R console>:
You have provided a sequence of keepX of length: 2 for block mrna and 2 for block protein and 2 for block methylation.
This results in 8 models being fitted for each component and each nrepeat, this may take some time to run, be patient!

```
> # Optimal features to select for each component and block
> optimal_keepX <- tuned$choice.keepX
> print(optimal_keepX)
$mrna
[1] 10  5  5
$protein
[1] 5 5 5
$methylation
[1] 5 5 5

> # Final DIABLO model
> final.diablo <- block.splsda(  X = X, Y = Y,
                                 design = design, ncomp = ncomp,keepX = optimal_keepX)
Design matrix has changed to include Y; each block will be
        linked to Y.
> # Performance assessment
> perf.diablo <- perf( final.diablo,  validation = "Mfold",  folds = 2,  nrepeat = 2  )
```

error.rate

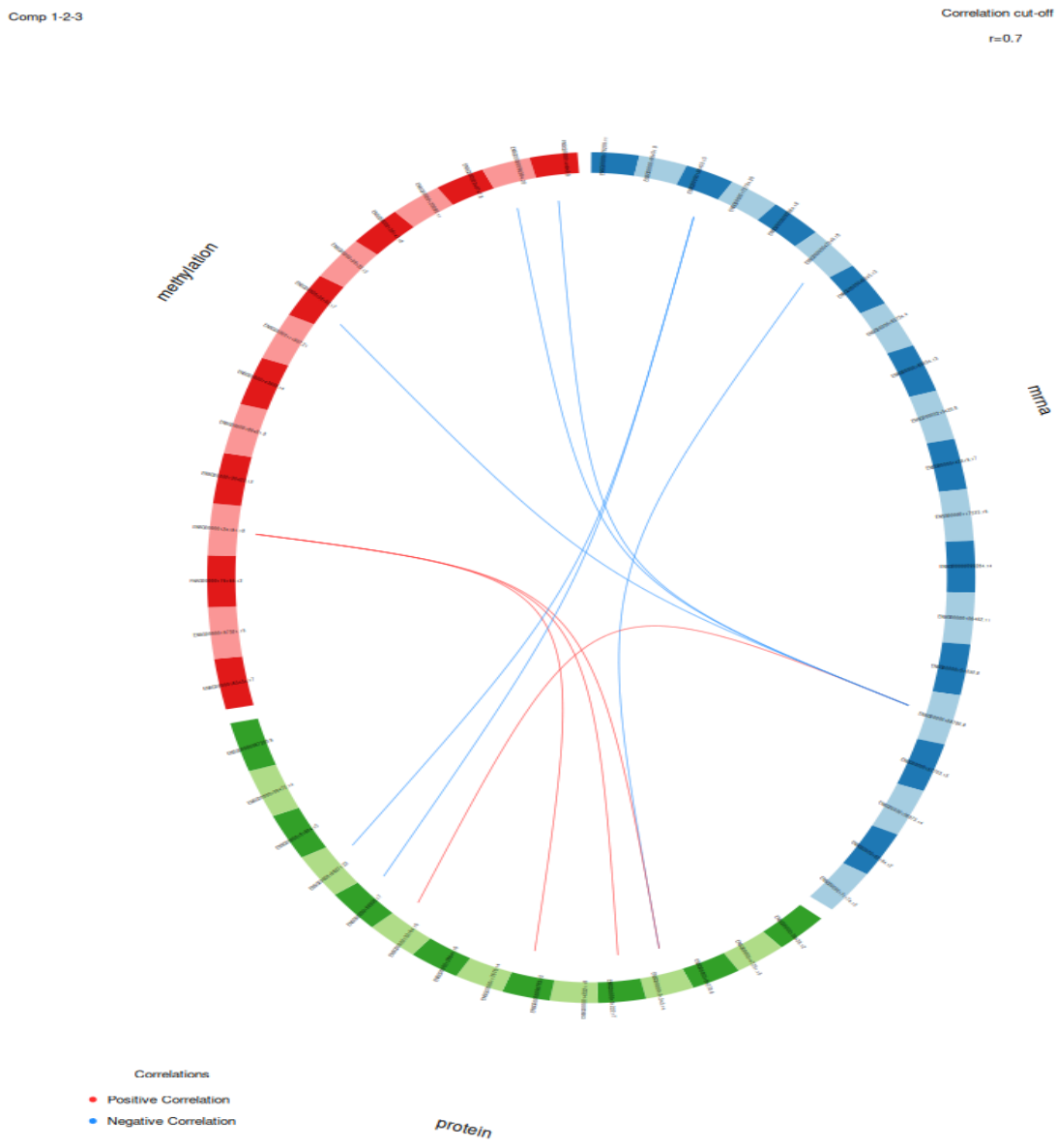
```
> print(perf.diablo$error.rate)
$mrna
      max.dist centroids.dist mahalanobis.dist
comp1 0.3402062      0.4175258       0.4175258
comp2 0.3453608      0.3711340       0.3711340
comp3 0.2938144      0.3453608       0.3247423

$protein
      max.dist centroids.dist mahalanobis.dist
comp1 0.3041237      0.3917526       0.3917526
comp2 0.3505155      0.3608247       0.3659794
comp3 0.3298969      0.3453608       0.3659794

$methylation
      max.dist centroids.dist mahalanobis.dist
comp1 0.3659794      0.4793814       0.4793814
comp2 0.3711340      0.4072165       0.3969072
comp3 0.3659794      0.3917526       0.3814433
```

Interpretation of Error Rates:

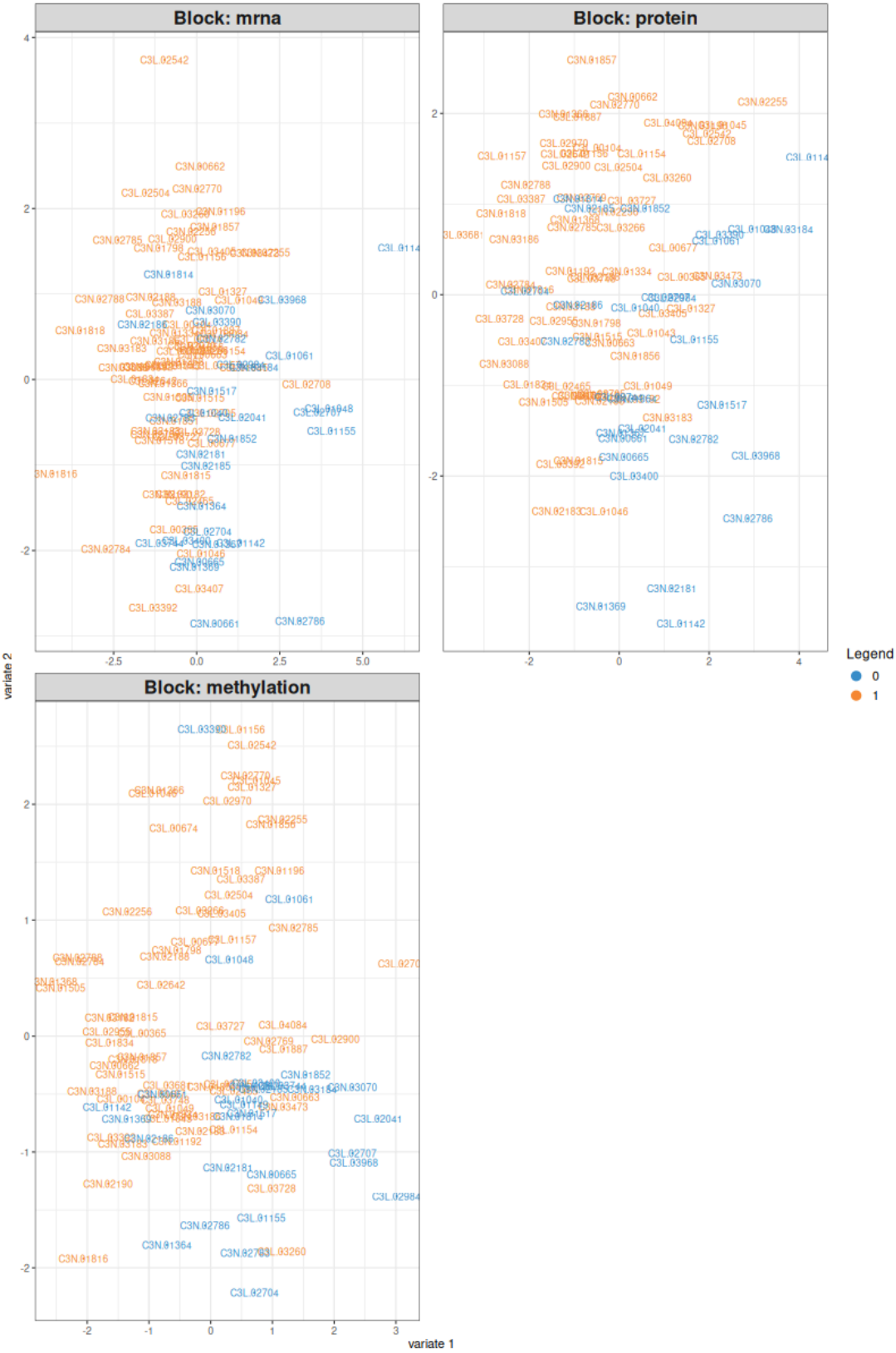
- **mRNA block**:
  - Error rates decrease from Component 1 (e.g., 0.34 for `max.dist`) to Component 3 (0.29), indicating improved classification as more components are included.
- **Protein block**:
  - Error rates are slightly higher overall but also decrease slightly from Component 1 to Component 3 (e.g., `max.dist`: 0.30 → 0.33).
- **Methylation block**:
  - Exhibits higher error rates than the other blocks (e.g., `max.dist` starts at 0.37), suggesting it may be less informative or harder to classify based on the selected features.

```
# Visualize results
plotIndiv(final.diablo, ind.names = FALSE, legend = TRUE,
        title = "GBM Samples by Survival Groups")
```

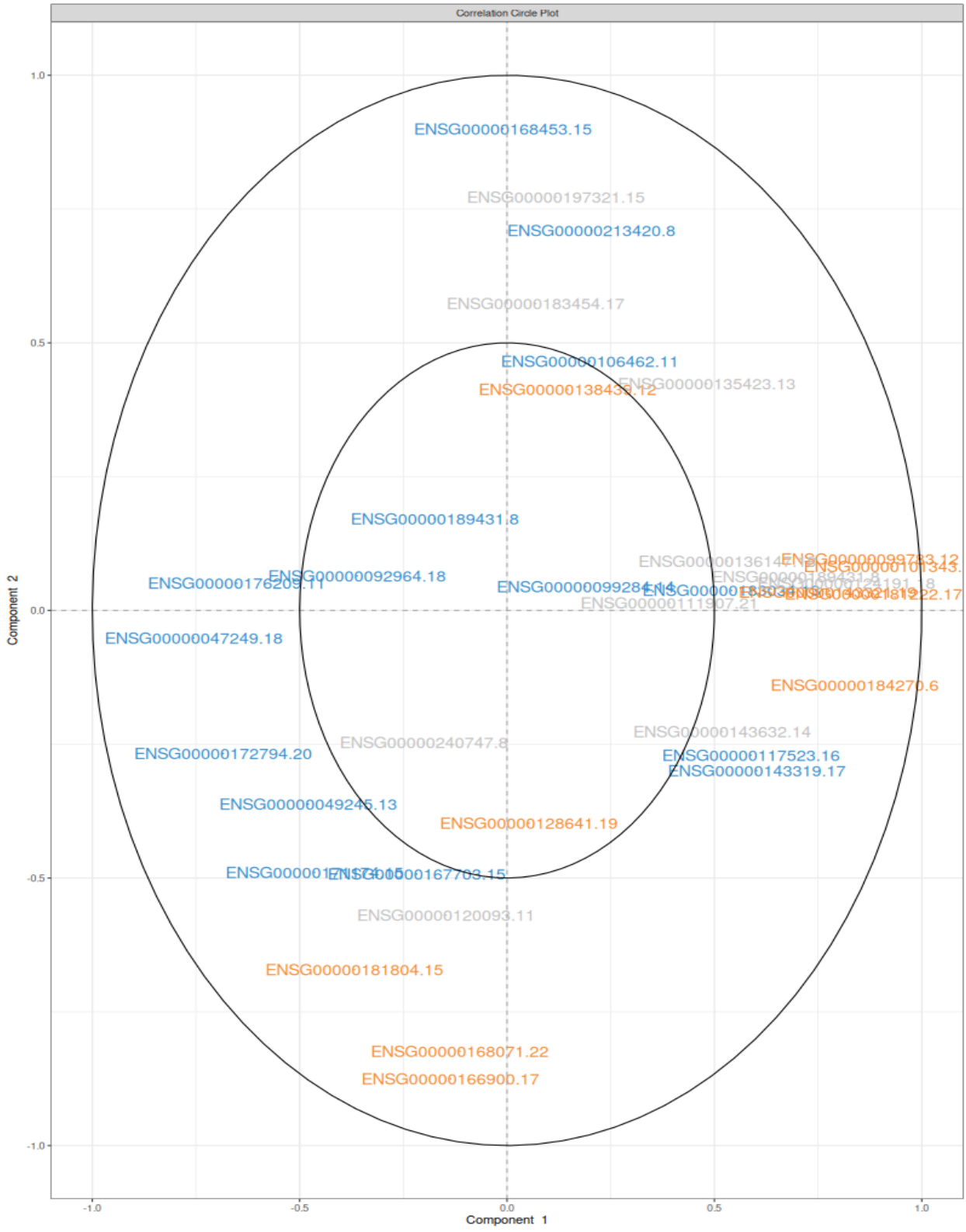
```
circosPlot(final.diablo, cutoff = 0.7, line.col = "blue")
```



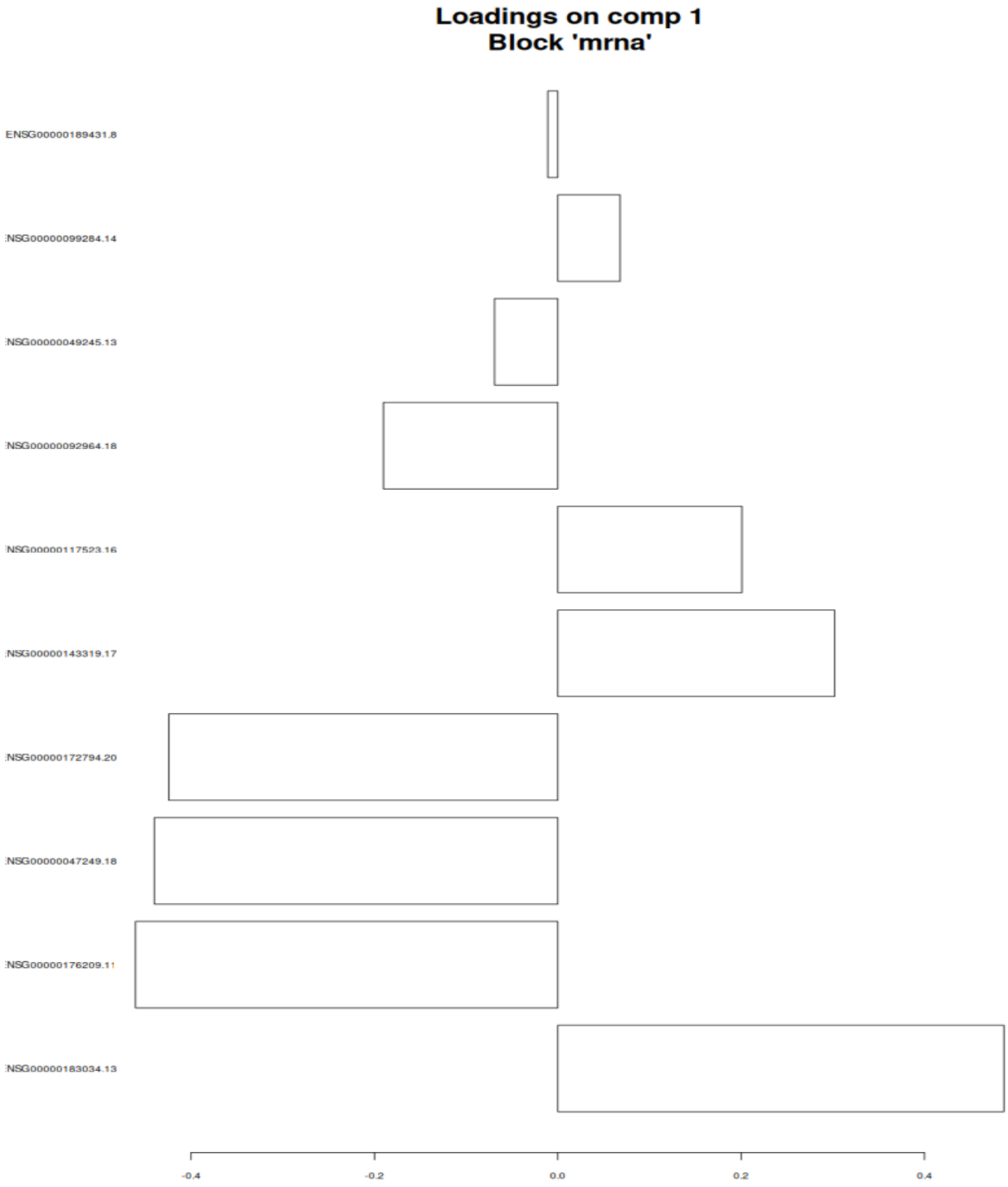
# Plot individual samples based on model components
plotIndiv(final.diablo, comp = 1:2, group = Y, legend = TRUE)



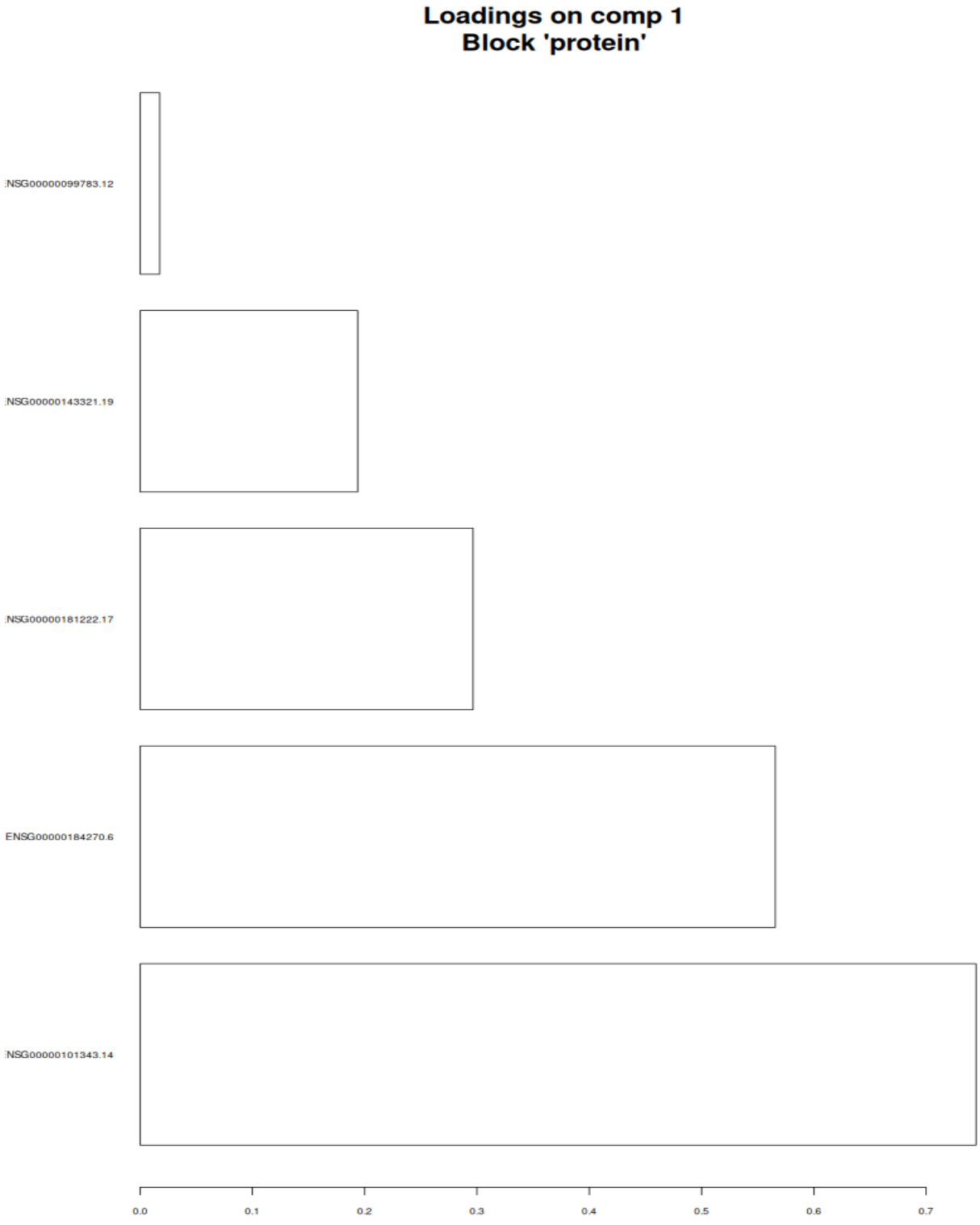
# Plot variable importance (features contributing to the components)
plotVar(final.diablo)



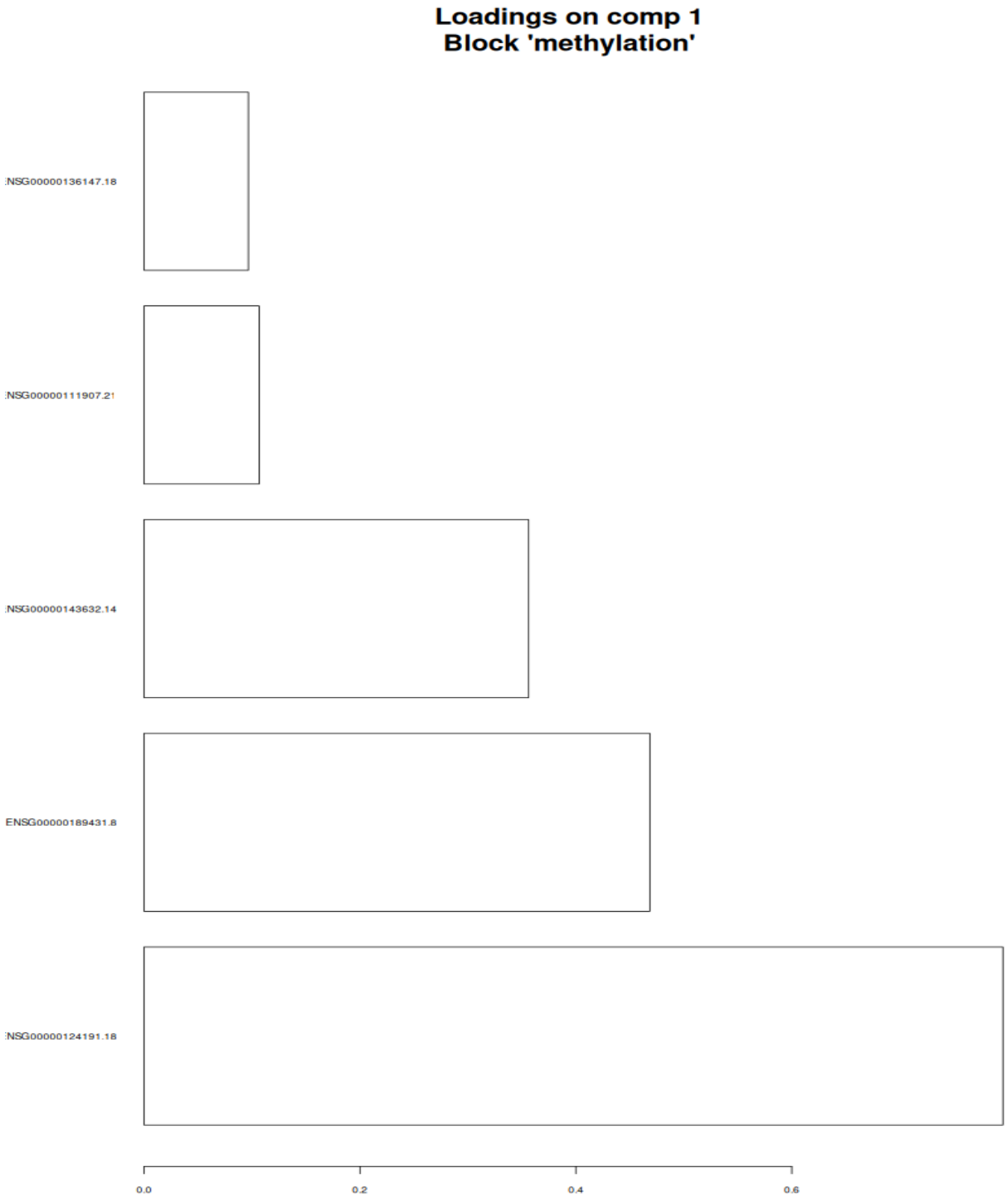
# Visualize the loadings for each block to understand feature importance
# mRNA, protein, and methylation block loadings
plotLoadings(final.diablo, block = "mrna")



plotLoadings(final.diablo, block = "protein")



plotLoadings(final.diablo, block = "methylation")

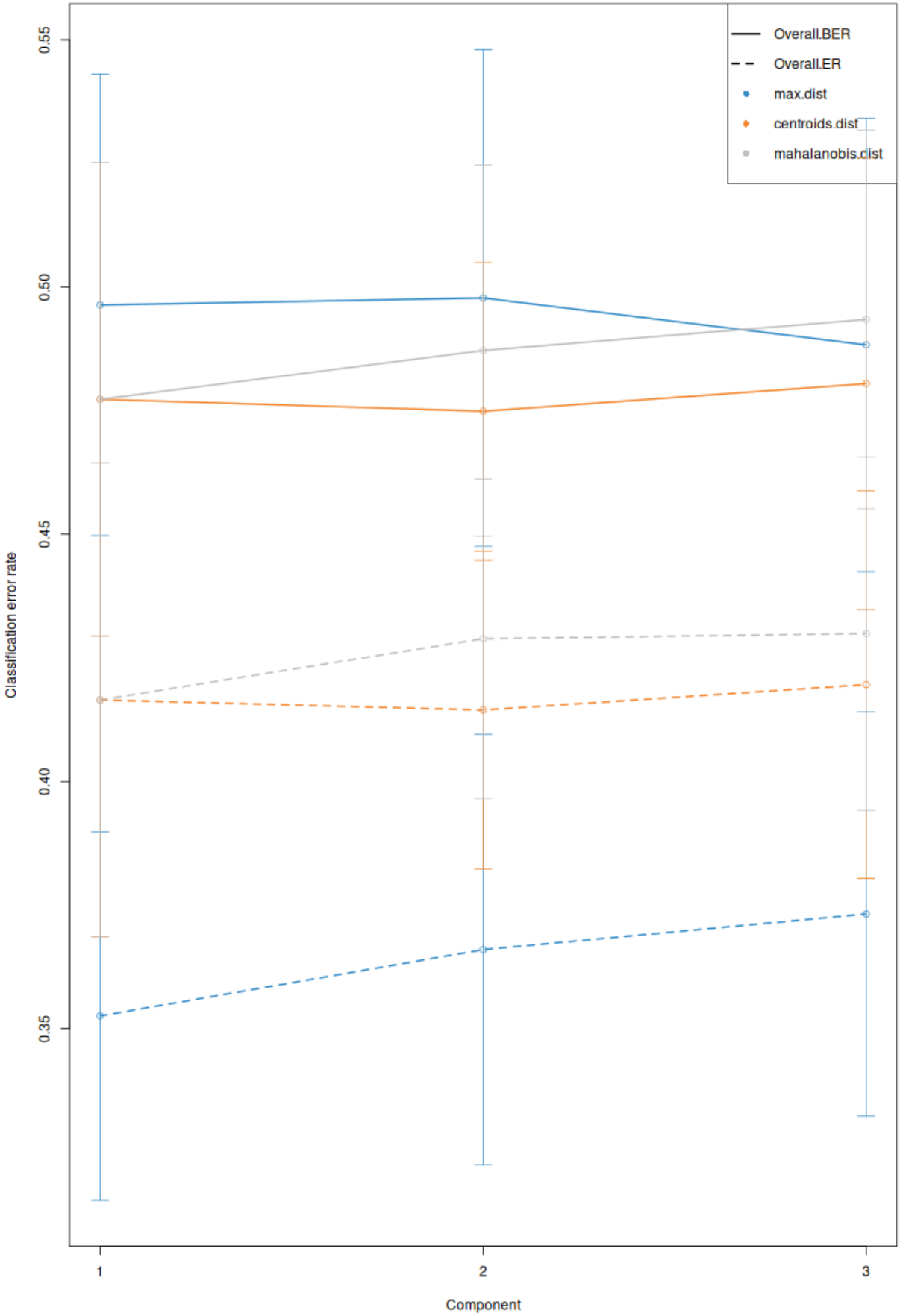


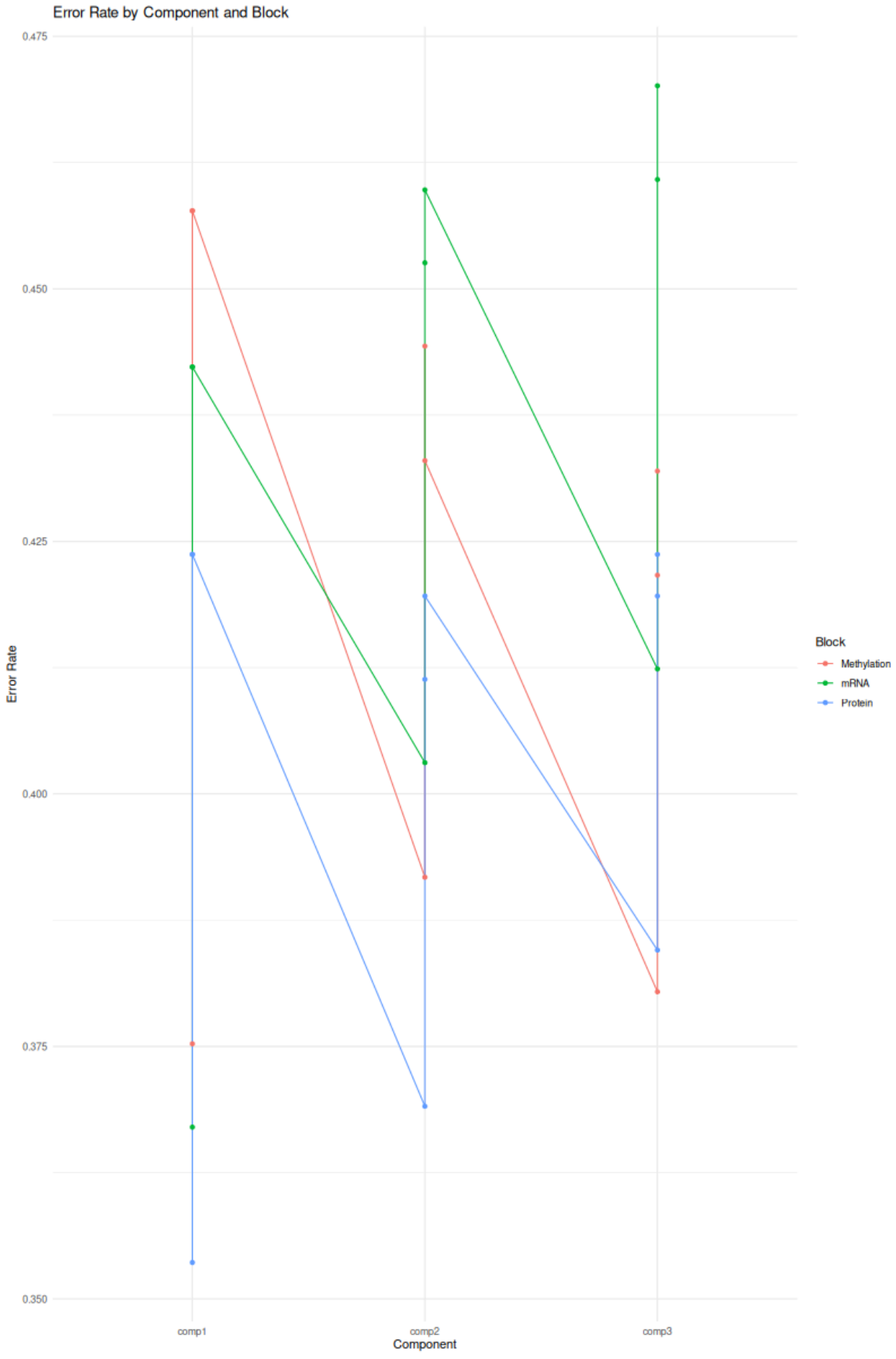
```
# Proportion of explained variance for each block
> prop_var <- block_splsda_result$prop_expl_var
> print(prop_var)
$mrna
    comp1      comp2      comp3
0.05690459 0.07049949 0.03821731

$protein
    comp1      comp2      comp3
0.10449740 0.07121244 0.04044533

$methylation
    comp1      comp2      comp3
0.04200126 0.01474492 0.02073153

$Y
   comp1     comp2     comp3
1.0000000 0.8776561 0.8201494



> # Weights for each component
> weights <- block_splsda_result$weights
> print(weights)
             comp1     comp2     comp3
mrna        0.4593712 0.4315652 0.4473601
protein     0.4255910 0.4353578 0.5350269
methylation 0.3706914 0.4460066 0.3340047



> # Cross-validation performance (e.g., classification)
> cv_result <- perf(block_splsda_result, validation = "Mfold", folds = 5, nrepeat = 10)
> print(cv_result)

Call:
 perf.sgccda(object = block_splsda_result, validation = "Mfold", folds = 5, nrepeat = 10)
```

Main numerical outputs:

Error rate (overall or BER) for each component and for each distance: see object$error.rate
Error rate per class, for each component and for each distance: see object$error.rate.class
Prediction values for each component: see object$predict
Classification of each sample, for each component and for each distance: see object$class
Stable features on each component: see object$features$stable
AUC values: see object$auc if auc = TRUE

```
> # 1. Error Rate for each component and each distance
> # View overall error rate
> print(cv_result$error.rate)
$mrna
      max.dist centroids.dist mahalanobis.dist
comp1 0.3670103      0.4422680       0.4422680
comp2 0.4030928      0.4525773       0.4597938
comp3 0.4123711      0.4608247       0.4701031

$protein
      max.dist centroids.dist mahalanobis.dist
comp1 0.3536082      0.4237113       0.4237113
comp2 0.3690722      0.4113402       0.4195876
comp3 0.3845361      0.4195876       0.4237113

$methylation
      max.dist centroids.dist mahalanobis.dist
comp1 0.3752577      0.4577320       0.4577320
comp2 0.3917526      0.4443299       0.4329897
comp3 0.3804124      0.4319588       0.4216495
```


Overall.BER
Overall.ER
max.dist
centroids.dist
mahalanobis.dist
Classification error rate
0.55
0.50
0.45
0.40
0.35
1
2
3
Component


Error Rate by Component and Block
0.475
0.450
0.425
0.400
0.375
0.350
Error Rate
comp1
comp2
comp3
Component
Block
Methylation
mRNA
Protein

top_mrna_genes

---

c(NA, NA, "ENSG00000183034.13", NA, NA, "ENSG00000143319.17",
"ENSG00000117523.16", "ENSG00000099284.14", NA, "ENSG00000160789.20"
)

ENSG00000183034.13
NCBI Gene Summary for OTOP2 Gene
Predicted to enable proton channel activity. Predicted to be involved in proton transmembrane transport. Predicted to be located in plasma membrane. Predicted to be integral component of membrane. [provided by Alliance of Genome Resources, Apr 2022]

ENSG00000099284.14
MacroH2A histone variants modulate enhancer activity to repress oncogenic programs and cellular reprogramming
https://www.nature.com/articles/s42003-023-04571-1

ENSG00000143319.17
Interferon-stimulated 20 kDa exonuclease-like 2

ENSG00000117523.16
NCBI Gene Summary for PRRC2C Gene
Enables protein C-terminus binding activity. Involved in stress granule assembly. Located in cytosol. [provided by Alliance of Genome Resources, Apr 2022]

ENSG00000160789.20
deregulation in cancer, where it contributes to genomic instability and aggressive tumor behavior, with potential as a biomarker for cancer diagnosis
https://pubmed.ncbi.nlm.nih.gov/33316938/