Package 'curatedCRCData'

April 8, 2015

Type Package

Title Colorectal Cancer Gene Expression Analysis

Version 1.1.2

Date 2013-10-21

Author Princy Parsana, Markus Riester, Curtis Huttenhower, Levi Waldron

Maintainer Princy Parsana <princyparsana@jhu.edu>

Description The curatedCRC package provides relevant functions and data for gene expression analysis in patients with colorectal cancer.

Depends R (>= 2.10.0), nlme

Suggests survival,sva,BiocStyle, xtable, genefilter, logging, Biobase

License Artistic-2.0

Namespace auto

biocViews Colorectal, Cancer, TCGA, ExperimentData

URL https://bitbucket.org/lwaldron/curatedcrcdata

R topics documented:

`curatedCRCData-package`

Clinically Annotated Data for the CRC Cancer Transcriptome

Description

The curatedCRCData package provides manually curated clinical data, uniformly processed expression data, and convenience functions for gene expression analysis in patients with colorectal cancer.

Details

Package:	curatedCRCData
Type:	Package
Version:	1.1.2
Date:	2013-10-21
License:	Artistic-2.0
Depends:	R (>= 2.10.0), affy

Author(s)

Princy Parsana, Markus Riester, Levi Waldron

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health

Maintainer: Levi Waldron <levi@jimmy.harvard.edu>

Examples

```
##List all datasets:
data(package="curatedCRCData")
##
```

GSE11237_eset *Celecoxib pre-treatment in human colorectal adenocarcinoma patients is associated with gene expression alterations suggestive of diminished cellular proliferation.*

Description

Cancer cells treated with the cyclooxygenase-2 inhibitor celecoxib show growth inhibition and induced apoptosis. This study was conducted to determine if the same processes are relevant to celecoxib's effects on human colorectal adenocarcinomas treated in vivo. A cohort of 23 patients with primary colorectal adenocarcinomas was randomised to receive a 7-d course of celecoxib (400mg b.i.d.) or no drug prior to surgical resection. Gene expression profiling was performed on resected adenocarcinomas from the cohort of patients. Using fold change (>1.5) and p-value (<0.05) cutoffs, 190 genes were differentially expressed between adenocarcinomas from patients receiving celecoxib and those that did not. The celecoxib pre-treated samples showed decreased expression levels in multiple genes involved in cellular lipid and glutathione metabolism; changes associated with diminished cellular proliferation. Celecoxib pre-treatment for 7 d in vivo is associated with alterations in colorectal adenocarcinoma gene expression which are suggestive of diminished cellular proliferation.

Usage

```
data( GSE11237_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Auman JT, Church R, Lee SY, Watson MA, Fleshman JW, Mcleod HL.??Celecoxib pre-treatm
  Laboratory: Auman, Mcleod 2008
  Contact information:
 Title: Celecoxib pre-treatment in human colorectal adenocarcinoma patients is associated with gene exp
  URL:
  PMIDs: 18653328
```

```
  Abstract: A 147 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      Affymetrix Human Genome U95 Version 2 Array
   platform_shorttitle:
      Affymetrix HG_U95Av2
   platform_summary:
      hg_u95av2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL8300
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: rma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: AADAC AAK1 ... ZZZ3 (8933 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 8933 features, 23 samples
Platform type: hg_u95av2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       23 character character

sample_type:
tumor
   23

summarygrade:
high  low
   4   19
```

```
G:
 1  2  3
 3 16  4

summarystage:
early  late
    4    19

T:
 1  2  3  4
 1  3 16  3

N:
 0  1  2
15  7  1

M:
 0  1
17  6

location:
   Length     Class      Mode
       23 character character

gender:
 f  m
13 10

stageall:
1 2 3 4
4 9 4 6

batch:
2003-08-19 2003-08-28
        15          8

drug_name:
celecoxib      NAs
       11        12

drug_treatment:
 n  y
12 11

uncurated_author_metadata:
   Length     Class      Mode
       23 character character
```

GSE12225.GPL3676_eset *Integrating chromosomal aberrations and gene expression profiles to dissect rectal tumorigenesis.*

Description

Accurate staging of rectal tumors is essential for making the correct treatment choice. In a previous study, we found that loss of 17p, 18q and gain of 8q, 13q and 20q could distinguish adenoma from carcinoma tissue and that gain of 1q was related to lymph node metastasis. In order to find markers for tumor staging, we searched for candidate genes on these specific chromosomes.We performed gene expression microarray analysis on 79 rectal tumors and integrated these data with genomic data from the same sample series. We performed supervised analysis to find candidate genes on affected chromosomes and validated the results with qRT-PCR and immunohistochemistry.Integration of gene expression and chromosomal instability data revealed similarity between these two data types. Supervised analysis identified up-regulation of EFNA1 in cases with 1q gain, and EFNA1 expression was correlated with the expression of a target gene (VEGF). The BOP1 gene, involved in ribosome biogenesis and related to chromosomal instability, was over-expressed in cases with 8q gain. SMAD2 was the most down-regulated gene on 18q, and on 20q, STMN3 and TGIF2 were highly up-regulated. Immunohistochemistry for SMAD4 correlated with SMAD2 gene expression and 18q loss.On basis of integrative analysis this study identified one well known CRC gene (SMAD2) and several other genes (EFNA1, BOP1, TGIF2 and STMN3) that possibly could be used for rectal cancer characterization.

Usage

```
data( GSE12225.GPL3676_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Lips EH, van Eijk R, de Graaf EJ, Oosting J et al.??Integrating chromosomal aberrati
  Laboratory: Lips,  Morreau 2008
  Contact information:
 Title: Integrating chromosomal aberrations and gene expression profiles to dissect rectal tumorigenesi
  URL:
  PMIDs: 18959792

  Abstract: A 221 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      NKI-CMF Homo sapiens 35k oligo array
   platform_shorttitle:
      NA
   platform_summary:
      nki-cmf homo sapiens 35k oligo array
```

```
   platform_manufacturer:
      Central Microarray Facility, NKI Amsterdam
   platform_distribution:
      non-commercial
   platform_accession:
      GPL3676
   platform_technology:
      spotted oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1CF ... ZZZ3 (19727 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19727 features, 79 samples
Platform type: nki-cmf homo sapiens 35k oligo array
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       79 character character

sample_type:
tumor
   79

T:
   0    1    2    3 NAs
  21   21   29    1    7

N:
 0  1  2  3
67  6  4  2

uncurated_author_metadata:
   Length     Class      Mode
       79 character character
```

GSE12945_eset *An expression module of WIPF1-coexpressed genes identifies patients with favorable prognosis in three tumor types.*

Description

Wiskott-Aldrich syndrome (WAS) predisposes patients to leukemia and lymphoma. WAS is caused by mutations in the protein WASP which impair its interaction with the WIPF1 protein. Here, we aim to identify a module of WIPF1-coexpressed genes and to assess its use as a prognostic signature for colorectal cancer, glioma, and breast cancer patients. Two public colorectal cancer microarray data sets were used for discovery and validation of the WIPF1 co-expression module. Based on expression of the WIPF1 signature, we classified more than 400 additional tumors with microarray data from our own experiments or from publicly available data sets according to their WIPF1 signature expression. This allowed us to separate patient populations for colorectal cancers, breast cancers, and gliomas for which clinical characteristics like survival times and times to relapse were analyzed. Groups of colorectal cancer, breast cancer, and glioma patients with low expression of the WIPF1 co-expression module generally had a favorable prognosis. In addition, the majority of WIPF1 signature genes are individually correlated with disease outcome in different studies. Literature gene network analysis revealed that among WIPF1 co-expressed genes known direct transcriptional targets of c-myc, ESR1 and p53 are enriched. The mean expression profile of WIPF1 signature genes is correlated with the profile of a proliferation signature. The WIPF1 signature is the first microarray-based prognostic expression signature primarily developed for colorectal cancer that is instrumental in other tumor types: low expression of the WIPF1 module is associated with better prognosis.

Usage

```
data( GSE12945_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Staub E, Groene J, Heinze M, Mennerich D et al.??An expression module of WIPF1-coexp
  Laboratory: Staub, Rosenthal 2009
  Contact information:
 Title: An expression module of WIPF1-coexpressed genes identifies patients with favorable prognosis in
  URL:
  PMIDs: 19399471

  Abstract: A 241 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133A] Affymetrix Human Genome U133A Array
   platform_shorttitle:
      Affymetrix HG-U133A
```

```
   platform_summary:
      hg-u133a
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL96
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1CF A2M ... ZZZ3 (12986 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 12986 features, 62 samples
Platform type: hg-u133a
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

records   n.max n.start  events  median 0.95LCL 0.95UCL
     62      62      62      12      NA      NA      NA

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

alt_sample_name:
   Length     Class      Mode
       62 character character

sample_type:
tumor
   62

primarysite:
co re
29 33

summarygrade:
high  low
```

```
  31   31

G:
 2  3
31 31

summarystage:
early  late
   16    46

T:
 2  3  4
16 42  4

N:
 0  1  2
36 14 12

M:
   0    1 NAs
  56    5    1

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  38.00   59.00   65.00   64.45   73.75   87.00

days_to_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    210    1005    1395    1267    1620    1920

vital_status:
deceased   living
      12       50

location:
   Length     Class      Mode
       62 character character

gender:
 f  m
28 34

lymphnodesremoved:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  11.00   13.25   16.50   19.00   22.75   42.00

lymphnodesinvaded:
  11   13   20 NAs
```

```
   2    1    1   58

stageall:
 1  2  3  4
13 23 21  5

batch:
   Length     Class      Mode
       62 character character

uncurated_author_metadata:
   Length     Class      Mode
       62 character character
```

GSE13067_eset *DNA copy-number alterations underlie gene expression differences between microsatellite stable and unstable colorectal cancers.*

Description

About 15% of colorectal cancers harbor microsatellite instability (MSI). MSI-associated gene expression changes have been identified in colorectal cancers, but little overlap exists between signatures hindering an assessment of overall consistency. Little is known about the causes and downstream effects of differential gene expression.DNA microarray data on 89 MSI and 140 microsatellite-stable (MSS) colorectal cancers from this study and 58 MSI and 77 MSS cases from three published reports were randomly divided into test and training sets. MSI-associated gene expression changes were assessed for cross-study consistency using training samples and validated as MSI classifier using test samples. Differences in biological pathways were identified by functional category analysis. Causation of differential gene expression was investigated by comparison to DNA copy-number data.MSI-associated gene expression changes in colorectal cancers were found to be highly consistent across multiple studies of primary tumors and cancer cell lines from patients of different ethnicities ($P < 0.001$). Clustering based on consistent changes separated additional test cases by MSI status, and classification of individual samples predicted MSI status with a sensitivity of 96% and specificity of 85%. Genes associated with immune response were up-regulated in MSI cancers, whereas genes associated with cell-cell adhesion, ion binding, and regulation of metabolism were down-regulated. Differential gene expression was shown to reflect systematic differences in DNA copy-number aberrations between MSI and MSS tumors ($P < 0.001$).Our results show cross-study consistency of MSI-associated gene expression changes in colorectal cancers. DNA copy-number alterations partly cause the differences in gene expression between MSI and MSS cancers.

Usage

```
data( GSE13067_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Jorissen RN, Lipton L, Gibbs P, Chapman M et al.??DNA copy-number alterations underl
  Laboratory: Jorissen and Sieber 2008
  Contact information:
 Title: DNA copy-number alterations underlie gene expression differences between microsatellite stable
  URL:
  PMIDs: 19088021

  Abstract: A 251 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 74 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
```

```
       74 character character

sample_type:
tumor
   74

msi:
 n  y
63 11

mss:
 n  y
11 63

batch:
   Length     Class      Mode
       74 character character

uncurated_author_metadata:
   Length     Class      Mode
       74 character character
```

GSE13294_eset *DNA copy-number alterations underlie gene expression differences between microsatellite stable and unstable colorectal cancers.*

Description

About 15% of colorectal cancers harbor microsatellite instability (MSI). MSI-associated gene expression changes have been identified in colorectal cancers, but little overlap exists between signatures hindering an assessment of overall consistency. Little is known about the causes and downstream effects of differential gene expression.DNA microarray data on 89 MSI and 140 microsatellite-stable (MSS) colorectal cancers from this study and 58 MSI and 77 MSS cases from three published reports were randomly divided into test and training sets. MSI-associated gene expression changes were assessed for cross-study consistency using training samples and validated as MSI classifier using test samples. Differences in biological pathways were identified by functional category analysis. Causation of differential gene expression was investigated by comparison to DNA copy-number data.MSI-associated gene expression changes in colorectal cancers were found to be highly consistent across multiple studies of primary tumors and cancer cell lines from patients of different ethnicities ($P < 0.001$). Clustering based on consistent changes separated additional test cases by MSI status, and classification of individual samples predicted MSI status with a sensitivity of 96% and specificity of 85%. Genes associated with immune response were up-regulated in MSI cancers, whereas genes associated with cell-cell adhesion, ion binding, and regulation of metabolism were down-regulated. Differential gene expression was shown to reflect systematic differences in DNA copy-number aberrations between MSI and MSS tumors ($P < 0.001$).Our results show cross-study consistency of MSI-associated gene expression changes in colorectal cancers. DNA copy-number alterations partly cause the differences in gene expression between MSI and MSS cancers.

Usage

```
data( GSE13294_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Jorissen RN, Lipton L, Gibbs P, Chapman M et al. DNA copy-number alterations underli
  Laboratory: Jorissen and Sieber 2008
  Contact information:
 Title: DNA copy-number alterations underlie gene expression differences between microsatellite stable
  URL:
  PMIDs: 19088021

  Abstract: A 251 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 155 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
```

```
---------------------------

alt_sample_name:
   Length     Class      Mode
      155 character character

sample_type:
tumor
  155

msi:
 n  y
77 78

mss:
 n  y
78 77

batch:
   Length     Class      Mode
      155 character character

uncurated_author_metadata:
   Length     Class      Mode
      155 character character
```

GSE14095_eset *Gene expression signature and response to the use of leucovorin, fluorouracil and oxaliplatin in colorectal cancer patients.*

Description

FOLFOX (a combination of leucovorin, fluorouracil and oxaliplatin) has achieved substantial success in the treatment of colorectal cancer (CRC) patients. However, about half of all patients show resistance to this regimen and some develop adverse symptoms such as neurotoxicity. In order to select patients who would benefit most from this therapy, we aimed to build a predictor for the response to FOLFOX using microarray gene expression profiles of primary CRC samples.Forty patients who underwent surgery for primary lesions were examined. All patients had metastatic or recurrent CRC and received modified FOLFOX6. Responders and nonresponders were determined according to the best observed response at the end of the first-line treatment. Gene-expression profiles of primary CRC were determined using Human Genome GeneChip arrays U133. We identified discriminating genes whose expression differed significantly between responders and nonresponders and then carried out supervised class prediction using the k-nearest-neighbour method.We identified 27 probes that were differentially expressed between responders and nonresponders at significant levels. Based on the expression of these genes, we constructed a FOLFOX response

predictor with an overall accuracy of 92.5%. The sensitivity, specificity, positive and negative predictive values were 78.6%, 100%, 100% and 89.7%, respectively.The present model suggests the possibility of selecting patients who would benefit from FOLFOX therapy both in the metastatic and the adjuvant setting. To our knowledge, this is the first study to establish a prediction model for the response to FOLFOX chemotherapy based on gene expression by microarray analysis.

Usage

```
data( GSE14095_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Watanabe T, Kobunai T, Yamamoto Y, Matsuda K et al.??Gene expression signature and r
  Laboratory: Watanabe, Hashimoto 2008
  Contact information:
 Title: Gene expression signature and response to the use of leucovorin, fluorouracil and oxaliplatin in
  URL:
  PMIDs: 21680303

  Abstract: A 241 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 189 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
      189 character character

uncurated_author_metadata:
   Length     Class      Mode
      189 character character
```

`GSE14333_eset` *Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer.*

Description

PURPOSE: Colorectal cancer prognosis is currently predicted from pathologic staging, providing limited discrimination for Dukes stage B and C disease. Additional markers for outcome are required to help guide therapy selection for individual patients. EXPERIMENTAL DESIGN: A multisite single-platform microarray study was done on 553 colorectal cancers. Gene expression changes were identified between stage A and D tumors (three training sets) and assessed as a prognosis signature in stage B and C tumors (independent test and external validation sets). RESULTS: One hundred twenty-eight genes showed reproducible expression changes between three sets of stage A and D cancers. Using consistent genes, stage B and C cancers clustered into two groups resembling early-stage and metastatic tumors. A Prediction Analysis of Microarray algorithm was developed to classify individual intermediate-stage cancers into stage A-like/good prognosis or stage D-like/poor prognosis types. For stage B patients, the treatment adjusted hazard ratio for 6-year recurrence in individuals with stage D-like cancers was 10.3 (95% confidence interval, 1.3-80.0; P = 0.011). For stage C patients, the adjusted hazard ratio was 2.9 (95% confidence interval, 1.1-7.6; P = 0.016). Similar results were obtained for an external set of stage B and C patients. The prognosis signature was enriched for downregulated immune response genes and upregulated cell signaling and extracellular matrix genes. Accordingly, sparse tumor infiltration with mononuclear chronic inflammatory cells was associated with poor outcome in independent patients. CONCLUSIONS: Metastasis-associated gene expression changes can be used to refine traditional outcome prediction, providing a rational approach for tailoring treatments to subsets of patients. (Clin Cancer Res 2009;15(24):7642-51).

Usage

```
data( GSE14333_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Jorissen RN, Gibbs P, Christie M, Prakash S et al.??Metastasis-Associated Gene Expr
  Laboratory: Jorissen and Sieber 2008
  Contact information:
 Title: Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stag
  URL:
  PMIDs: 19996206

  Abstract: A 257 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 290 samples
Platform type: hg-u133_plus_2
----------------------------
Available sample meta-data:
----------------------------

alt_sample_name:
   Length     Class      Mode
```

```
      290 character character

sample_type:
tumor
  290

summarystage:
early  late
  138   152

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  26.00   58.00   67.00   65.96   75.00   92.00

summarylocation:
   l    r NAs
 122  125   43

gender:
  f   m
126 164

stageall:
 1  2  3  4
44 94 91 61

dfs_status:
deceased_or_recurrence    living_norecurrence                   NAs
                   176                     50                    64

days_to_recurrence_or_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
   27.6   668.5  1154.0  1306.0  1785.0  4276.0      64

batch:
   Length     Class      Mode
      290 character character

uncurated_author_metadata:
   Length     Class      Mode
      290 character character
```

GSE16125.GPL5175_eset *Integrative approach for prioritizing cancer genes in sporadic colon cancer.*

Description

The current multistep carcinogenesis models of colon cancer do not fully capture the genetic heterogeneity of the disease, which is additionally complicated by the presence of passenger and driver genetic alterations. The aim of this study was to select in the context of this significant heterogeneity additional genes functionally related to colon cancer development. High-throughput copy number and gene expression data of 36 microsatellite stable sporadic colon cancers resected from patients of a single institution characterized for mutations in APC, KRAS, TP53 and loss of 18q were analyzed. Genes whose expression correlated with the underlying copy number pattern were selected, and their association with the above listed mutations and overall survival was evaluated. Gain of 20q was strongly associated with TP53 mutation, and overall survival with alterations on 7p, 8p, 13q, 18q, and 20q. An association with 18q loss and gain of 8q24 was also observed. New candidate genes with a potential role in colon cancer are PLCG1 on 20q, DBC1 on 8q21, and NDGR1 on 8p24. In addition, an unexpected pattern of loss and mutability was found in the region upstream of the KRAS gene. By integrating copy number alterations with gene expression and mutations in colon cancer associated genes, we have developed a strategy that identifies previously known molecular features and additional players in the molecular landscape of colon cancer.Copyright 2009 Wiley-Liss, Inc.

Usage

```
data( GSE16125.GPL5175_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Reid JF, Gariboldi M, Sokolova V, Capobianco P et al.??Integrative approach for pric
  Laboratory: Reid, Pierotti 2009
  Contact information:
 Title: Integrative approach for prioritizing cancer genes in sporadic colon cancer.
  URL:
  PMIDs: 19672874

  Abstract: A 225 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) versio
n]
   platform_shorttitle:
      Affymetrix HuEx-1_0-st
   platform_summary:
      huex-1_0-st
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
```

```
      GPL5175
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: ABCF1 ANAPC5 ... ZNF259P1///ZNF259 (282 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 282 features, 36 samples
Platform type: huex-1_0-st
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

   4 observations deleted due to missingness
records   n.max n.start  events  median 0.95LCL 0.95UCL
 32.000  32.000  32.000  22.000   1.397   0.822      NA

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

alt_sample_name:
   Length     Class      Mode
       36 character character

summarystage:
early  late  NAs
    6    27     3

T:
   1    2    3    4 NAs
   1    5    8   19    3

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  19.00   61.75   67.50   65.58   71.25   90.00

days_to_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
   30.0   202.5   390.0   926.2  1658.0  2970.0       4
```

```
vital_status:
deceased   living     NAs
      22       10        4

gender:
 f  m
20 16

kras:
 n  y
16 20

mutation_apc:
 n  y
19 17

batch:
   Length     Class      Mode
       36 character character

uncurated_author_metadata:
   Length     Class      Mode
       36 character character
```

GSE17536_eset *Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer.*

Description

Staging inadequately predicts metastatic risk in patients with colon cancer. We used a gene expression profile derived from invasive, murine colon cancer cells that were highly metastatic in an immunocompetent mouse model to identify patients with colon cancer at risk of recurrence.This phase 1, exploratory biomarker study used 55 patients with colorectal cancer from Vanderbilt Medical Center (VMC) as the training dataset and 177 patients from the Moffitt Cancer Center as the independent dataset. The metastasis-associated gene expression profile developed from the mouse model was refined with comparative functional genomics in the VMC gene expression profiles to identify a 34-gene classifier associated with high risk of metastasis and death from colon cancer. A metastasis score derived from the biologically based classifier was tested in the Moffitt dataset.A high score was significantly associated with increased risk of metastasis and death from colon cancer across all pathologic stages and specifically in stage II and stage III patients. The metastasis score was shown to independently predict risk of cancer recurrence and death in univariate and multivariate models. For example, among stage III patients, a high score translated to increased relative risk of cancer recurrence (hazard ratio, 4.7; 95% confidence interval, 1.566-14.05). Furthermore, the metastasis score identified patients with stage III disease whose 5-year recurrence-free survival was >88% and for whom adjuvant chemotherapy did not increase survival time.A gene expression

profile identified from an experimental model of colon cancer metastasis predicted cancer recurrence and death, independently of conventional measures, in patients with colon cancer.

Usage

```
data( GSE17536_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Smith JJ, Deane NG, Wu F, Merchant NB et al.??Experimentally derived metastasis gene
  Laboratory: Smith JJ,??Beauchamp RD 2009
  Contact information:
 Title: Experimentally derived metastasis gene expression profile predicts recurrence and death in pati
  URL:
  PMIDs: 19914252

  Abstract: A 260 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 177 samples
Platform type: hg-u133_plus_2
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

records   n.max n.start  events  median 0.95LCL 0.95UCL
 177.00  177.00  177.00   73.00   11.08    4.86      NA

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

alt_sample_name:
   Length     Class      Mode
      177 character character

summarygrade:
high  low
  27  150

G:
  1   2   3
 16 134  27

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  26.00   57.00   66.00   65.48   75.00   92.00

recurrence_status:
norecurrence   recurrence         NAs
         109           36           32

days_to_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   27.6   683.4  1268.0  1444.0  2035.0  4276.0

vital_status:
deceased   living
      73      104

gender:
 f  m
81 96

stageall:
 1  2  3  4
24 57 57 39
```

```
ethnicity:
    black caucasian  hispanic     other
        9       151         1        16

batch:
   Length     Class      Mode
      177 character character

uncurated_author_metadata:
   Length     Class      Mode
      177 character character
```

GSE17537_eset *Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer.*

Description

Staging inadequately predicts metastatic risk in patients with colon cancer. We used a gene expression profile derived from invasive, murine colon cancer cells that were highly metastatic in an immunocompetent mouse model to identify patients with colon cancer at risk of recurrence.This phase 1, exploratory biomarker study used 55 patients with colorectal cancer from Vanderbilt Medical Center (VMC) as the training dataset and 177 patients from the Moffitt Cancer Center as the independent dataset. The metastasis-associated gene expression profile developed from the mouse model was refined with comparative functional genomics in the VMC gene expression profiles to identify a 34-gene classifier associated with high risk of metastasis and death from colon cancer. A metastasis score derived from the biologically based classifier was tested in the Moffitt dataset.A high score was significantly associated with increased risk of metastasis and death from colon cancer across all pathologic stages and specifically in stage II and stage III patients. The metastasis score was shown to independently predict risk of cancer recurrence and death in univariate and multivariate models. For example, among stage III patients, a high score translated to increased relative risk of cancer recurrence (hazard ratio, 4.7; 95% confidence interval, 1.566-14.05). Furthermore, the metastasis score identified patients with stage III disease whose 5-year recurrence-free survival was >88% and for whom adjuvant chemotherapy did not increase survival time.A gene expression profile identified from an experimental model of colon cancer metastasis predicted cancer recurrence and death, independently of conventional measures, in patients with colon cancer.

Usage

```
data( GSE17537_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Smith JJ, Deane NG, Wu F, Merchant NB et al.??Experimentally derived metastasis gene
  Laboratory: Smith JJ,??Beauchamp RD 2009
  Contact information:
 Title: Experimentally derived metastasis gene expression profile predicts recurrence and death in pati
  URL:
  PMIDs: 19914252

  Abstract: A 260 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 55 samples
Platform type: hg-u133_plus_2
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

records   n.max n.start  events  median 0.95LCL 0.95UCL
     55      55      55      20      NA      NA      NA
```

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

alt_sample_name:
   Length     Class      Mode
       55 character character

summarygrade:
high  low NAs
   3   33   19

G:
   1    2    3 NAs
   1   32    3   19

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  23.00   54.00   62.00   62.31   72.00   94.00

days_to_tumor_recurrence:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   13.81 1243.00  988.10 1694.00 2303.00

recurrence_status:
norecurrence   recurrence
          36           19

days_to_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.82  950.30 1506.00 1357.00 1801.00 3345.00

vital_status:
deceased   living
      20       35

gender:
 f  m
29 26

stageall:
 1  2  3  4
 4 15 19 17

ethnicity:
    black caucasian  hispanic
        4        50         1
```

```
batch:
   Length     Class      Mode
       55 character character

uncurated_author_metadata:
   Length     Class      Mode
       55 character character
```

GSE17538.GPL570_eset *Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer.*

Description

Staging inadequately predicts metastatic risk in patients with colon cancer. We used a gene expression profile derived from invasive, murine colon cancer cells that were highly metastatic in an immunocompetent mouse model to identify patients with colon cancer at risk of recurrence.This phase 1, exploratory biomarker study used 55 patients with colorectal cancer from Vanderbilt Medical Center (VMC) as the training dataset and 177 patients from the Moffitt Cancer Center as the independent dataset. The metastasis-associated gene expression profile developed from the mouse model was refined with comparative functional genomics in the VMC gene expression profiles to identify a 34-gene classifier associated with high risk of metastasis and death from colon cancer. A metastasis score derived from the biologically based classifier was tested in the Moffitt dataset.A high score was significantly associated with increased risk of metastasis and death from colon cancer across all pathologic stages and specifically in stage II and stage III patients. The metastasis score was shown to independently predict risk of cancer recurrence and death in univariate and multivariate models. For example, among stage III patients, a high score translated to increased relative risk of cancer recurrence (hazard ratio, 4.7; 95% confidence interval, 1.566-14.05). Furthermore, the metastasis score identified patients with stage III disease whose 5-year recurrence-free survival was >88% and for whom adjuvant chemotherapy did not increase survival time.A gene expression profile identified from an experimental model of colon cancer metastasis predicted cancer recurrence and death, independently of conventional measures, in patients with colon cancer.

Usage

```
data( GSE17538.GPL570_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Smith JJ, Deane NG, Wu F, Merchant NB et al.??Experimentally derived metastasis gene
  Laboratory: Smith JJ,??Beauchamp RD 2009
  Contact information:
 Title: Experimentally derived metastasis gene expression profile predicts recurrence and death in pati
```

```
  URL:
  PMIDs: 19914252

  Abstract: A 260 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 238 samples
Platform type: hg-u133_plus_2
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

   6 observations deleted due to missingness
records   n.max n.start  events  median 0.95LCL 0.95UCL
 232.00  232.00  232.00   93.00   11.08    5.57      NA

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

alt_sample_name:
   Length     Class      Mode
      238 character character
```

```
summarygrade:
high  low NAs
  30  183   25

G:
   1    2    3 NAs
  17  166   30   25

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  23.00   56.00   65.00   64.56   74.00   94.00

days_to_tumor_recurrence:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
    0.0   216.2   906.8  1093.0  1713.0  4276.0       6

recurrence_status:
norecurrence   recurrence         NAs
         145           55           38

days_to_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
  12.82  699.70 1402.00 1423.00 1919.00 4276.00       6

vital_status:
deceased   living     NAs
      93      139        6

gender:
  f   m
114 124

stageall:
   1    2    3    4 NAs
  28   72   76   56    6

ethnicity:
    black caucasian  hispanic     other
       13       206         2        17

batch:
   Length     Class      Mode
      238 character character

uncurated_author_metadata:
   Length     Class      Mode
      238 character character
```

GSE18105_eset *MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal cancer.*

Description

Distant metastasis is the major cause of death in colorectal cancer (CRC) patients. To identify genes influencing the prognosis of patients with CRC, we compared gene expression in primary tumors with and without distant metastasis using an oligonucleotide microarray. We also examined the expression of the candidate gene in 100 CRC patients by quantitative real-time reverse transcription PCR and studied the relationship between its expression and the prognosis of patients with CRC. As a result, we identified MUC12 as a candidate gene involved in metastasis processes by microarray analysis. Quantitative real-time reverse transcription PCR showed that MUC12 expression was significantly lower in cancer tissues than in adjacent normal tissues ($p < 0.001$). In Stages II and III CRC, patients with low expression showed worse disease-free survival ($p = 0.020$). Multivariate analysis disclosed that MUC12 expression status was an independent prognostic factor in Stages II and III CRC (relative risk, 8.236; 95% confidence interval, 1.702-39.849 $p = 0.009$). Our study revealed the prognostic value of MUC12 expression in CRC patients. Moreover, our result suggests MUC12 expression is a possible candidate gene for assessing postoperative adjuvant therapy for CRC patients.

Usage

```
data( GSE18105_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Matsuyama T, Ishikawa T, Mogushi K, Yoshida T et al.??MUC12 mRNA expression is an in
  Laboratory: Matsuyama, Sugihara 2009
  Contact information:
 Title: MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal
  URL:
  PMIDs: 20162577

  Abstract: A 188 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
```

```
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 111 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
      111 character character

sample_type:
adjacentnormal          tumor
            17             94

M:
 0  1
67 44

batch:
   Length     Class      Mode
      111 character character

uncurated_author_metadata:
   Length     Class      Mode
      111 character character
```

GSE2109_eset *Expression Project for Oncology (expO)*

Description

EXpression Project for Oncology, International Genomics Consortium, www.intgen.org

Usage

```
data( GSE2109_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: EXpression Project for Oncology, International Genomics Consortium, www.intgen.org
  Laboratory: expO, IGC 2005
  Contact information:
  Title: Expression Project for Oncology (expO)
  URL:
  PMIDs: PMID unknown

  Abstract: A 8 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
```

```
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 428 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
      428 character character

summarygrade:
high  low NAs
  75  270   83

G:
   1    2    3    4 NAs
  10  260   71    4   83

summarystage:
early  late  NAs
  166   168    94

T:
   1    2    3    4 NAs
  14   55  244   31   84

N:
   0    1    2 NAs
 187  101   59   81

M:
   0    1 NAs
 274   64   90

family_history:
   n    y NAs
 184  241    3

gender:
   f    m NAs
 208  219    1

stageall:
   1    2    3    4 NAs
```

```
  52  114  104   64   94

ethnicity:
other
  428

batch:
   Length     Class      Mode
      428 character character

uncurated_author_metadata:
   Length     Class      Mode
      428 character character
```

GSE21510_eset *Clinical significance of osteoprotegerin expression in human colorectal cancer.*

Description

This study aimed to identify a novel biomarker or a target of treatment for colorectal cancer (CRC).The expression profiles of cancer cells in 104 patients with CRC were examined using laser microdissection and oligonucleotide microarray analysis. Overexpression in CRC cells, especially in patients with distant metastases, was a prerequisite to select candidate genes. The mRNA expression of candidate genes was investigated by quantitative reverse transcriptase PCR (RT-PCR) in 77 patients as a validation study. We analyzed the protein expression and localization of the candidate gene by immunohistochemical study and investigated the relationship between protein expression and clinicopathologic features in 274 CRC patients.Using microarray analysis, we identified 6 candidate genes related to distant metastases in CRC patients. Among these genes, osteoprotegerin (OPG) is known to be associated with aggressiveness in several cancers through inhibition of apoptosis via neutralization of the function of TNF-related apoptosis-inducing ligand. The mRNA expression of OPG in cancer tissues was significantly higher in patients with distant metastases than those without metastases. Overexpression of OPG protein was associated with significantly worse overall survival and relapse-free survival. Moreover, overexpression of the OPG protein was an independent risk factor for CRC recurrence.Overexpression of OPG may be a predictive biomarker of CRC recurrence and a target for treatment of this disease.??2011 AACR.

Usage

```
data( GSE21510_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Tsukamoto S, Ishikawa T, Iida S, Ishiguro M et al.??Clinical significance of osteopr
```

```
 Laboratory: Tsukamoto, Sugihara 2010
 Contact information:
Title: Clinical significance of osteoprotegerin expression in human colorectal cancer.
 URL:
 PMIDs: 21270110

 Abstract: A 211 word abstract is available. Use abstract method.
 Information is available on: preprocessing
 notes:
  platform_title:
     [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
  platform_shorttitle:
     Affymetrix HG-U133Plus2
  platform_summary:
     hg-u133_plus_2
  platform_manufacturer:
     Affymetrix
  platform_distribution:
     commercial
  platform_accession:
     GPL570
  platform_technology:
     in situ oligonucleotide
  warnings:
     No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 148 samples
Platform type: hg-u133_plus_2
----------------------------
Available sample meta-data:
----------------------------

alt_sample_name:
   Length     Class      Mode
      148 character character

sample_type:
adjacentnormal          tumor
            25            123
```

```
summarystage:
early  late  NAs
   73    74    1

T:
 0  1  2  3  4
 1 19 54 47 27

M:
 0  1
94 54

batch:
   Length     Class      Mode
      148 character character

uncurated_author_metadata:
   Length     Class      Mode
      148 character character
```

GSE21815_eset *Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers.*

Description

The functional impact of recently discovered long noncoding RNAs (ncRNAs) in human cancer remains to be clarified. One long ncRNA which has attracted attention is the Hox transcript antisense intergenic RNA termed HOTAIR, a long ncRNA expressed from the developmental HOXC locus located on chromosome 12q13.13. In cooperation with Polycomb complex PRC2, the HOTAIR long ncRNA is reported to reprogram chromatin organization and promote breast cancer metastasis. In this study, we examined the status and function of HOTAIR in patients with stage IV colorectal cancer (CRC) who have liver metastases and a poor prognosis. HOTAIR expression levels were higher in cancerous tissues than in corresponding noncancerous tissues and high HOTAIR expression correlated tightly with the presence of liver metastasis. Moreover, patients with high HOTAIR expression had a relatively poorer prognosis. In a subset of 32 CRC specimens, gene set enrichment analysis using cDNA array data revealed a close correlation between expression of HOTAIR and members of the PRC2 complex (SUZ12, EZH2, and H3K27me3). Our findings suggest that HOTAIR expression is associated with a genome-wide reprogramming of PRC2 function not only in breast cancer but also in CRC, where upregulation of this long ncRNA may be a critical element in metastatic progression.

Usage

```
data( GSE21815_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Kogo R, Shimamura T, Mimori K, Kawahara K et al.??Long noncoding RNA HOTAIR regulate
  Laboratory: Mori M,??Mimori K,??Yokobori T 2010
  Contact information:
 Title: Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated
  URL:
  PMIDs: 21862635

  Abstract: A 201 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name vers
ion)
   platform_shorttitle:
      Agilent G4112F
   platform_summary:
      agilent-014850 whole human genome microarray 4x44k g4112f
   platform_manufacturer:
      Agilent
   platform_distribution:
      commercial
   platform_accession:
      GPL6480
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19686 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19686 features, 141 samples
Platform type: agilent-014850 whole human genome microarray 4x44k g4112f
----------------------------
Available sample meta-data:
----------------------------

alt_sample_name:
```

```
   Length     Class      Mode
      141 character character

sample_type:
adjacentnormal          tumor           NAs
             5             66             70

summarystage:
early  late  NAs
   14    52    75

T:
   1    2    3    4 NAs
   4   10   31   21   75

N:
   0    1    2 NAs
  43   15    8   75

M:
   0    1 NAs
  56   10   75

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
  32.00   57.00   65.00   64.73   74.00   85.00       4

gender:
   f    m NAs
  59   78    4

stageall:
   1    2    3    4 NAs
  13   26   17   10   75

uncurated_author_metadata:
   Length     Class      Mode
      141 character character
```

GSE24549.GPL5175_eset *Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associations with splicing factor expression levels and patient survival.*

Description

Colorectal cancer (CRC) is a heterogeneous disease that, on the molecular level, can be characterized by inherent genomic instabilities; chromosome instability and microsatellite instability. In the present study we analyze genome-wide disruption of pre-mRNA splicing, and propose transcriptome instability as a characteristic that is analogous to genomic instability on the transcriptome level.Exon microarray profiles from two independent series including a total of 160 CRCs were investigated for their relative amounts of exon usage differences. Each exon in each sample was assigned an alternative splicing score calculated by the FIRMA algorithm. Amounts of deviating exon usage per sample were derived from exons with extreme splicing scores.There was great heterogeneity within both series in terms of sample-wise amounts of deviating exon usage. This was strongly associated with the expression levels of approximately half of 280 splicing factors (54% and 48% of splicing factors were significantly correlated to deviating exon usage amounts in the two series). Samples with high or low amounts of deviating exon usage, associated with overall transcriptome instability, were almost completely separated into their respective groups by hierarchical clustering analysis of splicing factor expression levels in both sample series. Samples showing a preferential tendency towards deviating exon skipping or inclusion were associated with skewed transcriptome instability. There were significant associations between transcriptome instability and reduced patient survival in both sample series. In the test series, patients with skewed transcriptome instability showed the strongest prognostic association (P = 0.001), while a combination of the two characteristics showed the strongest association with poor survival in the validation series (P = 0.03).We have described transcriptome instability as a characteristic of CRC. This transcriptome instability has associations with splicing factor expression levels and poor patient survival.

Usage

```
data( GSE24549.GPL5175_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Sveen A,????esen TH,??Rognum TO,??Lothe RA,??Skotheim RI. Transcriptome instabilit
  Laboratory: Sveen A,????esen TH,??Rognum TO,??Lothe RA,??Skotheim RI 2011
  Contact information:
 Title: Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associat
  URL:
  PMIDs: 21619627

  Abstract: A 282 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) versio
n]
   platform_shorttitle:
      Affymetrix HuEx-1_0-st
   platform_summary:
      huex-1_0-st
```

```
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL5175
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: ABCF1 ANAPC5 ... ZNF259P1///ZNF259 (282 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 282 features, 83 samples
Platform type: huex-1_0-st
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       83 character character

summarystage:
early  late
   46    37

T:
 2  3
46 37

msi:
   n    y NAs
  69   13    1

mss:
 n  y
14 69

batch:
   Length     Class      Mode
```

```
        83 character character

uncurated_author_metadata:
   Length     Class      Mode
        83 character character
```

GSE24550.GPL5175_eset *Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associations with splicing factor expression levels and patient survival.*

Description

Colorectal cancer (CRC) is a heterogeneous disease that, on the molecular level, can be characterized by inherent genomic instabilities; chromosome instability and microsatellite instability. In the present study we analyze genome-wide disruption of pre-mRNA splicing, and propose transcriptome instability as a characteristic that is analogous to genomic instability on the transcriptome level.Exon microarray profiles from two independent series including a total of 160 CRCs were investigated for their relative amounts of exon usage differences. Each exon in each sample was assigned an alternative splicing score calculated by the FIRMA algorithm. Amounts of deviating exon usage per sample were derived from exons with extreme splicing scores.There was great heterogeneity within both series in terms of sample-wise amounts of deviating exon usage. This was strongly associated with the expression levels of approximately half of 280 splicing factors (54% and 48% of splicing factors were significantly correlated to deviating exon usage amounts in the two series). Samples with high or low amounts of deviating exon usage, associated with overall transcriptome instability, were almost completely separated into their respective groups by hierarchical clustering analysis of splicing factor expression levels in both sample series. Samples showing a preferential tendency towards deviating exon skipping or inclusion were associated with skewed transcriptome instability. There were significant associations between transcriptome instability and reduced patient survival in both sample series. In the test series, patients with skewed transcriptome instability showed the strongest prognostic association (P = 0.001), while a combination of the two characteristics showed the strongest association with poor survival in the validation series (P = 0.03).We have described transcriptome instability as a characteristic of CRC. This transcriptome instability has associations with splicing factor expression levels and poor patient survival.

Usage

```
data( GSE24550.GPL5175_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Sveen A, Agesen TH, Nesbakken A, Rognum TO et al.??Transcriptome instability in colo
  Laboratory: Sveen A,????esen TH,??Rognum TO,??Lothe RA,??Skotheim RI 2011
  Contact information:
```

```
 Title: Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associat
  URL:
  PMIDs: 21619627

  Abstract: A 282 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) versio
n]
   platform_shorttitle:
      Affymetrix HuEx-1_0-st
   platform_summary:
      huex-1_0-st
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL5175
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: ABCF1 ANAPC5 ... ZNF259P1///ZNF259 (282 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 282 features, 90 samples
Platform type: huex-1_0-st
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       90 character character

sample_type:
adjacentnormal          tumor
            13             77
```

```
summarystage:
early  late  NAs
   44    33    13

T:
   2    3 NAs
  44   33   13

msi:
   n    y NAs
  41   24   25

mss:
 n  y
49 41

batch:
   Length     Class      Mode
       90 character character

uncurated_author_metadata:
   Length     Class      Mode
       90 character character
```

GSE2630_eset *A gene signature of 8 genes could identify the risk of recurrence and progression in Dukes' B colon cancer patients.*

Description

The benefit of postoperative adjuvant chemotherapy in patients with Dukes' B colorectal cancer is still uncertain and its routine use is not recommended. The five-year relapse rate is approximately 25-40% and the identification of patients at high risk of recurrence would represent an important strategy for the use of adjuvant chemotherapy. We retrospectively analyzed gene expression profiles in frozen tumor specimens from patients with Dukes' B colorectal cancer by using high density oligonucleotide microarrays. Our results show a subset of 48 genes differentially expressed with an associated probability <0.001 in the t-test. Another statistical procedure based on the Fisher criterion resulted in 11 genes able to separate both groups. We selected the 8 genes present in both subsets. The differential expression of five genes (CHD2, RPS5, ZNF148, BRI3 and MGC23401) in colon cancer progression was confirmed by real-time PCR in an independent set of patients of Dukes' B and C stages.

Usage

```
data( GSE2630_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Bandr??E,??Malumbres R,??Cubedo E,??Sola J,??Garc??Foncillas J,??Labarga A. A gene
 Laboratory: Bandr??E,??Malumbres R,??Cubedo E,??Sola J,??Garc??Foncillas J,??Labarga A 2005
  Contact information:
 Title: A gene signature of 8 genes could identify the risk of recurrence and progression in Dukes B colo
  URL:
  PMIDs: 17390049

  Abstract: A 151 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      Human 19K oligo array
   platform_shorttitle:
      NA
   platform_summary:
      human 19k oligo array
   platform_manufacturer:
      NA
   platform_distribution:
      non-commercial
   platform_accession:
      GPL2006
   platform_technology:
      spotted oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: 1 10 ... 9 (48 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 48 features, 16 samples
Platform type: human 19k oligo array
----------------------------
Available sample meta-data:
----------------------------

alt_sample_name:
   Length     Class      Mode
```

```
       16 character character

T:
 2  3
 6 10

N:
 0
16

M:
 0
16

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  42.00   59.50   67.00   64.38   72.00   81.00

recurrence_status:
norecurrence   recurrence
          10            6

summarylocation:
   l    r NAs
   2    4   10

gender:
 f  m
 5 11

uncurated_author_metadata:
   Length     Class      Mode
       16 character character
```

GSE26682.GPL570_eset *MRE11 deficiency increases sensitivity to poly(ADP-ribose) polymerase inhibition in microsatellite unstable colorectal cancers.*

Description

Microsatellite instability (MSI) is displayed by approximately 15% of colorectal cancers (CRC). Defective DNA mismatch repair generates mutations at repetitive DNA sequences such as those located in the double strand break (DSB) repair gene MRE11. We assessed the mutational status of MRE11 in a panel of 17 CRC cell lines and 46 primary tumors and found a strong correlation with MSI status in both cell lines and tumors. Therefore, we hypothesized that deficiency in MRE11 may sensitize CRC cells to poly(ADP-ribose) polymerase (PARP-1) inhibition based on the concept of

synthetic lethality. We further assessed the activity of the PARP-1 inhibitor, ABT-888, in CRC cell lines and observed preferential cytotoxicity in those MSI cell lines harboring mutations in MRE11 compared with both wild-type cell lines and microsatellite stable (MSS) cell lines. A significant correlation between MRE11 expression levels and cytotoxicity to ABT-888 at 10 ??M was observed (R?? = 0.915, P < 0.001). Using two experimental approaches, including short hairpin RNA knocking down MRE11 in the wild-type and MSS cell line SW-480 and a second cell line model transfected with mutant MRE11, we experimentally tried to confirm the role of MRE11 in conferring sensitivity to PARP-1 inhibition. Both models led to changes in proliferation in response to ABT-888 at different concentrations, and a drug-response effect was not observed, suggesting a possible contribution of additional genes. We conclude that MSI colorectal tumors deficient in DSB repair secondary to mutation in MRE11 show a higher sensitivity to PARP-1 inhibition. Further clinical investigation of PARP-1 inhibitors is warranted in MSI CRCs.

Usage

```
data( GSE26682.GPL570_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Vilar E,??Gruber SB,??Rennert G,??Bartnik CM,??Stenzel SL,??Iniesta MD,??Raskin L,
  Laboratory: Vilar E,??Morgan MA 2011
  Contact information:
 Title: MRE11 deficiency increases sensitivity to poly(ADP-ribose) polymerase inhibition in microsatel
  URL:
  PMIDs: 21300766

  Abstract: A 257 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
```

```
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1CF A2M ... ZZZ3 (11838 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 11838 features, 176 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
      176 character character

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  45.00   66.00   73.00   71.75   78.00   92.00

msi:
   n    y NAs
 119   41   16

gender:
 f  m
83 93

mss:
   n    y NAs
  41  119   16

batch:
   Length     Class      Mode
      176 character character

uncurated_author_metadata:
   Length     Class      Mode
      176 character character
```

GSE26682.GPL96_eset *MRE11 deficiency increases sensitivity to poly(ADP-ribose) polymerase inhibition in microsatellite unstable colorectal cancers.*

Description

Microsatellite instability (MSI) is displayed by approximately 15% of colorectal cancers (CRC). Defective DNA mismatch repair generates mutations at repetitive DNA sequences such as those located in the double strand break (DSB) repair gene MRE11. We assessed the mutational status of MRE11 in a panel of 17 CRC cell lines and 46 primary tumors and found a strong correlation with MSI status in both cell lines and tumors. Therefore, we hypothesized that deficiency in MRE11 may sensitize CRC cells to poly(ADP-ribose) polymerase (PARP-1) inhibition based on the concept of synthetic lethality. We further assessed the activity of the PARP-1 inhibitor, ABT-888, in CRC cell lines and observed preferential cytotoxicity in those MSI cell lines harboring mutations in MRE11 compared with both wild-type cell lines and microsatellite stable (MSS) cell lines. A significant correlation between MRE11 expression levels and cytotoxicity to ABT-888 at 10 ??M was observed (R?? = 0.915, P < 0.001). Using two experimental approaches, including short hairpin RNA knocking down MRE11 in the wild-type and MSS cell line SW-480 and a second cell line model transfected with mutant MRE11, we experimentally tried to confirm the role of MRE11 in conferring sensitivity to PARP-1 inhibition. Both models led to changes in proliferation in response to ABT-888 at different concentrations, and a drug-response effect was not observed, suggesting a possible contribution of additional genes. We conclude that MSI colorectal tumors deficient in DSB repair secondary to mutation in MRE11 show a higher sensitivity to PARP-1 inhibition. Further clinical investigation of PARP-1 inhibitors is warranted in MSI CRCs.

Usage

```
data( GSE26682.GPL96_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Vilar E,??Gruber SB,??Rennert G,??Bartnik CM,??Stenzel SL,??Iniesta MD,??Raskin L,
  Laboratory: Vilar E,??Morgan MA 2011
  Contact information:
 Title: MRE11 deficiency increases sensitivity to poly(ADP-ribose) polymerase inhibition in microsatel
  URL:
  PMIDs: 21300766

  Abstract: A 257 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133A] Affymetrix Human Genome U133A Array
   platform_shorttitle:
      Affymetrix HG-U133A
   platform_summary:
      hg-u133a
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
```

```
   platform_accession:
      GPL96
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1CF A2M ... ZZZ3 (12986 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 12986 features, 155 samples
Platform type: hg-u133a
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
      155 character character

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  21.00   66.50   75.00   72.61   80.50   94.00

msi:
   n    y NAs
  99   41   15

gender:
 f  m
69 86

mss:
   n    y NAs
  41   99   15

batch:
   Length     Class      Mode
      155 character character

uncurated_author_metadata:
   Length     Class      Mode
```

```
155 character character
```

GSE26906_eset *Expression Profiles in Stage II Colon Cancer According to APC Gene Status.*

Description

Colorectal cancer is one of the most common cancers in the world. Histoclinical staging is efficient, but combination with molecular markers may improve the classification of stage II cancers. Several tumor-suppressor genes have been associated with colorectal cancer, and the most frequent allelic losses have been extensively studied for their prognosis effect, but the results remain controversial. In a previous study, we found a possible influence of the chromosome 5 status in the development of liver metastases in stage II colon cancers. We have here investigated the role of the APC gene, located in chromosome arm 5q, in a series of 183 colon adenocarcinomas through a combined analysis of gene expression, mutation, allelic loss and promoter methylation, and metastasis occurrence. Point mutations were found in 73% of cases and allelic losses were found in 39%; 59% of tumors presented with a biallelic inactivation, with a very strong interdependence of the two APC hits (P = 2.1 x 10(-9)). No association was found between expression, number and type of APC alterations, and metastatic evolution. Our results show that the determination of APC status cannot help in the prediction of metastasis and cannot be used to subclassify stage II colon cancers.

Usage

```
data( GSE26906_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Birnbaum DJ, Laibe S, Ferrari A, Lagarde A et al.??Expression Profiles in Stage II C
  Laboratory: Olschwang S 2011
  Contact information:
 Title: Expression Profiles in Stage II Colon Cancer According to APC Gene Status.
  URL:
  PMIDs: 22496922

  Abstract: A 199 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
```

```
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 90 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       90 character character

sample_type:
tumor
   90

summarystage:
early
   90

T:
 2
90

M:
 0  1
69 21

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

```
  31.00   58.50   69.50   66.52   75.00   94.00

summarylocation:
 l  r
65 25

gender:
 f  m
47 43

mutation_apc:
 n  y
22 68

batch:
   Length     Class      Mode
       90 character character

uncurated_author_metadata:
   Length     Class      Mode
       90 character character
```

GSE27544_eset *Genome-wide profiling characterizes CRCs with genetic instability and specific routes to HLA class I loss and immunoescape*

Description

Bernal M,??Garc??Alcalde F,??Concha ????Blanco A,??Garrido F,??Ruiz-Cabello F

Usage

```
data( GSE27544_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Bernal M,??Garc??Alcalde F,??Concha ????Blanco A,??Garrido F,??Ruiz-Cabello F
 Laboratory: Bernal M,??Garc??Alcalde F,??Concha ????Blanco A,??Garrido F,??Ruiz-Cabello F 2011
  Contact information:
 Title: Genome-wide profiling characterizes CRCs with genetic instability and specific routes to HLA cl
  URL:
  PMIDs: PMID unknown

  Abstract: A 7 word abstract is available. Use abstract method.
```

```
  Information is available on: preprocessing
  notes:
   platform_title:
      [HT_HG-U133_Plus_PM] Affymetrix HT HG-U133+ PM Array Plate
   platform_shorttitle:
      Affymetrix HT HG-U133+ PM
   platform_summary:
      ht_hg-u133_plus_pm
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL13158
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1CF ... ZZZ3 (20741 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 20741 features, 22 samples
Platform type: ht_hg-u133_plus_pm
----------------------------
Available sample meta-data:
----------------------------

alt_sample_name:
   Length     Class      Mode
       22 character character

sample_type:
tumor
   22

msi:
 n  y
14  8

mss:
 n  y
```

```
 8 14

batch:
2009-10-07
        22

uncurated_author_metadata:
   Length     Class      Mode
       22 character character
```

GSE28702_eset *Potential responders to FOLFOX therapy for colorectal cancer by Random Forests analysis.*

Description

Molecular characterisation using gene-expression profiling will undoubtedly improve the prediction of treatment responses, and ultimately, the clinical outcome of cancer patients.To establish the procedures to identify responders to FOLFOX therapy, 83 colorectal cancer (CRC) patients including 42 responders and 41 non-responders were divided into training (54 patients) and test (29 patients) sets. Using Random Forests (RF) algorithm in the training set, predictor genes for FOLFOX therapy were identified, which were applied to test samples and sensitivity, specificity, and out-of-bag classification accuracy were calculated.In the training set, 22 of 27 responders (81.4% sensitivity) and 23 of 27 non-responders (85.1% specificity) were correctly classified. To improve the prediction model, we removed the outliers determined by RF, and the model could correctly classify 21 of 23 responders (91.3%) and 22 of 23 non-responders (95.6%) in the training set, and 80.0% sensitivity and 92.8% specificity, with an accuracy of 69.2% in 29 independent test samples.Random Forests on gene-expression data for CRC patients was effectively able to stratify responders to FOLFOX therapy with high accuracy, and use of pharmacogenomics in anticancer therapy is the first step in planning personalised therapy.

Usage

```
data( GSE28702_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Tsuji S, Midorikawa Y, Takahashi T, Yagi K et al.??Potential responders to FOLFOX th
  Laboratory: Tsuji 2011
  Contact information:
 Title: Potential responders to FOLFOX therapy for colorectal cancer by Random Forests analysis.
  URL:
  PMIDs: 22095227
```

```
  Abstract: A 185 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 59 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       59 character character

sample_type:
metastatic      tumor
         3         56

gender:
 f  m
20 39

batch:
```

```
   Length     Class      Mode
       59 character character

drug_name:
mfolfox6
      59

drug_response:
 n  y
25 34

uncurated_author_metadata:
   Length     Class      Mode
       59 character character
```

GSE3294_eset *Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa.*

Description

The causative molecular pathways underlying the pathogenesis of colorectal cancer (CRC) need to be better characterized. The purpose of our study was to better understand the genetic mechanism of oncogenesis for human colorectal cancer and to identify new potential tumor markers of use in clinical practice. We used cDNA microarrays to compare gene expression profiles of colorectal biopsies from 25 CRC patients and 13 normal mucosa from adjacent non-cancerous tissues. Findings were validated by real-time PCR; in addition, western blotting and immunochemistry analysis were carried out as further confirmation of differential expression at a protein level. Comparing cancerous tissues with normal colonic mucosa we identified 584 known genes differentially expressed to a significant degree ($p<0.001$). Many of the transcripts that were more abundant in tumors than in non-neoplastic tissues appear to reflect important events for colon carcinogenesis. For example, a significant number of these genes serve as apoptotic inhibitors (e.g. BFAR, BIRC1, BIRC6). Furthermore, we observed the simultaneous up-regulation of HLA-E and the down-regulation of beta2-microglobulin; these genes strongly support a potential tumor escape strategy from immune surveillance in colon cancer tissues. Our study provides new gene candidates in the pathogenesis of human CRC disease. From our results we hypothesize that CRC cells escape immune surveillance through a specific gene expression alteration; moreover, over-expression of several survival genes seems to confer a more anti-apoptotic phenotype. These genes are involved in pathways not previously implicated in CRC pathogenesis and they may provide new targets for therapy.

Usage

```
data( GSE3294_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Bianchini M, Levy E, Zucchini C, Pinski V et al.??Comparative study of gene expressi
  Laboratory: Bianchini 2005
  Contact information:
 Title: Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and n
  URL:
  PMIDs: 16773188

  Abstract: A 245 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      UHN SS-Human 19Kv7
   platform_shorttitle:
      UHN SS-Human 19Kv7
   platform_summary:
      uhn ss-human 19kv7
   platform_manufacturer:
      UHN
   platform_distribution:
      commercial
   platform_accession:
      GPL2829
   platform_technology:
      spotted DNA/cDNA
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: AA001103 AA001104 ... Z45302 (15437 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 15437 features, 24 samples
Platform type: uhn ss-human 19kv7
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
```

```
       24 character character

summarystage:
early  late
    4    20

T:
 1  2  3
 3  1 20

N:
 0  1  2
10  9  5

M:
 0
24

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  35.00   67.25   76.50   71.46   78.75   89.00

gender:
 f  m
 8 16

stageall:
 2  3
 6 18

uncurated_author_metadata:
   Length     Class      Mode
       24 character character
```

GSE33113_eset *Mutations in the Ras-Raf Axis underlie the prognostic value of CD133 in colorectal cancer.*

Description

High expression of cancer stem cell (CSC) marker CD133 has been used as a predictor for prognosis in colorectal cancer (CRC), suggesting that enumeration of CSCs, using CD133, is predictive for disease progression. However, we showed recently that both CD133 mRNA and protein are not downregulated during differentiation of colon CSCs, pointing to an alternative reason for the prognostic value of CD133. We therefore set out to delineate the relation between CD133 expression and prognosis.A CRC patient series was studied for expression of CD133 and other CSC markers

by microarray and quantitative PCR analysis. In addition, several common mutations were analyzed to determine the relation with CD133 expression.CD133 mRNA expression predicted relapse-free survival in our patient series, whereas several other CSC markers could not. Moreover, no correlation was found between expression of other CSC markers and CD133. Interestingly, high CD133 expression was related to mutations in K-Ras and B-Raf, and inhibition of mutant K-Ras or downstream mitogen-activated protein kinase kinase (MEK) signaling decreases CD133 expression. In addition, an activated K-Ras gene expression signature could predict CD133 expression in our patient set as well as data sets of other tumor types.CD133 expression is upregulated in CRC tumors that have a hyperactivated Ras-Raf-MEK-ERK pathway and is therefore related to mutations in K-Ras or B-Raf. As mutations in either gene have been related to poor prognosis, we conclude that CD133 expression is not indicative for CSC numbers but rather related to the mutation or activity status of the Ras-Raf pathway.

Usage

```
data( GSE33113_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: de Sousa E Melo F, Colak S, Buikhuisen J, Koster J et al.??Methylation of cancer-ste
  Laboratory: Medema JP,??Tanis PJ 2011
  Contact information:
 Title: Mutations in the Ras-Raf Axis underlie the prognostic value of CD133 in colorectal cancer.
  URL:
  PMIDs: 22496204

  Abstract: A 247 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
```

```
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 96 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       96 character character

sample_type:
adjacentnormal          tumor
             6             90

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
  35.00   60.00   72.50   69.73   79.00   95.00       2

gender:
   f    m NAs
  50   44    2

stageall:
 2
96

days_to_recurrence_or_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
   50.0   508.8  1180.0  1234.0  1633.0  3599.0       2

batch:
   Length     Class      Mode
       96 character character

uncurated_author_metadata:
   Length     Class      Mode
       96 character character
```

GSE39582_eset *Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value.*

Description

Colon cancer (CC) pathological staging fails to accurately predict recurrence, and to date, no gene expression signature has proven reliable for prognosis stratification in clinical practice, perhaps because CC is a heterogeneous disease. The aim of this study was to establish a comprehensive molecular classification of CC based on mRNA expression profile analyses.Fresh-frozen primary tumor samples from a large multicenter cohort of 750 patients with stage I to IV CC who underwent surgery between 1987 and 2007 in seven centers were characterized for common DNA alterations, including BRAF, KRAS, and TP53 mutations, CpG island methylator phenotype, mismatch repair status, and chromosomal instability status, and were screened with whole genome and transcriptome arrays. 566 samples fulfilled RNA quality requirements. Unsupervised consensus hierarchical clustering applied to gene expression data from a discovery subset of 443 CC samples identified six molecular subtypes. These subtypes were associated with distinct clinicopathological characteristics, molecular alterations, specific enrichments of supervised gene expression signatures (stem cell phenotype-like, normal-like, serrated CC phenotype-like), and deregulated signaling pathways. Based on their main biological characteristics, we distinguished a deficient mismatch repair subtype, a KRAS mutant subtype, a cancer stem cell subtype, and three chromosomal instability subtypes, including one associated with down-regulated immune pathways, one with up-regulation of the Wnt pathway, and one displaying a normal-like gene expression profile. The classification was validated in the remaining 123 samples plus an independent set of 1,058 CC samples, including eight public datasets. Furthermore, prognosis was analyzed in the subset of stage II-III CC samples. The subtypes C4 and C6, but not the subtypes C1, C2, C3, and C5, were independently associated with shorter relapse-free survival, even after adjusting for age, sex, stage, and the emerging prognostic classifier Oncotype DX Colon Cancer Assay recurrence score (hazard ratio 1.5, 95% CI 1.1-2.1, p???=???0.0097). However, a limitation of this study is that information on tumor grade and number of nodes examined was not available.We describe the first, to our knowledge, robust transcriptome-based classification of CC that improves the current disease stratification based on clinicopathological variables and common DNA markers. The biological relevance of these subtypes is illustrated by significant differences in prognosis. This analysis provides possibilities for improving prognostic models and therapeutic strategies. In conclusion, we report a new classification of CC into six molecular subtypes that arise through distinct biological pathways.

Usage

```
data( GSE39582_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Marisa L, de Reyni??A, Duval A, Selves J et al.??Gene expression classification of c
  Laboratory: Marisa, Boige 2012
  Contact information:
```

```
 Title: Gene expression classification of colon cancer into molecular subtypes: characterization, valic
  URL:
  PMIDs: 23700391

  Abstract: A 384 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 566 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
      566 character character

T:
  0   1   2   3   4
  4  33 264 205  60

age_at_initial_pathologic_diagnosis:
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
   22.0    59.0    68.0    66.9    76.0    97.0       1

location:
  distal proximal
     342      224

gender:
  f   m
256 310

stageall:
  2
566

batch:
   Length     Class      Mode
      566 character character

uncurated_author_metadata:
   Length     Class      Mode
      566 character character
```

GSE3964_eset *Deciphering cellular states of innate tumor drug responses.*

Description

The molecular mechanisms underlying innate tumor drug resistance, a major obstacle to successful cancer therapy, remain poorly understood. In colorectal cancer (CRC), molecular studies have focused on drug-selected tumor cell lines or individual candidate genes using samples derived from patients already treated with drugs, so that very little data are available prior to drug treatment.Transcriptional profiles of clinical samples collected from CRC patients prior to their exposure to a combined chemotherapy of folinic acid, 5-fluorouracil and irinotecan were established using microarrays. Vigilant experimental design, power simulations and robust statistics were used to restrain the rates of false negative and false positive hybridizations, allowing successful discrimination between drug resistance and sensitivity states with restricted sampling. A list of 679 genes was established that intrinsically differentiates, for the first time prior to drug exposure, subsequently diagnosed chemo-sensitive and resistant patients. Independent biological validation performed through quantitative PCR confirmed the expression pattern on two additional patients. Careful annotation of interconnected functional networks provided a unique representation of the cellular states underlying drug responses.Molecular interaction networks are described that provide a solid foundation on which to anchor working hypotheses about mechanisms underlying in vivo innate tumor drug responses. These broad-spectrum cellular signatures represent a starting point from which by-pass chemotherapy schemes, targeting simultaneously several of the molecular mechanisms involved, may be developed for critical therapeutic intervention in CRC patients. The demonstrated power of

this research strategy makes it generally applicable to other physiological and pathological situations.

Usage

```
data( GSE3964_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Graudens E, Boulanger V, Mollard C, Mariage-Samson R et al.??Deciphering cellular s
  Laboratory: Graudens, Imbeaud 2006
  Contact information:
  Title: Deciphering cellular states of innate tumor drug responses.
  URL:
  PMIDs: 16542501

  Abstract: A 242 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      11K_VJF-ARRAY
   platform_shorttitle:
      NA
   platform_summary:
      11k_vjf-array
   platform_manufacturer:
      Array s/IMAGE - Genexpress
   platform_distribution:
      non-commercial
   platform_accession:
      GPL3282
   platform_technology:
      spotted DNA/cDNA
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: 384D8-2 38600 ... ZYX (5845 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 5845 features, 15 samples
```

```
Platform type: 11k_vjf-array
----------------------------
Available sample meta-data:
----------------------------

alt_sample_name:
   Length     Class      Mode
       15 character character

sample_type:
adjacentnormal          tumor
             6              9

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  49.00   60.50   65.00   64.27   68.00   71.00

gender:
f m
7 8

stageall:
 4
15

uncurated_author_metadata:
   Length     Class      Mode
       15 character character
```

GSE4045_eset *Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis.*

Description

Serrated colorectal carcinomas (CRCs) are morphologically different from conventional CRCs and have been proposed to follow a distinct pathway of CRC formation. Despite studies of single molecular events in this tumor type, the diagnosis of serrated CRC relies on morphology and the putative unique biological character of these tumors has not been established. Here we show that the gene expression profiling of 37 CRCs separated serrated and conventional CRCs into two distinct branches in unsupervised hierarchical clustering (P-value 7.8 x 10(-7)), and revealed 201 differentially expressed genes representing potential biomarkers for serrated CRC. Immunohistochemistry was utilized to verify the key findings in the 37 CRCs examined by expression profiling, and a separate validation set of 37 serrated and 86 conventional CRCs was examined to evaluate the candidate biomarkers in an extended sample material. Ephrin receptor B2, hypoxia-inducible factor 1-alpha

and patched appeared as proteins important for genesis of serrated CRC. This study establishes serrated CRCs as a biologically distinct subclass of CRC and represents a step forward in the molecular classification of these cancers. The study also provides a platform to understand the molecular basis of serrated CRC and in long term may contribute to the development of specific treatment options for this tumor type.

Usage

```
data( GSE4045_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Laiho P, Kokko A, Vanharanta S, Salovaara R et al.??Serrated carcinomas form a subcl
  Laboratory: Laiho, Aaltonen 2007
  Contact information:
 Title: Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis.
  URL:
  PMIDs: 16819509

  Abstract: A 205 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133A] Affymetrix Human Genome U133A Array
   platform_shorttitle:
      Affymetrix HG-U133A
   platform_summary:
      hg-u133a
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL96
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1CF A2M ... ZZZ3 (12986 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 12986 features, 37 samples
Platform type: hg-u133a
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       37 character character

summarygrade:
 low NAs
  36    1

G:
   1    2    3 NAs
   4   28    4    1

family_history:
   n    y NAs
  32    3    2

msi:
 n  y
36  1

location:
  distal proximal
      22       15

gender:
 f  m
19 18

mss:
 n  y
 1 36

stageall:
 2  3  4
 2 33  2

batch:
   Length     Class      Mode
       37 character character

uncurated_author_metadata:
```

```
Length     Class      Mode
    37 character character
```

GSE4526_eset *Gene expression signature for recurrence in stage III colorectal cancers.*

Description

Colorectal cancer patients with lymph node metastases (stage III) show poorer prognosis than those without. Predicting development of recurrence may guide the need for intensive follow-up and/or adjuvant chemotherapy in such patients. The authors' objective was to identify a set of discriminating genes that could predict recurrence in stage III colorectal cancer.Thirty-six stage III colorectal cancer patients were studied. Tumor samples were obtained from surgically resected specimens. Thirteen patients developed recurrence, whereas 23 patients did not. Gene expression profiles were determined using human HG-U133 Plus 2.0 Gene Chip (Affymetrix, Santa Clara, Calif).The authors identified 45 discriminating genes between patients with and without recurrence. By using this gene set, they established a new model to predict recurrence with an accuracy of 90.9%. The discriminating genes included calcineurin-binding protein 1 (CABIN1), whose expression differed remarkably between patients with and without recurrence (P=.0073). The authors further examined the DNA copy number of CABIN1 and were able to show a significant relation with recurrence (P<.012). Patients having CABIN1 gene loss demonstrated a higher risk of recurrence (odds ratio, 18.8). DNA copy number of CABIN1 alone could predict recurrence with an accuracy of 80.0%.The results of the current study demonstrated that gene expression profiling is useful in predicting recurrence in stage III colorectal cancer. The authors identified CABIN1 among discriminating genes that may play a key role in the development of recurrence. These results may help to establish an individualized therapy for stage III colorectal cancer.Copyright (c) 2009 American Cancer Society.

Usage

```
data( GSE4526_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Watanabe T, Kobunai T, Sakamoto E, Yamamoto Y et al.??Gene expression signature for
  Laboratory: Watanabe T,??Kobunai T,??Toda E,??Oka T 2006
  Contact information:
 Title: Gene expression signature for recurrence in stage III colorectal cancers.
  URL:
  PMIDs: 19016304

  Abstract: A 246 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
```

```
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 36 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       36 character character

recurrence_status:
norecurrence   recurrence
          23           13

stageall:
 3
36

uncurated_author_metadata:
   Length     Class      Mode
       36 character character
```

GSE45270_eset *AMC tubular and serrated adenomas*

Description

Profiling project of a panel of tubular adenoma and serrated adenoma patient material collected in the Academic Medical Center

Usage

```
data( GSE45270_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Profiling project of a panel of tubular adenoma and serrated adenoma patient materia
 Laboratory: Medema JP,??de Sousa E Melo F,??Vermeulen L,??Jansen M,??Dekker E,??Van Noesel C,??Fessler
  Contact information:
  Title: AMC tubular and serrated adenomas
  URL:
  PMIDs: PMID unknown

  Abstract: A 19 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
   platform_shorttitle:
      Affymetrix HG-U133Plus2
   platform_summary:
      hg-u133_plus_2
   platform_manufacturer:
      Affymetrix
   platform_distribution:
      commercial
   platform_accession:
      GPL570
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: frma
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: A1BG A1BG-AS1 ... ZZZ3 (19320 total)
```

```
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 19320 features, 13 samples
Platform type: hg-u133_plus_2
---------------------------
Available sample meta-data:
---------------------------

alt_sample_name:
   Length     Class      Mode
       13 character character

batch:
2012-08-23
        13

uncurated_author_metadata:
   Length     Class      Mode
       13 character character
```

`TCGA.COAD_eset` *Comprehensive molecular characterization of human colon and rectal cancer.*

Description

To characterize somatic alterations in colorectal carcinoma, we conducted a genome-scale analysis of 276 samples, analysing exome sequence, DNA copy number, promoter methylation and messenger RNA and microRNA expression. A subset of these samples (97) underwent low-depth-of-coverage whole-genome sequencing. In total, 16% of colorectal carcinomas were found to be hypermutated: three-quarters of these had the expected high microsatellite instability, usually with hypermethylation and MLH1 silencing, and one-quarter had somatic mismatch-repair gene and polymerase ?? (POLE) mutations. Excluding the hypermutated cancers, colon and rectum cancers were found to have considerably similar patterns of genomic alteration. Twenty-four genes were significantly mutated, and in addition to the expected APC, TP53, SMAD4, PIK3CA and KRAS mutations, we found frequent mutations in ARID1A, SOX9 and FAM123B. Recurrent copy-number alterations include potentially drug-targetable amplifications of ERBB2 and newly discovered amplification of IGF2. Recurrent chromosomal translocations include the fusion of NAV2 and WNT pathway member TCF7L1. Integrative analyses suggest new markers for aggressive colorectal carcinoma and an important role for MYC-directed transcriptional activation and repression.

Usage

```
data( TCGA.COAD_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Comprehensive molecular characterization of human colon and rectal cancer. Nature 2
  Laboratory: The Cancer Genome Atlas Network 2012
  Contact information:
 Title: Comprehensive molecular characterization of human colon and rectal cancer.
  URL:
  PMIDs: 22810696

  Abstract: A 168 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      Agilent 244K Custom Gene Expression G4502A-07-3
   platform_shorttitle:
      Agilent G4502A-07-3
   platform_summary:
      agilent-014850 whole human genome microarray 4x44k g4112f
   platform_manufacturer:
      Agilent
   platform_distribution:
      commercial
   platform_accession:
      NA
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: 15E1.2 2-PDE ... ZZZ3 (17814 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 17814 features, 130 samples
Platform type: agilent-014850 whole human genome microarray 4x44k g4112f
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)
```

```
   124 observations deleted due to missingness
records   n.max n.start  events  median 0.95LCL 0.95UCL
   6.00    6.00    6.00    5.00    1.37    1.16      NA

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

unique_patient_ID:
   Length     Class      Mode
      130 character character

alt_sample_name:
   Length     Class      Mode
      130 character character

sample_type:
adjacentnormal          tumor
             7            123

primarysite:
  co NAs
 129    1

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  36.00   65.00   72.00   70.89   80.00   90.00

days_to_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
   29.0   442.0   500.0   494.5   681.0   774.0     124

vital_status:
deceased   living     NAs
       5       15      110

msi:
   n    y NAs
   4    1  125

location:
   Length     Class      Mode
      130 character character

gender:
 f  m
67 63
```

```
kras:
   n    y NAs
   1    2  127

mss:
   n    y NAs
   1    4  125

uncurated_author_metadata:
   Length     Class      Mode
      130 character character

fu:
   n NAs
  31   99

bevacizumab:
   n NAs
  31   99

irinotecan:
   n NAs
  31   99

capecitabine:
   n    y NAs
  25    6   99

cpt11:
   n NAs
  31   99

dexamethasone:
   n NAs
  31   99

erbitux:
   n NAs
  31   99

gcsf:
   n NAs
  31   99

fudr:
   n NAs
  31   99
```

```
folfiri:
   n NAs
  31   99

folfox:
   n NAs
  31   99

leucovorin:
   n NAs
  31   99

mitomycin:
   n    y NAs
  30    1   99

platin:
   n NAs
  31   99

panitumumab:
   n NAs
  31   99

pegfilgrastim:
   n NAs
  31   99

raltitrexed:
   n NAs
  31   99

xeloda:
   n NAs
  31   99

ancillary:
   n    y NAs
  30    1   99

chemotherapy:
   y NAs
  31   99

moltherapy:
   n    y NAs
  21   10   99
```

TCGA.READ_eset *Comprehensive molecular characterization of human colon and rectal cancer.*

Description

To characterize somatic alterations in colorectal carcinoma, we conducted a genome-scale analysis of 276 samples, analysing exome sequence, DNA copy number, promoter methylation and messenger RNA and microRNA expression. A subset of these samples (97) underwent low-depth-of-coverage whole-genome sequencing. In total, 16% of colorectal carcinomas were found to be hypermutated: three-quarters of these had the expected high microsatellite instability, usually with hypermethylation and MLH1 silencing, and one-quarter had somatic mismatch-repair gene and polymerase ?? (POLE) mutations. Excluding the hypermutated cancers, colon and rectum cancers were found to have considerably similar patterns of genomic alteration. Twenty-four genes were significantly mutated, and in addition to the expected APC, TP53, SMAD4, PIK3CA and KRAS mutations, we found frequent mutations in ARID1A, SOX9 and FAM123B. Recurrent copy-number alterations include potentially drug-targetable amplifications of ERBB2 and newly discovered amplification of IGF2. Recurrent chromosomal translocations include the fusion of NAV2 and WNT pathway member TCF7L1. Integrative analyses suggest new markers for aggressive colorectal carcinoma and an important role for MYC-directed transcriptional activation and repression.

Usage

```
data( TCGA.READ_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Comprehensive molecular characterization of human colon and rectal cancer. Nature 2
  Laboratory: The Cancer Genome Atlas Network 2012
  Contact information:
 Title: Comprehensive molecular characterization of human colon and rectal cancer.
  URL:
  PMIDs: 22810696

  Abstract: A 168 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      Agilent 244K Custom Gene Expression G4502A-07-3
   platform_shorttitle:
      Agilent G4502A-07-3
   platform_summary:
      agilent-014850 whole human genome microarray 4x44k g4112f
```

```
   platform_manufacturer:
      Agilent
   platform_distribution:
      commercial
   platform_accession:
      NA
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: 15E1.2 2-PDE ... ZZZ3 (17814 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 17814 features, 51 samples
Platform type: agilent-014850 whole human genome microarray 4x44k g4112f
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

   50 observations deleted due to missingness
records   n.max n.start  events  median 0.95LCL 0.95UCL
  1.000   1.000   1.000   1.000   0.866      NA      NA

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

unique_patient_ID:
   Length     Class      Mode
       51 character character

alt_sample_name:
   Length     Class      Mode
       51 character character

sample_type:
tumor
   51

primarysite:
  re NAs
  50    1
```

```
age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  41.00   62.00   67.00   67.67   73.50   89.00

days_to_death:
 316 NAs
   1   50

vital_status:
deceased   living     NAs
       1        4       46

msi:
   n NAs
   1   50

location:
Rectosigmoid Junction                rectum                   NAs
                    3                    47                     1

gender:
 f  m
24 27

kras:
   n    y NAs
   2    2   47

mss:
   y NAs
   1   50

uncurated_author_metadata:
   Length     Class      Mode
       51 character character

fu:
   n NAs
  16   35

bevacizumab:
   n NAs
  16   35

irinotecan:
   n NAs
  16   35
```

```
capecitabine:
   n    y NAs
  15    1   35

cpt11:
   n NAs
  16   35

dexamethasone:
   n NAs
  16   35

erbitux:
   n NAs
  16   35

gcsf:
   n NAs
  16   35

fudr:
   n NAs
  16   35

folfiri:
   n NAs
  16   35

folfox:
   n NAs
  16   35

leucovorin:
   n NAs
  16   35

mitomycin:
   n NAs
  16   35

platin:
   n NAs
  16   35

panitumumab:
   n NAs
  16   35
```

```
pegfilgrastim:
   n NAs
  16   35

raltitrexed:
   n NAs
  16   35

xeloda:
   n NAs
  16   35

ancillary:
   n NAs
  17   34

chemotherapy:
   y NAs
  17   34

moltherapy:
   n    y NAs
  15    2   34
```

`TCGA.RNASeqV2.READ_eset` *Comprehensive molecular characterization of human colon and rectal cancer.*

Description

To characterize somatic alterations in colorectal carcinoma, we conducted a genome-scale analysis of 276 samples, analysing exome sequence, DNA copy number, promoter methylation and messenger RNA and microRNA expression. A subset of these samples (97) underwent low-depth-of-coverage whole-genome sequencing. In total, 16% of colorectal carcinomas were found to be hypermutated: three-quarters of these had the expected high microsatellite instability, usually with hypermethylation and MLH1 silencing, and one-quarter had somatic mismatch-repair gene and polymerase ?? (POLE) mutations. Excluding the hypermutated cancers, colon and rectum cancers were found to have considerably similar patterns of genomic alteration. Twenty-four genes were significantly mutated, and in addition to the expected APC, TP53, SMAD4, PIK3CA and KRAS mutations, we found frequent mutations in ARID1A, SOX9 and FAM123B. Recurrent copy-number alterations include potentially drug-targetable amplifications of ERBB2 and newly discovered amplification of IGF2. Recurrent chromosomal translocations include the fusion of NAV2 and WNT pathway member TCF7L1. Integrative analyses suggest new markers for aggressive colorectal carcinoma and an important role for MYC-directed transcriptional activation and repression.

Usage

```
data( TCGA.RNASeqV2.READ_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Comprehensive molecular characterization of human colon and rectal cancer. Nature 2
  Laboratory: The Cancer Genome Atlas Network 2012
  Contact information:
 Title: Comprehensive molecular characterization of human colon and rectal cancer.
  URL:
  PMIDs: 22810696

  Abstract: A 168 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [RNASeqV2] Illumina HiSeq RNA sequencing
   platform_shorttitle:

   platform_summary:
      NA
   platform_manufacturer:
      Illumina
   platform_distribution:
      sequencing
   platform_accession:
      NA
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: ? A1BG ... ZZZ3 (20502 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 20502 features, 6 samples
Platform type: NA
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)
```

```
   3 observations deleted due to missingness
records   n.max n.start  events  median 0.95LCL 0.95UCL
   3.00    3.00    3.00    3.00    3.44    2.72      NA

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

unique_patient_ID:
   Length     Class      Mode
        6 character character

alt_sample_name:
   Length     Class      Mode
        6 character character

sample_type:
tumor
    6

primarysite:
re
 6

age_at_initial_pathologic_diagnosis:
56 57 72 73 77
 1  1  1  1  2

days_to_death:
 991 1257 1741 NAs
   1    1    1    3

vital_status:
deceased   living     NAs
       3        2        1

msi:
   n NAs
   5    1

location:
Rectosigmoid Junction               rectum
                   4                    2

gender:
f
6
```

```
kras:
n y
3 3

mss:
   y NAs
   5    1

uncurated_author_metadata:
   Length     Class      Mode
        6 character character
```

TCGA.RNASeqV2_eset *Comprehensive molecular characterization of human colon and rectal cancer.*

Description

To characterize somatic alterations in colorectal carcinoma, we conducted a genome-scale analysis of 276 samples, analysing exome sequence, DNA copy number, promoter methylation and messenger RNA and microRNA expression. A subset of these samples (97) underwent low-depth-of-coverage whole-genome sequencing. In total, 16% of colorectal carcinomas were found to be hypermutated: three-quarters of these had the expected high microsatellite instability, usually with hypermethylation and MLH1 silencing, and one-quarter had somatic mismatch-repair gene and polymerase ?? (POLE) mutations. Excluding the hypermutated cancers, colon and rectum cancers were found to have considerably similar patterns of genomic alteration. Twenty-four genes were significantly mutated, and in addition to the expected APC, TP53, SMAD4, PIK3CA and KRAS mutations, we found frequent mutations in ARID1A, SOX9 and FAM123B. Recurrent copy-number alterations include potentially drug-targetable amplifications of ERBB2 and newly discovered amplification of IGF2. Recurrent chromosomal translocations include the fusion of NAV2 and WNT pathway member TCF7L1. Integrative analyses suggest new markers for aggressive colorectal carcinoma and an important role for MYC-directed transcriptional activation and repression.

Usage

```
data( TCGA.RNASeqV2_eset )
```

Format

```
experimentData(eset):
Experiment data
 Experimenter name: Comprehensive molecular characterization of human colon and rectal cancer. Nature 2
  Laboratory: The Cancer Genome Atlas Network 2012
  Contact information:
 Title: Comprehensive molecular characterization of human colon and rectal cancer.
```

```
  URL:
  PMIDs: 22810696

  Abstract: A 168 word abstract is available. Use abstract method.
  Information is available on: preprocessing
  notes:
   platform_title:
      [RNASeqV2] Illumina HiSeq RNA sequencing
   platform_shorttitle:

   platform_summary:
      NA
   platform_manufacturer:
      Illumina
   platform_distribution:
      sequencing
   platform_accession:
      NA
   platform_technology:
      in situ oligonucleotide
   warnings:
      No warnings yet

Preprocessing: default
featureData(eset):
An object of class AnnotatedDataFrame
  featureNames: ? A1BG ... ZZZ3 (20502 total)
  varLabels: probeset gene
  varMetadata: labelDescription
```

Details

```
assayData: 20502 features, 195 samples
Platform type: NA
Overall survival time-to-event summary (in years):
Call: survfit(formula = Surv(time, cens) ~ -1)

   174 observations deleted due to missingness
records   n.max n.start  events  median 0.95LCL 0.95UCL
 21.000  21.000  21.000  18.000   1.208   0.715   5.605

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

unique_patient_ID:
   Length     Class      Mode
      195 character character
```

```
alt_sample_name:
   Length     Class      Mode
      195 character character

sample_type:
adjacentnormal          tumor
            14            181

primarysite:
 co
195

age_at_initial_pathologic_diagnosis:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  31.00   55.00   66.00   64.95   77.00   90.00

days_to_death:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NAs
   43.0   187.0   424.0   816.6  1422.0  2134.0     174

vital_status:
deceased   living     NAs
      18      119       58

msi:
   n    y NAs
  32    9  154

location:
   Length     Class      Mode
      195 character character

gender:
  f   m
 87 108

kras:
   n    y NAs
   9    4  182

mss:
   n    y NAs
   9   32  154

uncurated_author_metadata:
   Length     Class      Mode
      195 character character
```

```
fu:
   n NAs
  60  135

bevacizumab:
   n NAs
  60  135

irinotecan:
   n NAs
  60  135

capecitabine:
   n    y NAs
  59    1  135

cpt11:
   n    y NAs
  59    1  135

dexamethasone:
   n    y NAs
  59    1  135

erbitux:
   n NAs
  60  135

gcsf:
   n NAs
  60  135

fudr:
   n    y NAs
  57    3  135

folfiri:
   n    y NAs
  59    1  135

folfox:
   n NAs
  60  135

leucovorin:
   n NAs
  60  135
```

```
mitomycin:
   n NAs
  60  135

platin:
   n NAs
  60  135

panitumumab:
   n NAs
  60  135

pegfilgrastim:
   n    y NAs
  57    3  135

raltitrexed:
   n    y NAs
  59    1  135

xeloda:
   n NAs
  60  135

ancillary:
   n    y NAs
  58    2  135

chemotherapy:
   n    y NAs
   2   58  135

moltherapy:
   n    y NAs
  57    3  135
```

Index