Package ‘MetaGxBreast’

July 4, 2024

Type Package

Title Transcriptomic Breast Cancer Datasets

Version 1.24.0

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_19

git_last_commit ba1bd33

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-07-04

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>
## Contents

<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>duplciates</td>
<td>15</td>
</tr>
<tr>
<td>EMC2</td>
<td>16</td>
</tr>
<tr>
<td>EORTC10994</td>
<td>18</td>
</tr>
<tr>
<td>EXP0</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/
  PMIDs: 17157792
  No abstract available.
  notes:
    summary:
      Recurrent copy number abnormalities differ between tumor subtypes as defined by gene expression patterns. Accuracy of stratification by outcome can be improved by combining expression and copy number.
      mapping.method:
        maxRowVariance
      mapping.group:
        EntrezGene.ID
    preprocessing:
      As published by original author.

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

Details

assayData: 21169 features, 118 samples
Platform type:
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

  1 observation deleted due to missingness
  n  events  median  0.95LCL  0.95UCL
<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>age_at_initial_pathologic_diagnosis</td>
<td>31.00</td>
<td>44.00</td>
<td>51.00</td>
<td>55.06</td>
<td>66.00</td>
<td>88.00</td>
<td>1</td>
</tr>
<tr>
<td>grade</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>NA's</td>
<td>10</td>
<td>42</td>
<td>61</td>
</tr>
<tr>
<td>dmfs_days</td>
<td>0</td>
<td>767</td>
<td>2059</td>
<td>2094</td>
<td>3336</td>
<td>5183</td>
<td>1</td>
</tr>
<tr>
<td>dmfs_status</td>
<td>norecurrence</td>
<td>recurrence</td>
<td>NA's</td>
<td>91</td>
<td>26</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>days_to_tumor_recurrence</td>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
<td>3rd Qu.</td>
<td>Max.</td>
<td>NA's</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sample_name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sample_type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>er</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pgr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### DFHCC

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>767</th>
<th>2059</th>
<th>2094</th>
<th>3336</th>
<th>5183</th>
<th>1</th>
</tr>
</thead>
</table>

- **recurrence_status:**
  - norecurrence
  - recurrence
  - NA's

- **days_to_death:**
  - Min: 47
  - 1st Qu: 1117
  - Median: 2234
  - Mean: 2347
  - 3rd Qu: 3504
  - Max: 5183

- **vital_status:**
  - deceased
  - living

- **treatment:**
  - chemo.plus.hormono
  - chemotherapy
  - hormonotherapy
  - untreated

- **batch:**
  - CAL

- **uncurated_author_metadata:**
  - Length
  - Class
  - Mode

<table>
<thead>
<tr>
<th></th>
<th>118</th>
</tr>
</thead>
</table>

**Source**

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

---

**DFHCC**

**Description**

ExpressionSet for the DFHCC Dataset

**Format**

- `experimentData(eset)`
  - Experiment data
    - Experimenter name:
    - Laboratory:
    - Contact information: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2826790/
Title:
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)
  varLabels: probeset gene EntrezGene.ID best_probe
  varMetadata: labelDescription

**Details**

assayData: 42447 features, 40 samples
Platform type:
Available sample meta-data:

sample_name:
Length   Class   Mode
 40 character character

alt_sample_name:
Length   Class   Mode
 40 character character

class:
tumor
 40

batch:
DFHCC3
 40

uncurated_author_metadata:
Length   Class   Mode
 40 character character

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

tumor:
sample_type:
171
er:
negative positive
<table>
<thead>
<tr>
<th>pgr:</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative positive NA's</td>
</tr>
<tr>
<td>23 65 83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tumor_size:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. 1st Qu. Median 3rd Qu. Max. NA's</td>
</tr>
<tr>
<td>0.20 1.80 2.30 2.74 3.50 8.50 83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 NA's</td>
</tr>
<tr>
<td>53 36 82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>days_to_death:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. 1st Qu. Median 3rd Qu. Max. NA's</td>
</tr>
<tr>
<td>171.0 417.0 957.5 1235.0 1852.0 4069.0 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>vital_status:</th>
</tr>
</thead>
<tbody>
<tr>
<td>deceased living NA's</td>
</tr>
<tr>
<td>43 127 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUKE</td>
</tr>
<tr>
<td>171</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>uncurated_author_metadata:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length Class Mode</td>
</tr>
<tr>
<td>171 character character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>duplicates:</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUKE.DUKE_T00.622 DUKE.DUKE_T01.052 DUKE.DUKE_T01.522 DUKE.DUKE_T01.534</td>
</tr>
<tr>
<td>1 1 1 1</td>
</tr>
<tr>
<td>NA's</td>
</tr>
<tr>
<td>167</td>
</tr>
</tbody>
</table>

**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
sample_type: 
tumor
  160

er:
negative positive
  123  37

pgr:
negative positive  NA's
  133  25  2

N:
  0  1  NA's
  58  95  7

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

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

treatment:
chemotherapy
  160

batch:
DUKE2
  160

uncurated_author_metadata:
  Length  Class  Mode
  160  character  character

Source

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

experimentData(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.

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

assayData: 42447 features, 204 samples

Platform type:

Available sample meta-data:

sample_name:
<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>204</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>1.00</td>
<td>51.75</td>
<td>102.50</td>
<td>102.50</td>
<td>153.20</td>
<td>204.00</td>
</tr>
</tbody>
</table>

sample_type:
<table>
<thead>
<tr>
<th>tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>204</td>
</tr>
</tbody>
</table>

N:
0 NA's
48    156

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>0</td>
<td>335</td>
<td>640</td>
<td>799</td>
<td>1098</td>
<td>3507</td>
</tr>
</tbody>
</table>

dmfs_status:
norecurrence recurrence
| 19   | 185    |

treatment:
| chemotherapy | untreated |
| 156          | 48        |

batch:
<table>
<thead>
<tr>
<th>EMC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>204</td>
</tr>
</tbody>
</table>

uncurated_author_metadata:
<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>204</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

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: labelDescription

Details

assayData: 20967 features, 49 samples
Platform type:
-----------------------------
Available sample meta-data:
sample_name:
  Length  Class  Mode
  49  character  character

alt_sample_name:
  Length  Class  Mode
  49  character  character

sample_type:
tumor
  49

er:
negative  positive
  22  27

pgr:
negative  positive  NA's
  29  18  2

tumor_size:
  1  2  3  4
  4  23  14  8

N:
  0  1
  19  30

grade:
  1  2  3  NA's
  4  22  20  3

batch:
  EORTC10994
  49

uncurated_author_metadata:
  Length  Class  Mode
  49  character  character

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

tumor

353

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

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

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

age_at_initial_pathologic_diagnosis:
<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>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>EXPO.EXPO_GSM53027</th>
<th>EXPO.EXPO_GSM53059</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</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:
  PMIDs: 17659439
  No abstract available.
  notes:
    summary:
      A potentially more powerful clinicogenomic model was created by combining
      a subset of relevant genes from an already published gene expression signature
      and a commonly used clinical prognostic model (NPI). The genes in this model
      are known to have a role in breast cancer, carcinogenesis, or chemotherapy
      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:
  --------------------------------------
GSE25066

sample_name:
   Length    Class    Mode
   150 character character

alt_sample_name:
   Length    Class    Mode
   150 character character

sample_type:
tumor
   150

N:
   1
   150

treatment:
chemotherapy
   150

batch:
FNCLCC
   150

uncurated_author_metadata:
   Length    Class    Mode
   150 character character

Source

GSE25066     GSE25066

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
### her2:
- Negative: 485
- Positive: 6
- NA's: 17

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

### N:
- 0: 157
- 1: 351

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

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

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

### dmfs_status:
- No recurrence: 397
- Recurrence: 111

### batch:
- GSE25066: 508

### uncurated_author_metadata:
- Length: 508
- Class: character
- Mode: character

### chemosensitivity_prediction:
- Rx Insensitive: 339
- Rx Sensitive: 169

### GGI_prediction:
- High: 336
- Low: 172

### PAM50_prediction:
- Basal: 189
- Her2: 37
- LumA: 160
- LumB: 78
- Normal: 44
dlda30_prediction:
pCR  RD
196  312

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

Source

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:
---------------------------------------------
sample_name:
    Length  Class  Mode
    115 character character

sample_type:
tumor
    115

er:
negative positive
    44   71

pgr:
negative positive
    70   45

her2:
negative positive
    81   34

T:
    1  2  3  4
    5  87 18  5

N:
    0  1
    32 83

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

grade:
    1  2  3
    16 78 21

batch:
GSE32646
    115

uncurated_author_metadata:
Length Class Mode
115 character character

Source

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

GSE58644        GSE58644

Description

ExpressionSet for the GSE58644 Dataset

Format

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

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

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

Details

assayData: 21462 features, 321 samples
Platform type:
---------------------------
Available sample meta-data:
---------------------------

sample_name:
  Length  Class  Mode
    321 character  character

alt_sample_name:
  Length  Class  Mode
    321 character  character

sample_type:
tumor
    321

er:
negative  positive   NA's
       70      250        1

her2:
negative  positive   NA's
       256       58         7

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

T:
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</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></td>
<td>29.00</td>
<td>49.00</td>
<td>58.00</td>
<td>58.82</td>
<td>68.00</td>
<td>93.00</td>
</tr>
<tr>
<td>grade</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>NA's</td>
<td>26</td>
<td>135</td>
</tr>
<tr>
<td>dmfs_status</td>
<td></td>
<td>norecurrence</td>
<td>recurrence</td>
<td>295</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dmfs_days</td>
<td></td>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>9496</td>
<td>17900</td>
<td>21620</td>
<td>33600</td>
<td>52590</td>
</tr>
<tr>
<td>treatment</td>
<td></td>
<td>chemo.plus.hormono</td>
<td>chemotherapy</td>
<td>hormonotherapy</td>
<td>untreated</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>91</td>
<td>29</td>
<td>66</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NA's</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chemo</td>
<td></td>
<td>0</td>
<td>1</td>
<td>NA's</td>
<td>105</td>
<td>123</td>
<td>93</td>
</tr>
<tr>
<td>tamoxifen</td>
<td></td>
<td>0</td>
<td>1</td>
<td>NA's</td>
<td>39</td>
<td>157</td>
<td>125</td>
</tr>
<tr>
<td>herceptin</td>
<td></td>
<td>0</td>
<td>1</td>
<td>NA's</td>
<td>190</td>
<td>12</td>
<td>119</td>
</tr>
<tr>
<td>batch</td>
<td>GSE58644</td>
<td>321</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uncurated_author_metadata</td>
<td></td>
<td>Length</td>
<td>Class</td>
<td>Mode</td>
<td>321</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>
duplicates:
Length    Class    Mode
321 character character

Source

HLP

Description
ExpressionSet for the HLP Dataset

Format
experimentData(eset):
Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=19688261
  Title:
  URL: http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-543/
  PMIDs: 19688261
  No abstract available.
  notes:
      summary:
          The results show evidence of different patterns of genetic aberrations in
distinct molecular subtypes of breast cancer. Patterns of copy number aber
rations 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>
</tr>
</thead>
<tbody>
<tr>
<td>tumor</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>

| grade: | 3 |
|        | 53 |

<table>
<thead>
<tr>
<th>batch:</th>
</tr>
</thead>
</table>
IRB
53

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:
<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>129</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>129</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

sample_type:
tumor

er:
negative positive
| 53  | 76 |

her2:
negative positive
| 98  | 31 |

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>0.800</td>
<td>1.500</td>
<td>2.200</td>
<td>2.488</td>
<td>3.000</td>
<td>8.500</td>
</tr>
</tbody>
</table>

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
n the study.
mapping.method:
maxRowVariance
mapping.group:
EnterzGene.ID
preprocessing:
As published by original author.

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

assayData: 280 features, 88 samples
Platform type:

Available sample meta-data:

sample_name:
  Length  Class  Mode
  88 character  character

alt_sample_name:
  Length  Class  Mode
  88 character  character

sample_type:
tumor
  88

er:
negative  positive
  15   73

pgr:
negative  positive
  23   65

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

N:
  0  1
  19  69

treatment:
chemotherapy  untreated
  61   27

batch:
KOO
  88

uncurated_author_metadata:
  Length  Class  Mode
  88 character  character

duplicates:
loadBreastDatasets

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 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 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)

Source

Unavailable

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
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).

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 tcga datasets are loaded separately as they are very large and doing so will help prevent memory allocation errors for R windows. Furthermore, these datasets are so large that they dominate statistical analyses so it is best that they are analyzed separate of the 37 smaller datasets loaded with the string majority.
- `removeDuplicates` remove patients with a Spearman correlation greater than or equal to 0.98 with other patient expression profiles (default TRUE).
- `quantileCutoff` A numeric between 0 and 1 specifying to remove genes with standard deviation below the required quantile (default 0).
- `rescale` apply centering and scaling to the expression sets (default FALSE).
minNumberGenes an integer specifying to remove expression sets with less genes than this number (default 0)

minNumberEvents an integer specifying how many survival events must be in the dataset to keep the dataset (default 0)

minSampleSize an integer specifying the minimum number of patients required in an eset (default 0)

removeRetracted remove datasets from retracted papers (default TRUE, currently just PMID17290060 dataset)

removeSubsets remove datasets that are a subset of other datasets (default TRUE, currently just PMID19318476)

keepCommonOnly remove probes not common to all datasets (default FALSE)

imputeMissing remove patients from datasets with missing expression values

Value

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

Examples

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

Description

ExpressionSet for the LUND Dataset

Format

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

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

Details

assayData: 11154 features, 143 samples
Platform type:

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

sample_name:
Length Class Mode
143 character character

alt_sample_name:
Length Class Mode
143 character character

tumor:
143

er:
negative positive
29 114

pgr:
negative positive NA's
47 88 8

tumor_size:
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
0.200 1.100 1.500 1.486 1.800 4.000 2
N:
  0
  143

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

batch:
  LUNDS1 LUNDS2 LUNDS3 LUNDS4
  30   47   22   44

uncurated_author_metadata:
  Length  Class  Mode
  143  character  character

Source

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

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

<table>
<thead>
<tr>
<th>sample_type:</th>
<th>tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>er:</th>
<th>negative</th>
<th>positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>38</td>
<td>162</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>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.100</td>
<td>1.500</td>
<td>2.000</td>
<td>2.070</td>
<td>2.425</td>
<td>6.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>200</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>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25.00</td>
<td>50.00</td>
<td>60.00</td>
<td>59.98</td>
<td>69.00</td>
<td>90.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>grade:</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29</td>
<td>136</td>
<td>35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dmfs_days:</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>30</td>
<td>1905</td>
<td>2715</td>
<td>2816</td>
<td>3855</td>
<td>7200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dmfs_status:</th>
<th>norecurrence</th>
<th>recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>154</td>
<td>46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>treatment:</th>
<th>untreated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
<th></th>
</tr>
</thead>
</table>
MAINZ
200

uncurated_author_metadata:
Length Class Mode
200 character character

Source

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:
EnterzGene.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
Details

assayData: 20967 features, 230 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>230</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>230</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

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:

<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>26.00</td>
<td>45.00</td>
<td>51.00</td>
<td>52.02</td>
<td>59.00</td>
<td>79.00</td>
</tr>
</tbody>
</table>

grade:

| 1 | 2 | 3 |
| 13| 94| 123|

treatment:

chemotherapy
230

batch:
MAQC2
230

uncurated_author_metadata:
Length Class Mode
230 character character

Source

---

MCCC
MCCC

Description

ExpressionSet for the MCCC Dataset

Format

experimentData(eset):
Experiment data
  Experimenter name: 
  Laboratory: 
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=19960244
  Title: 
  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:

<table>
<thead>
<tr>
<th>sample_name</th>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>75</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>tumor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MCCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
<td></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>75</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

Source


Description

ExpressionSet for the MDA4 Dataset

Format

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

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

Details
assayData: 21169 features, 129 samples
Platform type:
-----------------------------------
Available sample meta-data:
-----------------------------------
sample_name:
  Length Class Mode
  129 character character

unique_patient_ID:
  Length Class Mode
  129 character character

sample_type:
tumor
  129

er:
negative positive NA's
  48   79   2
### METABRIC

**pgr:**
- negative: 73
- positive: 54
- NA's: 2

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

**tumor_size:**

<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>Value</td>
<td>0.000</td>
<td>0.500</td>
<td>1.800</td>
<td>2.162</td>
<td>3.000</td>
<td>10.000</td>
<td>8</td>
</tr>
</tbody>
</table>

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

**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>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>28.00</td>
<td>43.00</td>
<td>51.00</td>
<td>51.43</td>
<td>61.00</td>
<td>73.00</td>
</tr>
</tbody>
</table>

**treatment:**
- chemotherapy: 129

**batch:**
- MDA4: 129

**uncurated_author_metadata:**

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>129</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

**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/EGAS000000000083
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>
</tr>
</thead>
<tbody>
<tr>
<td>er:</td>
<td>negative</td>
<td>positive</td>
</tr>
<tr>
<td></td>
<td>440</td>
<td>1508</td>
</tr>
<tr>
<td>her2:</td>
<td>negative</td>
<td>positive</td>
</tr>
<tr>
<td></td>
<td>676</td>
<td>148</td>
</tr>
<tr>
<td>tumor_size:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min. 1st Qu. Median</td>
<td>0.000</td>
<td>1.700</td>
</tr>
<tr>
<td>Mean 3rd Qu. Max. NA's</td>
<td>2.621</td>
<td>3.000</td>
</tr>
<tr>
<td>N:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1042</td>
<td>950</td>
</tr>
<tr>
<td>age_at_initial_pathologic_diagnosis:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min. 1st Qu. Median</td>
<td>21.93</td>
<td>51.36</td>
</tr>
<tr>
<td>Mean 3rd Qu. Max. NA's</td>
<td>61.13</td>
<td>70.76</td>
</tr>
<tr>
<td>grade:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>170</td>
<td>775</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>234</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>days_to_death:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min. 1st Qu. Median</td>
<td>3</td>
<td>1498</td>
</tr>
<tr>
<td>Mean 3rd Qu. Max. NA's</td>
<td>2948</td>
<td>4357</td>
</tr>
<tr>
<td>vital_status:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>deceased</td>
<td>891</td>
<td></td>
</tr>
<tr>
<td>living</td>
<td>1081</td>
<td></td>
</tr>
<tr>
<td>NA's</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>treatment:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chemo.plus.hormono</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>chemotherapy</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>hormonotherapy</td>
<td></td>
<td>1029</td>
</tr>
<tr>
<td>untreated</td>
<td></td>
<td>685</td>
</tr>
<tr>
<td>batch:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min. 1st Qu. Median</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Mean 3rd Qu. Max. NA's</td>
<td>2.613</td>
<td>3.000</td>
</tr>
</tbody>
</table>

uncurated_author_metadata:
Length  Class  Mode
2136 character character

duplicates:
Length  Class  Mode
2136 character character

Source
https://www.ebi.ac.uk/ega/studies/EGAS0000000083

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

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

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

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

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

| er:                                      |
| negative positive                       |
| 42 57                                   |

| pgr:                                     |
| negative positive NA's                  |
| 55 43 1                                 |

| her2:                                    |
| positive NA's                            |
| 85 14                                   |

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

| N:                                       |
| 0 1                                     |
| 34 65                                   |

| age_at_initial_pathologic_diagnosis:    |
| Min. 1st Qu. Median 3rd Qu. Max.        |
| 30.00 46.50 55.00 55.81 63.50 87.00     |

<table>
<thead>
<tr>
<th>dmfs_days:</th>
</tr>
</thead>
</table>

<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>245</td>
<td>1279</td>
<td>1971</td>
<td>1888</td>
<td>2575</td>
<td>3924</td>
<td>17</td>
</tr>
</tbody>
</table>

dmfs_status:
- norecurrence
- recurrence
- NA's

| batch: | MSK | 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

Description
ExpressionSet for the MUG Dataset

Format
experimentData(eset):
- Experiment data
  - Experimenter name:
  - Laboratory:
  - Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=18592372
  - Title:
  - PMIDs: 18592372
- No abstract available.
- notes:
  - summary:
    - A method was developed to separate tumor cells and their microenvironment to test the prognostic abilities of the immune system. Results showed that lymphatic infiltration is beneficial for ER negative patients, but probably not beneficial for ER positive patients.
  - mapping.method:
    - maxRowVariance
  - mapping.group:
EntrezGene.ID
preprocessing:
As published by original author.

featureData(eset):
An object of class 'AnnotatedDataFrame'
featureNames: H200000001 H200000005 ... opHsV04TC000043 (14288 total)
varLabels: probeset gene EntrezGene.ID best_probe
varMetadata: labelDescription

Details

assayData: 14288 features, 152 samples
Platform type:
-------------------------------------------
Available sample meta-data:
-------------------------------------------
sample_name:
  Length  Class  Mode
  152 character character

alt_sample_name:
  Length  Class  Mode
  152 character character

sample_type:
tumor
  152

batch:
MUG
  152

uncurated_author_metadata:
  Length  Class  Mode
  152 character character

Source

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.
| 1.0 | 46.5 | 92.0 | 92.0 | 137.5 | 183.0 |

Sample type: tumor

Batch: NCCS

Uncurated author metadata:

<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>183</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

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

recurrence_status:
  norecurrence recurrence
  54       45

treatment:
  chemotherapy hormonotherapy untreated
  10       78       11

batch:
  NCI
  99

uncurated_author_metadata:
  Length Class Mode
  99 character character

Source
  Supplemental data from paper

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

<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>295</td>
<td>79</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Available sample meta-data:

<table>
<thead>
<tr>
<th>sample_name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
</tr>
<tr>
<td>337</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alt_sample_name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
</tr>
<tr>
<td>4.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type:</th>
</tr>
</thead>
</table>
tumor
337

<table>
<thead>
<tr>
<th>er:</th>
</tr>
</thead>
</table>
negative | positive |
| 88  | 249    |

<p>| tumor_size: |</p>
<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>Min. 1st Qu. Median Mean 3rd Qu. Max.</td>
<td>0.200</td>
<td>1.500</td>
<td>2.000</td>
<td>2.241</td>
<td>2.800</td>
<td>5.500</td>
</tr>
<tr>
<td>N:</td>
<td>0 1</td>
<td>193 144</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age_at_initial_pathologic_diagnosis:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min. 1st Qu. Median Mean 3rd Qu. Max.</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 2 3</td>
<td>79 109 149</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dmfs_days:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min. 1st Qu. Median Mean 3rd Qu. Max.</td>
<td>9 1252</td>
<td>2414</td>
<td>2546</td>
<td>3602</td>
<td>6699</td>
<td>18</td>
</tr>
<tr>
<td>dmfs_status:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no recurrence recurrence NA's</td>
<td>210 109</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>days_to_tumor_recurrence:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min. 1st Qu. Median Mean 3rd Qu. Max.</td>
<td>9 1252</td>
<td>2414</td>
<td>2546</td>
<td>3602</td>
<td>6699</td>
<td>18</td>
</tr>
<tr>
<td>recurrence_status:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no recurrence recurrence NA's</td>
<td>210 109</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>days_to_death:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min. 1st Qu. Median Mean 3rd Qu. Max.</td>
<td>20 1934</td>
<td>2637</td>
<td>2870</td>
<td>3763</td>
<td>6694</td>
<td>42</td>
</tr>
<tr>
<td>vital_status:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deceased living NA's</td>
<td>79</td>
<td>216</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatment:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chemotherapy hormonotherapy untreated</td>
<td>90</td>
<td>40</td>
<td>207</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>batch:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NKI NKI2</td>
<td>117</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uncurated_author_metadata:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Length Class Mode
337 character character

Source
Not available

Description
ExpressionSet for the PNC Dataset

Format
experimentData(eset):
Experiment data
Experimenter name:
Laboratory:
Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=21910250
Title:
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.

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

class: sample_name:
Length   Class   Mode
92   character   character

class: alt_sample_name:
Length   Class   Mode
92   character   character

class: sample_type:
tumor
92

class: er:
negative  positive  NA's
43   45   4

class: pgr:
negative  positive  NA's
43   40   9

class: her2:
negative  positive  NA's
64   26   2

class: 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

class: N:
0   1  NA's
43   40   9

class: age_at_initial_pathologic_diagnosis:
Min.  1st Qu.  Median  Mean  3rd Qu.  Max.  NA's
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

days_to_death:

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

vital_status:

dead living NA's
25 63 4

batch:
PNC
92

uncurated_author_metadata:

Length Class Mode
92 character character

Source


STK

STK

Description

ExpressionSet for the STK Dataset

Format

experimentData(eset):
Experiment data
Experimenter name:
Laboratory:
Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=16280042
Title:
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:
<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>28</td>
<td>58</td>
<td>61</td>
<td>12</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>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>66</td>
<td>2022</td>
<td>2467</td>
<td>2234</td>
<td>2846</td>
<td>3099</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>recurrence_status:</th>
<th>no recurrence</th>
<th>recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>113</td>
<td>46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>treatment:</th>
<th>chemotherapy</th>
<th>hormonotherapy</th>
<th>untreated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>89</td>
<td>48</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
<th>STK</th>
<th>159</th>
</tr>
</thead>
</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>159</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

**Source**


**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.

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

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

**Details**

- assayData: 3663 features, 118 samples
- Platform type:
- Overall survival time-to-event summary (in years):
  
  ```r
  Call: survfit(formula = Surv(time, cens) ~ -1)
  
  n  events  median 0.95LCL 0.95UCL
  118.00  46.00  4.67   3.34 NA
  ---------------------------
  ```

**Available sample meta-data:**

- sample_name:
  - Length  Class  Mode
  - 118 character character

- alt_sample_name:
  - Length  Class  Mode
  - 118 character character

- sample_type:
  - tumor
  - 118
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tumor_size:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>13</td>
<td>62</td>
<td>32</td>
<td>5</td>
<td></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>34</td>
<td>79</td>
<td>5</td>
<td></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>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21.00</td>
<td>46.25</td>
<td>58.00</td>
<td>58.47</td>
<td>71.75</td>
<td>85.00</td>
</tr>
</tbody>
</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>11</td>
<td>49</td>
<td>53</td>
<td>5</td>
<td></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>122.0</td>
<td>396.0</td>
<td>761.0</td>
<td>927.9</td>
<td>1233.0</td>
<td>2800.0</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>recurrence_status:</th>
<th>no recurrence</th>
<th>recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>days_to_death:</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>91</td>
<td>426</td>
<td>898</td>
<td>1019</td>
<td>1392</td>
<td>5722</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>vital_status:</th>
<th>deceased</th>
<th>living</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>46</td>
<td>72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>treatment:</th>
<th>chemotherapy</th>
<th>hormonotherapy</th>
<th>untreated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>23</td>
<td>73</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
</tr>
</thead>
<tbody>
<tr>
<td>STNO2</td>
</tr>
<tr>
<td>118</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>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**

```r
eperimentData(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**

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

         n  events  median 0.95LCL 0.95UCL
          1073 103.00 10.05     8.56    12.05
```
Available sample meta-data:

<table>
<thead>
<tr>
<th>sample_name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length   Class   Mode</td>
</tr>
<tr>
<td>1073      character character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alt_sample_name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length   Class   Mode</td>
</tr>
<tr>
<td>1073      character character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>unique_patient_ID:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length   Class   Mode</td>
</tr>
<tr>
<td>1073      character character</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample_type:</th>
</tr>
</thead>
<tbody>
<tr>
<td>tumor</td>
</tr>
<tr>
<td>1073</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>er:</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative  positive  NA's</td>
</tr>
<tr>
<td>233      790      50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pgr:</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative  positive  NA's</td>
</tr>
<tr>
<td>334      686      53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>her2:</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative  positive  NA's</td>
</tr>
<tr>
<td>549      161      363</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>age_at_initial_pathologic_diagnosis:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. 1st Qu. Median Mean 3rd Qu. Max.</td>
</tr>
<tr>
<td>26.00   49.00   58.00   58.48   68.00   90.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>days_to_death:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. 1st Qu. Median Mean 3rd Qu. Max.</td>
</tr>
<tr>
<td>-7.0     137.0    412.0    820.3   1180.0    6796.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>vital_status:</th>
</tr>
</thead>
<tbody>
<tr>
<td>deceased  living</td>
</tr>
<tr>
<td>103      970</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA 1073</td>
</tr>
</tbody>
</table>
uncurated_author_metadata:
  Length	Class	Mode
  1073 character character

Source

http://cancergenome.nih.gov/

Description

ExpressionSet for the TRANSBIG Dataset

Format

experimentData(eset):
  Experiment data
  Experimenter name:
  Laboratory:
  Contact information: http://www.ncbi.nlm.nih.gov/pubmed/?term=17545524
  Title:
  PMIDs: 17545524
  No abstract available.
  notes:
  summary:
    The 76-gene signature was validated. The results supports the hypothesis that utilizing the gene signature could reduce the number of patients who receive unnecessary adjuvant therapy.
  mapping.method:
    maxRowVariance
  mapping.group:
    EntrezGene.ID
  preprocessing:
    As published by original author.

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

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

\[
\begin{array}{cccc}
\text{n} & \text{events} & \text{median} & 0.95LCL \\ 198.0 & 56.0 & NA & 17.1 \\
\end{array}
\]

Available sample meta-data:

<table>
<thead>
<tr>
<th>sample_name:</th>
<th>sample_type:</th>
<th>er:</th>
<th>tumor_size:</th>
<th>age_at_initial_pathologic_diagnosis:</th>
<th>grade:</th>
<th>dmfs_days:</th>
<th>dmfs_status:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>Class</td>
<td>Mode</td>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
<td>Mean 3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td>198</td>
<td>character</td>
<td>character</td>
<td>0.600</td>
<td>1.700</td>
<td>2.000</td>
<td>2.181</td>
<td>2.500</td>
</tr>
<tr>
<td>198</td>
<td>tumor</td>
<td>198</td>
<td>negative</td>
<td>positive</td>
<td>64</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>0</td>
<td>198</td>
<td>24.00</td>
<td>42.00</td>
<td>46.00</td>
<td>46.39</td>
<td>51.00</td>
</tr>
<tr>
<td>198</td>
<td>1 2 3 NA's</td>
<td>30</td>
<td>83 83 83 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

147 51
<table>
<thead>
<tr>
<th>days_to_tumor_recurrence:</th>
<th>Min. 1st Qu. Median Mean 3rd Qu. Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>121 1528 3534 3399 5130 8711</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>recurrence_status:</th>
<th>norecurrence recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>112 86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>days_to_death:</th>
<th>Min. 1st Qu. Median Mean 3rd Qu. Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>146 2744 4562 4150 5610 9108</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>vital_status:</th>
<th>deceased living</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>56 142</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>treatment:</th>
<th>untreated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>198</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>batch:</th>
<th>VDXGUYU VDXIGRU VDXKIU VDXOXFU VDXRHU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>36 50 51 24 37</td>
</tr>
</tbody>
</table>

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

**Source**


**Description**

ExpressionSet for the UCSF Dataset

**Format**

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

PMIDs: 17428335, 14612510
No abstract available.

notes:
summary:
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

n events median 0.95LCL 0.95UCL
133.00 44.00 11.56 9.25 NA

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

sample_name:

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

sample_type:
tumor
162
<table>
<thead>
<tr>
<th></th>
<th>negative</th>
<th>positive</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>er:</strong></td>
<td>41</td>
<td>101</td>
<td>20</td>
</tr>
<tr>
<td><strong>pgr:</strong></td>
<td>46</td>
<td>94</td>
<td>22</td>
</tr>
<tr>
<td><strong>her2:</strong></td>
<td>35</td>
<td>19</td>
<td>108</td>
</tr>
<tr>
<td><strong>tumor_size:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Qu.</td>
<td>1.800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>2.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.682</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>3.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>11.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A's</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N:</strong></td>
<td>67</td>
<td>82</td>
<td>13</td>
</tr>
<tr>
<td><strong>age_at_initial_pathologic_diagnosis:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>28.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Qu.</td>
<td>44.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>53.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>56.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>70.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>88.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A's</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>grade:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A's</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>dmfs_days:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Qu.</td>
<td>897</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>2040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>8267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A's</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>dmfs_status:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no recurrence</td>
<td>140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>recurrence</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>days_to_tumor_recurrence:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Qu.</td>
<td>861</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1865</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>8267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A's</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>recurrence_status:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no recurrence</td>
<td>125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>recurrence</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>days_to_death:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Qu.</td>
<td>1087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>2054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>3087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>8267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A's</td>
<td>29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
vital_status:
deceased  living  NA's
 54       99       9

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

batch:
UCSF
162

uncurated_author_metadata:
Length  Class  Mode
162  character  character

Source
Not available

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

N:  
0 1 NA's
126 135 44

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

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

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

recurrence_status:  
norecurrence recurrence NA's
170 70 65

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

vital_status:  
deceased living NA's
51 189 65

batch:  
UNC4
305

uncurated_author_metadata:  
Length Class Mode
305 character character

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

Source
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
<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><strong>alt_sample_name:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length Class Mode</td>
<td></td>
<td></td>
<td>133</td>
<td>character character</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>sample_type:</strong></td>
<td></td>
<td></td>
<td>133</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>er:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>negative positive NA's</td>
<td>40</td>
<td>86</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>pgr:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>negative positive NA's</td>
<td>6</td>
<td>56</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>tumor_size:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>1.200</td>
<td>1.900</td>
<td>1.892</td>
<td>2.300</td>
<td>6.000</td>
</tr>
<tr>
<td><strong>N:</strong></td>
<td></td>
<td></td>
<td>133</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>age_at_initial_pathologic_diagnosis:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24.00</td>
<td>44.00</td>
<td>53.00</td>
<td>51.79</td>
<td>60.00</td>
<td>73.00</td>
</tr>
<tr>
<td><strong>grade:</strong></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3 NA's</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>51</td>
<td>29</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>dmfs_days:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>61</td>
<td>1338</td>
<td>2809</td>
<td>2724</td>
<td>4078</td>
<td>5305</td>
</tr>
<tr>
<td><strong>dmfs_status:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no recurrence recurrence NA's</td>
<td>97</td>
<td>28</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>days_to_tumor_recurrence:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>61</td>
<td>1338</td>
<td>2675</td>
<td>2687</td>
<td>3912</td>
<td>5305</td>
</tr>
<tr>
<td><strong>recurrence_status:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no recurrence recurrence NA's</td>
<td>76</td>
<td>49</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
treatment:
untreated
133

batch:
KIU OXFU
64 69

uncurated_author_metadata:
<table>
<thead>
<tr>
<th>Length</th>
<th>Class</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>133</td>
<td>character</td>
<td>character</td>
</tr>
</tbody>
</table>

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 p53 mutant and wild-type tumors. The signature is more effective than sequence-based assessments of p53 in predicting prognosis and therapeutic response.
   mapping.method:
      maxRowVariance
mapping.group:
   EntrezGene.ID
preprocessing:
As published by original author.

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

Details

assayData: 36178 features, 251 samples
Platform type:
---------------------------
Available sample meta-data:
---------------------------

sample_name:
  Length  Class     Mode
  251 character character

alt_sample_name:
  Length  Class     Mode
  251 character character

sample_type:
tumor
  251

er:
negative  positive  NA's
  34   213       4

pgr:
negative  positive
  61   190

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

N:
  0 1 NA's
  158 84 9

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


## VDX

### grades:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>128</td>
<td>54</td>
<td>2</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>30</td>
<td>1870</td>
<td>3711</td>
<td>3007</td>
<td>3985</td>
<td>4654</td>
<td>17</td>
</tr>
</tbody>
</table>

### recurrence_status:

<table>
<thead>
<tr>
<th>norecurrence</th>
<th>recurrence</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>181</td>
<td>55</td>
<td>15</td>
</tr>
</tbody>
</table>

### treatment:

<table>
<thead>
<tr>
<th>hormonotherapy</th>
<th>untreated</th>
<th>NA's</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>142</td>
<td>29</td>
</tr>
</tbody>
</table>

### batch:

UPPT UPPU

| 80 | 171 |

### uncurated_author_metadata:

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

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

sample_type:
tumor
  344

er:
negative positive
  135   209

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

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

grade:
1 2 3 NA's
7 42 148 147

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

dmfs_status:
norecurrence recurrence
226 118

treatment:
untreated
344

batch:
VDX VDXN
286 58

uncurated_author_metadata:
Length Class Mode
344 character character

Source
Index

* datasets
  CAL, 3
  DFHCC, 5
  DFHCC2, 8
  DFHCC3, 10
  DUKE, 11
  DUKE2, 13
duplicates, 15
  EMC2, 16
  EORTC10994, 18
  EXPO, 20
  FNCLCC, 22
  GSE25066, 23
  GSE32646, 26
  GSE48091, 28
  GSE58644, 29
  HLP, 32
  IRB, 34
  KOO, 36
  LUND, 40
  LUND2, 42
  MAINZ, 44
  MAQC2, 46
  MCCC, 48
  MDA4, 49
  METABRIC, 51
  MSK, 54
  MUG, 56
  NCCS, 57
  NCI, 59
  NKI, 61
  PNC, 64
  STK, 66
  STNO2, 68
  TCGA, 71
  TRANSBIG, 73
  UCSF, 75
  UNC4, 78
  UNT, 81
  UPP, 83
  VDX, 85
  CAL (CAL), 3
  DFHCC (DFHCC), 5
  DFHCC2 (DFHCC2), 8
  DFHCC3 (DFHCC3), 10
  DUKE (DUKE), 11
  DUKE2 (DUKE2), 13
  EMC2 (EMC2), 16
  EORTC10994 (EORTC10994), 18
  EXPO (EXPO), 20
  FNCLCC (FNCLCC), 22
  GSE25066 (GSE25066), 23
  GSE32646 (GSE32646), 26
  GSE48091 (GSE48091), 28
  GSE58644 (GSE58644), 29
  HLP (HLP), 32
  IRB (IRB), 34
  KOO (KOO), 36
  LUND (LUND), 40
  LUND2 (LUND2), 42
  MAINZ (MAINZ), 44
  MAQC2 (MAQC2), 46
  MCCC (MCCC), 48
  MDA4 (MDA4), 49
  METABRIC (METABRIC), 51
  MSK (MSK), 54
  MUG (MUG), 56
  NCCS (NCCS), 57
  NCI (NCI), 59
  NKI (NKI), 61
  PNC (PNC), 64
  STK (STK), 66
  STNO2 (STNO2), 68
  TCGA (TCGA), 71
  TRANSBIG (TRANSBIG), 73
  UCSF (UCSF), 75
  UNC4 (UNC4), 78
  UNT (UNT), 81
INDEX

UPP (UPP), 83
VDX (VDX), 85

TCGA, 71
TRANSBIG, 73

CAL, 3
UCSF, 75

DFHCC, 5
UNC4, 78

DFHCC2, 8
UPP, 83

DFHCC3, 10
VDX, 85

DUKE, 11
DUKE2, 13
duplicates, 15

EMC2, 16
EORTC18994, 18
EXPO, 20

FNCLCC, 22

GSE25066, 23
GSE32646, 26
GSE48091, 28
GSE58644, 29

HLP, 32
IRB, 34

KOO, 36

loadBreastDatasets, 38
loadBreastEsets, 39

LUND, 40
LUND2, 42

MAINZ, 44
MAQC2, 46
MCCC, 48
MDA4, 49
METABRIC, 51
MSK, 54
MUG, 56

NCCS, 57
NCI, 59
NKI, 61

PNC, 64

STK, 66
STNO2, 68