Package 'CloneStrat'

January 18, 2020

Title Multi-sample clonal deconvolution of tumor exome sequencing data

Version 0.0.9

Description
Functions to deconvolute clones and sub-clones in multi-regional/temporal whole exome sequencing data of solid tumor in presence of microarray based copy number profiles. Additional functions include estimation of said copy number profiles from exome sequencing.

Depends R (¿= 3.5.0)

URL https://https://github.com/Subhayan18/CloneStrat/

BugReports https://github.com/Subhayan18/CloneStrat/issues/

License GPL-3

Encoding UTF-8

LazyData true

Imports readxl,
mclust,
fpc,
KODAMA,
dplyr,
ggplot2

NeedsCompilation no

RoxygenNote 7.0.2

R topics documented:

cluster.doc *Clonal deconvolution*

Description

Clone / Sub-clone decomposition of WES data

Usage

```
cluster.doc(x)
```

Arguments

x	A dataframe with first column as sample IDs and second column as variant allele frequencies of corresponding variants obtained from WES

Value

A list of 6 objects is retuned that includes all the summary statistics, diagnositics and the predictions.

cluster.diagnostics is an object of S3 class which includes clustering diagnostics from the model-based clustering.

fitted cluster is a clustering object of class S3. This will be the fitted clustering either user driven/ system predicted or user over-ridden.

predicted.data is necessarily an extension to the input data x with the addition of the predicted clone and sub-clone status of each variant for corresponding samples.

optimum.clusters is the system predicted optimum number of clusters that was either fitted or suggested to the user. The detailed statistics used for this decision can be found in cluster.diagnostics$bic

diagnosed.dunn is the Dunn index for the suggested cluster.

fitted.dunn is the Dunn index for the fitted cluster.

Examples

```
cluster.doc(test.dat)
```

cluster.doubt *User overriden clonal deconvolution*

Description

Sample specific user curated Clone / Sub-clone decomposition of WES data

Usage

```
cluster.doubt(CD.obj, sample.name, cluster.num)
```

Arguments

`CD.obj`	A `cluster.doc` object
`sample.name`	a vector of sample IDs
`cluster.num`	a numeric vetor of clone/sub-clonal split of respective sample

Value

A list of 3 objects

`fitted.cluster` includes the clustering results from the final fit with user input

`predicted.data` is the original fit rendered from `cluster.doc`

`userfed.data` shows the changed clustering results due to the user defined clone / sub-clone smear for the selected samples

Examples

```
cd.res<-cluster.doc(test.dat)
cd.new<-cluster.doubt(cd.res,c("Sample_1","Sample_3"),c(2,2,3,2))
```

`CS.scale` *Probabilistic quotient normalization of WES data*

Description

A normalization technique described in *Dieterle, et al. (2006)* applied on the cancer cell fraction (CCF) to rescale variant allele frequencies (VAF). This method is particularly suggested if the quality of samples vary more than 0.1 all accross the board.

Usage

```
CS.scale(x, vaf, CCF)
```

Arguments

`x`	A `dataframe` of *WES* data with first column as sample IDs of corresponding variants
`vaf`	The column number of `x` that includes VAFs
`CCF`	The column number of `x` that includes CCFs

Value

A `dataframe` with all the elements of `x` with the new estimated VAFs in the column *scaled.vaf* and an additional column *unscaled.vaf* that includes the original VAFs

Examples

```
pqn.dat<-CS.scale(test.dat,vaf=2,CCF=3)
hist(pqn.dat$scaled.vaf)
```

`mutect2.qc` *Quality Control on Mutect2 output*

Description

A quality control (QC) and transformation on the WES output from the Mutect2 variant caller. This re-organizes the data in a way that is friendlier for using in *CloneStrat*

Usage

```
mutect2.qc(WES, sample.name)
```

Arguments

`WES`	A `dataframe` of the Mutect2 output
`sample.name`	a `vector` of sample names or IDs

Value

A transformed `dataframe` usable in *CloneStrat* that represents data on each variant of each sample in rows

Examples

```
res<-mutect2.qc(WES,sample.name)
```

`T.goodness.test` *Test of fit of clonal deconvolution*

Description

A chi square test to assess the *goodness of fit* of the clonal : sub-clonal clouds. This test can be used to obtain outliers that do not fit into the proposed clonal deconvolution space.

Usage

```
T.goodness.test(x)
```

Arguments

`x`	A `dataframe` with the first three columns in the specific order: sample name or ID of a variant, variant allele frquencies (VAF) and cancer cell fraction (CCF)

Value

A list of two objects. *x* is same as the input `dataframe` with addede columns named *expected VAF_*, *chi_sq_* and *P value_* corresponding to each cloud of clone : Sub-clone combination. *rej* is a subset of `x` containing variants that fail the test for at least one cloud.

*expected VAF_* represents estimated variant allele frequencies for a given cloud.

*chi_sq_* is the Chi square test statistic for the cloud.

*P value_* is the P value corresponding to the *chi_sq_* statistic.

Examples

```
T.goodness.test(test.dat)
```

`test.dat` *Random number generated WES data for eight hypothetical samples*

Description

Data generated with varying random normal probabilities. ideal chromosomal segmentation profile is assumed resulting in three separate distinct clouds of clones and sub-clones.

Usage

```
data(test.dat)
```

Format

An object of class `"dataframe"`

Value

`sample` is column of IDs corresponding to 8 distinct samples.

`vaf` denotes the variant allele frequencies of each variant (see `annotation`).

`CCF` are the cancer cell fractions of each sample.

`annotation` indicates corresponding variants for which observations are notes in each row. Variants can be shared among several samples as well as be private mutation.

Examples

```
data(test.dat)
table(test.dat$CCF)
table(test.dat$annotation)
hist(test.dat$vaf)
```

`variant.auto.plot` *Auromated Multi-sample plot*

Description

Automated plotting of all variants present in the WES data

Usage

```
variant.auto.plot(CD.obj, annotation.col)
```

Arguments

`CD.obj`	A `cluster.doc object`
`annotation.col`	name of the column containing annotations of the variants in original WES `dataframe` used in the clonal deconvolution using `cluster.doc`

Value

Plot objects with the relevant annotation highlighted.

This function plots all variants present in the sample. Depending on the number of variants this can generate a *lot* of plots. All of these plots will be saved under a new directory named `img` inside the working directory. Hence, it is important to check that there are no directory named `img` inside the working directory

Examples

```
cd.res<-cluster.doc(test.dat)
variant.auto.plot(cd.res,'annotation')
```

---

variant.plot *Multi-sample variant plot*

Description

Plotting a specific variant present in more than one WES sample

Usage

```
variant.plot(CD.obj, annotation.col, variant)
```

Arguments

`CD.obj`	A `cluster.doc object`
`annotation.col`	name of the column containing annotations of the variants in original WES `dataframe` used in the clonal deconvolution using `cluster.doc`
`variant`	a `character` string specifying *only one* annotation which is to be displayed

Value

A plot object with the relevant annotation highlighted

Examples

```
cd.res<-cluster.doc(test.dat)
variant.plot(cd.res,'annotation','variant_74')
```

Index