Package ‘plasmut’

February 22, 2024

Type  Package
Title  Stratifying mutations observed in cell-free DNA and white blood cells as germline, hematopoietic, or somatic
Version  1.0.0
Description  A Bayesian method for quantifying the likelihood that a given plasma mutation arises from clonal hematopoiesis or the underlying tumor. It requires sequencing data of the mutation in plasma and white blood cells with the number of distinct and mutant reads in both tissues. We implement a Monte Carlo importance sampling method to assess the likelihood that a mutation arises from the tumor relative to non-tumor origin.
License  Artistic-2.0
Encoding  UTF-8
biocViews  Bayesian, SomaticMutation, GermlineMutation, Sequencing
RoxygenNote  7.2.3
Imports  tibble, stats, dplyr
Depends  R (>= 4.3.0)
Suggests  knitr, markdown, tidyverse, ggrepel, magrittr, qpdf, BiocStyle, biocViews, testthat (>= 3.0.0)
VignetteBuilder  knitr
Config/testthat/edition  3
git_url  https://git.bioconductor.org/packages/plasmut
git_branch  RELEASE_3_18
git_last_commit  43c16fe
git_last_commit_date  2023-10-24
Repository  Bioconductor 3.18
Date/Publication  2024-02-21
Author  Adith Arun [aut, cre], Robert Scharpf [aut]
Maintainer  Adith Arun <adith.3.arun@gmail.com>
**R topics documented:**

- `crcseq` ................................................. 2
- `importance_sampler` .................................. 2
- `model_w` ............................................. 3
- `plasma_somatic` ..................................... 4
- `plasmut` .............................................. 5
- `wbc_somatic` ........................................... 5

**Index**

| crcseq | This data is an example dataset to show how to use the package |

**Description**

A cohort of metastatic colorectal cancer patients whose plasma and buffy coat were sequenced as part of the CAIRO5 trial. The cohort and analyses are described here: [https://pubmed.ncbi.nlm.nih.gov/36534496/](https://pubmed.ncbi.nlm.nih.gov/36534496/)

**Value**

An example DNA sequencing dataset of matched plasma and wbc colorectal cancer samples `crcseq`

**importance_sampler**

*Importance sampler to estimate marginal likelihoods and Bayes factors*

**Description**

Importance sampler to estimate marginal likelihoods and Bayes factors

**Usage**

`importance_sampler(dat, params, save_montecarlo = TRUE)`

**Arguments**

- `dat` data frame with observed mutant and total counts and the analyte (plasma or buffy coat) it was taken from and the identifiers on what the mutation is (e.g., KRASG12C) and pt id
- `params` list with ctc, ctdna and chip a and b beta parameters reflect beliefs on what fraction of fragments belong to each class; montecarlo.samples being the number of MC samples; prior weight is the prior.weight reflects how much importance sampling to implement, closer to zero means more importance density considered
- `save_montecarlo` save more indepth monte carlo results
implement importance sampling for a data set to assess probability of tumor derived mutations from sequencing results

Examples

```r
param.list <- list(ctc=list(a=1, b=9999),
                   ctdna=list(a=1, b=9),
                   chip=list(a=1, b=9),
                   montecarlo.samples=50e3,
                   prior.weight=0.1)

dat <- data.frame(y=c(4, 1),
                   n=c(1000, 1000),
                   analyte=c("plasma", "buffy coat"),
                   mutation="mutA",
                   sample_id="id1")

importance_sampler(dat, param.list)
```

Description

Estimate the marginal likelihood that mutations in buffy coat and cfDNA reflect CH or correspond to germline mutations. If germline, the allele frequency should be 50 percent. The prior should be diffuse enough to handle CHIP mutations which are potentially way less than 50 percent

Usage

```r
model_w(dat, params)
```

Arguments

dat tibble containing vectors y and n. y and n should be named

params a list with named elements that must include the following a which is the prior expectation for number of CH or germline variants observed in the sequencing data b which is the prior expectation for number of fragments reflecting CH or germline

Value

list of samples, probability densities, and likelihood for non-tumor assumption
Examples

```r
param.list <- list(ctc=list(a=1, b=9999),
                   ctdna=list(a=1, b=9),
                   chip=list(a=1, b=9),
                   montecarlo.samples=50e3,
                   prior.weight=0.1)
dat <- data.frame(y=c(4, 1),
                  n=c(1000, 1000),
                  analyte=c("plasma", "buffy coat"),
                  mutation="mutA",
                  sample_id="id1")
model_w(dat, param.list)
```

### plasma_somatic

Estimate the marginal likelihood that variants identified in cell-free DNA are derived from tumor cells (ctDNA-derived)

#### Description

Estimate the marginal likelihood that variants identified in cell-free DNA are derived from tumor cells (ctDNA-derived)

#### Usage

```r
plasma_somatic(dat, params)
```

#### Arguments

- **dat**
  - A tibble containing vectors `y` and `n`; `y` and `n` should be named
- **params**
  - A list with named elements that must include the following:
    - `a`: prior expectation for number of plasma somatic variants observed in the plasma sequencing data
    - `b`: prior expectation for number of plasma fragments not containing variants

#### Value

generate importance samples for plasma somatic model

#### Examples

```r
param.list <- list(ctc=list(a=1, b=9999),
                   ctdna=list(a=1, b=9),
                   chip=list(a=1, b=9),
                   montecarlo.samples=50e3,
                   prior.weight=0.1)
dat <- data.frame(y=c(4, 1),
                  n=c(1000, 1000),
                  analyte=c("plasma", "buffy coat"),
                  mutation="mutA",
                  sample_id="id1")
model_w(dat, param.list)
```
plasmut

```r
mutation="mutA",
sample_id="id1")
plasma_somatic(dat, param.list)
```

---

**plasmut**

*Bayesian models for estimating the origin of a sequenced DNA fragment*

**Description**

The plasmut package provides a Bayesian importance sampling based approach to estimate the likelihood of a mutation arising from clonal hematopoiesis or tumor.

---

**wbc_somatic**

*Estimate the marginal likelihood of observing somatic mutations from CTCs present in buffy coat*

\[
p(y_w | \theta_w, n_w, model_S) \times p(\theta_w | Model_S) \theta_w | model_S \sim \text{beta}(1, 999) \]

## sequencing error or CTC

**Description**

Estimate the marginal likelihood of observing somatic mutations from CTCs present in buffy coat

\[
p(y_w | \theta_w, n_w, model_S) \times p(\theta_w | Model_S) \theta_w | model_S \sim \text{beta}(1, 999) \]

## sequencing error or CTC

**Usage**

`wbc_somatic(dat, params)`

**Arguments**

- `dat` : tibble containing vectors 'y' and 'n'; 'y' and 'n' should be named
- `params` : a list with named elements that must include the following: 'a': prior expectation for number of somatic variants observed in the WBC sequencing data (either by error or from a CTC) 'b': prior expectation for number of WBCs not containing the variant

**Value**

generate importance samples for wbc somatic model
Examples

```r
param.list <- list(ctc=list(a=1, b=9999),
       ctdna=list(a=1, b=9),
       chip=list(a=1, b=9),
       montecarlo.samples=50e3,
       prior.weight=0.1)

dat <- data.frame(y=c(4, 1),
       n=c(1000, 1000),
       analyte=c("plasma", "buffy coat"),
       mutation="mutA",
       sample_id="id1")
wbc_somatic(dat, param.list)
```
Index

crcseq, 2
importance_sampler, 2
model_w, 3
plasma_somatic, 4
plasmut, 5
wbc_somatic, 5