

# User Manual

## 1. INTRODUCTION

FastDMA is a software analyzing Illumina Infinium HumanMethylation450 BeadChip data, which is featured as multiple core parallel computing.

## 2. INSTALLATION

FastDMA uses GNU standard to compile and install so that it can be easily installed on Linux system.

However, **for non-expert in Linux system**, we highly recommend you strictly follow the installation guideline to install FastDMA.

- I. Download the most recent fastdma package (nominated as fastdmr-1.2.0.tar.gz) and boost, gsl library (boost\_1\_52\_0.tar.gz and gsl-1.15.tar.gz)
- II. Unzip fastdma package to a directory (noted as <fastdma-version>) and the two libraries under this directory. Make sure that you now have <fastdma-version>/boost\_1\_52\_0 and <fastdma-version>/gsl-1.15 directories.
- III. Run "bash fastdma-install <inst\_dir>" command under <fastdma-version> directory (<inst\_dir> is the directory that you intended to install fastdma). This step is usually time-consuming because of the system must compile boost and gsl libraries.
- IV. If you can find <inst\_dir>/boost, <inst\_dir>/gsl, and <inst\_dir>/bin directory, the installation is successful.
- V. For the first time running fastdma, you have to run "**source ~/.bashrc**" command to configure the environment path. Now you are prepared to run FastDMA.

For **non-first time installing** FastDMA or **experts in Linux system**, please see the following guideline to install FastDMA.

- I. Install boost and gsl library. (For non-first time installing, you have already install these in <inst\_dir>/boost and <inst\_dir>/gsl. So, you can skip this step)

- II. Run `./configure --prefix=<inst_dir_2> --with-boost=<inst_dir>/boost/ --with-gsl=<inst_dir>/gsl/` under the `<fastdma-version>` directory to configure installation. (This time, you want to install fastdma at `<inst_dir_2>`)
- III. Run `make` and `make install` consecutively to install FastDMA.
- IV. Configure the boost and gsl runtime library (If you are only to reinstall FastDMA, skip this step).
- V. Please make sure that you have `<inst_dir_2>/bin` directory to check your installation.

### 3. USAGE EXAMPLE

We will take the test dataset as an example to inform you the usage of fastdmr. To accomplish this, you must download `test_file.tar.gz` file. It is assumed that you have already installed FastDMA successfully under `<inst_dir>` and prepared to run. Now, you should have `<inst_dir>/bin/split` and `<inst_dir>/bin/fastdmr` files. And when you run `"<inst_dir>/bin/fastdma"` command, it shows the help information.

Make a directory `<dataset>` and unzip `test_file.tar.gz` into it. Now you can see the **test\_data.txt** file. This file is just like the HumanMethylation450 BeadChip data after GenomeStudio software processing. There are many columns named as `<Name>.AVG_Beta` and `<Name>.Detection.Pval` specified at the first row. `<Name>` is the sample name in the study. `".AVG_Beta"` and `".Detection.Pval"` specify the columns as **beta value** and **detection p value**. Another column `"TargetID"` specifies the **probe ID** in HumanMethylation450 BeadChip.

Now you can run `"<inst_dir>/bin/split -t TargetID -b .AVG_Beta -p .Detection.Pval -d <dataset> <dataset>/test_data.txt"` to format the original file into the files required by FastDMA. The `-t -b -p` parameters require the column names specifying correlated data. The `-d` parameter specify where to put all the formatted files. After this step, you have **many formatted files named as the samples** in `<dataset>` directory.

Now please see the `<dataset>/test_configure.txt` file, this file configures the FastDMA input and sample information. The first line begins with `"#"` denotes that this line is comment and every comment line means nothing to FastDMA program. The first three columns specify the

**data file, sample name, and group.** The other columns specify covariables in the study. In the test study, age and sex are taken as covariables **but not batch.** (Note, every **covariable must be in real number format,** you can not write sex covariable as "Male" and "Female". Covariables must appear in front of non-covariable annotation). To run fastdma, you must **modify the first column** to specify the data files you just made.

Now please see the **test\_probe\_ann.txt** file. This is an annotation of each probe on BeadChip. There are 4 columns in this file, namely, **probe ID, chromosome, genome coordinate, annotation.** You can have your own annotation file with anything you want in the 4th column. Real annotation file is also provided at sourceforge. This file is intended to annotate the probes and calculate the genomic distance between probes.

The last file **test\_cpglisland.txt** is in standard bed 4 column format. This file specifies the regions you want to test differential methylation. By default, it contains all the CpG islands along the genome, but you can have your own.

Now make a directory <result> to store the result files and run "**<inst\_dir>/bin/fastdma -p 0.01 -q 0.01 --unify -n 2 -c <dataset>/test\_configure.txt --covariable 2 --probe <dataset>/test\_probe\_ann.txt --op <result>/probe.txt --probe-prefix <result>/probe --cgi <dataset>/test\_cpglisland.txt --oc <result>/cgi.txt --cgi-prefix <result>/cgi --od <result>/dmr.txt --dmr-prefix <result>/dmr**" to finish a run testing all the functions in FastDMA. Many files appears in the <result> directory. Files ended by ".txt" are user-friendly text results. Files ended by ".BedGraph" can be directly loaded into genome browsers to visualize the results.

## 4. OPTIONS

OPTIONS	EXPLANATION
-h, --help	Display help information
--normalize-off	Do not normalize the original beta values before further analyses
--cgi-off	Do not do DMR analysis with given regions
--dmr-off	Do not do DMR analysis without given region
-n, --thread	Number of parallel threads, default is 1
-p, --pvalue	Threshold of pvalue in statistical analysis, default is 0.01
-q, --fdr	Threshold of false discovery rate (FDR) in statistical analysis, default is 0.01
-c, --configure	Configuration file as test_configure.txt specifying data files, sample names, group information and covariables
--continuous	Regress the continuous variabe rather than the discrete group information
--filter	Reject probes with possibly poor behavior
--covariable	Covariable number, default is 0
--probe	Probe annotation file as test_probe_ann.txt
--op	Probe analysis result in text format
--probe-prefix	Probe analysis result in BedGraph format
--cgi	Given regions to do DMR analysis
--oc	CpG island (CGI) analysis result in text format
--cgi-prefix	CGI analysis result in BedGraph format
--unify	Use the unified method to do DMR analysis
--od	DMR analysis result in text format
--dmr-prefix	DMR analysis result in BedGraph format