Package 'TBSignatureProfiler'

March 30, 2021

Title Profile RA-Seq Data Using TB Pathway Signatures

Version 1.2.0

Description Signatures of TB progression, TB disease, and other TB disease states have been created. This package makes it easy to profile RNA-Seq data using these signatures and common signature profiling tools including ASSIGN, GSVA, and ssGSEA.

License MIT + file LICENSE

Encoding UTF-8

biocViews GeneExpression, DifferentialExpression

LazyData true

RoxygenNote 7.1.0

Depends R (>= 4.0.0)

Imports ASSIGN (>= 1.23.1), GSVA, SummarizedExperiment, S4Vectors, methods, BiocParallel, ComplexHeatmap, RColorBrewer, ggplot2, reshape2, circlize, glmnet, ROCit, bioDist, readr, boot, DESeq2, caret, ggfortify, e1071, DT, edgeR, singscore, gdata, shiny

Suggests testthat, spelling, lintr, covr, knitr, rmarkdown, BiocStyle

URL https://github.com/compbiomed/TBSignatureProfiler

BugReports https://github.com/compbiomed/TBSignatureProfiler/issues

Language en-US

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/TBSignatureProfiler

git_branch RELEASE_3_12

git_last_commit 16ccdd3

git_last_commit_date 2020-10-27

Date/Publication 2021-03-29

Author David Jenkins [aut, cre],
Yue Zhao [aut],
W. Evan Johnson [aut],
Aubrey Odom [aut],
Christian Love [aut]

Maintainer David Jenkins <dfj@bu.edu>

R topics documented:

`bootstrapAUC` *Bootstrap the AUC and conduct T-Tests for a collection of signatures.*

Description

Run bootstrapping of the AUC and derive the p-value for a 2-sample t-test for all signatures tested on a given dataset.

Usage

```
bootstrapAUC(
  SE_scored,
  annotationColName,
  signatureColNames,
  num.boot = 100,
  pb.show = TRUE
)
```

Arguments

SE_scored	a SummarizedExperiment object with genes as the row features and signature scores in the colData. There should also be a column of annotation data. Required.
annotationColName	a character string giving the column name in colData that contains the annotation data. Required.
signatureColNames	a vector of column names in the colData that contain the signature score data. Required.
num.boot	integer. The number of times to bootstrap the data. The default is 100.
pb.show	logical for whether to show a progress bar while running code. The default is TRUE.

Value

A list of length 3 returning a vector of p-values for a 2-sample t-test, bootstrapped AUC values, and an AUC value for using all scored values for all signatures specified in signatureColNames.

Examples

```
# Run signature profiling
choose_sigs <- TBsignatures[1]
prof_indian <- runTBsigProfiler(TB_indian, useAssay = "logcounts",
                                algorithm = "ssGSEA",
                                combineSigAndAlgorithm = TRUE,
                                signatures = choose_sigs,
                                parallel.sz = 1)
# Bootstrapping
booted <- bootstrapAUC(SE_scored = prof_indian, annotationColName = "label",
                       signatureColNames = names(choose_sigs), num.boot = 5)
booted
```

Bootstrap_LOOCV_LR_AUC

Bootstrap on Leave-one-out CV with Logistic Regression.

Description

Bootstrap on Leave-one-out CV with Logistic Regression.

Usage

```
Bootstrap_LOOCV_LR_AUC(df, targetVec, nboot)
```

Arguments

df	a data.frame of gene expression count data. Required.
targetVec	a binary vector of the response variable. Should be the same number of rows as df. Required.
nboot	an integer specifying the number of bootstrap iterations.

Value

A list of length 2 with elements

auc	A vector the length of nboot with the AUC from each bootstrap iteration.
byClass	A dataframe with number of rows equal to nboot. Each row contains the sensitivity, specificity, positive predictive value, negative predictive value, precision, recall, F1, prevalence, detection rate, detection prevalence and balanced accuracy for that bootstrap iteration.

---

common_sigAnnotData *Annotation information for published TB signatures.*

Description

A data.frame of annotation information for published tuberculosis signatures. This table differs from that of sigAnnotData as it refers to signatures via the name given in scientific publications, and via a consistent naming system otherwise. Currently, this table includes two variables, disease and tissue type.

Usage

```
common_sigAnnotData
```

Format

data.frame

Details

The disease variable indicates whether the signature was developed to distinguish TB from LTBI ("Disease"), TB from some combination of other diseases and possibly LTBI ("OD"), TB from Human Immunodeficiency Virus ("HIV"), TB from pneumonia ("PNA"), or identify risk of progression to TB ("RISK"), risk of TB treatment failure ("FAIL"), or classify treatment responses (i.e., failures from cures, "RES").

The tissue type variable denotes whether the signature was developed using samples of either whole blood/paxgene or peripheral blood mononuclear cells (PBMCs). Due to the manipulation of cells inherently required to obtain PBMCs, many scientists prefer to use only whole blood samples for analysis.

Source

- **Anderson_42**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. 10.1056/NEJMoa1303657
- **Anderson_OD_51**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. 10.1056/NEJMoa1303657

- **Berry_393**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. 10.1038/nature09247
- **Berry_OD_86**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. 10.1038/nature09247
- **Blankley_5**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. 10.1183/13993003.02121-2015
- **Blankley_380**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. 10.1183/13993003.02121-2015
- **Bloom_OD_144**: Bloom, Chloe I., Christine M. Graham, Matthew P. R. Berry, Fotini Rozakeas, Paul S. Redford, Yuanyuan Wang, Zhaohui Xu, et al. 2013. "Transcriptional Blood Signatures Distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung Cancers." PloS One 8 (8): e70630. 10.1371/journal.pone.0070630
- **Esmail_82**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Esmail_203**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Esmail_893**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Gliddon_OD_3**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org: . 10.1101/583674
- **Gliddon_OD_4**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J.,

Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org: . 10.1101/583674

- **Jacobsen_3**: Jacobsen, Marc, Dirk Repsilber, Andrea Gutschmidt, Albert Neher, Knut Feldmann, Hans J. Mollenkopf, Andreas Ziegler, and Stefan H. E. Kaufmann. 2007. "Candidate Biomarkers for Discrimination between Infection and Disease Caused by Mycobacterium Tuberculosis." Journal of Molecular Medicine 85 (6): 613-21. 10.1007/s00109-007-0157-6
- **Kaforou_27**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Kaforou_OD_44**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Kaforou_OD_53**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Lee_4**: Lee, Shih-Wei, Lawrence Shih-Hsin Wu, Guan-Mau Huang, Kai-Yao Huang, Tzong-Yi Lee, and Julia Tzu-Ya Weng. 2016. "Gene Expression Profiling Identifies Candidate Biomarkers for Active and Latent Tuberculosis." BMC Bioinformatics 17 Suppl 1 (January): 3. 10.1186/s12859-015-0848-x
- **PREDICT29**: Leong, S., Zhao, Y., Ribeiro-Rodrigues, R., Jones-López, E. C., Acuña-Villaorduña, C., Rodrigues, P. M., Palaci, M., Alland, D., Dietze, R., Ellner, J. J., Johnson, W. E., Salgame, P., Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB. Tuberculosis (Edinb). 2020 Jan;120:101898. 10.1016/j.tube.2020.101898
- **DIAG4**: Maertzdorf, Jeroen, Gayle McEwen, January Weiner 3rd, Song Tian, Eric Lader, Ulrich Schriek, Harriet Mayanja-Kizza, Martin Ota, John Kenneth, and Stefan He Kaufmann. 2016. "Concise Gene Signature for Point-of-Care Classification of Tuberculosis." EMBO Molecular Medicine 8 (2): 86-95. 10.15252/emmm.201505790
- **Maertzdorf_OD_100**: Maertzdorf, Jeroen, January Weiner 3rd, Hans-Joachim Mollenkopf, TBornot TB Network, Torsten Bauer, Antje Prasse, Joachim Müller-Quernheim, and Stefan H. E. Kaufmann. 2012. "Common Patterns and Disease-Related Signatures in Tuberculosis and Sarcoidosis." Proceedings of the National Academy of Sciences of the United States of America 109 (20): 7853-58. 10.1073/pnas.1121072109
- **Rajan_HIV_5**: Rajan, Jayant V., Semitala, Fred C., Kamya, Moses R., Yoon, Christina., Mehta, Tejas., Cattamanchi, Adithya., Seielstad, Mark., Montalvo, Lani., Andama, Alfred., Katende, Jane., Asege, Lucy., Nakaye, Martha., Mwebe, Sandra. 2018 "A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus" Clinical Infectious Diseases: 1-7. 10.1093/cid/ciy835
- **Roe_3**: Roe, Jennifer, Venturini, Cristina, Gupta, Rishi K., Gurry, Celine, Chain, Benjamin M., Sun, Yuxin, Southern, Jo, Jackson, Charlotte, Lipman, Marc, C., Miller, Robert F., Martineau, Adrian R., Abubakar, Ibrahim, Noursadeghi, Mahdad. 2019 "T1 Blood transcriptomic stratification of short-term risk in contacts of tuberculosis": . 10.1093/cid/ciz252
- **Roe_OD_4**: Roe, Jennifer K., Niclas Thomas, Eliza Gil, Katharine Best, Evdokia Tsaliki, Stephen Morris-Jones, Sian Stafford, et al. 2016. "Blood Transcriptomic Diagnosis of Pulmonary and Extrapulmonary Tuberculosis." JCI Insight 1 (16): e87238. 10.1172/jci.insight.87238

- **Sambarey_HIV_10**: Sambarey, Awanti, Abhinandan Devaprasad, Abhilash Mohan, Asma Ahmed, Soumya Nayak, Soumya Swaminathan, George D'Souza, et al. 2017. "Unbiased Identification of Blood-Based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks." EBioMedicine 15 (February): 112-26. 10.1016/j.ebiom.2016.12.009
- **Singhania_OD_20**: Singhania, Akul, Raman Verma, Christine M. Graham, Jo Lee, Trang Tran, Matthew Richardson, Patrick Lecine, et al. 2018. "A Modular Transcriptional Signature Identifies Phenotypic Heterogeneity of Human Tuberculosis Infection." Nature Communications 9 (1): 2308. 10.1038/s41467-018-04579-w
- **Sloot_HIV_2**: Sloot, Rosa, Maarten F. Schim van der Loeff, Erik W. van Zwet, Mariëlle C. Haks, Sytze T. Keizer, Maarten Scholing, Tom H. M. Ottenhoff, Martien W. Borgdorff, and Simone A. Joosten. 2015. "Biomarkers Can Identify Pulmonary Tuberculosis in HIV-Infected Drug Users Months Prior to Clinical Diagnosis." EBioMedicine 2 (2): 172-79. 10.1016/j.ebiom.2014.12.001
- **Suliman_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. 10.1164/rccm.201711-2340OC
- **RISK4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. 10.1164/rccm.201711-2340OC
- **DIAG3**: Sweeney, Timothy E., Lindsay Braviak, Cristina M. Tato, and Purvesh Khatri. 2016. "Genome-Wide Expression for Diagnosis of Pulmonary Tuberculosis: A Multicohort Analysis." The Lancet. Respiratory Medicine 4 (3): 213-24. 10.1016/S2213-2600(16)00048-5
- **DISEASE**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **FAILURE**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **RESPONSE5**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **Walter_51**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Walter_PNA_47**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Walter_PNA_119**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15

- **ACS_COR**: Zak, Daniel E., Adam Penn-Nicholson, Thomas J. Scriba, Ethan Thompson, Sara Suliman, Lynn M. Amon, Hassan Mahomed, et al. 2016. "A Blood RNA Signature for Tuberculosis Disease Risk: A Prospective Cohort Study." The Lancet 387 (10035): 231222. 10.1016/S0140-6736(15)01316-1

Examples

```
data("common_sigAnnotData")
```

`compareAlgs` *Compare scoring algorithms on a single signature via heatmap or boxplot.*

Description

It may be useful to compare the results of scoring across several different scoring algorithms via a method of visualization, such as a heatmap. The `compareSigs` function allows the input of a SummarizedExperiment data object and conducts profiling on each signature desired, and outputting a heatmap or boxplot for each signature.

Usage

```
compareAlgs(
  input,
  signatures = NULL,
  annotationColName,
  useAssay = "counts",
  algorithm = c("GSVA", "ssGSEA", "ASSIGN", "PLAGE", "Zscore", "singscore"),
  showColumnNames = TRUE,
  showRowNames = TRUE,
  scale = FALSE,
  colorSets = c("Set1", "Set2", "Set3", "Pastel1", "Pastel2", "Accent", "Dark2",
    "Paired"),
  choose_color = c("blue", "gray95", "red"),
  colList = list(),
  show.pb = FALSE,
  parallel.sz = 0,
  output = "heatmap",
  num.boot = 100
)
```

Arguments

`input`	an input data object of the class `"SummarizedExperiment"`. Required.
`signatures`	a `list` of signatures to run with their associated genes. This list should be in the same format as `TBsignatures`, included in the TBSignatureProfiler package. If `signatures = NULL`, the default set of signatures `TBsignatures` list is used. For details, run `?TBsignatures`. The default is `NULL`.
`annotationColName`	a character string giving the column name in `colData` that contains the annotation data. Required.

useAssay	a character string specifying the assay to use for signature profiling when input is a SummarizedExperiment. Required only for input data of the class SummarizedExperiment. If null, the assay used will be "counts". The default is NULL.
algorithm	a vector of algorithms to run, or character string if only one is desired. The default is c("GSVA","ssGSEA","ASSIGN","PLAGE","Zscore","singscore").
showColumnNames	logical. Setting showColumnNames = TRUE will show the column names (i.e. sample names) on the heatmap. The default is TRUE.
showRowNames	logical. Setting showColumnNames = TRUE will show the row names (i.e. signature names) on the heatmap. The default is TRUE.
scale	logical. Setting scale = TRUE scales the signature data. The default is FALSE.
colorSets	a vector of names listing the color sets in the order that they should be used in creating the heatmap. By default, this function will use the color sets in the order listed in Usage for annotation information. You may replace the default with the same collection of sets in order that you want to use them, or provide custom color sets with the colList parameter.
choose_color	a vector of color names to be interpolated for the heatmap gradient, or a colorRamp function produced by circlize::colorRamp2. The default is c("blue","gray95","red").
colList	a named list of named vectors specifying custom color information to pass to ComplexHeatmap::Heatmap(). The list should have as many elements as there are annotation columns, and each element name should correspond exactly with the name of each annotation column. The colors in the vector elements should be named according to the levels of the factor in that column's annotation data if the annotation is discrete, or it should be produced with circlize::colorRamp2 if the annotation is continuous. By default, ColorBrewer color sets will be used. See the the parameter colorSets for additional details.
show.pb	logical, whether warnings and other output from the profiling should be suppressed (including progress bar output). Default is FALSE.
parallel.sz	an integer identifying the number of processors to use when running the calculations in parallel for the GSVA and ssGSEA algorithms. If parallel.sz = 0, all cores are used. The default is 0.
output	a character string specifying whether the outputted plot should be a "heatmap" or "boxplot". The default is "heatmap".
num.boot	an integer indicating the number of times to bootstrap the data.

Value

A heatmap or boxplot for each signature specified comparing the enumerated algorithms.

Examples

```
# Example using the TB_hiv data set, two signatures, and 3 algorithms
data("TB_hiv")
compareAlgs(TB_hiv, signatures = TBsignatures[c(1,2)],
            annotationColName = "Disease",
            algorithm = c("GSVA", "ssGSEA", "PLAGE"),
            scale = TRUE, parallel.sz = 1, output = "heatmap")
```

compareBoxplots *Create a comparison plot of boxplots for bootstrapped AUC values.*

Description

Present the results of AUC bootstrapping for a collection of scored signatures via boxplots.

Usage

```
compareBoxplots(
  SE_scored,
  annotationColName,
  signatureColNames,
  num.boot = 100,
  name = "Boxplot Comparison of Signature AUCs",
  pb.show = TRUE,
  abline.col = "red",
  fill.col = "gray79",
  outline.col = "black",
  rotateLabels = FALSE
)
```

Arguments

SE_scored	a SummarizedExperiment object with genes as the row features and signature scores in the colData. There should also be a column of annotation data. Required.
annotationColName	a character string giving the column name in colData that contains the annotation data. Required.
signatureColNames	a vector of column names in the colData that contain the signature score data. Required.
num.boot	an integer indicating the number of times to bootstrap the data.
name	a character string giving the overall title for the plot. The default is "Boxplot Comparison of Signature AUCs".
pb.show	logical for whether to show a progress bar while running code. Default is TRUE.
abline.col	the color to be used for the dotted line at AUC = 0.5 (the chance line). The default is "red".
fill.col	the color to be used to fill the boxplots. The default is "white".
outline.col	the color to be used for the boxplot outlines. The default is "black".
rotateLabels	If TRUE, rotate labels. Default is FALSE.

Value

A plot with side-by-side boxplots of bootstrapped AUC values for each specified signature.

Examples

```
# Run signature profiling
 choose_sigs <- TBsignatures[c(1, 2)]
 prof_indian <- runTBsigProfiler(TB_indian[seq_len(25), ],
                                 useAssay = "logcounts",
                                 algorithm = "ssGSEA",
                                 signatures = choose_sigs,
                                 parallel.sz = 1)
 # Create boxplots
 compareBoxplots(prof_indian, annotationColName = "label",
                 signatureColNames = names(choose_sigs), rotateLabels = TRUE)
```

---

deseq2_norm_rle *Normalize gene expression count data.*

Description

Normalize gene expression count data.

Usage

```
deseq2_norm_rle(inputData)
```

Arguments

inputData	a data.frame or matrix of gene expression count data. Required.

Value

A data.frame or matrix of normalized count data.

Examples

```
## Example using the counts assay from a SummarizedExperiment
data_in <- SummarizedExperiment::assay(TB_indian, "counts")
res <- deseq2_norm_rle(data_in)
```

---

distinctColors *Generate a distinct palette for coloring different clusters.*

Description

Create a distinct palette for coloring different heatmap clusters. The function returns colors for input into ComplexHeatmap:Heatmap(), signatureGeneHeatmap() and signatureHeatmap().

Usage

```
distinctColors(
  n,
 hues = c("red", "cyan", "orange", "blue", "yellow", "purple", "green", "magenta"),
  saturation.range = c(0.7, 1),
  value.range = c(0.7, 1)
)
```

Arguments

n	an integer describing the number of colors to generate. Required.
hues	a vector of character strings indicating the R colors available from the colors() function. These will be used as the base colors for the clustering scheme. Different saturations and values (i.e. darkness) will be generated for each hue. Default is c("red","cyan","orange","blue","yellow","purple","green","magenta")
saturation.range	a numeric vector of length 2 with values between 0 and 1 giving the range of saturation. The default is c(0.25,1).
value.range	a numeric vector of length 2 with values between 0 and 1 giving the range of values. The default is c(0.5,1).

Value

A vector of distinct colors that have been converted to HEX from HSV.

Examples

```
distinctColors(10)
```

LOOAUC_simple_multiple_noplot_one_df

Perform Leave-one-out CV with Logistic Regression.

Description

Perform Leave-one-out CV with Logistic Regression.

Usage

```
LOOAUC_simple_multiple_noplot_one_df(df, targetVec)
```

Arguments

df	a data.frame of gene expression count data. Required.
targetVec	a binary vector of the response variable. Should be the same number of rows as df. Required.

Value

A list of length 3 with elements

auc	The AUC from the LOOCV procedure.
byClass	A vector containing the sensitivity, specificity, positive predictive value, negative predictive value, precision, recall, F1, prevalence, detection rate, detection prevalence and balanced accuracy.
prob	A vector of the test prediction probabilities.

---

mkAssay *Add SummarizedExperiment assays to the data structure.*

Description

Given an input of a Summarized Experiment with a counts or CPM assay, This function creates additional assays for a gene expression count dataset to be used in further analysis.

Usage

```
mkAssay(
  SE_obj,
  input_name = "counts",
  output_name = NULL,
  log = FALSE,
  counts_to_CPM = TRUE,
  prior_counts = 3
)
```

Arguments

SE_obj	a SummarizedExperiment object containing gene expression data. Required.
input_name	a character string specifying the name of the assay to be referenced for creating additional assays. Default is "counts".
output_name	a character string to concatenate to "log" when computing a log assay. If NULL, then input_name will be substituted. Only used if log = TRUE. Default is NULL.
log	logical. Indicate whether an assay returned should be the log of whichever assay is specified in "output_name". If counts_to_CPM = TRUE as well, then a log CPM assay will also be created. Default is FALSE.
counts_to_CPM	logical. This argument only applies if the input_type is a counts assay. If TRUE, then the output assays will include a normalized CPM assay. If log = TRUE as well, then a log CPM assay will also be created. Default is TRUE.
prior_counts	an integer specifying the average count to be added to each observation to avoid taking the log of zero. Used only if log = TRUE. The default is 3.

Value

This function returns a SummarizedExperiment object with up to 3 additional assay types attached to the original inputted object.

cpm	Counts per million
logcpm	Log counts per million
log_<output_name>	Log of original inputted assay. <output_name> will be replaced by inputted parameter.

Author(s)

Aubrey Odom

Examples

```
# Create a log assay of the original assay input
# TB_hiv dataset already has counts data
log_only <- mkAssay(TB_hiv, log = TRUE, counts_to_CPM = FALSE)
log_only

# Create a CPM assay
CPM_only <- mkAssay(TB_hiv)
CPM_only

# Create a logCPM, logcounts, and CPM assay
all_assays <- mkAssay(TB_hiv, log = TRUE)
all_assays
```

---

plotQuantitative *Create a boxplot using logistic regression and bootstrap LOOCV to evaluate signatures.*

Description

This function takes as input a data.frame with genetic expression count data, and uses a bootstrapped leave-one-out cross validation procedure with logistic regression to allow for numeric and graphical comparison across any number of genetic signatures. It creates a boxplot of bootstrapped AUC values.

Usage

```
plotQuantitative(
  df.input,
  targetVec.num,
  signature.list = NULL,
  signature.name.vec = NULL,
  num.boot = 100,
  pb.show = TRUE,
  name = "Signature Evaluation: Bootstrapped AUCs",
  fill.col = "white",
```

```
  outline.col = "black",
  abline.col = "red",
  rotateLabels = FALSE
)
```

Arguments

df.input	a data.frame of gene expression count data. Required.
targetVec.num	a numeric binary vector of the response variable. The vector should be the same number of rows as df. Required.
signature.list	a list of signatures to run with their associated genes. This list should be in the same format as TBsignatures, included in the TBSignatureProfiler package. If signature.list = NULL, the default set of signatures TBsignatures list is used. For details, run ?TBsignatures.
signature.name.vec	A vector specifying the names of the signatures to be compared. This should be the same length as signature.list. If signature.name.vec = NULL, the default set of signatures TBsignatures list is used.
num.boot	an integer specifying the number of bootstrap iterations.
pb.show	logical. If TRUE then a progress bar for the bootstrapping procedure will be displayed as output. The default is TRUE.
name	a character string giving a name for the outputted boxplot of bootstrapped AUCs. The default is "Signature Evaluation: Bootstrapped AUCs".
fill.col	the color to be used to fill the boxplots. The default is "white".
outline.col	the color to be used for the boxplot outlines. The default is "black".
abline.col	the color to be used for the dotted line at AUC = 0.5 (the chance line). The default is "red".
rotateLabels	logical. If TRUE, the x-axis labels will be rotated. The default is FALSE.

Value

a boxplot comparing the bootstrapped AUCs of inputted signatures

Examples

```
inputTest <- matrix(rnorm(1000), 100, 20,
                    dimnames = list(paste0("gene", seq.int(1, 100)),
                                    paste0("sample", seq.int(1, 20))))
inputTest <- as.data.frame(inputTest)
targetVec <- sample(c(0,1), replace = TRUE, size = 20)
signature.list <- list(sig1 = c("gene1", "gene2", "gene3"),
                       sig2 = c("gene4", "gene5", "gene6"))
signature.name.vec <- c("sig1", "sig2")
num.boot <- 5
plotQuantitative(inputTest, targetVec.num = targetVec,
                 signature.list = signature.list,
                 signature.name.vec = signature.name.vec,
                 num.boot = num.boot, rotateLabels = FALSE)
```

`runTBsigProfiler` *Run TB gene signature profiling.*

Description

Using some subset of the signatures listed in `TBsignatures` and specified scoring algorithms, this function runs gene signature profiling on an input gene expression dataset. It allows for scores to be computed for these signatures which can be compared using various visualization tools also provided in the TBSignatureProfiler package.

Usage

```
runTBsigProfiler(
  input,
  useAssay = NULL,
  signatures = NULL,
  algorithm = c("GSVA", "ssGSEA", "ASSIGN", "PLAGE", "Zscore", "singscore"),
  combineSigAndAlgorithm = FALSE,
  assignDir = NULL,
  outputFormat = NULL,
  parallel.sz = 0,
  ASSIGNiter = 1e+05,
  ASSIGNburnin = 50000
)
```

Arguments

`input`	an input data object of the class `SummarizedExperiment`, `data.frame`, or `matrix` containing gene expression data. Required.
`useAssay`	a character string specifying the assay to use for signature profiling when `input` is a `SummarizedExperiment`. Required only for input data of the class `SummarizedExperiment`. If null, the assay used will be `"counts"`. The default is `NULL`.
`signatures`	a `list` of signatures to run with their associated genes. This list should be in the same format as `TBsignatures`, included in the TBSignatureProfiler package. If `signatures = NULL`, the default set of signatures `TBsignatures` list is used. For details, run `?TBsignatures`. The default is `NULL`.
`algorithm`	a vector of algorithms to run, or character string if only one is desired. The default is `c("GSVA","ssGSEA","ASSIGN","PLAGE","Zscore","singscore")`.
`combineSigAndAlgorithm`	logical, not supported if `input` is a SummarizedExperiment object (in which case, the default is `TRUE`). For a matrix or data frame, if `TRUE`, the row names will be in the form <algorithm>_<signature>. If `FALSE`, there will be a column named 'algorithm' that lists which algorithm is used, and a column named 'pathway' that lists the signature profiled. If `NULL`, and one algorithm was used, the algorithm will not be listed. The default is `FALSE`.
`assignDir`	a character string naming a directory to save intermediate ASSIGN results if `algorithm` specifies `"ASSIGN"`. The default is `NULL`, in which case intermediate results will not be saved.

outputFormat	a character string specifying the output data format. Possible values are "SummarizedExperiment", "matrix", or "data.frame". The default is to return the same type as the input object.
parallel.sz	an integer identifying the number of processors to use when running the calculations in parallel for the GSVA and ssGSEA algorithms. If parallel.sz = 0, all cores are used. The default is 0.
ASSIGNiter	an integer indicating the number of iterations to use in the MCMC for the ASSIGN algorithm. The default is 100,000.
ASSIGNburnin	an integer indicating the number of burn-in iterations to use in the MCMC for the ASSIGN algorithm. These iterations are discarded when computing the posterior means of the model parameters. The default is 50,000.

Value

A SummarizedExperiment object, data.frame, or matrix of signature profiling results. The returned object will be of the format specified in outputFormat. If input is a SummarizedExperiment and outputFormat = "SummarizedExperiment", then the output will retain any input information stored in the input colData. In general, if outputFormat = "SummarizedExperiment" then columns in the colData will include the scores for each desired signature with samples on the rows. If input is a data.frame or matrix, then the returned object will have signatures on the rows and samples on the columns.

Source

Profiling for the Z-Score, PLAGE, GSVA, ssGSEA algorithms are all conducted with the Bioconductor GSVA package. Profiling for the singscore algorithm is conducted with the Bioconductor singscore package.

References

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-112. doi: 10.1038/nature08460.

Foroutan, M. et al. (2018). Single sample scoring of molecular phenotypes. *BMC Bioinformatics*, **19**. doi: 10.1186/s12859-018-2435-4.

Lee, E. et al. (2008). Inferring pathway activity toward precise disease classification. *PLoS Comp Biol*, 4(11):e1000217. doi: 10.1371/journal.pcbi.1000217

Shen, Y. et al. (2015). ASSIGN: context-specific genomic profiling of multiple heterogeneous biological pathways. *Bioinformatics*, **31**, 1745-1753. doi: 10.1093/bioinformatics/btv031.

Subramanian, A. et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, **102**, 15545-15550. doi: 10.1073/pnas.0506580102.

Tomfohr, J. et al. (2005). Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics*, 6:225. doi: 10.1186/1471-2105-6-225

Examples

```
## Using a data.frame input/output
 # Create some toy data to test Zak_RISK_16 signature, using 5 samples with low
 # expression & five samples with high expression of the signatures genes.
df_testdata <- as.data.frame(rbind(matrix(c(rnorm(80), rnorm(80) + 5), 16, 10,
                            dimnames = list(TBsignatures$Zak_RISK_16,
```

```
                            paste0("sample", seq_len(10)))),
                      matrix(rnorm(1000), 100, 10,
                            dimnames = list(paste0("gene", seq_len(100)),
                            paste0("sample", seq_len(10))))))
res <- runTBsigProfiler(input = df_testdata,
                        signatures = TBsignatures["Zak_RISK_16"],
                        algorithm = c("GSVA", "ssGSEA"),
                        combineSigAndAlgorithm = FALSE,
                        parallel.sz = 1)
subset(res, res$pathway == "Zak_RISK_16")

## Using a SummarizedExperiment input/output
 # The TB_indian SummarizedExperiment data is included in the package.
GSVA_res <- runTBsigProfiler(input = TB_indian,
                            useAssay = "logcounts",
                            signatures = TBsignatures["Zak_RISK_16"],
                            algorithm = c("GSVA"),
                            combineSigAndAlgorithm = FALSE,
                            parallel.sz = 1)
GSVA_res$Zak_RISK_16
```

`sigAnnotData` *Annotation information for published TB signatures.*

Description

A data.frame of annotation information for published tuberculosis signatures. Currently, this table includes two variables, `disease` and `tissue type`.

Usage

```
sigAnnotData
```

Format

data.frame

Details

The `disease` variable indicates whether the signature was developed to distinguish TB from LTBI ("Disease"), TB from some combination of other diseases and possibly LTBI ("OD"), TB from Human Immunodeficiency Virus ("HIV"), TB from pneumonia ("PNA"), or identify risk of progression to TB ("RISK"), risk of TB treatment failure ("FAIL"), or classify treatment responses (i.e., failures from cures, "RES").

The `tissue type` variable denotes whether the signature was developed using samples of either whole blood/paxgene or peripheral blood mononuclear cells (PBMCs). Due to the manipulation of cells inherently required to obtain PBMCs, many scientists prefer to use only whole blood samples for analysis.

Source

- **Anderson_42**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. 10.1056/NEJMoa1303657
- **Anderson_OD_51**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. 10.1056/NEJMoa1303657
- **Berry_393**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. 10.1038/nature09247
- **Berry_OD_86**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. 10.1038/nature09247
- **Blankley_5**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. 10.1183/13993003.02121-2015
- **Blankley_380**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. 10.1183/13993003.02121-2015
- **Bloom_OD_144**: Bloom, Chloe I., Christine M. Graham, Matthew P. R. Berry, Fotini Rozakeas, Paul S. Redford, Yuanyuan Wang, Zhaohui Xu, et al. 2013. "Transcriptional Blood Signatures Distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung Cancers." PloS One 8 (8): e70630. 10.1371/journal.pone.0070630
- **Esmail_82**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Esmail_203**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Esmail_893**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115

- **Gliddon_OD_3**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org:10.1101/583674
- **Gliddon_OD_4**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org:10.1101/583674
- **Jacobsen_3**: Jacobsen, Marc, Dirk Repsilber, Andrea Gutschmidt, Albert Neher, Knut Feldmann, Hans J. Mollenkopf, Andreas Ziegler, and Stefan H. E. Kaufmann. 2007. "Candidate Biomarkers for Discrimination between Infection and Disease Caused by Mycobacterium Tuberculosis." Journal of Molecular Medicine 85 (6): 613-21. 10.1007/s00109-007-0157-6
- **Kaforou_27**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Kaforou_OD_44**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Kaforou_OD_53**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Lee_4**: Lee, Shih-Wei, Lawrence Shih-Hsin Wu, Guan-Mau Huang, Kai-Yao Huang, Tzong-Yi Lee, and Julia Tzu-Ya Weng. 2016. "Gene Expression Profiling Identifies Candidate Biomarkers for Active and Latent Tuberculosis." BMC Bioinformatics 17 Suppl 1 (January): 3. 10.1186/s12859-015-0848-x
- **Leong_RISK_29**: Leong, S., Zhao, Y., Ribeiro-Rodrigues, R., Jones-López, E. C., Acuña-Villaorduña, C., Rodrigues, P. M., Palaci, M., Alland, D., Dietze, R., Ellner, J. J., Johnson, W. E., Salgame, P., Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB. Tuberculosis (Edinb). 2020 Jan;120:101898. 10.1016/j.tube.2020.101898
- **Maertzdorf_4**: Maertzdorf, Jeroen, Gayle McEwen, January Weiner 3rd, Song Tian, Eric Lader, Ulrich Schriek, Harriet Mayanja-Kizza, Martin Ota, John Kenneth, and Stefan He Kaufmann. 2016. "Concise Gene Signature for Point-of-Care Classification of Tuberculosis." EMBO Molecular Medicine 8 (2): 86-95. 10.15252/emmm.201505790
- **Maertzdorf_OD_100**: Maertzdorf, Jeroen, January Weiner 3rd, Hans-Joachim Mollenkopf, TBornot TB Network, Torsten Bauer, Antje Prasse, Joachim Müller-Quernheim, and Stefan H. E. Kaufmann. 2012. "Common Patterns and Disease-Related Signatures in Tuberculosis and Sarcoidosis." Proceedings of the National Academy of Sciences of the United States of America 109 (20): 7853-58. 10.1073/pnas.1121072109
- **Rajan_HIV_5**: Rajan, Jayant V., Semitala, Fred C., Kamya, Moses R., Yoon, Christina., Mehta, Tejas., Cattamanchi, Adithya., Seielstad, Mark., Montalvo, Lani., Andama, Alfred.,

Katende, Jane., Asege, Lucy., Nakaye, Martha., Mwebe, Sandra. 2018 "A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus" Clinical Infectious Diseases: 1-7. 10.1093/cid/ciy835

- **Roe_3**: Roe, Jennifer, Venturini, Cristina, Gupta, Rishi K., Gurry, Celine, Chain, Benjamin M., Sun, Yuxin, Southern, Jo, Jackson, Charlotte, Lipman, Marc, C., Miller, Robert F., Martineau, Adrian R., Abubakar, Ibrahim, Noursadeghi, Mahdad. 2019 "T1 Blood transcriptomic stratification of short-term risk in contacts of tuberculosis": . 10.1093/cid/ciz252
- **Roe_OD_4**: Roe, Jennifer K., Niclas Thomas, Eliza Gil, Katharine Best, Evdokia Tsaliki, Stephen Morris-Jones, Sian Stafford, et al. 2016. "Blood Transcriptomic Diagnosis of Pulmonary and Extrapulmonary Tuberculosis." JCI Insight 1 (16): e87238. 10.1172/jci.insight.87238
- **Sambarey_HIV_10**: Sambarey, Awanti, Abhinandan Devaprasad, Abhilash Mohan, Asma Ahmed, Soumya Nayak, Soumya Swaminathan, George D'Souza, et al. 2017. "Unbiased Identification of Blood-Based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks." EBioMedicine 15 (February): 112-26. 10.1016/j.ebiom.2016.12.009
- **Singhania_OD_20**: Singhania, Akul, Raman Verma, Christine M. Graham, Jo Lee, Trang Tran, Matthew Richardson, Patrick Lecine, et al. 2018. "A Modular Transcriptional Signature Identifies Phenotypic Heterogeneity of Human Tuberculosis Infection." Nature Communications 9 (1): 2308. 10.1038/s41467-018-04579-w
- **Sloot_HIV_2**: Sloot, Rosa, Maarten F. Schim van der Loeff, Erik W. van Zwet, Mariëlle C. Haks, Sytze T. Keizer, Maarten Scholing, Tom H. M. Ottenhoff, Martien W. Borgdorff, and Simone A. Joosten. 2015. "Biomarkers Can Identify Pulmonary Tuberculosis in HIV-Infected Drug Users Months Prior to Clinical Diagnosis." EBioMedicine 2 (2): 172-79. 10.1016/j.ebiom.2014.12.001
- **Suliman_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. 10.1164/rccm.201711-2340OC
- **Suliman_RISK_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. 10.1164/rccm.201711-2340OC
- **Sweeney_OD_3**: Sweeney, Timothy E., Lindsay Braviak, Cristina M. Tato, and Purvesh Khatri. 2016. "Genome-Wide Expression for Diagnosis of Pulmonary Tuberculosis: A Multicohort Analysis." The Lancet. Respiratory Medicine 4 (3): 213-24. 10.1016/S2213-2600(16)00048-5
- **Thompson_9**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **Thompson_FAIL_13**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **Thompson_RES_5**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **Walter_51**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15

- **Walter_PNA_47**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Walter_PNA_119**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Zak_RISK_16**: Zak, Daniel E., Adam Penn-Nicholson, Thomas J. Scriba, Ethan Thompson, Sara Suliman, Lynn M. Amon, Hassan Mahomed, et al. 2016. "A Blood RNA Signature for Tuberculosis Disease Risk: A Prospective Cohort Study." The Lancet 387 (10035): 231222. 10.1016/S0140-6736(15)01316-1

Examples

```
data("sigAnnotData")
```

`signatureBoxplot` *Plot a boxplot of signature genes.*

Description

Plot a boxplot of signature genes.

Usage

```
signatureBoxplot(
  inputData,
  annotationData,
  signatureColNames,
  annotationColName,
  name = "Signatures",
  scale = FALSE,
  includePoints = TRUE,
  notch = FALSE,
  rotateLabels = FALSE,
  nrow = NULL,
  ncol = NULL,
  fill_colors = NULL
)
```

Arguments

`inputData`	an input data object. It should either be of the class `SummarizedExperiment` and contain the profiled signature data and annotation data as columns in the `colData`, or alternatively be of the classes `data.frame` or `matrix` and contain only the gene expression data. Required.
`annotationData`	a `data.frame` or `matrix` of annotation data, with one column. Only required if `inputData` is a `data.frame` or `matrix` of signature data.

signatureColNames
: a vector of the column names in colData that contain the signature data. Only required if inputData is a SummarizedExperiment object.

annotationColName
: a character string naming the column name in the colData that contains the annotation data to be used in making the boxplot. Only required if inputData is a SummarizedExperiment object.

name
: a character string giving the title of the boxplot. The default is "Signatures".

scale
: logical. Setting scale = TRUE scales the signature data. The default is FALSE.

includePoints
: logical. If TRUE, points will be included over the boxplots. The default is TRUE.

notch
: logical. Notches are used to compare groups; if the notches of two boxes do not overlap, this suggests that the medians are significantly different. If TRUE, the boxplot will be notched. The default is FALSE.

rotateLabels
: logical. If TRUE, the x-axis labels will be rotated. The default is FALSE.

nrow
: integer giving the number of rows in the resulting array.

ncol
: integer giving the number of columns in the resulting array.

fill_colors
: a vector of color names to be used as the fill colors for the boxplot. If NULL, colors will be supplied via RColorBrewer. The default is fill_colors = NULL.

Value

A ggplot2 boxplot of the signature data using the provided annotation information.

Examples

```
library(SummarizedExperiment)

# Generate some artificial data that shows a difference in Zak_RISK_16
mat_testdata <- rbind(matrix(c(rnorm(80), rnorm(80) + 5), 16, 10,
                             dimnames = list(TBsignatures$Zak_RISK_16,
                                             paste0("sample", seq_len(10)))),
                      matrix(rnorm(1000), 100, 10,
                             dimnames = list(paste0("gene", seq_len(100)),
                                             paste0("sample", seq_len(10)))))

# Create a SummarizedExperiment object that contains the data
testdataSE <- SummarizedExperiment(assays = SimpleList(data = mat_testdata),
                                   colData = DataFrame(sample =
                                                         c(rep("down", 5),
                                                           rep("up", 5))))

# Run profiler using GSVA and ssGSEA on Zak_RISK_16 signature
res <- runTBsigProfiler(testdataSE, useAssay = "data",
                        signatures = TBsignatures["Zak_RISK_16"],
                        algorithm = c("GSVA", "ssGSEA"), parallel.sz = 1,
                        combineSigAndAlgorithm = TRUE)
signatureBoxplot(res, signatureColNames = c("GSVA_Zak_RISK_16",
                                            "ssGSEA_Zak_RISK_16"),
                 annotationColName = "sample", name = "Zak_RISK_16 Signature")
```

signatureGeneHeatmap *Plot a heatmap of a single signature score with individual gene expression levels.*

Description

This function takes the profiled gene expression data for a single signature and creates a heatmap based on the expression scores.

Usage

```
signatureGeneHeatmap(
  inputData,
  useAssay,
  sigGenes,
  name = "Signature",
  signatureColNames = NULL,
  annotationColNames = NULL,
  scale = TRUE,
  showColumnNames = TRUE,
  showRowNames = TRUE,
  colList = list(),
  colorSets = c("Set1", "Set2", "Set3", "Pastel1", "Pastel2", "Accent", "Dark2",
    "Paired"),
  choose_color = c("blue", "gray95", "red"),
  ...
)
```

Arguments

inputData	a SummarizedExperiment object containing the profiled signature data and annotation data as columns in the colData. Required.
useAssay	a character string specifying the assay to use for the gene expression data. Required.
sigGenes	a vector identifying the genes in the signature to use in the heatmap. For inbuilt signatures, you can use TBsignatures (e.g., TBsignatures[["ACS_COR"]]). Required.
name	a character string with the plot title of the heatmap. The default is "Signatures".
signatureColNames	a vector of the column names in the colData that contain the signature data. Required.
annotationColNames	a vector of the column names in the colData that contain the annotation data. If NULL, no annotation bar besides those of the scoring algorithms will be drawn on the heatmap. The default is NULL.
scale	logical. Setting scale = TRUE scales the signature data. The default is TRUE.
showColumnNames	logical. Setting showColumnNames = TRUE will show the column names (i.e. sample names) on the heatmap. The default is TRUE.

showRowNames — logical. Setting showColumnNames = TRUE will show the row names (i.e. signature names) on the heatmap. The default is TRUE.

colList — a named list of named vectors specifying custom color information to pass to ComplexHeatmap::Heatmap(). The list should have as many elements as there are annotation columns and gene signatures (i.e. sigGenes), and each element name should correspond exactly with the name of each annotation column/signature. The colors in the vector elements should be named according to the levels of the factor in that column's annotation data if the annotation is discrete, or it should be produced with circlize::colorRamp2 if the annotation/gene is continuous. By default, ColorBrewer color sets will be used. See the the parameter colorSets for additional details.

colorSets — a vector of names listing the color sets in the order that they should be used in creating the heatmap. By default, this function will use the color sets in the order listed in Usage for annotation information. You may replace the default with the same collection of sets in order that you want to use them, or provide custom color sets with the colList parameter.

choose_color — a vector of color names to be interpolated for the heatmap gradient, or a colorRamp function produced by circlize::colorRamp2. The default is c("blue","gray95","red").

... — Additional parameters to pass to ComplexHeatmap::Heatmap().

Value

A ComplexHeatmap plot.

Examples

```
library(SummarizedExperiment)
# Generate some artificial data that shows a difference in Zak_RISK_16
mat_testdata <- rbind(matrix(c(rnorm(80), rnorm(80) + 5), 16, 10,
                             dimnames = list(TBsignatures$Zak_RISK_16,
                                             paste0("sample", seq_len(10)))),
                      matrix(rnorm(1000), 100, 10,
                             dimnames = list(paste0("gene", seq_len(100)),
                                             paste0("sample", seq_len(10)))))

# Create a SummarizedExperiment object that contains the data
testdataSE <- SummarizedExperiment(assays = SimpleList(data = mat_testdata),
                                   colData = DataFrame(sample =
                                                         c(rep("down", 5),
                                                           rep("up", 5))))
# Run profiler using GSVA and ssGSEA on Zak_RISK_16
res <- runTBsigProfiler(testdataSE, useAssay = "data",
                        signatures = TBsignatures["Zak_RISK_16"],
                        algorithm = c("GSVA", "ssGSEA"), parallel.sz = 1,
                        combineSigAndAlgorithm = TRUE)

# Plot a heatmap of signature genes and pathway predictions
signatureGeneHeatmap(res, useAssay = "data",
                     sigGenes = TBsignatures[["Zak_RISK_16"]],
                     signatureColNames = c("GSVA_Zak_RISK_16",
                                           "ssGSEA_Zak_RISK_16"),
                     annotationColNames = c("sample"), showColumnNames = FALSE,
                     name = "Zak_RISK_16")
```

signatureHeatmap *Plot a heatmap of signature scores.*

Description

This function takes a dataset of scored gene expression data as an input and returns a ComplexHeatmap plot for for visual comparison of signature performance.

Usage

```
signatureHeatmap(
  inputData,
  annotationData = NULL,
  name = "Signatures",
  signatureColNames,
  annotationColNames = NULL,
  colList = list(),
  scale = FALSE,
  showColumnNames = TRUE,
  showRowNames = TRUE,
  colorSets = c("Set1", "Set2", "Set3", "Pastel1", "Pastel2", "Accent", "Dark2",
    "Paired"),
  choose_color = c("blue", "gray95", "red"),
  split_heatmap = "disease",
  annotationSignature = sigAnnotData,
  ...
)
```

Arguments

inputData	an input data object. It should either be of the class SummarizedExperiment and contain the profiled signature data and annotation data as columns in the colData, or alternatively be of the classes data.frame or matrix and contain only the gene expression data. Required.
annotationData	a data.frame or matrix of annotation data, with one column. Only required if inputData is a data.frame or matrix of signature data. The row names must equal those of the inputData column names. Default is NULL.
name	a character string with the plot title of the heatmap. The default is "Signatures".
signatureColNames	a vector of the column names in colData that contain the signature data. Only required if inputData is a SummarizedExperiment object.
annotationColNames	a vector of the column names in colData that contain the annotation data. Only required if inputData is a SummarizedExperiment. Default is NULL.
colList	a named list of named vectors specifying custom color information to pass to ComplexHeatmap::Heatmap(). The list should have as many elements as there are annotation columns, and each element name should correspond exactly with the name of each annotation column. The colors in the vector elements should be named according to the levels of the factor in that column's annotation data if the annotation is discrete, or it should be produced with circlize::colorRamp2 if

the annotation is continuous. By default, ColorBrewer color sets will be used. See the the parameter colorSets for additional details.

scale — logical. Setting scale = TRUE scales the signature data. The default is FALSE.

showColumnNames — logical. Setting showColumnNames = TRUE will show the column names (i.e. sample names) on the heatmap. The default is TRUE.

showRowNames — logical. Setting showColumnNames = TRUE will show the row names (i.e. signature names) on the heatmap. The default is TRUE.

colorSets — a vector of names listing the color sets in the order that they should be used in creating the heatmap. By default, this function will use the color sets in the order listed in Usage for annotation information. You may replace the default with the same collection of sets in order that you want to use them, or provide custom color sets with the colList parameter.

choose_color — a vector of color names to be interpolated for the heatmap gradient, or a colorRamp function produced by circlize::colorRamp2. The default is c("blue","gray95","red").

split_heatmap — a character string either giving the column title of annotationSignature containing annotation data for which to split the heatmap rows (i.e., signatures), or "none" if no split is desired. The default is "disease".

annotationSignature — a data.frame or matrix with information to be used in splitting the heatmap. The first column should signature names. The column of annotation information should be specified in split_heatmap. Other columns will be ignored. The default is sigAnnotData.

... — Additional arguments to be passed to ComplexHeatmap::Heatmap().

Details

If both annotationData = NULL and annotationColNames = NULL, no annotation bar will be drawn on the heatmap.

Value

A ComplexHeatmap plot.

Examples

```
library(SummarizedExperiment)
# Generate some artificial data that shows a difference in Zak_RISK_16
mat_testdata <- rbind(matrix(c(rnorm(80), rnorm(80) + 5), 16, 10,
                             dimnames = list(TBsignatures$Zak_RISK_16,
                                             paste0("sample", seq_len(10)))),
                      matrix(rnorm(1000), 100, 10,
                             dimnames = list(paste0("gene", seq_len(100)),
                                             paste0("sample", seq_len(10)))))
# Create a SummarizedExperiment object that contains the data
testdataSE <- SummarizedExperiment(assays = SimpleList(data = mat_testdata),
                                   colData = DataFrame(sample =
                                                         c(rep("down", 5),
                                                           rep("up", 5))))
res <- runTBsigProfiler(testdataSE, useAssay = "data",
                        signatures = TBsignatures["Zak_RISK_16"],
                        algorithm = c("GSVA", "ssGSEA"), parallel.sz = 1,
```

```
                        combineSigAndAlgorithm = TRUE)
signatureHeatmap(res, signatureColNames = c("GSVA_Zak_RISK_16",
                                            "ssGSEA_Zak_RISK_16"),
                 annotationColNames = "sample", scale = TRUE,
                 showColumnNames = FALSE, split_heatmap = "none")

# Example using custom colors for the annotation information
color2 <- stats::setNames(c("purple", "black"), c("down", "up"))
color.list <- list("sample" = color2)

signatureHeatmap(res, signatureColNames = c("GSVA_Zak_RISK_16",
                                            "ssGSEA_Zak_RISK_16"),
                 annotationColNames = "sample", scale = TRUE,
                 showColumnNames = FALSE,
                 colList = color.list, split_heatmap = "none")
```

`SignatureQuantitative` *Use logistic regression and bootstrap LOOCV to evaluate signatures.*

Description

This function takes as input a `data.frame` with genetic expression count data, and uses a bootstrapped leave-one-out cross validation procedure with logistic regression to allow for numeric and graphical comparison across any number of genetic signatures.

Usage

```
SignatureQuantitative(
  df.input,
  targetVec.num,
  signature.list = NULL,
  signature.name.vec = NULL,
  num.boot = 100,
  pb.show = TRUE
)
```

Arguments

`df.input`	a `data.frame` of gene expression count data. Required.
`targetVec.num`	a numeric binary vector of the response variable. The vector should be the same number of rows as `df`. Required.
`signature.list`	a `list` of signatures to run with their associated genes. This list should be in the same format as `TBsignatures`, included in the TBSignatureProfiler package. If `signature.list = NULL`, the default set of signatures `TBsignatures` list is used. For details, run `?TBsignatures`.
`signature.name.vec`	A vector specifying the names of the signatures to be compared. This should be the same length as `signature.list`. If `signature.name.vec = NULL`, the default set of signatures `TBsignatures` list is used.
`num.boot`	an integer specifying the number of bootstrap iterations.

pb.show — logical. If TRUE then a progress bar for the bootstrapping procedure will be displayed as output. The default is TRUE.

name — a character string giving a name for the outputted boxplot of bootstrapped AUCs. The default is "Quantitative Evaluation of Signatures via Bootstrapped AUCs".

Value

the AUC, sensitivity and specificity

Examples

```
inputTest <- matrix(rnorm(1000), 100, 20,
                    dimnames = list(paste0("gene", seq.int(1, 100)),
                                    paste0("sample", seq.int(1, 20))))
inputTest <- as.data.frame(inputTest)
targetVec <- sample(c(0,1), replace = TRUE, size = 20)
signature.list <- list(sig1 = c("gene1", "gene2", "gene3"),
                       sig2 = c("gene4", "gene5", "gene6"))
signature.name.vec <- c("sig1", "sig2")
num.boot <- 5
SignatureQuantitative(inputTest, targetVec.num = targetVec,
                      signature.list = signature.list,
                      signature.name.vec = signature.name.vec,
                      num.boot = num.boot)
```

---

signatureROCplot *Create an array of ROC plots to compare signatures.*

Description

Create an array of ROC plots to compare signatures.

Usage

```
signatureROCplot(
  inputData,
  annotationData,
  signatureColNames,
  annotationColName,
  scale = FALSE,
  choose_colors = c("cornflowerblue", "gray24"),
  name = "Signatures",
  nrow = NULL,
  ncol = NULL
)
```

Arguments

inputData	an input data object. It should either be of the class SummarizedExperiment and contain the profiled signature data and annotation data as columns in the colData, or alternatively be of the classes data.frame or matrix and contain only the gene expression data. Required.
annotationData	a data.frame or matrix of annotation data, with one column. Only required if inputData is a data.frame or matrix of signature data.
signatureColNames	a vector of the column names of inputData that contain the signature data. If inputData is a SummarizedExperiment object, these are the column names of the object colData.
annotationColName	a character string naming the column name in the colData that contains the annotation data to be used in making the boxplot. Only required if inputData is a SummarizedExperiment object.
scale	logical. Setting scale = TRUE scales the signature data. The default is FALSE.
choose_colors	a vector of length 2 defining the colors to be used in the ROC plots. The default is c("cornflowerblue","gray24").
name	a character string giving the title of the boxplot. The default is "Signatures".
nrow	integer giving the number of rows in the resulting array.
ncol	integer giving the number of columns in the resulting array.

Value

An array of ROC plots.

Examples

```
# Run signature profiling
 choose_sigs <- subset(TBsignatures,
                       !(names(TBsignatures) %in% c("Lee_4", "Roe_OD_4")))[c(1,2)]
 prof_indian <- runTBsigProfiler(TB_indian, useAssay = "logcounts",
                                 algorithm = "ssGSEA",
                                 signatures = choose_sigs,
                                 parallel.sz = 1)

# Create ROC plots
signatureROCplot(prof_indian, signatureColNames = names(choose_sigs),
                 annotationColName = "label")
```

---

signatureROCplot_CI *Create an array of ROC plots with confidence interval bands to compare signatures.*

Description

Create an array of ROC plots with confidence interval bands to compare signatures.

Usage

```
signatureROCplot_CI(
  inputData,
  annotationData,
  signatureColNames,
  annotationColName,
  scale = FALSE,
  choose_colors = c("cornflowerblue", "gray50", "gray79"),
  name = NULL,
  nrow = NULL,
  ncol = NULL,
  ci.lev = 0.95,
  pb.show = TRUE
)
```

Arguments

inputData	an input data object. It should either be of the class SummarizedExperiment and contain the profiled signature data and annotation data as columns in the colData, or alternatively be of the classes data.frame or matrix and contain only the gene expression data. Required.
annotationData	a data.frame or matrix of annotation data, with one column. Only required if inputData is a data.frame or matrix of signature data.
signatureColNames	a vector of the column names of inputData that contain the signature data. If inputData is a SummarizedExperiment object, these are the column names of the object colData.
annotationColName	a character string naming the column name in the colData that contains the annotation data to be used in making the boxplot. Only required if inputData is a SummarizedExperiment object.
scale	logical. Setting scale = TRUE scales the signature data. The default is FALSE.
choose_colors	a vector of length 3 defining the colors to be used in the ROC plots. The default is c("cornflowerblue","gray50","gray79").
name	a character string giving the title of the boxplot. If NULL, the plot title will be "ROC Plots for Gene Signatures,<ci.lev>% Confidence". The default is NULL.
nrow	integer giving the number of rows in the resulting array.
ncol	integer giving the number of columns in the resulting array.
ci.lev	a number between 0 and 1 giving the desired level of confidence for computing ROC curve estimations.
pb.show	logical for whether to show a progress bar while running code. The default is TRUE.

Value

An array of ROC plots.

Examples

```
# Run signature profiling

 choose_sigs <- TBsignatures[c(1, 2)]
 prof_indian <- runTBsigProfiler(TB_indian, useAssay = "logcounts",
                                 algorithm = "Zscore",
                                 signatures = choose_sigs,
                                 parallel.sz = 1)

# Create ROC plots with confidence intervals
signatureROCplot_CI(prof_indian, signatureColNames = names(choose_sigs),
                    annotationColName = "label")
```

`tableAUC` *Create a table of results for t-tests and bootstrapped AUCs for multiple scored signatures.*

Description

This function collects the results of bootstrapping and t-tests for a scored gene expression dataset and presents them using a JavaScript table with an R interface, or as a `data.frame`.

Usage

```
tableAUC(
  SE_scored,
  annotationColName,
  signatureColNames,
  num.boot = 100,
  pb.show = TRUE,
  output = "DataTable"
)
```

Arguments

`SE_scored`	a `SummarizedExperiment` object with genes as the row features and signature scores in the `colData`. There should also be a column of annotation data. Required.
`annotationColName`	a character string giving the column name in `colData` that contains the annotation data. Required.
`signatureColNames`	a vector of column names in the `colData` that contain the signature score data. Required.
`num.boot`	integer. The number of times to bootstrap the data. The default is `100`.
`pb.show`	logical for whether to show a progress bar while running code. The default is `TRUE`.
`output`	a character string indicating the table output format. Possible values are `DataTable` and `data.frame`. The default is `DataTable`.

Value

A JavaScript table with an R interface using the DT package.

Examples

```
# Run signature profiling
 choose_sigs <- TBsignatures[c(1, 2)]
 prof_indian <- runTBsigProfiler(TB_indian, useAssay = "logcounts",
                                 algorithm = "ssGSEA",
                                 signatures = choose_sigs,
                                 parallel.sz = 1)
 # Create table
 tableAUC(SE_scored = prof_indian, annotationColName = "label",
          signatureColNames = names(choose_sigs))

 # Create data.frame object
h <-  tableAUC(SE_scored = prof_indian, annotationColName = "label",
          signatureColNames = names(choose_sigs), output = "data.frame",
          num.boot = 5)
head(h)
```

TBcommon *A list of published TB signatures, using author-given names.*

Description

A set of 34 Tuberculosis gene signatures from various publications. This set of signatures uses gene symbols. Attempts have been made to use updated gene symbols and remove symbols that did not match the most recent annotation. Additional sets for Entrez IDs and Ensembl IDs are forthcoming.

Usage

```
TBcommon
```

Format

list

Details

This list differs from TBsignatures in that signatures with names specified in their originating publication (or that of a peer) are given that common name rather than using the TBSignatureProfiler naming system. Otherwise, signature names are composed of the last name of the primary author, followed by a possible context for the signature, and ending with either the number of gene transcripts or genes in the signature with respect to however it was described in the original publication.

Possible signature contexts:

- OD: Other diseases
- HIV: Human Immunodeficiency Virus
- PNA: Pneumonia

- RISK: Risk of developing active TB
- RES: Response to TB treatment
- FAIL: Failure of TB treatment

Note that in some cases signatures will be positive identifiers of TB whereas others are negative identifiers; this should be taken into account when creating ROC curves and computing any AUC estimates.

Source

- **Anderson_42**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. 10.1056/NEJMoa1303657
- **Anderson_OD_51**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. 10.1056/NEJMoa1303657
- **Berry_393**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. 10.1038/nature09247
- **Berry_OD_86**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. 10.1038/nature09247
- **Blankley_5**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. 10.1183/13993003.02121-2015
- **Blankley_380**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. 10.1183/13993003.02121-2015
- **Bloom_OD_144**: Bloom, Chloe I., Christine M. Graham, Matthew P. R. Berry, Fotini Rozakeas, Paul S. Redford, Yuanyuan Wang, Zhaohui Xu, et al. 2013. "Transcriptional Blood Signatures Distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung Cancers." PloS One 8 (8): e70630. 10.1371/journal.pone.0070630
- **Esmail_82**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Esmail_203**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating

Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115

- **Esmail_893**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Gliddon_OD_3**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org: . 10.1101/583674
- **Gliddon_OD_4**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org: . 10.1101/583674
- **Jacobsen_3**: Jacobsen, Marc, Dirk Repsilber, Andrea Gutschmidt, Albert Neher, Knut Feldmann, Hans J. Mollenkopf, Andreas Ziegler, and Stefan H. E. Kaufmann. 2007. "Candidate Biomarkers for Discrimination between Infection and Disease Caused by Mycobacterium Tuberculosis." Journal of Molecular Medicine 85 (6): 613-21. 10.1007/s00109-007-0157-6
- **Kaforou_27**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Kaforou_OD_44**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Kaforou_OD_53**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Lee_4**: Lee, Shih-Wei, Lawrence Shih-Hsin Wu, Guan-Mau Huang, Kai-Yao Huang, Tzong-Yi Lee, and Julia Tzu-Ya Weng. 2016. "Gene Expression Profiling Identifies Candidate Biomarkers for Active and Latent Tuberculosis." BMC Bioinformatics 17 Suppl 1 (January): 3. 10.1186/s12859-015-0848-x
- **PREDICT29**: Leong, S., Zhao, Y., Ribeiro-Rodrigues, R., Jones-López, E. C., Acuña-Villaorduña, C., Rodrigues, P. M., Palaci, M., Alland, D., Dietze, R., Ellner, J. J., Johnson, W. E., Salgame, P., Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB. Tuberculosis (Edinb). 2020 Jan;120:101898. 10.1016/j.tube.2020.101898
- **DIAG4**: Maertzdorf, Jeroen, Gayle McEwen, January Weiner 3rd, Song Tian, Eric Lader, Ulrich Schriek, Harriet Mayanja-Kizza, Martin Ota, John Kenneth, and Stefan He Kaufmann.

2016. "Concise Gene Signature for Point-of-Care Classification of Tuberculosis." EMBO Molecular Medicine 8 (2): 86-95. 10.15252/emmm.201505790

- **Maertzdorf_OD_100**: Maertzdorf, Jeroen, January Weiner 3rd, Hans-Joachim Mollenkopf, TBornot TB Network, Torsten Bauer, Antje Prasse, Joachim Müller-Quernheim, and Stefan H. E. Kaufmann. 2012. "Common Patterns and Disease-Related Signatures in Tuberculosis and Sarcoidosis." Proceedings of the National Academy of Sciences of the United States of America 109 (20): 7853-58. 10.1073/pnas.1121072109
- **Rajan_HIV_5**: Rajan, Jayant V., Semitala, Fred C., Kamya, Moses R., Yoon, Christina., Mehta, Tejas., Cattamanchi, Adithya., Seielstad, Mark., Montalvo, Lani., Andama, Alfred., Katende, Jane., Asege, Lucy., Nakaye, Martha., Mwebe, Sandra. 2018 "A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus" Clinical Infectious Diseases: 1-7. 10.1093/cid/ciy835
- **Roe_3**: Roe, Jennifer, Venturini, Cristina, Gupta, Rishi K., Gurry, Celine, Chain, Benjamin M., Sun, Yuxin, Southern, Jo, Jackson, Charlotte, Lipman, Marc, C., Miller, Robert F., Martineau, Adrian R., Abubakar, Ibrahim, Noursadeghi, Mahdad. 2019 "T1 Blood transcriptomic stratification of short-term risk in contacts of tuberculosis": . 10.1093/cid/ciz252
- **Roe_OD_4**: Roe, Jennifer K., Niclas Thomas, Eliza Gil, Katharine Best, Evdokia Tsaliki, Stephen Morris-Jones, Sian Stafford, et al. 2016. "Blood Transcriptomic Diagnosis of Pulmonary and Extrapulmonary Tuberculosis." JCI Insight 1 (16): e87238. 10.1172/jci.insight.87238
- **Sambarey_HIV_10**: Sambarey, Awanti, Abhinandan Devaprasad, Abhilash Mohan, Asma Ahmed, Soumya Nayak, Soumya Swaminathan, George D'Souza, et al. 2017. "Unbiased Identification of Blood-Based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks." EBioMedicine 15 (February): 112-26. 10.1016/j.ebiom.2016.12.009
- **Singhania_OD_20**: Singhania, Akul, Raman Verma, Christine M. Graham, Jo Lee, Trang Tran, Matthew Richardson, Patrick Lecine, et al. 2018. "A Modular Transcriptional Signature Identifies Phenotypic Heterogeneity of Human Tuberculosis Infection." Nature Communications 9 (1): 2308. 10.1038/s41467-018-04579-w
- **Sloot_HIV_2**: Sloot, Rosa, Maarten F. Schim van der Loeff, Erik W. van Zwet, Mariëlle C. Haks, Sytze T. Keizer, Maarten Scholing, Tom H. M. Ottenhoff, Martien W. Borgdorff, and Simone A. Joosten. 2015. "Biomarkers Can Identify Pulmonary Tuberculosis in HIV-Infected Drug Users Months Prior to Clinical Diagnosis." EBioMedicine 2 (2): 172-79. 10.1016/j.ebiom.2014.12.001
- **Suliman_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. 10.1164/rccm.201711-2340OC
- **RISK4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. 10.1164/rccm.201711-2340OC
- **DIAG3**: Sweeney, Timothy E., Lindsay Braviak, Cristina M. Tato, and Purvesh Khatri. 2016. "Genome-Wide Expression for Diagnosis of Pulmonary Tuberculosis: A Multicohort Analysis." The Lancet. Respiratory Medicine 4 (3): 213-24. 10.1016/S2213-2600(16)00048-5
- **DISEASE**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **FAILURE**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004

- **RESPONSE5**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **Walter_51**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Walter_PNA_47**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Walter_PNA_119**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **ACS_COR**: Zak, Daniel E., Adam Penn-Nicholson, Thomas J. Scriba, Ethan Thompson, Sara Suliman, Lynn M. Amon, Hassan Mahomed, et al. 2016. "A Blood RNA Signature for Tuberculosis Disease Risk: A Prospective Cohort Study." The Lancet 387 (10035): 231222. 10.1016/S0140-6736(15)01316-1

Examples

```
data("TBcommon")
```

`TBsignatures` *A list of published TB signatures.*

Description

A set of 34 Tuberculosis gene signatures from various publications. This set of signatures uses gene symbols. Attempts have been made to use updated gene symbols and remove symbols that did not match the most recent annotation. Additional sets for Entrez IDs and Ensembl IDs are forthcoming.

Usage

```
TBsignatures
```

Format

list

Details

Signature names are composed of the last name of the primary author, followed by a possible context for the signature, and ending with either the number of gene transcripts or genes in the signature, with respect to however it was described in the signature in the original publication.

Possible signature contexts:

- OD: Other diseases
- HIV: Human Immunodeficiency Virus
- PNA: Pneumonia
- RISK: Risk of developing active TB
- RES: Response to TB treatment
- FAIL: Failure of TB treatment

Note that in some cases signatures will be positive identifiers of TB whereas others are negative identifiers; this should be taken into account when creating ROC curves and computing any AUC estimates.

Source

- **Anderson_42**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. 10.1056/NEJMoa1303657
- **Anderson_OD_51**: Anderson, Suzanne T., Myrsini Kaforou, Andrew J. Brent, Victoria J. Wright, Claire M. Banwell, George Chagaluka, Amelia C. Crampin, et al. 2014. "Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa." The New England Journal of Medicine 370 (18): 1712-23. 10.1056/NEJMoa1303657
- **Berry_393**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. 10.1038/nature09247
- **Berry_OD_86**: Berry, Matthew P. R., Christine M. Graham, Finlay W. McNab, Zhaohui Xu, Susannah A. A. Bloch, Tolu Oni, Katalin A. Wilkinson, et al. 2010. "An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis." Nature 466 (7309): 973-77. 10.1038/nature09247
- **Blankley_5**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. 10.1183/13993003.02121-2015
- **Blankley_380**: Blankley, Simon, Christine M. Graham, Joe Levin, Jacob Turner, Matthew P. R. Berry, Chloe I. Bloom, Zhaohui Xu, et al. 2016. "A 380-Gene Meta-Signature of Active Tuberculosis Compared with Healthy Controls." The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology 47 (6): 1873-76. 10.1183/13993003.02121-2015
- **Bloom_OD_144**: Bloom, Chloe I., Christine M. Graham, Matthew P. R. Berry, Fotini Rozakeas, Paul S. Redford, Yuanyuan Wang, Zhaohui Xu, et al. 2013. "Transcriptional Blood Signatures Distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung Cancers." PloS One 8 (8): e70630. 10.1371/journal.pone.0070630
- **Esmail_82**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115

- **Esmail_203**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Esmail_893**: Esmail, Hanif, Rachel P. Lai, Maia Lesosky, Katalin A. Wilkinson, Christine M. Graham, Stuart Horswell, Anna K. Coussens, Clifton E. Barry 3rd, Anne O'Garra, and Robert J. Wilkinson. 2018. "Complement Pathway Gene Activation and Rising Circulating Immune Complexes Characterize Early Disease in HIV-Associated Tuberculosis." Proceedings of the National Academy of Sciences of the United States of America 115 (5): E964-73. 10.1073/pnas.1711853115
- **Gliddon_OD_3**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org:10.1101/583674
- **Gliddon_OD_4**: Gliddon, Harriet D., Kaforou, Myrsini, Alikian, Mary, Habgood-Coote, Dominic, Zhou, Chenxi, Oni, Tolu, Anderson, Suzanne T., Brent, Andrew J., Crampin, Amelia C., Eley, Brian, Kern, Florian, Langford, Paul R., Ottenhoff, Tom H. M., Hibberd, Martin L., French, Neil, Wright, Victoria J., Dockrell, Hazel M., Coin, Lachlan J., Wilkinson, Robert J., Levin, Michael. 2019 "Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification" biorxiv.org:10.1101/583674
- **Jacobsen_3**: Jacobsen, Marc, Dirk Repsilber, Andrea Gutschmidt, Albert Neher, Knut Feldmann, Hans J. Mollenkopf, Andreas Ziegler, and Stefan H. E. Kaufmann. 2007. "Candidate Biomarkers for Discrimination between Infection and Disease Caused by Mycobacterium Tuberculosis." Journal of Molecular Medicine 85 (6): 613-21. 10.1007/s00109-007-0157-6
- **Kaforou_27**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Kaforou_OD_44**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Kaforou_OD_53**: Kaforou, Myrsini, Victoria J. Wright, Tolu Oni, Neil French, Suzanne T. Anderson, Nonzwakazi Bangani, Claire M. Banwell, et al. 2013. "Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study." PLoS Medicine 10 (10): e1001538. 10.1371/journal.pmed.1001538
- **Lee_4**: Lee, Shih-Wei, Lawrence Shih-Hsin Wu, Guan-Mau Huang, Kai-Yao Huang, Tzong-Yi Lee, and Julia Tzu-Ya Weng. 2016. "Gene Expression Profiling Identifies Candidate Biomarkers for Active and Latent Tuberculosis." BMC Bioinformatics 17 Suppl 1 (January): 3. 10.1186/s12859-015-0848-x
- **Leong_RISK_29**: Leong, S., Zhao, Y., Ribeiro-Rodrigues, R., Jones-López, E. C., Acuña-Villaorduña, C., Rodrigues, P. M., Palaci, M., Alland, D., Dietze, R., Ellner, J. J., Johnson, W. E., Salgame, P., Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB. Tuberculosis (Edinb). 2020 Jan;120:101898. 10.1016/j.tube.2020.101898

- **Maertzdorf_4**: Maertzdorf, Jeroen, Gayle McEwen, January Weiner 3rd, Song Tian, Eric Lader, Ulrich Schriek, Harriet Mayanja-Kizza, Martin Ota, John Kenneth, and Stefan He Kaufmann. 2016. "Concise Gene Signature for Point-of-Care Classification of Tuberculosis." EMBO Molecular Medicine 8 (2): 86-95. 10.15252/emmm.201505790
- **Maertzdorf_OD_100**: Maertzdorf, Jeroen, January Weiner 3rd, Hans-Joachim Mollenkopf, TBornot TB Network, Torsten Bauer, Antje Prasse, Joachim Müller-Quernheim, and Stefan H. E. Kaufmann. 2012. "Common Patterns and Disease-Related Signatures in Tuberculosis and Sarcoidosis." Proceedings of the National Academy of Sciences of the United States of America 109 (20): 7853-58. 10.1073/pnas.1121072109
- **Rajan_HIV_5**: Rajan, Jayant V., Semitala, Fred C., Kamya, Moses R., Yoon, Christina., Mehta, Tejas., Cattamanchi, Adithya., Seielstad, Mark., Montalvo, Lani., Andama, Alfred., Katende, Jane., Asege, Lucy., Nakaye, Martha., Mwebe, Sandra. 2018 "A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus" Clinical Infectious Diseases: 1-7. 10.1093/cid/ciy835
- **Roe_3**: Roe, Jennifer, Venturini, Cristina, Gupta, Rishi K., Gurry, Celine, Chain, Benjamin M., Sun, Yuxin, Southern, Jo, Jackson, Charlotte, Lipman, Marc, C., Miller, Robert F., Martineau, Adrian R., Abubakar, Ibrahim, Noursadeghi, Mahdad. 2019 "T1 Blood transcriptomic stratification of short-term risk in contacts of tuberculosis": . 10.1093/cid/ciz252
- **Roe_OD_4**: Roe, Jennifer K., Niclas Thomas, Eliza Gil, Katharine Best, Evdokia Tsaliki, Stephen Morris-Jones, Sian Stafford, et al. 2016. "Blood Transcriptomic Diagnosis of Pulmonary and Extrapulmonary Tuberculosis." JCI Insight 1 (16): e87238. 10.1172/jci.insight.87238
- **Sambarey_HIV_10**: Sambarey, Awanti, Abhinandan Devaprasad, Abhilash Mohan, Asma Ahmed, Soumya Nayak, Soumya Swaminathan, George D'Souza, et al. 2017. "Unbiased Identification of Blood-Based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks." EBioMedicine 15 (February): 112-26. 10.1016/j.ebiom.2016.12.009
- **Singhania_OD_20**: Singhania, Akul, Raman Verma, Christine M. Graham, Jo Lee, Trang Tran, Matthew Richardson, Patrick Lecine, et al. 2018. "A Modular Transcriptional Signature Identifies Phenotypic Heterogeneity of Human Tuberculosis Infection." Nature Communications 9 (1): 2308. 10.1038/s41467-018-04579-w
- **Sloot_HIV_2**: Sloot, Rosa, Maarten F. Schim van der Loeff, Erik W. van Zwet, Mariëlle C. Haks, Sytze T. Keizer, Maarten Scholing, Tom H. M. Ottenhoff, Martien W. Borgdorff, and Simone A. Joosten. 2015. "Biomarkers Can Identify Pulmonary Tuberculosis in HIV-Infected Drug Users Months Prior to Clinical Diagnosis." EBioMedicine 2 (2): 172-79. 10.1016/j.ebiom.2014.12.001
- **Suliman_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. 10.1164/rccm.201711-2340OC
- **Suliman_RISK_4**: Suliman, Sara, Ethan Thompson, Jayne Sutherland, January Weiner Rd, Martin O. C. Ota, Smitha Shankar, Adam Penn-Nicholson, et al. 2018. "Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis." American Journal of Respiratory and Critical Care Medicine, April. https://doi.org/10.1164/rccm.201711-2340OC. 10.1164/rccm.201711-2340OC
- **Sweeney_OD_3**: Sweeney, Timothy E., Lindsay Braviak, Cristina M. Tato, and Purvesh Khatri. 2016. "Genome-Wide Expression for Diagnosis of Pulmonary Tuberculosis: A Multicohort Analysis." The Lancet. Respiratory Medicine 4 (3): 213-24. 10.1016/S2213-2600(16)00048-5
- **Thompson_9**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004

- **Thompson_FAIL_13**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **Thompson_RES_5**: Thompson, Ethan G., Ying Du, Stephanus T. Malherbe, Smitha Shankar, Jackie Braun, Joe Valvo, Katharina Ronacher, et al. 2017. "Host Blood RNA Signatures Predict the Outcome of Tuberculosis Treatment." Tuberculosis 107 (December): 48-58. 10.1016/j.tube.2017.08.004
- **Walter_51**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Walter_PNA_47**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Walter_PNA_119**: Walter, Nicholas D., Mikaela A. Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, et al. 2016. "Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: A Case-Control Study with Systematic Cross-Classifier Evaluation." Journal of Clinical Microbiology 54 (2): 274-82. 10.1128/JCM.01990-15
- **Zak_RISK_16**: Zak, Daniel E., Adam Penn-Nicholson, Thomas J. Scriba, Ethan Thompson, Sara Suliman, Lynn M. Amon, Hassan Mahomed, et al. 2016. "A Blood RNA Signature for Tuberculosis Disease Risk: A Prospective Cohort Study." The Lancet 387 (10035): 231222. 10.1016/S0140-6736(15)01316-1

Examples

```
data("TBsignatures")
```

`TBSPapp` *Run the TBSignatureProfiler Shiny application.*

Description

Use this function to run the TBSignatureProfiler application.

Usage

```
TBSPapp()
```

Value

The Shiny application will open.

Examples

```
# Upload data through the app
if (interactive()){
  TBSPapp()
  }
```

TB_hiv *An example TB dataset with TB/HIV data.*

Description

An example dataset containing the gene expression and metadata in a SummarizedExperiment object for 31 subjects with HIV and/or Tuberculosis diseases. Information on subject infection status can be accessed with `TB_hiv$Disease`. Samples with both TB and HIV contamination are marked as `tb_hiv`, while samples with HIV and no TB are marked as `hiv_only`.

Usage

```
TB_hiv
```

Format

SummarizedExperiment

Details

This dataset was published as part of a study to assess whether gene expression signatures and cytokine levels would distinguish active TB in advanced HIV in a cohort residing in Sub-Saharan Africa (Verma et. al 2018). Participants were severely immunosuppressed TB-HIV patients who had not yet received TB treatment or anti-retroviral therapy (ART). The dataset included in this package has been lightly edited from the originally published dataset due to the removal of one participant who was HIV positive, on ART and developed TB during follow-up. Whole blood RNA-Seq analysis was performed on all 31 participants.

References

Verma S., Du P., et. al. (2018). Tuberculosis in advanced HIV infection is associated with increased expression of IFN and its downstream targets. *BMC Infectious Diseases* **18:220**. doi: 10.1186/s12879-018-3127-4.

Examples

```
data("TB_hiv")
```

TB_indian *An example TB dataset with Indian population data.*

Description

An example dataset containing the gene expression and metadata in a SummarizedExperiment object for an Indian population. Active TB contamination of the 44 subjects is denoted for each as a "1"(active) or "0" (latent/not present), and can be accessed via `TB_indian$label`. The SummarizedExperiment object contains 2 assays (counts and log(counts)), and the column names give the unique subject identification number along with the subject's gender.

Usage

```
TB_indian
```

Format

SummarizedExperiment

Details

This dataset was published as part of a study to assess performance of published TB signatures in a South Indian population (Leong et. al 2018). RNA sequencing was performed on whole blood PAX gene samples collected from 28 TB patients and 16 latent TB infected (LTBI) subjects enrolled as part of an ongoing household contact study. Whole blood RNA-Seq analysis was performed on all 44 participants.

References

Leong S., Zhao Y., et. al. (2018). Existing blood transcriptional classifiers accurately discriminate active tuberculosis from latent infection in individuals from south India. *Tuberculosis* **109**, 41-51. doi: 10.1016/j.tube.2018.01.002.

Examples

```
data("TB_indian")
```

Index