DMDtoolkit manual

Jul, 2016

The functions of DMDtoolkit include:

1) assisted diagnosis for DMD / BMD using genetic testing;

2) drawing the mutated protein sequence and motifs;

3) drawing pedigree of DMD family;

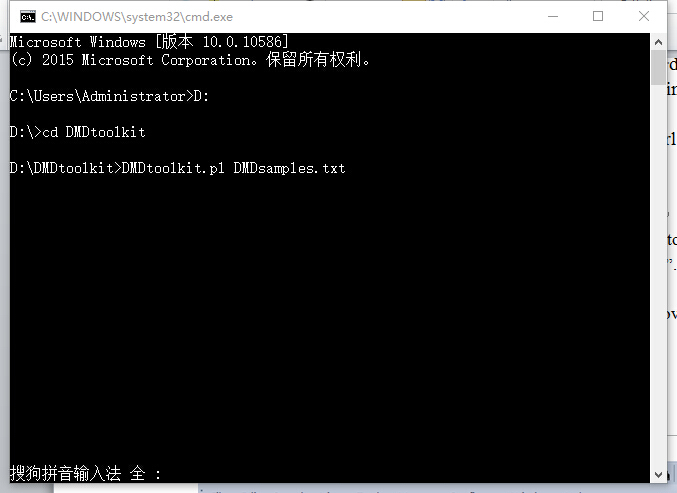
4) smartly screening the data to maximize the use of existing data;

5) performing statistics for the DMD population and visualizing the results.

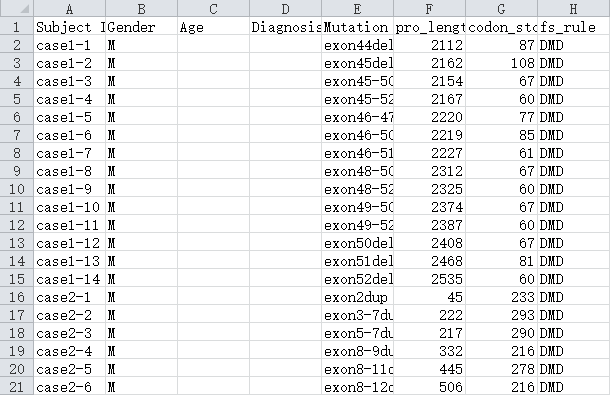
Note: Please install R (https://cran.r-project.org/) and Perl (https://www.perl.org/) before running the following commands.

1. assisted diagnosis for DMD / BMD using genetic testing

For Windows users, open the dos/cmd window, and move to the working directory, e.g. D:/DMDtoolkit by typing “D:” and “cd DMDtoolkit”. Then use the command: DMDtoolkit.pl DMDsamples.txt. For Linux/Unix users, open the terminal window and move to the working directory, then use the same command.



After several seconds, you will get six output files: “DMDsamples.Dp427m.\*”(rdata/pros/stats/diag/diag2/diag3) and open the diag3 file by Excel/WPS to see the diagnosis.

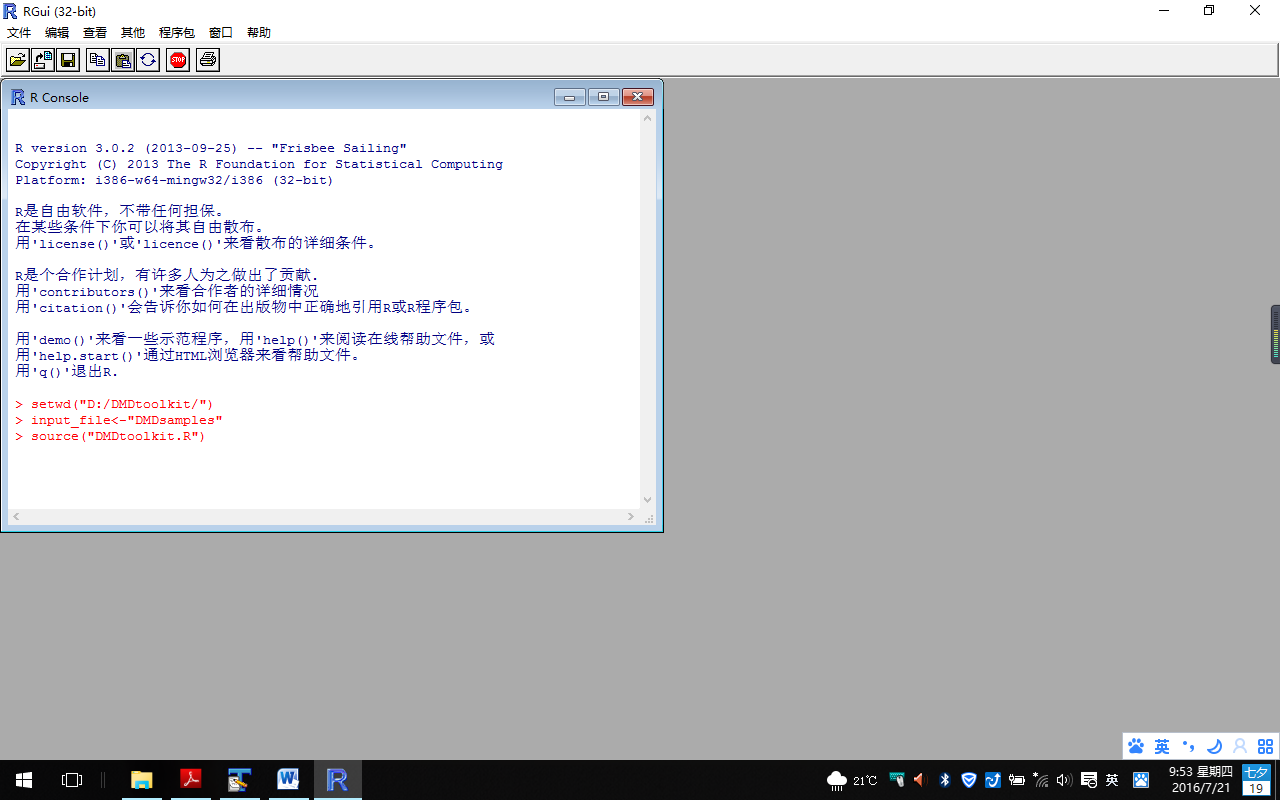


You can create your own input file according to the format of DMDsamples.txt. Five columns are required: Subject ID, Gender, Age, Diagnosis and Mutation, separated by a tab. Missing data is allow.

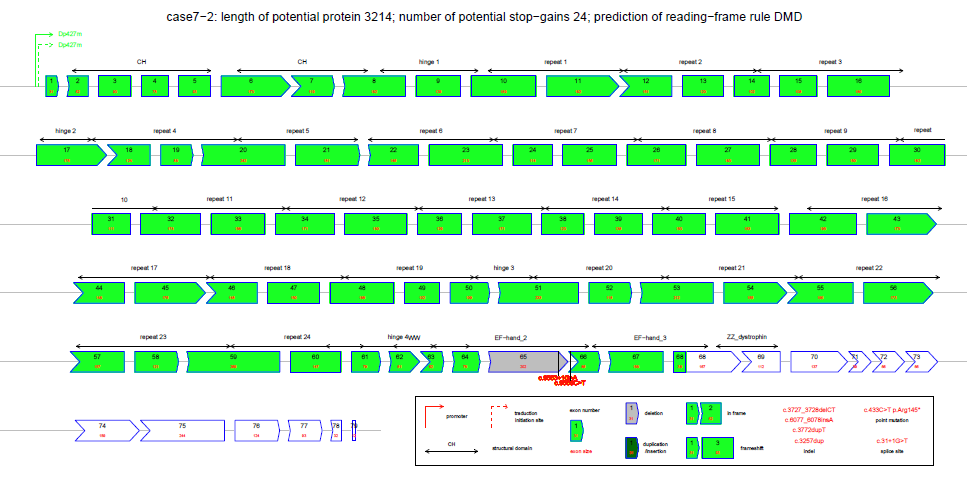
Note: Please ensure that the following files --- “codon list.txt”, “DMD gene.fa”, “Dp427m CDs.fa”, “Dp427m protein.fa”, “Dp427m CDs.txt”, “Dp427m Domains.txt” and “ESE matrices.txt” --- are in the same working directory.

1. drawing the mutated protein sequence and motifs

Use the commands under the R console: setwd(“the/working/directory”) to move to the working directory; input\_file<-“file name” to read the input file; source(“DMDtoolkit.R”) to perform the graphing.

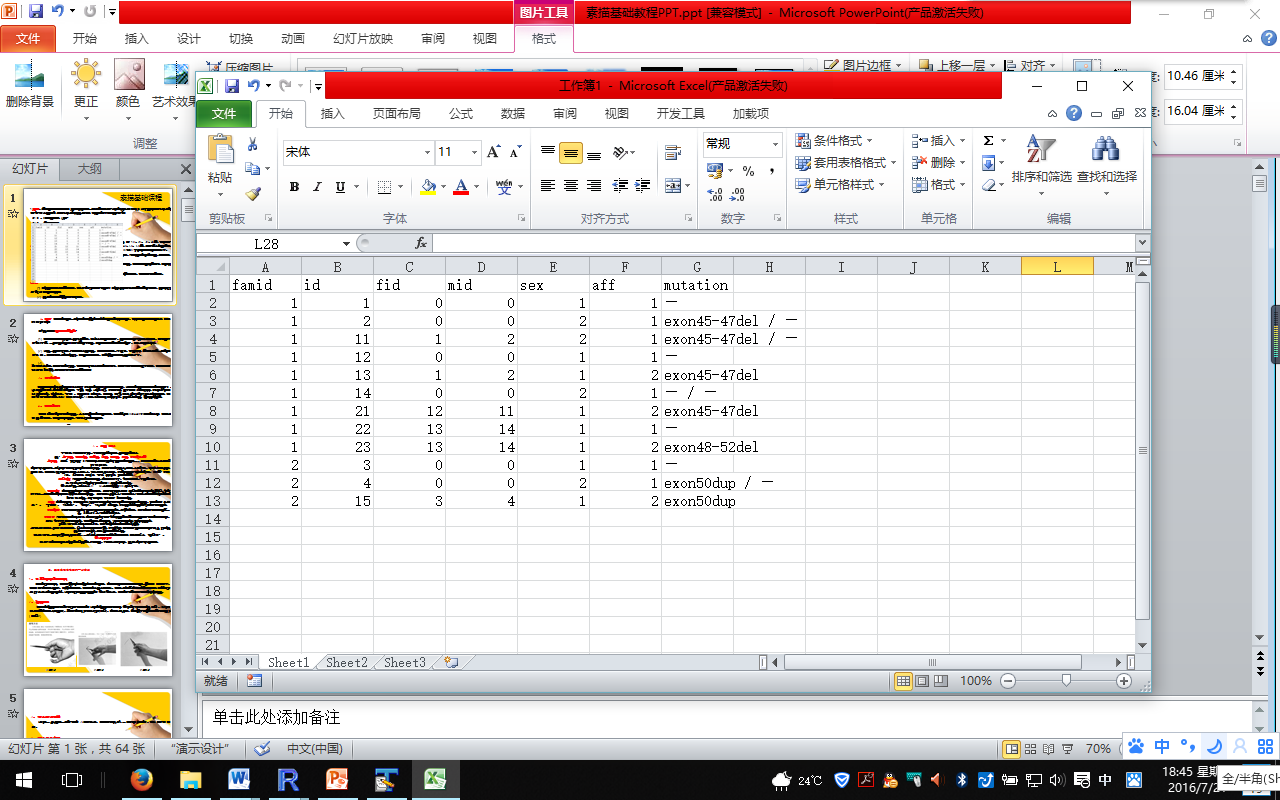


This command will call the files “DMDsamples.Dp427m.\*”(rdata/pros/stats/diag/diag2/diag3) and create 64 graph files in pdf format automatically. One of them is as follows:

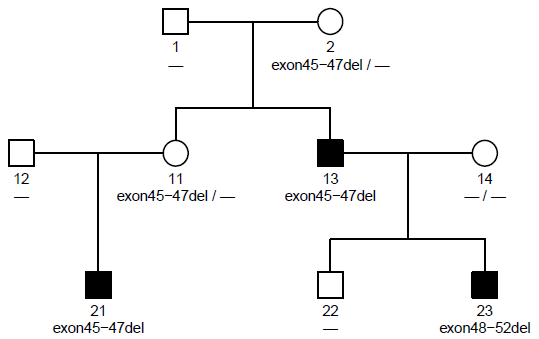


1. drawing pedigree of DMD family

Use the commands under the R console: setwd(“the/working/directory”) to move to the working directory; source(“DMDtoolkit.R”) to call the program; plot.ped("file name") to draw the pedigrees in pdf format automatically. Take “pedigree.txt” as an example. Seven columns are required: famid --- family ID, id --- individual ID, fid --- father ID, mid --- mother ID, sex, aff --- affected or not (1 no / 2 yes), and mutation. Missing data is allow.

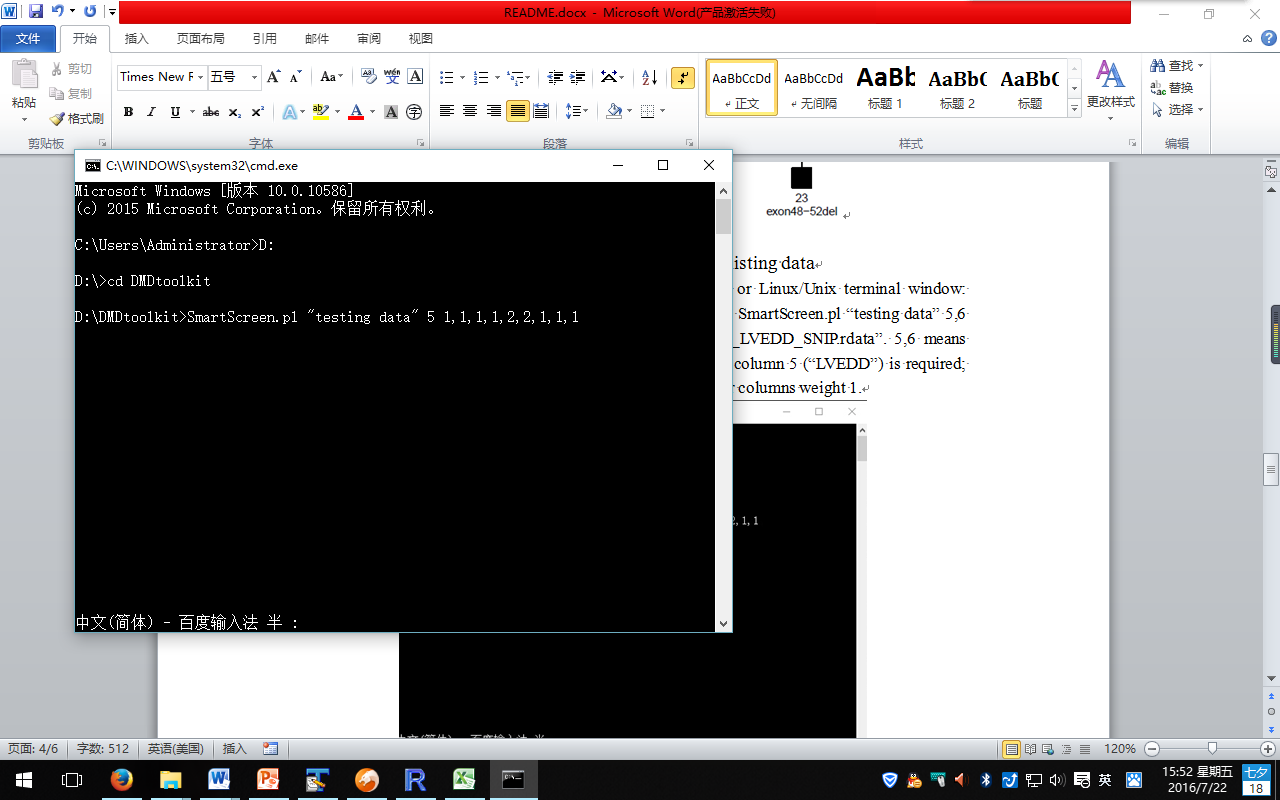


One of the pedigrees is as follows:

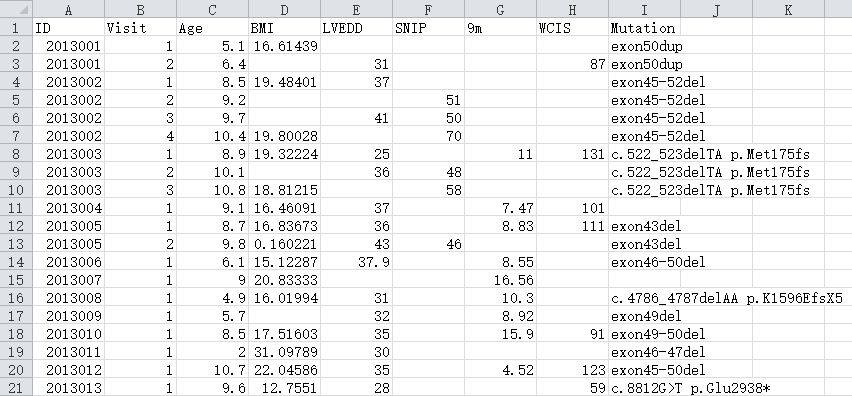


1. smartly screening the data to maximize the use of existing data

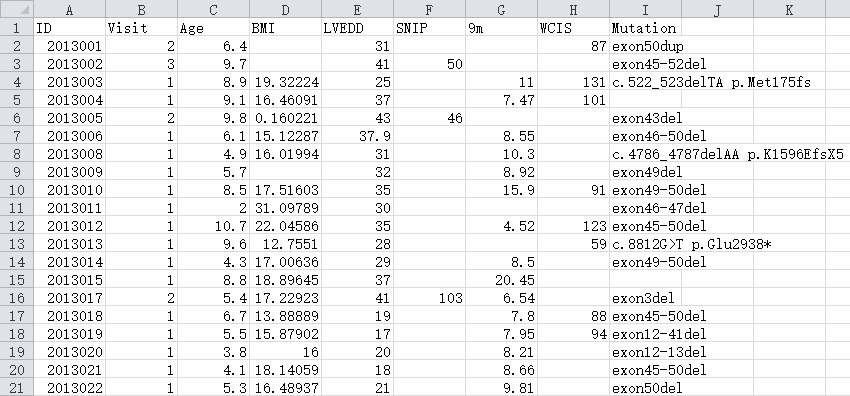
Use the command under the Windows dos/cmd window or Linux/Unix terminal window: SmartScreen.pl “file name” column\_No weights. For example: SmartScreen.pl “testing data” 5 1,1,1,1,2,2,1,1 will create a filtered file named “testing data\_LVEDD.rdata”. 5 means column 5 (“LVEDD”) is required (5,6 means columns 5 (“LVEDD”) and 6 (“SNIP”) are required); 1,1,1,1,2,2,1,1,1 means the 5th and 6th columns weight 2 and other columns weight 1.



The input file “testing data.txt” is shown as follows:

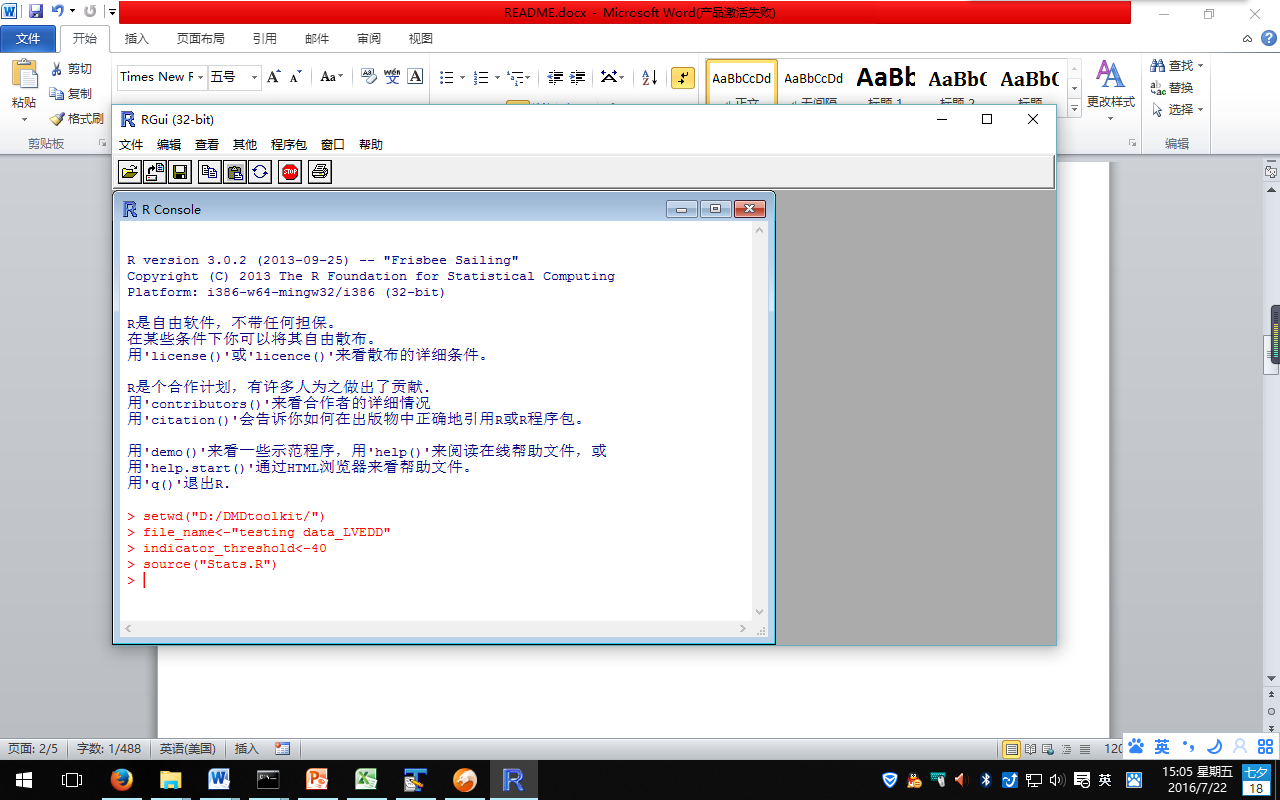


The output file “testing data\_LVEDD.rdata” is shown as follows:

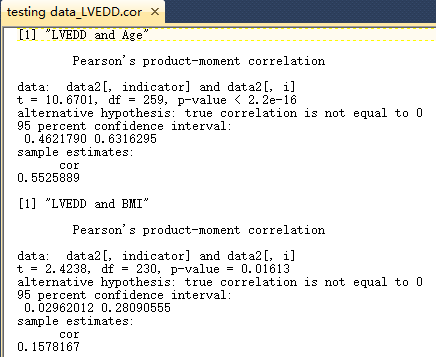
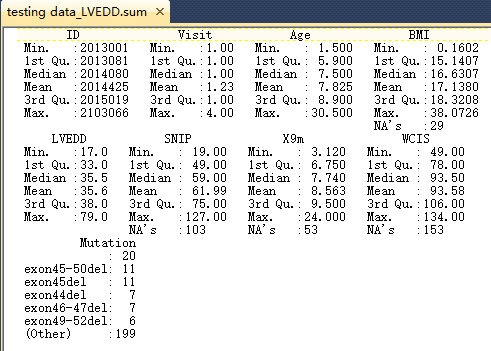
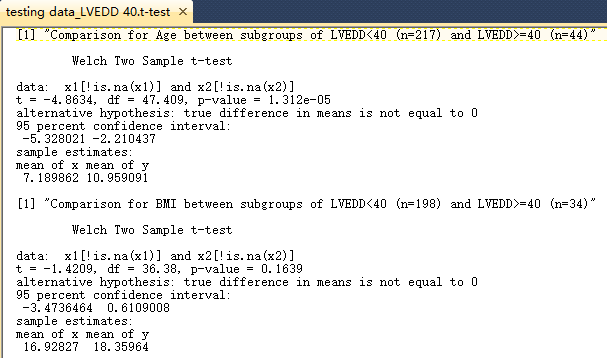
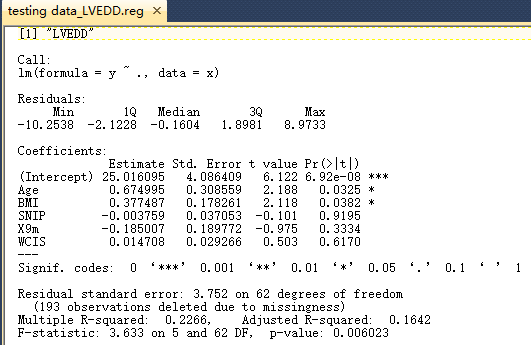


1. performing statistics for the DMD population and visualizing the results

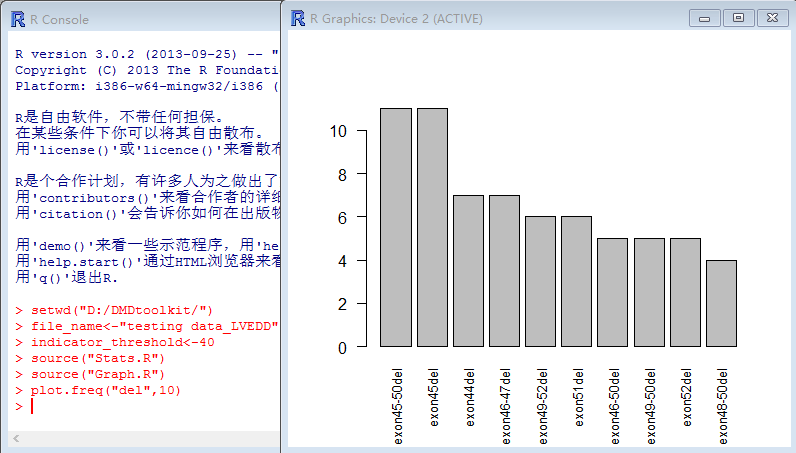
Use the commands under the R console: setwd(“the/working/directory”) to move to the working directory; file\_name<-"testing data\_LVEDD" to set the input file; indicator\_threshold<-# (e.g. 40) to set the threshold of subgroups for t-test; source(“Stats.R”) to perform the statistics.



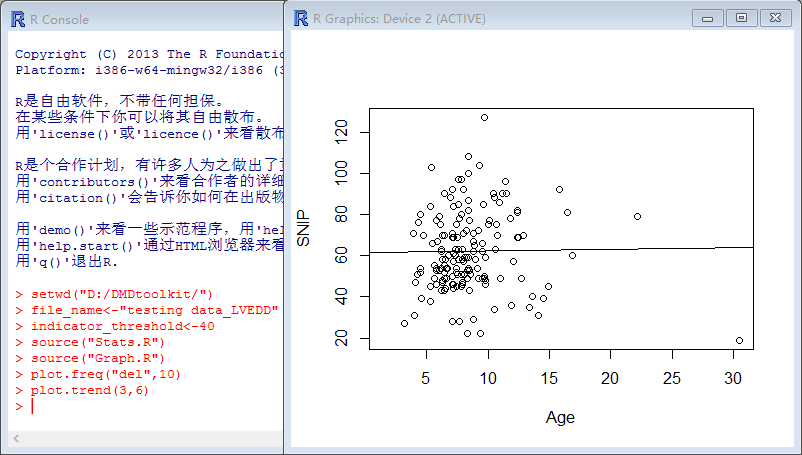
After the above commands, four output files will be created: “testing data\_LVEDD.\*” (sum/cor/reg) and “testing data\_LVEDD 40.t-test”, indicating summary, correlation, regression and t-test results.

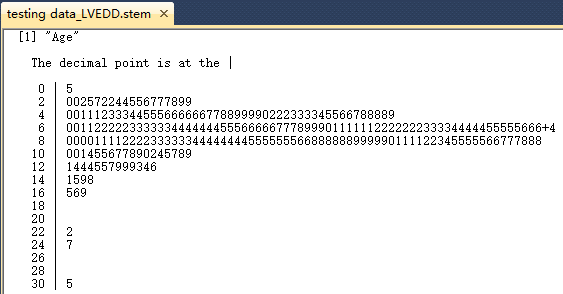
Use the command source(“Graph.R”), and then call plot.freq(type, num) to draw the mutation frequency histogram. type: “del”, “dup” and “all”; num: Arabic numbers.



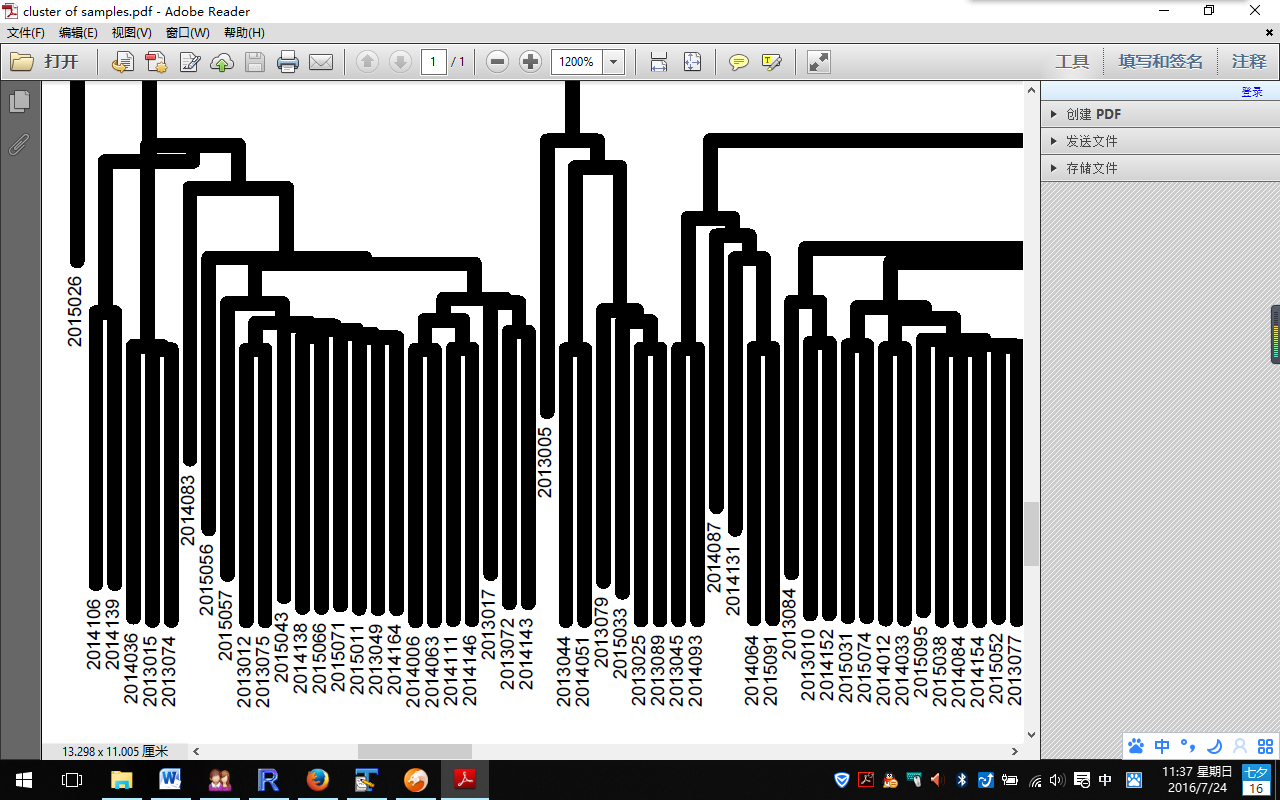
Call plot.trend(clmn\_1, clmn\_2) to draw the scatter plot and trend line of column 2 against column 1. Take plot.trend(3, 6) as an example, 3 means column 3 (“Age”), and 6 means column 6 (“SNIP”).



Call plot.stem(clmn\_no) to draw the stem and leaf plot. For example, plot.stem(3) will draw the stem and leaf plot of column 3 (“Age”), plot.stem(c(3,6)) will draw the plots of columns 3 & 6, and plot.stem(c(3:6)) will draw the plots of columns 3 to 6.



Call plot.clust(clmn\_no, cex\_no) to draw the cluster dendrogram. clmn\_no means column number; cex\_no is a numerical value giving the amount by which plotting text and symbols should be magnified relative to the default 1. plot.clust(1:6,0.4) will create the following pdf file.



Thank you for using DMDtoolkit. Any questions, please don't hesitate to contact [zhoujp111@126.com](mailto:zhoujp111@126.com).