BIOINFORMATICS

Vol. 23 ISMB/ECCB 2007, pages i282–i288
doi:10.1093/bioinformatics/btm201

Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks

Wei Keat Lim[1,2], Kai Wang[1,2], Celine Lefebvre[2] and Andrea Califano[1,2,*]

[1]Department of Biomedical Informatics, Columbia University, 622 West 168th Street, Vanderbilt Clinic 5th Floor and [2]Center for Computational Biology and Bioinformatics, Columbia University, 1130 Saint Nicholas Avenue, New York, NY 10032, USA

ABSTRACT

Motivation: An increasingly common application of gene expression profile data is the reverse engineering of cellular networks. However, common procedures to normalize expression profiles generated using the Affymetrix GeneChips technology were originally developed for a rather different purpose, namely the accurate measure of differential gene expression between two or more phenotypes. As a result, current evaluation strategies lack comprehensive metrics to assess the suitability of available normalization procedures for reverse engineering and, in general, for measuring correlation between the expression profiles of a gene pair.

Results: We benchmark four commonly used normalization procedures (MAS5, RMA, GCRMA and Li-Wong) in the context of established algorithms for the reverse engineering of protein–protein and protein–DNA interactions. Replicate sample, randomized and human B-cell data sets are used as an input. Surprisingly, our study suggests that MAS5 provides the most faithful cellular network reconstruction. Furthermore, we identify a crucial step in GCRMA responsible for introducing severe artifacts in the data leading to a systematic overestimate of pairwise correlation. This has key implications not only for reverse engineering but also for other methods, such as hierarchical clustering, relying on accurate measurements of pairwise expression profile correlation. We propose an alternative implementation to eliminate such side effect.

Contact: califano@c2b2.columbia.edu

1 INTRODUCTION

Affymetrix Genechip® arrays are currently among the most widely used high-throughput technologies for the genome-wide measurement of expression profiles. To minimize mis- and cross-hybridization problems, this technology includes both perfect match (PM) and mismatch (MM) probe pairs as well as multiple probes per gene (Lipshutz *et al.*, 1999). As a result, significant preprocessing is required before an absolute expression level for a specific gene may be accurately assessed. Such data preprocessing steps—which combine multiple probe signals into a single absolute call—are known as normalization procedures. They usually involve three steps: (a) background adjustment, (b) normalization and (c) summarization (Gautier *et al.*, 2004). Various methods have been devised for each of the three steps and thus a great number of possible combinations exist, facing the microarray user community with a complex and often daunting set of choices. We summarize some of the commonly used procedures in Table 1.

As more and more preprocessing methods become available, it is increasingly important to rigorously and systematically benchmark their performance. Cope and colleagues (Cope *et al.*, 2004) developed a graphical tool to evaluate normalization procedures that benefits users in identifying the best method in their study. The benchmarking system took advantage of dilution and spike-in experimental procedures, yielding materials where the actual concentrations of some mRNA were known a priori. The performance of a normalization method would then be ranked based on the overall error estimate in the prediction of the concentration of these mRNAs (Bolstad *et al.*, 2003; Liu *et al.*, 2005). A different evaluation framework was recently proposed, which is based on the analysis of the correlation between the expression levels of genes in replicate samples as well as the correlation among same-operon genes in bacteria (Harr and Schlotterer, 2006). Correlation-based analysis was also investigated by varying the normalization methods of RMA procedure, in order to provide a quantitative assessment of their effects on gene–gene correlation structure (Qiu *et al.*, 2005). While the former type of comparative approach identifies method that best differentiates concentration levels of RNA transcript, the latter favors methods that can optimally identify an expected correlation between gene pairs. However, none of these comparative frameworks studies whether the normalization procedure may introduce correlation artifacts for gene pairs that are not expected to be co-expressed. As a result, they also fail to address the suitability of these methods to the reconstruction of cellular networks from expression profile data, including the inference of networks topological properties and gene functional relationships based on co-expression measurements (Basso *et al.*, 2005; Butte and Kohane, 2000; Hughes *et al.*, 2000). In these methods, artifacts in the correlation measure can dramatically increase the number of inferred false-positive interactions.

In this article, we summarize the effects of various normalization procedures on the accurate estimate of gene expression profiles, both in terms of accuracy and of artifact minimization. Furthermore, we study their efficacy of protein–protein interaction (PPI) inference in a reverse engineering context. The flowchart shown in Figure 1 illustrates the comparative methodology adopted in this article. In particular, we compare the Spearman rank correlation between gene expression profile pairs from replicate samples as well as from samples with randomly permuted probe values. This allows to assess

*To whom correspondence should be addressed.

Table 1. Summary of four commonly used normalization procedures

Procedure	Background correction	Normalization	Summarization	Reference
MAS5	Ideal (full or partial) MM subtraction	Constant	Tukey biweight	Hubbell *et al.*, 2002
RMA	Signal (exponential) and noise (normal) close-form transformation	Quantile	Median polish	Irizarry *et al.*, 2003
GCRMA	Optical noise, probe affinity and MM adjustment	Quantile	Median polish	Wu *et al.*, 2004
Li–Wong	None	Invariant set	Multiplicative model fitting	Li and Wong, 2001

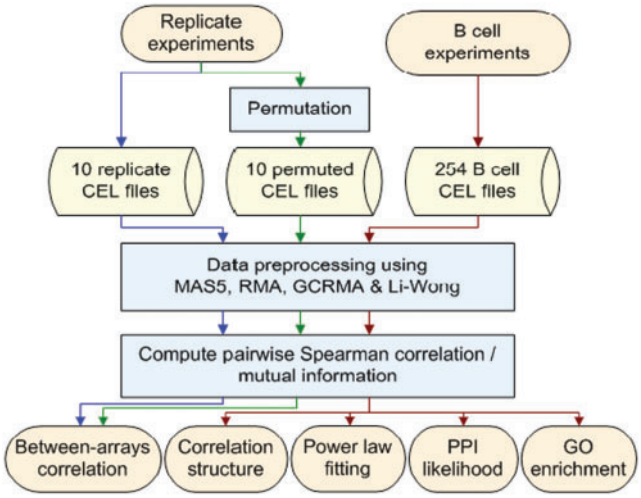


Fig. 1. Flowchart for the comparative analysis of normalization procedures. Arrows in the chart show the flow of the data sets (blue: data set with replicate samples, green: randomized data set, red: B-cell data set).

both true and artifact correlations. One unique feature of the analysis is that we permuted the raw intensity values stored in the Affymetrix *CEL* files to estimate deviations from the null hypothesis, where the expression profile data is completely uncorrelated before normalization. We also utilize a data set that consists of 254 expression profiles from normal and tumor related human B-cells to investigate the correlation structure among the gene expression profiles, as well as the global gene network connectivity. This data set has been used extensively in the literature (Margolin *et al.*, 2006; Wang *et al.*, 2006) and, as a result, it provides a unique opportunity to evaluate correct and incorrect inferences in a reverse-engineering settings.

Gene co-expression has been successfully used to infer functional relationship (Roberts *et al.*, 2000; Stuart *et al.*, 2003). We thus tested, for each normalization procedure, the hypothesis that highly co-expressed gene pairs are more likely to participate in the same biological pathways than those uncorrelated, by using biological process annotations from Gene Ontology (GO) (Ashburner *et al.*, 2000). To further address the issue of whether higher correlation reflects a higher probability of physical interaction, we exploit the approach as in (Jansen *et al.*, 2003) to compute a likelihood ratio for PPIs for gene pairs showing various degrees of correlation. The method relies on the well-justified hypothesis that proteins involved in a complex tend to be encoded by co-regulated genes, because it is energetically advantageous for the cell to synthesize them in stoichiometric balance (Ge *et al.*, 2001). Thus, an increasing PPI likelihood ratio should reflect an increasing probability of a bona fide physical interaction and correlation artifacts should dilute that relationship. The proposed evaluation strategies finally assess how well these normalization procedures fit in the context of algorithms that rely on statistical dependencies among gene expression profiles, such as the ones used to reverse engineer gene networks.

2 METHODS

2.1 Microarray data

Generations of microarray replicates are described in detail in Tu *et al.* (2002). In brief, mRNA from the Ramos human Burkitt's lymphoma cell line is used for the experiments. The purified sample is separated equally into several subgroups and each subgroup independently goes through the preparation steps. The final target sample is then divided into several samples and independently hybridized to 10 different Affymetrix HGU95A arrays. The data set used for investigating gene co-expression consists of gene expression profiles from 254 naturally occurring phenotypic variations of human B-cell. It represents a wide variety of homogenous B-cell phenotypes derived from normal and tumor related populations. The microarray experiments are described in Basso *et al.* (2005) and the *CEL* files are available on the Gene Expression Omnibus website (series accession number: GSE2350).

2.2 Permutation of *CEL* file

Raw signal intensities for each probe pairs were randomly permuted to create uninformative *CEL* files. We retained the relative position between PM and MM for every probe pairs, in order to ensure fair comparison between normalization procedures that utilize MM information to correct for non-specific binding and those that rely entirely on PM intensities. However, shuffling the probe pairs has been sufficient to destroy real signal of the probe sets as they now consist of random probes values. This data is crucial in our comparative study as the null set should not contain any information.

2.3 Normalization procedures

We compared the four normalization procedures MAS5, RMA, GCRMA and Li–Wong, and all the normalization were implemented using software packages available from Bioconductor (http://www.bioconductor.org). We used the default parameters from the software

packages unless otherwise specified. The term 'Li–Wong' refers to the procedure that normalizes arrays using invariant set of genes and then fits a parametric model to the probe set data, as described in Li and Wong (2001).

2.4 Evaluation of biological function relationship

GO annotations of the genes were extracted from Affymetrix HGU95 annotation file. There are 10 369 terms for biological process in total and 61 general terms were removed. We are interested only in specific terms that are shared by <5% of the genes in the microarray. A gene pair sharing a common GO term is then deemed functionally related.

2.5 Likelihood ratio of protein–protein interaction

We assembled a set of gold-standard positive interactions by taking the union of interaction data from the Human Protein Reference Database (HPRD), the Biomolecular Interaction Network Database (BIND), the Database of Interacting Proteins (DIP) and IntAct (Bader *et al.*, 2003; Hermjakob *et al.*, 2004; Peri *et al.*, 2003; Xenarios *et al.*, 2002). The resulting gold-standard positive set consists of 21 509 unique PPIs (heterodimers only) that could possibly pair up among genes in the Human Genome U95 array. A negative gold-standard is harder to define, but we took the common approach by taking the lists of protein pairs that are unlikely to interact given their cellular localization. The assembled negative set contains 6 101 360 pairs of proteins encoded by genes represented on the U95 array. The likelihood ratio is computed as the fraction of conditional probabilities for a set of protein pairs, here the top predicted gene pairs ranked by statistical dependency between expression profiles, given the gold-standard positive (*pos*) and negative (*neg*) sets:

$$LR = \frac{P(\text{coexpressed pairs}|pos)}{P(\text{coexpressed pairs}|neg)}$$

3 RESULTS

A common approach used to evaluate a normalization procedure is to compare correlation coefficient between replicate samples. We compared four normalization procedures, MAS5, RMA, GCRMA and Li–Wong, on gene expression measurements of 10 replicate samples as well as on their permuted data files. The randomized data set plays the role of a negative control (null-hypothesis) such that any significant correlation measured on the permuted dataset could be deemed an artifact of the specific normalization procedure. Figure 2a shows the comparison of between-sample Spearman rank correlation among the four normalization procedures. While all four procedures achieve correlation >0.9, GCRMA seems to produce higher overall correlation measures than the other methods, while MAS5 appears to produce the lowest overall correlation measures. This may be incorrectly interpreted to imply that GCRMA normalization outperforms the other methods. However, Figure 2b provides a completely different interpretation for these observations. It shows that both RMA and especially GCRMA produce highly significant correlation measurements even when applied to the randomized set. The conclusion is that the higher overall correlation after normalization with these two methods is an artifact and will likely skew the results of reverse-engineering methods.

In several methods for the reverse engineering of cellular networks, physical interactions—including protein–protein and protein–DNA interactions—are inferred from the statistical dependencies between gene expression profiles (Basso

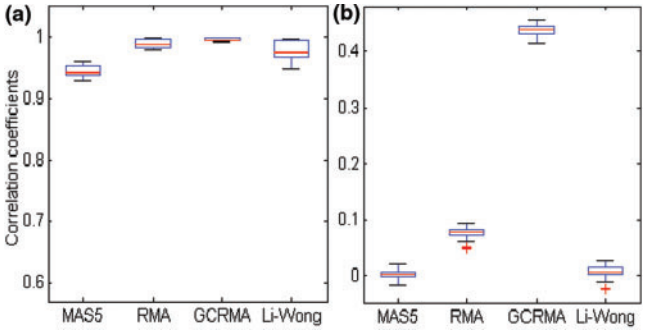


Fig. 2. Comparison of Spearman rank correlation between arrays. Each box plot represents a distribution of 45 points of correlation coefficients in (**a**) replicate data set, and (**b**) randomized data set. RMA and GCRMA are both significantly deviate from zero in (b), with *P*-values 3×10^{-17} and 0 (below MATLAB computational precision), respectively.

et al., 2005; Ge *et al.*, 2001). A correct estimate of the correlation structure in the gene expression profile data is thus one of the most crucial ingredients of a successful reverse-engineering algorithm. To estimate the impact of normalization procedures on the correlation structure, we preprocessed a set of 254 Affymetrix arrays using the four normalization procedures and then computed correlation between all pairs of probe sets. Figure 3 provides a global view of the correlation structure in these data sets. In particular, out of 77 millions possible probe set pairs, there are 5.2 millions (6.7%) expression pairs with correlation coefficient, $|\rho|$, >0.75 in GCRMA-normalized data set. In contrast, MAS5-normalized data set contains only 0.04% or ~33 000 pairs above this cutoff value. This is an extraordinary difference which warrants further investigation, especially in light of the previous results from the analysis of a randomized set. Since most biological networks are known to be scale-free, or possess a degree distribution that can be approximated by a power law (Barabasi and Oltvai, 2004), we varied threshold of a relevance network (Butte and Kohane, 2000) and fit the global network connectivity to a power-law distribution. Pairwise mutual information (MI) of the network was estimated using a Gaussian kernel method (Margolin *et al.*, 2006). Figure 4 compares R^2-value of the fitting in each of the four normalized data sets. With the exception of GCRMA, all other networks show a good fit of scale-free distribution and the R^2 reach a plateau above threshold *P*-value of 1×10^{-10}. While this is just a sanity check rather than a fully quantitative result, it should be clear that a normalization method resulting in a large deviation from the accepted model of topological connectivity in the cell should be approached with caution.

We finally proceeded to assess whether higher correlation, after a specific normalization procedure, would reflect a higher chance of either functional or physical interaction between two genes. Two non-parametric gene-pair correlation measures were tested, including Spearman rank correlation and MI. Here we only report results for MI analysis, as both measurements produce consistent results in all the tests we performed. MI was chosen as it arguably provides

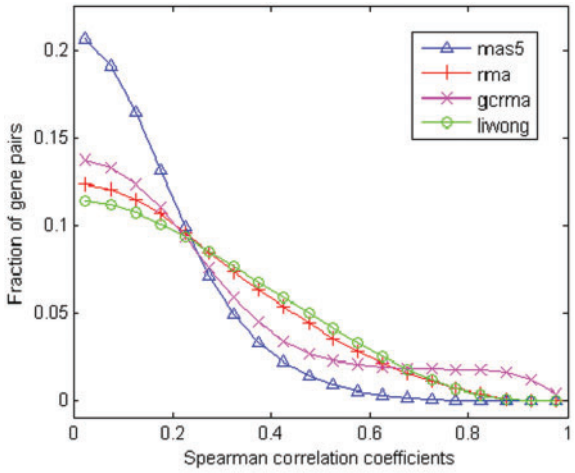


Fig. 3. Histogram of the correlation coefficients between gene expression profiles in the data sets produced by four different normalization procedures. *X*-axis corresponds to the Spearman correlation coefficient of 20 equal-size bins and *y*-axis corresponds to the count of each bin as a fraction of the total number of all possible pairs.

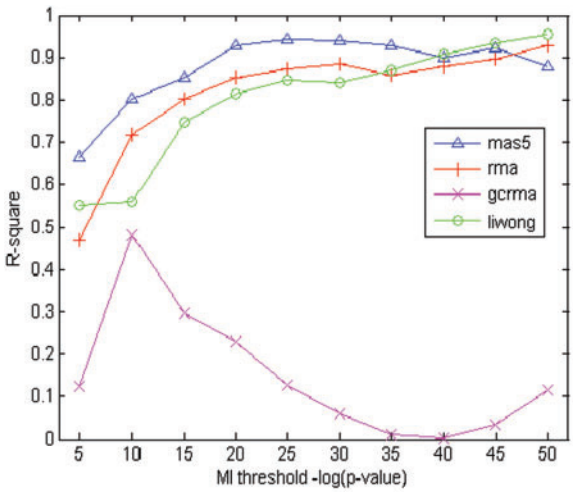


Fig. 4. Fitting of the networks connectivity to a power-law distribution.

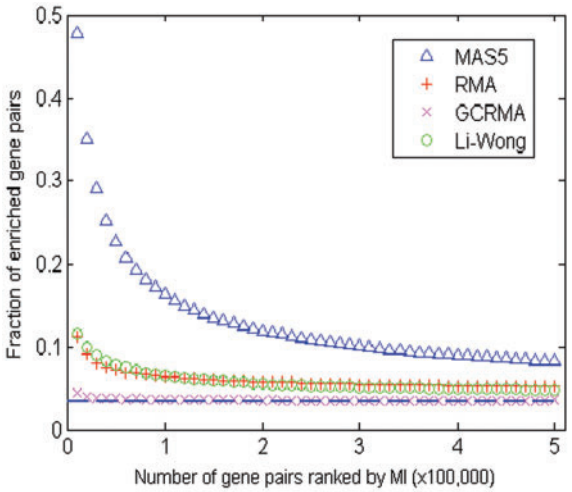


Fig. 5. Fraction of the highly correlated gene pairs sharing the same GO biological process. Gene pairs are ranked by mutual information.

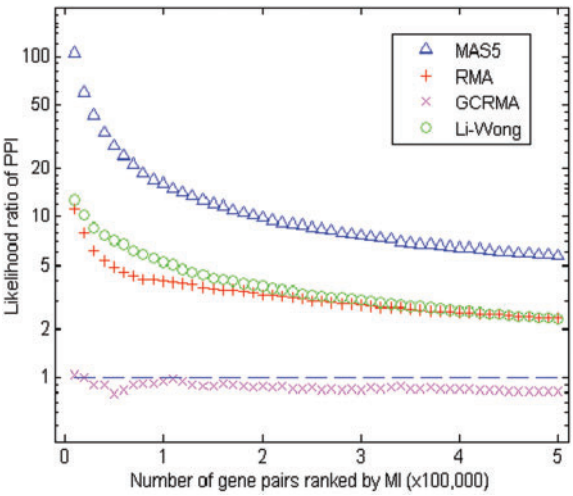
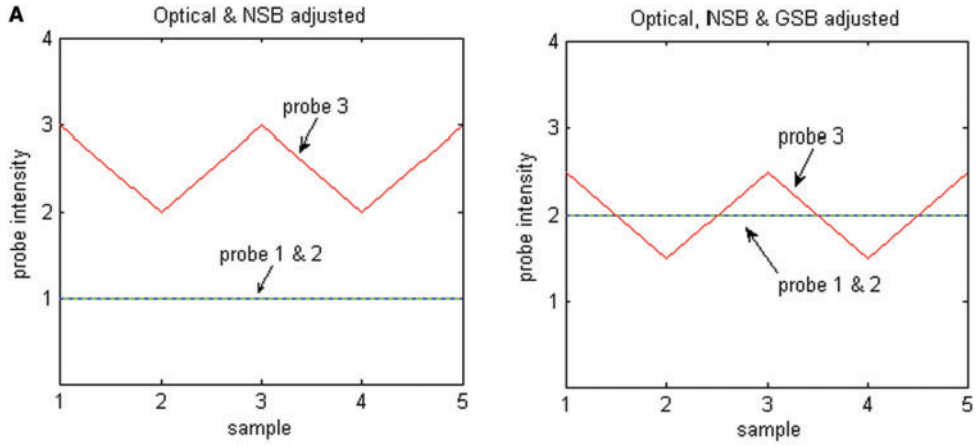


Fig. 6. Likelihood ratio of PPI for various ranges of the gene-pair correlation.

the best estimate of pairwise statistical dependency in a non-linear setting. We first evaluated functional relationship between genes by examining whether they share a common GO biological process annotation. Figure 5 compares the fraction of gene pairs with the same GO annotation in a cumulative equal-frequency histogram. MAS5 demonstrated the best result with ~48% of the top 10 000 pairs having a common GO biological process term, followed by RMA and Li–Wong, while GCRMA produces a fraction that is comparable to the background level. We then computed likelihood ratios of PPIs for the top correlated gene pairs using the gold-standard PPI interaction sets described in the Methods section. Although this is a rather naïve method that directly correlates physical interaction with gene co-expression, the results should still be able to provide a fair comparison among the four normalization procedures given that the same assumption is made for all of them. Figure 6 shows the plots of the PPI likelihood ratio as a function of various ranges of gene–gene correlations. Note that the discrepancy of performance between the curves implies that the MI ranks of the gene pairs are not consistent among the four normalization procedures. Gene pairs found to be highly correlated in one data set may not be significant in another data set. Our results show that MAS5-normalized data provide by far the best platform for inferring PPIs. To our surprise, yet consistently with all other tests in this article, data normalized by GCRMA dramatically scrambled the ranks of correlation among gene pairs, i.e. highly correlated gene pairs are equally likely to be a positive PPI or a negative PPI. This is likely the consequence of correlation artifacts resulting in the introduction of a large number of gene pairs that are not truly correlated among the top most correlated ones.



B

	s1	s2	s3	s4	s5
p3	3	3	3	3	3
p2	1	1	1	1	1
p1	1	1	1	1	1

	s1	s2	s3	s4	s5
p3	3	1	3	1	3
p2	1	2	1	2	1
p1	1	2	1	2	1

Fig. 7. A hypothetical case explaining the cause of spurious correlation in GCRMA-normalized data set. (**A**) Intensity profiles, and (**B**) intensity ranks, for three probes before (left) and after (right) GSB adjustment. Before GSB adjustment, probe 1 and 2 have the lowest intensities, $m = 1$, and the lowest ranks in the data set. If probe 1 and 2 were adjusted for the same value due to their similarity in probe affinity, and probe 3 was adjusted for a different value such that the intensity profile crosses over the other two profiles, the expression ranks of p1 and p2 change over the samples. Pairwise rank correlation between p1 and p2 is then tremendously increased. The effect of probe 3 is overly simplified in this hypothetical case and the actual data should contain a combinatorial effect of many other possible probes in the array.

The results of this analysis are both surprising and concerning. GCRMA has been a popular procedure used to convert raw microarray data into gene expression profiles and it was shown to outperform other normalization procedures in detecting differentially expressed genes (Wu *et al.*, 2004). However, we observed the opposite result when correlation artifacts are considered. This does not just affect reverse-engineering methods, but any other method that relies on an accurate measure of gene-pair expression profile correlation, such as hierarchical clustering among many others.

It is rather obvious that the dramatic under-performance of GCRMA is due to its background adjustment step, since (a) it is the only step where GCRMA and RMA differ and (b) RMA has been performing much better than GCRMA in our study, albeit not as well as MAS5. The background adjustment in GCRMA consists of three sequential steps: (1) optical background correction, (2) probe intensity adjustment through non-specific binding (NSB) utilizing affinity information and optical noise-adjusted MM intensities and (3) probe intensity adjustment through gene-specific binding (GSB), where NSB-adjusted PM intensities are further corrected for the effect of PM probe affinities. In step (2), the default GCRMA procedure implemented in the R statistical package truncates PM intensities to a minimum value, m, if the NSB-adjusted intensity values are less than m, and in step (3), these truncated PM probes are further adjusted for GSB.

We suggest that the GSB adjustment of truncated values is a significant flaw in the design of the GCRMA normalization procedure and that it is directly responsible for the difference in performance between RMA and GCRMA. From a theoretical perspective, once a probe intensity is truncated at m, it should be deemed uninformative and further adjustments should be avoided. More specifically, if any two probes with similar affinities are truncated in the same subset of samples, GSB adjustment could introduce correlation artifacts between the two probes. Figure 7 demonstrates a simplified scenario, where two probes with similar affinity, and intensities both truncated to m, could gain a high correlation after GSB adjustment. The GCRMA procedure applies quantile normalization after the background adjustment steps, where all probe intensities are essentially transformed to ranks within each sample. Without GSB-adjustment, these two probes will both rank the lowest in all samples as their intensities are truncated. However, if GSB adjustment is applied, they may switch ranks with other untruncated probes in some samples after the adjustment, and such rank switches can be highly correlated between the two probes owing to their similar probe affinity. One possible solution here is to decrease the value of m, in order to reduce the truncated regions as well as the number of the affected probes. However, in practice, we found that the most effective way to reduce this problem is to avoid further GSB adjustment altogether on the probes with truncated intensities. Note that any algorithm used to compute non-parametric statistical dependencies between probe-pairs should randomize the rank of equal-value entries such that any probe pair with a constant expression level across samples should not contribute to the correlation.

To test our speculations, we reimplemented the GCRMA procedure without adjusting GSB for uninformative probes—i.e. probes that are truncated to m after NSB adjustment.

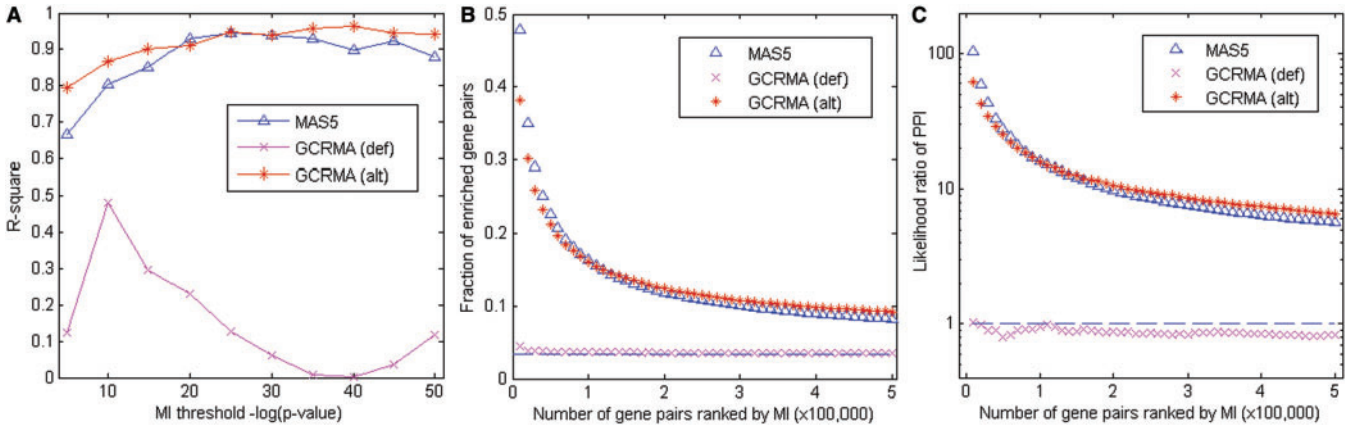


Fig. 8. Comparison of the GCRMA default (def) normalization procedure, GCRMA alternative (alt) implementation and MAS5 in terms of (**A**) fitness of network connectivity to a power-law distribution, (**B**) fraction of gene pairs sharing a common GO biological process annotation and (**C**) likelihood ratio of PPI.

To ensure the lowest intensity rank of these probes, any other probes with GSB-adjusted value less than m were also truncated at m. Finally, an infinitesimal amount of uniformly distributed noise was added to truncated probes to avoid rank-order correlation issues. The new implementation successfully removed the artificial correlation induced in the default version and, as shown in Figure 8, performed much better than the original GCRMA and almost at par with MAS5.

4 CONCLUSIONS

The use of GCRMA and RMA normalization procedures for Affymetrix GeneChip® technology has received a remarkably broad adoption in the community due to previous benchmarks demonstrating their superiority with respect to other methods. However, while these methods perform well in the assessment of differential expression analysis, we found that they also introduce correlation artifacts in the data. This seriously undermines their utilization, at least in their standard form, upstream of reverse engineering algorithms or any other method relying on the estimate of expression profile correlation. Thus, our results raise issues on the validity of many studies obtained on the basis of correlation measures after these normalization procedures were applied. Specifically we suggest that the implementation of a specific step in GCRMA—the GSB adjustment of truncated values—introduces artificial correlation among the probesets. Unfortunately, according to our analysis, these artifacts are not dataset specific and can survive even after the use of additional probe sets postprocessing filters such as those based on mean, SD and coefficient of variation.

Results were completely consistent across four classes of tests, including (a) a direct assessment of correlation artifacts from replicate and randomized samples, (b) an evaluation of the global topological properties of reverse engineered networks, (c) a study of the functional clustering of correlated genes and (d) a study of the relationship between gene-pair expression profile correlation and membership in stable protein complexes. The unequivocal result is that normalization with GCRMA substantially reduces the ability to distinguish between actual and incorrect functional and physical interactions. In particular, GCRMA is likely to introduce an extraordinary number of false positives, while MAS5 appears to perform optimally with respect to these tests.

We conclude that the choice of normalization procedure strongly affects the correlation structure in the data. Thus, choosing the right normalization procedure is a key step towards the inference of accurate cellular networks. Our comparative analysis favors MAS5 in this context even though (or probably because) it infers fewer interactions but with the highest functional and physical interaction enrichment.

Finally, we suggest that a specific correction to the default implementation of GCRMA in the R package appears to substantially improve its performance, making it competitive with that of MAS5. With this correction, we believe that GCRMA can be properly utilized in the context of reverse engineering gene networks.

ACKNOWLEDGEMENTS

We thank Drs R.Dalla-Favera, K.Basso and U.Klein for sharing the B-cell gene expression profile dataset and Dr M.S.Carro for an insightful discussion. This work was supported by the National Cancer Institute (R01CA109755), the National Institute of Allergy and Infectious Diseases (R01AI066116), and the National Centers for Biomedical Computing NIH Roadmap Initiative (U54CA121852).

Conflict of Interest: none declared.

REFERENCES

Ashburner,M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

Bader,G.D. *et al.* (2003) BIND: the biomolecular interaction network database. *Nucleic Acids Res.*, **31**, 248–250.

Barabasi,A.L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.

Basso,K. *et al*. (2005) Reverse engineering of regulatory networks in human B cells. *Nat*. *Genet*., **37**, 382–390.

Bolstad,B.M. *et al*. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

Butte,A.J. and Kohane,I.S. (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac*. *Symp*. *Biocomput*., **5**, 418–429.

Cope,L.M. *et al*. (2004) A benchmark for Affymetrix GeneChip expression measures. *Bioinformatics*, **20**, 323–331.

Gautier,L. *et al*. (2004) Affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.

Ge,H. *et al*. (2001) Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat*. *Genet*., **29**, 482–486.

Harr,B. and Schlotterer,C. (2006) Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons. *Nucleic Acids Res*., **34**, e8.

Hermjakob,H. *et al*. (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res*., **32**, D452–455.

Hubbell,E. *et al*. (2002) Robust estimators for expression analysis. *Bioinformatics*, **18**, 1585–1592.

Hughes,T.R. *et al*. (2000) Functional discovery via a compendium of expression profiles. *Cell*, **102**, 109–126.

Irizarry,R.A. *et al*. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

Jansen,R. *et al*. (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, **302**, 449–453.

Li,C. and Wong,W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc*. *Natl Acad*. *Sci*. *USA*, **98**, 31–36.

Lipshutz,R.J. *et al*. (1999) High density synthetic oligonucleotide arrays. *Nat*. *Genet*., **21**, 20–24.

Liu,X. *et al*. (2005) A tractable probabilistic model for Affymetrix probe-level analysis across multiple chips. *Bioinformatics*, **21**, 3637–3644.

Margolin,A.A. *et al*. (2006) Reverse engineering cellular networks. *Nat*. *Protocols*, **1**, 662–671.

Peri,S. *et al*. (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res*., **13**, 2363–2371.

Qiu,X. *et al*. (2005) The effects of normalization on the correlation structure of microarray data. *BMC Bioinformat*., **6**, 120.

Roberts,C.J. *et al*. (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science*, **287**, 873–880.

Stuart,J.M. *et al*. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–255.

Tu,Y. *et al*. (2002) Quantitative noise analysis for gene expression microarray experiments. *Proc*. *Natl Acad*. *Sci*. *USA*, **99**, 14031–14036.

Wang,K. *et al*. (2006) Genome-wide discovery of modulators of transcriptional interactions in human B lymphocytes. *Lect*. *Notes Comput*. *Sci*. (RECOMB), **3909**, 348–362.

Wu,Z. *et al*. (2004) A model-based background adjustment for oligonucleotide expression arrays. *J*. *Am*. *Stat*. *Assoc*., **99**, 909–917.

Xenarios,I. *et al*. (2002) DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res*., **30**, 303–305.