How to use cghMCR

Jianhua Zhang Bin Feng

March 7, 2006

1 Overview

This vignette demonstrate how to use *cghMCR* to locate minimum common regions (MCR) across arrayCGH profiles derived from different samples. MCR was initially proposed by Dr. Lynda Chin's lab (Aguirre et. al. 2004) to identify chromosome regions showing common gains/losses across samples using arrayCGH platform. The *cghMCR* pacakge implements the algothrim.

2 Getting Started

The example data used in this vignette are artificially constructed following the *Agilent* arrayCGH format within 5000 probes to maintain speed.

2.1 Read the sample data

The sample data are stroed in the *data* subdirectory and can be read in using `read.Agilent` of *marray*.

```
> require("cghMCR")

Loading required package: cghMCR
Loading required package: DNAcopy
Loading required package: marray
Loading required package: limma

Attaching package: 'cghMCR'


        The following object(s) are masked from package:DNAcopy :

         plot.DNAcopy
```

```
[1] TRUE

> require("marray")

[1] TRUE

> sampleData <- read.Agilent(list.files(file.path(.path.package("cghMCR"),
+     "sampledata"), full.name = TRUE, pattern = "sample"), path = "")

Reading ...  //tmp/Rinst.30852/cghMCR/sampledata/sample1.agi
Reading ...  //tmp/Rinst.30852/cghMCR/sampledata/sample2.agi
Reading ...  //tmp/Rinst.30852/cghMCR/sampledata/sample3.agi
```

`sampleData` has three samples with intensity measures for 5000 probes.

```
> maNsamples(sampleData)

[1] 3

> length(maLabels(maGnames(sampleData)))

[1] 5000
```

It normally takes a few steps such as quality assessment and normalization before engaging data analysis. Here we conduct a loess normliztion that is supported by `marray`. Readers are referred to packages *marray* and *limma* for further information.

```
> normedData <- maNorm(sampleData, norm = "loess")
```

2.2 Identify chromosome segments

For each sample, we need to first identify chromosome segments having similar intensity measures. The function `getSegments` is a wrapper around the main functions provided by *DNAcopy* that are capable of detecting chromosome regions within which probe intensities remain similar.

```
> segments <- getSegments(normedData)

Analyzing: X..tmp.Rinst.30852.cghMCR.sampledata.sample1.agi
Analyzing: X..tmp.Rinst.30852.cghMCR.sampledata.sample2.agi
Analyzing: X..tmp.Rinst.30852.cghMCR.sampledata.sample3.agi
```

Results from the segmentation analysis (`segments`) is a list with three elements:

```
> names(segments)

[1] "data"   "output" "call"
```

The *data* element contain the normalized data, the *output* element contains the chromosome segments identified, and the *call* elements contains the function call with parameters passed indicated. Now, let's plot the original data and the segments to see what the segments look like.

```
> plot(segments)
```

2.3 Identify MCRs

The element - `output` of the `segments` object generated in the previous section contains the segmentation data and will be used to get the MCRs. The parameter `margin` is numeric indicating how many basepairs should two adjancent segments be allowed to apart to be considered as one locus. Parameters `gain.threshold` and `loss.threshold` are also numerics indicating the minimum positive or maximum negative values for chromsome segments to be considered as gains or losses.

```
> cghmcr <- cghMCR(segments[["output"]], margin = 0, gain.threshold = 0.8,
+     loss.threshold = -0.8)
> mcrs <- MCR(cghmcr)
```

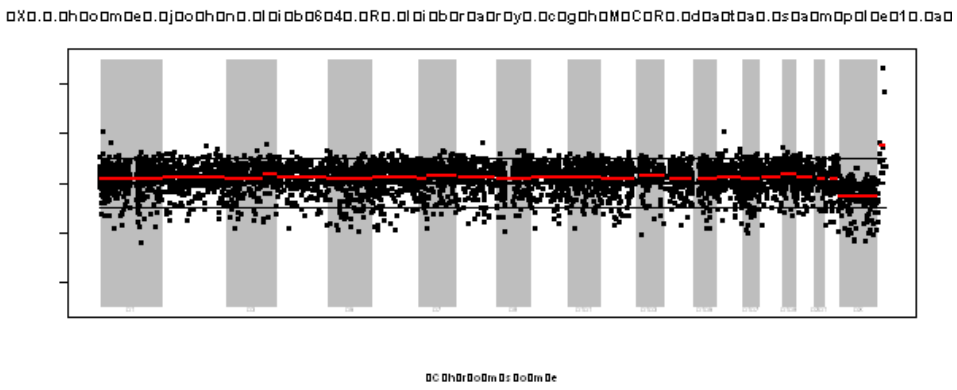
Using the above settings, we get four MCRs but only one (on chromosome 7) of them are common to two samples.

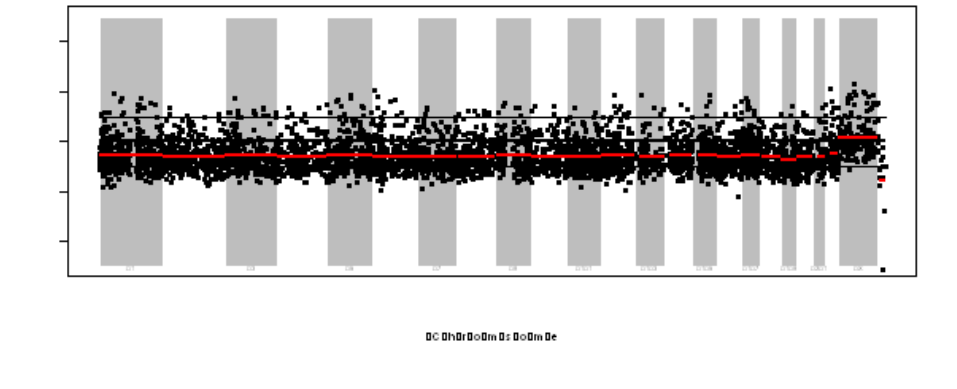
```
> print(cbind(mcrs[, c("chromosome", "status", "mcr.start", "mcr.end",
+     "samples")]))
```

```
  chromosome status mcr.start mcr.end
Y "Y"        "gain" "4283463" "26983049"
Y "Y"        "loss" "4283463" "26983049"
  samples
Y "X..tmp.Rinst.30852.cghMCR.sampledata.sample1.agi"
Y "X..tmp.Rinst.30852.cghMCR.sampledata.sample2.agi"
```

To include probe ids for the MCRs identified, we can call the function `mergeMCRProbes` to have probe ids within each MCR appended. Multiple probes are separated by a ",".

```
> mcrs <- mergeMCRProbes(mcrs, segments[["data"]])
> print(cbind(mcrs[, c("chromosome", "status", "mcr.start", "mcr.end",
+     "probes")]))
```





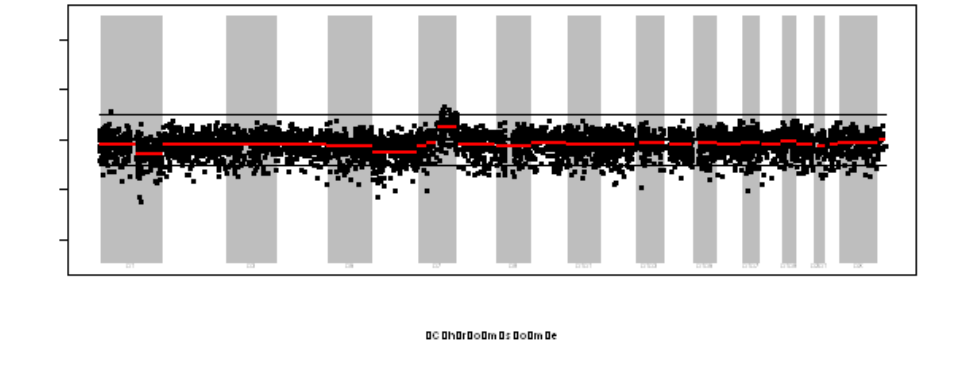


Figure 1:

```
  chromosome status mcr.start mcr.end
Y "Y"        "gain" "4283463" "26983049"
Y "Y"        "loss" "4283463" "26983049"
  probes
Y "A_14_P120393,A_14_P124965,A_14_P127818,A_14_P120961,A_14_P106297,A_14_P112182,A_14_P
Y "A_14_P120393,A_14_P124965,A_14_P127818,A_14_P120961,A_14_P106297,A_14_P112182,A_14_P
```

3 Session Information

The version number of R and packages loaded for generating the vignette were:

```
R version 2.2.0, 2005-10-06, x86_64-unknown-linux-gnu

attached base packages:
[1] "tools"     "methods"   "stats"     "graphics"  "grDevices" "utils"
[7] "datasets"  "base"

other attached packages:
 cghMCR  marray   limma DNAcopy
"1.0.0" "1.8.0" "2.2.0" "1.4.0"
```

4 References

Aguirre, AJ, C. Brennan, G. Bailey, R. Sinha, B. Feng, C. Leo, Y. Zhang, J. Zhang, N. Bardeesy, C. Cauwels, C. Cordon-Cardo, MS Redston, RA DePinho and L. Chin. High-resolution Characterization of the Pancreatic Adenocarcinoma Genome. Proc Natl Acad Sci U S A. 2004. 101(24):9067-9072.