Package 'SALMON'

October 13, 2022

Title SALMON (Signature Analyzer for Low Mutation cOuNts)

Version 0.0.4

Date 2022-10-13

Description To perform mutational signature analysis for targeted sequenced tumors.

Imports stats, glmnet

Depends R (>= 4.1.0)

License GPL-2

NeedsCompilation yes

Author DongHyuk Lee [aut],
Bin Zhu [aut],
Bill Wheeler [cre]

Maintainer Bill Wheeler <wheelerb@imsweb.com>

R topics documented:

SALMON-package   *SALMON (Signature Analyzer for Low Mutation cOuNts)*

Description

To perform mutational signature analysis for targeted sequenced tumors.

Details

This packages includes a novel algorithm, SALMON (Signature Analyzer for Low Mutation cOuNts), to perform mutational signature analysis for targeted sequenced tumors. The algorithm consists of first applying the signeR algorithm to extract profiles of de novo mutational signatures by appropriately adjusting for various panel sizes. Next, the profiles of identified de novo mutational signatures are mapped to the profiles of catalog signatures of tumor mutation burden (TMB), in the unit of the number of mutations per million base pairs, using penalized non-negative least squares. Then, given the panel sizes and profiles of mapped TMB catalog signatures, signature activities are estimated for all samples simultaneously through the Expectation-Maximization (EM) algorithm. Finally, the expected number of mutations attributed by a signature, namely signature expectancy, is calculated for each targeted sequenced tumor.

The main functions in this package are EstimateSigActivity, CalculateSigExpectancy, and mappingSignature.

Author(s)

Donghyuk Lee <dhyuklee@pusan.ac.kr> and Bin Zhu <bin.zhu@nih.gov>

References

Lee, D., Wang, D., Yang, X., Shi, J., Landi, M., Zhu, B. (2021) SUITOR: selecting the number of mutational signatures through cross-validation. bioRxiv, doi: https://doi.org/10.1101/2021.07.28.454269.

CalculateSigExpectancy

CalculateSigExpectancy

Description

Estimate signature expectancy

Usage

```
CalculateSigExpectancy(L, W, H)
```

Arguments

L	Panel size matrix or data frame with samples in columns
W	Catalog signature profiles matrix or data frame with signatures in columns
H	Activity matrix or data frame with samples in columns

Details

If K is the number of signatures and N is the number of samples, then H must be of dimension K X N, ncol(L) = N, and ncol(W) = K.

Value

A matrix of dimension K X N, where K is the number of signatures and N is the number of samples.

Author(s)

Donghyuk Lee <dhyuklee@pusan.ac.kr> and Bin Zhu <bin.zhu@nih.gov>

See Also

[EstimateSigActivity](#)

Examples

```
data(SimData, package="SALMON")

CalculateSigExpectancy(SimData$L, SimData$TrueW_TMB, SimData$TrueH)
```

EstimateSigActivity   *EstimateSigActivity*

Description

Estimate signature activities

Usage

```
EstimateSigActivity(V, L, W, n.start=50, iter.max=5000, eps=1e-5)
```

Arguments

V	Mutation type matrix or data frame with samples in columns
L	Panel size matrix or data frame with samples in columns
W	Catalog signature profiles matrix or data frame with signatures in columns
n.start	Number of initializations. The default is 50.
iter.max	Maximum number iterations in the EM algorithm. The default is 5000.
eps	Stopping tolerance in the EM algorithm. The default is 1e-5.

Details

For the objects codeV, L, and W, we must have that $dim(V) = dim(L)$ and $ncol(W) = K$, where K is the number of signatures.

Value

A list containing the estimated activity matrix H, the log-likelihood loglike, and the logical value converged.

Author(s)

Donghyuk Lee <dhyuklee@pusan.ac.kr> and Bin Zhu <bin.zhu@nih.gov>

See Also

[CalculateSigExpectancy](#)

Examples

```
    data(SimData, package="SALMON")

  ## Not run:
    EstimateSigActivity(SimData$V, SimData$L, SimData$W_TMB)

## End(Not run)
```

---

mappingSignature   *Find a subset of TMB-based catalog SBS signatures*

Description

This function finds a subset of TMB-based catalog SBS signatures whose linear combination approximate de novo SBS signatures detected by signeR.

Usage

```
 mappingSignature(W_hat, W_ref=NULL, niter=100, cutoff.I2=0.1, min.repeats=80)
```

Arguments

W_hat	Matrix or data frame of de novo signatures from signeR
W_ref	NULL or a matrix or data frame of TMB-based catalog signatures. If NULL, then it will default to SimData$W_TMB (see SimData).
niter	Number of iterations. The default is 100.
cutoff.I2	Cutoff value to select signatures. The default is 0.1.
min.repeats	Minimum number of iterations to select signatures with $I^2 >$ cutoff.I2 . The default is 80.

Value

A data frame with columns COSMIC and freq.

Author(s)

Donghyuk Lee <dhyuklee@pusan.ac.kr> and Bin Zhu <bin.zhu@nih.gov>

SimData	*Data for examples*

Description

Simulated data for examples

Details

This file consists of the list `SimData` with the following objects:

- V : Simulated mutation catalog matrix of size 96 by 10027
- L : Panel size matrix of size 96 by 10027
- W_TMB : Tumor mutation burden based signatures size 96 by 76
- TrueH : Simulated signature activity matrix of size 6 by 10027 used to generate V
- TrueW_TMB : Tumor mutation burden based signatures size 96 by 6 used to generate V

Index