

How to Use LFSPRO

R installation

LFSPRO is an R package. We must run it in an R environment. We need to install R first.

We can download R from the following link:

<http://cran.rstudio.com/>

Download R 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:

```
$ sudo apt-get install r-base
```

Red Hat or Fedora

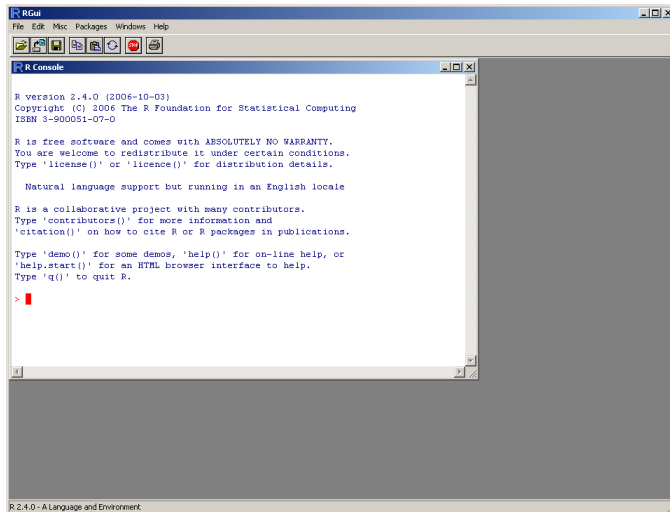
```
$ sudo yum install R.i386
```

LFSPRO installation

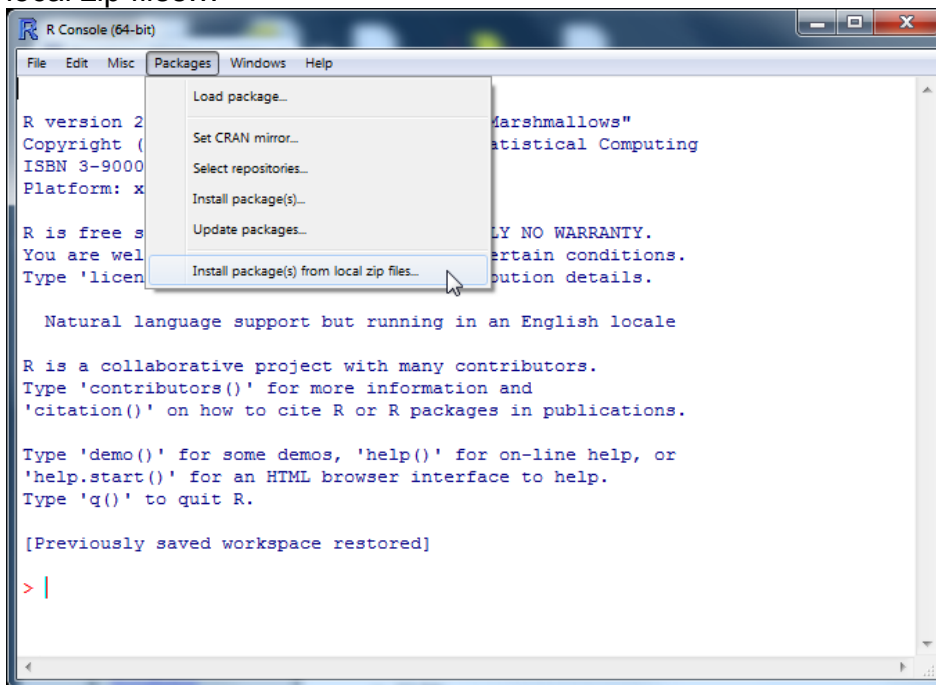
LFSPRO is an R package. We can download it from <http://bioinformatics.mdanderson.org/main/LFSPRO>

Windows

First run R from start menu, it looks like:



Then we can install LFSPRO from the menu “Package/Install package(s) from local zip files...”



Choose the LFSPRO_1.0.4.zip you just downloaded to install the package. Or you can type `install.packages("PATH/LFSPRO_1.0.4.zip")` in the R console to install the package.

Mac OS

First run R from Applications, it looks like

```
R Console
R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin10.8.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

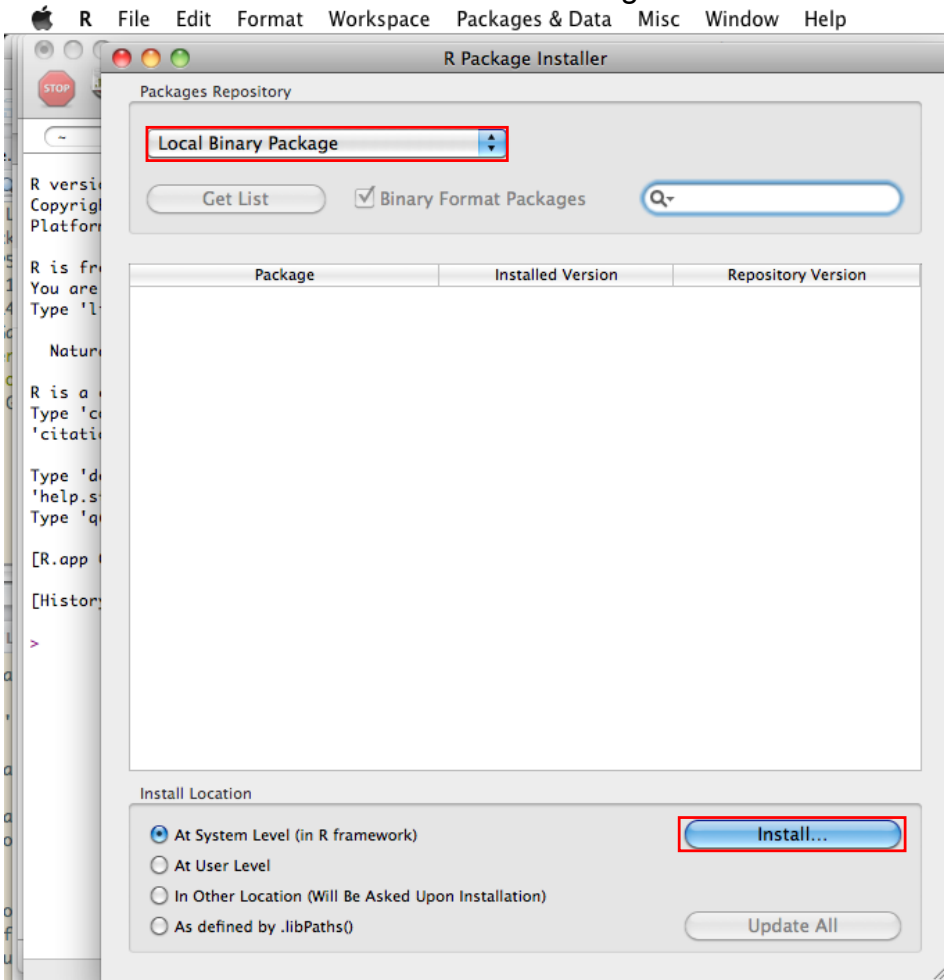
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.65 (6784) x86_64-apple-darwin10.8.0]
[History restored from /Users/gpeng1/.Rapp.history]
>|
```

Then we can install R package from menu “Packages & Data/Package installer”. There will be a new window like the following:



Choose Local Binary Package from the top left of the window and then click the Install button at the bottom. Choose LFSPRO_1.0.4.tgz you just downloaded to install the package. Or you can type `install.packages("PATH/LFSPRO_1.0.4.tgz")` in the R console to install the package.

Linux and Unix

After starting R, you can type `install.packages("PATH/LFSPRO_1.0.4.tar.gz")` to install the package. Or you can type `R CMD INSTALL PATH/LFSPRO_1.0.4.tar.gz` to install the package in Linux/Unix console.

Using LFSPRO in R

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

In LFSPRO, we need two data sets as input: family information data and cancer information data. They are two tables.

You can type

`fam.data`

and

`cancer.data`

in R to check the sample data.

For family information data, there are 6 columns:

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 are 4 columns:

fam.id

Family id

id

Individual id

cancer.type

Cancer type. We classified the cancers into 11 groups according to NCCCN 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 put sts in this column.

diag.age

The age when the individual was diagnosed with cancer

You can prepare your data in two csv files.

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

fam1,14,leukemia,2
fam1,36,non.lfs,24
fam1,37,non.lfs,54
fam2,0,ost,14
fam2,0,breast,25
fam2,104,non.lfs,55
fam2,205,non.lfs,58
fam2,205,breast,60
fam2,300,non.lfs,87

Then read these to two data sets into R.

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

```
cancer.data <- read.csv("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 PATH to the file name.

After preparing the data, you can use the following code to run LFSPRO:

```
allef.g <- list(c(0.9999,0.0001))  
mRate.g <- 5e-4  
options(stringsAsFactors = FALSE)  
fam.id <- c("fam1", "fam2", "fam2", "fam2", "fam2")  
id <- c(0,0,2,100,200)  
counselee.id <- data.frame(fam.id, id)  
lfspro(fam.data, cancer.data, LFSpenet.2010, counselee.id, allef.g, 1, mRate.g)
```

Let's used the sample data as an example. If you want to check the status of sample 0 in fam1 and sample 0, 2, 100, 200 in fam2, you first need to generate a data frame of counselee.id as following:

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

Then set allele frequency in population

```
allef.g <- list(c(0.9999,0.0001))
```

Here the TP53 mutation allele frequency is set to 0.0001.

Then set the de novo mutation rate (per gene per generation)

```
mRate.g <- 5e-4
```

It is recommended to use the above two parameters, but you can change the above two parameters according to your need.

After that, type

```
rlt <- lfspro(fam.data, cancer.data, LFSpenet.2010, counselee.id, allef.g,  
1,mRate.g)
```

to estimate the posterior probability for TP53 mutation. If it is larger than a cut off (0.2 is recommended here), this individual is a TP53 mutation carrier.

LFSpenet.2010 is the penetrance of TP53 mutation that has been included in the library.

Of course, you can also use the following two command to test the LFS classic and Chompret criteria:

```
LFSClassic(fam.data, cancer.data, counselee.id)  
LFSChompret(fam.data, cancer.data, counselee.id)
```

You can find an introduction to R in

<http://cran.r-project.org/doc/manuals/R-intro.pdf>

And you can also check the details of LFSPRO package in the LFSPRO R manual or in the LFSPRO help file in R.