HumMeth27QCReport: A Package to Generate QC Reports for Infinium HumanMethylation27 BeadChip Methylation Assay Data

Francesco M. Mancuso

November 24, 2010

Contents

1 Introduction

This document describes an R package for generating QC reports. The goal of this project is to create a tool to allow users of Illumina Infinium HumanMethylation27 BeadChip Methylation Assay[1] to quickly access the data quality of a batch of processed arrays. The package makes use of the *methylumi* package for reading files exported from BeadStudio software, generating intensity plots and normalizing Beta values. Several new plots are generated and a printable pdf files are created. To run properly and generate the summary Excel file, the script needs that a working versions of Perl is installed on your machine.

2 Usage

After starting R, the package should be loaded using the following.

```
> library(HumMeth27QCReport)
```

[1] www.illumina.com/

This will load *HumMeth27QCReport* as well as the *methylumi*, *lumi*, *amap*, *Hmisc*, *gplots*, *plotrix*, *WriteXLS* and *tcltk* packages and their dependencies.

```
R> HumMeth27QCReport(Dir)
```

Dir: character string containing the location of the directory in which the input file are.

If desired the analysis can be split in three parts by a single function call:

* **QCRep** creates all the histogram plots relative to the internal controls of the Illumina assay into pdf files.

```
R> QCRep(Dir)
```

* **QCCheck** creates all the plots relatve to the quality of the samples.

```
R> QCCheck(Dir)
```

* **NormCheck** normalize the Beta Values and plot a PCA and a hierarchical Clustering of the samples using the noralized data

```
R> NormCheck(Dir)
```

3 Inputs and Outputs files

HumMeth27QCReport takes in input the three files from BeadStudio plus an optional text file with the chip control samples to discard from the normalization step:

* **Sample table** (it is compulsory that the file name contains the word "Sample", case sensitive, and not the others reserved words)

* **Control table** (it is compulsory that the file name contains the word "Control", case sensitive, and not the others reserved words)

* **BetaAverage table** (it is compulsory that the file name contains the word "Avg", case sensitive, and not the others reserved words)

* **Discard.txt** (compulsory name)

Sample table - Required columns from BeadStudio:
- Index
- Sample ID

- Sample Group
- Sentrix Barcode
- Sample Section
- Detected Genes (0.01)
- Detected Genes (0.05)
- Signal Average GRN
- Signal Average RED
- Signal P05 GRN
- Signal P05 RED
- Signal P25 GRN
- Signal P25 RED
- Signal P50 GRN
- Signal P50 RED
- Signal P75 GRN
- Signal P75 RED
- Signal P95 GRN
- Signal P95 RED
- Sample_Well
- Sample_Plate

Control table - Required columns from BeadStudio (<Sn> = Sample Name):
- Index
- TargetID
- ProbeID
- <Sn>.Signal_Grn
- <Sn>.Signal_Red
- <Sn>.Detection Pval
- ...

Required controls (rows):
* BISULFITE CONVERSION (4 rows)
* EXTENSION (4 rows)
* HYBRIDIZATION (3 rows)
* NEGATIVE (16 rows)
* NON-POLYMORPHIC (4 rows)
* SPECIFICITY (4 rows)
* STAINING (4 rows)
* TARGET REMOVAL

AverageBeta table - Required columns from BeadStudio (<Sn> = Sample Name):
- Index
- TargetID
- <Sn>.AVG_Beta

- <Sn>.Intensity
- <Sn>.Signal_A
- <Sn>.Signal_B
- <Sn>.BEAD_STDERR_A
- <Sn>.BEAD_STDERR_B
- <Sn>.Avg_NBEADS_A
- <Sn>.Avg_NBEADS_B
- <Sn>.Detection Pval
- ...
- SYMBOL

Discard.txt Text file that contains the name of the samples (the same name present in the Sample table; one sample per row) you want to discard from normalization. i.e. sample controls to see if chip worked properly like un-methylated samples.

HumMeth27QCReport creates as output different plots (saved in pdf files) to asses the quality of the samples:
- a histogram foreach internal control.
- an Intensity Graph plot foreach sample recalling the "plotSampleIntensities" function of methylumi package.
- a histogram with the percentage of non dectected CPG (that is the CPGs tha have a detection p-value bigger than 0.05 or 0.01.
- a histogram with the average p-value for each sample.
- a PCA of normalized Beta values
- a Cluster of normalized Beta values

As ulterior output, an Excel file is provided. It contains the normalized BetaValues, a summary of the Internal Controls and of the gene detection and different lists of non-detected CPGs.

The methods to normalize the data are described further in the *lumi* package documentation.

4 Acknowledgments

I would like to thanks Manuela Hummel for the technical support.