Model-Based Multifactor Dimensionality Reduction

MBMDR-2.7.5 is a software that is able to detect multiple sets of significant gene-gene and/or gene-environment interactions in relation to a trait of interest, while efficiently controlling type I error rates. The trait can either be expressed on a binary scale, or on a continuous scale.

In this document, we present the sequential usage of the program and the main used options. If you are interested in a parallel usage of our software, the analysis of multiple-traits or censoring data, or other options, please contact the main author of this C++ program at f.vanlishout@ulg.ac.be.

1) Representation of the problem

Our software uses its own optimised representation of the problem. However, to facilitate its usage, we have developed a synopsis to convert a *ped* and a *map* file into a file coded in our internal representation:

```
mbmdr-2.7.5-mac-64bit.out --plink2mbmdr "--binary or --continuous"
-ped 'plinkPedFile' -map 'plinkMapFile' -o 'mbmdrFile' -tr 'traductFile'
```

Example:

Suppose that you want to convert the files *MY_FILE.map* and *MY_FILE.ped* containg the genotype and phenotype (represented on a continuous scale) of a study, into a new file called *MY_INPUT.txt*. The following command line solves the problem. Note that the software will also produce a file called *MY_TRADUCT.txt*, describing how the different genotypes of the SNPs of your dataset have been coded.

```
mbmdr-2.7.5-mac-64bit.out --plink2mbmdr --continuous -ped MY_FILE.ped
-map MY_FILE.map -o MY_INPUT.txt -tr MY_TRADUCT.txt
```

Here is a detailed description of the content of a file coded in our internal representation:

```
Trait  S1  S2  ... Sm
X1 Y11 Y12 ... Y1m
... ... ... ... ... ... ... ...
Xk Yk1 Yk2 ... Ykm
```

Where Si are the names of the markers
Xj is 'NA' if the value of the trait is non-available and a number otherwise
Yij is -9 if the value of the genotype is missing and 0, 1 or 2 otherwise

2) Execution of the software

Our software takes as argument a list of options described below and an input file in the internal representation described above.

```
mbmdr-2.7.5-mac-64bit.out [options] MY_INPUT.txt
```

ALGORITHM
- --maxT use our improvement of the max-T step-down permutation algorithm
- --minP use the classical min-P step-down permutation algorithm
- --margP use the classical marginal permutation algorithm

EXECUTION
- --sequential use the sequential version

TYPE OF DATA
- --binary the input file contains only one trait with binary values
- --continuous the input file contains only one trait with continuous values

TEST-STATISTIC COMPUTATION
- --hlo-mode uses the HLO method

STEP I: RISK CELL PRIORITIZATION
- --one-cell-approach generates the HLO matrix using prioritization by cell tests
- [-c DOUBLE] sets the p-value cut-off used in the cell tests (default: 0.1)

STEP II: HLO CONSTRUCT ASSOCATION TEST
- --h-vs-l analyses the HLO matrix using the H vs L technique
- --two-tests analyses the HLO matrix using the TWO TESTS technique

OUTPUT
- [-o STRING] sets the name of the output file (default: 'infileprefix'_output.txt)
- [--verbose] verbose intermediate results in the standard output

PARAMETERS
- -d INT sets the dimension (order of multi-locus model)
- -n INT sets the number of pairs in the result
- -p INT sets the number of permutations
- [-r INT] sets the starting random sequence (default: random value auto-generated)

Example:

```
mbmdr-2.7.5-mac-64bit.out --maxT --sequential --continuous
--hlo-mode  --one-cell-approach -c 0.1 --h-vs-l -d 2 -n 50 -p 999 -r 1980
–o MY_OUTPUT.txt MY_INPUT.txt
```