Investigating the effects of simple filtering on detecting differential gene expression

B. M. Bolstad

December 9, 2002

Introduction

Previously we have seen that MAS 5.0 like algorithms suffer in comparison to RMA like algorithms when using fold-change to detect differential gene exression. It was shown that a MAS 5.0 style background allows one to better predict observed vs true fold-change in a spike-in experiment, but raises the variability the observed foldchange for non spike-in probes. Using ROC curves we found that the mas 5.0 background with Ideal Mismatch correction did not help when using fold-chage to pick differential probesets.

Background correction adds variability at the low end of the intensities, we will investigate what effect removing these probesets has on detecting differential expression.

Methods

Computing expression

We will use three methods (from the vast array of methods surveyed previously) to compute measures of expression for each probeset. We will describe them in terms of a three step process. One method will be by RMA, that is RMA background correction, Quantile Normalization and summarization by median polish. The second will be an MAS 5.0 implementation: MAS 5.0 Background, no normalization and summarization using a 1-step Tukey biweight. The third will be RMA without background correction, ie normalize using the quantile method and then summarize using median polish.

Filtering

It is been found previously that background correction adds variability to low intensity probes and thus to the computed measures of expression. To remove the effects that this might have on detecting differential expression we will use a simple filtering method. We will filter out low intensity probesets, by filtering out fixed proportions of the data.

MAS 5.0 has probeset specific filtering based upon P/A calls. The MAS 5.0 P/A system uses information about the probes in a probeset from a single array. The filtering system we will uses uses information about the measure of expression across arrays. Clearly our filtering system is a little different, but we should capture many of the same probesets.

We will use two distinct methods of filtering and then combine the two together to form a third method of filtering.

Filtering on mean intensity

We will use average expression intensity across chips to measure probeset intensity. Using our filtering system, a probeset is called present if its mean expression across all chips is higher than a certain level.

Filtering on variability

This aim is this method is to filter out noisy probesets. We will compute the variability of probesets across arrays. Our filtering will be based on removing genes that have variability above a certain level (ie these are the probesets that will be called absent). To be fair to the spike-in probesets, which should be variable across chips, we will use their variability after taking account of the spike-in concentrations (ie use the ANOVA within spike-in concentration variabilty.

Joint Filtering

We will try to filter out genes that have both low mean and high variance. To do this we will rank probesets by variance across arrays from highest to lowest and rank by mean expression across arrays from lowest to highest. Combining the two sets of ranks by summing, to get a joint rank on both mean and variance. Use these ranks to filter out probesets.

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 haveing 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

We compare the ability of expression measures to detect differential expression by looking at Receiver Operating Characteristic (ROC) curves. We will also explore the effect that removing low intensity probesets has on the ROC curves.

Filtering on mean intensity

As expected filtering has the greatest effect on the MAS 5.0 like expression summary. This can be seen in figure 1 where we have plotted ROC curves for the tukey biweight summarization method. We have filtered out the lowest 1%, 5%, 10%, 20%, 30%, 40% and 50% of probeset intensities. The higher the intensity we filter out, the better we can pick out differential probesets. Even at the highest level of filtering no spike in probesets were removed in the filtering.

The same analysis is repeated for the RMA expression measure in figure 2. As expected the level of filtering really has no effect on the ability to discern differential expression.

We compare the three expression measure methods in figure 3. In particular we compare the two unfiltered RMA methods to the 50% filtered Tukey biweight method. Even with heavy filtering

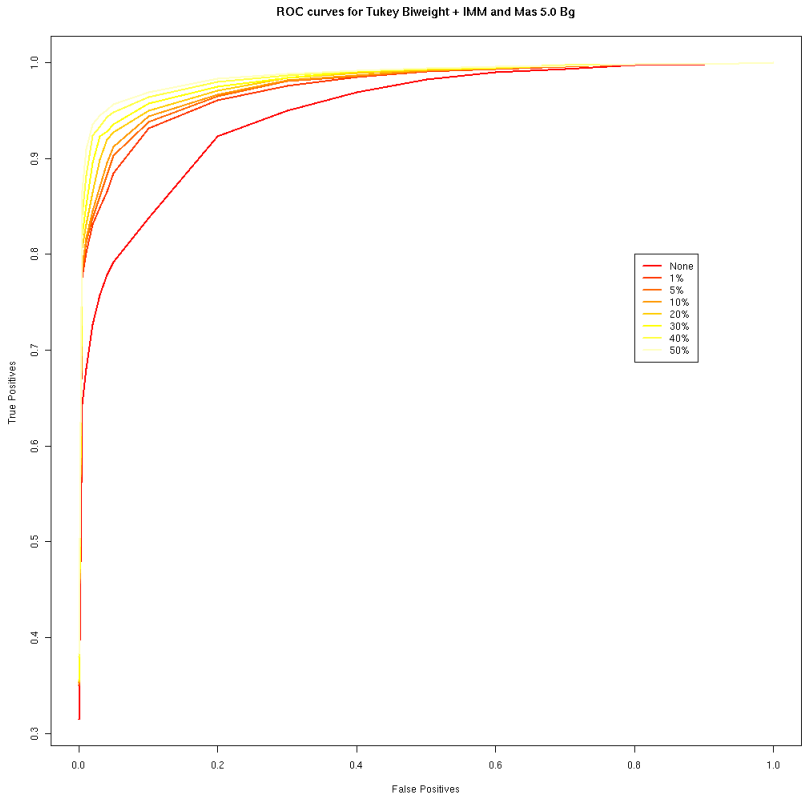

Figure 1: ROC curves for Tukey Biweight + Mas5.0bg + IMM at different levels of filtering

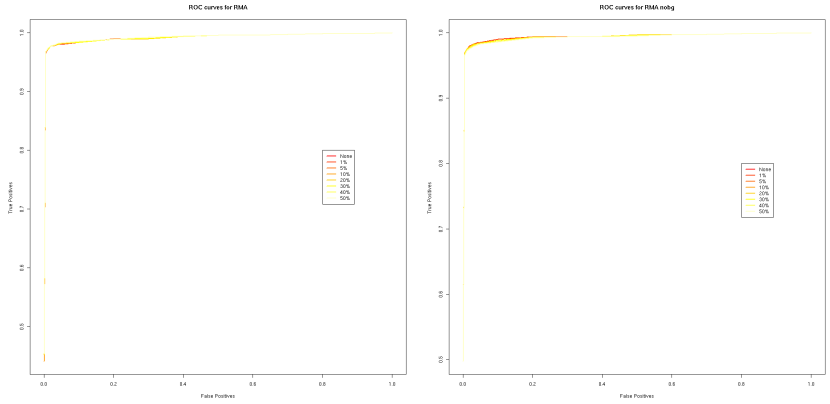

Figure 2: ROC curves for RMA methods accross levels of filtering

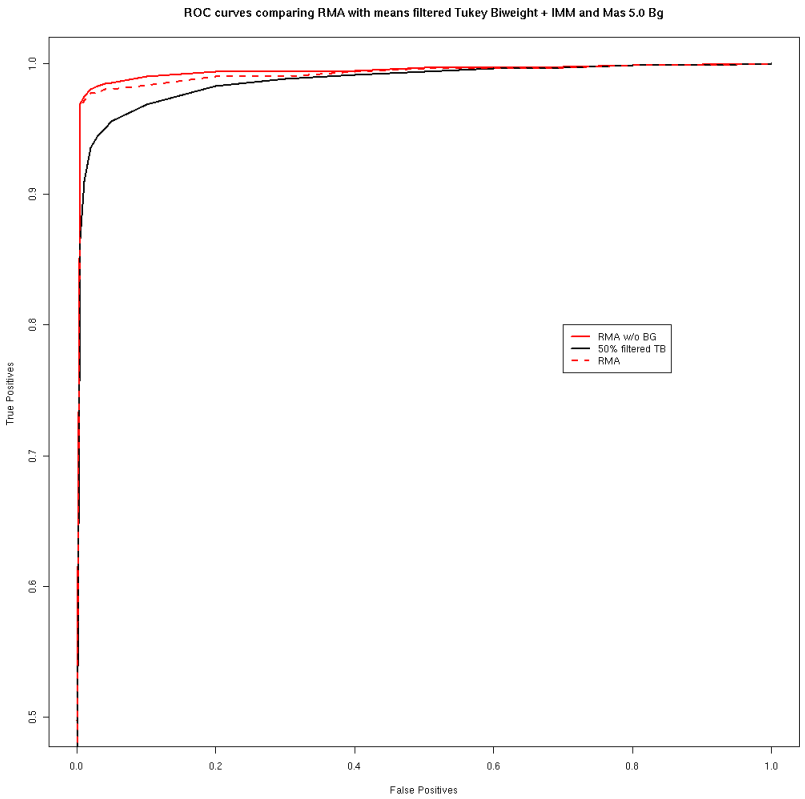

Figure 3: ROC curves comparing RMA with means filtered Tukey Biweight Mas 5.0 bg +IMM

Filtering level (%)	RMA	Tukey Biweight	RMA no bg
0	14	14	14
1	14	14	10
2.5	14	14	3
5	14	14	2
7.5	13	14	1
10	13	14	1
20	8	14	0
30	6	14	0
40	5	14	0
50	3	14	0

Table 1: Number of spikeins present at each level of variance filtering

the Biweight method with the MAS5.0 background does not detect differential expression well using fold change.

Filtering on variability

We repeat the same analysis using our variability criteria to filter out genes. There was one problem with this method when applied to the RMA methods, even after adjusting the variance of the spikeins, the spikein variances were still a little higher than average, so filtering on the variances resulted in some of the spikein probesets being filtered out. This effect can be seen in table 1.

Figure 4 shows the effects of the filtering on ability to detect differential expression using the tukey biweight measure. Heavy filtering improves our ability to detect differential expression, just as it did when using filtering by mean expression. Figure 5 shows the effects of the filtering on detecting differential expression using the RMA methods. The filtering really had no major effect on detecting differential expression when using an RMA measure.

We compare the variance filtered Tukey Biweight with with the unfiltered RMA methods in figure 6. Even after filtering 50% of the probesets out, the RMA methods are still detecting differential expression better.

Joint filtering

Finally we apply the joint filtering method. The levels of filtering can be see in figure 7. This figure plots $\log_2$ variance versus mean expression for the Tukey biweight expression measure. Our previous filtering methods would correspond to horizontal or vertical strips on this plot. Given that mean and variance filtering really left the RMA measures ability to detect differential expression unchanged we have applied the joint filtering only to the tukey biweight method.

Figure 8 shows the effects of joint filtering on the ability to detect differential expression with the Tukey Biweight data. As before filtering out more data, improves out ability to detect differential expression.

Now comparing the three filtering methods at the 50% level in figure 9 we see that using mean filtering gave the best results for detecting differential expression for the tukey biweight. But As we have seen before the RMA methods outperform all the Tukey Biweight methods.

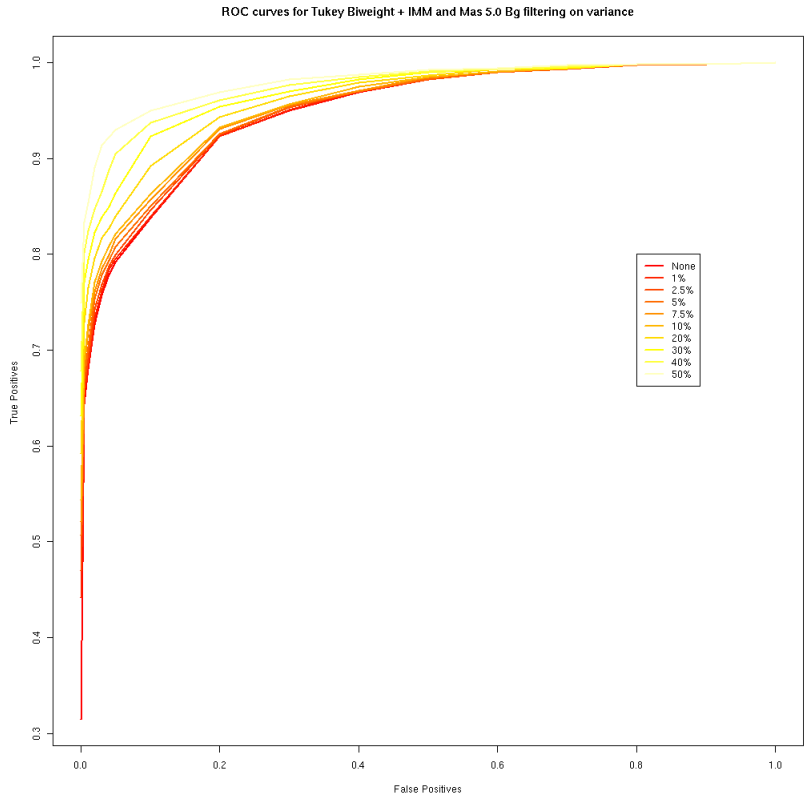

Figure 4: ROC curves for Tukey Biweight + Mas5.0bg + IMM at different levels of variance filtering

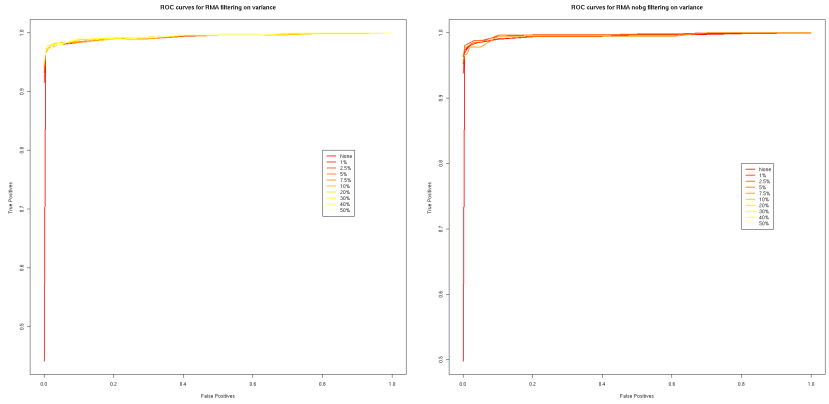

Figure 5: ROC curves for RMA methods across levels of variance filtering

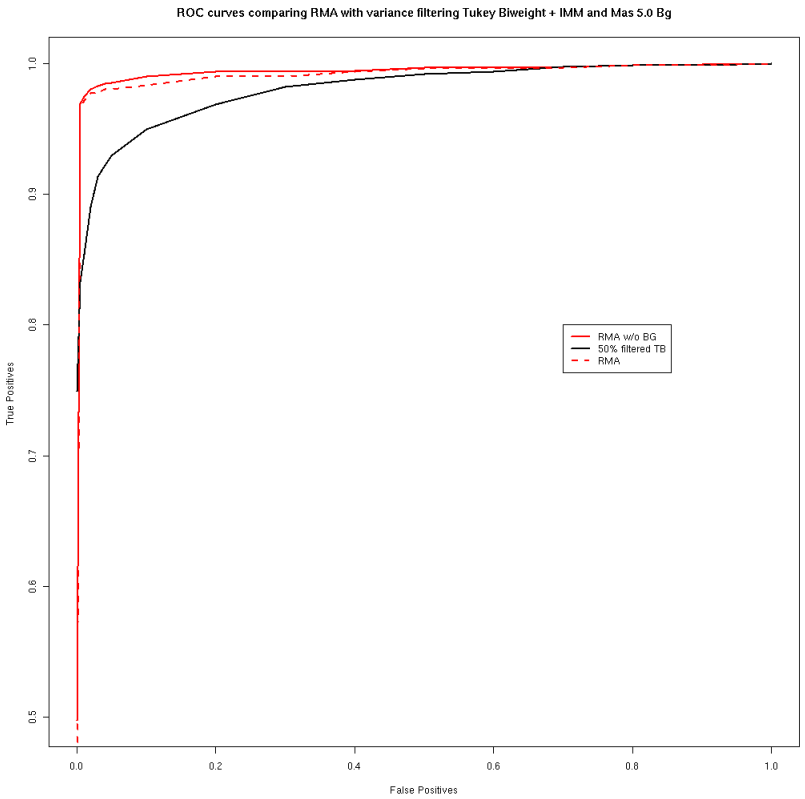

Figure 6: ROC curves comparing RMA with variance filtered Tukey Biweight Mas 5.0 bg +IMM

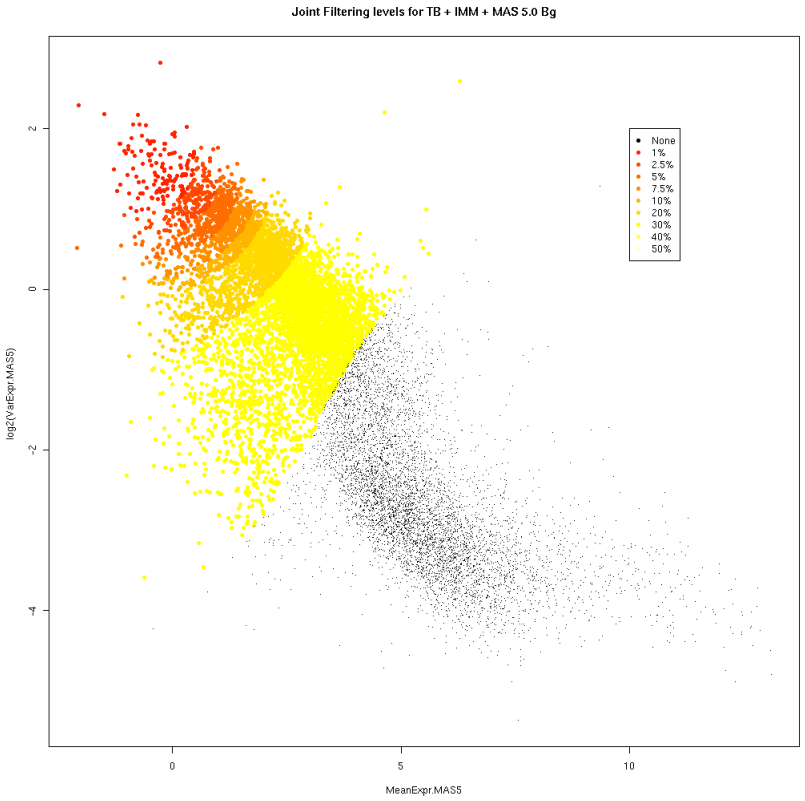

Figure 7: $\log_2$ Variance vs Mean joint filtering levels

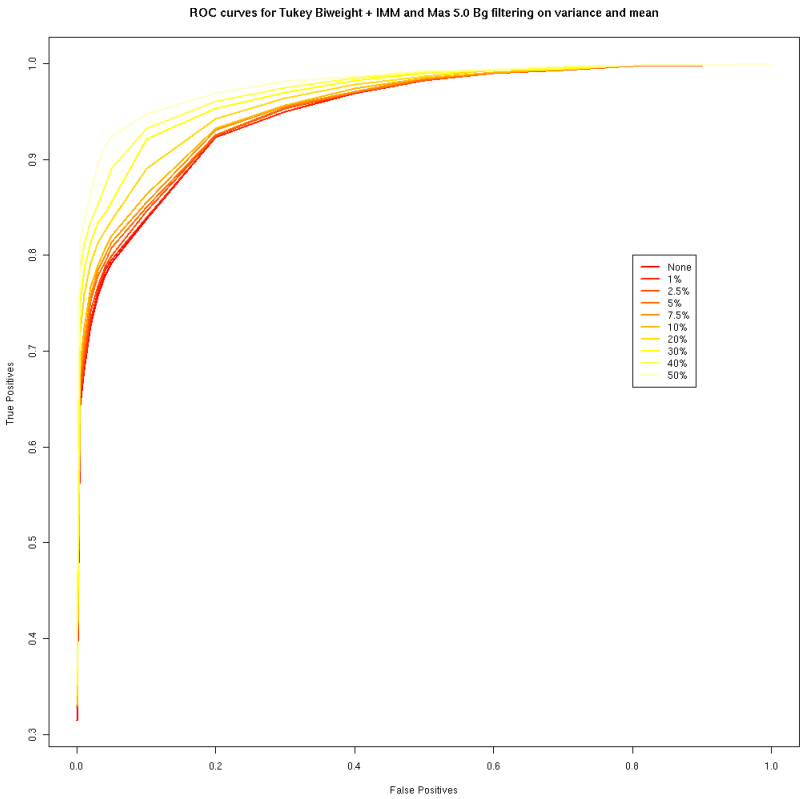

Figure 8: ROC curves for Tukey Biweight + Mas5.0bg + IMM at different levels of joint filtering

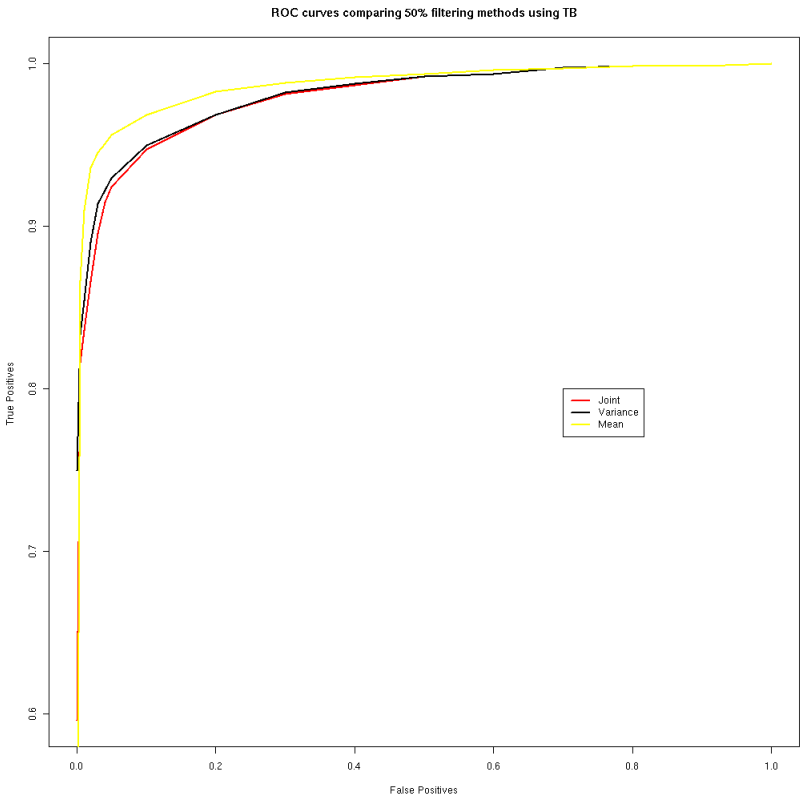

Figure 9: ROC curves for Tukey Biweight + Mas5.0bg + IMM using 50% filtering across filtering methods

Conclusions

For the MAS 5.0 like (Tukey Biweight) method filtering out low intensity probesets and/or high variabilty probesets did help when using observed fold change to detect differential probesets. These filtering methods, as one would expect, did not have much effect on the RMA measures. Filtering on mean probeset intensity provided the best results.

We comparing ability to detect differential expression across methods the RMA methods still did better then the best case examined when using filtering using the Tukey biweight 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