

# Package ‘COPDSexualDimorphism’

October 7, 2014

**Type** Package

**Title** Sexual dimorphic and COPD differential analysis for gene expression and methylation.

**Version** 1.0.0

**Date** 2014-03-30

**Author** J Fah Sathirapongsasuti

**Maintainer** J Fah Sathirapongsasuti <fah@cs.stanford.edu>

**Description** Sexual dimorphic and COPD differential (SDCD) analysis contrasts regression coefficients from two stratified analysis. Stratification can be done in two ways: by COPD status or by sex. For COPD-stratified analysis, SDCD analysis contrasts sexual dimorphism between cases and controls, while sex-stratified SDCD analysis contrasts COPD differential expression pattern between males and females. The package is meant to be used in conjunction with the package limma.

**License** LGPL-2.1

**LazyLoad** yes

**Depends** COPDSexualDimorphism.data, NCBI2R, RColorBrewer, beeswarm,limma, GenomicRanges, gplots, gtools

**biocViews**

Software, AssayDomain, Microarray, GeneExpression, DNAMethylation, DifferentialExpression

## R topics documented:

COPDSexualDimorphism-package . . . . .	2
do.sdcd.boxplot . . . . .	3
lgrc.sdcd.genes . . . . .	4
sdcd . . . . .	5

<b>Index</b>	<b>8</b>
--------------	----------

COPDSexualDimorphism-package

*Sexual Dimorphic and COPD Differential Analysis*

---

**Description**

Sexual dimorphic and COPD differential (SDCD) analysis contrasts regression coefficients from two stratified analysis. Stratification can be done in two ways: by COPD status or by sex. For COPD-stratified analysis, SDCD analysis contrasts sexual dimorphism between cases and controls, while sex-stratified SDCD analysis contrasts COPD differential expression pattern between males and females. The package is meant to be used in conjunction with the package limma.

**Details**

Package: COPDSexualDimorphism  
Type: Package  
Version: 1.0  
Date: 2013-09-02  
License: LGPL-2.1

Data and main methods to accompany "Integrative Genomics of Sexual Dimorphism in COPD." The main functions are `sdcd`, which contrasts regression coefficients from sex-stratified and COPD-stratified analyses.

**Author(s)**

J Fah Sathirapongsasuti

Maintainer: J Fah Sathirapongsasuti <fah@cs.stanford.edu>

**References**

Sathirapongsasuti JF, Glass K, Huttenhower C, Quackenbush J, DeMeo DL. Integrative Genomics of Sexual Dimorphism in COPD. (In Prep).

[www.lung-genomics.org](http://www.lung-genomics.org)

**Examples**

```
data(lgrc.expr.meta)
data(lgrc.expr)
data(lgrc.genes)

library(limma)

## Sex-stratified
design.mtx = cbind(ctrl=1,
```

```

copd=as.integer(grepl("COPD",colnames(expr))),
age=expr.meta$age,
pkyr=expr.meta$pkyrs)

good.idx = apply(design.mtx,1,function(x){!any(is.na(x))}) & (expr.meta$gender == "1-Male")
male.fit = lmFit(log(expr)[,good.idx], design.mtx[good.idx,])
male.fit = eBayes(male.fit)

good.idx = apply(design.mtx,1,function(x){!any(is.na(x))}) & (expr.meta$gender == "2-Female")
female.fit = lmFit(log(expr)[,good.idx], design.mtx[good.idx,])
female.fit = eBayes(female.fit)

male.female.copd.beta.diff.genes = sdcd(male.fit, female.fit, "copd", lgrc.genes, fdr.cutoff=0.25, file.prefix="n

## COPD-stratified
design.mtx = cbind(ctrl=1,
gender=expr.meta$gender,
age=expr.meta$age,
pkyr=expr.meta$pkyrs)

good.idx = apply(design.mtx,1,function(x){!any(is.na(x))}) & grepl("COPD",colnames(expr))
copd.fit = lmFit(log(expr)[,good.idx], design.mtx[good.idx,])
copd.fit = eBayes(copd.fit)

good.idx = apply(design.mtx,1,function(x){!any(is.na(x))}) & grepl("CTRL",colnames(expr))
ctrl.fit = lmFit(log(expr)[,good.idx], design.mtx[good.idx,])
ctrl.fit = eBayes(ctrl.fit)

copd.ctrl.gender.beta.diff.genes = sdcd(copd.fit, ctrl.fit, "gender", lgrc.genes, fdr.cutoff=0.25, file.prefix="n

## Combine
sdcd.genes = merge(copd.ctrl.gender.beta.diff.genes, male.female.copd.beta.diff.genes, by=setdiff(intersect(names(male.female.copd.beta.diff.genes),names(copd.ctrl.gender.beta.diff.genes))))
sdcd.genes = unique(sdcd.genes)
print(paste("There are", nrow(sdcd.genes),"SDCD genes"))

```

---

do.sdcd.boxplot

*Represent sexual dimorphic data as boxplots.*


---

## Description

Stratify and plot data by gender and by disease status.

## Usage

```
do.sdcd.boxplot(marker, data, copd.bool, male.bool, symbol = marker, filename = paste(marker, ".pdf", sep = ""))
```

## Arguments

**marker** Character string representing the name of the marker (e.g. Ensembl gene, VMR) to plot.

data	Data to be plot. E.g. gene expression or percent methylation.
copd.bool	Boolean array corresponding to the COPD subject columns of data.
male.bool	Boolean array corresponding to the male subject columns of data.
symbol	Gene symbol associated with the marker.
filename	File name. If NA, display plot on screen.
take.log	Boolean. If true, plot in log scale.

**Author(s)**

J Fah Sathirapongsausti

**References**

Sathirapongsausti JF, Glass K, Huttenhower C, Quackenbush J, DeMeo DL. Integrative Genomics of Sexual Dimorphism in COPD. (In Prep).

**Examples**

```
data(lgrc.sdcd.genes)
data(lgrc.expr.meta)
data(lgrc.expr)
copd.bool = (expr.meta$diagmaj == "2-COPD/Emphysema")
male.bool = (expr.meta$gender == "1-Male")
do.sdcd.boxplot("ENSG00000182472", expr, copd.bool, male.bool, symbol="CAPN12", filename=NA)
```

---

lgrc.sdcd.genes                      *Sexual dimorphic and COPD differential genes*

---

**Description**

A list of 959 genes whose expressions were significantly different between males and females and between COPD and controls. These were identified by the function `sdcd`.

**Usage**

```
data(lgrc.sdcd.genes)
```

**Format**

The format is: num [1:14557, 1:229] 10.45 1.88 12 9.02 8.81 ... - attr(\*, "dimnames")=List of 2 ..\$ :  
chr [1:14557] "ENSG00000000003" "ENSG00000000005" "ENSG000000000419" "ENSG000000000457"  
... ..\$: chr [1:229] "LT001098RU\_COPD" "LT001796RU\_CTRL" "LT005419RU\_COPD" "LT007392RU\_COPD"  
...

**References**

Sathirapongsausti JF, Glass K, Huttenhower C, Quackenbush J, DeMeo DL. Integrative Genomics of Sexual Dimorphism in COPD. (In Prep).

**Examples**

```
data(lgrc.sdcd.genes)
head(sdcd.genes)
```

sdcd

*Sexual dimorphic and COPD differential analysis***Description**

Given linear models (limma) from a stratified analysis, sdcd compares the coefficients of the main effects across sexes. It then reports on the markers with significant differences in the coefficients. Typically used in conjunction with limma.

**Usage**

```
sdcd(male.fit, female.fit, coeff, genes, fdr.cutoff = 0.25, stat = c("z", "t"), file.prefix = "male.fem
sdcd.vmr(male.fit, female.fit, coeff, genes, fdr.cutoff = 0.25, stat = c("z", "t"), annotate = FALSE, an
sdcd.core(male.fit, female.fit, coeff, stat = c("z", "t"))
```

**Arguments**

male.fit, female.fit	Objects of type limma, as generated by the function eBayes. OR a list with fields: coefficients (matrix), stdev.unscaled (matrix), sigma (numeric vector), df.residual (numeric), and df.prior (numeric).
coeff	Coefficients of the main effect of interest. sex for COPD-stratified analysis and COPD for sex-stratified analysis. This should correspond to a column name of the matrices coefficients and stdev.unscaled in male.fit and female.fit.
genes	Annotation of the gene expression probes.
fdr.cutoff	Numeric cutoff for FDR q-values
stat	Choices between "z" and "t". For "z", a z-test is used to assess significance of the difference between the regression coefficients. For "t", the t-statistics, as opposed to the coefficients themselves, are contrasted.
annotate	For SDCD methylation analysis, a boolean to determin if the methylated regions (VMRs) should be annotated by genes within 10kb or not.
annotate.with	For SDCD methylation analysis, when annotate == TRUE this option indicates how to annotate the methylated regions. The "genes" option uses the input object while the "NCBI" option uses function GetNeighGenes in the package NCBI2R.
file.prefix	Prefix for output file name.
class.names	An array of character strings of length two representing the two strata.
write.file	A boolean that determined weather the results should be written out as files.

## Details

This is to be used in conjunction with the package `limma`. Linear model fits are passed to the `sdcd` function. The main function is for gene expression data while the function `sdcd.vmr` is for methylation data. The main functionality is the same between the two, but the data annotation and output are slightly different. See intended usage in the vignette.

## Value

A data frame with gene information and the following columns:

<code>*.beta</code>	Coefficients of the main effects in the two strata
<code>*.sd</code>	Standard deviation of the coefficients of the main effects in the two strata
<code>*.t</code>	T-statistics associated with the coefficients of the main effects in the two strata
<code>*.p.value</code>	P-value associated with the coefficients of the main effects in the two strata
<code>beta.diff</code>	The difference between the coefficients in the two strata
<code>beta.diff.pooled.sd</code>	Standard deviation of <code>beta.diff</code>
<code>stratum1.stratum2.p</code>	P-value associated with <code>beta.diff</code>
<code>stratum1.stratum2.p.adj</code>	Benjamini-Hochberg FDR corrected p-value

## Author(s)

J Fah Sathirapongsasuti

## References

Sathirapongsasuti JF, Glass K, Huttenhower C, Quackenbush J, DeMeo DL. Integrative Genomics of Sexual Dimorphism in COPD. (In Prep).

## See Also

[lgrc.expr](#)

## Examples

```
data(lgrc.expr.meta)
data(lgrc.expr)
data(lgrc.genes)

## Sex-stratified
design.mtx = cbind(ctrl=1,
  copd=as.integer(grepl("COPD", colnames(expr))),
  age=expr.meta$age,
  pkyr=expr.meta$pkyrs)

good.idx = apply(design.mtx, 1, function(x){!any(is.na(x))}) & (expr.meta$gender == "1-Male")
male.fit = lmFit(log(expr)[,good.idx], design.mtx[good.idx,])
```

```
male.fit = eBayes(male.fit)

good.idx = apply(design.mtx,1,function(x){!any(is.na(x))}) & (expr.meta$gender == "2-Female")
female.fit = lmFit(log(expr)[,good.idx], design.mtx[good.idx,])
female.fit = eBayes(female.fit)

male.female.copd.beta.diff.genes = sdcd(male.fit, female.fit, "copd", lgrc.genes, fdr.cutoff=0.25, file.prefix="r")

## COPD-stratified
design.mtx = cbind(ctrl=1,
gender=expr.meta$gender,
age=expr.meta$age,
pkyr=expr.meta$pkyrs)

good.idx = apply(design.mtx,1,function(x){!any(is.na(x))}) & grepl("COPD",colnames(expr))
copd.fit = lmFit(log(expr)[,good.idx], design.mtx[good.idx,])
copd.fit = eBayes(copd.fit)

good.idx = apply(design.mtx,1,function(x){!any(is.na(x))}) & grepl("CTRL",colnames(expr))
ctrl.fit = lmFit(log(expr)[,good.idx], design.mtx[good.idx,])
ctrl.fit = eBayes(ctrl.fit)

copd.ctrl.gender.beta.diff.genes = sdcd(copd.fit, ctrl.fit, "gender", lgrc.genes, fdr.cutoff=0.25, file.prefix="r")

## Combine
sdcd.genes = merge(copd.ctrl.gender.beta.diff.genes, male.female.copd.beta.diff.genes, by=setdiff(intersect(names(male.female.copd.beta.diff.genes), names(male.female.copd.beta.diff.genes))), all=TRUE)
sdcd.genes = unique(sdcd.genes)
print(paste("There are ", nrow(sdcd.genes), " SDCD genes"))
```

# Index

## \*Topic **COPD**

COPDSexualDimorphism-package, 2  
do.sdcd.boxplot, 3  
sdcd, 5

## \*Topic **SDCD**

COPDSexualDimorphism-package, 2  
sdcd, 5

## \*Topic **boxplot**

do.sdcd.boxplot, 3

## \*Topic **datasets**

lgrc.sdcd.genes, 4

## \*Topic **package**

COPDSexualDimorphism-package, 2

## \*Topic **sexual dimorphism**

COPDSexualDimorphism-package, 2  
do.sdcd.boxplot, 3  
sdcd, 5

COPDSexualDimorphism  
(COPDSexualDimorphism-package),  
2

COPDSexualDimorphism-package, 2

do.sdcd.boxplot, 3

getNeighbors (sdcd), 5

lgrc.expr, 6

lgrc.sdcd.genes, 4

my.smart.plot (do.sdcd.boxplot), 3

pool.sd (sdcd), 5

sdcd, 5

sdcd.genes (lgrc.sdcd.genes), 4