

LaSPU Manual

The assoc-LaSPU 1.0 is a Linux command line operated program compatible with most Unix-like system. The main executable is 'assoc-aSPU.r'. There are several folders under the root path. Folder 'AnnotationData' stores the example genotype annotation file; 'CovariateNames' stores the example file appointing covariate names; 'GenotypeData' stores the example genotype file; 'PhenotypeCovariateData' stores the example file including both phenotype and covariate data, stored as a long format matrix; 'NodesProfile' stores the configuration file for multiple nodes, only applicable when users enable the corresponding parallel computing schema; 'R' stores the kernel R functions; 'OutputData' stores the output from the program; 'Demo' stores the explanation file for those examples files, and the bash script file with example command lines to operate on the example data (including example annotation file, example covariate file, example genotype file, etc.); When 'assoc-aSPU.r' is executed with proper inputs and options, all necessary functions will be called and implemented automatically. Results from the program will be saved to the folder "OutputData". If any error occurs during the program run, the error message together with normal output messages of the program will be logged in a '*.log' file under program root path. In the same time, an '*.rda' file containing R runtime environment will be saved to the 'OutputData' folder for further error debugging purpose.

Steps of operating the program:

Note: "\$" starts a Linux shell command line.

1. Install all the packages as in "./R/head.r" manually (only for the first time).
2. The first line of 'assoc-aSPU.r', i.e., the shebang of Linux, should automatically call your local 'Rscript' executable; otherwise, modify it to your current customized path of 'Rscript':
#!/your-path-of-Rscript -slave -vanilla

3. “\$ chmod u+x assoc-aSPU.r” to add executable privilege to current user if not existed yet.
4. “\$./assoc-aSPU.r -h” to see help documents.
5. Check 'Demo' folder:
'ExampleData_demo.txt': shows the format of all kinds of input data you need to prepare.
'commandLine_demo.sh': shows the example command line operating on the example data.
6. You are ready to prepare your own data into preferred format and run the program.

Some Usage Notes:

1. For fastest speed, use '-include_aSPU FALSE' to only include five asymptotic tests (which are fast).
2. When including LaSPU family tests, you are recommended to use parallel computing schema as specified in '-parallel_scheme' argument.
3. Check the '-parallel_scheme' and '-parallel_over_gene' arguments to figure out a parallel computing scheme best for your dataset.
4. With '-usePermutedU TRUE', LaSPU family tests will use permutation method instead of simulation method to calculate the p value, which will cost significantly more computation time. It is advised that in the real data application involving rare variants, users could first run simulation based LaSPU test to obtain the global significant genetic loci, and then apply the more time-consuming permutation based LaSPU test on those selected loci as validation.
5. If you have multiple nodes available, check '-jobExecuteOnMultipleNodes' and '-MultipleNodes_profile' arguments to learn how to enable parallel run on multiple nodes with multiple cores. The example node profile configuration is deposited under the folder “NodesProfile”, which can be easily adapted to work under your cluster of nodes. Using multiple nodes is usually only necessary when you want to run LaSPU family tests with '-usePermutedU TRUE'.