Statistical Challenges in Preprocessing in Microarray Experiments in Cancer

Kouros Owzar, William T. Barry, Sin-Ho Jung, Insuk Sohn, and Stephen L. George

Abstract Many clinical studies incorporate genomic experiments to investigate the potential associations between high-dimensional molecular data and clinical outcome. A critical first step in the statistical analyses of these experiments is that the molecular data are preprocessed. This article provides an overview of preprocessing methods, including summary algorithms and quality control metrics for microarrays. Some of the ramifications and effects that preprocessing methods have on the statistical results are illustrated. The discussions are centered around a microarray experiment based on lung cancer tumor samples with survival as the clinical outcome of interest. The procedures that are presented focus on the array platform used in this study. However, many of these issues are more general and are applicable to other instruments for genome-wide investigation. The discussions here will provide insight into the statistical challenges in preprocessing microarrays used in clinical studies of cancer. These challenges should not be viewed as inconsequential nuisances but rather as important issues that need to be addressed so that informed conclusions can be drawn.

In recent years, there has been a surge of genome-wide experiments using high-throughput technologies included as companions to clinical studies in cancer. This situation reflects an increased understanding in the cancer research community of the valuable additional information that can be garnered from these experiments. From a practical standpoint, these technologies have been made practical through a reduction in overall cost and the availability of improved software and hardware computing resources. In addition, cancer researchers can conduct preliminary investigations on publicly available data, and use online facilities for querying annotation and biological pathway information for a better understanding of the findings of genome-wide experiments.

High-throughput technologies enable the investigation of characteristics of the genome such as single nucleotide polymorphisms (1), DNA copy number changes (2), and mRNA expression levels (3). Traditionally, the statistical objective for the experiments using these technologies is the investigation of potential associations between a large number of molecular markers with clinical end points such as tumor response or time to death. The corresponding statistical analyses generally fall into three categories. (*a*) Association studies, the aim of which is the construction of panels of interesting genes (4) or biological pathways (5); (*b*) prognostic or prediction studies, the aim of which is the construction of models based on molecular markers to classify patients with respect to clinical end points; and (*c*) class discovery studies, the aim of which is to discover clusters based on molecular data. Throughout this article, we shall use the term "features" to refer to molecular markers on the array platforms.

On many array platforms, each feature is quantitatively represented by several measures of intensity. To carry out statistical analyses, these intensities need to be adequately preprocessed. This involves reducing a very large set of intensities to a matrix of summary measures so that each feature is quantified using a single representative measure (e.g., expression; ref. 6). The literature regarding statistical methodology is heavily geared towards the development of new and the assessment of existing statistical methods based on the summary measures. These are often referred to as high-level analyses. Likewise, in the cancer research literature, the statistical method sections are mainly devoted to descriptions of high-level analyses with only a token description of the preprocessing methods. In both cases, preprocessing, or low-level analyses, is relegated to the status of a nuisance factor thought to be of little importance. George (7) provides an overview of statistical issues arising in the application of genomics and biomarkers in clinical trials. In this article, we will illustrate some of the challenges and explore the implications of the preprocessing on the conclusions.

We begin by considering an example using a data set originally analyzed and discussed by Beer et al. (8). They conduct an extensive set of analyses for investigating the association between overall survival and features from the Affymetrix hu6800 chip. This data set has been analyzed extensively in the literature, including an article by Jung et al. (9), who conduct an analysis using a rank-covariance estimator, which can be thought of as a robust nonparametric counterpart

Authors' Affiliation: Department of Biostatistics and Bioinformatics, and the Cancer and Leukemia Group B Statistical Center, Duke University Medical Center, Durham, North Carolina

Received 2/15/08; revised 5/13/08; accepted 6/6/08.

Requests for reprints: Kouros Owzar, Department of Biostatistics and Bioinformatics, Duke University Medical Center, 2424 Erwin Road, Suite 802, Room 8031, Durham, NC 27710. Phone: 919-681-1829; E-mail: kouros.owzar@duke.edu.

doi:10.1158/1078-0432.CCR-07-4532

Table 1. The top 10 genes based on an analysis of the Beer et al. (8) data using the method described in Jung et al. (9)

RMA		MAS5		Beer et al.	
Symbol	***P***	**Symbol**	***P***	**Symbol**	***P***
CD8B	0.0697	*RAFTLIN*	0.0245	*RAFTLIN*	0.0187
SLC2A1	0.1270	*TMSB4X*	0.0465	*NP*	0.0993
CCR2	0.2111	*SLC2A1*	0.0559	*KLHDC3*	0.2968
PLD3	0.2224	*IHPK1*	0.3312	*TMSB4X*	0.3808
RAFTLIN	0.2433	*MLL*	0.3414	*CXCL3*	0.4084
HNRPL	0.2787	*NP*	0.3492	*SELP*	0.4441
BCL2	0.3106	*PRKACB*	0.4494	*STX1A*	0.5026
PFKP	0.3223	*<NA>*	0.4787	*SEC31L1*	0.5068
STX1A	0.3610	*E2F4*	0.5528	*PRKACB*	0.5355
INPP5D	0.3690	*P2RX5*	0.5846	*PBXIP1*	0.5571

Note: *P* values refer to the family-wise error-adjusted rates.

of univariate Cox regression, to identify features associated with survival. The top 10 features according to this analysis method, ranked according to the family-wise error rate–adjusted *P* values (10), are listed in Table 1 using summary measures obtained from three different preprocessing methods: robust multichip algorithm (RMA; refs. 11, 12), MAS5 (13), and a method used by Beer et al. (described in the supplementary document for ref. 8). At the 10% level, there is one significant feature (*CD8B*) based on the RMA method, three significant features (*RAFTLIN*, *TMSB4X*, and *SLC2A1*) based on the MAS5 method, and two significant features (*RAFTLIN* and *NP*) based on the Beer et al. method. Often, features are excluded based on nonphenotypic criteria during the preprocessing method. For this illustration, we employed the filter used by Beer et al. The results are also sensitive to the choice of the filter.

In the discussions to follow, we will focus our attention on illustrating some of the challenges regarding preprocessing within the framework of the Beer et al. example data. Although the discussions will focus on Affymetrix RNA arrays, as a consequence of choosing this example, many of the concepts apply to other types of microarrays.

Preprocessing: From Image to Measure

To carry out high-level statistical analyses, raw imaging data are first quantified as intensities in the hybridization of sample to probe. The data then go through a series of preprocessing steps to generate a summary measure for each feature. These steps consist of some or all of the following:

- *Background correction.* For DNA-based arrays, it is important to apply adjustments for background noise that can result from nonspecific hybridization, incomplete washing of the slide, or other technical artifacts in the generation of scanned images. These corrections are done at the probe level to remove spatial effects within each chip.
- *Normalization.* As with many other lab measurements, the collection of intensities must be globally standardized such that features are comparable across all chips.
- *Summary measure calculation.* When an array platform contains several probes for each feature, a summary measure must be obtained in order to quantify the amount of RNA expression, the change in DNA copy number, or to call the genotype.
- *Filtering.* Some features should be excluded from the association studies. For example, features such as housekeeping or control genes are for quality control purposes and should be excluded from high-level analyses. Filtering is also often used to reduce the number of features in the final analyses by removing features, which, for example, have relatively low variability across the samples.

The Affymetrix oligonucleotide array used by Beer et al. is a common platform for measuring mRNA expression levels. Affymetrix arrays are comprised of short sequences (25 bp in length) that are synthesized directly to glass slides using a photolithographic process. This technique can produce high-density chips with hundreds of thousands of unique oligomers. This allows multiple probes, collectively termed a "probe set," to represent a single feature on the array. A probe set typically consists of anywhere from 5 to 20 probe pairs that correspond to distinct sequences within the transcript. Each probe pair consists of a "perfect match" probe and a "mismatch" probe in which the nucleotide in the 13th position is switched.

Preprocessing of Affymetrix arrays commonly involves generating a summary measure for each probe set. Affymetrix has released a series of algorithms [MAS4.0, MAS5.0 (ref. 13), and PLIER (ref. 14)], that quantify expression from increased binding to perfect match over mismatch probes. However, there is considerable debate as to whether mismatch probes detect only nonspecific hybridization, and alternative algorithms have been proposed by academic investigators. For example, the model-based expression index (MBEI) proposed by Li and Wong (15, 16), uses parametric multiplicative models for probe-specific rates of hybridization. This is defined for perfect match intensities only, the difference in perfect match and mismatch, or both. RMA, proposed by Irizarry et al. (11, 12), uses parametric background correction followed by quantile normalization and robust fitting of a log-linear additive model based on perfect match only. GeneChip RMA (17) extends the RMA algorithm to use probe sequence information in estimating nonspecific hybridization during background correction. Numerous other algorithms have been employed in the literature including the method used in Beer et al. (8).

With the increasing number of preprocessing methods, control experiments have been done and made available as benchmarks for evaluating relative performance. This includes dilution and mixture experiments (18), and spike-in experiments as well as the Affymetrix Latin square data sets. In order to develop a standardized approach for comparing the various methodologies, the Affycomp project (19, 20) has provided an application for each algorithm to the Affymetrix Latin square data sets. The relative performances of these methods are

assessed using a number of metrics for quantifying accuracy and precision, reflecting a bias-variance tradeoff among these methods.

Quality Control and Outlier Detection

Quality control at the array level is a critical step in detecting potential outliers and batch differences. Here, we describe two different approaches. Plots are used to visualize aberrant hybridization patterns and to display poor correlation within the set(s) of arrays under investigation. Also, the quality of the arrays can be quantified using heuristic measures. In many cancer studies, replicate arrays cannot be run for defective arrays due to either cost constraints or a lack of additional biospecimens. At the same time, poor-quality arrays should not have undue influence on differential expression or classification algorithms. Therefore, one must be cautious in determining the level of stringency for excluding arrays from the analyses.

Plots of density estimates and principal components (21) are useful means of visualizing the data at both the probe intensity and transcript levels. In an extensive quality assessment analysis of the Beer et al. (8) microarray data set, we have used these graphical devices to identify a batch effect and two sets of outliers in the data and their impact on several popular preprocessing algorithms. Some of the findings of this investigation are shown in Fig. 1 (*red* and *blue*, batch effects; *purple* and *green*, two sets of outliers). The density estimates of the perfect match intensities are drawn for each array (*top*). A distinct difference in the distribution of the intensities may be noted between the two batches of arrays (*red* and *blue*). Likewise, the outlier arrays (*purple* and *green*) have intensity profiles that are much brighter than the others (*red* or *blue*). In order to assess whether preprocessing would remedy these global effects, a plot of the first two principal components was produced. The results based on RMA preprocessing is shown in Fig. 1 (*bottom left*). The four clusters observed at the probe level are still present despite preprocessing. Next, the influence of the outlier arrays is examined by their removal from the RMA preprocessing procedure. The corresponding principal components plot is shown in Fig. 1 (*bottom right*). The segregation of the two batches has vanished, suggesting that the presence of these outliers prevents RMA from correcting the observed global difference.

A visual examination, provided by the Bioconductor package affyPLM, of one of the arrays drawn in purple is shown in Fig. 2. A raw image of the chip is also shown (*top left*). The rest of the figure plots various types of residuals obtained after subtracting off a probe-level linear model [additional details for this method are found in section 3.5 of Bolstad et al. (22)]. A distinct spatial artifact is visible in the middle of this chip and the other three purple arrays (images not shown).

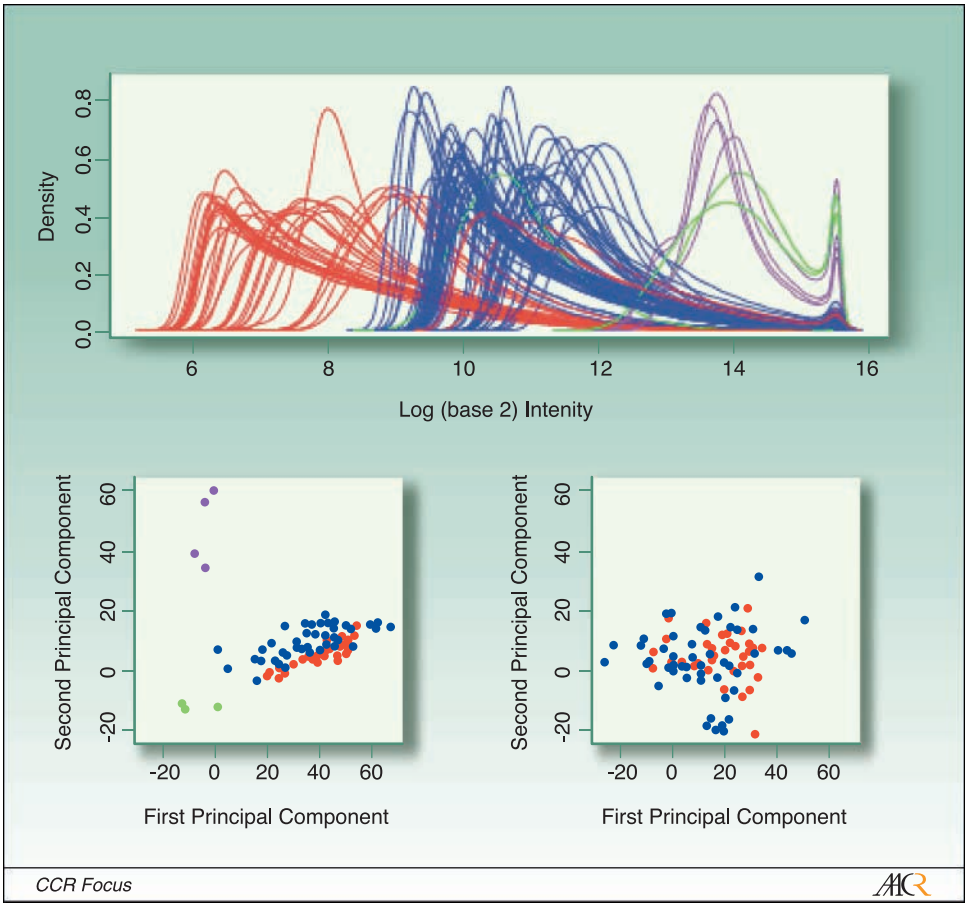

Fig. 1. *Top row*, density estimates of the probe intensities for each of the 96 CEL files from Beer et al. (8). The PCA plots of expression values obtained when all arrays are RMA preprocessed (*bottom left*) and when seven outlier arrays are removed (*bottom right*).

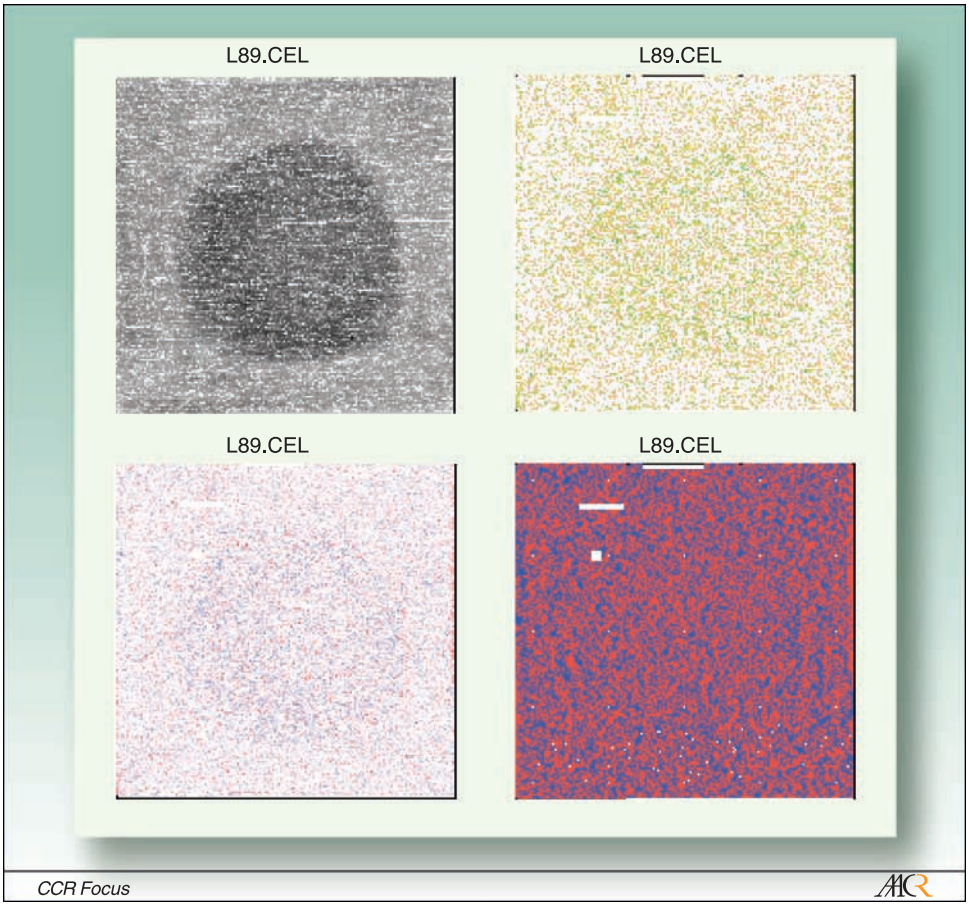

Fig. 2. *Top left*, raw image of the (*purple*) outlier chips. The rest of the figure plots various types of residuals obtained after subtracting off a probe-level linear model.

Summary measures of quality control are provided in Fig. 3 for the outlier arrays. Several common measures in a graphical display supplied by the Bioconductor package simpleaffy are provided (Fig. 3A). These include (*a*) the percentage of probe sets with intensities above background, or "present" as defined by Affymetrix (23); (*b*) the average background intensity of the array; (*c*) the scale factor and an acceptable range displayed as a shaded interval on the log 2-scale; (*d*) 3′/5′ ratio glyceraldehyde-3-phosphate dehydrogenase; and (*e*) 3′/5′ ratio of β-actin. For each measure, values that exceed typical ranges of acceptability are highlighted in red. These ratios are commonly used measures of sample quality, in which elevated levels indicate the integrity of starting RNA, efficiency of cDNA synthesis, and/or transcription of cRNA in running arrays. Thus, global patterns of RNA degradation can also be plotted using functions supplied be the affy package from Bioconductor (Fig. 3B), in which the average probe intensity of all probe sets are ordered from the 5′ to 3′ end. This plotting function is supplied by the affy package from Bioconductor as a means of evaluating global RNA degradation patterns in the samples. These results show that the outlier samples were not immediately detectable from summary measures of quality controls alone.

It is important to mention that the point of this discussion is not to suggest the removal of any array that seems to be an outlier from the statistical analyses. Rather, we set out to summarize the effect of a set of outliers on the performance of a popular preprocessing method for this specific data set. Removal of the outliers seemingly improves the performance of this preprocessing method. However, the removal of an outlier array would remove a patient from the statistical analyses, which may result in bias. More specifically, unless the array is deemed to be technically defective beyond a reasonable doubt, its removal from the analysis is not something that can be recommended. This emphasizes the importance of using a statistical methodology that is robust with respect to outliers.

Challenges in the Prospective Setting

As microarrays and other high-throughput biotechnologies are increasingly used in the study of cancer therapeutics, a particular interest has been the identification of genomic signatures that classify tumor subtypes according to clinical outcome. Retrospective analyses of tumor samples have generated genomic signatures for many types of cancer, including lymphoma (24), breast (25), and lung carcinomas (26). Clinical trials are required to evaluate and properly validate the prognostic or predictive capability of signatures, and much effort has gone into developing design strategies for testing and validating genomics in a prospective manner (27–30). However, there has been little discussion on the

ramifications of preprocessing algorithms used in the development of each signature in the prospective setting.

Many of the common preprocessing and quality control strategies presented above are designed to simultaneously investigate the full set of samples in a study. This includes the model-based approaches to summarizing expression at the transcript level (e.g., RMA, MBEI, and PLIER), quality control plots for visualizing outliers and batch effects, and the quality control summary measures for relative scale factor and average background. All preprocessing could be limited to single-array methods. However, the relative performance of these methods in Affycomp suggests that this would result in decreased precision and accuracy. Adaptations to the summarization approaches of MBEI and RMA have been proposed in which probe-level variables are first estimated by a training set, and then applied to the incoming sample as fixed effects (16, 31). As an alternative approach, one uses a set of "standardization" samples that are selected prior to the initiation of the trial; then, each microarray collected over the course of the trial is preprocessed in conjunction with the set. In this way, the incoming sample informs the probe-level variable estimates, yet the principle of exchangeability is maintained. Whether a training or standardizing set is used in generating postprocessed data, it is critically important that the samples are representative of the patient population. However, the adequacy of the set must be determined by the investigators, and will depend on the experimental design such that universal standards have not been identified in the field. The following simulations represent one mechanism whereby the quality of postprocessed data can be assessed once a standardization set is in place.

One important determination in selecting a standardizing set is the minimum necessary number of arrays. We conducted a comprehensive bootstrap analysis of the Beer et al. data set to evaluate the sensitivity of postprocessed data to set size. For each array, standardizing sets of n = 5, 10, 15, 20, 25, or 30 were randomly selected, and the RMA preprocessing algorithm was applied. Probe set–specific variances in expression were computed from 200 bootstrap replicates, and then averaged across all arrays. Box plots in Fig. 4 show that variability is substantially reduced when the standardizing set consists of at least 20 arrays. Furthermore, variances are attenuated when one or both sets of outlier samples are removed, and variance stabilization seems to be achieved when all seven outlier samples are removed (*bottom right*). These results illustrate the sensitivity of RMA to standardizing set size and outlier arrays. Further analyses are required using data sets with technical replicates and spike-in genes to evaluate precision and accuracy.

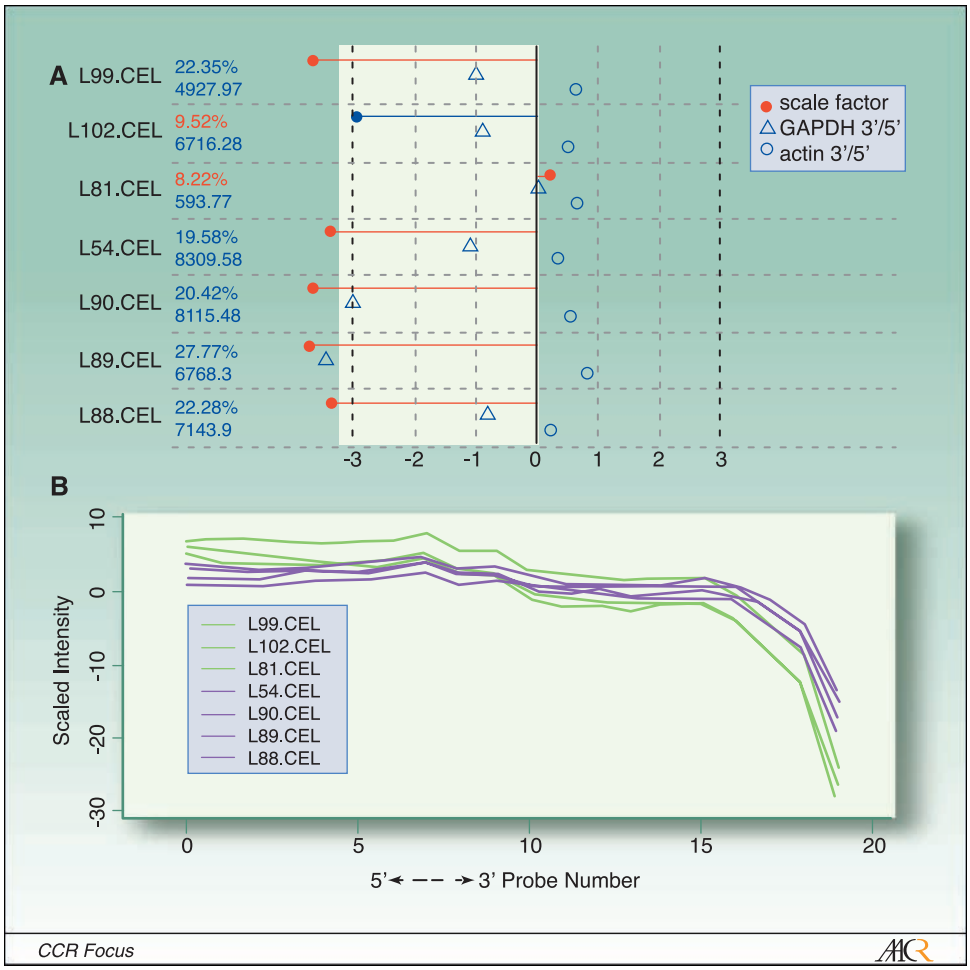

Fig. 3. Graphical representations of summary measures for array quality. *A,* output from the quality control reports generated by Bioconductor/simpleaffy for the six outlying arrays identified in Fig. 1. The percentage present and average background are printed, and the scale factor, β-actin 3′/5′ ratio, and glyceraldehyde-3-phosphate dehydrogenase 3′/5′ ratio are plotted on the log 2-scale. Red, values that cross typical thresholds. *B,* RNA degradation plots from Bioconductor/affy. For each transcript, probe pairs are ordered from 5′ to 3′, and the average position-specific perfect match value is plotted for each array to indicate any global patterns of sample degradation.

Computational Tools

Preprocessing of microarray data is a computationally intensive task requiring access to appropriate computing hardware and software. There are a number of commercial and open-source products that can be used to carry out the preprocessing steps presented in this report.

The analyses presented using the example data set were done using the open-source R (32) statistical environment along with packages from Bioconductor (33). The affy package provides functions for MAS5 and RMA preprocessing and GeneChip RMA and PLIER packages provide functions for GeneChip RMA and PLIER preprocessing. The expresso function in affy allows the user to mix and match from a set of predefined background correction and normalization methods. The reduced models in MBEI have been implemented in the freely available software dCHIP.

Discussion

In this article, we have reviewed several preprocessing and quality control methods and applied them to an example data set microarray in lung cancer. For a detailed and accessible discussion on various aspects of preprocessing, the reader may refer to the monograph by Simon et al. (34) and the articles by Quakenbush (35), Hoffmann et al. (36), McClintick et al. (37), McClintick and Edenberg (38), Jones et al. (39), and Seo and Hoffman (40). The results in our article indicate the presence of outlier arrays and batch effects in the data. A number of postprocessing methods have been proposed to address these issues. Suárez-Fariñas et al. (41) propose a corrective method for removing spatial artifacts. For batch effects correction, Johnson et al. (42) propose an empirical Bayes method, and Benito et al. (43) propose a method using support vector machines.

If a study in development plans to use information from a cancer gene list of signatures constructed based on an older chip, then one has to decide how to map the features from the old chip to the newer chip. The mapping may be done by matching on gene symbols or homologues queried from public databases. However, one should be concerned that the marginal or joint distributions of the intensities may differ between the platforms, and consequently, the summary measures may not be comparable. More importantly, besides these potential statistical caveats, the matching may not be biologically relevant if the probe sequences for the features on the two chips do not match.

Single nucleotide polymorphism arrays are used to investigate DNA polymorphisms. The fact that the final outcome is not a continuous expression measure but rather a number of copies of an allele for each feature, may give the erroneous impression that preprocessing of single nucleotide polymorphism arrays is more straightforward than that of, e.g., RNA microarrays. The genotypes are not determined but rather

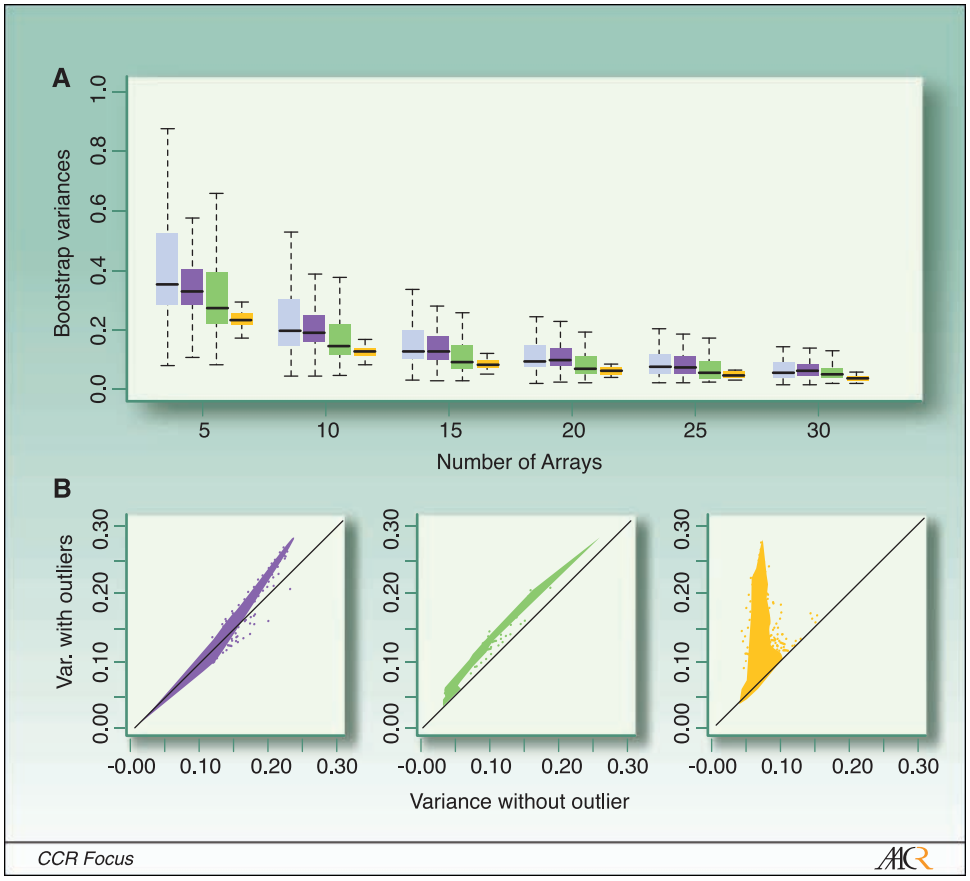

Fig. 4. Bootstrap variance estimates for expression values generated from RMA. *A,* box plots of the average variance in expression across all arrays when preprocessed with 200 random standardizing sets of size $n = 5, 10, 15, 20, 25, 30$. Arrays were either selected from the full set of samples from Beer et al. (8), after removing outlier set 1 (*purple*), outlier set 2 (*green*), or both (*yellow*). *B,* scatter plots of bootstrap variance estimates from standardization sets of size $n = 20$ with or without removing outlier samples.

called based on intensities. The issues raised related to preprocessing for the RNA microarrays applies to these instruments as well.

In some cancer studies, multiple chips are produced for some of the patients. These could be replicates generated for quality control and reproducibility. Unless there is overwhelming and definitive evidence that a replicate chip is defective, it is likely to be inappropriate to exclude it from the analysis. As such, in the case of replicate arrays, it may be necessary to aggregate the arrays. One simple approach is to average the arrays for each patient across the features.

The results discussed regarding the outliers in the Beer et al. data set are part of a more extensive study which we have carried out. One of the conclusions from this study is that the most influential aspect of outlier effects is the method used for background correction. This conclusion agrees with that of the Affycomp report (20).

As microarray experiments are increasingly used in cancer trials, the MicroArray Quality Control project was formed between the Food and Drug Administration and academic institutions to provide quality control tools (44). In the first phase of the MicroArray Quality Control project, the relative performance of different array platforms was assessed using several large sets of technical replicates that were run across multiple sites and determined to be comparable (45). Future efforts of the MicroArray Quality Control are to examine diagnostics of array reliability and to explore the utility of microarray technology in the development and validation of predictive models.

In summary, we outline a list of general recommendations.

- The examples discussed illustrate that it is difficult to assess the quality of the data solely based on summary measures. For any study, the investigators should be provided the files containing the raw data (e.g., Affymetrix *.CEL, Illumina *.idat or aCGH *.sproc files) rather than a spreadsheet with expressions.
- Standardized quantitative quality control measures, such as those provided by the chip manufacturer, are useful and should be considered as part of the preprocessing package. These are, however, not substitutes for graphical tools such those considered in this report.
- Often the physical file names of the arrays reveals experimental factors such as treatment assignment or cell line type. The lab generating the arrays should be blinded to the experimental factors to avoid unintentional induction of batch effects.
- The lab should be asked to provide information (e.g., date or time) that can be used in the identification of potential batch effects.
- In the case of posttreatment arrays, investigators should avoid confounding batch and experimental factors by not sending the specimens from each group of the factor in batches to the lab.
- As Chau et al. (6) point out, sample processing can affect the quality of the biospecimens. Consequently, batch effects may be introduced at the institution obtaining the biospecimens as well as at the repository responsible for receiving, storing, and processing the biospecimens for shipment to the microarray lab. As such, it is important for the investigators to understand the flow of the biospecimens.
- The preprocessing steps should be reproducible. For R and Bioconductor users, the Sweave tool (46) provides the facilities to simultaneously carry out and document the entire preprocessing procedure by intertwining R with the typesetting system LaTeX (47). The resulting document, which can be submitted as supplementary material, will also facilitate the article review process. More importantly, this document will provide an important quality control component of the study.
- Arrays could be generated based on various types of biospecimens such as frozen tumor tissue, paraffin-embedded tumor tissue, or cancer cell lines. Care should be taken before jointly preprocessing arrays based on different types of biospecimens. Many preprocessing algorithms require the provision of input variables or thresholds. The defaults may not be appropriate for all tissues.

We have provided a glimpse of some of the basic challenges investigators face during the preprocessing phase of high-dimensional molecular data in cancer studies. It is inappropriate to designate these challenges as unimportant or inconsequential nuisances, especially considering the potential ethical ramifications of using models based on these data to assign treatment in a prospective manner. These challenges should be welcomed as an opportunity for additional research on the development of improved methodologies and to pave the way for better understanding of the results from microarray experiments.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank the two reviewers for providing insightful and helpful comments leading to substantial improvements of the manuscript.

References

1. Mei R, Galipeau PC, Prass C, et al. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. Genome Res 2000;10:1126–37.
2. Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 1999;23:41–6.
3. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. Science 1995;270:467–470.
4. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2000;98:5116–21.
5. Barry WT, Nobel AB, Wright FA. Significance analysis of functional categories in gene expression studies: a structured permutation approach. Bioinformatics 2005;21:1943–9.
6. Chau CH, Rixe O, McLeod H, Figg WD. Validation of analytical methods for biomarkers employed in drug development. Clin Cancer Res 2008;18:5967–76.

7. George SL. Statistical issues in translational cancer research. Clin Cancer Res 2008;18:5954–8.
8. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816–24.
9. Jung SH, Owzar K, George SL. A multiple testing procedure to associate gene expression levels with survival. Stat Med 2005;24:3077–88.
10. Westfall PH, Young SS: Resampling-based multiple testing: examples and methods for *P*-value adjustment. Wiley Series in Probability & Mathematical Statistics: Applied Probability & Statistics. John Wiley & Sons; 1992.
11. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 2003;31:e15.
12. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249–64.
13. Hubbell E, Liu WM, Mei R. Robust estimators for expression analysis. Bioinformatics 2002;18:1585–92.
14. Hubbell E. PLIER: an M-estimator for expression array. Affymetrix Inc. 2005, Santa Clara, CA.
15. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol 2001;2:1–11.
16. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A 2001;98:31–6.
17. Wu Z, Irizarry RA. Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J Comput Biol 2005;12:882–93.
18. Lemon WJ, Palatini JJT, Krahe R, Wright FA. Theoretical and experimental comparisons of gene expression indexes for oligonucleotide arrays. Bioinformatics 2002;18:1470–6.
19. Cope LM, Irizarry RA, Jaffee HA, Wu Z, Speed TP. A benchmark for Affymetrix GeneChip expression measures. Bioinformatics 2004;20:323–31.
20. Irizarry RA, Wu Z, Jaffee HA. Comparison of Affymetrix GeneChip expression measures. Bioinformatics 2006;22:789–94.
21. Mardia KV, Kent JT, and Bibby JM. Multivariate analysis. Academic Press 1979.
22. Bolstad BM, Irizarry R, Gautier L, Wu Z. Preprocessing high-density oligonucleotide arrays. In: Gentleman RC, Carey VJ, Huber W, Irizarry R, Dudoit S, editors. Bioinformatics and computational biology solutions using R and Bioconductor (Statistics for Biology and Health). Springer-Verlag; 2005.
23. Affymetrix: statistical algorithms description document (white paper). Affymetrix Inc., Santa Clara, CA, 2002.
24. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 2002;8:68–74.
25. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
26. Hayes DN, Monti S, Parmigiani G, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol 2006;24:5079–90.
27. Simon R. Using genomics in clinical trial design. Clin Cancer Res 2008;18:5984–94.
28. Taylor JMG, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. Clin Cancer Res 2008;18;5977–83.
29. Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. Clin Cancer Res 2005;11:7872–8.
30. Simon R, Wang SJ. Use of genomic signatures in therapeutics development in oncology and other diseases. Pharmacogenomics J 2006;6:166–73.
31. Katz S, Irizarry RA, Lin X, Tripputi M, Porter MW. A summarization approach for Affymetrix GeneChip data using a reference training set from a large, biologically diverse database. BMC Bioinformatics 2006;7.
32. R Development Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
33. Gentleman R, Carey V, Bates D, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
34. Simon R, Korn EL, McShane LM, Radmacher MD, Wright GW, Zhao Y. Design and analysis of DNA microarray investigations. Springer-Verlag; 2004.
35. Quackenbush J. Microarray data normalization and transformation. Nat Genet 2002;32 Suppl:496–501.
36. Hoffmann R, Seidl T, Dugas M. Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. Genome Biol 2002;3.
37. McClintick JN, Jerome RE, Nicholson CR, et al. Edenberg HJ. Reproducibility of oligonucleotide arrays using small samples. BMC Genomics 2003;4:4.
38. McClintick JN, Edenberg HJ. Effects of filtering by present call on analysis of microarray experiments. BMC Bioinformatics 2006;7:49.
39. Jones L, Goldstein DR, Hughes G, et al. Assessment of the relationship between pre-chip and post-chip quality measures for Affymetrix GeneChip expression data. BMC Bioinformatics 2006;7:211.
40. Seo J, Hoffman EP. Probe set algorithms: is there a rational best bet? BMC Bioinformatics 2006;7:395.
41. Suárez-Fariñas M, Pellegrino M, Wittkowski K, et al. Harshlight: a corrective make-up program for microarray chips. BMC Bioinformatics 2006;6:294.
42. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118–27.
43. Benito M, Parker J, Du Q, et al. Adjustment of systematic microarray data biases. Bioinformatics 2004;20:105–14.
44. Shi LM, Reid LH, Jones WD, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 2006;24:1151–61.
45. Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, et al. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. Nat Biotechnol 2006;24:1140–50.
46. Leisch F. Sweave: dynamic generation of statistical reports using literate data analysis. In: Härdle W, Rönz B, editors. Compstat 2002—Proceedings in Computational Statistics, pages 575–580. Physica Verlag, Heidelberg, 2002. ISBN 3–7908–1517–9.
47. Lamport L. LaTeX: a document preparation system. 2nd ed. Addison-Wesley. 1994.