Bioconductor mAPKL Package

Argiris Sakellariou\textsuperscript{1,2}

\textsuperscript{1} Biomedical Informatics Unit, Biomedical Research Foundation of the Academy of Athens, Athens, Greece
\textsuperscript{2} Department of Informatics and Telecommunications, National and Kapodistrian Univ. of Athens, Athens, Greece

argisake@gmail.com

October 19, 2016

Contents

1 Introduction 1

2 Identification of deferentially expressed genes 2
  2.1 Breast cancer data ................................................. 2
  2.2 Data normalization and transformation .................................. 3
  2.3 mAPKL gene selection .................................................... 5
  2.4 Building and evaluating classification models ............................ 5

3 Advanced usage of the package 7
  3.1 Annotation analysis ....................................................... 7
  3.2 Network characteristics .................................................. 9

4 Reporting 11

5 Session info 11

6 Reference 12

1 Introduction

The mAPKL bioconductor R package implements a hybrid gene selection method, which is based on the hypothesis that among the statistically significant genes in a ranked list, there should be clusters of
genes that share similar biological functions related to the investigated disease. Thus, instead of keeping a number of $N$ top ranked genes, it would be more appropriate to define and keep a number of gene cluster exemplars.

The proposed methodology combines filtering and cluster analysis to select a small yet highly discriminatory set of non-redundant genes. Regarding the filtering step, a multiple hypothesis testing approach from *multtest* r-package (maxT) is employed to rank the genes of the training set according to their differential expression. Then the top $N$ genes (e.g. $N = 200$) are reserved for cluster analysis. First the index of Krzanowski and Lai as included in the *ClusterSim* r-package is applied on the disease samples of the training set to determine the number of clusters. The Krzanowski and Lai index is defined by $\text{DIFF}(k) = (k - 1)^2 W_{k-1} - k^2 W_k$ when choosing the number of clusters ($k$) to maximize the quantity $KL(k) = \left| \frac{\text{DIFF}(k)}{\text{DIFF}(k+1)} \right|$. The $W_k$ denotes the within-cluster sum of squared errors.

Finally, cluster analysis is performed with the aid of Affinity Propagation (AP) clustering algorithm, which detects $n$ (where $n = k$ the Krzanowski and Lai index) clusters among the top $N$ genes, the so-called exemplars. Those $n$ exemplars are expected to form a classifier that shall discriminate between normal and disease samples (Sakellariou et al. 2012, *BMC Bioinformatics* 13:270). This package implements the pre-mentioned methodology through a core function, the *mAPKL*. In the upcoming sections follows a guidance of the included functions and its functionality through differential expression analysis scenarios on a breast cancer dataset (GSE5764) which is part of the *mAPKLDat*a package.

## 2 Identification of deferentially expressed genes

### 2.1 Breast cancer data

Throught this tutorial we utilized a publicly available breast cancer dataset comprised of 30 samples, where 20 of them represent normal cases and the remaining 10 samples stand for tumor cases. We first load the package and then the breast cancer data. Then with the aid of the *sampling* function we create a separate training and validation sets where 60% of the samples will be used for training and the rest 40% of the samples will be used for evaluation purposes.

```r
library(mAPKL)
library(mAPKLDat)
data(mAPKLDat)
varLabels(mAPKLDat)
breast <- sampling(Data=mAPKLDat, valPercent=40, classLabels="type", seed=135)
```

The *sampling* function returns an S3 class (breast) with two eSet class objects that nest the relevant training and validation samples. Those two objects are used throughout the rest of the analysis.

```r
breast
```

```r
## $trainData
```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 54675 features, 18 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM134584 GSM134690 ... GSM134695 (18 total)
##   varLabels: title type
##   varMetadata: labelDescription
## featureData: none
## experimentData: use 'experimentData(object)'
##   pubMedIds: 17389037
## Annotation: hgu133plus2

## $testData
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 54675 features, 12 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM134691 GSM134588 ... GSM134692 (12 total)
##   varLabels: title type
##   varMetadata: labelDescription
## featureData: none
## experimentData: use 'experimentData(object)'
##   pubMedIds: 17389037
## Annotation: hgu133plus2

### 2.2 Data normalization and transformation

We perform normalization to the expression values through the `preprocess` function.

```r
normTrainData <- preprocess(breast$trainData)
normTestData <- preprocess(breast$testData)
```

The `preprocess` function produces a list of several available normalization and transformation options. Besides density plots per method are produced and saved to current working directory to assist the user to decide upon which method to select before proceeding to mAPKL analysis.
Figure 1: Density plots of normalized intensity values

The following graph presents the density plots of 8 possible normalization processes with or without log2 transformation. The `preprocess` function applies all of them and it is up to the user, which one will engage for the rest of the analysis. In brief, the available approaches are mean-centering, z-score, quantile, and cyclic loess. During this case study we will proceed with the expression values following log2 transformation and cyclic loess normalization.
2.3 mAPKL gene selection

In this example we employ the expression values of log2 transformation and cyclic loess normalization to proceed with the mAPKL analysis.

```r
exprs(breast$trainData) <- normTrainData$clL2.normdata
dataType = 7)
```

## b=10 b=20 b=30 b=40 b=50 b=60 b=70 b=80 b=90 b=100
## b=110 b=120 b=130 b=140 b=150 b=160 b=170 b=180 b=190 b=200
## b=210 b=220 b=230 b=240 b=250 b=260 b=270 b=280 b=290 b=300
## b=310 b=320 b=330 b=340 b=350 b=360 b=370 b=380 b=390 b=400
## b=410 b=420 b=430 b=440 b=450 b=460 b=470 b=480 b=490 b=500
## b=510 b=520 b=530 b=540 b=550 b=560 b=570 b=580 b=590 b=600
## b=610 b=620 b=630 b=640 b=650 b=660 b=670 b=680 b=690 b=700
## b=710 b=720 b=730 b=740 b=750 b=760 b=770 b=780 b=790 b=800
## b=810 b=820 b=830 b=840 b=850 b=860 b=870 b=880 b=890 b=900
## b=910 b=920 b=930 b=940 b=950 b=960 b=970 b=980 b=990 b=1000

## Please wait! The (KL) cluster indexing may take several minutes...
## Asking for 15 number of clusters
## fc according to limma
```

2.4 Building and evaluating classification models

After having get the exemplars from mAPKL analysis we build an SVM classifier to test their discriminatory performance. Regarding the SVM setup, we utilize a linear kernel for which the cost attribute is inferred by the tune.svm function. however, the user may freely use another kernel and a different Cross Validation approach than 5-folds.

```r
classPred <- classification(out.clL2@exemplTrain, "type", out.clL2@exemplTest)
```

## The training set has 12 Negative and 6 Positive samples. Using k-fold=5 C-V
## THE BEST PARAMETERS TUNING STAGE
## THE TRAINING STAGE
```

## Call:
## svm.default(x = train mtx, y = lbs, scale = FALSE, type = "C-classification")
## kernel = "linear", gamma = best_gamma, cost = best_cost, cross = k_fold)
##
## Parameters:
## SVM-Type:  C-classification
## SVM-Kernel:  linear
## cost:  2
## gamma:  0.125
##
## Number of Support Vectors:  5

## ############# THE PREDICTION STAGE #######################

## Test Labels Prediction Labels
## GSM134691  0  0
## GSM134588  0  0
## GSM134688  0  0
## GSM134694  0  1
## GSM134697  0  0
## GSM134700  0  0
## GSM134687  0  0
## GSM134709  0  0
## GSM134710  1  1
## GSM134698  1  1
## GSM134689  1  1
## GSM134692  1  1

## Negative samples:  8
## Positive samples:  4

## TN=7
## FP=1
## TP=4
## FN=0
## AUC=0.94
## Accuracy=92.00
## MCC=0.84
## Specificity=0.88
## Sensitivity=1.00

The output of the `classification` function informs us about the SVM setup, the number of Support Vectors and finally shows the predicted labels along with the initial. In this example, there is a validation set different from the training set and therefore we may use the respective labels to obtain the performance characteristics. The relevant function `metrics` called inside the `classification` function, calculates five key measures: the Area Under the ROC curve AUC, the classification accuracy, the Matthews correlation coefficient MCC classification measure, the degree of true negative's identification Specificity, and finally the degree of true positive's identification Sensitivity.

### 3 Advanced usage of the package

#### 3.1 Annotation analysis

For each contemporary chip technology, there is a relevant annotation file, in which the user may drag several genome oriented information. Regarding the breast cancer microarray data, the gene expression values were stored on Affymetrix gene chips. Using the `annotate` function, the user may obtain several info related to probe id, gene symbol, Entrez id, ensembl id, and chromosomal location.

```r
gene.info <- annotate(out.clL2@exemplars, "hgu133plus2.db")
gene.info@results
```

<table>
<thead>
<tr>
<th>PROBEID</th>
<th>SYMBOL</th>
<th>ENTREZID</th>
<th>ENSEMBL</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>215717_s_at</td>
<td>FBN2</td>
<td>2201</td>
<td>ENSG00000138829</td>
<td>5q23.3</td>
</tr>
<tr>
<td>1561358_at</td>
<td>TXLNA</td>
<td>200081</td>
<td>ENSG00000084652</td>
<td>1p35.1</td>
</tr>
<tr>
<td>222752_s_at</td>
<td>TMEM206</td>
<td>55248</td>
<td>ENSG00000065600</td>
<td>1q32.3</td>
</tr>
<tr>
<td>233922_at</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>218871_x_at</td>
<td>CSGALNACT2</td>
<td>55454</td>
<td>ENSG00000169826</td>
<td>10q11.21</td>
</tr>
<tr>
<td>33323_r_at</td>
<td>SFN</td>
<td>2810</td>
<td>ENSG00000175793</td>
<td>1p36.11</td>
</tr>
<tr>
<td>244311_at</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>220932_at</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>205508_at</td>
<td>SCN1B</td>
<td>6324</td>
<td>ENSG00000105711</td>
<td>19q13.1</td>
</tr>
<tr>
<td>209596_at</td>
<td>MXRA5</td>
<td>25878</td>
<td>ENSG00000101825</td>
<td>Xp22.33</td>
</tr>
<tr>
<td>215180_at</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>1560638_a_at</td>
<td>LOC105375839</td>
<td>105375839</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>201852_x_at</td>
<td>CQL3A1</td>
<td>1281</td>
<td>ENSG00000168542</td>
<td>2q31</td>
</tr>
<tr>
<td>229947_at</td>
<td>PI15</td>
<td>51050</td>
<td>ENSG00000137558</td>
<td>8q21.11</td>
</tr>
<tr>
<td>221731_x_at</td>
<td>VCAN</td>
<td>1462</td>
<td>ENSG00000038427</td>
<td>5q14.3</td>
</tr>
</tbody>
</table>

We may exploit the output of the `annotate` function to extent our analysis. For instance, we may perform
*pathway analysis* on the exemplars. For this purpose we will utilize the `probes2pathways` function that utilizes the `reactome.db` package. This function employs the probe ids to identify the relevant pathways.

```r
probes2pathways(gene.info)
##
## 14742281
## "Homo sapiens: Degradation of the extracellular matrix"
## 14742282
## "Homo sapiens: Degradation of the extracellular matrix"
## 15669481
## "Homo sapiens: Elastic fibre formation"
## 15669482
## "Homo sapiens: Elastic fibre formation"
## 14742441
## "Homo sapiens: Extracellular matrix organization"
## 14742442
## "Homo sapiens: Extracellular matrix organization"
## 21293791
## "Homo sapiens: Molecules associated with elastic fibres"
## 21293792
## "Homo sapiens: Molecules associated with elastic fibres"
## 14742281
## "Homo sapiens: Degradation of the extracellular matrix"
## 14742282
## "Homo sapiens: Degradation of the extracellular matrix"
## 15669481
## "Homo sapiens: Elastic fibre formation"
## 15669482
## "Homo sapiens: Elastic fibre formation"
## 14742441
## "Homo sapiens: Extracellular matrix organization"
## 14742442
## "Homo sapiens: Extracellular matrix organization"
## 21293791
## "Homo sapiens: Molecules associated with elastic fibres"
## 21293792
## "Homo sapiens: Molecules associated with elastic fibres"
```
### 3.2 Network characteristics

Regarding the network characteristics, we compute through the `netwAttr` function three different types of centralities (degree, closeness, betweenness) and a measure for clustering coefficient called transitivity. The degree centrality of a node refers to the number of connections or edges of that node to other nodes. The closeness centrality describes the reciprocal accumulated shortest length distance from a node to all other connected nodes. The betweenness centrality depicts the number of times a node intervenes along the shortest path of two other nodes. Transitivity measures the degree of nodes to create clusters within a network. For all four network measures we provide both global and local values. Furthermore, we compose an edge list (Node1-Node2-weight) based on the $N$ top ranked genes. We may exploit that measures to depict the exemplars’ network characteristics.

```r
net.attr <- netwAttr(out.clL2)
wd-degreeL <- net.attr@degree$WdegreeL[out.clL2@exemplars]
wclosenessL <- net.attr@closeness$WclosenessL[out.clL2@exemplars]
wbetweennessL <- net.attr@betweenness$WbetweennessL[out.clL2@exemplars]
wtransitivityL <- net.attr@transitivity$WtransitivityL[out.clL2@exemplars]

Global.val <- c(net.attr@degree$WdegreeG, net.attr@closeness$WclosenessG, net.attr@betweenness$WbetweennessG, net.attr@transitivity$WtransitivityG)

Global.val <- round(Global.val, 2)
exempl.netattr <- rbind(wDegreeL, wclosenessL, wbetweennessL, wtransitivityL)
netAttr <- cbind(Global.val, exempl.netattr)
netAttr <- t(netAttr)
netAttr
```

<table>
<thead>
<tr>
<th></th>
<th>wDegreeL</th>
<th>wclosenessL</th>
<th>wbetweennessL</th>
<th>wtransitivityL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global.val</td>
<td>330.18</td>
<td>0.93</td>
<td>741.81</td>
<td>0.57</td>
</tr>
<tr>
<td>215717_s_at</td>
<td>308.35</td>
<td>1.25</td>
<td>886.00</td>
<td>0.14</td>
</tr>
<tr>
<td>1561358_at</td>
<td>346.92</td>
<td>1.34</td>
<td>1141.00</td>
<td>0.14</td>
</tr>
<tr>
<td>222752_s_at</td>
<td>327.89</td>
<td>0.65</td>
<td>0.00</td>
<td>0.14</td>
</tr>
<tr>
<td>233922_at</td>
<td>317.58</td>
<td>0.79</td>
<td>2.00</td>
<td>0.15</td>
</tr>
<tr>
<td>218871_x_at</td>
<td>293.73</td>
<td>0.53</td>
<td>768.00</td>
<td>0.14</td>
</tr>
<tr>
<td>33323_r_at</td>
<td>338.19</td>
<td>0.27</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>244311_at</td>
<td>294.80</td>
<td>0.63</td>
<td>0.00</td>
<td>0.15</td>
</tr>
<tr>
<td>220932_at</td>
<td>359.10</td>
<td>0.66</td>
<td>0.00</td>
<td>0.14</td>
</tr>
<tr>
<td>205508_at</td>
<td>309.07</td>
<td>0.89</td>
<td>4.00</td>
<td>0.14</td>
</tr>
<tr>
<td>209596_at</td>
<td>345.13</td>
<td>1.34</td>
<td>278.00</td>
<td>0.14</td>
</tr>
<tr>
<td>215180_at</td>
<td>333.37</td>
<td>1.37</td>
<td>1440.00</td>
<td>0.14</td>
</tr>
<tr>
<td>1560638_a_at</td>
<td>368.23</td>
<td>1.38</td>
<td>4615.00</td>
<td>0.14</td>
</tr>
</tbody>
</table>
and identify potential hubs. The calculations of this example are based on the "clr" network reconstruction method. However, there are for the time being two more options, including the "aracne.a" and "aracne.m".

```r
# For local degree > global + standard deviation
sdev <- sd(net.attr$WdegreeL)
msd <- net.attr$WdegreeG + sdev
hubs <- wDegreeL[which(wDegreeL > msd)]
hubs
##  220932_at 1560638_a_at
## 359.10 368.23
```

Finally, we may plot the network for those nodes that their local weighted degree is greater than Global weighted degree plus 2 times the standard deviation. We set this rule for both significance and illustartion purposes (that edge list has dimension 604 x 3).

```r
sdev <- sd(net.attr$WdegreeL)
ms2d <- net.attr$WdegreeG + 2 * sdev
net <- net.attr$WdegreeL[which(net.attr$WdegreeL >
ms2d)]
```
4 Reporting

The overall analysis is summarized in an html report produced by the report function. It covers the dataset representation depicting the samples’ names and their respective class labels, the exemplars section where statistical results and network characteristics are included. The classification performance section illustrates the performance metrics achieved in either cross-validation or hold-out validation. Finally, several annotation info are presented if an annotation analysis has occurred.

5 Session info

sessionInfo()

## R Under development (unstable) (2016-09-29 r71410)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 16.04.1 LTS
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C LC_TIME=en_US.UTF-8
## [4] LC_COLLATE=C LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8 LC_NAME=C LC_ADDRESS=C
## [10] LC_TELEPHONE=C LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats4 parallel stats graphics grDevices utils datasets methods
## [9] base
##
## other attached packages:
## Libraries

```r
## [1] igraph_1.0.1  hgu133plus2.db_3.2.3  org.Hs.eg.db_3.4.0  AnnotationDbi_1.37.0
## [5] IRanges_2.9.0  S4Vectors_0.13.0  mAPKLD ata_1.5.2  mAPKL_1.5.2
## [9] Biobase_2.35.0  BiocGenerics_0.21.0  knitr_1.14
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.7  formatR_1.4  highr_0.6  class_7.3-14
## [5] tools_3.4.0  digest_0.6.10  parmigene_1.0.2  jsonlite_1.1
## [9] evaluate_0.10  RSQ Lite_1.0.0  lattice_0.20-34  Matrix_1.2-7.1
## [13] shiny_0.14.1  DBI_0.5-1  R2HTML_2.3.2  e1071_1.6-7
## [17] apcluster_1.4.3  stringr_1.1.0  cluster_2.0.5  htmlwidgets_0.7
## [21] ade4_1.7-4  multtest_2.31.0  grid_3.4.0  modeest_2.1
## [25] R6_2.2.0  rgl_0.96.0  survival_2.39-5  limma_3.31.0
## [29] reactome.db_1.55.0  magrittr_1.5  htmltools_0.3.5  MASS_7.3-45
## [33] splines_3.4.0  BiocStyle_2.3.1  mime_0.5  xtable_1.8-2
## [37] httpuv_1.3.3  stringi_1.1.2  clusterSim_0.44-5
```

## Reference


### Data Samples

<table>
<thead>
<tr>
<th>Learning Set</th>
<th>Class Labels</th>
<th>Validation Set</th>
<th>Class Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMR134004</td>
<td>0</td>
<td>GMR134105</td>
<td>0</td>
</tr>
<tr>
<td>GMR134690</td>
<td>0</td>
<td>GMR134698</td>
<td>0</td>
</tr>
<tr>
<td>GMR134650</td>
<td>0</td>
<td>GMR134634</td>
<td>0</td>
</tr>
<tr>
<td>GMR134696</td>
<td>0</td>
<td>GMR134697</td>
<td>0</td>
</tr>
<tr>
<td>GMR134659</td>
<td>0</td>
<td>GMR134701</td>
<td>0</td>
</tr>
<tr>
<td>GMR134703</td>
<td>0</td>
<td>GMR134709</td>
<td>0</td>
</tr>
<tr>
<td>GMR134706</td>
<td>0</td>
<td>GMR134711</td>
<td>1</td>
</tr>
<tr>
<td>GMR134708</td>
<td>0</td>
<td>GMR134718</td>
<td>1</td>
</tr>
<tr>
<td>GMR134709</td>
<td>0</td>
<td>GMR134847</td>
<td>1</td>
</tr>
<tr>
<td>GMR134706</td>
<td>0</td>
<td>GMR134892</td>
<td>1</td>
</tr>
<tr>
<td>GMR134701</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMR134704</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMR136977</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMR134951</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMR134707</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMR136965</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Exemplars

<table>
<thead>
<tr>
<th>Examples</th>
<th>Additional Attributes</th>
<th>FC</th>
<th>wt. degree</th>
<th>wt. influence</th>
<th>wt. similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>215172_e</td>
<td>3.939</td>
<td>2.765</td>
<td>315.92</td>
<td>1.97</td>
<td>0.68</td>
</tr>
<tr>
<td>214100_d</td>
<td>3.512</td>
<td>2.149</td>
<td>868.65</td>
<td>1.14</td>
<td>0.68</td>
</tr>
<tr>
<td>233752_d</td>
<td>3.621</td>
<td>0.283</td>
<td>321.69</td>
<td>1.65</td>
<td>0.14</td>
</tr>
<tr>
<td>218973_e</td>
<td>3.584</td>
<td>1.235</td>
<td>317.50</td>
<td>1.97</td>
<td>2</td>
</tr>
<tr>
<td>332122_e</td>
<td>3.587</td>
<td>0.283</td>
<td>338.19</td>
<td>1.46</td>
<td>0.14</td>
</tr>
<tr>
<td>244131_d</td>
<td>3.675</td>
<td>0.59</td>
<td>317.12</td>
<td>1.97</td>
<td>0.14</td>
</tr>
<tr>
<td>225752_d</td>
<td>3.595</td>
<td>0.59</td>
<td>346.12</td>
<td>1.97</td>
<td>0.14</td>
</tr>
<tr>
<td>235106_e</td>
<td>1.002</td>
<td>0.283</td>
<td>330.27</td>
<td>1.97</td>
<td>0.14</td>
</tr>
<tr>
<td>235856_d</td>
<td>1.002</td>
<td>0.283</td>
<td>330.27</td>
<td>1.97</td>
<td>0.14</td>
</tr>
<tr>
<td>235106_e</td>
<td>1.002</td>
<td>0.283</td>
<td>330.27</td>
<td>1.97</td>
<td>0.14</td>
</tr>
<tr>
<td>215100_d</td>
<td>1.002</td>
<td>0.283</td>
<td>330.27</td>
<td>1.97</td>
<td>0.14</td>
</tr>
<tr>
<td>235106_e</td>
<td>1.002</td>
<td>0.283</td>
<td>330.27</td>
<td>1.97</td>
<td>0.14</td>
</tr>
</tbody>
</table>

### Classification Performance

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>Accuracy</th>
<th>MCC</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5-fold validation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Genome Annotation

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>UNIGENE.ID</th>
<th>SYMBOL</th>
<th>ENSEMBL.ID</th>
<th>SYMBOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>P8925</td>
<td>2327</td>
<td>P3910</td>
<td>ENSG00000003825</td>
<td>P3910</td>
</tr>
<tr>
<td>M6539</td>
<td>227152</td>
<td>T2M66</td>
<td>ENSG000000000920</td>
<td>T2M66</td>
</tr>
<tr>
<td>233752_d</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>233752_d</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>233752_d</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>233752_d</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>233752_d</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>233752_d</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>233752_d</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Figure 3: mAPKL analysis report