How to Use LFSPRO

Jasper Chen, Gang Peng

2/14/2019

R Installation

LFSPRO is an R package, must be run in an R environment. You need to install R/Rstudio first from the following link: http://cran.rstudio.com/. RStudio is just an IDE (Integrated Development Environment) in which you are using the R language.

Download R/Rstudio according to your operating system. Make sure you download the latest version.

- Windows:

  Download the windows version of R. It is an .exe file. After downloading, double- click the exe file to install R.

- Mac OS:

  Download the Mac OS version of R. It is a .pkg file. After downloading, double- click the pkg file to install R.

- Linux or Unix:

    - Ubuntu or Debian

  ```
  $ sudu apt-get install r-base
  ```
    - Red Hat or Fedora:

  ```
  $ sudo yum install R.i386
  ```

LFSPRO Installation

You can download LFSPRO either from GitHub (recommended) directly or from the MD Anderson LFSPRO webpage.

1. GitHub (Recommended)

You can type the codes below in the R console to download and install LFSPRO package from GitHub as picture.

```
install.packages("devtools")
devtools::install_github("wwylab/LFSPRO")
```

Download LFSPRO from GitHub

2. LFSPRO Webpage

You can download LFSPRO package from https://bioinformatics.mdanderson.org/public-software/lfspro/ according to your operation system.

Windows

Run R
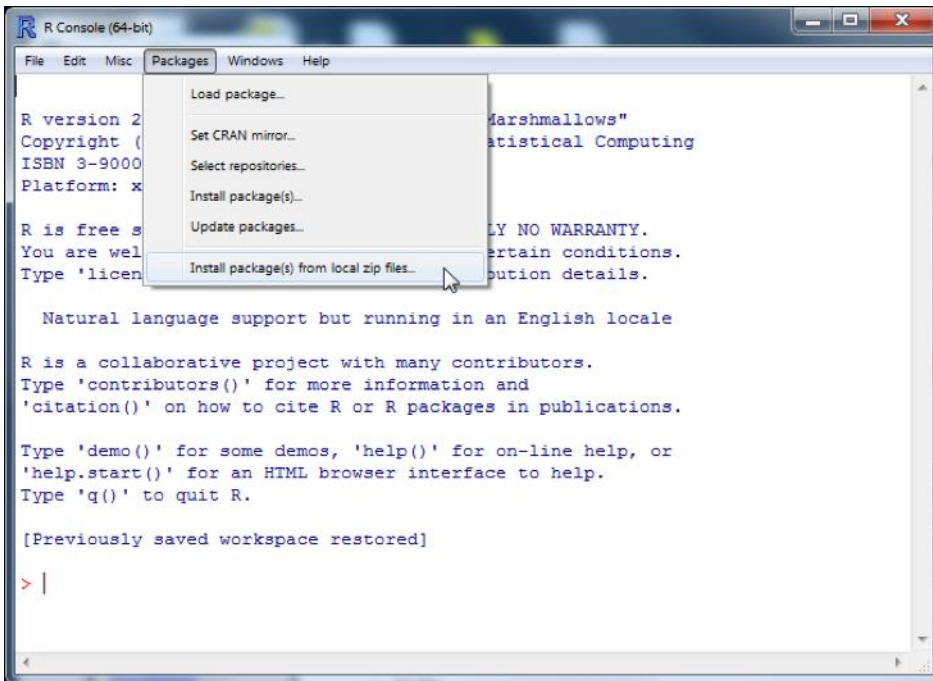
Then click drop-down menu "Package", "Install package(s) from local zip files…"

Choose the `LFSPRO_2.0.0.zip` you just downloaded. Or you can type `install.packages("YOUR/PATH/LFSPRO_2.0.0.zip")` in the R console to install the package.

Mac OS
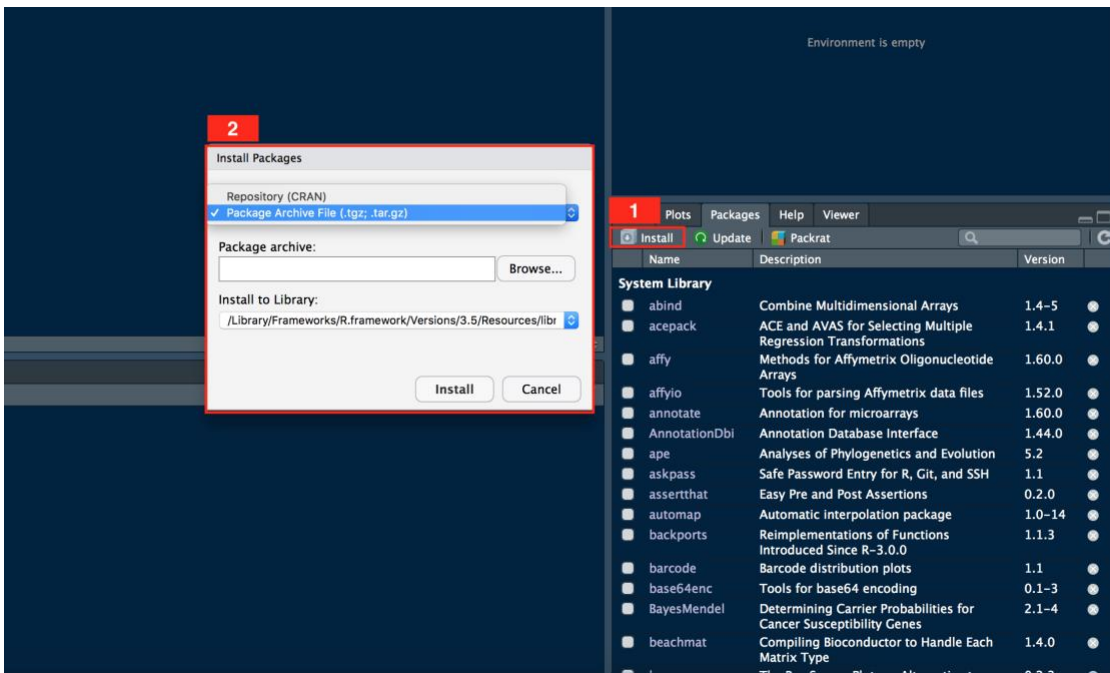
Run R, then click "Install" and upload the downloaded LFSPRO .tgz file

Or you can type `install.packages("YOUR/PATH/LFSPRO_2.0.0.tgz")` in the R console to install the package.

Linux and Unix

After running R, you can type `install.packages("YOUR/PATH/LFSPRO_2.0.0.tar.gz")` to install the package. Or you can type `R CMD INSTALL PATH/LFSPRO_2.0.0.tar.gz` to install the package in Linux/Unix console.

Using LFSPRO in R

Input Data of LFSPRO

After installing, type `library("LFSPRO")` in the R console to load the package.

To run LFSPRO, you need to prepare two input datasets: family information data and cancer information data. Type `fam.data` and `cancer.data` to load built-in dataset examples.

There should be 6 columns in family information data.

- fam.id: family ID
- id: individual id
- fid: father ID
- mid: mother ID
- gender: gender. 0: female, 1: male
- age: If the individual was alive at the last contact date, use the age at that time. otherwise, use the age at death.

For cancer information data, there should be 4 columns.

- fam.id: family id
- id: Individual id
- cancer.type: We classified the cancers into 11 groups as below according to NCCN Guidelines Version 1.2012 Li-Fraumeni Syndrome criteria.
    - sts (soft tissue sarcoma)
    - ost (osterosarcoma)
    - brain(brain tumor other than choroid plexus carcinoma)
    - breast (breast cancer)
    - acc (adrenocortical carcinoma)
    - leukemia
    - lung (lung bronchoalveolar cancer)
    - choroid (choroid plexus carcinoma)
    - other.lfs (other LFS spectrum cancers)
    - non.lfs (non LFS spectrum invasive cancers) benign(benign tumors)

For example, if the patient has soft tissue sarcoma, just use `sts` in this column.

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

Input Data Preparation

You can prepare your own family information data and cancer information data with Excel and save them into separate .csv files. (`NA` for missing data)

For example:

fam.data.csv

fam.id	id	fid	mid	gender	age
fam1	0	5	6	0	24
fam1	1	NA	NA	1	35
fam1	2	NA	NA	0	77
fam1	3	NA	NA	1	63
fam1	4	NA	NA	0	87
fam1	5	1	2	1	53
fam1	6	3	4	0	74
fam1	9	5	6	1	13
fam1	10	NA	NA	1	34
fam1	11	5	6	0	21
fam1	12	5	6	1	5
fam1	13	5	6	1	41
fam1	14	10	11	1	2

cancer.data.csv

fam.id	id	cancer.type	diag.age
fam1	0	ost	15
fam1	0	breast	23
fam1	5	brain	50
fam1	5	lung	50
fam1	9	ost	11
fam1	12	ost	4

Load family and cancer data in R.

```
fam.data <- read.csv("YOUR/PATH/fam.data.csv",
                     colClasses =
c("character","integer","integer","integer","numeric","numeric"), header = TRUE)

cancer.data <- read.csv("YOUR/PATH/cancer.data.csv",
                     colClasses = c("character","integer","character","numeric"),
header = TRUE)
```

NOTE: if the two csv files are not in the R working directory, please add YOUR PATH to the file name.

Estimate TP53 Mutation Carrier Probability and Future Risks

Now you are ready to use LFSPRO to calculate TP53 mutation carrier probability and predict future cancer risk!

Simply, tell R whose probabilities that you want to calculate by specifying the counselee's family ID and individual ID.

If you want to check the status of multiple individuals, you can generate a dataframe of counselee ID as following:

```
fam.id <- c("fam1","fam2","fam2","fam2","fam2")
id <- c(0,0,2,100,200)
counselee.id <- data.frame(fam.id, id)
```

In the new `lfspro.mode` function, we integrated different in one function. You can specify what you want to predict by selecting mode, "1st.all", "1st.cs" or "mpc".

- "1st.all" for only predicting the probability of carrying TP53 germline mutation (Peng, G., 2017)
- "1st.cs" for predicting cancer-specific risks (Shin, S. J, 2018)
- "mpc" for predicting risks of developing multiple primary cancers (Shin, S. J, 2018).

```
lfspro.mode(fam.data, cancer.data,counselee.id,"1st.all")
lfspro.mode(fam.data, cancer.data,counselee.id,"1st.cs")
lfspro.mode(fam.data, cancer.data,counselee.id,"mpc")
```

Note: TP53 mutation allele frequency (MAF) has been set as 0.0006 as default (Peng, G., 2017).

TP53 mutation rate (mRate) has been set as 0.00012 as default.

It is recommended to use the default parameters, but you can specify the mutation allele frequency and mutation rate as you need with code as below before.

allef <- list(c(1-MAF, MAF))

mRate <- 5e-4

The output consists of posterior probability and future cancer risks. If the posterior probability is larger than a cut off (0.2 is recommended here), then probably he is a TP53 mutation carrier.

Of course, you can also use the following two commands to use the LFS classic (Li, F., et al., 1988) and Chompret criteria (Chompret, A., et al., 2001):

```
lfsClassic(fam.data, cancer.data, counselee.id)
lfsChompret(fam.data, cancer.data, counselee.id)
```

You can find more introduction to R in http://cran.r-project.org/doc/manuals/R-intro.pdf

And also check the details of LFSPRO package in the LFSPRO R manual or in the LFSPRO help file in R (`?LFSPRO`).

References

Peng, G., Bojadzieva, J., Ballinger, M. L., Li, J., Blackford, A. L., Mai, P. L., Savage, S. A., Thomas, D. M., Strong, L. C., ... Wang, W. (2017). Estimating TP53 Mutation Carrier Probability in Families with Li-Fraumeni Syndrome Using LFSPRO. , , 837-844. DOI: 10.1158/1055-9965.EPI-16-0695

Shin, S. J., Ning, J., Bojadzieva, J., Strong, L. C., and Wang, W. (2018). Bayesian estimation of a semiparametric recurrent event model with applications to the penetrance estimation of multiple primary cancers in Li-Fraumeni syndrome. , 00, 1–16. https://doi.org/10.1093/biostatistics/kxy066

Shin, S. J., Yuan, Y., Strong, L. C., Bojadzieva, J., and Wang, W. (2018). Bayesian Semiparametric Estimation of Cancer-Specific Age-at-Onset Penetrance With Application to Li-Fraumeni Syndrome. , 1–12. https://doi.org/10.1080/01621459.2018.1482749

Li, F. P., et al. (1988). A cancer family syndrome in twenty-four kindreds. , : 5358-5362.

Chompret, A., et al. (2001). Sensitivity and predictive value of criteria for p53 germline mutation screening. , : 43-47.

Tinat, J., et al. (2009). 2009 version of the Chompret criteria for Li Fraumeni syndrome. , : e108-109.

Bougeard, G., et al.(2015). Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. , : p. 2345-52.