NORDI (NORmal DIscretization for gene expression data)
Program User's Guide

This simple guide allows the compilation and manipulation of the NORDI Source program in R language. It is suggested to follow these three steps before running NORDI:

0 Preliminary:
- Installation of the R program. A comprehensive guide can be found in: **http://cran.r-project.org/** . NORDI has been made with the (2007-11-26) R-2.6.1 release, but it can be compiled with older releases of R.

1 Data Format:
- Your data has to be presented in this format:

	Condition 1	Condition 2	Condition 3	Condition 4	...	Condition M
GENE 1	-0.01	0.29	-0.42	-0.29	...	0.06
GENE 2	0	0.19	0.12	-0.36	...	-0.15
GENE 3	0.06	0.03	0.07	-0.34	...	-0.07
…	0.15	0	0.11	-0.23	...	-0.03
…	0.3	0.07	0.19	-0.23	...	-0.1
...	0.2	0.07	0.11	-0.32	...	-0.23
GENE N	0.24	0.15	0.12	-0.15	...	-0.04

  That is with one empty space at the beginning, with the first row for column names and the first column for gene names, then the numeric gene expression measures fill the data (as shown in the table).
- It is recommended to have the data in a text file with data separated either by tabs or by spaces.

2 R First Steps:
- Open the R screen and choose the working directory. Go to the principal menu, then File, and then change dir…
- Put the data file (*.txt) and the NORDI source program file, *"NORDISource.txt"*, in the working directory. You can use the *"DerisiData.txt"* file as example.
- Use the R console and install the basic Bioconductor packages. Go to the principal menu, then Packages, and then Install package(s)…
- Install the three special packages used in NORDI program: *"outliers"*, *"tseries"* and *"nortest"*.

3 NORDI run and results
- Load the data file by executing this command in the R console:

  T<-read.table("DerisiData.txt", header = TRUE, sep = "\t", quote="\"", dec=".", row.names=1, fill=TRUE, blank.lines.skip = TRUE);
- Run the NORDI source program by executing this command in the R console

  source("NORDISource.txt")
- The result *"DiscretizedGenExpressionMatrix.txt"* is in the working directory.

Attention! NORDI uses two basic and crucial parameters:
- The *p-value* for determining the gene expression matrix column outliers using the Grubbs outliers test.
- The quantiles for a standardized normal distribution corresponding to certain α for determining the upper and lower discretization thresholds in each column matrix.

In the source program these parameters are fixed to *p-value* $=.01$ and $\alpha=.05$, but the user is able to change them in the source program: NORDISource.txt.