Package ‘TPES’

July 21, 2025

Type Package

Title Tumor Purity Estimation using SNVs

Version 1.0.0

Date 2019-01-09

Author Alessio Locallo <`alessio.locallo@gmail.com`>, Davide Prandi <`davide.prandi@unitn.it`>, Francesca Demichelis <`f.demichelis@unitn.it`>

Maintainer Alessio Locallo <`alessio.locallo@gmail.com`>

Description A bioinformatics tool for the estimation of the tumor purity from sequencing data. It uses the set of putative clonal somatic single nucleotide variants within copy number neutral segments to call tumor cellularity.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Depends R (>= 2.10)

Imports grDevices, graphics, stats

NeedsCompilation no

Repository CRAN

Date/Publication 2019-01-11 17:40:12 UTC

Contents

`TCGA_A8_A0A7_maf` *SNVsReadCountsFile for sample TCGA-A8-A0A7*

Description

A data frame object containing the read counts data of somatic single nucleotide variants (SNVs) loci for sample TCGA-A8-A0A7. The header contains the chromosme that harbors the SNV ("chr" column), the position of the SNV (defined by the "start" and "end" columns), the informations about the reference and alternative base counts ("ref.count" and "alt.count" columns, respectively) and finally the sample ID ("sample" column). For more information please visit MAF file format.

Format

A data.frame object.

`TCGA_A8_A0A7_ploidy` *Ploidy data for sample TCGA-A8-A0A7*

Description

A data frame containing the ploidy status of a sample. It must contains at least the sample ID ("sample" column) and the ploidy status ("ploidy" column).

Format

A data.frame object.

`TCGA_A8_A0A7_seg` *SEG file (segmented data) for sample TCGA-A8-A0A7*

Description

A data frame object that lists loci and associated numeric values. The header must be compatible with the standard format defined by the Broad Institute. For more information please visit SEG file format.

Format

A data.frame object.

`TCGA_HT_8564_maf` *SNVsReadCountsFile for sample TCGA-HT-8564*

Description

A data frame object containing the read counts data of somatic single nucleotide variants (SNVs) loci for sample TCGA-HT-8564. The header contains the chromosme that harbors the SNV ("chr" column), the position of the SNV (defined by the "start" and "end" columns), the informations about the reference and alternative base counts ("ref.count" and "alt.count" columns, respectively) and finally the sample ID ("sample" column). For more information please visit MAF file format.

Format

A data.frame object.

`TCGA_HT_8564_ploidy` *Ploidy data for sample TCGA-HT-8564*

Description

A data frame containing the ploidy status of a sample. It must contains at least the sample ID ("sample" column) and the ploidy status ("ploidy" column).

Format

A data.frame object.

`TCGA_HT_8564_seg` *SEG file (segmented data) for sample TCGA-HT-8564*

Description

A data frame object that lists loci and associated numeric values. The header must be compatible with the standard format defined by the Broad Institute. For more information please visit SEG file format.

Format

A data.frame object.

`TPES_purity` *Tumor Purity Estimation using SNVs*

Description

TPES_purity function estimates tumor purity.

Usage

```
TPES_purity(ID, SEGfile, SNVsReadCountsFile, ploidy, RMB = 0.47,
  maxAF = 0.55, minCov = 10, minAltReads = 5, minSNVs = 10)
```

Arguments

`ID`	Sample ID. Must be the same ID as in SEGfile, SNVsReadCountsFile and ploidy.
`SEGfile`	A standard SEG file (segmented data). It is a data frame object that lists loci and associated numeric values. The header must be compatible with the standard format defined by the Broad Institute. For more information please visit SEG file format.
`SNVsReadCountsFile`	A standard MAF (Mutation Annotation Format) file. It is a data frame object containing the read counts data of somatic single nucleotide variants (SNVs) loci. The header must contains at least informations about the chromosme that harbors the SNV ("chr" column), the position of the SNV (defined by the "start" and "end" columns), the sample ID ("sample" column) and finally the informations about the reference and alternative base counts ("ref.count" and "alt.count" columns, respectively). For more information please visit MAF file format.
`ploidy`	A data frame containing the ploidy status of a sample. It must contain at least the sample ID ("sample" column) and the ploidy status ("ploidy" column).
`RMB`	The Reference Mapping Bias value. The reference genome contains only one allele at any given locus, so reads that carry a non-reference allele are less likely to be mapped during alignment; this causes a shift from 0.5. It can be estimated as: $1 - medAF$, where medAF is the median value of the allelic fraction of the sample's germline heterozygous SNPs. Default is set to 0.47. For more informations see: PMID: 19808877.
`maxAF`	The filter on the allelic fraction (AF) distribution of SNVs. This is necessary to be sure to keep only heterozygous SNVs. Clonal and subclonal SNVs, which have an AF greater than maxAF, will be removed.
`minCov`	The minimum coverage for a SNV to be retained.
`minAltReads`	The minimum coverage for the alternative base of a SNV to be retained.
`minSNVs`	The minimum number of SNVs required to make a purity call.

Value

TPES returns a data.frame object with one row per sample and the following columns:

`sample`	The sample ID;
`purity`	The sample purity estimated by TPES;
`purity.min`	The sample minimum purity estimated by TPES;
`purity.max`	The sample maximum purity estimated by TPES;
`n.segs`	The number of copy number neutral segments used by TPES;
`n.SNVs`	The number of SNVs used by TPES;
`RMB`	The Reference Mapping Bias value used to estimate the tumor purity;
`BandWidth`	The smoothing bandwidth value of the `density` function chosen by TPES.
`log`	Reports if the run was successful; otherwise provides debugging information.

Examples

```
## Compute tumor purity for samples "TCGA-A8-A0A7" and "TCGA-HT-8564"
## https://cancergenome.nih.gov/
## Please copy and paste the following lines:
library(TPES)
TPES_purity(ID = "TCGA-A8-A0A7", SEGfile = TCGA_A8_A0A7_seg,
SNVsReadCountsFile = TCGA_A8_A0A7_maf, ploidy = TCGA_A8_A0A7_ploidy,
RMB = 0.47, maxAF = 0.55, minCov = 10, minAltReads = 5, minSNVs = 10)

TPES_purity(ID = "TCGA-HT-8564", SEGfile = TCGA_HT_8564_seg,
SNVsReadCountsFile = TCGA_HT_8564_maf, ploidy = TCGA_HT_8564_ploidy,
RMB = 0.47, maxAF = 0.55, minCov = 10, minAltReads = 5, minSNVs = 10)
```

`TPES_report` *Tumor Purity Estimation using SNVs*

Description

TPES_report function produces a graphical report regarding the allelic fraction values of the putative clonal SNVs used by TPES_purity and the density function(s) computed by TPES_purity.

Usage

```
TPES_report(ID, SEGfile, SNVsReadCountsFile, ploidy, RMB = 0.47,
  maxAF = 0.55, minCov = 10, minAltReads = 5, minSNVs = 10)
```

Arguments

`ID`	Sample ID. Must be the same ID as in SEGfile, SNVsReadCountsFile and ploidy.
`SEGfile`	A standard SEG file (segmented data). It is a data frame object that lists loci and associated numeric values. The header must be compatible with the standard format defined by the Broad Institute. For more information please visit SEG file format.
`SNVsReadCountsFile`	A standard MAF (Mutation Annotation Format) file. It is a data frame object containing the read counts data of somatic single nucleotide variants (SNVs) loci. The header must contains at least informations about the chromosme that harbors the SNV ("chr" column), the position of the SNV (defined by the "start" and "end" columns), the sample ID ("sample" column) and finally the informations about the reference and alternative base counts ("ref.count" and "alt.count" columns, respectively). For more information please visit MAF file format.
`ploidy`	A data frame containing the ploidy status of a sample. It must contain at least the sample ID ("sample" column) and the ploidy status ("ploidy" column).
`RMB`	The Reference Mapping Bias value. The reference genome contains only one allele at any given locus, so reads that carry a non-reference allele are less likely to be mapped during alignment; this causes a shift from 0.5. It can be estimated as: $1 - medAF$, where medAF is the median value of the allelic fraction of the sample's germline heterozygous SNPs. Default is set to 0.47. For more informations see: PMID: 19808877.
`maxAF`	The filter on the allelic fraction (AF) distribution of SNVs. This is necessary to be sure to keep only heterozygous SNVs. Clonal and subclonal SNVs, which have an AF greater than maxAF, will be removed.
`minCov`	The minimum coverage for a SNV to be retained.
`minAltReads`	The minimum coverage for the alternative base of a SNV to be retained.
`minSNVs`	The minimum number of SNVs required to make a purity call.

Value

A plot with:

`histogram`	Represents the allelic fraction distribution of putative clonal and subclonal (if presents) SNVs within copy number neutral segments and the peak(s) detected by TPES;
`density plot`	Represents how the density function varies according to different bandwidth values (for more information see `density`); only the bandwidth values that result in at most 2 peaks are considered.

Examples

```
## Generate TPES report for samples "TCGA-A8-A0A7" and "TCGA-HT-8564"
## https://cancergenome.nih.gov/
## Please copy and paste the following lines:
library(TPES)
```

```
TPES_report(ID = "TCGA-A8-A0A7", SEGfile = TCGA_A8_A0A7_seg,
SNVsReadCountsFile = TCGA_A8_A0A7_maf, ploidy = TCGA_A8_A0A7_ploidy,
RMB = 0.47, maxAF = 0.55, minCov = 10, minAltReads = 5, minSNVs = 10)

TPES_report(ID = "TCGA-HT-8564", SEGfile = TCGA_HT_8564_seg,
SNVsReadCountsFile = TCGA_HT_8564_maf, ploidy = TCGA_HT_8564_ploidy,
RMB = 0.47, maxAF = 0.55, minCov = 10, minAltReads = 5, minSNVs = 10)
```

Index