

Package ‘MetaGxBreast’

October 14, 2018

Type Package

Title Transcriptomic Breast Cancer Datasets

Version 1.0.0

Date 2018-02-25

Author Michael Zon <michaelzon7@gmail.com>, Deena M.A. Gendoo <deena.gendoo@utoronto.ca>, Benjamin Haibe-Kains <benjamin.haibe.kains@utoronto.ca>

Maintainer Michael Zon <michaelzon7@gmail.com>

Description A collection of Breast Cancer Transcriptomic Datasets that are part of the MetaGx-Data package compendium.

License Artistic-2.0

Depends Biobase, stats, lattice, impute, AnnotationHub, ExperimentHub, R (>= 3.5.0)

Suggests testthat, xtable

NeedsCompilation no

biocViews ExpressionData, CancerData, Homo_sapiens_Data, ArrayExpress, GEO, NCI, MicroarrayData, ExperimentData, BreastCancerData

LazyData yes

RoxygenNote 6.0.1

git_url <https://git.bioconductor.org/packages/MetaGxBreast>

git_branch RELEASE_3_7

git_last_commit e6eea5a

git_last_commit_date 2018-04-30

Date/Publication 2018-10-14

R topics documented:

CAL	2
DFHCC	5
DFHCC2	7
DFHCC3	9
DUKE	10
DUKE2	12
duplicates	14

EMC2	15
EORTC10994	16
EXPO	18
FNCLCC	20
GSE25066	21
GSE32646	24
GSE48091	26
GSE58644	27
HLP	29
IRB	31
KOO	33
loadBreastEsets	35
LUND	36
LUND2	38
MAINZ	39
MAQC2	41
MCCC	43
MDA4	44
METABRIC	46
MSK	49
MUG	51
NCCS	52
NCI	53
NKI	55
PNC	58
STK	60
STNO2	62
TCGA	64
TRANSBIG	66
UCSF	68
UNC4	71
UNT	73
UPP	76
VDX	78

Index	81
--------------	-----------

CAL

CAL

Description

ExpressionSet for the CAL Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/17157792
  Title:

```

URL: <http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-158/>

PMIDs: 17157792

No abstract available.

notes:

summary:

Recurrent copy number abnormalities differ between tumor subtypes as defined by gene expression patterns. Accuracy of stratification by outcome can be improved by combining expression and copy number.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
(21169 total)

varLabels: probeset gene EntrezGene.ID best_probe

varMetadata: labelDescription

Details

assayData: 21169 features, 118 samples

Platform type:

Overall survival time-to-event summary (in years):

Call: survfit(formula = Surv(time, cens) ~ -1)

1 observation deleted due to missingness

n	events	median	0.95LCL	0.95UCL
117.00	77.00	8.96	8.33	9.71

Available sample meta-data:

sample_name:

Length	Class	Mode
118	character	character

sample_type:

tumor
118

er:

negative	positive
43	75

pgr:

negative	positive	NA's
51	66	1

tumor_size:
 Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
 0.300 1.675 2.300 2.729 3.500 7.500 2

N:
 0 1
 51 67

age_at_initial_pathologic_diagnosis:
 Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
 31.00 44.00 51.00 55.06 66.00 88.00 1

grade:
 1 2 3 NA's
 10 42 61 5

dmfs_days:
 Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
 0 767 2059 2094 3336 5183 1

dmfs_status:
 norecurrence recurrence NA's
 91 26 1

days_to_tumor_recurrence:
 Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
 0 767 2059 2094 3336 5183 1

recurrence_status:
 norecurrence recurrence NA's
 81 36 1

days_to_death:
 Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
 47 1117 2234 2347 3504 5183 1

vital_status:
 deceased living
 77 41

treatment:
 chemo.plus.hormono chemotherapy hormonotherapy untreated
 25 36 40 14
 NA's
 3

batch:
 CAL
 118

uncurated_author_metadata:

Length Class Mode
 118 character character

Source

<http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-158/>

DFHCC

DFHCC

Description

ExpressionSet for the DFHCC Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2826790/
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19615
  PMIDs: 20098429
  No abstract available.
  notes:
  summary:
    A small number of over-expressed and over-amplified genes were significant
ly associated with early recurrence despite adjuvant therapy. This was ve
rified in independent cohorts.
  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
  (42447 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

assayData: 42447 features, 115 samples
 Platform type:

Available sample meta-data:

sample_name:

Length	Class	Mode
115	character	character

alt_sample_name:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
6.0	155.0	230.0	293.3	398.5	828.0

sample_type:

tumor
115

er:

negative	positive
45	70

pgr:

negative	positive
51	64

her2:

negative	positive
79	36

tumor_size:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.800	1.350	2.100	2.312	2.850	6.500

N:

0	1
62	53

age_at_initial_pathologic_diagnosis:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
32.00	45.00	53.00	53.89	60.00	85.00

grade:

1	2	3
23	28	64

dmfs_days:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
30	1500	1920	1799	2325	2640

dmfs_status:

norecurrence	recurrence
101	14

treatment:

chemo.plus.hormono	chemotherapy	hormonotherapy	untreated
42	38	22	7
NA's			
6			

batch:
DFHCC
115

uncurated_author_metadata:
Length Class Mode
115 character character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19615>

DFHCC2

DFHCC2

Description

Test the efficacy of treating TNBC with neoadjuvant cisplatin; explore biomarkers to identify predictors of response

Format

```
experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2834466/
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18864
  PMIDs: 20100965
```

Abstract: A 16 word abstract is available. Use 'abstract' method.

notes:

summary:

A subset of the patients experienced a response induced by cisplatin and biomarkers were identified that could predict response to cisplatin.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

```
featureData(eset):
```

An object of class 'AnnotatedDataFrame'

```
featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
(42447 total)
varLabels: probeset gene EntrezGene.ID best_probe
varMetadata: labelDescription
```

Details

```
assayData: 42447 features, 84 samples
Platform type:
```

```
-----
Available sample meta-data:
-----
```

```
sample_name:
  Length      Class      Mode
      84 character character
```

```
unique_patient_ID:
  Length      Class      Mode
      84 character character
```

```
sample_type:
tumor
   84
```

```
er:
negative positive
   53      31
```

```
pgr:
negative positive
   53      31
```

```
her2:
negative positive
   66      18
```

```
age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
 29.00  45.00   53.00   52.89  59.00   85.00
```

```
grade:
  1  2  3
10 16 58
```

```
treatment:
chemotherapy
      84
```

```
batch:
DFHCC2_CISPLATIN DFHCC2_REFERENCE
                24                60
```



```

uncurated_author_metadata:
  Length      Class      Mode
    84 character character

```

```

duplicates:
  Length      Class      Mode
    84 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18864>

DFHCC3	<i>DFHCC3</i>
--------	---------------

Description

ExpressionSet for the DFHCC3 Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/16473279
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3744
  PMIDs: 16473279
  No abstract available.
  notes:
    summary:
      Basal like cancerse_often lack an inactivated X chromosome.e_Other markers
      found were duplication of the active X chromosome ande_nonheterochromatin
      ized X chromosomal DNA. A small subset of X chromosomal genes were overexp
      ressed. These abnormalities are thought to led to the pathogenesis of basa
      l like cancers.
    mapping.method:
      maxRowVariance
    mapping.group:
      EntrezGene.ID
    preprocessing:
      As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
  (42447 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

assayData: 42447 features, 40 samples

Platform type:

Available sample meta-data:

sample_name:

Length	Class	Mode
40	character	character

alt_sample_name:

Length	Class	Mode
40	character	character

sample_type:

tumor
40

batch:

DFHCC3
40

uncurated_author_metadata:

Length	Class	Mode
40	character	character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3744>

DUKE

DUKE

Description

ExpressionSet for the DUKE Dataset

Format

experimentData(eset):

Experiment data

 Experimenter name:

 Laboratory:

 Contact information: <http://www.ncbi.nlm.nih.gov/pubmed/16273092>

 Title:

 URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse3143>

 PMIDs: 16273092

 No abstract available.

 notes:

summary:

It was shown that the activation_status of several oncogenic pathways can be identified by gene expression signatures. These gene signatures identify deregulation of pathways, associations with clinically relevant outcomes, and characteristics of specific cancers and tumor subtypes.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

featureNames: 1000_at 1001_at ... AFX-MurIL4_at (12085 total)

varLabels: probeset gene EntrezGene.ID best_probe

varMetadata: labelDescription

Details

assayData: 12085 features, 171 samples

Platform type:

Overall survival time-to-event summary (in years):

Call: survfit(formula = Surv(time, cens) ~ -1)

1 observation deleted due to missingness

n	events	median	0.95LCL	0.95UCL
170.00	43.00	9.01	6.22	NA

Available sample meta-data:

sample_name:

Length	Class	Mode
171	character	character

alt_sample_name:

Length	Class	Mode
171	character	character

sample_type:

tumor
171

er:

negative	positive
57	114

pgr:

negative	positive	NA's
23	65	83

```
tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
  0.20   1.80   2.30   2.74   3.50   8.50    83
```

```
N:
  0    1 NA's
 53  36  82
```

```
days_to_death:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
 171.0  417.0  957.5 1235.0 1852.0 4069.0    1
```

```
vital_status:
deceased  living    NA's
      43     127      1
```

```
batch:
DUKE
171
```

```
uncurated_author_metadata:
  Length    Class    Mode
  171 character character
```

```
duplicates:
DUKE.DUKE_T00.622 DUKE.DUKE_T01.052 DUKE.DUKE_T01.522 DUKE.DUKE_T01.534
              1              1              1              1
              NA's
              167
```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse3143>

DUKE2

DUKE2

Description

Predicting response with gene signature

Format

```
experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/18024211
  Title:
```

URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse6861>
 PMIDs: 18024211

Abstract: A 5 word abstract is available. Use 'abstract' method.

notes:

summary:

Retraction in Lancet Feb 2011 (21277543); Regimen specific signatures were able to predict pathological complete response. Selecting patients with these gene signataures could increase the proportion of patients with pCR than by basing clinical decisions on clinical factors.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

featureNames: 1053_3p_at 117_3p_at ... X79510cds_3p_s_at (45490 total)

varLabels: probeset gene EntrezGene.ID best_probe

varMetadata: labelDescription

Details

assayData: 45490 features, 160 samples

Platform type:

 Available sample meta-data:

sample_name:

Length	Class	Mode
160	character	character

alt_sample_name:

Length	Class	Mode
160	character	character

sample_type:

tumor

160

er:

negative positive

123	37
-----	----

pgr:

negative positive NA's

133	25	2
-----	----	---

```

N:
  0    1 NA's
58  95    7

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
26.00  43.00   49.00   49.41  56.00   70.00    35

grade:
  1    2    3 NA's
  2   37   70   51

treatment:
chemotherapy
      160

batch:
DUKE2
      160

uncurated_author_metadata:
  Length    Class    Mode
      160 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse6861>

duplicates	<i>a list containing the names of patients that are believed to be duplicates across datasets</i>
------------	---

Description

The object is a list where each element is a patient ID that is believed to be a duplicate of a patient in another dataset. Patients are designated as duplicated if they have Spearman correlations greater than or equal to 0.98 with other patient expression profiles

Format

A list with 107 elements, each of which is a patient ID.

EMC2

*EMC2***Description**

ExpressionSet for the EMC2 Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/19421193
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse12276
  PMIDs: 19421193
  No abstract available.
  notes:
    summary:
      Genes were identified that may increase the ability of breast cancer cells
      to infiltrate the blood-brain barrier.
    mapping.method:
      maxRowVariance
    mapping.group:
      EntrezGene.ID
    preprocessing:
      As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
  (42447 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 42447 features, 204 samples
Platform type:
-----
Available sample meta-data:
-----

sample_name:
  Length      Class      Mode
  204 character character

alt_sample_name:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.

```

1.00 51.75 102.50 102.50 153.20 204.00

sample_type:

tumor
204

N:

0 NA's
48 156

dmfs_days:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	335	640	799	1098	3507

dmfs_status:

norecurrence	recurrence
19	185

treatment:

chemotherapy	untreated
156	48

batch:

EMC2
204

uncurated_author_metadata:

Length	Class	Mode
204	character	character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse12276>

EORTC10994

EORTC10994

Description

ExpressionSet for the EORTC10994 Dataset

Format

experimentData(eset):

Experiment data

Experimenter name:

Laboratory:

Contact information: <http://www.ncbi.nlm.nih.gov/pubmed/?term=15897907>

Title:

URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1561>

PMIDs: 15897907

No abstract available.

notes:

summary:

The tumors with an apocrine gene expression profile had strong histological apocrine features. These tumors were androgen receptor positive and were all ER negative, creating further classifications of tumor cells based on steroid receptor activity- luminal which are ER and AR positive, basal that are ER and AR negative, and molecular apocrine that are ER negative and AR positive.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
(20967 total)

varLabels: probeset gene EntrezGene.ID best_probe

varMetadata: labelDescription

Details

assayData: 20967 features, 49 samples

Platform type:

Available sample meta-data:

sample_name:

Length	Class	Mode
49	character	character

alt_sample_name:

Length	Class	Mode
49	character	character

sample_type:

tumor
49

er:

negative	positive
22	27

pgr:

negative	positive	NA's
29	18	2

tumor_size:

1 2 3 4
4 23 14 8

N:

0 1
19 30

grade:

1 2 3 NA's
4 22 20 3

batch:

EORTC10994
49

uncurated_author_metadata:

Length Class Mode
49 character character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1561>

EXPO

EXPO

Description

ExpressionSet for the EXPO Dataset

Format

experimentData(eset):

Experiment data

 Experimenter name:

 Laboratory:

 Contact information:

 Title:

 URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2109>

 PMIDs:

 No abstract available.

 notes:

 summary:

 N/A

 mapping.method:

 maxRowVariance

 mapping.group:

 EntrezGene.ID

 preprocessing:

As published by original author.

```
featureData(eset):
An object of class 'AnnotatedDataFrame'
 featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
 (42447 total)
 varLabels: probeset gene EntrezGene.ID best_probe
 varMetadata: labelDescription
```

Details

```
assayData: 42447 features, 353 samples
Platform type:
```

```
-----
Available sample meta-data:
-----
```

```
sample_name:
  Length      Class      Mode
    353 character character
```

```
alt_sample_name:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  1005  21640  101100  134700  215900  486200
```

```
sample_type:
tumor
  353
```

```
er:
negative positive  NA's
    85      161    107
```

```
pgr:
negative positive  NA's
    114      129    110
```

```
her2:
negative positive  NA's
    166      61    126
```

```
age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
  25.00  45.00  55.00  59.44  67.50  95.00    1
```

```
grade:
  1  2  3 NA's
  32 114 151 56
```

```
batch:
EXPO
  353
```

```

uncurated_author_metadata:
  Length      Class      Mode
  353 character character

```

```

duplicates:
EXPO.EXPO_GSM53027 EXPO.EXPO_GSM53059      NA's
                   1                   1      351

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2109>

FNCLCC

FNCLCC

Description

ExpressionSet for the FNCLCC Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=17659439
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7017
  PMIDs: 17659439
  No abstract available.
  notes:
    summary:
      A potentially more powerful clinicogenomic model was created by combining
      a subset of relevant genes from an already published gene expression signa
      ture and a commonly used clinical prognostic model (NPI). The genes in thi
      s model are known to have a role in breast cancer, carcinogenesis, or chem
      ootherapy resistance.
    mapping.method:
      maxRowVariance
    mapping.group:
      EntrezGene.ID
    preprocessing:
      As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: UMGC_00005 UMGC_00007 ... UMGC_09018 (6064 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

assayData: 6064 features, 150 samples

Platform type:

Available sample meta-data:

sample_name:

Length	Class	Mode
150	character	character

alt_sample_name:

Length	Class	Mode
150	character	character

sample_type:

tumor
150

N:

1
150

treatment:

chemotherapy
150

batch:

FNCLCC
150

uncurated_author_metadata:

Length	Class	Mode
150	character	character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7017>

GSE25066

GSE25066

Description

ExpressionSet for the GSE25066 Dataset

Format

experimentData(eset):
Experiment data

```

Experimenter name:
Laboratory:
Contact information:
Title:
URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25066
PMIDs: 21558518
No abstract available.
notes:
  summary:

  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
  (20967 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 20967 features, 508 samples
Platform type:
-----

```

```

Available sample meta-data:
-----

```

```

sample_name:
  Length      Class      Mode
    508 character character

```

```

alt_sample_name:
  Length      Class      Mode
    508 character character

```

```

sample_type:
tumor
  508

```

```

er:
negative positive      NA's
    205      297      6

```

```

pgr:
negative positive      NA's
    258      243      7

```

```

her2:
negative positive  NA's
      485         6    17

```

```

T:
T0 T1 T2 T3 T4
  3 30 255 145 75

```

```

N:
  0  1
157 351

```

```

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  24.0   42.0   49.0   49.8   58.0   75.0

```

```

grade:
  1  2  3  4 NA's
 32 180 259 15 22

```

```

dmfs_days:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0.0   636.5   999.9  1088.0  1500.0  2717.0

```

```

dmfs_status:
norecurrence  recurrence
           397           111

```

```

batch:
GSE25066
      508

```

```

uncurated_author_metadata:
  Length    Class    Mode
     508 character character

```

```

chemosensitivity_prediction:
Rx Insensitive  Rx Sensitive
           339           169

```

```

GGI_prediction:
High Low
  336 172

```

```

PAM50_prediction:
  Basal  Her2  LumA  LumB  Normal
    189   37  160   78   44

```

```

dlda30_prediction:
pCR RD
  196 312

```

RCB_prediction:
 RCB-0/I RCB-II/III
 230 278

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25066>

GSE32646

GSE32646

Description

ExpressionSet for the GSE32646 Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information:
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32646
  PMIDs: 22320227
  No abstract available.
  notes:
  summary:

  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... 91952_at (42437 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 42437 features, 115 samples
Platform type:
-----
Available sample meta-data:
-----

```



```

sample_name:
  Length   Class      Mode
    115 character character

sample_type:
tumor
  115

er:
negative positive
   44      71

pgr:
negative positive
   70      45

her2:
negative positive
   81      34

T:
  1  2  3  4
  5 87 18  5

N:
  0  1
 32 83

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
 27.00  45.00   51.00   51.49  59.00   73.00

grade:
  1  2  3
 16 78 21

batch:
GSE32646
  115

uncurated_author_metadata:
  Length   Class      Mode
    115 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32646>

GSE48091

*GSE48091***Description**

ExpressionSet for the GSE48091 Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information:
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48091
  PMIDs: 26077471
  No abstract available.
notes:
  summary:

  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 100121619_TGI_at 100121620_TGI_at ... 100314044_TGI_at
    (23246 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 23246 features, 623 samples
Platform type:
-----
Available sample meta-data:
-----

sample_name:
  Length      Class      Mode
    623 character character

sample_type:
tumor
  623

```

```

batch:
GSE48091
  623

uncurated_author_metadata:
  Length      Class      Mode
  623 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48091>

GSE58644

GSE58644

Description

ExpressionSet for the GSE58644 Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information:
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58644
  PMIDs: 25284793
  No abstract available.
  notes:
  summary:

  mapping.method:
  maxRowVariance
  mapping.group:
  EntrezGene.ID
  preprocessing:
  As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 7896756 7896759 ... 8180179 (21462 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

assayData: 21462 features, 321 samples
Platform type:

Available sample meta-data:

sample_name:
Length Class Mode
321 character character

alt_sample_name:
Length Class Mode
321 character character

sample_type:
tumor
321

er:
negative positive NA's
70 250 1

her2:
negative positive NA's
256 58 7

tumor_size:
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.600 1.600 2.100 2.354 2.600 15.000

T:
1 2 3 4 NA's
43 59 13 1 205

N:
0 1 NA's
138 151 32

age_at_initial_pathologic_diagnosis:
Min. 1st Qu. Median Mean 3rd Qu. Max.
29.00 49.00 58.00 58.82 68.00 93.00

grade:
1 2 3 NA's
26 135 159 1

dmfs_status:
norecurrence recurrence
295 26

dmfs_days:

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0     9496   17900   21620   33600   52590

treatment:
chemo.plus.hormono      chemotherapy      hormonotherapy      untreated
           91                29                66                10
           NA's
           125

chemo:
  0  1 NA's
105 123  93

tamoxifen:
  0  1 NA's
 39 157 125

herceptin:
  0  1 NA's
190  12 119

batch:
GSE58644
  321

uncurated_author_metadata:
  Length    Class      Mode
  321 character character

duplicates:
  Length    Class      Mode
  321 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58644>

HLP

HLP

Description

ExpressionSet for the HLP Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:

```

Contact information: <http://www.ncbi.nlm.nih.gov/pubmed/?term=19688261>

Title:

URL: <http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-543/>

PMIDs: 19688261

No abstract available.

notes:

summary:

The results show evidence of different patterns of genetic aberrations in distinct molecular subtypes of breast cancer. Patterns of copy number aberrations may drive biological phenomena characteristic to each subtype.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

featureNames: 9g8cQB1TZtuiix.u1U fJUdX0IAn_P9VLTgJU ...

xopB7pPn18FJ067uDs (26536 total)

varLabels: probeset gene EntrezGene.ID best_probe

varMetadata: labelDescription

Details

assayData: 26536 features, 53 samples

Platform type:

Available sample meta-data:

sample_name:

Length	Class	Mode
53	character	character

alt_sample_name:

Length	Class	Mode
53	character	character

sample_type:

tumor
53

er:

negative	positive
28	25

pgr:

negative	positive
33	20

```

her2:
negative positive
      40      13

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
  1.200  1.800  2.450  2.648  3.000  8.000    5

N:
  0    1 NA's
 27  25   1

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
 30.00  47.50  53.50  54.96  64.25  81.00    5

grade:
  3
 53

batch:
HLP
 53

uncurated_author_metadata:
  Length    Class      Mode
    53 character character

```

Source

<http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-543/>

 IRB

 IRB

Description

ExpressionSet for the IRB Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/18297396
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5460
  PMIDs: 18297396
  No abstract available.

```

```

notes:
  summary:

  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
  (42447 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 42447 features, 129 samples
Platform type:
-----
Available sample meta-data:
-----

sample_name:
  Length      Class      Mode
    129 character character

alt_sample_name:
  Length      Class      Mode
    129 character character

sample_type:
tumor
  129

er:
negative positive
   53      76

her2:
negative positive
   98      31

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
 0.800  1.500  2.200  2.488  3.000  8.500

N:
  0 1
64 65

```



```

grade:
  1 2 3
27 32 70

treatment:
untreated
  129

batch:
IRB
129

uncurated_author_metadata:
  Length      Class      Mode
  129 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5460>

K00

KOO

Description

link does not work, in progress⁸

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/12747878
  Title:
  URL: Unavailable
  PMIDs: 12747878

```

Abstract: A 6 word abstract is available. Use 'abstract' method.

notes:

summary:

A new gene signature was used to accurately predict 90
n the study.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

```
featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at (280
  total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription
```

Details

assayData: 280 features, 88 samples

Platform type:

 Available sample meta-data:

```
sample_name:
  Length      Class      Mode
      88 character character
```

```
alt_sample_name:
  Length      Class      Mode
      88 character character
```

sample_type:

```
tumor
  88
```

er:

```
negative positive
      15      73
```

pgr:

```
negative positive
      23      65
```

tumor_size:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.20	1.80	2.30	2.74	3.50	8.50

N:

```
0 1
19 69
```

treatment:

```
chemotherapy      untreated
      61      27
```

batch:

```
KOO
  88
```

```

uncurated_author_metadata:
  Length      Class      Mode
      88 character character

```

```

duplicates:
  Length      Class      Mode
      88 character character

```

Source

Unavailable

loadBreastEsets	<i>Function to load breast cancer expression sets from the Experiment Hub</i>
-----------------	---

Description

This function returns breast cancer datasets from the hub and a vector of patients from the datasets that are most likely duplicates

Usage

```

loadBreastEsets(loadString = "majority", removeDuplicates = TRUE,
  quantileCutoff = 0, rescale = FALSE, minNumberGenes = 0,
  minNumberEvents = 0, minSampleSize = 0, removeRetracted = TRUE,
  removeSubsets = TRUE, keepCommonOnly = FALSE, imputeMissing = FALSE)

```

Arguments

loadString	a character vector specifying which data will be loaded. The default is "majority", which loads in 37 of the 39 datasets. The other option is to provide a character vector of the names of the datasets to load. The metabric and tcga datasets are loaded separately as they are very large and doing so will help prevent memory allocation errors for R windows. Furthermore, these datasets are so large that they dominate statistical analyses so it is best that they are analyzed separate of the 37 smaller datasets loaded with the string majority
removeDuplicates	remove patients with a Spearman correlation greater than or equal to 0.98 with other patient expression profiles (default TRUE)
quantileCutoff	A numeric between 0 and 1 specifying to remove genes with standard deviation below the required quantile (default 0)
rescale	apply centering and scaling to the expression sets (default FALSE)
minNumberGenes	an integer specifying to remove expression sets with less genes than this number (default 0)
minNumberEvents	an integer specifying how many survival events must be in the dataset to keep the dataset (default 0)

minSampleSize an integer specifying the minimum number of patients required in an eset (default 0)
removeRetracted remove datasets from retracted papers (default TRUE, currently just PMID17290060 dataset)
removeSubsets remove datasets that are a subset of other datasets (default TRUE, currently just PMID19318476)
keepCommonOnly remove probes not common to all datasets (default FALSE)
imputeMissing remove patients from datasets with missing expression values

Value

a list with 2 elements. The First element named `esets` contains the datasets. The second element named `duplicates` contains a vector with patient IDs for the duplicate patients (those with Spearman correlation greater than or equal to 0.98 with other patient expression profiles).

Examples

```
#Use the default loadString = "majority" if you want the 37 smaller datasets
esetsAndDups = loadBreastEsets(loadString = c("CAL", "DFHCC", "DFHCC2", "DFHCC3", "DUKE", "DUKE2", "EMC2"))
```

LUND

LUND

Description

ExpressionSet for the LUND Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=18430221
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31863
  PMIDs: 18430221
  No abstract available.
  notes:
    summary:
      A significant difference was found between the ER positive subgroup and ER
      negative subgroup in the gene expression profiles.
    mapping.method:
      maxRowVariance
    mapping.group:
      EntrezGene.ID
  preprocessing:
    As published by original author.
  
```

```
featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: H200006618 H200006808 ... H300022925 (11154 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription
```

Details

```
assayData: 11154 features, 143 samples
Platform type:
```

```
-----
Available sample meta-data:
-----
```

```
sample_name:
  Length      Class      Mode
    143 character character
```

```
alt_sample_name:
  Length      Class      Mode
    143 character character
```

```
sample_type:
tumor
  143
```

```
er:
negative positive
    29      114
```

```
pgr:
negative positive      NA's
    47      88      8
```

```
tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.   NA's
  0.200  1.100  1.500  1.486  1.800  4.000    2
```

```
N:
  0
143
```

```
age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  27.00  47.50  56.00  54.76  63.00  73.00
```

```
batch:
LUNDS1 LUNDS2 LUNDS3 LUNDS4
    30    47    22    44
```

```
uncurated_author_metadata:
  Length      Class      Mode
```

143 character character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31863>

LUND2

LUND2

Description

ExpressionSet for the LUND2 Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=17452630
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5325
  PMIDs: 17452630
  No abstract available.
  notes:
    summary:
      Microarray signature was able to show PTEN mRNA losse_when IHC was unable,
      even though tumors exhibited PTEN loss behavior. Stathmim was an accurate
      IHC marker of the signature and had prognostic significance.
    mapping.method:
      maxRowVariance
    mapping.group:
      EntrezGene.ID
    preprocessing:
      As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1 2 ... 27648 (22008 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 22008 features, 105 samples
Platform type:
-----
Available sample meta-data:
-----

```

```

sample_name:
  Length   Class   Mode
    105 character character

alt_sample_name:
  Length   Class   Mode
    105 character character

sample_type:
tumor
  105

er:
negative positive
   60      45

treatment:
hormonotherapy
      105

batch:
LUND2
  105

uncurated_author_metadata:
  Length   Class   Mode
    105 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5325>

MAINZ

MAINZ

Description

ExpressionSet for the MAINZ Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=18593943
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11121
  PMIDs: 18593943

```

No abstract available.

notes:

summary:

Poor prognosis is noted in tumors with low ER expression, showing the highest level of proliferative activity. In some tumors with highly expressed B-cell or T-cell metagenes, metastases rarely occurred, even with high proliferation and low ER expression.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
(20967 total)

varLabels: probeset gene EntrezGene.ID best_probe

varMetadata: labelDescription

Details

assayData: 20967 features, 200 samples

Platform type:

Available sample meta-data:

sample_name:

Length	Class	Mode
200	character	character

alt_sample_name:

Length	Class	Mode
200	character	character

sample_type:

tumor
200

er:

negative	positive
38	162

tumor_size:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.100	1.500	2.000	2.070	2.425	6.000

N:

0
200


```
age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  25.00  50.00   60.00   59.98  69.00   90.00
```

```
grade:
  1  2  3
  29 136 35
```

```
dmfs_days:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  30   1905   2715   2816   3855   7200
```

```
dmfs_status:
  norecurrence  recurrence
             154             46
```

```
treatment:
  untreated
      200
```

```
batch:
  MAINZ
      200
```

```
uncurated_author_metadata:
  Length    Class      Mode
  200 character character
```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11121>

MAQC2

MAQC2

Description

ExpressionSet for the MAQC2 Dataset

Format

```
experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=20064235
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20194
  PMIDs: 20064235
```

No abstract available.

notes:

summary:

It is possible to build multi-gene classifiers of clinical outcome. Prediction accuracy depends on training sample size and classification difficulty.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
(20967 total)

varLabels: probeset gene EntrezGene.ID best_probe

varMetadata: labelDescription

Details

assayData: 20967 features, 230 samples

Platform type:

Available sample meta-data:

sample_name:

Length	Class	Mode
230	character	character

alt_sample_name:

Length	Class	Mode
230	character	character

sample_type:

tumor
230

er:

negative	positive
89	141

pgr:

negative	positive
126	104

her2:

negative	positive
190	40

```

N:
  0  1
66 164

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 26.00  45.00   51.00   52.02  59.00   79.00

grade:
  1  2  3
13  94 123

treatment:
chemotherapy
      230

batch:
MAQC2
      230

uncurated_author_metadata:
  Length    Class      Mode
      230 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20194>

MCCC

MCCC

Description

ExpressionSet for the MCCC Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=19960244
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19177
  PMIDs: 19960244
  No abstract available.
notes:
  summary:
    Overall, expression and copy number profiling of familial tumors have shown that the tumors show molecular heterogeneity similar to sporadic tumors

```

and are defined by their molecular subtypes rather than BRCA1 or BRCA2 germline mutation status.

```
mapping.method:
  maxRowVariance
mapping.group:
  EntrezGene.ID
preprocessing:
  As published by original author.
```

featureData(eset):

```
An object of class 'AnnotatedDataFrame'
 featureNames: probe_10017 probe_10021 ... probe_7650767 (19048 total)
 varLabels: probeset gene EntrezGene.ID best_probe
 varMetadata: labelDescription
```

Details

assayData: 19048 features, 75 samples

Platform type:

Available sample meta-data:

sample_name:

Length	Class	Mode
75	character	character

sample_type:

tumor
75

batch:

MCCC
75

uncurated_author_metadata:

Length	Class	Mode
75	character	character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19177>

MDA4

MDA4

Description

ExpressionSet for the MDA4 Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=16896004
  Title:
  URL: http://bioinformatics.mdanderson.org/pubdata.html
  PMIDs: 16896004
  No abstract available.
notes:
  summary:
    The developed 30-probe set has high sensitivity and negative predictive value, accurately identifying 12 out of 13 patients with pCR and 27 out of 28 patients with residual disease.
  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
    (21169 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 21169 features, 129 samples
Platform type:
-----
Available sample meta-data:
-----

sample_name:
  Length      Class      Mode
    129 character character

unique_patient_ID:
  Length      Class      Mode
    129 character character

sample_type:
tumor
  129

er:
negative positive      NA's

```

```

      48      79      2

pgr:
negative positive  NA's
      73      54      2

her2:
negative positive
      114      15

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.   NA's
  0.000  0.500  1.800  2.162  3.000 10.000    8

N:
  0  1 NA's
59 62  8

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
 28.00  43.00  51.00  51.43  61.00  73.00

treatment:
chemotherapy
      129

batch:
MDA4
  129

uncurated_author_metadata:
  Length    Class    Mode
  129 character character

duplicates:
MDA4.MDA4_M207 MDA4.MDA4_M400    NA's
              1              1    127

```

Source

<http://bioinformatics.mdanderson.org/pubdata.html>

METABRIC

METABRIC

Description

ExpressionSet for the METABRIC Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/22522925
  Title:
  URL: https://www.ebi.ac.uk/ega/studies/EGAS00000000083
  PMIDs: 22522925
  No abstract available.
notes:
  summary:

  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: ILMN_1802380 ILMN_1736104 ... ILMN_1709472 (36155
    total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 36155 features, 2136 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

```

```

      165 observations deleted due to missingness
      n events median 0.95LCL 0.95UCL
1971.0  891.0   12.3   11.6   13.2

```

```

-----
Available sample meta-data:
-----

```

```

sample_name:
  Length      Class      Mode
  2136 character character

```

```

alt_sample_name:
  Length      Class      Mode
  2136 character character

```

```

sample_type:

```

healthy tumor
144 1992

er:
negative positive NA's
440 1508 188

her2:
negative positive NA's
676 148 1312

tumor_size:
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.000 1.700 2.300 2.621 3.000 18.200 164

N:
0 1 NA's
1042 950 144

age_at_initial_pathologic_diagnosis:
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
21.93 51.36 61.78 61.13 70.76 96.29 13

grade:
1 2 3 NA's
170 775 957 234

days_to_death:
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
3 1498 2632 2948 4357 9218 147

vital_status:
deceased living NA's
891 1081 164

treatment:
chemo.plus.hormono chemotherapy hormonotherapy untreated
196 226 1029 685

batch:
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
1.000 1.000 3.000 2.613 3.000 5.000 144

uncurated_author_metadata:
Length Class Mode
2136 character character

duplicates:
Length Class Mode
2136 character character

Source

<https://www.ebi.ac.uk/ega/studies/EGAS00000000083>

MSK

MSK

Description

ExpressionSet for the MSK Dataset

Format

experimentData(eset):

Experiment data

 Experimenter name:

 Laboratory:

 Contact information: <http://www.ncbi.nlm.nih.gov/pubmed/?term=16049480>

 Title:

 URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2603>

 PMIDs: 16049480

 No abstract available.

 notes:

 summary:

 A set of genes were identified that mark and mediate metastasis to the lung. Some genes confer growth advantages to both the breast tumor and lung environment, while others contribute to aggressive growth specifically in the lung.

 mapping.method:

 maxRowVariance

 mapping.group:

 EntrezGene.ID

 preprocessing:

 As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

 featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
 (20967 total)

 varLabels: probeset gene EntrezGene.ID best_probe

 varMetadata: labelDescription

Details

assayData: 20967 features, 99 samples

Platform type:

Available sample meta-data:

sample_name:

```

      Length   Class      Mode
      99 character character

alt_sample_name:
      Length   Class      Mode
      99 character character

sample_type:
tumor
  99

er:
negative positive
  42      57

pgr:
negative positive  NA's
  55      43      1

her2:
positive  NA's
  85      14

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  1.100  2.450  3.200  3.624  4.300 10.000

N:
  0 1
34 65

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  30.00  46.50  56.00  55.81  63.50  87.00

dmfs_days:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
  245  1279  1971  1888  2575  3924  17

dmfs_status:
norecurrence  recurrence  NA's
      55      27      17

batch:
MSK
  99

uncurated_author_metadata:
  Length   Class      Mode
  99 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2603>

MUG

MUG

Description

ExpressionSet for the MUG Dataset

Format

experimentData(eset):

Experiment data

 Experimenter name:

 Laboratory:

 Contact information: <http://www.ncbi.nlm.nih.gov/pubmed/?term=18592372>

 Title:

 URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse10510>

 PMIDs: 18592372

 No abstract available.

 notes:

 summary:

 A method was developed to separate tumor cells and their microenvironment to test the prognostic abilities of the immune system. Results showed that lymphatic infiltration is beneficial for ER negative patients, but probably not beneficial for ER positive patients.

 mapping.method:

 maxRowVariance

 mapping.group:

 EntrezGene.ID

 preprocessing:

 As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

 featureNames: H200000001 H200000005 ... opHsV04TC000043 (14288 total)

 varLabels: probeset gene EntrezGene.ID best_probe

 varMetadata: labelDescription

Details

assayData: 14288 features, 152 samples

Platform type:

Available sample meta-data:

sample_name:

 Length Class Mode

```

152 character character

alt_sample_name:
  Length   Class   Mode
152 character character

sample_type:
tumor
152

batch:
MUG
152

uncurated_author_metadata:
  Length   Class   Mode
152 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse10510>

NCCS

NCCS

Description

ExpressionSet for the NCCS Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=18636107
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse5364
  PMIDs: 18636107
  No abstract available.
notes:
  summary:
    48 genes were identified that displayed highly restricted levels of expres-
    sion in tumors compared to normal tissues. This was validated in 11 indepe-
    ndent cohorts of different cancer types.
  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
preprocessing:

```

As published by original author.

```
featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
  (20967 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription
```

Details

```
assayData: 20967 features, 183 samples
Platform type:
```

```
-----
Available sample meta-data:
```

```
-----
sample_name:
  Length      Class      Mode
    183 character character
```

```
alt_sample_name:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.0   46.5   92.0   92.0  137.5   183.0
```

```
sample_type:
tumor
  183
```

```
batch:
NCCS
  183
```

```
uncurated_author_metadata:
  Length      Class      Mode
    183 character character
```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse5364>

NCI

NCI

Description

ExpressionSet for the NCI Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=12917485
  Title:
  URL: Supplemental data from paper
  PMIDs: 12917485
  No abstract available.
notes:
  summary:
    Expression patterns were strongly associated with ER status, moderately as
    sociated with grade, but not associated with menopausal state, node status
    , or tumor size. Genes that were significantly associated with survival we
    re identified.
  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: AF106966 AF217974 ... Y12473 (5154 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 5154 features, 99 samples
Platform type:
-----
Available sample meta-data:
-----

sample_name:
  Length      Class      Mode
    99 character character

alt_sample_name:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
 21580  21610   21640   21650   21670   21830

sample_type:
tumor
  99

er:
negative positive

```

```

      34      65

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0.80  1.80    2.50    2.82  3.00    8.00

N:
  0  1
46 53

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
 33.00  49.00   57.00   57.47  64.50   90.00

grade:
  1  2  3
16 38 45

days_to_tumor_recurrence:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
    8    967    2057    1969  2930    4067

recurrence_status:
norecurrence  recurrence
           54           45

treatment:
  chemotherapy  hormonotherapy  untreated
             10             78             11

batch:
NCI
99

uncurated_author_metadata:
  Length    Class    Mode
    99 character character

```

Source

Supplemental data from paper

 NKI

 NKI

Description

ExpressionSet for the NKI Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=12490681; http://www.ncbi.nlm.nih.gov/pubmed/?term=11823860
  Title:
  URL: Not available
  PMIDs: 12490681, 11823860
  No abstract available.
  notes:
    summary:
      It was found that the gene expression profile that was studied was more powerful in predicting outcome of disease in younger patients than using standard clinical and pathological criteria.
    mapping.method:
      maxRowVariance
    mapping.group:
      EntrezGene.ID
    preprocessing:
      As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: Contig45645_RC Contig44916_RC ... Contig62037_RC (14960 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 14960 features, 337 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

```

```

42 observations deleted due to missingness
  n   events  median 0.95LCL 0.95UCL
295     79     NA      NA      NA

```

```

-----
Available sample meta-data:
-----

```

```

sample_name:
  Length      Class      Mode
  337 character character

```

```

alt_sample_name:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  4.0   123.0   215.0   214.1  312.0   404.0

```



```

sample_type:
tumor
  337

er:
negative positive
  88      249

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0.200  1.500   2.000   2.241  2.800   5.500

N:
  0  1
193 144

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  26.0  40.0   45.0   44.2  49.0   62.0

grade:
  1  2  3
  79 109 149

dmfs_days:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
  9  1252   2414   2546  3602   6699   18

dmfs_status:
norecurrence  recurrence  NA's
  210         109      18

days_to_tumor_recurrence:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
  9  1252   2414   2546  3602   6699   18

recurrence_status:
norecurrence  recurrence  NA's
  210         109      18

days_to_death:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
  20  1934   2637   2870  3763   6694   42

vital_status:
deceased  living  NA's
  79      216   42

treatment:
  chemotherapy  hormonotherapy  untreated
  90            40            207

```

```
batch:
  NKI NKI2
  117 220
```

```
uncurated_author_metadata:
  Length      Class      Mode
  337 character character
```

Source

Not available

PNC	<i>PNC</i>
-----	------------

Description

ExpressionSet for the PNC Dataset

Format

```
experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=21910250
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20711
  PMIDs: GSE20711, PMID 21910250
  No abstract available.
  notes:
    summary:
      Breast tumors can be further divided than the currently known expression s
ubtypes based on DNA methylation profiles.
    mapping.method:
      maxRowVariance
    mapping.group:
      EntrezGene.ID
    preprocessing:
      As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
  (42447 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription
```

Details

```

assayData: 42447 features, 92 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

      4 observations deleted due to missingness
      n events median 0.95LCL 0.95UCL
88.0   25.0     NA   11.3     NA

-----
Available sample meta-data:
-----

sample_name:
  Length   Class   Mode
    92 character character

alt_sample_name:
  Length   Class   Mode
    92 character character

sample_type:
tumor
  92

er:
negative positive  NA's
   43     45     4

pgr:
negative positive  NA's
   43     40     9

her2:
negative positive  NA's
   64     26     2

tumor_size:
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.  NA's
  0.900  1.700  2.500  2.758  3.000 10.000   6

N:
  0  1 NA's
43 40  9

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.  NA's
 32.16  48.57  53.90  55.97  64.84  82.13   4

grade:
  1  2  3 NA's

```

13 5 70 4

days_to_tumor_recurrence:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
29	967	2216	2122	2931	5139	7

recurrence_status:

norecurrence	recurrence	NA's
49	36	7

days_to_death:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
318	1940	2372	2525	3043	5139	4

vital_status:

deceased	living	NA's
25	63	4

batch:

PNC
92

uncurated_author_metadata:

Length	Class	Mode
92	character	character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20711>

STK

STK

Description

ExpressionSet for the STK Dataset

Format

experimentData(eset):

Experiment data

Experimenter name:

Laboratory:

Contact information: <http://www.ncbi.nlm.nih.gov/pubmed/?term=16280042>

Title:

URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse1456>

PMIDs: 16280042

No abstract available.

notes:

summary:

Expression profiling was able to better predict prognosis compared to histological staging.

```
mapping.method:
  maxRowVariance
mapping.group:
  EntrezGene.ID
preprocessing:
  As published by original author.
```

```
featureData(eset):
An object of class 'AnnotatedDataFrame'
 featureNames: 1007_s_at 1053_at ... 244889_at (36178 total)
 varLabels: probeset gene EntrezGene.ID best_probe
 varMetadata: labelDescription
```

Details

```
assayData: 36178 features, 159 samples
Platform type:
```

```
-----
Available sample meta-data:
```

```
sample_name:
  Length      Class      Mode
  159 character character
```

```
alt_sample_name:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  1.0   67.0   136.0   138.3  208.5   277.0
```

```
sample_type:
tumor
  159
```

```
er:
negative positive
  29      130
```

```
age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  31.0   48.0   56.0   57.8   68.5   87.0
```

```
grade:
  1    2    3 NA's
  28  58  61  12
```

```
days_to_tumor_recurrence:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  66   2022   2467   2234   2846   3099
```

```
recurrence_status:
```

```

norecurrence  recurrence
           113           46

treatment:
  chemotherapy  hormonotherapy  untreated
             89             48             22

batch:
STK
159

uncurated_author_metadata:
  Length  Class  Mode
    159 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse1456>

STNO2

STNO2

Description

ExpressionSet for the STNO2 Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=12829800
  Title:
  URL: http://smd.princeton.edu/cgi-bin/publication/viewPublication.pl?pub\_no=248
  PMIDs: 12829800
  No abstract available.
  notes:
    summary:
      Distinct breast cancer subtypes were determined by gene expression profile
s and were validated in other published datasets.
    mapping.method:
      maxRowVariance
    mapping.group:
      EntrezGene.ID
    preprocessing:
      As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'

```

```
featureNames: IMAGE:1020315 IMAGE:1030271 ... IMAGE:971399 (3663
total)
varLabels: probeset gene EntrezGene.ID best_probe
varMetadata: labelDescription
```

Details

```
assayData: 3663 features, 118 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)
```

n	events	median	0.95LCL	0.95UCL
118.00	46.00	4.67	3.34	NA

```
-----
Available sample meta-data:
-----
```

```
sample_name:
  Length      Class      Mode
  118 character character
```

```
alt_sample_name:
  Length      Class      Mode
  118 character character
```

```
sample_type:
tumor
  118
```

```
er:
negative positive  NA's
  31      82      5
```

```
tumor_size:
  1  2  3  4 NA's
  6 13 62 32  5
```

```
N:
  0  1 NA's
 34 79  5
```

```
age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
 21.00  46.25   58.00   58.47  71.75   85.00
```

```
grade:
  1  2  3 NA's
 11 49 53  5
```

```
days_to_tumor_recurrence:
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
122.0	396.0	761.0	927.9	1233.0	2800.0	23

```
recurrence_status:
norecurrence  recurrence
                58         60
```

```
days_to_death:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   91   426   898   1019  1392   5722
```

```
vital_status:
deceased  living
          46     72
```

```
treatment:
  chemotherapy  hormonotherapy  untreated
                23             73             22
```

```
batch:
STN02
  118
```

```
uncurated_author_metadata:
  Length    Class    Mode
  118 character character
```

Source

http://smd.princeton.edu/cgi-bin/publication/viewPublication.pl?pub_no=248

TCGA

TCGA

Description

ExpressionSet for the TCGA Dataset

Format

```
experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/23000897
  Title:
  URL: http://cancergenome.nih.gov/
  PMIDs: 23000897
  No abstract available.
  notes:
```


summary:

```
mapping.method:
  maxRowVariance
mapping.group:
  EntrezGene.ID
preprocessing:
  As published by original author.
```

featureData(eset):

```
An object of class 'AnnotatedDataFrame'
 featureNames: A1BG A2M ... ARHGAP11A.2 (19504 total)
 varLabels: probeset gene EntrezGene.ID best_probe
 varMetadata: labelDescription
```

Details

```
assayData: 19504 features, 1073 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)
```

	n	events	median	0.95LCL	0.95UCL
	1073.00	103.00	10.05	8.56	12.05

```
-----
Available sample meta-data:
-----
```

```
sample_name:
  Length      Class      Mode
  1073 character character
```

```
alt_sample_name:
  Length      Class      Mode
  1073 character character
```

```
unique_patient_ID:
  Length      Class      Mode
  1073 character character
```

```
sample_type:
tumor
  1073
```

```
er:
negative positive  NA's
  233      790      50
```

```
pgr:
negative positive  NA's
  334      686      53
```

```

her2:
negative positive    NA's
      549      161      363

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 26.00  49.00   58.00   58.48  68.00   90.00

days_to_death:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -7.0  137.0   412.0   820.3 1180.0  6796.0

vital_status:
deceased  living
      103     970

batch:
TCGA
1073

uncurated_author_metadata:
  Length    Class      Mode
  1073 character character

```

Source

<http://cancergenome.nih.gov/>

TRANSBIG

TRANSBIG

Description

ExpressionSet for the TRANSBIG Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=17545524
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gsE7390
  PMIDs: 17545524
  No abstract available.
notes:
  summary:
    The 76-gene signature was validated. The results supports the hypothesis t

```

hat utilizing the gene signature could reduce the number of patients who receive unnecessary adjuvant therapy.

```
mapping.method:
  maxRowVariance
mapping.group:
  EntrezGene.ID
preprocessing:
  As published by original author.
```

```
featureData(eset):
An object of class 'AnnotatedDataFrame'
 featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
 (20967 total)
 varLabels: probeset gene EntrezGene.ID best_probe
 varMetadata: labelDescription
```

Details

```
assayData: 20967 features, 198 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)
```

n	events	median	0.95LCL	0.95UCL
198.0	56.0	NA	17.1	NA

```
-----
Available sample meta-data:
-----
```

```
sample_name:
  Length      Class      Mode
  198 character character
```

```
sample_type:
tumor
  198
```

```
er:
negative positive
  64      134
```

```
tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0.600  1.700  2.000  2.181  2.500  5.000
```

```
N:
  0
  198
```

```
age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
```

24.00 42.00 46.00 46.39 51.00 60.00

grade:

1	2	3	NA's
30	83	83	2

dmfs_days:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
125	2375	4384	3954	5566	9108

dmfs_status:

norecurrence	recurrence
147	51

days_to_tumor_recurrence:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
121	1528	3534	3399	5130	8711

recurrence_status:

norecurrence	recurrence
112	86

days_to_death:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
146	2744	4562	4150	5610	9108

vital_status:

deceased	living
56	142

treatment:

untreated
198

batch:

VDXGUYU	VDXIGRU	VDXKIU	VDXOXFU	VDXRHU
36	50	51	24	37

uncurated_author_metadata:

Length	Class	Mode
198	character	character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gsE7390>

Description

ExpressionSet for the UCSF Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=17428335; http://www.ncbi.nlm.nih.gov/pubmed/?term=14612510
  Title:
  URL: Not available
  PMIDs: 17428335, 14612510
  No abstract available.
notes:
  summary:
    A gene set was identified that correctly predicted outcomes more effective
ly than using histological markers.
  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: probe_1 probe_3 ... probe_10365 (8015 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 8015 features, 162 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

```

```

      29 observations deleted due to missingness
      n  events  median 0.95LCL 0.95UCL
133.00  44.00  11.56   9.25     NA

```

```

-----
Available sample meta-data:
-----

```

```

sample_name:
  Length      Class      Mode
    162 character character

```

```

alt_sample_name:

```

```

      Length   Class      Mode
      162 character character

sample_type:
tumor
  162

er:
negative positive  NA's
   41      101      20

pgr:
negative positive  NA's
   46      94      22

her2:
negative positive  NA's
   35      19      108

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
  0.000  1.800   2.000   2.682  3.200 11.000    7

N:
  0  1 NA's
 67 82 13

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
 28.00  44.00   53.00   56.61  70.00  88.00    9

grade:
  1  2  3 NA's
 14 62 74 12

dmfs_days:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
   47   897   2040   2084  2992  8267   29

dmfs_status:
norecurrence  recurrence
      140           22

days_to_tumor_recurrence:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
   47   861   1865   1985  2847  8267   29

recurrence_status:
norecurrence  recurrence
      125           37

days_to_death:

```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
47	1087	2054	2140	3087	8267	29

vital_status:

deceased	living	NA's
54	99	9

treatment:

chemo.plus.hormono	chemotherapy	hormonotherapy	untreated
31	38	61	22
NA's			
10			

batch:

UCSF
162

uncurated_author_metadata:

Length	Class	Mode
162	character	character

Source

Not available

UNC4

UNC4

Description

ExpressionSet for the UNC4 Dataset

Format

```
experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=20813035
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse18229
  PMIDs: 20813035
  No abstract available.
notes:
  summary:
    Clinically, this subtype is usually triple negative invasive ductal carcinomas with a poor prognosis. Response to standard of care preoperative chemotherapy is intermediate between basal-like and luminal tumors. The claudin-low subtype most closely resembles the mammary epithelial stem cell.
  mapping.method:
```

```

      maxRowVariance
mapping.group:
  EntrezGene.ID
preprocessing:
  As published by original author.

```

```

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: probe.10 probe.12 ... probe.79701 (5420 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 5420 features, 305 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

```

```

      65 observations deleted due to missingness
      n events median 0.95LCL 0.95UCL
240.00  51.00   7.73   6.82    NA

```

```

-----
Available sample meta-data:
-----

```

```

sample_name:
  Length      Class      Mode
    305 character character

```

```

sample_type:
tumor
  305

```

```

er:
negative positive  NA's
    99      154     52

```

```

pgr:
negative positive  NA's
    126      109     70

```

```

her2:
negative positive  NA's
    203      58     44

```

```

tumor_size:
  1  1.5  3  6 NA's
60  1 129 43 72

```

```

N:

```



```

0      1 NA's
126 135  44

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
 24.00  46.00   55.00   56.73  68.00   89.00    59

grade:
  1      2      3 NA's
 25     80   138   62

days_to_tumor_recurrence:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
  30.0  450.0   750.0   954.3 1380.0  3540.0    64

recurrence_status:
norecurrence  recurrence          NA's
           170           70           65

days_to_death:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
   30    540    885   1104   1590   5190    65

vital_status:
deceased  living    NA's
     51    189     65

batch:
UNC4
305

uncurated_author_metadata:
  Length    Class      Mode
    305 character character

duplicates:
K00.K00_KF_105 K00.K00_T01_514          NA's
              1              1          303

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse18229>

 UNT

 UNT

Description

ExpressionSet for the UNT Dataset

Format

```

experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=16478745; http://www.ncbi.nlm.nih.gov/pubmed/16478745
  Title:
  URL: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2990
  PMIDs: 16478745, 17401012
  No abstract available.
notes:
  summary:
    A gene expression grading index (GGI) was developed. The GGI reclassified
    grade 2 patients into two groups with low and high risks of recurrence.
  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
  featureNames: 1007_s_at 1053_at ... 244889_at (36084 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

```

Details

```

assayData: 36084 features, 133 samples
Platform type:
-----
Available sample meta-data:
-----

sample_name:
  Length      Class      Mode
  133 character character

alt_sample_name:
  Length      Class      Mode
  133 character character

sample_type:
tumor
  133

er:
negative positive      NA's
  40      86      7

```

```

pgr:
negative positive  NA's
      6      56      71

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0.000  1.200  1.900  1.892  2.300  6.000

N:
  0
133

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  24.00  44.00  53.00  51.79  60.00  73.00

grade:
  1  2  3 NA's
  32 51 29 21

dmfs_days:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
   61  1338  2809  2724  4078  5305

dmfs_status:
norecurrence  recurrence  NA's
              97          28          8

days_to_tumor_recurrence:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
   61  1338  2675  2687  3912  5305

recurrence_status:
norecurrence  recurrence  NA's
              76          49          8

treatment:
untreated
  133

batch:
KIU OXFU
  64  69

uncurated_author_metadata:
  Length  Class  Mode
  133 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2990>

UPP

UPP

Description

ExpressionSet for the UPP Dataset

Format

experimentData(eset):

Experiment data

Experimenter name:

Laboratory:

 Contact information: <http://www.ncbi.nlm.nih.gov/pubmed/?term=16141321>

Title:

 URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse3494>

PMIDs: 16141321

No abstract available.

notes:

summary:

 A 32-gene expression signature of p53 was identified that differentiates p53 mutant and wild-type tumors. The signature is more effective than sequence-based assessments of p53 in predicting prognosis and therapeutic response.

mapping.method:

maxRowVariance

mapping.group:

EntrezGene.ID

preprocessing:

As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

featureNames: 1007_s_at 1053_at ... 244889_at (36178 total)

varLabels: probeset gene EntrezGene.ID best_probe

varMetadata: labelDescription

Details

assayData: 36178 features, 251 samples

Platform type:

Available sample meta-data:

sample_name:

Length Class Mode

251 character character

alt_sample_name:

```

      Length      Class      Mode
      251 character character

sample_type:
tumor
  251

er:
negative positive  NA's
   34      213      4

pgr:
negative positive
   61      190

tumor_size:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0.200  1.500   2.000   2.243  2.562  13.000

N:
  0      1 NA's
 158  84   9

age_at_initial_pathologic_diagnosis:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  28.00  52.00   64.00   62.11  72.00   93.00

grade:
  1      2      3 NA's
  67  128   54   2

days_to_tumor_recurrence:
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.  NA's
   30   1870   3711   3007   3985   4654   17

recurrence_status:
norecurrence  recurrence      NA's
           181           55           15

treatment:
hormonotherapy  untreated      NA's
           80           142           29

batch:
UPPT UPPU
  80  171

uncurated_author_metadata:
  Length      Class      Mode
  251 character character

```

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse3494>

VDX

VDX

Description

ExpressionSet for the VDX Dataset

Format

experimentData(eset):

Experiment data

 Experimenter name:

 Laboratory:

 Contact information: <http://www.ncbi.nlm.nih.gov/pubmed/?term=15721472>; <http://www.ncbi.nlm.nih.gov/pubmed/?term=17420468>

 Title:

 URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2034>; <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse3494>

 PMIDs: 15721472, 17420468

 No abstract available.

 notes:

 summary:

 15721472: A gene signature was identified that can accurately predict distant metastases in node-negative cases. 17420468: Tumors with a lung metastatic gene signature were shown to be larger.

 mapping.method:

 maxRowVariance

 mapping.group:

 EntrezGene.ID

 preprocessing:

 As published by original author.

featureData(eset):

An object of class 'AnnotatedDataFrame'

 featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at
 (21169 total)

 varLabels: probeset gene EntrezGene.ID best_probe

 varMetadata: labelDescription

Details

assayData: 21169 features, 344 samples

Platform type:

Available sample meta-data:

sample_name:

 Length Class Mode

344 character character

alt_sample_name:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.0	122.8	605.5	575.7	836.5	2038.0

sample_type:

tumor
344

er:

negative	positive
135	209

tumor_size:

1	2	3	4	NA's
146	132	5	3	58

N:

0
344

age_at_initial_pathologic_diagnosis:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
26.00	44.00	52.00	53.88	63.00	83.00	58

grade:

1	2	3	NA's
7	42	148	147

dmfs_days:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
61	1254	2616	2377	3285	5201

dmfs_status:

norecurrence	recurrence
226	118

treatment:

untreated
344

batch:

VDX	VDXN
286	58

uncurated_author_metadata:

Length	Class	Mode
344	character	character

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2034>; <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse5>

Index

*Topic **datasets**

CAL, [2](#)
DFHCC, [5](#)
DFHCC2, [7](#)
DFHCC3, [9](#)
DUKE, [10](#)
DUKE2, [12](#)
duplicates, [14](#)
EMC2, [15](#)
EORTC10994, [16](#)
EXPO, [18](#)
FNCLCC, [20](#)
GSE25066, [21](#)
GSE32646, [24](#)
GSE48091, [26](#)
GSE58644, [27](#)
HLP, [29](#)
IRB, [31](#)
KOO, [33](#)
LUND, [36](#)
LUND2, [38](#)
MAINZ, [39](#)
MAQC2, [41](#)
MCCC, [43](#)
MDA4, [44](#)
METABRIC, [46](#)
MSK, [49](#)
MUG, [51](#)
NCCS, [52](#)
NCI, [53](#)
NKI, [55](#)
PNC, [58](#)
STK, [60](#)
STNO2, [62](#)
TCGA, [64](#)
TRANSBIG, [66](#)
UCSF, [68](#)
UNC4, [71](#)
UNT, [73](#)
UPP, [76](#)
VDX, [78](#)

CAL (CAL), [2](#)
DFHCC (DFHCC), [5](#)
DFHCC2 (DFHCC2), [7](#)
DFHCC3 (DFHCC3), [9](#)
DUKE (DUKE), [10](#)
DUKE2 (DUKE2), [12](#)
EMC2 (EMC2), [15](#)
EORTC10994 (EORTC10994), [16](#)
EXPO (EXPO), [18](#)
FNCLCC (FNCLCC), [20](#)
GSE25066 (GSE25066), [21](#)
GSE32646 (GSE32646), [24](#)
GSE48091 (GSE48091), [26](#)
GSE58644 (GSE58644), [27](#)
HLP (HLP), [29](#)
IRB (IRB), [31](#)
KOO (KOO), [33](#)
LUND (LUND), [36](#)
LUND2 (LUND2), [38](#)
MAINZ (MAINZ), [39](#)
MAQC2 (MAQC2), [41](#)
MCCC (MCCC), [43](#)
MDA4 (MDA4), [44](#)
METABRIC (METABRIC), [46](#)
MSK (MSK), [49](#)
MUG (MUG), [51](#)
NCCS (NCCS), [52](#)
NCI (NCI), [53](#)
NKI (NKI), [55](#)
PNC (PNC), [58](#)
STK (STK), [60](#)
STNO2 (STNO2), [62](#)
TCGA (TCGA), [64](#)
TRANSBIG (TRANSBIG), [66](#)
UCSF (UCSF), [68](#)
UNC4 (UNC4), [71](#)
UNT (UNT), [73](#)
UPP (UPP), [76](#)
VDX (VDX), [78](#)

CAL, [2](#)
DFHCC, [5](#)
DFHCC2, [7](#)
DFHCC3, [9](#)
DUKE, [10](#)

DUKE2, [12](#)
duplicates, [14](#)

EMC2, [15](#)
EORTC10994, [16](#)
EXPO, [18](#)

FNCLCC, [20](#)

GSE25066, [21](#)
GSE32646, [24](#)
GSE48091, [26](#)
GSE58644, [27](#)

HLP, [29](#)

IRB, [31](#)

KOO, [33](#)

loadBreastEsets, [35](#)
LUND, [36](#)
LUND2, [38](#)

MAINZ, [39](#)
MAQC2, [41](#)
MCCC, [43](#)
MDA4, [44](#)
METABRIC, [46](#)
MSK, [49](#)
MUG, [51](#)

NCCS, [52](#)
NCI, [53](#)
NKI, [55](#)

PNC, [58](#)

STK, [60](#)
STNO2, [62](#)

TCGA, [64](#)
TRANSBIG, [66](#)

UCSF, [68](#)
UNC4, [71](#)
UNT, [73](#)
UPP, [76](#)

VDX, [78](#)