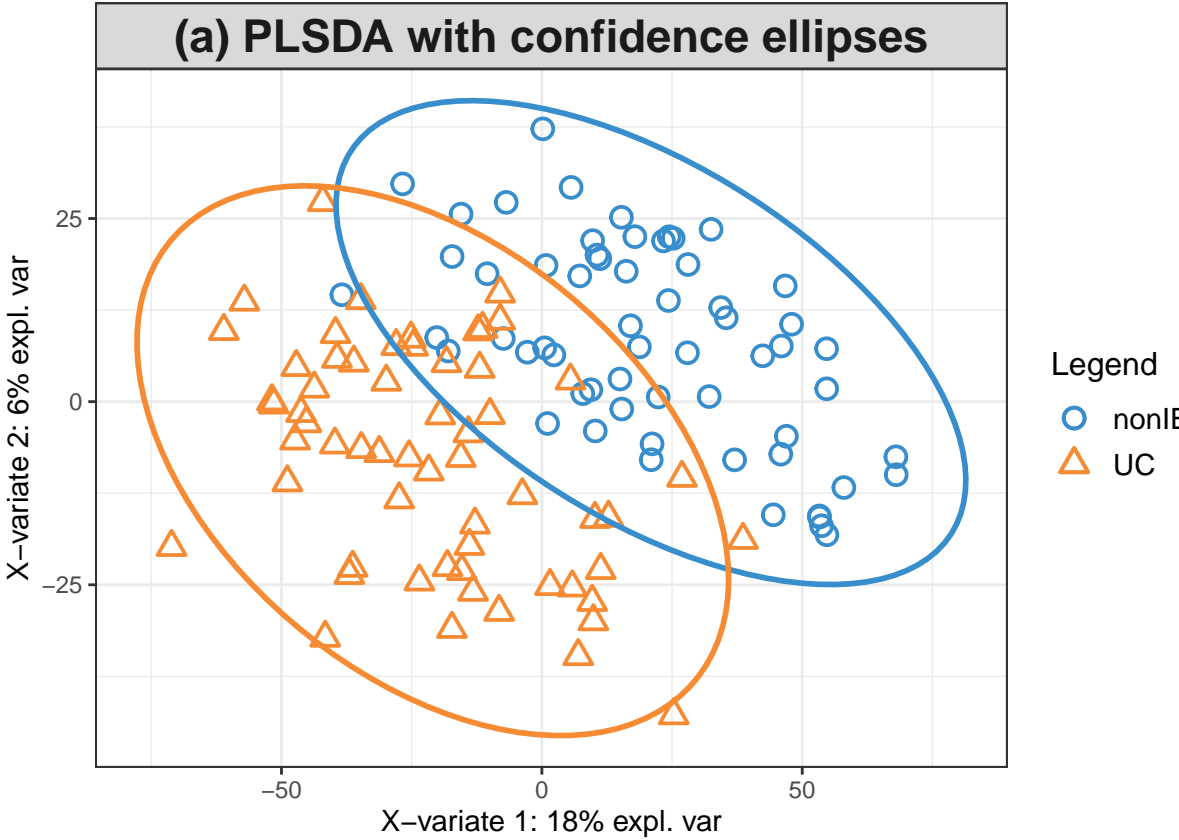
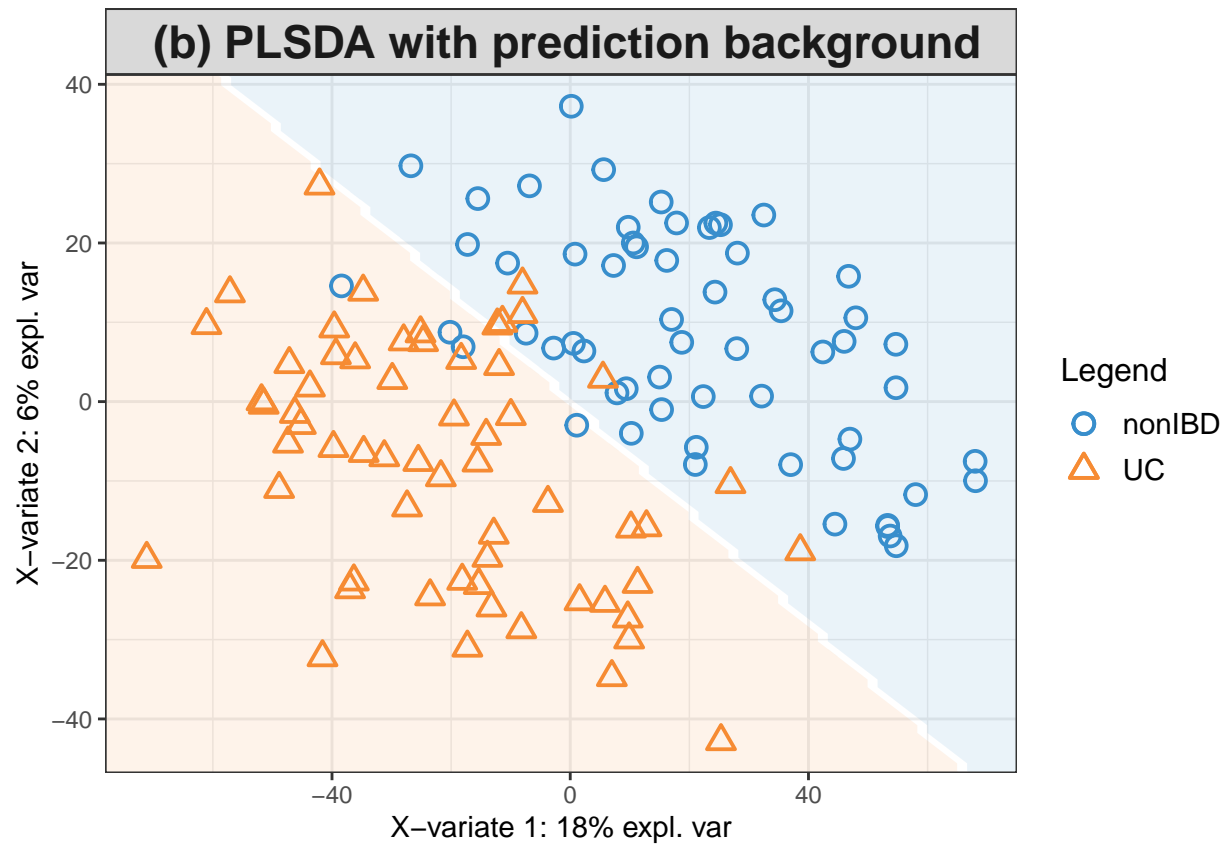


FIGURE 3: Sample plots of the UC metabolomic data after a basic PLS-DA model was oper-



ated on this data.



## Tuning sPLS-DA

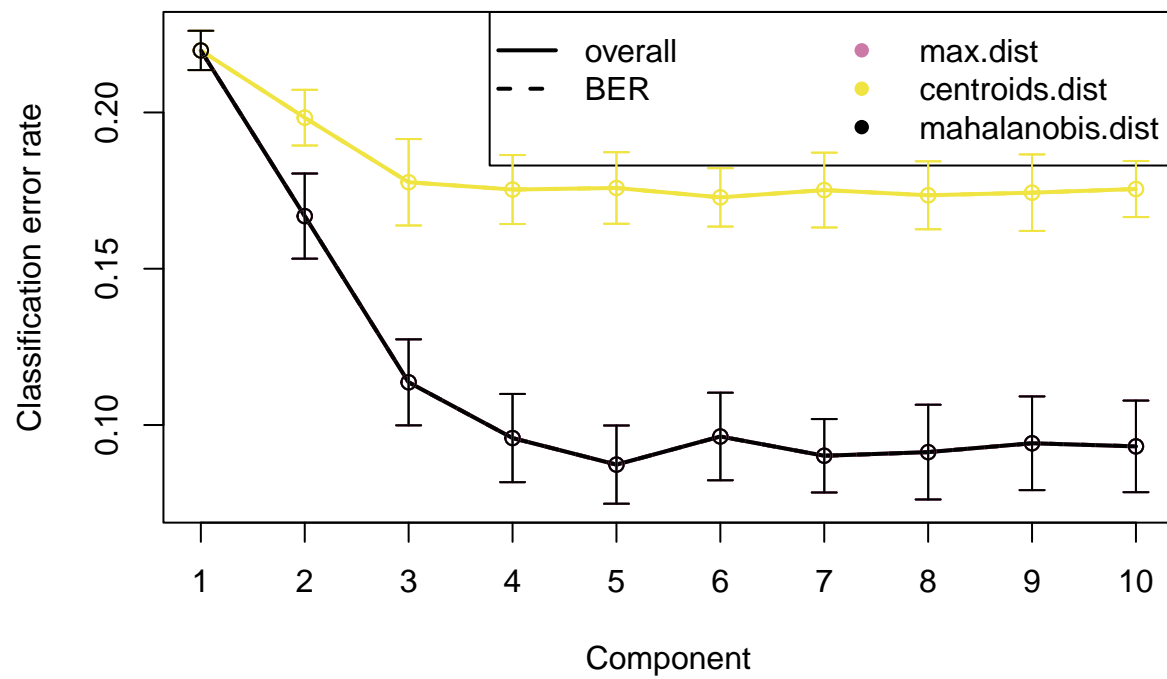
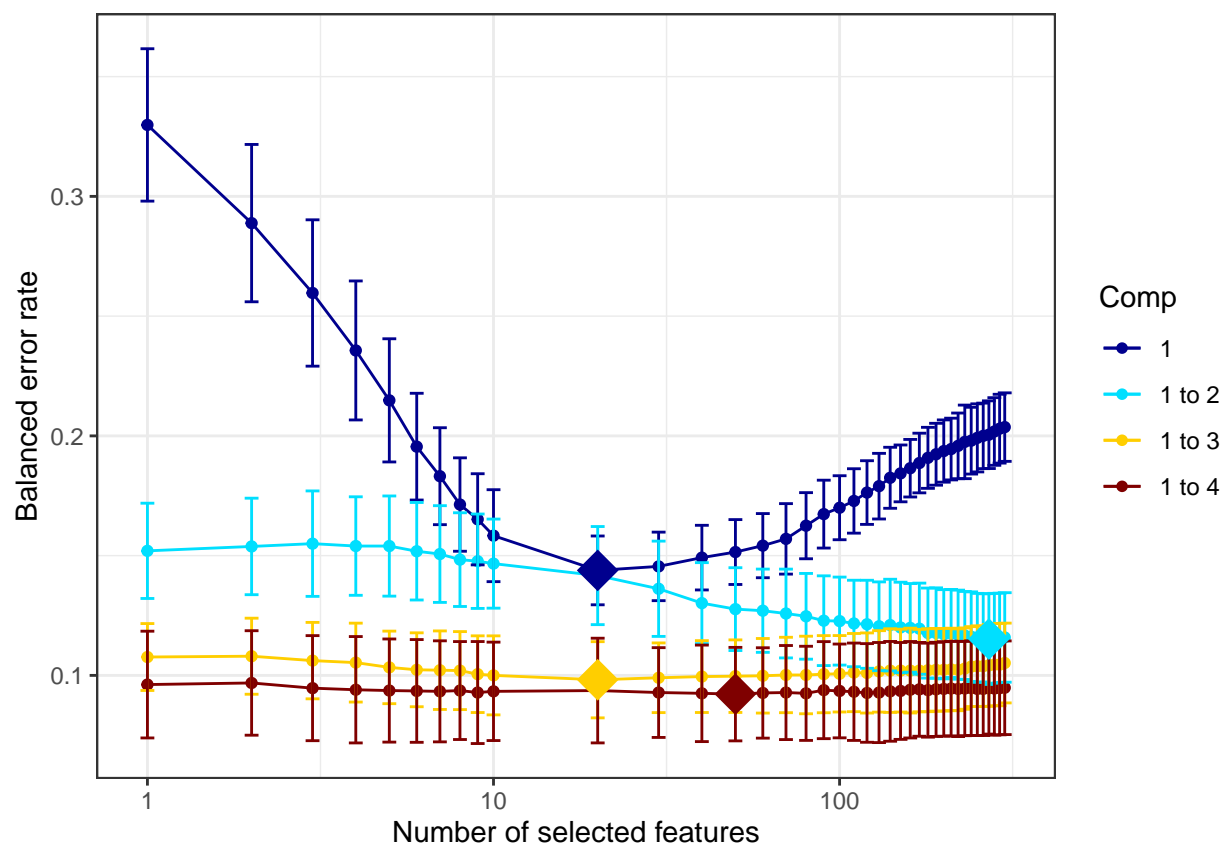
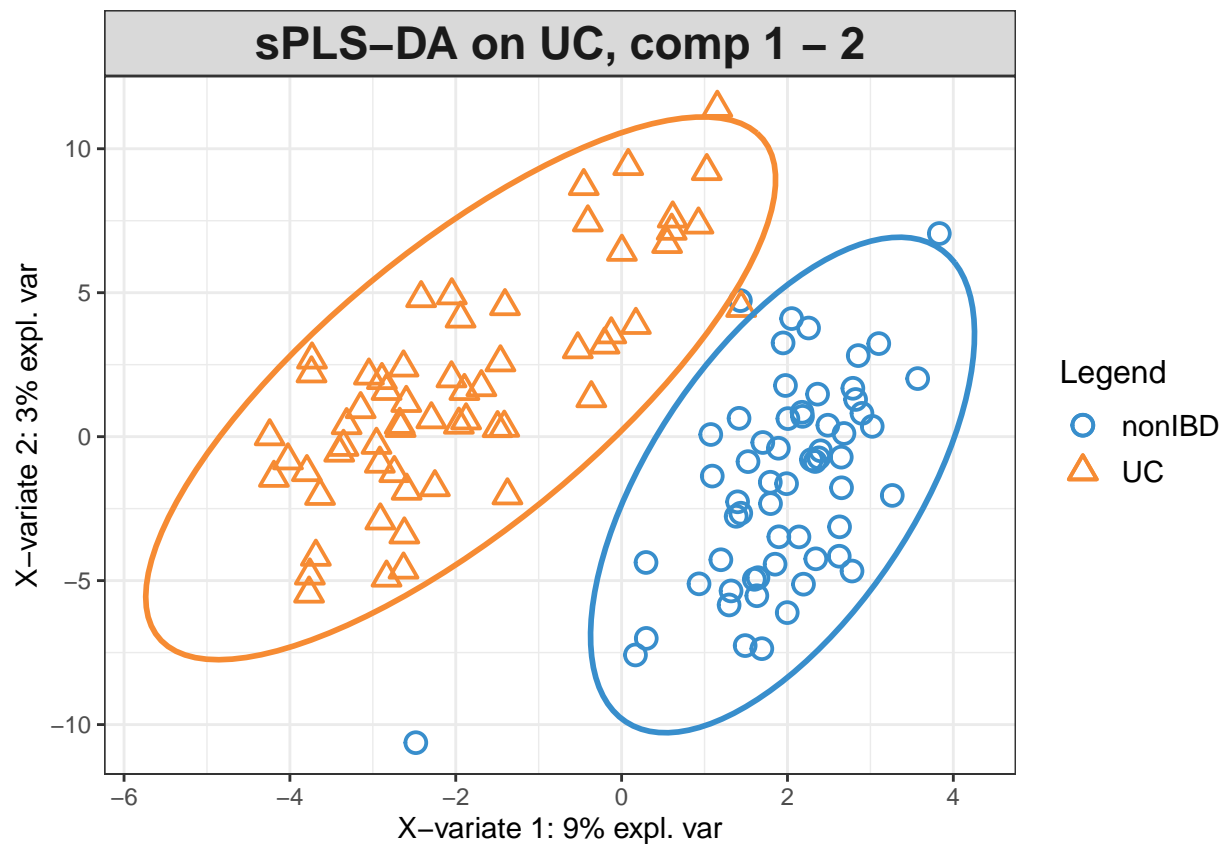
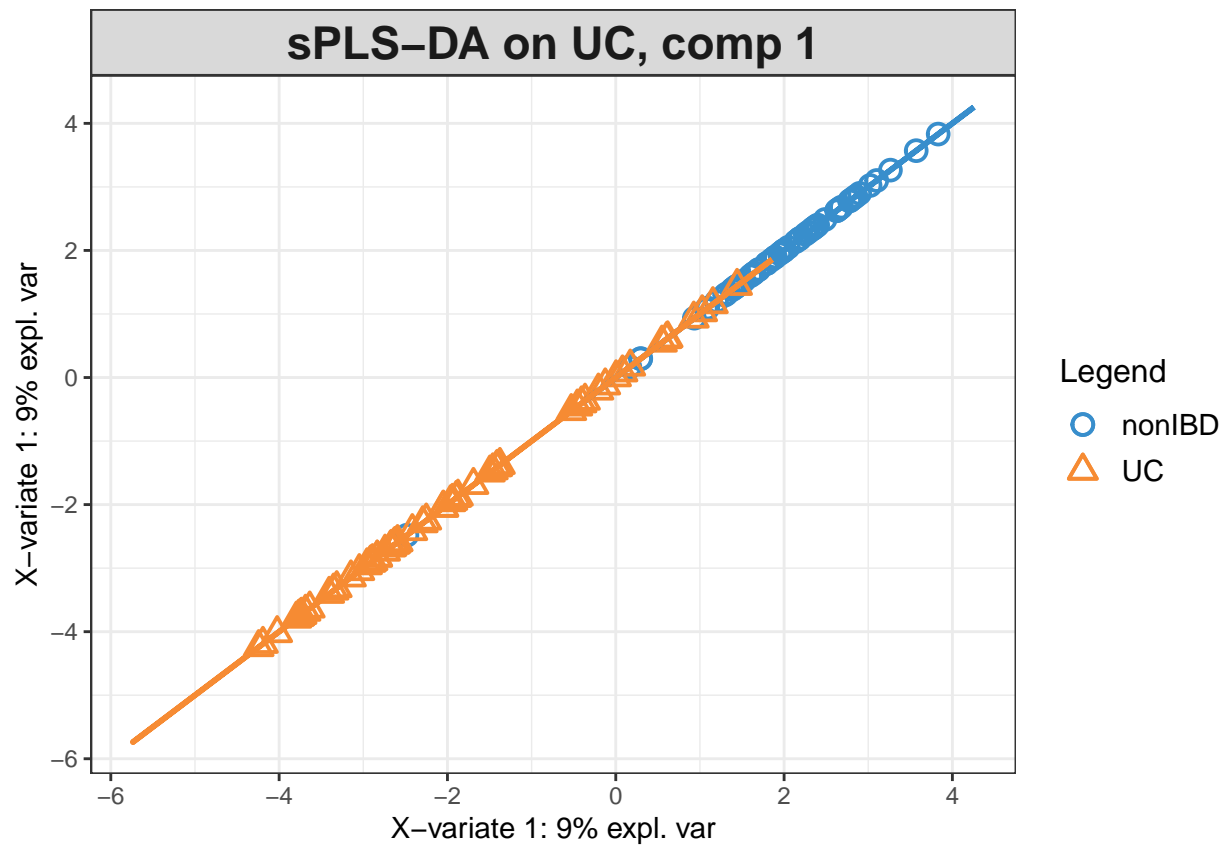


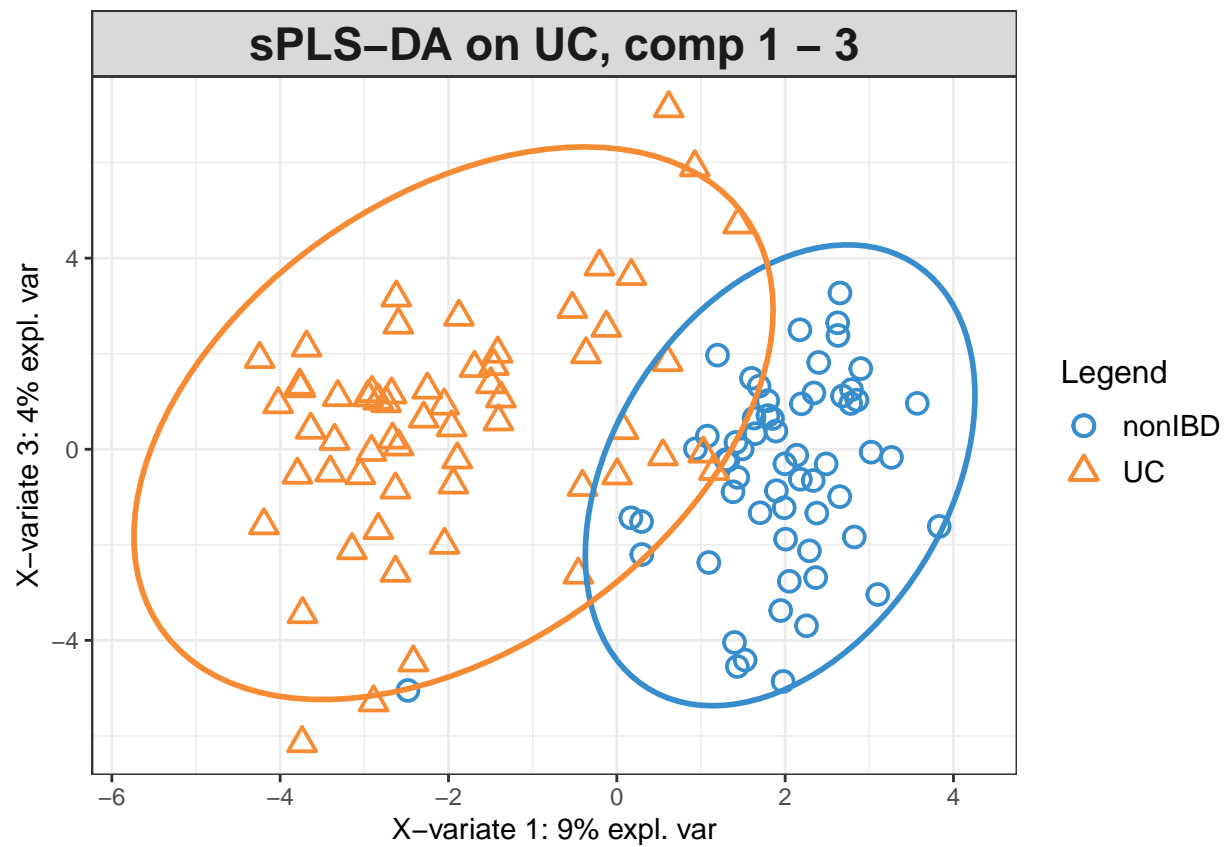


FIGURE 4: Tuning the number of components in PLS-DA on the UC metabolomic data



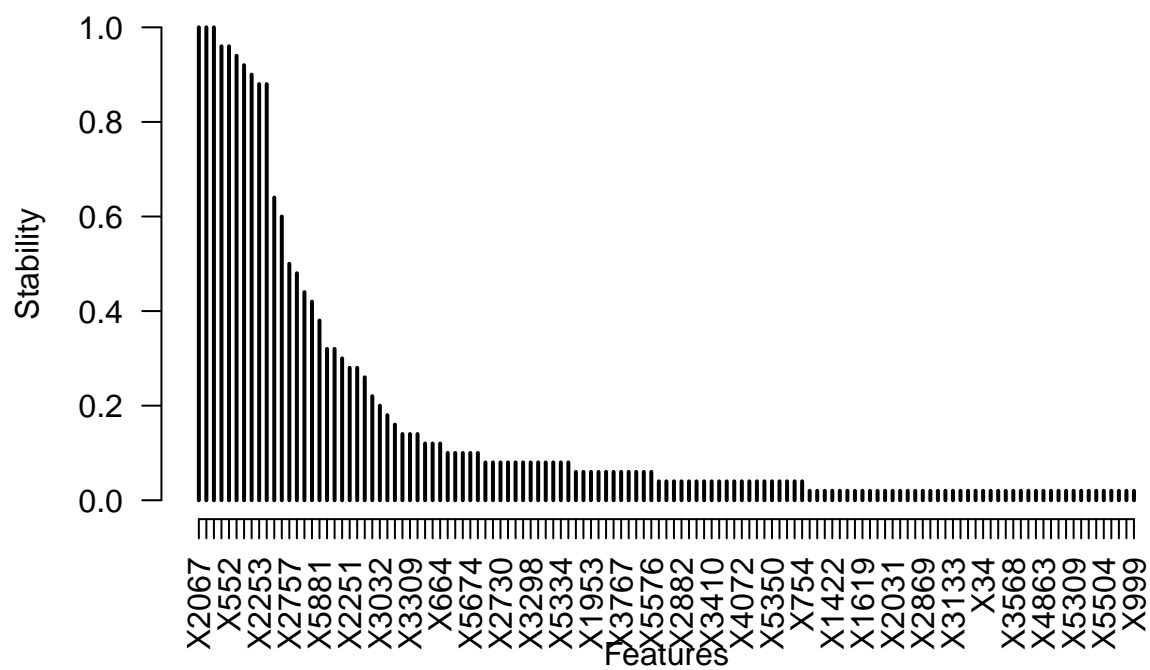
## Creating the final model



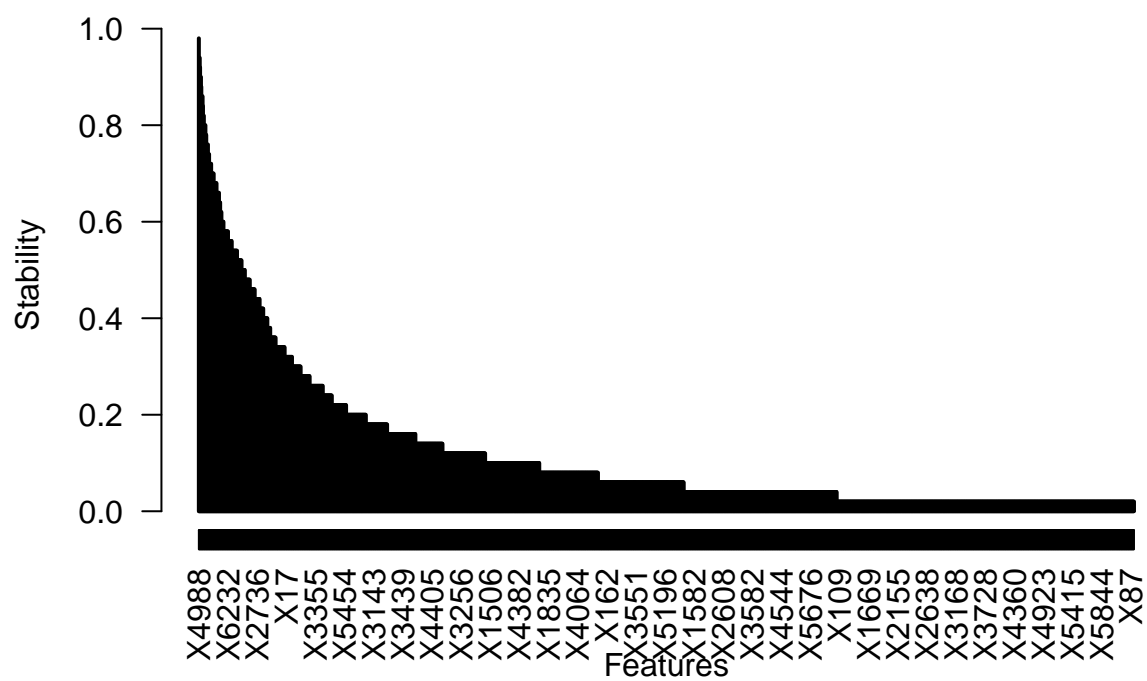


## Error in cim plot: figure margins too large. See ?cim for help.

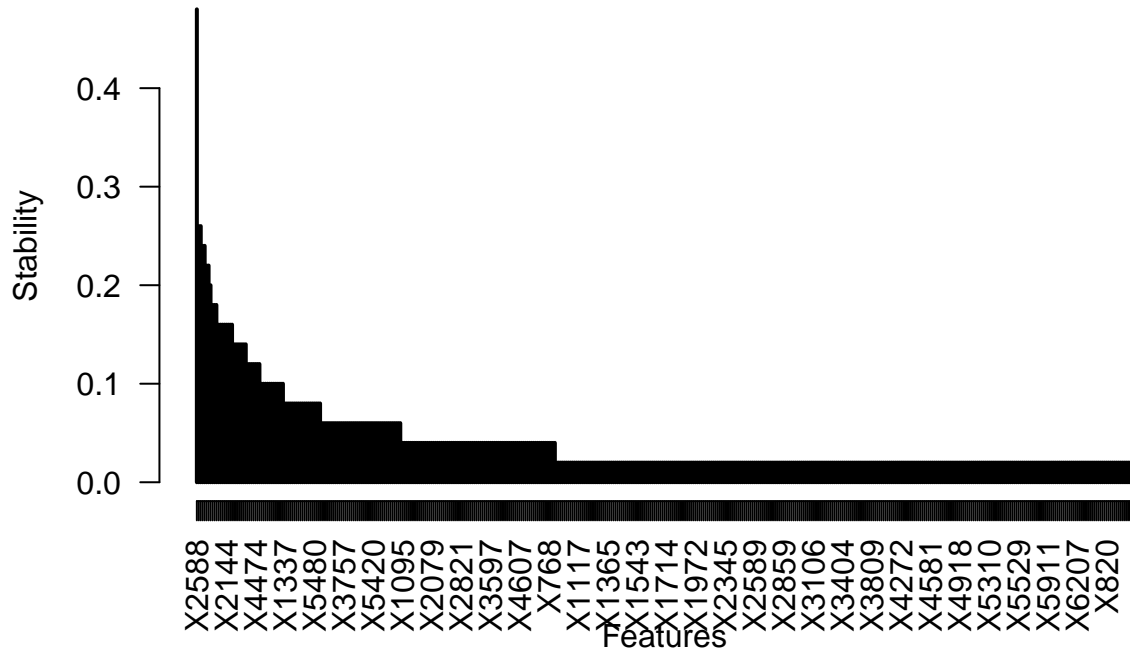
(a) Comp 1



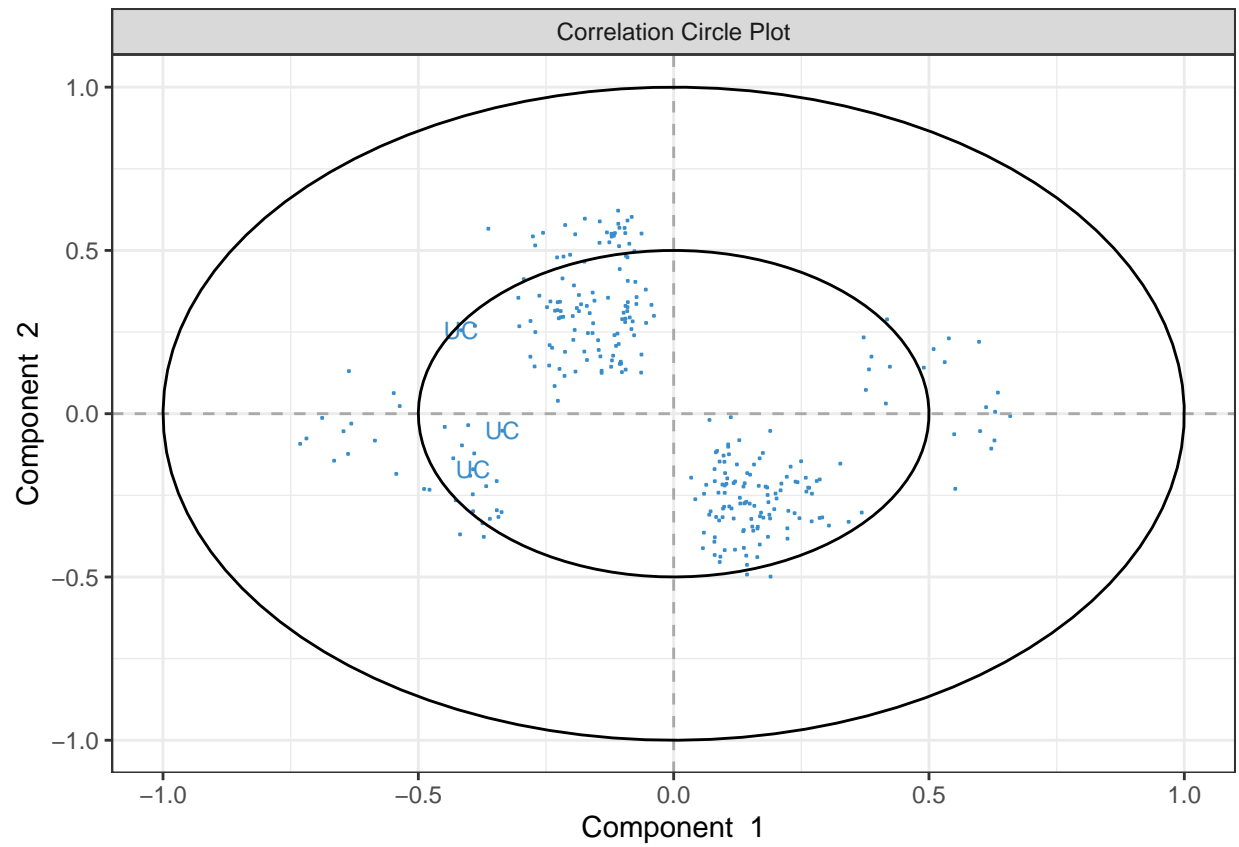
(b) Comp 2



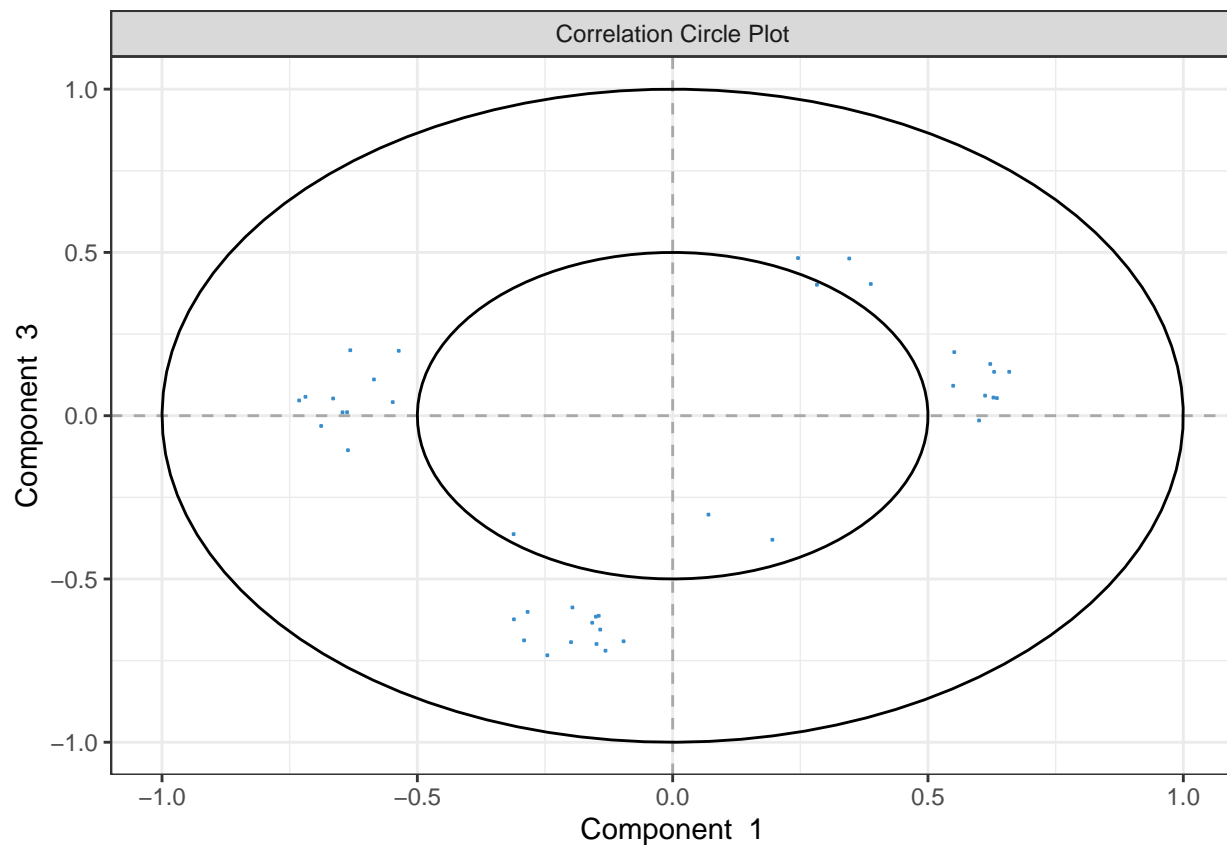
### (c) Comp 3



## Warning: Removed 287 rows containing missing values ('geom\_text()').



```
## Warning: Removed 40 rows containing missing values ('geom_text()').
```



Training the model

Confusion matrix

Table 1: Confusion matrix

	nonIBD	UC
nonIBD	29	4
UC	0	24

```
##
## Attaching package: 'caret'

## The following objects are masked from 'package:mixOmics':
##
##   nearZeroVar, plsda, splsda

## Confusion Matrix and Statistics
##
##           Reference
## Prediction nonIBD UC
```



```

##      nonIBD      29  4
##      UC          0 24
##
##              Accuracy : 0.9298
##              95% CI : (0.83, 0.9805)
##      No Information Rate : 0.5088
##      P-Value [Acc > NIR] : 6.94e-12
##
##              Kappa : 0.8593
##
##      Mcnemar's Test P-Value : 0.1336
##
##              Sensitivity : 0.8571
##              Specificity : 1.0000
##      Pos Pred Value : 1.0000
##      Neg Pred Value : 0.8788
##              Prevalence : 0.4912
##      Detection Rate : 0.4211
##      Detection Prevalence : 0.4211
##      Balanced Accuracy : 0.9286
##
##      'Positive' Class : UC
##

```

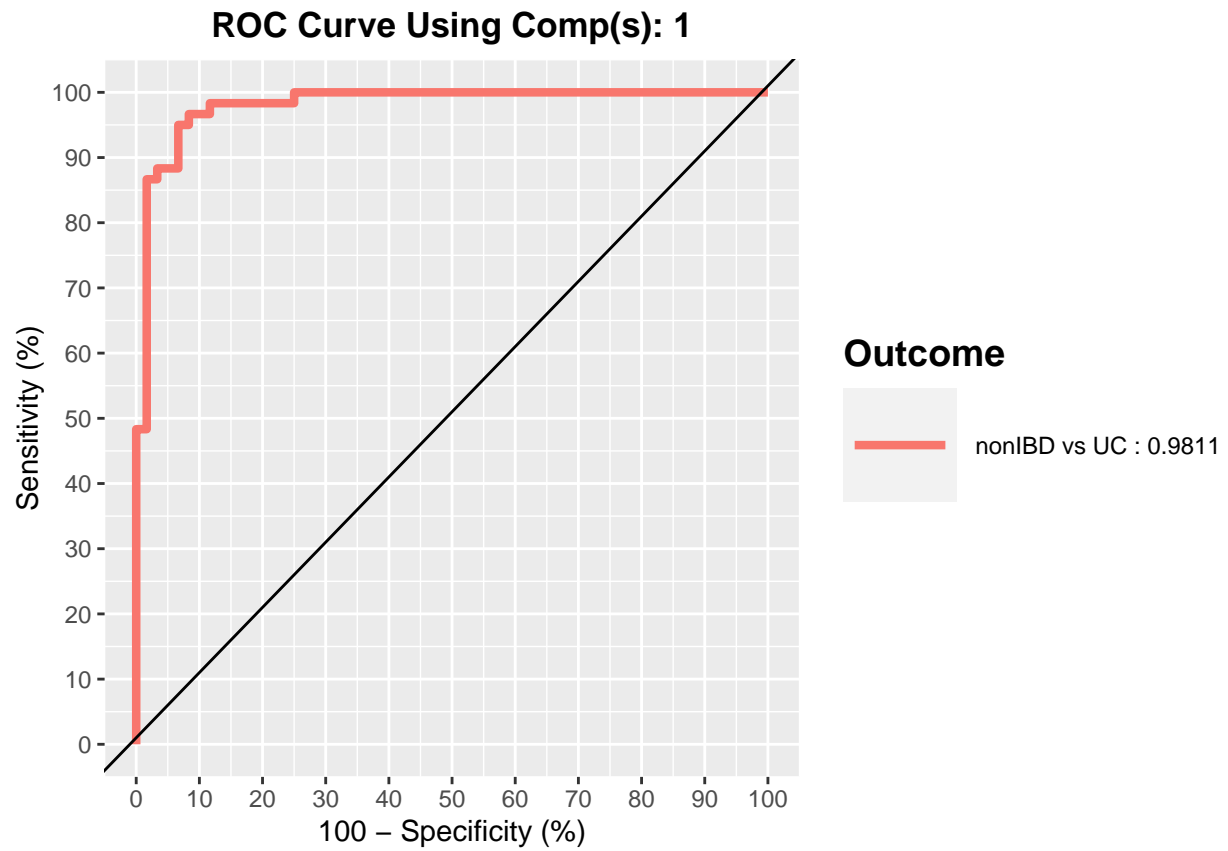
Performance plots

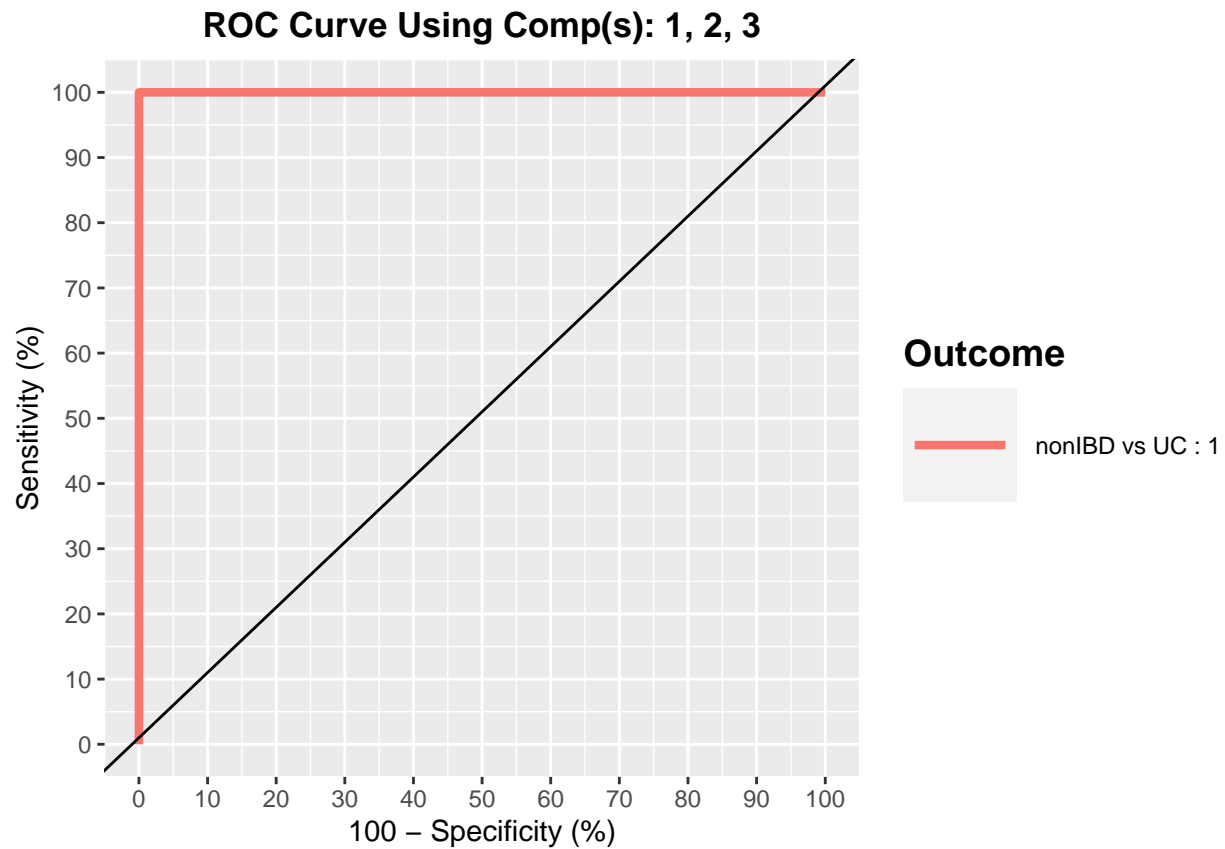
## Confusion Matrix

		Actual	
		nonIBD	UC
Predicted	nonIBD	29	4
	UC	0	24

## Confusion matrix statistics

Sensitivity	Specificity	Precision	Recall	Balanced Accuracy
0.86	1.00	1.00	0.88	0.93
Accuracy	0.93		Kappa	





### Odds ratio for possible confounding variables

```
##               Odds ratio    2.5 %   97.5 %
## (Intercept)           1 0.6028659 1.658744
## samples$sexMale       1 0.4888607 2.045572
```

```
##               Odds ratio    2.5 %   97.5 %
## (Intercept)           0.893486 0.4715531 1.692953
## samples$consent_age   1.004058 0.9851150 1.023366
```

```
##               Odds ratio 2.5 % 97.5 %
## (Intercept)           4.254481e+07    0    Inf
## samples$raceMore than one race 1.762847e-08    0    Inf
## samples$raceWhite      2.056655e-08    0    Inf
```