mmgmos User Guide

Xuejun Liu, Magnus Rattray, Marta Milo, Neil D. Lawrence

April 25, 2006

1 Introduction

The *mmgmos* package is to estimate the expression levels and the confidence of measures for multiple arrays of the same type of Affymetrix GeneChips using the Multi-chip modified gamma Model for Oligonucleotide Signal (multi-mgMOS) (2) and the modified gamma Model for Oligonucleotide Signal (mgMOS) (1). The results obtains from *mmgmos* package include the mean, standard deviation and 5, 25, 50, 75 and 95 percentiles of the estimated expression measurements. This software package uses the optimaisation program donlp2 (3).

2 From probe level data to expression values

The main task of *mmgmos* package is to go from probe level data (*Cel* files) to expression measures. *mmgmos* requires *affy* package. Before start make sure the `affy` package is installed.

1. Create a directory, move all the relevant *CEL* files to that directory.

2. Start R in that directory.

3. If using the Rgui for Microsoft Windows make sure your working directory contains the *Cel* files (use "File -> Change Dir" menu item).

4. Load the library.

   ```
   R> library(mmgmos) ##load the mmgmos package
   ```

5. Read in the data.

   ```
   R> Data <-  ReadAffy() ##read data in working directory
   ```

6. Calculate the expression levels and related confidence.

```
R> eset <- mmgmos(Data)   ##using multi-mgMOS
R> eset <- mgmos(Data)    ##using mgMOS alternatively
```

7. Save results into files. Do not need to specify the extension of the file name here.

```
R> write.reslts(eset, file="myresults")
```

The results will be saved in the following 7 files:

   (a) **myresults_exprs.csv** contains the mean of the expression values.

   (b) **myresults_se.csv** contains the standard deviation of the expression values.

   (c) **myresults_prctile5.csv** contains the 5 percentile of the expression values.

   (d) **myresults_prctile25.csv** contains the 25 percentile of the expression values.

   (e) **myresults_prctile50.csv** contains the 50 percentile of the expression values.

   (f) **myresults_prctile75.csv** contains the 75 percentile of the expression values.

   (g) **myresults_prctile95.csv** contains the 95 percentile of the expression values.

Alternatively, for large data sets step 5 and 6 can be replaced by a single step

```
R> eset <- justmmgMOS()  ##using multi-mgMOS
R> eset <- justmgMOS()   ##using mgMOS alternatively
```

which is more memory efficient and makes large data sets properly processed. In this case, ReadAffy() shouldn't be called.

References

[1] Milo,M., Niranjan,M., Holley,M.C., Rattray,M. and Lawrence,N.D. (2004) A probabilistic approach for summarising oligonucleotide gene expression data. Technical report available upon request.

[2] Liu,X., Milo,M., Lawrence,N.D. and Rattray,M. (2005) A tractable probabilistic model for Affymetrix probe-level analysis across multiple chips. Bioinformatics, 21:3637-3644.

[3] Peter Spellucci. DONLP2 code and accompanying documentation. Electronically available via http://plato.la.asu.edu/donlp2.html.