

## Preparation data and set up R for analysis

R is a language and environment for statistical computing and graphics. MIAMI data can be analyzed by using R source which can be download from MIAMI site (<http://grc.dept.med.gunma-u.ac.jp/~gene/indexe.html>).

- (1) Labeling control sample with Cy3 and case sample with Cy5.
- (2) Hybridization and wash.
- (3) Scan with appropriate PMT to avoid saturated spots and feature extraction using Agilent scanner (CGH setting).
- (4) Install and setup R from (<http://cran.r-project.org/>).
- (5) Download the R source file (Lm\_Prot\_AJ\_Cmp2\_NDC.R) from MIAMI site (<http://grc.dept.med.gunma-u.ac.jp/~gene/indexe.html>).

# How to analyze Agilent scanner data with R

- (1) Start R.
- (2) Load source "Lm\_Prot\_AJ\_Cmp2\_NDC.R" with file menu.
- (3) Change directory to your data folder by file menu and press enter.
- (4) Input the file name for Hpa II-cleavage difference and press enter.
- (5) Input the file name for Msp I-cleavage difference and press enter.
- (6) Input a name added to output files and press enter.
- (7) After several minutes, calculation will be ended with the notification of "R process all finished."
- (8) Output files will be made in the same folder as your data files. Following are the detail of the files.

```
> source("C:/RScript/Lm_Prot_AJ_Cmp2_NDC.R")
Thu Jun 26 16:00:24 2008

Current working directory is : C:/Program Files/R/R-2.3.0
Change directory to your data folder by file menu and press enter.
Current working directory is : C:/Documents and Settings/畑田出穂/デスクトップ/test
Input file name(A): A_14716_10246.txt
Input file name(B): B_14716_10247.txt
fname(A): A_14716_10246.txt fname(B): B_14716_10247.txt
Input a name added to output files: P19
Name added to output files: P19
Thu Jun 26 16:01:12 2008

A_14716_10246.txt has 243496 lines, B_14716_10247.txt has 243496 lines.
After merge and selecting(Flags, ControlType) process, 237133 lines remained.
Thu Jun 26 16:15:24 2008

Finish the process omit double or triple lines
Thu Jun 26 16:15:25 2008

sum50: 11857586 11857891 11857293 11857336
Thu Jun 26 16:15:31 2008
ave50: 1815.307 2037.438 1247.217 1372.536
TH: 90.76535 101.8719 62.36085 68.62678
ab_min: 0.9171256 ab_max: 1.206065
a: 0.9007496 b: -0.01432636
Ac: 5 Bc: 2 Dc: 5
3000 entry processed.
6000 entry processed.
9000 entry processed.
12000 entry processed.
15000 entry processed.
18000 entry processed.
21000 entry processed.
24000 entry processed.
27000 entry processed.
30000 entry processed.
Finish the calculations for hyper, hypo
Thu Jun 26 16:18:04 2008

Res: 30586 Hyper: 252 Hypo: 3 finished.
UCSC: 255 finished.
R process all finished.
Thu Jun 26 16:18:37 2008

> █
```

Res: Data file of the results.

Hyper: List of hypermethylated probes when the control sample is labeled with Cy3.

Hypo: List of hypomethylated probes when the control sample is labeled with Cy3.

Statics: Statics file.

UCSC: File which can be used for UCSC genome browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>).

## Definitions of terms in files

ProbeName	
GeneName	
gProcessedSignal_A	Hpa II signal of control
rProcessedSignal_A	Hpa II signal of case
gProcessedSignal_B	Msp I signal of control
rProcessedSignal_B	Msp I signal of case
Flags_A	
Flags_B	
ControlType	
ProbeUID	
SystematicName	
Description	
PValueA	$P$ value of $\log[(rProcessedSignal\_A)/(gProcessedSignal\_A)]$
PValueB	$P$ value of $\log[(rProcessedSignal\_B)/(gProcessedSignal\_B)]$
Cy3A	Hpa II signal of control (Normalized)
Cy5A	Hpa II signal of case (Normalized)
Cy3B	Msp I signal of control (Normalized)
Cy5B	Msp I signal of case (Normalized)
A	$Cy3A/Cy5A$
B	$Cy3B/Cy5B$
logA	
logB	
DRL	Horizontal distance from the regression line
D10	$10^{(DRL)}$
judge	judge=1(hypermethylated): $D10 > 5, A > 5, 0.5 < B < 2$
	judge=-1(hypomethylated): $D10 < 0.2, A < 0.2, 0.5 < B < 2$
	judge=0: If Judge is not equal to 1 or -1, then 0.