Package 'curatedPCaData'

November 10, 2026

Title Curated Prostate Cancer Data

Version 1.8.0

Date 2024-01-10

Description The package curatedPCaData offers a selection of annotated prostate cancer datasets featuring multiple omics, manually curated metadata, and derived downstream variables. The studies are offered as MultiAssayExperiment (MAE) objects via ExperimentHub, and comprise of clinical characteristics tied to gene expression, copy number alteration and somatic mutation data. Further, downstream features computed from these multi-omics data are offered. Multiple vignettes help grasp characteristics of the various studies and provide example exploratory and meta-analysis of leveraging the multiple studies provided here-in.

License CC BY 4.0

Encoding UTF-8

RoxygenNote 7.2.3

VignetteBuilder knitr

biocViews ExperimentHub, ExperimentData, ProstateCancerData, CancerData, Homo_sapiens_Data, MicroarrayData, RNASeqData, ExpressionData, CopyNumberVariationData, Somatic, GEO, ReproducibleResearch

URL https://github.com/Syksy/curatedPCaData

BugReports https://github.com/Syksy/curatedPCaData/issues

Depends R (>= 4.4.0), S4Vectors (>= 0.23.18), MultiAssayExperiment, RaggedExperiment

Imports ExperimentHub, AnnotationHub, utils, methods, rlang, stats, reshape2

Suggests BiocStyle, knitr, ggplot2, rmarkdown, survival, survminer, ComplexHeatmap, corrplot, formatR

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

git_branch RELEASE_3_23

git_last_commit 060f061

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-10

Author Teemu Daniel Laajala [aut, cre] (ORCID:
<https://orcid.org/0000-0002-7016-7354>),
Jordan Creed [ctb],
Christelle Colin Leitzinger [ctb],
Varsha Sreekanth [ctb],
Federico Calboli [ctb],
Kalaimathy Singaravelu [ctb],
Michael Orman [ctb],
Alex Soupir [ctb],
Anni Halkola [ctb]

Maintainer Teemu Daniel Laajala <teelaa@utu.fi>

Contents

curatedPCaData . . . 3
curatedPCaDatasets_abida . . . 4
curatedPCaDatasets_baca . . . 5
curatedPCaDatasets_barbieri . . . 5
curatedPCaDatasets_barwick . . . 7
curatedPCaDatasets_chandran . . . 8
curatedPCaDatasets_friedrich . . . 9
curatedPCaDatasets_hieronymus . . . 10
curatedPCaDatasets_icgcca . . . 10
curatedPCaDatasets_igc . . . 11
curatedPCaDatasets_kim . . . 12
curatedPCaDatasets_kunderfranco . . . 13
curatedPCaDatasets_ren . . . 14
curatedPCaDatasets_sun . . . 15
curatedPCaDatasets_taylor . . . 16
curatedPCaDatasets_tcga . . . 17
curatedPCaDatasets_true . . . 18
curatedPCaDatasets_wallace . . . 19
curatedPCaDatasets_wang . . . 20
curatedPCaDatasets_weiner . . . 21
getPCa . . . 22
getPCaStudies . . . 24
getPCaSummarySamples . . . 25
getPCaSummaryStudies . . . 25
getPCaSummarySurv . . . 26
getPCaSummaryTable . . . 27
template_prad . . . 27

Index **29**

curatedPCaData *curatedPCaData*

Description

The package curatedPCaData offers a selection of annotated prostate cancer datasets featuring multiple omics, manually curated metadata, and derived downstream variables. The studies are offered as MultiAssayExperiment (MAE) objects via ExperimentHub, and comprise of clinical characteristics tied to gene expression, copy number alteration and somatic mutation data. Further, downstream features computed from these multi-omics data are offered. Multiple vignettes help grasp characteristics of the various studies and provide example exploratory and meta-analysis of leveraging the multiple studies provided here-in.

Author(s)

Maintainer: Teemu Daniel Laajala <teelaa@utu.fi> (ORCID)

Other contributors:

- Jordan Creed <jordan.h.creed@moffitt.org> [contributor]
- Christelle Colin Leitzinger <christelle.colinleitzinger@moffitt.org> [contributor]
- Varsha Sreekanth <varsha.sreekanth@cuanschutz.edu> [contributor]
- Federico Calboli <federico.calboli@luke.fi> [contributor]
- Kalaimathy Singaravelu <kalsin@utu.fi> [contributor]
- Michael Orman <michael.orman@cuanschutz.edu> [contributor]
- Alex Soupir <Alex.Soupir@moffitt.org> [contributor]
- Anni Halkola <ansuha@utu.fi> [contributor]

See Also

Useful links:

- https://github.com/Syksy/curatedPCaData
- Report bugs at https://github.com/Syksy/curatedPCaData/issues

`curatedPCaDatasets_abida` *Abida et al. MAE-object*

Description

MultiAssayExperiment object containing gene expression (gex), copy number alteration (cna), mutations (mut) and immune cell estimates for Abida et al.

Format

A MAE object spanning 444 castrate resistant prostate cancer samples.

cna.gistic matrix with 20291 rows and 444 columns, from GISTIC discretized copy number alteration calls.

gex.relz matrix with 18971 rows and 266 columns, from z-score normalized expression in relative to paired normals.

mut RaggedExperiment with 63184 rows and 444 columns, mutation data from cbioportal

cibersort matrix with 22 rows and 261 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 261 columns, of xcell based deconvolution data

epic matrix with 8 rows and 261 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 261 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 261 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 261 columns, of risk scores and AR scores

mcp mmatrix with 11 rows and 261 columns, of mcp-counter based deconvolution data

Details

the clinical data refers to a sample of 444 tumors collected in 429 patients. The tissue was collected primarily at metastatic sites rather than from the prostate.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.cbioportal.org/study/summary?id=prad_su2c_2019

References

(PubMed)

Examples

```
mae_abida <- getPCa('abida')
```

curatedPCaDatasets_baca

Baca et al. MAE-object

Description

MultiAssayExperiment object containing copy number alteration (cna) and mutations for Baca et al.

Format

A MAE object spanning 56 prostate cancer samples.

cna.gistic matrix with 20124 rows and 56 columns, from GISTIC discretized copy number alteration calls.

mut RaggedExperiment with 2734 rows and 57 columns, mutation data from cbioportal

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.cbioportal.org/study/summary?id=prad_broad_2013

Examples

```
mae_baca <- getPCa('baca')
```

curatedPCaDatasets_barbieri

Barbieri et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex), copy number alteration (cna), mutations (mut) and immune cell estimates for Barbieri et al.

Format

A MAE object spanning prostate adenocarcinomas from Barbieri et. al

cna.gistic matrix with 20844 rows and 109 columns, from GISTIC discretized copy number alteration calls

gex.relz matrix with 17917 rows and 31 columns, from z-score normalized expression in relative to paired normals.

mut RaggedExperiment with 5737 rows and 112 columns, mutation data from cbioportal

cibersort matrix with 22 rows and 31 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 31 columns, of xcell based deconvolution data

epic matrix with 8 rows and 31 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 31 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 31 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 31 columns, of risk scores and AR scores

mcp matrix with 11 rows and 31 columns, of mcp-counter based deconvolution data

Details

The data comprises of primary localised prostate tumors from two cohorts, the Weill Cornell Medical College (WCMC; New York, NY), and the Uropath (Perth, Australia), which commercially provides banked urological tissues. None of the samples comes from patients who had received prior treatment for prostate cancer.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.cbioportal.org/study/summary?id=prad_broad

References

(PubMed)

Examples

```
mae_barbieri <- getPCa('barbieri')
```

curatedPCaDatasets_barwick

Barwick et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates for Barwick et al.

Format

A MAE object spanning 146 prostate cancer samples.

gex.logq matrix with 482 rows and 146 columns, for the log-quantile normalized gene expression data

cibersort matrix with 22 rows and 139 columns, of cibersort based deconvolution data

quantiseq matrix with 11 rows and 139 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 139 columns, of cell types based on ESTIMATE method

scores matrix with 3 rows and 139 columns, of risk scores and AR scores

mcp matrix with 4 rows and 139 columns, of mcp-counter based deconvolution data

Details

Barwick et al. uses an older customized DASL array; therefore its gene coverage is lower, and many downstream methods fail due to lack of gene overlap.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18655

References

(PubMed)

Examples

```
mae_barwick <- getPCa('barwick')
```

curatedPCaDatasets_chandran

Chandran et al., Yu et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates for Chandran et al., Yu et al.

Format

An MAE object spanning 171 men with prostate cancer

gex.rma matrix with 9007 rows and 171 columns, for the gene expression data

cibersort matrix with 22 rows and 171 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 171 columns, of xcell based deconvolution data

epic matrix with 8 rows and 171 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 171 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 171 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 171 columns, of risk scores and AR scores

mcp matrix with 11 rows and 171 columns, of mcp-counter based deconvolution data

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6919

References

(PubMed) (PubMed)

Examples

```
mae_chandran <- getPCa('chandran')
```

curatedPCaDatasets_friedrich

Friedrich et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates for Friedrich et al.

Format

An MAE object spanning 255 men with prostate cancer

gex.logq matrix with 23097 rows and 255 columns, for the log-quantile normalized gene expression data

cibersort matrix with 22 rows and 255 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 255 columns, of xcell based deconvolution data

epic matrix with 8 rows and 255 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 255 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 255 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 255 columns, of risk scores and AR scores

mcp matrix with 11 rows and 255 columns, of mcp-counter based deconvolution data

Details

The data comprises of 255 samples, with 164 primary tumors samples, 52 adjacent normal samples, and 39 benign prostate hyperplasia samples. This dataset includes in its totality the 164 samples analysed in Kreutz et al. 2020.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134051

References

(PubMed) (PubMed)

Examples

```
mae_friedrich <- getPCa('friedrich')
```

curatedPCaDatasets_hieronymus
Hieronymus et al. MAE-object

Description

MultiAssayExperiment object containing copy number alteration (cna) for Hieronymus et al.

Format

A MAE object spanning 104 tumor samples

cna.gistic matrix with 18026 rows and 104 columns, from GISTIC discretized copy number alteration calls

Details

The data comprises of 104 samples, for which are available clinical data and Copy Number Alteration, but no gene expression data – thus no deconvolution results are available.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54691

References

(PubMed)

Examples

```
mae_hieronymus <- getPCa('hieronymus')
```

curatedPCaDatasets_icgcca
ICGC CA MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates for the ICGC CA (Canadian) cohort.

Format

An MAE object spanning 213 men with prostate cancer

gex.rma matrix with 17208 rows and 213 columns, of gene expression data

cibersort matrix with 22 rows and 213 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 213 columns, of xcell based deconvolution data

epic matrix with 8 rows and 213 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 213 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 213 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 213 columns, of risk scores and AR scores

mcp matrix with 11 rows and 213 columns, of mcp-counter based deconvolution data

Details

The data refers to samples from the ICGC Canadian Prostate Cancer Genome Network (ICGC-PRAD-CA)

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://dcc.icgc.org/projects/PRAD-CA

References

(PubMed)

Examples

```
mae_icgcca <- getPCa('icgcca')
```

`curatedPCaDatasets_igc` *International Genomics Consortium (IGC) MAE-object*

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates from IGC's Expression Project for Oncology (expO) with a subset for prostate cancer

Format

An MAE object spanning 83 men with prostate cancer

gex.rma matrix with 12798 rows and 83 columns, of gene expression data

cibersort matrix with 22 rows and 83 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 83 columns, of xcell based deconvolution data

epic matrix with 8 rows and 83 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 83 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 83 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 83 columns, of risk scores and AR scores

mcp matrix with 11 rows and 83 columns, of mcp-counter based deconvolution data

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2109

Examples

```
mae_igc <- getPCa('igc')
```

---

curatedPCaDatasets_kim

Kim et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates from Kim et al.

Format

An MAE object spanning 266 men with prostate cancer

gex.rma matrix with 17638 rows and 266 columns, of gene expression data

cibersort matrix with 22 rows and 266 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 266 columns, of xcell based deconvolution data

epic matrix with 8 rows and 266 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 266 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 266 columns, of cell types based on ESTIMATE method

scores matrix with 2 rows and 266 columns, of risk scores and AR scores

mcp matrix with 11 rows and 266 columns, of mcp-counter based deconvolution data

Details

The dataset consists of 266 NCCN very low/low or favorable-intermediate risk PCa patients

Value

A MultiAssayExperiment corresponding to the study and its available

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119616

References

(PubMed)

Examples

```
mae_kim <- getPCa('kim')
```

---

curatedPCaDatasets_kunderfranco

Kunderfranco et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates from Kunderfranco et al.

Format

An MAE object spanning 67 samples of normal prostate samples and prostate cancer samples

gex.logr matrix with 16546 rows and 67 columns, of gene expression data

cibersort matrix with 22 rows and 67 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 67 columns, of xcell based deconvolution data

epic matrix with 8 rows and 67 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 67 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 67 columns, of cell types based on ESTIMATE method

scores matrix with 2 rows and 67 columns, of risk scores and AR scores

mcp matrix with 11 rows and 67 columns, of mcp-counter based deconvolution data

Details

The data contains 14 disease free benign prostate hyperplasia samples and 53 prostate cancer samples.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14206

References

(PubMed)

Examples

```
mae_kunderfranco <- getPCa('kunderfranco')
```

---

curatedPCaDatasets_ren

Ren et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex), mutations and immune cell estimates from Ren et al.

Format

An MAE object spanning 65 men with prostate cancer

gex.relz matrix with 21046 rows and 65 columns, of gene expression data

mut RaggedExperiment with 50625 rows and 65 columns, of mutation data from cbioportal

cibersort matrix with 22 rows and 65 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 65 columns, of xcell based deconvolution data

epic matrix with 8 rows and 65 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 65 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 65 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 65 columns, of risk scores and AR scores

mcp matrix with 11 rows and 65 columns, of mcp-counter based deconvolution data

Details

The data contains 14 disease free benign prostate hyperplasia samples and 53 prostate cancer samples.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.cbioportal.org/study/summary?id=prad_eururol_2017

References

(PubMed)

Examples

```
mae_ren <- getPCa('ren')
```

---

curatedPCaDatasets_sun

Sun et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates from Sun et al.

Format

An MAE object spanning 79 men with prostate cancer

gex.rma matrix with 12798 rows and 79 columns, of gene expression data

cibersort matrix with 22 rows and 79 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 79 columns, of xcell based deconvolution data

epic matrix with 8 rows and 79 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 79 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 79 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 79 columns, of risk scores and AR scores

mcp matrix with 11 rows and 79 columns, of mcp-counter based deconvolution data

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25136

References

(PubMed)

Examples

```
mae_sun <- getPCa('sun')
```

curatedPCaDatasets_taylor

Taylor et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex), copy number alteration (cna), mutations and immune cell estimates from Taylor et al.

Format

An MAE object spanning 218 men with prostate cancer

cna.gistic matrix with 17832 rows and 194 columns, of gistic values for copy number alteration data

cna.logr matrix with 18062 rows and 218 columns, of log-ratios for copy number alteration data

gex.rma matrix with 17410 rows and 179 columns, of gene expression data

mut RaggedExperiment with 90 rows and 43 columns, of mutation data

cibersort matrix with 22 rows and 179 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 179 columns, of xcell based deconvolution data

epic matrix with 8 rows and 179 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 179 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 179 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 179 columns, of risk scores and AR scores

mcp matrix with 11 rows and 179 columns, of mcp-counter based deconvolution data

Details

Note that there is lack of overlap between the omics provided for each sample.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21035

References

(PubMed)

Examples

```
mae_taylor <- getPCa('taylor')
```

curatedPCaDatasets_tcga *TCGA MAE-object*

Description

MultiAssayExperiment object containing gene expression (gex), copy number alteration (cna), mutations and immune cell estimates from GDC TCGA.

Format

An MAE object spanning 369 men with prostate cancer

cna.gistic matrix with 23151 rows and 492 columns, from GISTIC discretized copy number alteration calls

gex.rsem.log matrix with 19658 rows and 461 columns, of gene expression data

mut RaggedExperiment with 30897 rows and 495 columns, of mutation data

cibersort matrix with 22 rows and 550 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 461 columns, of xcell based deconvolution data

epic matrix with 8 rows and 461 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 461 columns, of quantiseq based deconvolution data

mcp matrix with 11 rows and 461 columns, of mcp-counter based deconvolution data

estimate data.frame with 4 rows and 461 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 416 columns, of risk scores and AR scores

Details

TCGA data was obtained from the latest GDC's XenaBrowser release.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Prostate%20Cancer%20(PRAD)

References

(PubMed)

Examples

```
mae_tcga <- getPCa('tcga')
```

curatedPCaDatasets_true

True et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates from True et al.

Format

An MAE object spanning 32 men with prostate cancer

gex.logr matrix with 3615 rows and 32 columns, of gene expression data

cibersort matrix with 22 rows and 32 columns, of cibersort based deconvolution data

epic matrix with 8 rows and 32 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 32 columns, of quantiseq based deconvolution data

mcp matrix with 7 rows and 32 columns, of mcp-counter based deconvolution data

scores matrix with 2 rows and 32 columns, of risk scores and AR scores

estimate data.frame with 4 rows and 32 columns, of cell types based on ESTIMATE method

Details

The expression data for this dataset was produced by an older two colour chip that provides the relative expression levels for between tumor and normal. Due to the chip type, there was a lack of gene overlap with some of the downstream methods, thus some downstream methods are missing.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5132

Examples

```
mae_true <- getPCa('true')
```

curatedPCaDatasets_wallace

Wallace et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates from Wallace et al.

Format

An MAE object spanning 83 men with prostate cancer

gex.rma matrix with 12783 rows and 89 columns, of gene expression data

cibersort matrix with 22 rows and 83 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 89 columns, of xcell based deconvolution data

epic matrix with 8 rows and 89 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 89 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 89 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 89 columns, of risk scores and AR scores

mcp matrix with 11 rows and 89 columns, of mcp-counter based deconvolution data

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6956

References

(PubMed)

Examples

```
mae_wallace <- getPCa('wallace')
```

curatedPCaDatasets_wang

Wang et al. MAE-object

Description

MultiAssayExperiment object containing gene expression (gex) and immune cell estimates from Wang et al.

Format

An MAE object spanning 148 men with prostate cancer

gex.rma matrix with 12798 rows and 148 columns, of gene expression data

cibersort matrix with 22 rows and 148 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 148 columns, of xcell based deconvolution data

epic matrix with 8 rows and 148 columns, of epic based deconvolution data

quantiseq matrix with 11 rows and 148 columns, of quantiseq based deconvolution data

estimate data.frame with 4 rows and 148 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 148 columns, of risk scores and AR scores

mcp matrix with 11 rows and 148 columns, of mcp-counter based deconvolution data

Details

148 prostate samples, with various amounts of tumor, stroma, BPH and atrophic gland, were used for this study.

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8218

References

(PubMed)

Examples

```
mae_wang <- getPCa('wang')
```

curatedPCaDatasets_weiner *Weiner et al. MAE-object*

Description

A MultiAssayExperiment object containing gene expression (gex) and immune cell estimates for Weiner et al.

Format

An MAE spanning 838 prostate cancer samples of two cohorts

gex.rma matrix of 17410 rows and 838 columns of gene expression data

cibersort matrix with 22 rows and 838 columns, of cibersort based deconvolution data

xcell matrix with 39 rows and 838 columns, the xcell deconvolution of the expression data

epic matrix with 8 rows and 838 columns, the epic deconvolution of the expression data

quantiseq matrix with 11 rows and 838 columns, the quantiseq deconvolution of the expression data

estimate data.frame with 4 rows and 838 columns, of cell types based on ESTIMATE method

scores matrix with 4 rows and 838 columns, of risk scores and AR scores

mcp matrix with 11 rows and 838 columns, the mcp-counter deconvolution of the expression data

Value

A MultiAssayExperiment corresponding to the study and its available omics.

Source

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157548

References

(PubMed)

Examples

```
mae_weiner <- getPCa('weiner')
```

getPCa *Construct Prostate Cancer MultiAssayExperiment object from specific cohort*

Description

curatedPCaData provides MultiAssayExperiment container objects that are constructed from ExperimentHub. User provides PCa dataset name (see list on the overview-vignette or help files).

Usage

```
getPCa(dataset, assays, timestamp, sampletypes, verbose = FALSE, ...)
```

Arguments

dataset	character() of PCa cancer cohort names (e.g., 'abida')
assays	character() A vector of PCa assays. If not included, returns all available for the selected dataset; see below for more details.
timestamp	character(1) "20230215" indicating the data version to obtain from 'ExperimentHub'. See 'timestamp' section details.
sampletypes	A character vector of sample types to include in the MAE. If the parameter is missing, then all samples are returned. Allowed values are: 'primary', 'metastatic', 'normal', 'BPH', 'atrophic'
verbose	logical(1) Whether to show the dataset currenlty being (down)loaded (default TRUE)
...	Additional arguments passed on to the ExperimentHub constructor

Details

This function will check against available resources in ExperimentHub. For a list of datasets, see the overview-vignette or below dataset listing.

Value

a MultiAssayExperiment of the specified assays and cancer cohort

Available datasets

Following shortnames can be used with getPCa to query for their omics from ExperimentHub. For more detailed description of the studies, see their corresponding R documentation file. Further details are reported in the overview-vignette.

```
Shortname      Rd documentation file          Longer name(s)
-------------------------------------------------------------------------
```

```
abida        curatedPCaDatasets_abida        Abida et al.
baca         curatedPCaDatasets_baca         Baca et al.
barbieri     curatedPCaDatasets_barbieri     Barbieri et al.
barwick      curatedPCaDatasets_barwick      Barwick et al.
chandran     curatedPCaDatasets_chandran     Chandran et al., Yu et al.
friedrich    curatedPCaDatasets_friedrich    Friedrich et al.
hieronymus   curatedPCaDatasets_hieronymus   Hieronymus et al.
icgcca       curatedPCaDatasets_icgcca       ICGC Canadian subset
igc       curatedPCaDatasets_igc       International Genomics Consortium (PCa subset)
kim          curatedPCaDatasets_kim          Kim et al.
kunderfranco curatedPCaDatasets_kunderfranco Kunderfranco et al.
ren          curatedPCaDatasets_ren          Ren et al.
sun          curatedPCaDatasets_sun          Sun et al.
taylor       curatedPCaDatasets_taylor       Taylor et al.
tcga      curatedPCaDatasets_tcga      The Cancer Genome Atlas (PCa subset, PRAD-prefix)
true         curatedPCaDatasets_true         True et al.
wallace      curatedPCaDatasets_wallace      Wallace et al.
wang         curatedPCaDatasets_wang         Wang et al.
weiner       curatedPCaDatasets_weiner       Weiner et al.
```

Available Assays

The list of ExperimentList assay names and their descriptions. These assays can be entered as part of the assays argument in the main function.

```
ExperimentList data types   Description
------------------------------------------------------------------------------

gex.rma            Gene expression values (RMA normalized)
gex.logq                                  (Log-quantile)
gex.relz                                  (Relative z-scores)
gex.logr                                  (Log-ratios)
gex.rsem.log                              (Log-RSEM)
cna.gistic         Copy number alteration (GISTIC calls)
cna.logr                                  (Log-ratios)
mut                Somatic mutation calls
cibersort          Immune deconvolution as estimated by CIBERSORTx
xcell                                                   xCell
epic                                                    EPIC
quantiseq                                               quanTIseq
mcp                                                     MCP-counter
estimate                                                ESTIMATE
scores             Various risk and AR scores; see vignette references.
```

timestamp

"20230215" The timestamp is updated in case the data is updated. In this case, this section describes the changes made for new timestamps. At this time, the only data deposit is from 2023, Feb 15th,

indicated by "20230215".

See Also

curatedPCaData-package

Examples

```
mae_taylor <- getPCa(
    dataset = "taylor", timestamp = "20230215"
)
```

`getPCaStudies` *Create a vector of unique study identifiers available in curatedPCaData*

Description

This function accesses the metadata available together with the curatedPCaData package and creates a vector of unique study identifier names available in the ExperimentHub resources. This vector is useful, for example, when applying the 'getPCa' to create MAE from the studies.

Usage

```
getPCaStudies(uniqs = TRUE)
```

Arguments

`uniqs`	Should only unique instances be returned; by default TRUE

Details

Uses the internal 'metadata.csv' file, which is also used to provide the metadata available in ExperimentHub.

Value

A character vector with short study identifiers, e.g. 'abida', 'baca', ...

Examples

```
studyids <- getPCaStudies()
```

getPCaSummarySamples *Create a summary table for sample counts for different omics across the datasets in curatedPCaData*

Description

Create two types of summary matrices; first one takes all unique assay names, and summarizes sample counts available for all provided studies. The second matrix takes gene expression, copy number alteration, and somatic mutation data, and displays sample counts for each omic individually, as well as all combinations of intersections for samples available for multi-omics.

Usage

```
getPCaSummarySamples(maes)
```

Arguments

maes	List of MultiAssayExperiment objects to summarize

Value

A list with two elements: first is a matrix depicting sample counts in each unique assay name for the studies; second one is a matrix with sample count overlap between the three main omics (GEX = gene expression; CNA = copy number alteration; MUT = somatic mutations).

Examples

```
mae_taylor <- getPCa('taylor')
mae_tcga <- getPCa('tcga')
getPCaSummarySamples(maes = list(Taylor = mae_taylor, TCGA = mae_tcga))
```

getPCaSummaryStudies *Create a summary table of key characteristics for each study*

Description

Create a summary table of key characteristics for each study

Usage

```
getPCaSummaryStudies(maes)
```

Arguments

maes	A list of MAE objects or a character vector with study short ids

Value

A table with study names, references, sample counts, data source, and other notes.

Examples

```
studies <- getPCaStudies(uniqs = TRUE)
tab <- getPCaSummaryStudies(studies)
```

`getPCaSummarySurv` *Create survival summary for metadata in curatedPCaData*

Description

Create summary for survival-like columns (0/1 event with follow-up time) for colData metadata in curatedPCaData

Usage

```
getPCaSummarySurv(maes, event.name, time.name)
```

Arguments

`maes`	List of MultiAssayExperiment objects to summarize
`event.name`	Name for the colData column for the end-point event; typically one of 'days_to_disease_specific_recurren or 'days_to_overall_survival'
`time.name`	Name for the colData column for the time for the end-point until event or censoring; typically one of 'disease_specific_recurrence_status' or 'overall_survival_status'

Value

A matrix describing the event and follow-up information as character strings

Examples

```
mae_taylor <- getPCa('taylor')
mae_tcga <- getPCa('tcga')
getPCaSummarySurv(maes = list(Taylor = mae_taylor, TCGA = mae_tcga),
    event.name = 'disease_specific_recurrence_status', time.name =
    'days_to_disease_specific_recurrence')
```

getPCaSummaryTable *Create summary tables for colData metadata in curatedPCaData*

Description

Create a variable value availability tables; NA-value in addition to provided 'vals' and Other-values for debugging or anomalies

Usage

```
getPCaSummaryTable(maes, var.name, vals, nas = TRUE)
```

Arguments

maes	List of MultiAssayExperiment objects to summarize
var.name	Name of the metadata variable to look for in colData
vals	Possible values for the metadata variable to tabulate
nas	Whether NA or values other than those contained in 'vals' should be appended as extra columns

Value

A matrix describing the tabulates 'vals' and their proportions along with missingness

Examples

```
mae_taylor <- getPCa('taylor')
mae_tcga <- getPCa('tcga')
getPCaSummaryTable(maes = list(Taylor = mae_taylor, TCGA = mae_tcga),
    var.name = 'gleason_grade', vals = 5:10)
```

template_prad *Template for curatedPCaData clinical fields*

Description

This data frame contains the clinical fields which were the aim for extraction when gathering metadata for each dataset. Variable types, ranges, and description is reported.

Usage

```
template_prad
```

Format

An object of class `data.frame` with 68 rows and 6 columns.

Value

A data.frame of the PCa template

Examples

```
data(template_prad)
head(template_prad)
```

Index

∗ **datasets**
curatedPCaDatasets_abida, 4
curatedPCaDatasets_baca, 5
curatedPCaDatasets_barbieri, 5
curatedPCaDatasets_barwick, 7
curatedPCaDatasets_chandran, 8
curatedPCaDatasets_friedrich, 9
curatedPCaDatasets_hieronymus, 10
curatedPCaDatasets_icgcca, 10
curatedPCaDatasets_igc, 11
curatedPCaDatasets_kim, 12
curatedPCaDatasets_kunderfranco, 13
curatedPCaDatasets_ren, 14
curatedPCaDatasets_sun, 15
curatedPCaDatasets_taylor, 16
curatedPCaDatasets_tcga, 17
curatedPCaDatasets_true, 18
curatedPCaDatasets_wallace, 19
curatedPCaDatasets_wang, 20
curatedPCaDatasets_weiner, 21
template_prad, 27
∗ **internal**
curatedPCaData, 3

curatedPCaData, 3
curatedPCaData-package *(*curatedPCaData*)*, 3
curatedPCaDatasets_abida, 4
curatedPCaDatasets_baca, 5
curatedPCaDatasets_barbieri, 5
curatedPCaDatasets_barwick, 7
curatedPCaDatasets_chandran, 8
curatedPCaDatasets_friedrich, 9
curatedPCaDatasets_hieronymus, 10
curatedPCaDatasets_icgcca, 10
curatedPCaDatasets_igc, 11
curatedPCaDatasets_kim, 12
curatedPCaDatasets_kunderfranco, 13
curatedPCaDatasets_ren, 14
curatedPCaDatasets_sun, 15
curatedPCaDatasets_taylor, 16
curatedPCaDatasets_tcga, 17
curatedPCaDatasets_true, 18
curatedPCaDatasets_wallace, 19
curatedPCaDatasets_wang, 20
curatedPCaDatasets_weiner, 21

ExperimentHub, *22*

getPCa, 22
getPCaStudies, 24
getPCaSummarySamples, 25
getPCaSummaryStudies, 25
getPCaSummarySurv, 26
getPCaSummaryTable, 27

MultiAssayExperiment, *22*

template_prad, 27