Model-Based Multifactor Dimensionality Reduction

MBMDR-4.4.1 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 be expressed either on a binary or a continuous scale, or as a censored trait. To see the command line help, type

mbmdr.out help

The instructions to run MBMDR-4.4.1 are (depending on the data type) as follows:

```
mbmdr.out --binary [options] 'mbmdrFile'
mbmdr.out --continuous [options] 'mbmdrFile'
mbmdr.out --survival [options] 'mbmdrFile'
```

If your data is expressed on a binary or continuous scale, then the 'mbmdrFile' must be represented using the following structure (for censored trait see *--help --survival*)

Tr1	…	Trx	Cv1	…	Cvy	Ma1	...	Maz
T11	…	T1x	C11	…	C1y	M11	…	M1z
…	…	…	…	…	…	…	…	…
Tk1	…	Tkx	Ck1	…	Cky	Mk1	…	Mkz

The first line is a title line: the Trj's are the names of the x traits (x>=1), the Cvj's are the names of the y covariates (y>=0) and the Maj's are the names of the z markers, i.e. SNPs and/or environment variables (z>=2).

The first x columns contain the trait values: in the binary case, Tij is 1 if the i^{th} subject is a case for the j^{th} trait and 0 if it is a control; in the continuous case Tij is a continuous value representing the state of the i^{th} subject for the j^{th} trait. The next y columns are covariate values (missing values are not allowed). The last z columns are markers values (missing values must be coded '-9'):

- ➔ if Maj is a SNP: Mij is 0 if the ith subject is homozygous for the first allele, 1 if heterozygous and 2 if homozygous for the second allele.
- ➔ if Maj is an environment variable: the X different possible values of the environment variables should be coded 0, 1, ..., X-1.

If your dataset is in PLINK format, you can first use the following command line to create the 'mbmdrFile' (replace --binary by --continuous or --survival depending on your trait)

```
mbmdr.out --plink2mbmdr --binary -ped 'pedFile' -map 'mapFile' -o 'mbmdrFile' -tr 'trFile'
```

The file ‘trFile’ is an output file giving the chosen labels for the genotypes of each SNP. The ‘pedFile’ must contain a title line (see *--help --plink2mbmdr* for more options, if you have a ‘pheFile’ the header has to be “ID sex trait cov1 cov2 …”).

The different options of the program are: (the options between square brackets are not mandatory)

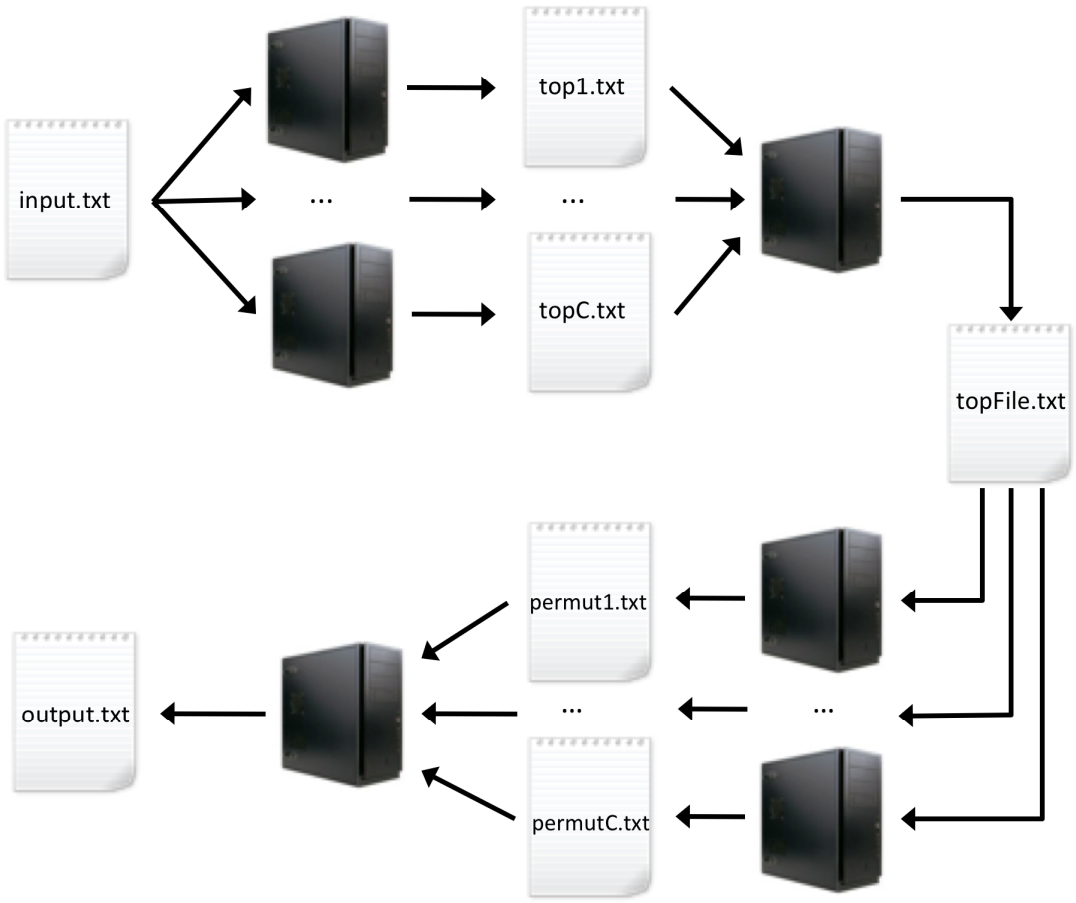
[-n INT]	number of top pairs in the output (default: 1000)
[-p INT]	permutation amount for multiple-testing (default: 999)
[-r INT]	random seed parameter (default: random value)
[-m INT]	minimum group size to be statistically relevant (default: 10)
[-at INT]	amount traits (default: 1)
[-ct INT]	current trait (default: 1)
[-ac INT]	amount covariates (default: 0)
[-x DOUBLE]	cutoff value for the statistical test (default: 0.1)
[-mt STRING]	multiple testing correction algorithm: NONE, MAXT, MINP, RAWP, STRAT1, STRAT2 or gammaMAXT (default)
[-rc STRING]	regress covariates: RESIDUALS (default), ONTHEFLY
[-o STRING]	output file name (default: 'inputprefix'_output.txt)
[-o2 STRING]	models file name (default: 'inputprefix'_models.txt)
[-a STRING]	adjust: CODOMINANT (default), ADDITIVE, ONESTEP or NONE
[-d STRING]	dimension of interactions: 1D, 2D (default) or 3D
[-pb STRING]	progress bar: NONE or NORMAL (default)
[-v STRING]	verbose in models file: SHORT, MEDIUM (default) or LONG
[-if STRING]	input format: MBMDR (default), MDR or ISALIVE
[-e LIST]	erase markers (LIST: comma-separated list of marker names)
[-E FILE]	erase markers (FILE: composed of one marker name per line)
[-k LIST]	keep only the markers from the comma-separated list
[-K FILE]	keep only the markers from the file
[-s FILE]	second stage of a discovery-replication analysis: keep only the pairs from the given output FILE of the first stage
[-f LIST]	filter: analyse only the pairs composed of exactly one marker from the given comma-separated list of marker names
[-F FILE]	filter: analyse only the pairs composed of exactly one marker from the given file and one marker from the input file
[-rt STRING]	rank transformation (continuous trait only): NONE (default) or RANK_TRANSFORM

Parallel Workflows

Users analysing big datasets should use the gammaMAXT parallel workflow.
(to consult the online manual see *--help --parallel*)

WARNING: please use the same set of computing options at each step!!

This workflow is composed of four steps:



STEP 1: compute partial top vectors on N CPUs (1, 2, ..., N)

```
mbmdr.out --continuous --gammastep1 -i INT -N INT [options] 'mbmdrFile'
```

SPECIFIC OPTIONS

-i INT	sets the current CPU id
-N INT	sets the total amount of CPUs
[-ti STRING]	sets the prefix of the temporary top files (default: top)

STEP 2: create the final top vector on one CPU

```
mbmdr.out --continuous --gammastep2 -N INT 'mbmdrFile'
```

SPECIFIC OPTIONS

-N INT	sets the total amount of CPUs
[-t STRING]	sets the top file name (default: topFile.txt)
[-ti STRING]	sets the prefix of the temporary top files (default: top)

STEP 3: compute the permutations on N CPUs (1, 2, ..., N)

```
mbmdr.out --continuous --gammastep3 -p INT -o STRING [options] 'mbmdrFile'
```

SPECIFIC OPTIONS

-p INT	sets the permutation amount to be run on the current CPU
-o STRING	sets the output file name (all CPUs must use 'xxxi.txt' where xxx is a common prefix and i the CPU id)
[-t STRING]	sets the top file name (default: topFile.txt)

STEP 4: create the final output file on one CPU

```
mbmdr.out --continuous --gammastep4 -c STRING -q INT [options] 'mbmdrFile'
```

SPECIFIC OPTIONS

-c STRING	sets the common prefix 'xxx' of the files generated at step 3
-q INT	sets the quantity of files generated at step 3
[-p INT]	sets the permutation amount (default: 999)
[-o STRING]	sets the output file name (default: 'inputprefix'_output.txt the file will be created in the directory of the input file)
[-t STRING]	sets the top file name (default: topFile.txt)