

Using MM in place of PM for detecting differential gene expression

B. M. Bolstad

December 12, 2002

Introduction

We shall look at the effect of using MM in place of PM for picking differential probesets. We will use RMA methods with and without background correction.

Methods

Expression will be computed using RMA but instead on using PM we will use MM to compute the expression value. Under the three-step computation of expression measures we will background correct using either RMA (or not at all), normalize across chips using quantile normalization and then compute expression using median polish on \log_2 probe intensities.

Differential expression is the difference in log scale expression values. We will average across spike in groups and compare the groups.

Data

The data used is a spikein latin square experiment carried out by Affymetrix. This dataset was used in the creation of the MAS 5.0 algorithm. It consists of 59 chips, where 14 probesets have been spiked in a known concentrations. The concentration of the spike-ins are 0, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024pM. In most cases there are three replicate chips having the same concentration profile, however there are also two concentration profiles that are replicated 12 times each and one concentration profile is represented only twice.

Results

Table 1 shows regression slopes for observed expression difference versus true expression difference. Using MM rather than PM gave us lower slopes and lower R^2 . The MAS 5.0 like measure did the

Method	Slope	R-squares
RMA	0.607	0.966
RMA using MM	0.427	0.894
RMA nobg	0.484	0.955
RMA nobg using MM	0.273	0.821
Tukey Biweight w IMM and MAS 5 BG	0.717	0.965

Table 1: Regression slopes

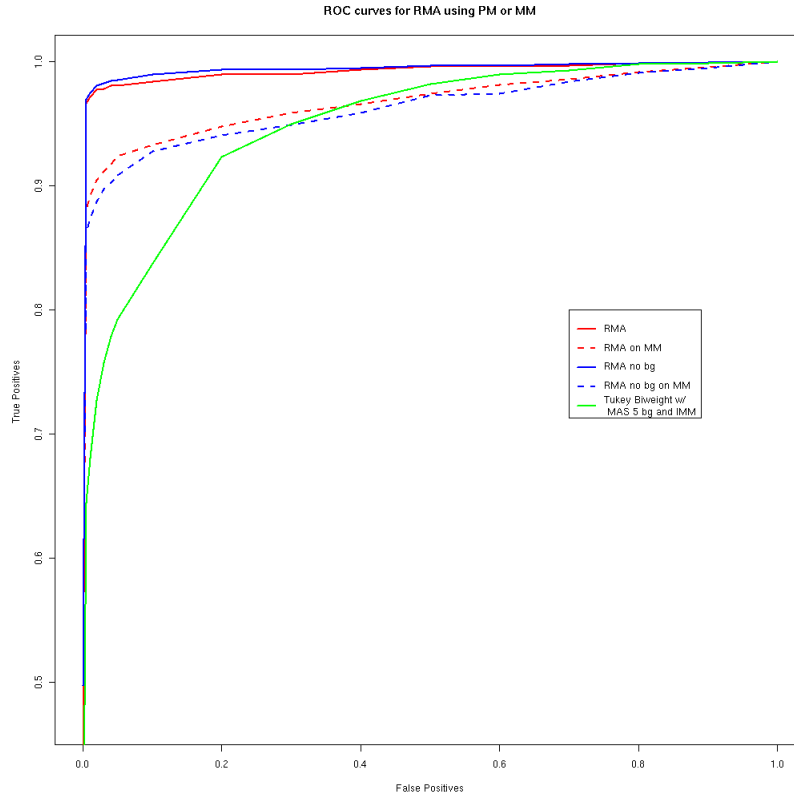


Figure 1: ROC curves comparing RMA based on MM with RMA based on PM

best in predicting true fold change.

Figure 1 presents ROC curves using observed fold change for each of the four methods, along with the MAS 5.0-like method. Using PM outperformed using MM when detecting differential probes using observed fold change. Both PM and MM RMA methods performed better than the MAS like method.

Conclusions

There is information in MM that can be used to detect differential probesets. Using MM was not quite as good as using PM in the RMA framework. All the RMA methods outperformed the MAS 5.0 alike method.

References

- [1] Bolstad, B.M., Irizarry R. A., Astrand, M., and Speed, T.P. (2003), A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. To appear in *Bioinformatics*
- [2] Affymetrix (2002) Statistical Algorithms Description Document http://www.affymetrix.com/support/technical/whitepapers/sadd_whitepaper.pdf

- [3] Irizarry, RA, Hobbs, B, Collin, F, Beazer-Barclay, YD, Antonellis, KJ, Scherf, U, Speed, TP (2003) Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. To appear in *Biostatistics*
- [4] Rafael A. Irizarry, Benjamin M. Bolstad, Francois Collin, Leslie M. Cope, Bridget Hobbs, and Terence P. Speed (2002) Summaries of Affymetrix GeneChip Probe Level Data. Accepted to *Nucleic Acids Research*