Package ‘Famdenovo’

Type Package
Title TP53 mutation carrier estimation
Version 0.1.0
Date 2018-02-04
Author Fan Gao, Wenyi Wang, Xuedong Pan
Maintainer Fan Gao <fgao3@mdanderson.org>, Wenyi Wang <wwang7@mdanderson.org>
Description This is a software that predicts the de novo status of a germline mutation in familial diseases based on family history. Currently we apply Famdenovo to the de novo TP53 mutations.
License GPL-3

How to install

Step 1. Install dependency “LFSPRO” package. If you have not installed "LFSPRO", download it from "http://bioinformatics.mdanderson.org/main/LFSPRO" and install it from local source in your R console by typing:

```
install.packages("where_you_saved_the_file/LFSPRO_0.1.5.tar.gz", repos =
NULL, type = "source")
```

Or, install from GitHub:

```
library(devtools)
install_github("wwylab/LFSPRO")
```

Step 2. Install dependency “Famdenovo” package. if you have not installed "Famdenovo", download it from "http://bioinformatics.mdanderson.org/main/Famdenovo" and install it from local source in your R console by typing:

```
install.packages("where_you_saved_the_file/Famdenovo_0.1.0.tar.gz", repos =
NULL, type = "source")
```

Or, install from GitHub:

```
library(devtools)
install_github("wwylab/Famdenovo")
```

How to use

Step 1. Load package

```
library(Famdenovo)
```

Step 2. Call "Famdenovo()" function

```
Famdenovo(family, cancer, mutation, person.id, gene = "TP53")
```

Format of the input files

Famdenovo requires three data sets as input: family, cancer, mutation, person.id, mutation, and gene.

family: family Information Data

The input should be a data frame. The family data should include the following columns with the corresponding column names:
id: index of the person. All individuals should from one family.
fid: index of the person's father. If the individual is the founder of the pedigree, set it as NA.
mid: index of the person's mother. If the individual is the founder of the pedigree, set it as NA.
gender: gender of the person. 0 - female; 1 - male
age: age of the person. If the individual is alive, it is as the current age. Otherwise, set it as the age of death.

Example Code:

```
Data("TP53.test1.family")
TP53.test1.family
```

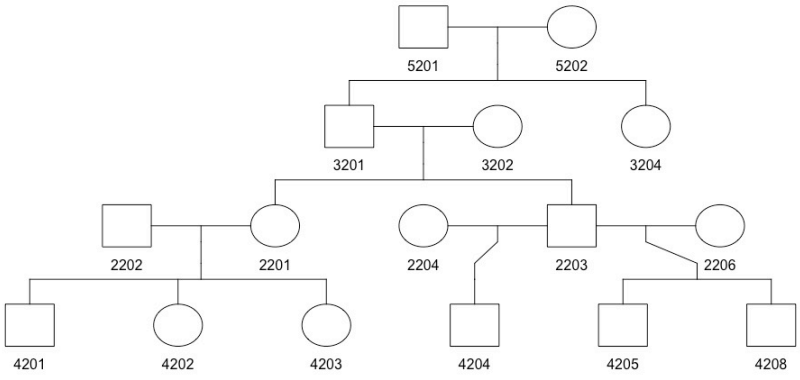

Figure 1. An example of a pedigree structure

	id	fid	mid	gender	age
1	2201	3201	3202	0	49
2	2202	NA	NA	1	47
3	2203	3201	3202	1	47
4	2204	NA	NA	0	45
5	2206	NA	NA	0	43
6	3201	5201	5202	1	68
7	3202	NA	NA	0	70
8	3204	5201	5202	0	78
9	4201	2202	2201	1	26
10	4202	2202	2201	0	23
11	4203	2202	2201	0	21
12	4204	2203	2204	1	12
13	4205	2203	2206	1	3
14	4208	2203	2206	1	2
15	5201	NA	NA	1	35
16	5202	NA	NA	0	48

Figure 2. The example of the family data for a pedigree in Figure 1.

cancer: Cancer Information Data

The input should be a data frame. The cancer data should include the following columns with the corresponding column names:

id: index of the person

cancer.type: type of the cancer. We divided all the cancers into 11 groups according to NCCCN Guidelines Version 1.2012 Li-Fraumeni Syndrome criteria. Check "LFSpro.cancer.type" for details.

diag.age: The age when the individual was diagnosed with cancer.

Example Code:

```
Data("TP53.test1.cancer")
TP53.test1.cancer
```

	id	cancer.type	diag.age
1	2201	breast	41
2	2201	breast	41
3	2201	sts	39
4	2203	non.lfs	43
5	3201	non.lfs	60
6	3201	non.lfs	50
7	3201	ost	23
8	3201	non.lfs	63
9	3201	sts	67
10	3204	breast	61
11	5202	non.lfs	47
12	5202	non.lfs	30

Figure 3. An example of the cancer file

mutation: Mutation Information Data

The input should be a data frame. The mutation data should include the following columns with the corresponding column names:
id: index of the person
mut.state: mutation status of the person. "W" - wild type; "M" - mutated. Individuals who are not sequenced are not included in the mutation information data.

Example Code:

```
Data("TP53.test1.mutation")
TP53.test1.mutation
```

```
    id mut.state
1 2201         M
2 2203         M
3 3201         M
4 3204         W
5 4204         W
```

Figure 4. An example of the mutation file

person.id

The input should be either character string(s) or numrical value(s) of the person(s) you want to analyze.

Gene

The input should be character string(s). The default value is "TP53". We will add other genes in the future.

Format of the output file

The output is the probability of any TP53 mutation being de novo, one TP53 mutation carrier per line. Each line contains three elements: "family id", "individual id" and "prob.denovo", respectively.

Here is an example:

```
 [1] "The following ids are not carriers: 1002, 1003"
   id   prob.denovo
1 2201 0.0001205471
2 2203 0.0001105599
3 3201 0.0126455556
```