MiChip

Jonathon Blake

May 1, 2024

Contents

1 Introduction

MiChip is a microarray platform using locked oligonucleotides for the analysis of the expression of mi-croRNAs in a variety of species Castoldi et al. (2008). The MiChip library provides a set of functions for loading data from several MiChip hybridizations, flag correction, filtering and summarizing the data. The data is then packaged as a Bioconductor *ExpressionSet* object where it can easily be further analyzed with the Bioconductor toolset *http://www.bioconductor.org/*.

First load the library.

```
> library(MiChip)
```

2 Reading the Hybridization Files

MiChip is scanned as a single colour cy3 hybridization and the output is gridded using Genepix software. To load the data from a set of MiChip hybridization Genepix files into bioconductor, use the *parseRawData()* method.

```
> datadir <-system.file("extdata", package="MiChip")
> defaultRawData <- parseRawData(datadir)

2 files found
Dealing with file  B_676.gpr
Dealing with file  B_677.gpr
```

The defaults are current directory "." And the "gpr" file extension. Loading data from a directory other than the current directory requires sending the directory to the method e.g. *otherDirectoryData* <-parseRawData(datadir="/myDemoDir", pat ="gpr") .

All files in the directory with the matching extension will be parsed and combined into an *ExpressionSet* containing all features on the chip with the background subtracted intensity from the scanner and quality flags. All hybridizations in the directory should be of the same type otherwise an error will be thrown.

3 Removing Unwanted Rows and Correcting for Flags

Due to the spotting configuration of MiChip and the probes supplied in the Exiqon probe library there are several data points which can be removed from the data set before analysis. Some of the spots on the chip are empty, others contain various controls and probes relating to microRNAs from different species possibly not relevant to the analysis in at hand. To remove data from these points use the *removeUnwantedRows()* method. This takes an array of strings and removes rows containing any of these strings in the gene name annotation of the data.

Remove all empty spots from data set

```
> noEmptiesDataSet <- removeUnwantedRows(defaultRawData, c("Empty"))

Raw Data  4608
Number of filters  1
Filtering special case empties
Filtered Data  4504
```

Use the helper method to produce the standard set of data rows for human MiChip experiments

```
> humanDataSet <- standardRemoveRows(defaultRawData)

Raw Data  4608
Number of filters  13
Filtering special case empties
Filtered Data  1118
```

Flags for the MiChip hybridizations are 0 for passes and negative values for spots that are marked absent. Data points with flag values less than zero are set to NA using the correctForFlags method.

```
> flagCorrectedDataSet <- correctForFlags(humanDataSet)

B_676  had  55  bad flags from  1118  data points
B_677  had  54  bad flags from  1118  data points
```

Positive but low intensities may lead to readings near background being taken as positive. Therefore an intensity cutoff can be sent to the *correctForFlags()* to set all the intensities under a set value to NA.

```
> flagCorrectedDataSet <- correctForFlags(humanDataSet, intensityCutoff = 50)

B_676  had  55  bad flags  and  77  values below cutoff from  1118  data points
B_677  had  54  bad flags  and  86  values below cutoff from  1118  data points
```

4 Summarizing Intensities

The MiChip probes are spotted in either duplicate or quadruplicate on the array. The individual readings of the data can be combined to give a single intensity value. The combined intensity is taken as the median of the individual intensities, omitting NAs. A minimum length for the acceptable number of present values is supplied to prevent features with only a low number of positive calls being accepted. Summarized intensities where the median absolute deviation is greater than the median intensity can be set to NA on the grounds of being too variable. This is done by setting the madAdjust argument to TRUE.

```
> summedData <- summarizeIntensitiesAsMedian(flagCorrectedDataSet,minSumlength = 0, madAdjust=FALSl
```

5 Plotting Functions

MiChip contains two functions for plotting intensity data, both are wrappers for standard plotting functions. The data however produced is written to a file allowing intensity plots and box plots to be produced automatically.

```
> plotIntensitiesScatter(exprs(summedData), NULL, "MiChipDemX", "SummarizedScatter")
```

File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/MiChipDemX_SummarizedScatter.jpg

Figure 1 shows scatter plots of the intensites of the hybrdizations.

```
> boxplotData(exprs(summedData), "MiChipDemX", "Summarized")
```

File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/MiChipDemX_Summarized.jpg
```
          used (Mb) gc trigger (Mb) max used (Mb)
Ncells  679592 36.3    1392714 74.4  1392714 74.4
Vcells 1403653 10.8    8388608 64.0  4196339 32.1
```

Figure 2 shows boxplots of the summarized intensity data.

6 Normalization

The major advantage of the MiChip library is to parse MiChip hybridization data sets into an *ExpressionSet* so that existing methods for normalization and hybridization within Bioconductor can be used. Median normalization per chip is implemented in the MiChip.

```
> mednormedData <- normalizePerChipMedian(summedData)
```

7 Writing Output Files

The *outputAnnotatedDataMatrix()* method combines the annotation and expression data from an *ExpressionSet*. This produces a tab delimited file containing data annotation in the left hand columns and expression data in the right for distribution or analysis with other applications.

```
> outputAnnotatedDataMatrix(mednormedData, "MiChipDemo", "medNormedIntensity",  "exprs")
```

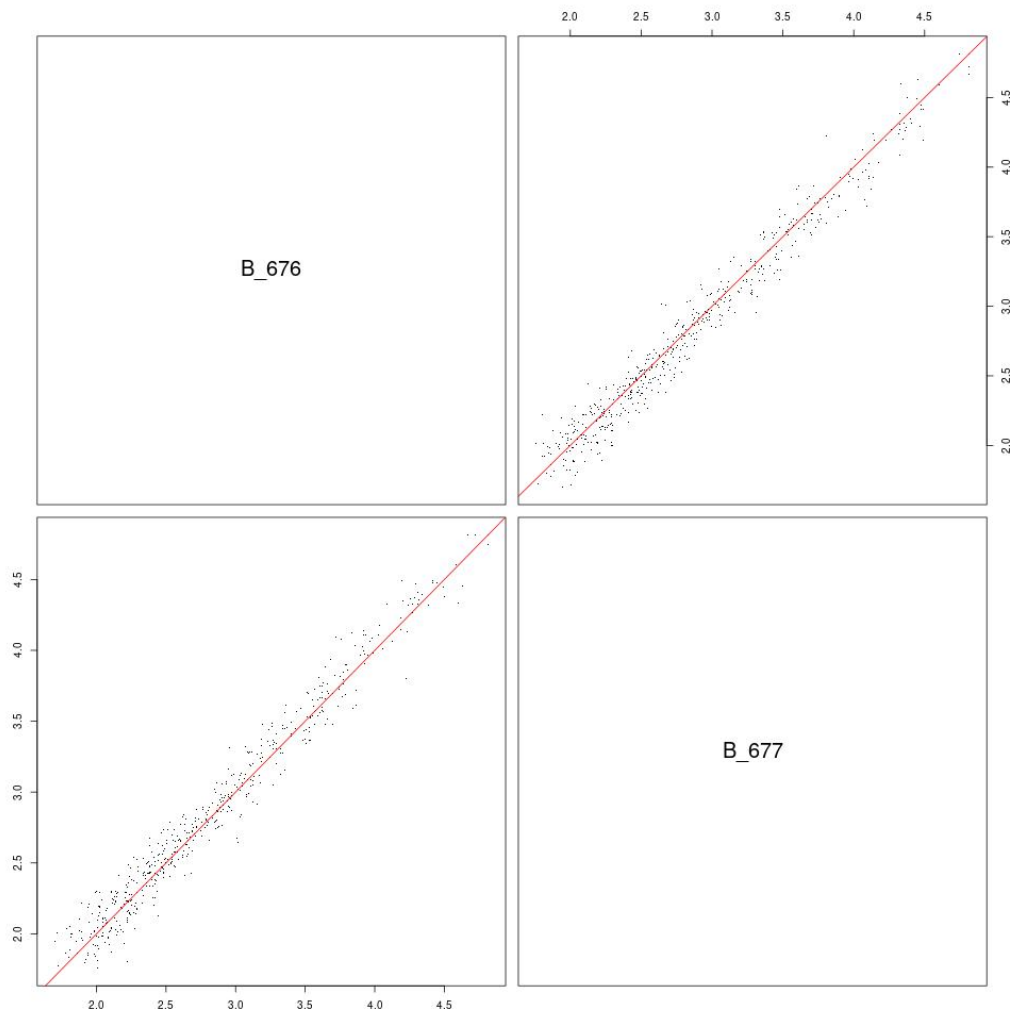File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/MiChipDemo_medNormedIntensity.txt

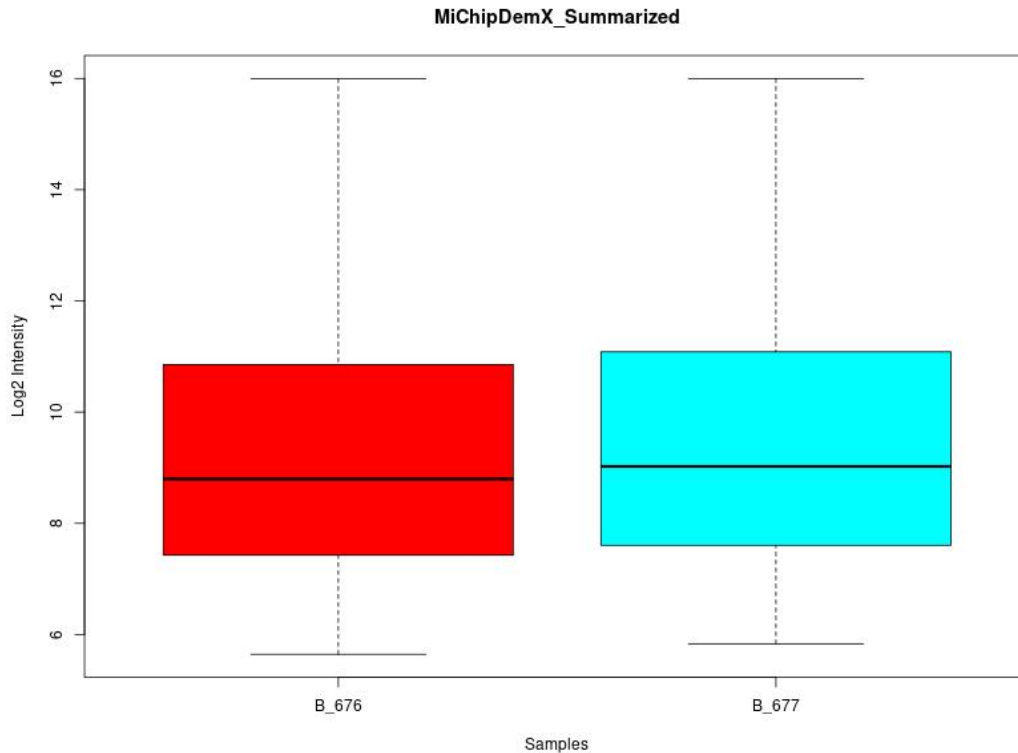Figure 1: Scatterplots of pairwise intensies per hybridization

Figure 2: Boxplot of Summarized Intensity Data

8 Combination of processes

The MiChip library has been developed to automate and simplify the analysis of MiChip hybridizations and provide a basis for incorporating the MiChip into analysis pipelines. A worked example of the analysis from file parsing to median normalization of the expression data is given in the workedExampleMedian-Normalization method.

```
> datadir <-system.file("extdata", package="MiChip")
>  myNormedEset <- workedExampleMedianNormalize("NormedDemo", intensityCutoff = 50,datadir)

2 files found
Dealing with file  B_676.gpr
Dealing with file  B_677.gpr
Raw Data  4608
Number of filters  13
Filtering special case empties
Filtered Data  1118
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_FilteredData.txt
B_676  had  55  bad flags  and  77  values below cutoff from  1118  data points
B_677  had  54  bad flags  and  86  values below cutoff from  1118  data points
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_FlagCorrected.jpg
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_FlagCorrectedIntensity
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_Summarized.jpg
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_summarizedIntensity.t
```

```
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_SummarizedScatter.jpg
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_Median Normalized.jpg
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_MedNormedScatter.jpg
File saved to /tmp/RtmposuLny/Rbuild9dea06196752/MiChip/vignettes/NormedDemo_medNormedIntensity.txt
```

References

M. Castoldi, S. Schmidt, V. Benes, M. W. Hentze, and M. U. Muckenthaler. michip: an array-based method for microrna expression profiling using locked nucleic acid capture probes. *Nature Protocols*, 3 (2):321–329, 2008. 1

M. Castoldi, S. Schmidt, V. Benes, M Noerholm, A. E. Kulozik, M. W. Hentze, and M. U. Muckenthaler. A sensitive array for microrna expression profiling (michip) based on locked nucleic acids (lna). *RNA*, 12(5):913–920, 2006.