Package ‘MetaGxBreast’

February 27, 2024

Type Package

Title Transcriptomic Breast Cancer Datasets

Version 1.22.0

Date 2020-04-23

Description A collection of Breast Cancer Transcriptomic Datasets that are part of the MetaGxData package compendium.

License Apache License (>= 2)

Depends R (>= 3.6.0), Biobase, AnnotationHub, ExperimentHub

Imports stats, lattice, impute, SummarizedExperiment

Suggests testthat, xtable, tinytex

NeedsCompilation no

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

LazyData yes

RoxygenNote 7.1.1

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

git_branch RELEASE_3_18

git_last_commit a3c220a

git_last_commit_date 2023-10-24

Repository Bioconductor 3.18

Date/Publication 2024-02-27

Author Michael Zon [aut], Deena M.A. Gendoo [aut], Christopher Eeles [ctb], Benjamin Haibe-Kains [aut, cre]

Maintainer Benjamin Haibe-Kains <benjamin.haibe.kains@utoronto.ca>
### R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAL</td>
<td>3</td>
</tr>
<tr>
<td>DFHCC</td>
<td>5</td>
</tr>
<tr>
<td>DFHCC2</td>
<td>8</td>
</tr>
<tr>
<td>DFHCC3</td>
<td>10</td>
</tr>
<tr>
<td>DUKE</td>
<td>11</td>
</tr>
<tr>
<td>DUKE2</td>
<td>13</td>
</tr>
<tr>
<td>duplicates</td>
<td>15</td>
</tr>
<tr>
<td>EMC2</td>
<td>16</td>
</tr>
<tr>
<td>EORTC10994</td>
<td>18</td>
</tr>
<tr>
<td>EXPO</td>
<td>20</td>
</tr>
<tr>
<td>FNCLCC</td>
<td>22</td>
</tr>
<tr>
<td>GSE25066</td>
<td>23</td>
</tr>
<tr>
<td>GSE32646</td>
<td>26</td>
</tr>
<tr>
<td>GSE48091</td>
<td>28</td>
</tr>
<tr>
<td>GSE58644</td>
<td>29</td>
</tr>
<tr>
<td>HLP</td>
<td>32</td>
</tr>
<tr>
<td>IRB</td>
<td>34</td>
</tr>
<tr>
<td>KOO</td>
<td>36</td>
</tr>
<tr>
<td>loadBreastDatasets</td>
<td>38</td>
</tr>
<tr>
<td>loadBreastEsets</td>
<td>39</td>
</tr>
<tr>
<td>LUND</td>
<td>40</td>
</tr>
<tr>
<td>LUND2</td>
<td>42</td>
</tr>
<tr>
<td>MAINZ</td>
<td>44</td>
</tr>
<tr>
<td>MAQC2</td>
<td>46</td>
</tr>
<tr>
<td>MCCC</td>
<td>48</td>
</tr>
<tr>
<td>MDA4</td>
<td>49</td>
</tr>
<tr>
<td>METABRIC</td>
<td>51</td>
</tr>
<tr>
<td>MSK</td>
<td>54</td>
</tr>
<tr>
<td>MUG</td>
<td>56</td>
</tr>
<tr>
<td>NCCS</td>
<td>57</td>
</tr>
<tr>
<td>NCI</td>
<td>59</td>
</tr>
<tr>
<td>NKI</td>
<td>61</td>
</tr>
<tr>
<td>PNC</td>
<td>64</td>
</tr>
<tr>
<td>STK</td>
<td>66</td>
</tr>
<tr>
<td>STNO2</td>
<td>68</td>
</tr>
<tr>
<td>TCGA</td>
<td>71</td>
</tr>
<tr>
<td>TRANSBIG</td>
<td>73</td>
</tr>
<tr>
<td>UCSF</td>
<td>75</td>
</tr>
<tr>
<td>UNC4</td>
<td>78</td>
</tr>
<tr>
<td>UNT</td>
<td>81</td>
</tr>
<tr>
<td>UPP</td>
<td>83</td>
</tr>
<tr>
<td>VDX</td>
<td>85</td>
</tr>
</tbody>
</table>

**Index** 88
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/
  PMID: 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
<table>
<thead>
<tr>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-158/">http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-158/</a></td>
</tr>
</tbody>
</table>

**Description**

ExpressionSet for the DFHCC Dataset

**Format**

```r
experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2826790/
```
Title:
PMIDs: 20098429
No abstract available.
notes:
summary:
A small number of over-expressed and over-amplified genes were significantly associated with early recurrence despite adjuvant therapy. This was verified 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
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:
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


---

**DFHCC3**

---

**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:
- PMIDs: 16473279
- No abstract available.

notes:

summary:

Basal like cancers often lack an inactivated X chromosome. Other markers found were duplication of the active X chromosome and nonheterochromatinized X chromosomal DNA. A small subset of X chromosomal genes were overexpressed. These abnormalities are thought to lead to the pathogenesis of basal-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)

dataLabels: probeset gene EntrezGene.ID best_probe

dataMetadata: labelDescription

---

**Details**

assayData: 42447 features, 40 samples

Platform type:
DUKE

Available sample meta-data:

sample_name:

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

alt_sample_name:

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

sample_type:
tumor

40

batch:
DFHCC3

40

uncurated_author_metadata:

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

Source


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:
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 ... AFFX-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

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:
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 signatures 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
<table>
<thead>
<tr>
<th>Sample Type</th>
<th>Tumor</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>ER</th>
<th>Negative</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>123</td>
<td>37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PR</th>
<th>Negative</th>
<th>Positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>133</td>
<td>25</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>0</th>
<th>1</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58</td>
<td>95</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age at Initial Pathologic Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>26.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Chemotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Batch</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUKE2</td>
</tr>
<tr>
<td>160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Uncurated Author Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>160</td>
</tr>
</tbody>
</table>

**Source**


**Duplicates**

A list containing the names of patients that are believed to be duplicates across datasets.
**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.

---

**Description**

ExpressionSet for the EMC2 Dataset

**Format**

```r
eperimentData(eset):
  Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/19421193
  Title:
  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.

eatureData(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

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:
  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: labelIDDescription

Details

assayData: 20967 features, 49 samples
Platform type:

Available sample meta-data:
<table>
<thead>
<tr>
<th>sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>49 character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alt_sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>49 character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type:</th>
<th>tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>er:</th>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>22</td>
<td>27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pgr:</th>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29</td>
<td>18</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tumor_size:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>4</th>
<th>23</th>
<th>14</th>
<th>8</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>N:</th>
<th>0</th>
<th>1</th>
<th>19</th>
<th>30</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>grade:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>22</td>
<td>20</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
<th>EORTC10994</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>uncurated_author_metadata:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>49 character</td>
</tr>
</tbody>
</table>

**Source**

Description

ExpressionSet for the EXPO Dataset

Format

experimentData(eset):
Experiment data
   Experimenter name:
   Laboratory:
   Contact information:
   Title:
   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:

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1005</td>
<td>21640</td>
<td>101100</td>
<td>134700</td>
<td>215900</td>
<td>486200</td>
</tr>
</tbody>
</table>

sample_type:
tumor

353

er:

<table>
<thead>
<tr>
<th></th>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>85</td>
<td>161</td>
<td>107</td>
</tr>
</tbody>
</table>

pgr:

<table>
<thead>
<tr>
<th></th>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>114</td>
<td>129</td>
<td>110</td>
</tr>
</tbody>
</table>

her2:

<table>
<thead>
<tr>
<th></th>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>166</td>
<td>61</td>
<td>126</td>
</tr>
</tbody>
</table>

age_at_initial_pathologic_diagnosis:

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25.00</td>
<td>45.00</td>
<td>55.00</td>
<td>59.44</td>
<td>67.50</td>
<td>95.00</td>
<td>1</td>
</tr>
</tbody>
</table>

grade:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>114</td>
<td>151</td>
<td>56</td>
</tr>
</tbody>
</table>

batch:

EXPO

353

uncurated_author_metadata:

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>353</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

duplicates:

<table>
<thead>
<tr>
<th>duplicates</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPO.EXPO_GSM53027 EXPO.EXPO_GSM53059</td>
<td>1</td>
</tr>
<tr>
<td>EXPO.EXPO_GSM53027 EXPO.GSM53059</td>
<td>1</td>
</tr>
<tr>
<td>EXPO.GSM53027 EXPO.GSM53059</td>
<td>351</td>
</tr>
</tbody>
</table>

Source

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:
   PMID: 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
         otherapy 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:
-----------------------------
<table>
<thead>
<tr>
<th>attribute</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample_name:</td>
<td>Length Class Mode 150 character character</td>
</tr>
<tr>
<td>alt_sample_name:</td>
<td>Length Class Mode 150 character character</td>
</tr>
<tr>
<td>sample_type:</td>
<td>tumor 150</td>
</tr>
<tr>
<td>N:</td>
<td>1 150</td>
</tr>
<tr>
<td>treatment:</td>
<td>chemotherapy 150</td>
</tr>
<tr>
<td>batch:</td>
<td>FNCLCC 150</td>
</tr>
<tr>
<td>uncurated_author_metadata:</td>
<td>Length Class Mode 150 character character</td>
</tr>
</tbody>
</table>

**Source**


**Description**

ExpressionSet for the GSE25066 Dataset

**Format**

experimentData(eset):

- Experiment data
  - Experimenter name:
  - Laboratory:
  - Contact information:
Title:
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
GSE25066

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


GSE32646

Description

ExpressionSet for the GSE32646 Dataset

Format

experimentData(eset):
  Experiment data
  Experimenter name:
  Laboratory:
  Contact information:
  Title:
  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:

<table>
<thead>
<tr>
<th>sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>115 character</td>
<td>character</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type:</th>
<th>tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>115</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>er:</th>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>44</td>
<td>71</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pgr:</th>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>her2:</th>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>81</td>
<td>34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T:</th>
<th>1 2 3 4 5 87 18 5</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>N:</th>
<th>0 1 32 83</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>age_at_initial_pathologic_diagnosis:</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27.00</td>
<td>45.00</td>
<td>51.00</td>
<td>51.49</td>
<td>59.00</td>
<td>73.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>grade:</th>
<th>1 2 3 16 78 21</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
<th>GSE32646 115</th>
</tr>
</thead>
</table>

| uncurated_author_metadata: |
Length Class Mode
115 character character

Source


Description

ExpressionSet for the GSE48091 Dataset

Format

experimentData(eset):
Experiment data
Experimenter name:
Laboratory:
Contact information:
Title:
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


Description

ExpressionSet for the GSE58644 Dataset

Format

experimentData(eset):
Experiment data
   Experimenter name:
   Laboratory:
   Contact information:
   Title:
   PMIDs: 25284793
   No abstract available.
   notes:
GSE58644

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
Source


<table>
<thead>
<tr>
<th>HLP</th>
<th>HLP</th>
</tr>
</thead>
</table>

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 ... xopB7pPn18FJ067u0s (26536 total)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription
Details

 assayData: 26536 features, 53 samples

 Platform type:

 Available sample meta-data:

<table>
<thead>
<tr>
<th>sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alt_sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type:</th>
<th>tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>er:</th>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pgr:</th>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>her2:</th>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tumor_size:</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.200</td>
<td>1.800</td>
<td>2.450</td>
<td>2.648</td>
<td>3.000</td>
<td>8.000</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N:</th>
<th>0</th>
<th>1</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27</td>
<td>25</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>age_at_initial_pathologic_diagnosis:</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30.00</td>
<td>47.50</td>
<td>53.50</td>
<td>54.96</td>
<td>64.25</td>
<td>81.00</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>grade:</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
</tr>
</thead>
</table>
uncurated_author_metadata:
  Length  Class  Mode
  53  character  character

Source

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

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:
  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


Description

link does not work, in progress

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
      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

<table>
<thead>
<tr>
<th>Platform type:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available sample meta-data:</td>
</tr>
</tbody>
</table>

| 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 |

<p>| tumor_size: |</p>
<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>1.80</td>
<td>2.30</td>
<td>2.74</td>
<td>3.50</td>
<td>8.50</td>
</tr>
</tbody>
</table>

| N: |
| 0 1 |
| 19 69 |

| treatment: |
| chemotherapy | untreated |
| 61 | 27 |

| batch: |
| KOO |
| 88 |

| uncurated_author_metadata: |
| Length Class Mode |
| 88 character character |

| duplicates: |
loadBreastDatasets

Function to load breast cancer SummarizedExperiment objects from the Experiment Hub

Description

This function returns breast cancer datasets from the hub and a vector of patients from the datasets that are duplicates based on a spearman correlation > 0.98

Usage

loadBreastDatasets(
  rescale = FALSE,
  minNumberGenes = 0,
  minNumberEvents = 0,
  minSampleSize = 0,
  keepCommonOnly = FALSE,
  imputeMissing = FALSE,
  removeDuplicates = FALSE
)

Arguments

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 a SummarizedExperiment (default 0)
keepCommonOnly  remove entrezIDs not common to all datasets (default FALSE)
imputeMissing   remove patients from datasets with missing expression values
removeDuplicates remove patients with a Spearman correlation greater than or equal to 0.98 with other patient expression profiles (default TRUE)
Value

A `list` with 2 elements. The First element named `SummarizedExperiment`'s 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).

---

`loadBreastEsets`  
*Function to load breast cancer expression sets from the Experiment Hub*

Description

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

Usage

```r
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 tega 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 man 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

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

---

**LUND**

ExpressionSet for the LUND Dataset

**Format**

- **experimentData(eiset):**
  - **Experiment data**
    - Experimenter name:
    - Laboratory:
  - **Title:**
  - **PMIDs:** 18430221
  - No abstract available.
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:

<table>
<thead>
<tr>
<th>sample_name</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample_name</td>
<td>143</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alt_sample_name</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>alt_sample_name</td>
<td>143</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample_type</td>
<td>143</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>er</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>er</td>
<td>143</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pgr</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>pgr</td>
<td>143</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tumor_size</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>tumor_size</td>
<td>0.200</td>
<td>1.100</td>
<td>1.500</td>
<td>1.486</td>
<td>1.800</td>
<td>4.000</td>
<td>2</td>
</tr>
</tbody>
</table>
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

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:
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**


---

**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:
  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:

<table>
<thead>
<tr>
<th></th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample_name:</td>
<td>200</td>
<td>character</td>
<td>character</td>
</tr>
<tr>
<td>alt_sample_name:</td>
<td>200</td>
<td>character</td>
<td>character</td>
</tr>
<tr>
<td>sample_type:</td>
<td>tumor</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>er:</td>
<td>negative</td>
<td>positive</td>
<td>38</td>
</tr>
<tr>
<td>tumor_size:</td>
<td>Min. 1st Qu.</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>1.500</td>
<td>2.000</td>
</tr>
<tr>
<td>N:</td>
<td>0</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>age_at_initial_pathologic_diagnosis:</td>
<td>Min. 1st Qu.</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>25.00</td>
<td>50.00</td>
<td>60.00</td>
</tr>
<tr>
<td>grade:</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>136</td>
<td>35</td>
</tr>
<tr>
<td>dmfs_days:</td>
<td>Min. 1st Qu.</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1905</td>
<td>2715</td>
</tr>
<tr>
<td>dmfs_status:</td>
<td>no recurrence</td>
<td>recurrence</td>
<td>154</td>
</tr>
<tr>
<td>treatment:</td>
<td>untreated</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>batch:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MAQC2

200

uncurated_author_metadata:
Length  Class  Mode
200 character character

Source

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:
  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
batch:
MAQC2
230

uncurated_author_metadata:
Length Class Mode
230 character character

Source


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:
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

MDA4
---

Description
ExpressionSet for the MDA4 Dataset

Format
eperimentData(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: 73
- positive: 54
- NA's: 2

**her2:**
- negative: 114
- positive: 15

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

**N:**
- 0: 1
- 1: 59
- NA's: 62

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

**treatment:**
- chemotherapy: 129

**batch:**
- MDA4: 129

**uncurated_author_metadata:**
- Length: 129
- Class: character
- Mode: character

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

---

**Source**

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

---

**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:
<table>
<thead>
<tr>
<th>Sample Type</th>
<th>Healthy</th>
<th>Tumor</th>
<th>144</th>
<th>1992</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>Negative</td>
<td>Positive</td>
<td>NA's</td>
<td></td>
</tr>
<tr>
<td></td>
<td>440</td>
<td>1508</td>
<td>188</td>
<td></td>
</tr>
<tr>
<td>HER2</td>
<td>Negative</td>
<td>Positive</td>
<td>NA's</td>
<td></td>
</tr>
<tr>
<td></td>
<td>676</td>
<td>148</td>
<td>1312</td>
<td></td>
</tr>
<tr>
<td>Tumor Size</td>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
<td>3rd Qu.</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>1.700</td>
<td>2.300</td>
<td>2.621</td>
</tr>
<tr>
<td>N</td>
<td>0</td>
<td>1</td>
<td>NA's</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1042</td>
<td>950</td>
<td>144</td>
<td></td>
</tr>
<tr>
<td>Age at Initial Pathologic Diagnosis</td>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
<td>3rd Qu.</td>
</tr>
<tr>
<td></td>
<td>21.93</td>
<td>51.36</td>
<td>61.78</td>
<td>61.13</td>
</tr>
<tr>
<td>Grade</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>NA's</td>
</tr>
<tr>
<td></td>
<td>170</td>
<td>775</td>
<td>957</td>
<td>234</td>
</tr>
<tr>
<td>Days to Death</td>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
<td>3rd Qu.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1498</td>
<td>2632</td>
<td>2948</td>
</tr>
<tr>
<td>Vital Status</td>
<td>Deceased</td>
<td>Living</td>
<td>NA's</td>
<td></td>
</tr>
<tr>
<td></td>
<td>891</td>
<td>1081</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>Chemo Plus Hormono</td>
<td>Chemotherapy</td>
<td>Hormonotherapy</td>
<td>Untreated</td>
</tr>
<tr>
<td></td>
<td>196</td>
<td>226</td>
<td>1029</td>
<td>685</td>
</tr>
<tr>
<td>Batch</td>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
<td>3rd Qu.</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>3.000</td>
<td>2.613</td>
</tr>
</tbody>
</table>
Length   Class  Mode
2136   character  character

duplicates:
Length   Class  Mode
2136   character  character

Source
https://www.ebi.ac.uk/ega/studies/EGAS00000000083

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:
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:
-----------------------------

<table>
<thead>
<tr>
<th>sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99 character</td>
<td>character</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alt_sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99 character</td>
<td>character</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type:</th>
<th>tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99</td>
</tr>
</tbody>
</table>

er:
<table>
<thead>
<tr>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>57</td>
</tr>
</tbody>
</table>

pgr:
<table>
<thead>
<tr>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>43</td>
<td>1</td>
</tr>
</tbody>
</table>

her2:
<table>
<thead>
<tr>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>14</td>
</tr>
</tbody>
</table>

tumor_size:
<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.100</td>
<td>2.450</td>
<td>3.200</td>
<td>3.624</td>
<td>4.300</td>
<td>10.000</td>
</tr>
</tbody>
</table>

N:
<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>65</td>
</tr>
</tbody>
</table>

age_at_initial_pathologic_diagnosis:
<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.00</td>
<td>46.50</td>
<td>56.00</td>
<td>55.81</td>
<td>63.50</td>
<td>87.00</td>
</tr>
</tbody>
</table>

dmfs_days:
### Description

ExpressionSet for the MUG Dataset

### Format

```r
eperimentData(eset):
  Experiment data
  Experimenter name:
  Laboratory:
  Title:
  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.
```

```
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

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:
PMIDs: 18636107
No abstract available.
notes:
  summary:
    48 genes were identified that displayed highly restricted levels of expression in tumors compared to normal tissues. This was validated in 11 independent 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.
sample_type:
tumor
183

batch:
NCCS
183

uncurated_author_metadata:
Length  Class  Mode
183 character character

Source

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 associated with grade, but not associated with menopausal state, node status, or tumor size. Genes that were significantly associated with survival were 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:

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8</td>
<td>967</td>
<td>2057</td>
<td>1969</td>
<td>2930</td>
<td>4067</td>
</tr>
</tbody>
</table>

recurrence_status:

<table>
<thead>
<tr>
<th></th>
<th>norecurrence</th>
<th>recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>54</td>
<td>45</td>
</tr>
</tbody>
</table>

treatment:

<table>
<thead>
<tr>
<th></th>
<th>chemotherapy</th>
<th>hormonotherapy</th>
<th>untreated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>78</td>
<td>11</td>
</tr>
</tbody>
</table>

batch:

NCI 99

uncurated_author_metadata:

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

Source

Supplemental data from paper

Description

ExpressionSet for the NKI Dataset

Format

experimentData(eset):

Experiment data
Experimenter name:
Laboratory:
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 po
werful 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 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:**
<table>
<thead>
<tr>
<th>Variable</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>age_at_initial_pathologic_diagnosis</td>
<td>26.0</td>
<td>40.0</td>
<td>45.0</td>
<td>44.2</td>
<td>49.0</td>
<td>62.0</td>
</tr>
<tr>
<td>grade</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dmfs_days</td>
<td>9</td>
<td>1252</td>
<td>2414</td>
<td>2546</td>
<td>3602</td>
<td>6699</td>
</tr>
<tr>
<td>dmfs_status</td>
<td>norecurrence</td>
<td>recurrence</td>
<td>NA's</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>days_to_tumor_recurrence</td>
<td>9</td>
<td>1252</td>
<td>2414</td>
<td>2546</td>
<td>3602</td>
<td>6699</td>
</tr>
<tr>
<td>recurrence_status</td>
<td>norecurrence</td>
<td>recurrence</td>
<td>NA's</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>days_to_death</td>
<td>20</td>
<td>1934</td>
<td>2637</td>
<td>2870</td>
<td>3763</td>
<td>6694</td>
</tr>
<tr>
<td>vital_status</td>
<td>deceased</td>
<td>living</td>
<td>NA's</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatment</td>
<td>chemotherapy</td>
<td>hormonotherapy</td>
<td>untreated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>batch</td>
<td>NKI</td>
<td>NKI2</td>
<td>117</td>
<td>220</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

uncurated_author_metadata:
**Source**

Not available

**Description**

ExpressionSet for the PNC Dataset

**Format**

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

```r
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

<table>
<thead>
<tr>
<th>n</th>
<th>events</th>
<th>median</th>
<th>0.95LCL</th>
<th>0.95UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>88.0</td>
<td>25.0</td>
<td>NA</td>
<td>11.3</td>
<td>NA</td>
</tr>
</tbody>
</table>

---

Available sample meta-data:
---

<table>
<thead>
<tr>
<th>sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>92</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alt_sample_name:</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>92</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>tumor</td>
<td>92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>er:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td>positive</td>
</tr>
<tr>
<td>43</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pgr:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td>positive</td>
</tr>
<tr>
<td>43</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>her2:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td>positive</td>
</tr>
<tr>
<td>64</td>
<td>26</td>
</tr>
</tbody>
</table>

| 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:            | 43    | 40      | 9      |

<table>
<thead>
<tr>
<th>age_at_initial_pathologic_diagnosis:</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
</table>
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

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:
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:
no-recurrence recurrence
113 46

treatment:
chemotherapy hormonotherapy untreated
89 48 22

batch:
STK
159

uncurated_author_metadata:
Length Class Mode
159 character character

Source

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:
Distinct breast cancer subtypes were determined by gene expression profiles and were validated in other published datasets.

**Mapping method:**
- `maxRowVariance`

**Mapping group:**
- `EntrezGene.ID`

**Preprocessing:**
- As published by original author.

**Feature Data:**
- An object of class 'AnnotatedDataFrame'
- Feature names: IMAGE:1020315 IMAGE:1030271 ... IMAGE:971399 (3663 total)
- Variable labels: probeset gene EntrezGene.ID best_probe
- Variable metadata: labelDescription

**Details**

- Assay data: 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:
<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>118</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

alt_sample_name:
<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>118</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

sample_type:
tumor
118
```
<table>
<thead>
<tr>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>82</td>
<td>5</td>
</tr>
<tr>
<td>tumor_size:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>62</td>
</tr>
<tr>
<td>N:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>NA's</td>
</tr>
<tr>
<td>34</td>
<td>79</td>
<td>5</td>
</tr>
<tr>
<td>age_at_initial_pathologic_diagnosis:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
</tr>
<tr>
<td>21.00</td>
<td>46.25</td>
<td>58.00</td>
</tr>
<tr>
<td>grade:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>49</td>
<td>53</td>
</tr>
<tr>
<td>days_to_tumor_recurrence:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
</tr>
<tr>
<td>122.0</td>
<td>396.0</td>
<td>761.0</td>
</tr>
<tr>
<td>recurrence_status:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>norecurrence</td>
<td>recurrence</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>days_to_death:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
</tr>
<tr>
<td>91</td>
<td>426</td>
<td>898</td>
</tr>
<tr>
<td>vital_status:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>deceased</td>
<td>living</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>treatment:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chemotherapy</td>
<td>hormonotherapy</td>
<td>untreated</td>
</tr>
<tr>
<td>23</td>
<td>73</td>
<td>22</td>
</tr>
<tr>
<td>batch:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STNO2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uncurated_author_metadata:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>Class</td>
<td>Mode</td>
</tr>
<tr>
<td>118</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>
Source

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

---

**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)

<table>
<thead>
<tr>
<th>n</th>
<th>events</th>
<th>median</th>
<th>.95LCL</th>
<th>.95UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1073.00</td>
<td>103.00</td>
<td>10.05</td>
<td>8.56</td>
<td>12.05</td>
</tr>
</tbody>
</table>
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.
  -7.00  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/

Description

ExpressionSet for the TRANSBIG Dataset

Format

eperimentData(eset):
  Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=17545524
  Title:
  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 r
      eceive 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)

<table>
<thead>
<tr>
<th>n</th>
<th>events</th>
<th>median</th>
<th>0.95LCL</th>
<th>0.95UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>198.0</td>
<td>56.0</td>
<td>NA</td>
<td>17.1</td>
<td>NA</td>
</tr>
</tbody>
</table>

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

sample_name:
<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>198</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

ds_data:

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

UCSF

Description
  ExpressionSet for the UCSF Dataset

Format
  experimentData(eset):
    Experiment data
      Experimenter name:
        Laboratory:
A gene set was identified that correctly predicted outcomes more effectively 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

<table>
<thead>
<tr>
<th>n</th>
<th>events</th>
<th>median</th>
<th>0.95LCL</th>
<th>0.95UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>133.00</td>
<td>44.00</td>
<td>11.56</td>
<td>9.25</td>
<td>NA</td>
</tr>
</tbody>
</table>

Available sample meta-data:

**sample_name:**

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
<th>162 character</th>
</tr>
</thead>
</table>

**alt_sample_name:**

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
<th>162 character</th>
</tr>
</thead>
</table>

**sample_type:**
tumor

162
er:
<table>
<thead>
<tr>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>101</td>
<td>20</td>
</tr>
</tbody>
</table>

pgr:
<table>
<thead>
<tr>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>94</td>
<td>22</td>
</tr>
</tbody>
</table>

her2:
<table>
<thead>
<tr>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>19</td>
<td>108</td>
</tr>
</tbody>
</table>

tumor_size:

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>1.800</td>
<td>2.000</td>
<td>2.682</td>
<td>3.200</td>
<td>11.000</td>
<td>7</td>
</tr>
</tbody>
</table>

N:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>82</td>
<td>13</td>
</tr>
</tbody>
</table>

age_at_initial_pathologic_diagnosis:

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.00</td>
<td>44.00</td>
<td>53.00</td>
<td>56.61</td>
<td>70.00</td>
<td>88.00</td>
<td>9</td>
</tr>
</tbody>
</table>

grade:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>62</td>
<td>74</td>
<td>12</td>
</tr>
</tbody>
</table>

dmfs_days:

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>897</td>
<td>2040</td>
<td>2084</td>
<td>2992</td>
<td>8267</td>
<td>29</td>
</tr>
</tbody>
</table>

dmfs_status:

<table>
<thead>
<tr>
<th>norecurrence</th>
<th>recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>22</td>
</tr>
</tbody>
</table>

days_to_tumor_recurrence:

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>861</td>
<td>1865</td>
<td>1985</td>
<td>2847</td>
<td>8267</td>
<td>29</td>
</tr>
</tbody>
</table>

recurrence_status:

<table>
<thead>
<tr>
<th>norecurrence</th>
<th>recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>37</td>
</tr>
</tbody>
</table>

days_to_death:

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>1087</td>
<td>2054</td>
<td>2140</td>
<td>3087</td>
<td>8267</td>
<td>29</td>
</tr>
</tbody>
</table>
vital_status:
dead  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

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:
  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 claudi
n-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:
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1.5</th>
<th>3</th>
<th>6</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60</td>
<td>1</td>
<td>129</td>
<td>43</td>
<td>72</td>
</tr>
</tbody>
</table>

N:
<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>126</td>
<td>135</td>
<td>44</td>
</tr>
</tbody>
</table>

age_at_initial_pathologic_diagnosis:
<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24.00</td>
<td>46.00</td>
<td>55.00</td>
<td>56.73</td>
<td>68.00</td>
<td>89.00</td>
<td>59</td>
</tr>
</tbody>
</table>

grade:
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>80</td>
<td>138</td>
<td>62</td>
</tr>
</tbody>
</table>

days_to_tumor_recurrence:
<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30.0</td>
<td>450.0</td>
<td>750.0</td>
<td>954.3</td>
<td>1380.0</td>
<td>3540.0</td>
<td>64</td>
</tr>
</tbody>
</table>

recurrence_status:
<table>
<thead>
<tr>
<th></th>
<th>no recurrence</th>
<th>recurrence</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>170</td>
<td>70</td>
<td>65</td>
</tr>
</tbody>
</table>

days_to_death:
<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>540</td>
<td>885</td>
<td>1104</td>
<td>1590</td>
<td>5190</td>
<td>65</td>
</tr>
</tbody>
</table>

vital_status:
<table>
<thead>
<tr>
<th></th>
<th>deceased</th>
<th>living</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>51</td>
<td>189</td>
<td>65</td>
</tr>
</tbody>
</table>

batch:
<table>
<thead>
<tr>
<th></th>
<th>UNC4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>305</td>
</tr>
</tbody>
</table>

uncurated_author_metadata:
<table>
<thead>
<tr>
<th></th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>305</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

duplicates:
<table>
<thead>
<tr>
<th></th>
<th>KOO.KOO_KF_105</th>
<th>KOO.KOO_T01_514</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>303</td>
</tr>
</tbody>
</table>

Source

Description

ExpressionSet for the UNT Dataset

Format

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

batch:
KIU OXFU

uncurated_author_metadata:
Length  Class  Mode
133    character  character

Source


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:
PMIDs: 16141321
No abstract available.
notes:
summary:
A 32-gene expression signature of p53 was identified that differentiates p
53 mutant and wild-type tumors. The signature is more effective than sequ
ence-based assessments of p53 in predicting prognosis and therapeutic resp
onse.
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
<table>
<thead>
<tr>
<th>Grade</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>67</td>
<td>128</td>
<td>54</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Days to Tumor Recurrence</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>1870</td>
<td>3711</td>
<td>3007</td>
<td>3985</td>
<td>4654</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recurrence Status</th>
<th>Norecurrence</th>
<th>Recurrence</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>181</td>
<td>55</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Hormonotherapy</th>
<th>Untreated</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80</td>
<td>142</td>
<td>29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Batch</th>
<th>UPPU</th>
<th>UPPT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80</td>
<td>171</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Uncurated Author Metadata</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>251</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

**Source**


**Description**

ExpressionSet for the VDX Dataset

**Format**

experimentData(eset):

Experiment data
- Experimenter name:
- Laboratory:
- Title:
- 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:
<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>344 character</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

alt_sample_name:
<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>122.8</td>
<td>605.5</td>
<td>575.7</td>
<td>836.5</td>
<td>2038.0</td>
</tr>
</tbody>
</table>

sample_type:
tumor
| 344 |

er:
negative positive
| 135 | 209 |

tumor_size:
| 1 2 3 4 NA's |
| 146 132 5 3 58 |

N:
age_at_initial_pathologic_diagnosis:

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>26.00</td>
<td>44.00</td>
<td>52.00</td>
<td>53.88</td>
<td>63.00</td>
<td>83.00</td>
<td>58</td>
</tr>
</tbody>
</table>

grade:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>42</td>
<td>148</td>
<td>147</td>
</tr>
</tbody>
</table>

dmfs_days:

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>1254</td>
<td>2616</td>
<td>2377</td>
<td>3285</td>
<td>5201</td>
</tr>
</tbody>
</table>

dmfs_status:

norecurrence recurrence

226 118

treatment:

untreated

344

batch:

VDX VDXN

286 58

uncurated_author_metadata:

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>344</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

Source

# Index

* datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAL</td>
<td>3</td>
</tr>
<tr>
<td>DFHCC</td>
<td>5</td>
</tr>
<tr>
<td>DFHCC2</td>
<td>8</td>
</tr>
<tr>
<td>DFHCC3</td>
<td>10</td>
</tr>
<tr>
<td>DUKE</td>
<td>11</td>
</tr>
<tr>
<td>DUKE2</td>
<td>13</td>
</tr>
<tr>
<td>duplicates</td>
<td>15</td>
</tr>
<tr>
<td>EMC2</td>
<td>16</td>
</tr>
<tr>
<td>EORTC10994</td>
<td>18</td>
</tr>
<tr>
<td>EXPO</td>
<td>20</td>
</tr>
<tr>
<td>FNCLCC</td>
<td>22</td>
</tr>
<tr>
<td>GSE25066</td>
<td>23</td>
</tr>
<tr>
<td>GSE32646</td>
<td>26</td>
</tr>
<tr>
<td>GSE48091</td>
<td>28</td>
</tr>
<tr>
<td>GSE58644</td>
<td>29</td>
</tr>
<tr>
<td>HLP</td>
<td>32</td>
</tr>
<tr>
<td>IRB</td>
<td>34</td>
</tr>
<tr>
<td>KOO</td>
<td>36</td>
</tr>
<tr>
<td>LUND</td>
<td>40</td>
</tr>
<tr>
<td>LUND2</td>
<td>42</td>
</tr>
<tr>
<td>MAINZ</td>
<td>44</td>
</tr>
<tr>
<td>MAQC2</td>
<td>46</td>
</tr>
<tr>
<td>MCCC</td>
<td>48</td>
</tr>
<tr>
<td>MDA4</td>
<td>49</td>
</tr>
<tr>
<td>METABRIC</td>
<td>51</td>
</tr>
<tr>
<td>MSK</td>
<td>54</td>
</tr>
<tr>
<td>MUG</td>
<td>56</td>
</tr>
<tr>
<td>NCCS</td>
<td>57</td>
</tr>
<tr>
<td>NCI</td>
<td>59</td>
</tr>
<tr>
<td>NKI</td>
<td>61</td>
</tr>
<tr>
<td>PNC</td>
<td>64</td>
</tr>
<tr>
<td>STK</td>
<td>66</td>
</tr>
<tr>
<td>STNO2</td>
<td>68</td>
</tr>
<tr>
<td>TCGA</td>
<td>71</td>
</tr>
<tr>
<td>TRANSBIG</td>
<td>73</td>
</tr>
<tr>
<td>UCSF</td>
<td>75</td>
</tr>
<tr>
<td>UNC4</td>
<td>78</td>
</tr>
<tr>
<td>UNT</td>
<td>81</td>
</tr>
</tbody>
</table>

UPP, 83
VDX, 85

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAL (CAL)</td>
<td>3</td>
</tr>
<tr>
<td>DFHCC (DFHCC)</td>
<td>5</td>
</tr>
<tr>
<td>DFHCC2 (DFHCC2)</td>
<td>8</td>
</tr>
<tr>
<td>DFHCC3 (DFHCC3)</td>
<td>10</td>
</tr>
<tr>
<td>DUKE (DUKE)</td>
<td>11</td>
</tr>
<tr>
<td>DUKE2 (DUKE2)</td>
<td>13</td>
</tr>
<tr>
<td>EMC2 (EMC2)</td>
<td>16</td>
</tr>
<tr>
<td>EORTC10994 (EORTC10994)</td>
<td>18</td>
</tr>
<tr>
<td>EXPO (EXPO)</td>
<td>20</td>
</tr>
<tr>
<td>FNCLCC (FNCLCC)</td>
<td>22</td>
</tr>
<tr>
<td>GSE25066 (GSE25066)</td>
<td>23</td>
</tr>
<tr>
<td>GSE32646 (GSE32646)</td>
<td>26</td>
</tr>
<tr>
<td>GSE48091 (GSE48091)</td>
<td>28</td>
</tr>
<tr>
<td>GSE58644 (GSE58644)</td>
<td>29</td>
</tr>
<tr>
<td>HLP (HLP)</td>
<td>32</td>
</tr>
<tr>
<td>IRB (IRB)</td>
<td>34</td>
</tr>
<tr>
<td>KOO (KOO)</td>
<td>36</td>
</tr>
<tr>
<td>LUND (LUND)</td>
<td>40</td>
</tr>
<tr>
<td>LUND2 (LUND2)</td>
<td>42</td>
</tr>
<tr>
<td>MAINZ (MAINZ)</td>
<td>44</td>
</tr>
<tr>
<td>MAQC2 (MAQC2)</td>
<td>46</td>
</tr>
<tr>
<td>MCCC (MCCC)</td>
<td>48</td>
</tr>
<tr>
<td>MDA4 (MDA4)</td>
<td>49</td>
</tr>
<tr>
<td>METABRIC (METABRIC)</td>
<td>51</td>
</tr>
<tr>
<td>MSK (MSK)</td>
<td>54</td>
</tr>
<tr>
<td>MUG (MUG)</td>
<td>56</td>
</tr>
<tr>
<td>NCCS (NCCS)</td>
<td>57</td>
</tr>
<tr>
<td>NCI (NCI)</td>
<td>59</td>
</tr>
<tr>
<td>NKI (NKI)</td>
<td>61</td>
</tr>
<tr>
<td>PNC (PNC)</td>
<td>64</td>
</tr>
<tr>
<td>STK (STK)</td>
<td>66</td>
</tr>
<tr>
<td>STNO2 (STNO2)</td>
<td>68</td>
</tr>
<tr>
<td>TCGA (TCGA)</td>
<td>71</td>
</tr>
<tr>
<td>TRANSBIG (TRANSBIG)</td>
<td>73</td>
</tr>
<tr>
<td>UCSF (UCSF)</td>
<td>75</td>
</tr>
<tr>
<td>UNC4 (UNC4)</td>
<td>78</td>
</tr>
<tr>
<td>UNT (UNT)</td>
<td>81</td>
</tr>
<tr>
<td>INDEX</td>
<td>Page</td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
</tr>
<tr>
<td>UPP (UPP), 83</td>
<td></td>
</tr>
<tr>
<td>VDX (VDX), 85</td>
<td></td>
</tr>
<tr>
<td>CAL, 3</td>
<td></td>
</tr>
<tr>
<td>DFHCC, 5</td>
<td></td>
</tr>
<tr>
<td>DFHCC2, 8</td>
<td></td>
</tr>
<tr>
<td>DFHCC3, 10</td>
<td></td>
</tr>
<tr>
<td>DUKE, 11</td>
<td></td>
</tr>
<tr>
<td>DUKE2, 13</td>
<td></td>
</tr>
<tr>
<td>dupplicates, 15</td>
<td></td>
</tr>
<tr>
<td>EMC2, 16</td>
<td></td>
</tr>
<tr>
<td>EORTC10994, 18</td>
<td></td>
</tr>
<tr>
<td>EXPO, 20</td>
<td></td>
</tr>
<tr>
<td>FNCLCC, 22</td>
<td></td>
</tr>
<tr>
<td>GSE25066, 23</td>
<td></td>
</tr>
<tr>
<td>GSE32646, 26</td>
<td></td>
</tr>
<tr>
<td>GSE48091, 28</td>
<td></td>
</tr>
<tr>
<td>GSE58644, 29</td>
<td></td>
</tr>
<tr>
<td>HLP, 32</td>
<td></td>
</tr>
<tr>
<td>IRB, 34</td>
<td></td>
</tr>
<tr>
<td>KOO, 36</td>
<td></td>
</tr>
<tr>
<td>loadBreastDatasets, 38</td>
<td></td>
</tr>
<tr>
<td>loadBreastEsets, 39</td>
<td></td>
</tr>
<tr>
<td>LUND, 40</td>
<td></td>
</tr>
<tr>
<td>LUND2, 42</td>
<td></td>
</tr>
<tr>
<td>MAINZ, 44</td>
<td></td>
</tr>
<tr>
<td>MAQC2, 46</td>
<td></td>
</tr>
<tr>
<td>MCCC, 48</td>
<td></td>
</tr>
<tr>
<td>MDA4, 49</td>
<td></td>
</tr>
<tr>
<td>METABRIC, 51</td>
<td></td>
</tr>
<tr>
<td>MSK, 54</td>
<td></td>
</tr>
<tr>
<td>MUG, 56</td>
<td></td>
</tr>
<tr>
<td>NCCS, 57</td>
<td></td>
</tr>
<tr>
<td>NCI, 59</td>
<td></td>
</tr>
<tr>
<td>NKI, 61</td>
<td></td>
</tr>
<tr>
<td>PNC, 64</td>
<td></td>
</tr>
<tr>
<td>STK, 66</td>
<td></td>
</tr>
<tr>
<td>STNO2, 68</td>
<td></td>
</tr>
<tr>
<td>TCGA, 71</td>
<td></td>
</tr>
<tr>
<td>TRANSBIG, 73</td>
<td></td>
</tr>
<tr>
<td>UCSF, 75</td>
<td></td>
</tr>
<tr>
<td>UNCC, 78</td>
<td></td>
</tr>
<tr>
<td>UNT, 81</td>
<td></td>
</tr>
<tr>
<td>UPP, 83</td>
<td></td>
</tr>
<tr>
<td>VDX, 85</td>
<td></td>
</tr>
</tbody>
</table>