R documentation

of 'mol_time.Rd'

January 18, 2018

---

mol_time *Molecular timing from CN and SNV profiles*

Description

This function uses the copy number (CN) profile along with the variant allele frequencies (VAFs) of single nucleotide variants (SNVs) to estimate timing of structural variation events. For more details see the methods section of our paper (citation details to be added on publication). The function uses inputs from ASCAT, copy number (Battenburg) and VAF calling.

Usage

```
mol_time(data.dir="data", res.dir="res-thresh_0.2")
```

Arguments

res.dir="../res-gth_0.2-hth_0.4-rsd_3"	directory for output files (should not exist, else the script stops). the directory name is constructed by bits of the parameters. The example shown above is produced if all parameters are the default ones
samples=NA	list of samples for the analysis, else all samples in dataset are analysed
data.dir="data"	inputs directory, should contain files: copy_number.txt, shared_clo-nal_mutations.txt and sample_normal_contamination.txt
grey.thresh=0.2	data points that belong to an alternative cluster (mclust() docs) with probability $> grey.thresh$ are excluded from the timing calculations
hclust.thresh=0.4	level at which clusters are created from the hclust tree
seed=3	set random seed at the top of function
iter=1000	bootstrap iterations
warn=FALSE	when TRUE it prints messages about removed samples (due to missing CN profile)
eps=FALSE	when TRUE also produces .eps versions of the _clusters plots

Details

During a copy number gain all SNVs already present on the involved allele will be duplicated as well. Consequently, the VAF of all these substitutions will change from 50% to ~66%, being present on two alleles of three. All mutations occurred anytime on the not duplicated allele and on one of duplicated alleles will have a VAF ~ 33%. Consequently, all subclonal substitutions will occur on one single allele with a $VAF < 33\%$.

This functions assigns each substitution to the 2 (pre gain) or 1 (post gain) alleles using the mclust R function. The input to the function is the corrected VAF profiles. To avoid noise between the different clusters the function uses a threshold below which data points that belong to an alternative cluster with probability larger than the threshold are excluded from the timing calculation. The timing calculation is based on the assumption that each patient has a constant mutation rate (r). This function estimates the relative time of each chromosomal gain occurrence ("molecular time"; t) using the model and formulas presented in the materials section of our paper.

Value

The function returns no value. But produces a number of output files in res.dir.

Note

version 0.1, 2017/9/5

Author(s)

Peter Campbell, Francesco Maura and Nicos Angelopoulos.
This code is maintained by Nicos Angelopoulos.
http://www.sanger.ac.uk/science/groups/cancer-genome-project

See Also

See methods in the following publication: (To be added on publication of current manuscript.)

Examples

```
 ## Not run: mol_time(samples="PD26401a")
 ## Not run: mol_time(samples=c("PD26400a","PD26401a"))
## Not run: mol_time(grey.thresh=0.1, data.dir="data", res.dir="res-thresh_0.1",warn=T)
```

Index