

Package ‘BayesSigRefitting’

April 22, 2024

Title A Bayesian Approach to Select Mutational Signatures

Version 0.0.3

Date 2024-04-22

Description A Bayesian approach to select and refit an optimal set of signatures from a collection of reference signatures. BayesSigRefitting specifies a Bayesian hierarchical model for signature selection, imposes model sparsity using Laplace prior, and utilizes the Shotgun Stochastic Search algorithm to select reference signatures.

Imports stats, parallel, glmnet, prodlim, Matrix

Depends R (>= 4.1.0)

License GPL-2

NeedsCompilation no

Author Min Hua [aut],
Bin Zhu [aut],
Bill Wheeler [cre]

Maintainer Bill Wheeler <wheelerb@imsweb.com>

R topics documented:

BayesSigRefitting-package	1
BayesSigRef	2
BSR.obj	4
Sig.W	4
SigRefitting	5
SimData_TrueH	6
SimData_TrueW	6
SimData_V	7
Index	8

BayesSigRefitting-package
A Bayesian Approach to Select Mutational Signatures

Description

A Bayesian approach to select and refit an optimal set of signatures from a collection of reference signatures. BayesSigRefitting specifies a Bayesian hierarchical model for signature selection, imposes model sparsity using Laplace prior, and utilizes the Shotgun Stochastic Search algorithm to select reference signatures.

Details

Mutational signatures are the patterns of somatic mutations within the cancer genome. Analysis of mutational signatures helps understand causes of cancer and inform treatment strategies. When the sample size is limited, signature refitting is commonly used. It requires selecting a set of reference signatures to be refitted, which has not been sufficiently explored. This package utilizes a Bayesian approach to select and refit an optimal set of signatures from a collection of reference signatures. **BayesSigRefitting** specifies a Bayesian hierarchical model for signature selection, imposes model sparsity using Laplace prior, and utilizes the Shotgun Stochastic Search algorithm to select reference signatures.

The two main functions in this package are [BayesSigRef](#) and [SigRefitting](#).

Author(s)

Min Hua <min.hua@nih.gov> and Bin Zhu <bin.zhu@nih.gov>

BayesSigRef	<i>BayesSigRef</i>
-------------	--------------------

Description

A Bayesian approach to select and refit an optimal set of signatures from a collection of reference signatures.

Usage

```
BayesSigRef(V, Sig.W="SBS", n.iter=3000, tol=0.05, n.core=1, sig.select=0.8)
```

Arguments

V	A numeric matrix with 96 rows and n columns with the rows representing the 96 mutation types and the columns representing the tumors. This is the Single Base Substitution (SBS) mutation counts matrix. The order of the mutation types should match the order of mutation types in COSMIC SBS signatures, users need to rearrange the data if the order does not match.
Sig.W	The string "SBS" or a numeric matrix with 96 rows where the rows represent the 96 mutation types, and the columns represent the signatures. This is the reference signatures from which signatures for refitting will be selected. The default setting is "SBS". If set to "SBS", then the COSMIC SBS 96 V3.1 signatures will be used. Users have the freedom to set it to the reference signatures of their own choice, however, the order of the mutation types should match that of the COSMIC SBS 96 category.
n.iter	The number of iterations the algorithm will repeat to conduct model searching. A larger value gives better convergence, however, requires longer run time. The default is 3000.

<code>tol</code>	The convergence tolerance for the algorithm. A smaller value gives better convergence, however requires a longer run time. The default is 0.05.
<code>n.core</code>	The number of cpu cores to use. The default is 1.
<code>sig.select</code>	The cut-off of the probability the signature will be selected. The default is 0.8, meaning a signature will be selected only if its posterior probability is greater than 0.8. Users can change the value according to their own domain knowledge and/or tolerance.

Details

The algorithm consists of two steps, signature selection and signature refitting. In the first step, BayesSigRef specifies a Bayesian hierarchical model for signature selection, imposes model sparsity using Laplace prior, and utilizes the Shotgun Stochastic Search algorithm to select reference signatures. The second step estimates the signature activities for each tumor using the selected signatures.

Value

A list containing the objects:

- `selection`: A named vector of the selected signatures, where the names are the signatures and the values are the locations of the selected signatures in the reference signatures. For example, (SBS13, 19) means the selected signature is SBS13 and it is the 19th signature in the reference signature set.
- `refitting`: A list containing the results of the signature selection and refitting, `codeW` and `H` matrices. `W` is the corresponding signature matrix, and `H` is the corresponding activity matrix.
- `objects`: A list containing objects needed for calling [SigRefitting](#) and the results of the signature selection step.

Author(s)

Min Hua <min.hua@nih.gov> and Bin Zhu <bin.zhu@nih.gov>

See Also

[SigRefitting](#)

Examples

```
# Importing data
data(SimData_V, package="BayesSigRefitting")

# Signature selection and refitting (the default options could take considerable time)
ref <- BayesSigRef(SimData_V, Sig.W="SBS", n.iter=3000, tol=0.05,
                  n.core=5, sig.select=0.8)

# Show the selected signatures
ref$selection

# Show the W matrix
ref$refitting$W

# Show the H matrix
ref$refitting$H
```

`BSR.obj`*Return object from [BayesSigRef](#)*

Description

An example return object from [BayesSigRef](#).

Details

See [BayesSigRef](#) for details.

See Also

[BayesSigRef](#)

Examples

```
data(BSR.obj, package="BayesSigRefitting")
names(BSR.obj)
```

`Sig.W`*COSMIC SBS Signatures*

Description

The set of 96 COSMIC Single Base Substitution (SBS) Signatures

Details

A 96 x 60 matrix, which contains 60 known reference SBS signatures from COSMIC SBS Mutational Signatures V3.3.1.

See Also

[BayesSigRef](#)

Examples

```
data(Sig.W, package="BayesSigRefitting")

# Display a subset of data
Sig.W[1:5, 1:5]
```

SigRefitting*Signature Refitting*

Description

Estimate the signature activities for each tumor using the selected signatures.

Usage

```
SigRefitting(BSR.obj, sig.select=0.8)
```

Arguments

BSR.obj	The returned object from BayesSigRef .
sig.select	The lower bound to select signatures. The default is 0.8.

Details

This function allows users to change the sig.select option without re-running the signature selection.

Value

A list containing the objects:

- selection: A named vector of the selected signatures, where the names are the signatures and the values are the locations of the selected signatures in the reference signatures. For example, (SBS13, 19) means the selected signature is SBS13 and it is the 19th signature in the reference signature set.
- W: A numeric matrix with 96 rows and k columns containing the selected signatures, where k is the number of the selected signatures.
- H: A numeric matrix with k rows containing the signature activities of the selected signatures in W.

Author(s)

Min Hua <min.hua@nih.gov> and Bin Zhu <bin.zhu@nih.gov>

See Also

[BayesSigRef](#)

Examples

```
data(BSR.obj, package="BayesSigRefitting")  
ret <- SigRefitting(BSR.obj, sig.select=0.95)
```

SimData_TrueH	<i>Signature Activity Matrix</i>
---------------	----------------------------------

Description

A signature activity matrix of dimension 12 x 200 used in the construction of [SimData_V](#).

See Also

[SimData_TrueW](#), [SimData_V](#)

Examples

```
data(SimData_TrueH, package="BayesSigRefitting")

# Display a subset of data
SimData_TrueH[1:5, 1:5]
```

SimData_TrueW	<i>Profile Matrix of Signatures</i>
---------------	-------------------------------------

Description

A 96 x 12 profile matrix of predefined ground truth signatures.

Details

The 96 x 12 profile matrix of the predefined ground truth signatures (SBS1, SBS2, SBS3, SBS5, SBS6, SBS8, SBS13, SBS17b, SBS18, SBS20, SBS26, SBS30) used in the construction of [SimData_V](#).

See Also

[SimData_V](#), [SimData_TrueH](#)

Examples

```
data(SimData_TrueW, package="BayesSigRefitting")

# Display a subset of data
SimData_TrueW[1:5, 1:5]
```

SimData_V*Mutation Catalog Matrix V*

Description

A 96 x 200 matrix of simulated mutation counts.

Details

The matrix contains 200 tumors for 96 single base substitution (SBS) types. The simulated mutation counts mimic the real data in the BRCA560 breast cancer dataset. Each element in the matrix is generated from the Poisson distribution with the mean corresponding to each element of $W_t H_t$, where W_t is the 96 x 12 profile matrix of the predefined ground truth signatures (SBS1, SBS2, SBS3, SBS5, SBS6, SBS8, SBS13, SBS17b, SBS18, SBS20, SBS26, SBS30), and H_t is the 12 x 200 signature activity matrix. The matrices W_t and H_t are also included in the data folder as [SimData_TrueW](#) and [SimData_TrueH](#).

See Also

[SimData_TrueW](#), [SimData_TrueH](#)

Examples

```
data(SimData_V, package="BayesSigRefitting")

# Display a subset of data
SimData_V[1:5, 1:5]
```

Index

* data

BSR.obj, [4](#)

Sig.W, [4](#)

SimData_TrueH, [6](#)

SimData_TrueW, [6](#)

SimData_V, [7](#)

* mutational signatures

BayesSigRef, [2](#)

SigRefitting, [5](#)

* package

BayesSigRefitting-package, [1](#)

BayesSigRef, [2](#), [2](#), [4](#), [5](#)

BayesSigRefitting

(BayesSigRefitting-package), [1](#)

BayesSigRefitting-package, [1](#)

BSR.obj, [4](#)

Sig.W, [4](#)

SigRefitting, [2](#), [3](#), [5](#)

SimData_TrueH, [6](#), [6](#), [7](#)

SimData_TrueW, [6](#), [6](#), [7](#)

SimData_V, [6](#), [7](#)