

RMA+ and RMA++ using the RefPlus package

Kai-Ming Chang
kai-ming.chang@astrazeneca.com
kaiming@gmail.com

Oct 25, 2006

Abstract

In this vignette, we introduce the ideas behind **RMA+** and **RMA++** methods, and give examples of using the functions in this package.

1 Introduction

RMA+ and **RMA++** are Affymetrix GeneChip microarray data pre-processing methods proposed in Chang, Harbron and South (2006). **RMA+** is an extension to the **RMA** algorithm by Irizarry et al. (2004), and **RMA++** is a further extension based on the **RMA+** method.

The **RMA+** algorithm calculates the microarray intensities using a pre-stored **RMA** model trained on a reference microarray set (can be standard reference microarrays, microarrays from an independent study, or an incomplete set of microarrays in a study). **RMA+** measurements of a microarray can be considered as an approximation to the **RMA** measurements of this microarray when the microarray is **RMA**ed with the reference set microarrays in one batch.

RMA++ measurements of a microarray are the average of multiple **RMA+** measurements of a microarray based on several reference sets. If the reference sets cover more information of the microarrays to be pre-processed than a single reference set does, the **RMA++** measurements will provide a better approximation to the **RMA** measurements.

2 RMA+

RMA+ procedure:

1. Fit the **RMA** model on the reference set and store the normalizing quantiles and the estimated probe effects;
2. Background correct the probe intensities of the microarrays to be pre-processed;
3. Normalize the background-corrected probe intensities to the normalizing quantiles (reference quantiles);
4. Derive the probeset intensity using the estimated probe effects and normalized background-corrected probe intensity data.

Step 1 can be done using the `rma.para` function in the package. The normalizing quantiles and the estimated probe effects are returned. Step 2-4 can be done using the `rmaplus` function.

Both functions provide an option of skipping the background correction step. In this case, the microarrays can be background-corrected independently.

3 RMA++

RMA++ procedure

1. Fit multiple RMA models on several reference sets and store the normalizing quantiles and the estimated probe effects of these reference sets;
2. Calculate the RMA+ measurements of the microarrays of interest for each reference set;
3. Average multiple RMA+ measurements of the microarray based on these reference sets.

4 Example

4.1 RMA+

The Dilution dataset in the `affydata` package consists of 4 microarray samples.

```
> library(RefPlus)
> library(affydata)
> data(Dilution)
> sampleNames(Dilution)

[1] "20A" "20B" "10A" "10B"
```

Firstly, we calculate the RMA measurements of the 4 microarrays *Ex0*:

```
> Ex0 <- exprs(rma(Dilution))
```

Background correcting
Normalizing
Calculating Expression

Secondly, we form a reference set using the first 3 samples and derive the reference quantiles and the reference probe effects:

```
> Para <- rma.para(Dilution[, 1:3], bg = TRUE, exp = TRUE)
> Ex1 <- Para[[3]]
```

Then, we calculate the RMA+ measurements of all microarrays *Ex2*. Figure 1 compares the RMA measurements and the RMA+ measurements of these 4 microarrays.

```
> Ex2 <- rmaplus(Dilution, rmapara = Para, bg = TRUE)
```

Use `rmapara`.

```

> par(mfrow = c(2, 2))
> plot(Ex0[, 1], Ex2[, 1], pch = ".", main = sampleNames(Dilution)[1])
> plot(Ex0[, 2], Ex2[, 2], pch = ".", main = sampleNames(Dilution)[2])
> plot(Ex0[, 3], Ex2[, 3], pch = ".", main = sampleNames(Dilution)[3])
> plot(Ex0[, 4], Ex2[, 4], pch = ".", main = sampleNames(Dilution)[4])

```

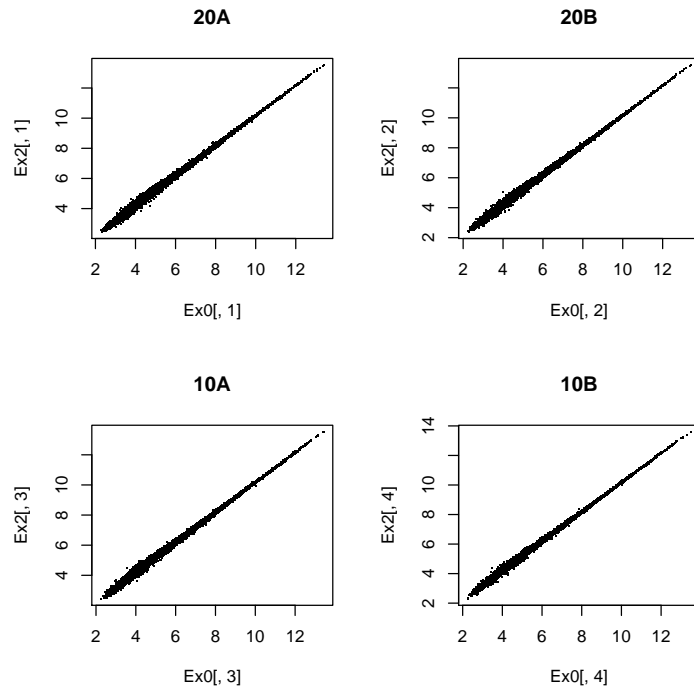


Figure 1: RMA (Ex0) vs. RMA+ (Ex2).

4.2 RMA++

Now, we form another reference set using the 2-4 samples and calculate a new set of RMA+ measurements *Ex3*.

```
> Para2 <- rma.para(Dilution[, 2:4], bg = TRUE, exp = TRUE)
> Ex3 <- rmaplus(Dilution, rmapara = Para2, bg = TRUE)
```

Use `rmapara`.

We can then obtain a set of RMA++ measurements by averaging these two sets of RMA+ measurements *Ex4*. Figure 2 compares the RMA measurements and the RMA++ measurements of these 4 microarrays.

```
> Ex4 <- (Ex2 + Ex3)/2

> par(mfrow = c(2, 2))
> plot(Ex0[, 1], Ex4[, 1], pch = ".", main = sampleNames(Dilution)[1])
> plot(Ex0[, 2], Ex4[, 2], pch = ".", main = sampleNames(Dilution)[2])
> plot(Ex0[, 3], Ex4[, 3], pch = ".", main = sampleNames(Dilution)[3])
> plot(Ex0[, 4], Ex4[, 4], pch = ".", main = sampleNames(Dilution)[4])
```

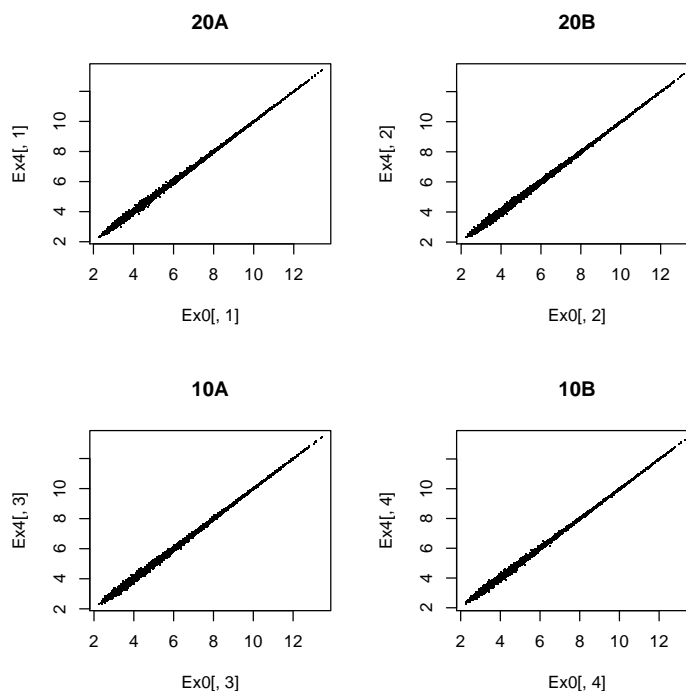


Figure 2: RMA (Ex0) vs. RMA++ (Ex4).

The root mean squares differences(RMSD) between RMA measurements and 2 RMA+ measurements, are

```
> sqrt(mean((Ex0 - Ex2)^2))
```

```
[1] 0.2138401
```

```
> sqrt(mean((Ex0 - Ex3)^2))
```

```
[1] 0.2283817
```

and the RMSD between RMA measurements and RMA++ measurements is

```
> sqrt(mean((Ex0 - Ex4)^2))
```

```
[1] 0.06549509
```

We can see that the RMA++ measurements can provide a better approximation to the RMA measurements, which is consistent with the comparison between figure 1 and figure 2.

References

- Chang, K.M., Harbron, C., South, M.C. (2006) "An Exploration of Extensions to the RMA Algorithm," *submitted*.
- Irizarry,R.A., Hobbs,B., Collin,F., Beazer-Barclay,Y.D., Antonellis,K.J., Scherf,U. and Speed,T.P. (2003) "Exploration, Normalization, and Summaries of High density Oligonucleotide Array Probe Level Data," *Biostatistics*, 4, 249-264.