

SMETHILLIUM: Spatial normalisation METHod for ILLumina InfinIUM HumanMethylation BeadChip

Camille Sabbah^{1,2,3}, Gildas Mazo^{1,2,3}, Caroline Paccard^{1,2,3}, Fabien Reyat^{1,4,5}, Philippe Hupé^{1,2,3,4}

January 27, 2011

- 1 Institut Curie, 26 rue d'Ulm, Paris, F-75248 France
- 2 INSERM, U900, Paris, F-75248 France
- 3 Mines ParisTech, Fontainebleau, F-77300 France
- 4 CNRS UMR144, 26 rue d'Ulm, Paris, F-75248 France
- 5 Institut Curie, Department of Surgery, 26 rue d'Ulm, Paris, F-75248 France

Contents

1	Description of the smethillium source code	2
2	Description of input data	2
2.1	Annotation files from Illumina	2
2.2	Raw data for the sample under investigation	2
3	Description of output data	2
3.1	beads	2
3.2	targets	3
4	Code example	5

1 Description of the smethillium source code

The source code is organised as follows:

data contains all the data needed to run the example

documentation contains the present documentation

script contains the code source from the SMETHILLIUM function and an example

2 Description of input data

2.1 Annotation files from Illumina

HumanMethylation27_270596_v.1.2.csv File containing information about the targets.
Available at "<http://www.illumina.com/forms/ftp.illumina.com>". See Illumina for details.

2.2 Raw data for the sample under investigation

The RCABeadStudio.txt (and met1.1BeadStudio.txt) file contains the raw data at the target level:

Index	row number
TargetID	ID of the target
ProbeID_A	address of the unmethylated probe
ProbeID_B	address of the methylated probe
COLOR_CHANNEL	color channel of the target
ADDRESSA_ID	address of the unmethylated probe
ADDRESSB_ID	address of the methylated probe
RCA.N46LEG.AVG_Beta	See BeadStudio User Guide for details (optional)
RCA.N46LEG.Intensity	See BeadStudio User Guide for details (optional)
RCA.N46LEG.Avg_NBEADS_A	See BeadStudio User Guide for details (optional)
RCA.N46LEG.Avg_NBEADS_B	See BeadStudio User Guide for details (optional)
RCA.N46LEG.BEAD_STDERR_A	See BeadStudio User Guide for details (optional)
RCA.N46LEG.BEAD_STDERR_B	See BeadStudio User Guide for details (optional)
RCA.N46LEG.Signal_A	See BeadStudio User Guide for details (optional)
RCA.N46LEG.Signal_B	See BeadStudio User Guide for details (optional)
RCA.N46LEG.Detection.Pval	See BeadStudio User Guide for details (optional)

The 4513233020.L.txt (and 4513233021.C.txt) file contains the raw data at the bead level:

Code	Address of the bead
Grn	Intensity for the green channel
GrnX, GrnY	position over the array for the green channel
Red	intensity for the red channel
RedX, RedY	position over the array for the red channel

3 Description of output data

The SMETHILLIUM function provides a list with the following information:

3.1 beads

Each row represents a bead. The controls are present and corresponds to the beads for which the variable COLOR_CHANNEL is "Control Bead".

Code	address of the bead
Grn	intensity for the green channel
GrnX, GrnY	position over the array for the green channel
Red	intensity for the red channel
RedX, RedY	position over the array for the red channel
TargetID	ID of the target
M_or_U	M if the probe tests the methylation state, U otherwise
COLOR_CHANNEL	color channel of the bead
NOISE	estimated background of the bead by SMETHILLIUM method
NORM	normalized intensity of the bead (after SMETHILLIUM method)
Log2NORM	log-intensity in base 2 of "NORM" (after SMETHILLIUM method)

3.2 targets

This data frame contains the data at the target level. It contains among others informations the beta value for the methylation state.

TargetID	ID of the target
Name	ID of the target
ProbeU	Code of the Unmethylated probe interrogating the locus
ProbeM	Code of the Methylated probe interrogating the locus
Chr	data from BeadStudio
MapInfo	data from BeadStudio
COLOR_CHANNEL	data from BeadStudio
TSS.Coordinate	data from BeadStudio
Gene_Strand	data from BeadStudio
Gene_ID	data from BeadStudio
Symbol	data from BeadStudio
Distance_to_TSS	data from BeadStudio
CPG_ISLAND	data from BeadStudio
CPG_ISLAND_LOCATIONS	data from BeadStudio
MIR_CPG_ISLAND	data from BeadStudio
MIR_NAMES	data from BeadStudio
IntNormMeanU	Normalized intensity summarized by mean whitout the outliers intensity for the Unmethylated probes (see the supplementary data)
IntNormMeanM	Normalized intensity summarized by mean whitout the outliers intensity for the Methylated probes (see the supplementary data)
IntNormMedianU	Normalized intensity summarized by median for the Unmethylated probes
IntNormMedianM	Normalized intensity summarized by median for the Methylated probes
BetaNormMean	beta value computed after normalisation where the intensities are summrized by mean whitout the outliers intensity (see the supplementary data)
BetaNormMedian	beta value computed after normalisation where the intensities are summrized by median
IntRawU	raw intensity summarized by mean for the Unmethylated probes
IntRawM	raw intensity summarized by mean for the Methylated probes
BetaRaw	beta value computed without normalisation where the intensities are summrized by mean
LogitBetaRaw	Logit transformation of the Raw Beta
LogitBetaNormMean	Logit transformation of the normalized mean Beta
LogitBetaNormMedian	Logit transformation of the normalized median Beta
Pvalue_M	p-value from the t-test calculated without the variance of the controls beads for the methylated beads (see the supplementary data)
Pvalue_M_Control	p-value from the t-test calculated with the variance of the controls beads for the methylated beads (see the supplementary data)
Pvalue_U	p-value from the t-test calculated without the variance of the controls beads for the unmethylated beads (see the supplementary data)
Pvalue_U_Control	p-value from the t-test calculated with the variance of the controls beads for the unmethylated beads (see the supplementary data)
Pvalue_NPMedian	p-value from the Wilcoxon-Mann-Whitney test calculated with all the controls beads

4 Code example

In order to run the example you have to:

- untar the SMETHILLIUM source code archive
- cd the SMETHILLIUM directory
- source the Example_SMETHILLIUM.R file.

```
source("script/smethillium.R")
```

```
##### Unmethylated Example #####
```

```
WorkableData <- read.csv("data/RCA.N46LEG_4513233020_L/4513233020_L.txt", header=TRUE, sep="", check.names = F)
```

```
InfoGenBeads <- read.csv("data/HumanMethylation27_270596_v.1.2.csv", as.is=TRUE, header=TRUE)
```

```
RCAExample <- smethillium(WorkableData=WorkableData, InfoGenBeads=InfoGenBeads, delta=1)
```

```
##### Methylated Example #####
```

```
WorkableData <- read.csv("data/met1.1_4513233021_C/4513233021_C.txt", header=TRUE, sep="", check.names = F)
```

```
InfoGenBeads <- read.csv("data/HumanMethylation27_270596_v.1.2.csv", as.is=TRUE, header=TRUE)
```

```
metExample <-smethillium(WorkableData=WorkableData, InfoGenBeads=InfoGenBeads, delta=1)
```