# Package 'RNAetc'

August 3, 2021

**Type** Package

**Title** What the Package Does (Title Case)

**Version** 0.1.0

**Author** Kim Dill-McFarland

**Maintainer** Kim Dill-McFarland <kadm@uw.edu>

**Description** More about what it does (maybe more than one line)
Use four spaces when indenting paragraphs within the Description.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**Imports** dplyr, edgeR, forcats, ggplot2, limma, magrittr, readr, tibble, tidyr, WGCNA

**Depends** R (>= 2.10)

## R topics documented:



---

`align_metrics` *Extract and format cleaning and alignment metrics*

---

### Description

Extract data from FastQC trim settings, Picard, samtools flagstat, and featureCounts output by the RNA-seq fastq pipeline

### Usage

```
align_metrics(
  data.dir = NULL,
  trim = TRUE,
  bam = TRUE,
  picard = TRUE,
  bam.filter = TRUE,
  count = TRUE
)
```

### Arguments

| | |
|---|---|
| data.dir | Character string of directory containing all associated files |
| trim | Logical if should include FastQC trim settings |
| bam | Logical if should include samtools flagstat for raw alignments |
| picard | Logical if should include Picard for raw alignments |
| bam.filter | Logical if should include samtools flagstat for filtered alignments |
| count | Logical if should include featureCounts total reads in genes |

### Value

Data frame cleaning and alignment metrics for all libraries

---

## example.dat *kimma example DGEList.*

---

### Description

An edgeR DGEList data set containing unnormalized RNA-seq counts. RNA-seq of human dendritic cells cultured with and without virus. Samples from 3 donors and a random subset of 1000 genes were selected. Counts are unnormalized.

### Usage

```
example.dat
```

### Format

Formal class 'DGEList' [package "edgeR"] with 1 slot:

1. **counts** A matrix with 1000 rows and 6 columns

   **rownames** character. ENSEMBL gene ID.
   **lib1** integer. Counts in library 1.
   **lib2** integer. Counts in library 2.
   **lib3** integer. Counts in library 3.
   **lib4** integer. Counts in library 4.
   **lib5** integer. Counts in library 5.
   **lib6** integer. Counts in library 6.

2. **samples** A data frame with 6 rows and 7 columns

   **group** factor. No grouping was provided. All = 1.

   **lib.size** numeric. Total library size for this 1000 gene subset.

   **norm.factors** numeric. Normalizatin factors. No normalization was completed. All = 1.

   **libID** character. Unique library ID. Matches column names in counts.

   **donorID** character. Donor ID.

   **median_cv_coverage** numeric. Median coefficient of variation of coverage. Quality metric for sequencing libraries calculated from original full data set.

   **virus** character. A for media samples with no virus. B for virus-infected samples.

3. **genes** A data frame with 1000 rows and 5 columns

   **hgnc_symbol** character. Current approved HGNC symbol.

   **Previous symbols** character. Previous HGNC symbols.

   **Alias symbols** character. Alias HGNC symbols.

   **gene_biotype** character. Gene product type. All = protein-coding.

   **geneName** character. ENSEMBL gene ID. Matches row names in counts.

## Source

https://github.com/altman-lab/P259_pDC_public

## References

Dill-McFarland et al. 2021. Eosinophil-mediated suppression and Anti-IL-5 enhancement of plasmacytoid dendritic cell interferon responses in asthma. J Allergy Clin Immunol. In revision

---

# example.kin *kimma example kinship.*

---

## Description

Matrix of pairwise kinship values between donor 1,2,3. Values are dummy data with 1 for self comparison, 0.5 for siblings, and 0.1 for unrelated.

## Usage

```
example.kin
```

## Format

A matrix with 3 rows and 3 variables:

**rowname** Donor ID. Same as column names

**donor1** numeric kinship (0-1) with donor 1

**donor2** numeric kinship (0-1) with donor 2

**donor3** numeric kinship (0-1) with donor 3

`example.voom` *kimma example EList.*

## Description

A limma EList data set containing normalized log2 RNA-seq counts. RNA-seq of human dendritic cells cultured with and without virus. Samples from 3 donors and a random subset of 1000 genes were selected. Counts are TMM normalized log2 counts per million (CPM).

## Usage

```
example.voom
```

## Format

Formal class 'EList' [package "limma"] with 1 slot:

1. **genes** A data frame with 1000 rows and 5 columns
   **hgnc_symbol** character. Current approved HGNC symbol.
   **Previous symbols** character. Previous HGNC symbols.
   **Alias symbols** character. Alias HGNC symbols.
   **gene_biotype** character. Gene product type. All = protein-coding.
   **geneName** character. ENSEMBL gene ID. Matches row names in E.
2. **targets** A data frame with 6 rows and 7 columns
   **group** factor. No grouping was provided. All = 1.
   **lib.size** numeric. Total library size for this 1000 gene subset.
   **norm.factors** numeric. TMM normalizatin factors.
   **libID** character. Unique library ID. Matches column names in E.
   **donorID** character. Donor ID.
   **median_cv_coverage** numeric. Median coefficient of variation of coverage. Quality metric for sequencing libraries calculated from original full data set.
   **virus** character. A for media samples with no virus. B for virus-infected samples.
3. **E** A matrix with 1000 rows and 6 columns
   **rownames** character. ENSEMBL gene ID.
   **lib1** integer. log2 CPM in library 1.
   **lib2** integer. log2 CPM in library 2.
   **lib3** integer. log2 CPM in library 3.
   **lib4** integer. log2 CPM in library 4.
   **lib5** integer. log2 CPM in library 5.
   **lib6** integer. log2 CPM in library 6.
4. **weights** A matrix with 1000 rows and 6 columns
   **1** numeric. limma gene weights for library 1.
   **2** numeric. limma gene weights for library 2.
   **3** numeric. limma gene weights for library 3.
   **4** numeric. limma gene weights for library 4.
   **5** numeric. limma gene weights for library 5.
   **6** numeric. limma gene weights for library 6.
5. **design** A matrix with 6 rows and 1 column
   **GrandMean** numeric. limma default design matrix.

## Source

https://github.com/altman-lab/P259_pDC_public

## References

Dill-McFarland et al. 2021. Eosinophil-mediated suppression and Anti-IL-5 enhancement of plasmacytoid dendritic cell interferon responses in asthma. J Allergy Clin Immunol. In revision

---

`filter_rare` *Filter rare and low abundance genes*

---

## Description

Filter genes at a minimum counts per million (CPM) in a minmum number or percent of total samples.

## Usage

```
filter_rare(
  dat,
  min.CPM,
  gene.var = "geneName",
  min.sample = NULL,
  min.pct = NULL,
  plot = FALSE
)
```

## Arguments

| | |
|---|---|
| `dat` | DGEList output by edgeR::DEGList( ) |
| `min.CPM` | numeric minimum counts per million (CPM) |
| `gene.var` | character name for column with gene names in dat$genes that matches names in expression data dat$E. Default "geneName" |
| `min.sample` | numeric minimum number of samples |
| `min.pct` | numeric minimum percent of samples (0-100) |
| `plot` | logical if should plot mean variance trends |

## Value

DGEList object filtered to not rare genes

## Examples

```
dat.filter <- filter_rare(dat = example.dat, min.CPM = 0.1, min.sample = 3)
dat.filter <- filter_rare(dat = example.dat, min.CPM = 0.1, min.pct = 10, plot = TRUE)
```

make_modules *Construct WGCNA modules and associated data*

## Description

Make WGCNA modules from gene expression data with dynamic soft threshold selection. Also outputs mean module expression and DAVID formatted gene lists

## Usage

```
make_modules(
  dat,
  genes = NULL,
  Rsq.min = NULL,
  sft.value = NULL,
  minModuleSize = 20,
  deepSplit = 3,
  nThread = 2
)
```

## Arguments

| | |
|---|---|
| dat | limma EList output by voom( ) |
| genes | Character vector of genes to used in module building. Must match rownames in dat. If not set, all genes in dat are used |
| Rsq.min | Numeric minimum R-squared for soft threshold selection. If set, sft.value is not used |
| sft.value | Numeric soft threshold. Set when minimum R-squared is no used |
| minModuleSize | Numeric minimum module size |
| deepSplit | Integer value between 0 and 4. Provides a simplified control over how sensitive module detection should be to module splitting, with 0 least and 4 most sensitive |
| nThread | Integer for number of threads to use |

## Value

List including:

- genes Character vector of genes used in module building
- sft Data frame with soft thresholding selected for module building. Includes power, minimum R-squared, and connectivity
- sft.plot ggplot object of soft thresholding topology and connectivity
- mods Data frame of genes in modules
- mods.voom Data frame of mean module expression in each library
- david DAVID formatted data frame of genes in modules

## Examples

```
dat.mods <- make_modules(dat = example.voom, sft.value = 1)
```

# Index