

Package ‘MM2S’

February 25, 2016

Type Package

Version 1.0.5

Date 2015-02-24

Title Single-Sample Classifier of Medulloblastoma Subtypes for
Medulloblastoma Patient Samples, Mouse Models, and Cell Lines

Description Description: A single-sample classifier that generates Medulloblastoma (MB) subtype predictions for single-samples of human MB patients and model systems, including cell lines and mouse-models. The MM2S algorithm uses a systems-based methodology that facilitates application of the algorithm on samples irrespective of their platform or source of origin. MM2S demonstrates > 96% accuracy for patients of well-characterized normal cerebellum, Wingless (WNT), or Sonic hedgehog (SHH) subtypes, and the less-characterized Group4 (86%) and Group3 (78.2%). MM2S also enables classification of MB cell lines and mouse models into their human counterparts. This package contains function for implementing the classifier onto human data and mouse data, as well as graphical rendering of the results as PCA plots and heatmaps.

Author Deena M.A. Gendoo <deena.gendoo@utoronto.ca>, Benjamin Haibe-Kains <benjamin.haibe.kains@utoronto.ca>

Maintainer Deena M.A. Gendoo <deena.gendoo@utoronto.ca>

Depends GSVA (>= 1.13.1), kknn (>= 1.2-5), parallel, lattice,
pheatmap, R (>= 2.10)

Suggests knitr, MM2Sdata

Imports datasets, grDevices, graphics, stats, utils

VignetteBuilder knitr

License GPL-3

biocViews MM2Sdata, Cancer, Classification, Microarray, Software,
Visualization

URL <http://www.sciencedirect.com/science/article/pii/S0888754315000774>,
<http://www.pmgenomics.ca/bhklab/software/mm2s>

LazyData yes

NeedsCompilation no

Repository CRAN

Date/Publication 2016-02-25 00:39:19

R topics documented:

GTML_Mouse_Preds	2
Human_Cho_Expr	3
MM2S.human	4
MM2S.mouse	5
MM2Ssetup	7
PCARender	8
PredictionsBarplot	9
PredictionsDistributionBoxplot	10
PredictionsDistributionPie	11
PredictionsHeatmap	12
WNT_Mouse_Expr	13

Index	15
--------------	-----------

GTML_Mouse_Preds	<i>Table of Pre-computed MM2S Predictions for 32 replicates of the GTML mouse model</i>
------------------	---

Description

Table containing percent confidence scores for 32 mouse model replicates, across 5 Medulloblastoma subtypes (WNT, SHH, Group3, Group4, and Normal)

Usage

```
data(GTML_Mouse_Preds)
```

Format

A data frame with 32 observations on the following 5 variables.

Group3 a numeric vector

Group4 a numeric vector

SHH a numeric vector

WNT a numeric vector

NORMAL a numeric vector

Details

Gene Expression data of the GTML model was obtained from GSE36594 and normalized as described in (references). The table is an example of MM2S output for that set of samples.

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

See Also

[PredictionsHeatmap](#)

Examples

```
data(GTML_Mouse_Preds)
PredictionsHeatmap(InputMatrix=GTML_Mouse_Preds, pdf_output=TRUE, pdfheight=20, pdfwidth=5)
```

Human_Cho_Expr	<i>Gene Expression data from selected samples of the Human CHO Medulloblastoma dataset</i>
----------------	--

Description

Gene Expression data (formatted and normalized using BrainArray CDFs) for 10 selected samples of the CHO dataset (see references for details).

Usage

```
data(Human_Cho_Expr)
```

Format

A data frame with 12129 observations on the following 10 variables.

MD_269 WNT sample
MD_172 WNT sample
NC_1 Normal Cerebellum
NC_3 Normal Cerebellum
MD_151 SHH sample
MD_153 SHH sample
MD_268 Group4 sample
MD_161 Group4 sample
MD_242 Group3 sample
MD_75 Group3 sample

Source

CHO dataset: Y.J. Cho, A. Tsherniak, P. Tamayo, S. Santagata, A. Ligon, H. Greulich, R. Berhoukim, V. Amani, L. Goumnerova, C.G. Eberhart, C.C. Lau, J.M. Olson, R.J. Gilbertson, A. Gajjar, O. Delattre, M. Kool, K. Ligon, M. Meyerson, J.P. Mesirov, S.L. Pomeroy, Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome, *J Clin Oncol*, 29 (2011) 1424-1430.

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

Examples

```
data(Human_Cho_Expr)
```

MM2S.human

MM2S Prediction of Human Medulloblastoma Samples

Description

This function generates MM2S subtype predictions for human samples of interest. Users are provided the option to save this the predictions as a XLS file.

Usage

```
MM2S.human(InputMatrix, xls_output, parallelize)
```

Arguments

InputMatrix	Matrix containing normalized gene expression data, with genes in rows and samples in columns. Gene IDs MUST be represented as Entres Gene IDs.
xls_output	Option to save MM2S subtype predictions as a XLS file.
parallelize	Option to set number of cores to run ssGSEA calculations in parallel. Default is 1 (no parallelization)

Value

Predictions	MM2S Percent Confidence Predictions of Human Subtypes (Group3, Group4, Normal, SHH, WNT) for a given sample.
MM2S_Subtype	List of Sample names the designated Human MB subtype from MM2S classification.

RankMatrixTesting

ssGSEA rank matrix of the test data, using selected genesets common between the test data and training set. These genesets are based on a filtered list using the top24 common genesets, for each subtype

RankMatrixTraining

ssGSEA rank matrix of the trained MM2S human data, using selected genesets common between the test data and training set. These genesets are based on a filtered list using the top24 common genesets, for each subtype

Author(s)

Deena M.A. Gendoo

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

See Also

[MM2S.mouse](#), [PredictionsHeatmap](#), [PCARender](#)

Examples

```
## load Human gene expression data containing selected samples from GSE37418
# data(Human_Cho_Expr)
# SubtypePreds<-MM2S.human(InputMatrix=Human_Cho_Expr[1:3],xls_output=TRUE,parallelize=2)

#### Not Run
## Generate Heatmap of Predictions
## PredictionsHeatmap(InputMatrix=SubtypePreds$Predictions, pdf_output=TRUE,pdfheight=5,pdfwidth=5)
```

MM2S.mouse

MM2S Prediction of Mouse Medulloblastoma Samples

Description

This function generates MM2S subtype predictions for Mouse samples of interest. Users are provided the option to save this the predictions as a XLS file.

Usage

```
MM2S.mouse(InputMatrix,xls_output,parallelize)
```

Arguments

InputMatrix	Matrix containing normalized gene expression data, with genes in rows and samples in columns. Gene IDs MUST be represented as Entres Gene IDs.
xls_output	Option to save MM2S subtype predictions as a XLS file.
parallelize	Option to set number of cores to run ssGSEA calculations in parallel. Default is 1 (no parallelization)

Value

Predictions	MM2S Percent Confidence Predictions of Human Subtypes (Group3, Group4, Normal, SHH, WNT) for a given Mouse sample.
MM2S_Subtype	List of Sample names the designated Human MB subtype from MM2S classification.
RankMatrixTesting	ssGSEA rank matrix of the test data, using selected genesets common between the test data and training set. These genesets are based on a filtered list using the top24 common genesets, for each subtype
RankMatrixTraining	ssGSEA rank matrix of the trained MM2S human data, using selected genesets common between the test data and training set. These genesets are based on a filtered list using the top24 common genesets, for each subtype

Author(s)

Deena M.A. Gendoo

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

See Also

[MM2S.human,PredictionsHeatmap,PCARender](#)

Examples

```
# Load Mouse gene expression data
# Data contains selected samples from a heterogenous WNT Mouse model
# data(WNT_Mouse_Expr)
# Generate Subtype Predictions
# SubtypePreds<-MM2S.mouse(InputMatrix=WNT_Mouse_Expr[2:3],xls_output=TRUE,parallelize=2)

## Not Run
## Generate Heatmap of Predictions
# PredictionsHeatmap(InputMatrix=SubtypePreds$Predictions, pdf_output=TRUE,pdfheight=5,pdfwidth=5)
```

```
## Generate projections of the selected genesets from Mouse model onto the training set, using PCA
# PCARender(GSVAmatrixTesting=SubtypePreds$RankMatrixTesting,
# GSVAmatrixTraining=SubtypePreds$RankMatrixTraining)
```

MM2Ssetup	<i>MM2S ssGSEA Training Matrix, MB Sample Group information for the training matrix, GMT files for Human and Mouse, and list of geneset names</i>
-----------	---

Description

This dataset contains the GMT files required to generate ssGSEA as part of MM2S along with a list of 674 geneset names that constitute the files, as well as the precomputed ssGSEA matrix on Human training data (347 human samples) as well as a list of the original MB subtype classifications for those samples.

Usage

```
data(MM2Ssetup)
```

Format

[MM2Ssetup](#) is a data file containing four objects:

HumanGMT GMT file 694 Human Genesets, based on Gene Ontology Biological Processes (GO-BP) for MM2S subtype prediction

MouseGMT GMT file 694 Mouse Genesets, based on Gene Ontology Biological Processes (GO-BP) for MM2S subtype prediction

MB_SampleInfo Sample Names and Subtype for 347 Human Samples of the MM2S training set

genesetHuman List of 694 genesets that are common to mouse and human, and which are part of the GMT files

GenesetStatNormal List of 674 genesets and p-values (from the Wilcoxon test) for the Normal Subtype. Of these, the top 24 genesets specific to this group, and which overlap with the testing datasets, are selected for generating the ssGSEA rank matrix

GenesetStatWNT List of 674 genesets and p-values (from the Wilcoxon test) for the WNT Subtype. Of these, the top 24 genesets specific to this group, and which overlap with the testing datasets, are selected for generating the ssGSEA rank matrix

GenesetStatSHH List of 674 genesets and p-values (from the Wilcoxon test) for the SHH Subtype. Of these, the top 24 genesets specific to this group, and which overlap with the testing datasets, are selected for generating the ssGSEA rank matrix

GenesetStatGroup3 List of 674 genesets and p-values (from the Wilcoxon test) for the Group3 Subtype. Of these, the top 24 genesets specific to this group, and which overlap with the testing datasets, are selected for generating the ssGSEA rank matrix

GenesetStatGroup4 List of 674 genesets and p-values (from the Wilcoxon test) for the Group4 Subtype. Of these, the top 24 genesets specific to this group, and which overlap with the testing datasets, are selected for generating the ssGSEA rank matrix

Frozen_ES_Rank_Matrix Ranked list of 674 genesets per sample of the human training set, sorted in descending order of enrichment scores.

Author(s)

Deena M.A. Gendoo

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

Examples

```
data(MM2Ssetup)
```

PCARender

PCA Projections of ssGSEA Rank Matrix for Test Samples Onto the MM2S Training Set

Description

This function projects the ssGSEA ranked matrix for the Test Data onto the ssGSEA ranked matrix of the MM2S training set. The projections are done using ssGSEA ranked matrix containing the selected genesets common to both the Training set and test data.

Usage

```
PCARender(GSVAmatrixTesting,GSVAmatrixTraining)
```

Arguments

GSVAmatrixTesting

Matrix of ranked enrichment scores for the tested datasets or data sample, containing the sample(s) in rows and genesets in columns

GSVAmatrixTraining

Matrix of ranked enrichment scores for the training datasets, containing the samples in rows and genesets in columns

Value

3 PDF files of projected test data onto the MM2S training set, using PCA (Principal Component Analysis) based on the selected genesets: PC1-PC2: Projection of the ssGSEA rank matrix from the testing set onto the training data, first and second principal component PC2-PC3: Projection of the ssGSEA rank matrix from the testing set onto the training data, second and third principal component Lattice: Lattice matrix with the projections onto PCA1-PC3

Author(s)

Deena M.A. Gendoo

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

Examples

```
## Running raw expression data through MM2S
# load Mouse gene expression data for the potential WNT mouse model
# data(WNT_Mouse_Expr)
# SubtypePreds<-MM2S.mouse(InputMatrix=WNT_Mouse_Expr[2:3],xls_output=TRUE,parallelize=1)
## Generate Heatmap
# PCARender(GSVAmatrixTesting=SubtypePreds$RankMatrixTesting,
#           GSVAmatrixTraining=SubtypePreds$RankMatrixTraining)
```

PredictionsBarplot *Stacked Barplot of MM2S Subtype Predictions for Given Samples*

Description

This function generates a stacked barplot of MM2S subtype predictions for samples of interest. Users are provided the option to save this heatmap as a PDF file.

Usage

```
PredictionsBarplot(InputMatrix, pdf_output, pdfheight, pdfwidth)
```

Arguments

InputMatrix	Matrix with samples in rows, and columns with MM2S percentage predictions for each subtype (Gr4,Gr3,SHH,WNT, and Normal)
pdf_output	Option to save the heatmap as a PDF file
pdfheight	User-defined specification for PDF height size
pdfwidth	User-defined specification for PDF width size

Value

Generated Stacked Barplot of MM2S subtype predictions. Samples are in columns. Stacks are reflective of prediction percentages across MB subtypes for a given sample.

Author(s)

Deena M.A. Gendoo

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

Examples

```
# Generate heatmap from already-computed predictions for the GTML Mouse Model
## load computed MM2S predictions for GTML mouse model
data(GTML_Mouse_Preds)
## Generate Barplot
PredictionsBarplot(InputMatrix=GTML_Mouse_Preds, pdf_output=TRUE, pdfheight=5, pdfwidth=5)

## Not Run
# Generate heatmap after running raw expression data through MM2S
# load Mouse gene expression data for the potential WNT mouse model
# data(WNT_Mouse_Expr)
# SubtypePreds<-MM2S.mouse(InputMatrix=WNT_Mouse_Expr[2:3], xls_output=TRUE, parallelize=1)
# Generate Heatmap
# PredictionsBarplot(InputMatrix=SubtypePreds$Predictions, pdf_output=TRUE, pdfheight=5, pdfwidth=5)
```

PredictionsDistributionBoxplot

Boxplot of MM2S Subtype Predictions for Given Samples

Description

This function generates a boxplot of MM2S subtype predictions for samples of interest. Users are provided the option to save this plot as a PDF file.

Usage

```
PredictionsDistributionBoxplot(InputMatrix, pdf_output, pdfheight, pdfwidth)
```

Arguments

InputMatrix	Output object from calling MM2S.Human or MM2S.Mouse predictions
pdf_output	Option to save the boxplot as a PDF file
pdfheight	User-defined specification for PDF height size
pdfwidth	User-defined specification for PDF width size

Value

Generated Pie of all the MM2S subtype predictions in a testing set.

Author(s)

Deena M.A. Gendoo

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

Examples

```
# First run raw expression data through MM2S
# load Mouse gene expression data for the potential WNT mouse model
# data(WNT_Mouse_Expr)
# SubtypePreds<-MM2S.mouse(InputMatrix=WNT_Mouse_Expr,xls_output=TRUE,parallelize=2)
# Generate Heatmap
# PredictionsDistributionBoxplot(InputMatrix=SubtypePreds,pdf_output=TRUE,pdfheight=5,pdfwidth=5)
```

PredictionsDistributionPie

Heatmap of MM2S Subtype Predictions for Given Samples

Description

This function generates a pie chart of MM2S subtype predictions for samples of interest. Users are provided the option to save this chart as a PDF file.

Usage

```
PredictionsDistributionPie(InputMatrix,pdf_output,pdfheight,pdfwidth)
```

Arguments

InputMatrix	Output object from calling MM2S.Human or MM2S.Mouse predictions
pdf_output	Option to save the heatmap as a PDF file
pdfheight	User-defined specification for PDF height size
pdfwidth	User-defined specification for PDF width size

Value

Generated Pie of all the MM2S subtype predictions in a testing set.

Author(s)

Deena M.A. Gendoo

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

Examples

```
# First run raw expression data through MM2S
# load Mouse gene expression data for the potential WNT mouse model
# data(WNT_Mouse_Expr)
# SubtypePreds<-MM2S.mouse(InputMatrix=WNT_Mouse_Expr,xls_output=TRUE,parallelize=2)
# Generate Heatmap
# PredictionsDistributionPie(InputMatrix=SubtypePreds,pdf_output=TRUE,pdfheight=5,pdfwidth=5)
```

PredictionsHeatmap *Heatmap of MM2S Subtype Predictions for Given Samples*

Description

This function generates a graphical heatmap of MM2S subtype predictions for samples of interest. Users are provided the option to save this heatmap as a PDF file.

Usage

```
PredictionsHeatmap(InputMatrix,pdf_output,pdfheight,pdfwidth)
```

Arguments

InputMatrix	Matrix containing the samples in rows and columns containing MM2S percentage predictions for each subtype (Gr4,Gr3,SHH,WNT, and Normal)
pdf_output	Option to save the heatmap as a PDF file
pdfheight	User-defined specification for PDF height size
pdfwidth	User-defined specification for PDF width size

Value

Generated Heatmap of MM2S subtype predictions. Samples are in rows, and prediction percentages are in columns.

Author(s)

Deena M.A. Gendoo

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

Examples

```
# Generate heatmap from already-computed predictions for the GTML Mouse Model
## load computed MM2S predictions for GTML mouse model
data(GTML_Mouse_Preds)
## Generate Heatmap
PredictionsHeatmap(InputMatrix=GTML_Mouse_Preds, pdf_output=TRUE, pdfheight=20, pdfwidth=5)

## Not Run
# Generate heatmap after running raw expression data through MM2S
# load Mouse gene expression data for the potential WNT mouse model
# data(WNT_Mouse_Expr)
# SubtypePreds<-MM2S.mouse(InputMatrix=WNT_Mouse_Expr[2:3], xls_output=TRUE, parallelize=1)
# Generate Heatmap
# PredictionsHeatmap(InputMatrix=SubtypePreds$Predictions,
# pdf_output=TRUE, pdfheight=5, pdfwidth=5)
```

WNT_Mouse_Expr

Gene Expression data pertaining to the GMTL mouse mode

Description

Gene Expression data (formatted and normalized using BrainArray CDFs) for 32 samples of the GTML mouse model (see references for details). These samples were obtained from GSE33199 and GSE24628.

Usage

```
data(WNT_Mouse_Expr)
```

Format

A data frame with 17607 observations on the following 6 variables.

```
GSE33199_GSM821968 a numeric vector
GSE33199_GSM821969 a numeric vector
GSE33199_GSM821970 a numeric vector
GSE24628_GSM607429 a numeric vector
GSE24628_GSM607436 a numeric vector
GSE24628_GSM607444 a numeric vector
```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33199> <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2>

References

Gendoo, D. M., Smirnov, P., Lupien, M. & Haibe-Kains, B. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics*, doi:10.1016/j.ygeno.2015.05.002 (2015)

Manuscript URL: <http://www.sciencedirect.com/science/article/pii/S0888754315000774>

Examples

```
data(WNT_Mouse_Expr)
```

Index

- *Topic **PCA**
 - PCARender, [8](#)
- *Topic **classification**
 - MM2S.human, [4](#)
 - MM2S.mouse, [5](#)
- *Topic **datasets**
 - GTML_Mouse_Preds, [2](#)
 - Human_Cho_Expr, [3](#)
 - WNT_Mouse_Expr, [13](#)
- *Topic **data**
 - MM2Ssetup, [7](#)
- *Topic **heatmap**
 - PredictionsBarplot, [9](#)
 - PredictionsDistributionBoxplot, [10](#)
 - PredictionsDistributionPie, [11](#)
 - PredictionsHeatmap, [12](#)
- Frozen_ES_Rank_Matrix (MM2Ssetup), [7](#)
- genesetHuman (MM2Ssetup), [7](#)
- GenesetStatGroup3 (MM2Ssetup), [7](#)
- GenesetStatGroup4 (MM2Ssetup), [7](#)
- GenesetStatNormal1 (MM2Ssetup), [7](#)
- GenesetStatSHH (MM2Ssetup), [7](#)
- GenesetStatWNT (MM2Ssetup), [7](#)
- GTML_Mouse_Preds, [2](#)
- Human_Cho_Expr, [3](#)
- HumanGMT (MM2Ssetup), [7](#)
- MB_SampleInfo (MM2Ssetup), [7](#)
- MM2S.human, [4](#), [6](#)
- MM2S.mouse, [5](#), [5](#)
- MM2Ssetup, [7](#), [7](#)
- MouseGMT (MM2Ssetup), [7](#)
- PCARender, [5](#), [6](#), [8](#)
- PredictionsBarplot, [9](#)
- PredictionsDistributionBoxplot, [10](#)
- PredictionsDistributionPie, [11](#)
- PredictionsHeatmap, [3](#), [5](#), [6](#), [12](#)
- WNT_Mouse_Expr, [13](#)